

# QUANTIFYING BAYESIAN FILTER PERFORMANCE FOR TURBULENT DYNAMICAL SYSTEMS THROUGH INFORMATION THEORY\*

M. BRANICKI<sup>†</sup> AND A.J. MAJDA<sup>‡</sup>

**Abstract.** Incomplete knowledge of the true dynamics and its partial observability pose a notoriously difficult problem in many scientific applications which require predictions of high-dimensional dynamical systems with instabilities and energy fluxes across a wide range of scales. In such cases assimilation of real data into the modeled dynamics is necessary for mitigating model error and for improving the stability and predictive skill of imperfect models. However, the practically implementable data assimilation/filtering strategies are also imperfect and not optimal due to the formidably complex nature of the underlying dynamics. Here, the connections between information theory and the filtering problem are exploited in order to establish bounds on the filter error statistics, and to systematically study the statistical accuracy of various Kalman filters with model error for estimating the dynamics of spatially extended, partially observed turbulent systems. The effects of model error on filter stability and accuracy in this high-dimensional setting are analyzed through appropriate information measures which naturally extend the common path-wise estimates of filter performance, like the mean-square error or pattern correlation, to the statistical superensemble setting that involves all possible initial conditions and all realizations of noisy observations of the truth signal. Particular emphasis is on the notion of practically achievable filter skill which requires trade-offs between different facets of filter performance; a new information criterion is introduced in this context. This information-theoretic framework for assessment of filter performance has natural generalizations to Kalman filtering with non-Gaussian statistically exactly solvable forecast models. Here, this approach is utilized to study the performance of imperfect, reduced-order filters involving Gaussian forecast models which use various spatio-temporal discretizations to approximate the dynamics of the stochastically forced advection-diffusion equation; important examples in this configuration include effects of biases due to model error in the filter estimates for the mean dynamics which are quantified through appropriate information measures.

**Key words.** Kalman filter, information theory, uncertainty quantification, turbulent dynamical systems.

**AMS subject classifications.** 93E11, 62B10, 94A17, 60G35, 60H10, 93E10, 60H30, 35C20, 76F55.

## 1. Introduction

This paper begins the analysis of the performance of Kalman filters with model error for estimating the discretized dynamics of spatially extended turbulent systems through the methods of information theory (e.g., [68, 29, 20]). Here, the term *filter* refers to a sequential Bayesian procedure which aims to provide optimal statistical estimates of the true state of high-dimensional turbulent dynamical system based on partial noisy observations and an imperfect model. Appropriate use of concepts from Shannon's information theory (e.g., [68, 29, 20]), which are widely applied in communication theory and coding theory, provides a natural framework for assessing the statistical accuracy of imperfect filters in this notoriously difficult setting with many degrees of freedom and a large number of positive Lyapunov exponents. When filtering linear Gaussian dynamics with no model error and subject to linear Gaussian observations, the classical Kalman filter [43, 44] provides the optimal, minimum error variance

---

\*Received: March 2, 2013; accepted (in revised form): June 29, 2013. Communicated by Shi Jin.

<sup>†</sup>Corresponding author, Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, USA (branicki@cims.nyu.edu).

<sup>‡</sup>Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, USA.

solution to such an estimation problem. Clearly, the ‘perfect model’ and Gaussianity assumptions are hardly ever justified in practice and a successful application of the Kalman filter framework usually relies on the ability to design an imperfect filter which, in spite of being suboptimal, performs ‘adequately’ and is computationally cheap. The strategy behind the design of such imperfect filtering techniques in geophysical and engineering applications is a purely pragmatic one since it currently offers the only realistic option that is capable of beating the ‘curse of dimension’ [6, 9, 69] and/or the ‘curse of ensemble size’ [40, 61] for real time filtering of sparsely observed spatially extended turbulent systems. For example, in the coupled atmosphere-ocean system, the current operational models for prediction of both weather and climate involve general circulation dynamics which are unavoidably coarse-grained and discretized in space and time with the multitude of unresolved processes and often ad-hoc parametrizations of the turbulent backscatter of the unresolved fluxes impacting the resolved scales. The resulting models for the prediction of weather and climate are high-dimensional chaotic dynamical systems with many positive Lyapunov exponents which need to be appropriately constrained by the observations for their stability and predictive skill. Designing statistically accurate imperfect Kalman filters for complex high-dimensional systems is difficult and, while the design process may sometimes be guided by physical intuition, it invariably involves a considerable amount of costly trial and error. The present information-theoretic framework is designed to provide systematic yet practical guidelines for assessing the statistical skill of imperfect high-dimensional Kalman filters and for improving their performance; rigorous treatment of filter error using related techniques will follow soon.

The concept of imperfect filtering has been omnipresent in engineering and geoscience applications. Bayesian hierarchical modeling [8] and reduced order filtering strategies [62, 3, 4, 16, 25, 66] drawing from the classical Kalman filter in the physical space have been developed with some success in these extremely complex systems. An alternative radical strategy, exploiting imperfect Kalman filters in the Fourier domain and capable of avoiding the ‘curse of ensemble size’, was proposed and studied in [59, 15, 37, 38, 60] for various reduced-order forecast models with judicious model error; these techniques were later used to efficiently filter sparsely observed geophysical flows in [38, 36]. An extension of the latter framework to Fourier domain filters utilizing non-Gaussian exactly solvable stochastic forecast models was introduced in [31, 30, 61, 34, 60]; these cheap stochastic filters, employing ‘on the fly’ bias estimation for reducing judicious model error, were shown to offer a computationally cheap and skillful alternative for filtering turbulent systems with many spatio-temporal scales and intermittent instabilities in the unresolved dynamics. The common feature of all the aforementioned filters is the use of imperfect forecast models and simple Kalman estimates for the posterior statistics for filtering the high-dimensional turbulent systems from sparse observations. As shown in [37, 60, 15, 59], several important mathematical problems arise in the practical application of these imperfect filtering strategies to complex spatially extended systems. For example, many subtle issues associated with violation of the classical observability and controllability criteria (e.g., [2, 60]) occur in Kalman filtering even in the relatively simple setting that involves imperfect linear Gaussian forecast models. Moreover, additional issues affecting the filter accuracy and stability arise due to the aliasing and correlations in the imperfect filter estimates [12, 37, 60] when only sparse observations are available; in high-dimensional filtering these effects are difficult to capture and quantify in the traditional path-wise framework which is commonly used in applica-

tions. We show below that the use of concepts and techniques rooted in information theory allows for a systematic assessment of the error in imperfect Kalman filters in the statistical superensemble setting; this new framework incorporates the ensemble-averaged effects of intermittent interactions between the mean state and fluctuations, which is important in accurate assessment of filter error in high-dimensional turbulent systems. On the other hand, the path-wise approach is more suitable for studying the filter skill in the presence of ‘rare event’-type phenomena in the turbulent dynamics which are marginalized in the ensemble-averaged skill measures. Thus, the information-theoretic framework proposed here for assessing the filter error is complementary to the standard path-wise approach and, in general, a synergistic use of the two methods is needed.

The information-theoretic approach to filtering has received surprisingly little attention in the literature, especially compared to various applications in control and dynamical systems. The scarce few contributions to this subject that we were able to trace over the past 50 years are the works of Bucy & Joseph [14], Kailath [42], Weidemann & Stear [71], Duncan [24], Zakai & Ziv [72], Tomita et al. [70], Galdos & Gustafson [28] and Newton [64, 65, 63]. Following the prevalent mathematical trend in stochastic filtering theory, most of these works, except [72, 28], did not consider model error in the filter and sought optimal information estimates on the filter performance without addressing the practically important issue of suboptimal but achievable estimates. Moreover, these works approached the problem in the spirit of the coding theory through seeking optimal estimates on information transfer; an important limitation of this approach in the context of filtering is that the coding theory framework allows for controlling both the input signal and the decoder/filter in order to obtain optimal data transfers. In [42, 24, 70] various bounds for the mutual information between the truth and its noisy observations were derived in terms of the optimal least-squares estimates. More recently, the statistical mechanical properties of Kalman filters were discussed in [63] and in [64, 65] for linear and nonlinear filters; in those papers, the signal-filter pair is described by a dissipative system with rates of information supply and dissipation, and with information flow from the observation process into the filter state.

An important theme present throughout this paper is that of practical optimality of imperfect Kalman filters in terms of best achievable filter skill given structural constraints in the filter and the computational cost. Imbedding the filtering problem in the appropriate information-theoretic framework allows for a systematic assessment of these issues and provides guidelines for filter design. It is well known [14, 29, 70] for the classical Kalman filter with no model error that minimizing the path-wise error variance also maximizes the pattern correlation between the truth and its filter estimate. However, these globally optimal imperfect filters may be still impossible to implement due to practical constraints on the forecast model dynamics or due to an unrealistic computational cost; information barriers arising from such structural constraints in modeling the dynamics are commonplace in geoscience applications (e.g., [58]). Consequently, trade-offs between the relevant facets of filter skill are often needed in high-dimensional imperfect filtering and a new information criterion for practically achievable filter optimality is introduced. We begin the analysis of performance and stability for Kalman filters by considering the linear Gaussian dynamics of the truth and the forecast models. Extensions of this promising framework to study the accuracy of Kalman filters with non-Gaussian exactly solvable forecast models [60, 61] will be reported soon elsewhere.

This paper is structured as follows: We start by describing the building blocks of the information-theoretic framework for assessing the filter performance in Section 2. First, the augmented system of equations for the evolution of the Kalman filter with model error is derived in Section 2.1. The information measures of filter skill are introduced and discussed in Section 2.2; the important link between the asymptotic path-wise measures of filter skill and their statistical ensemble counterparts is outlined in Section 2.2.2, while the practically important bounds on the filter error statistics are presented in Section 2.3. In sections 3 and 4 the skill of various reduced Kalman filter algorithms is discussed for estimating the spatially extended dynamics; in Section 3 the case of filtering the turbulent system with observations of all resolved modes is analyzed while Section 4 deals with the more difficult problem of sparse observations when high wave number information is aliased into the observed waveband. The canonical test model for filtering the turbulent dynamics in the Fourier domain, based on the stochastically driven advection-diffusion dynamics of a turbulent tracer, is described in Section 3.1 and important differences between filtering with plentiful and spatially sparse observations are outlined at the beginning of Section 4. We close in Section 5 by summarizing the main results and remarking on future developments.

## **2. General information-theoretic framework for quantifying performance of Kalman filters with model error**

The goal here is to develop a framework, rooted in information theory, for quantifying the skill of imperfect Kalman filters for estimating the dynamics of spatially extended and sparsely observed turbulent systems. There are a number of distinct features in this setup which make this topical problem both mathematically challenging and important for practical applications. First, in contrast to classical filtering theory (e.g., [49, 5, 41, 2, 17]), we allow for the presence of model error in the forecast model of the filter which introduces additional subtleties into the filtering problem even in the linear Gaussian setting. Thus, it is crucial to develop a systematic framework capable of quantifying the effects of model error and providing bounds on the filter performance in the presence of model error. While rigorous analysis of the imperfect filter stability and accuracy is still in its infancy, the framework developed below provides the means for analyzing the skill of imperfect Kalman filters and leads to new results and insight, complementing the earlier path-wise analysis of [59, 15, 37, 61, 60]. We begin this line of research by constructing information-theoretic tools for quantifying the skill of imperfect Kalman filtering of a linear Gaussian constant-coefficient PDE; this setup is similar in spirit to the previous work of Majda and collaborators [59, 15, 37, 61, 60] and is motivated by the classical linear stability tests for finite difference approximations in systems of nonlinear partial differential equations. Akin to the classical stability analysis of finite difference schemes, this approach allows detailed Fourier analysis of systems of Langevin equations and provides guidelines for filtering turbulent systems with both plentiful and sparse observations which are gradually introduced in the subsequent sections.

The key new formulation of the Kalman filtering problem is presented in Section 2.1 where the augmented linear Gaussian system is derived for the coupled evolution of the truth and its Kalman filter estimate. We then introduce a suite of information measures of the filter error in Section 2.2 which take advantage of the statistical ensemble framework in formulating the filtering problem. The estimates of the statistics of the filter error are derived in Section 2.3 and an information-based approach to optimizing imperfect Kalman filters is discussed in Section 2.4. While the derivations below are carried out for Kalman filtering with linear Gaussian forecast models,

extensions of this framework are possible to address similar issues in linear Bayesian filtering with nonlinear non-Gaussian statistically exactly solvable forecast models.

**2.1. Augmented dynamical system for the Kalman filter.** A prerequisite for the study of the statistical skill of a filtering algorithm is the derivation of a stochastic dynamical system for the combined evolution of the truth and its estimate; such a setup allows for computing biases and correlations between the two processes which are crucial in assessing the filter skill. Here, we consider a simple canonical setup, analogous to that in [60, 41, 2], in which the discrete dynamics of the truth signal is given by the linear Gaussian system

$$\mathbf{u}_{m+1} = F\mathbf{u}_m + \mathcal{F}_{m+1} + \boldsymbol{\sigma}_{m+1}, \quad \mathbb{E}[\boldsymbol{\sigma}_{m+1} \otimes \boldsymbol{\sigma}_{m+1}^*] \equiv r_{m+1}, \quad (2.1)$$

where  $F$  is the discrete-time evolution operator for the state vector  $\mathbf{u}$  whose dynamics is driven by the deterministic forcing,  $\mathcal{F}$ , and a Gaussian noise,  $\boldsymbol{\sigma}$ , with covariance  $r$ . *Filtering* the dynamics in (2.1) refers to the statistical estimation of the state  $\mathbf{u}$  at the consecutive times  $\{t_m\}_{m \in \mathbb{Z}^+}$  given noisy *observations* and a *forecast model* for the true dynamics in (2.1). Here, we assume linear Gaussian observations of  $\mathbf{u}$  at each time  $t_m$  represented by

$$\mathbf{v}_{m+1} = G\mathbf{u}_{m+1} + \boldsymbol{\sigma}_{m+1}^o, \quad \mathbb{E}[\boldsymbol{\sigma}_{m+1}^o \otimes (\boldsymbol{\sigma}_{m+1}^o)^*] \equiv r_{m+1}^o, \quad (2.2)$$

with the linear observation operator  $G$  and the observation noise variance  $r^o$ . The linear Gaussian forecast model used to approximate the true dynamics in (2.1) is given by

$$\mathbf{u}_{m+1}^M = F^M \mathbf{u}_m^M + \mathcal{F}_{m+1}^M + \boldsymbol{\sigma}_{m+1}^M, \quad \mathbb{E}[\boldsymbol{\sigma}_{m+1}^M \otimes (\boldsymbol{\sigma}_{m+1}^M)^*] \equiv r_{m+1}^M, \quad (2.3)$$

where the forward operator,  $F^M$ , the deterministic forcing  $\mathcal{F}^M$ , and the model noise variance,  $r^M$ , are not necessarily identical to  $F$ ,  $\mathcal{F}$ , and  $r$  in the true dynamics (2.1); model error is present in the forecast dynamics (2.3) when at least one of the following holds true:  $F^M \neq F$ , and/or  $\mathcal{F}^M \neq \mathcal{F}$ , and/or  $r^M \neq r$ .

The discrete-time, discrete-update Kalman filter [43, 41, 2, 17, 60] is a well-known iterative procedure which provides the maximum likelihood estimate,  $\bar{\mathbf{u}}_{m+1|m+1}$ , for the true state  $\mathbf{u}_{m+1}$  at time  $t_{m+1}$  via a Bayesian update which combines the observations up to  $\mathbf{v}_{m+1}$  and the predicted mean estimate  $\bar{\mathbf{u}}_{m+1|m} = F^M \bar{\mathbf{u}}_{m|m}$ , so that

$$\bar{\mathbf{u}}_{m+1|m+1} = F^M \bar{\mathbf{u}}_{m|m} + K_{m+1} (\mathbf{v}_{m+1} - GF^M \bar{\mathbf{u}}_{m|m}), \quad (2.4)$$

where the Kalman gain,  $0 \leq K_m \leq 1$ , is given by

$$K_{m+1} = \mathcal{E}_{m+1|m} G^* \left( G \mathcal{E}_{m+1|m} G^* + r_{m+1}^o \right)^{-1}, \quad (2.5)$$

with the forecast error covariance,  $\mathcal{E}_{m+1|m} \equiv \mathbb{E}[(\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m}) \otimes (\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m})^*]$ . The Kalman gain weights the estimate in (2.4) towards the model prediction for  $K_m \sim 0$  and towards the observations for  $K_m \sim 1$ ; some important issues related to the necessary approximations of the Kalman gain in imperfect filtering are discussed at the end of this section. The important fact pertaining to the estimate in (2.4) is that in the absence of model error, i.e., when  $(F^M = F, \mathcal{F}^M = \mathcal{F}, r^M = r)$ , the *posterior* mean  $\bar{\mathbf{u}}_{m+1|m+1}$  provides the optimal estimate for the true state  $\mathbf{u}_{m+1}$  given the sequence of noisy observations  $\{\mathbf{v}_i\}_{i \leq m+1}$  (see, e.g., [43, 44, 41, 2, 17, 60]); here, the *filter*

*optimality* is understood in the sense of minimum error variance and maximum path-wise correlation between the truth and the filter estimate. However, the optimality of the estimate in (2.4) no longer holds in the presence of model error, which introduces many additional mathematical subtleties into the above filtering framework, as we show in the following sections. Nevertheless, this type of ‘perfect model’ assumption is prevalent in the classical mathematical or engineering formulations of the filtering problem (e.g., [49, 5, 41, 2, 17]).

The information-theoretic evaluation of the performance of imperfect filters is achieved by considering the augmented dynamics of the true signal  $\mathbf{u}_m$  and its filter estimate  $\bar{\mathbf{u}}_{m|m}$ ; the coupled evolution of the augmented state  $\mathbf{y}_m \equiv (\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})^T$  is easily derived by combining (2.1) and (2.4) in the form

$$\begin{aligned} \begin{bmatrix} \mathbf{u}_{m+1} \\ \bar{\mathbf{u}}_{m+1|m+1} \end{bmatrix} &= \begin{bmatrix} F & 0 \\ K_{m+1}GF & (\mathcal{I} - K_{m+1}G)F^M \end{bmatrix} \begin{bmatrix} \mathbf{u}_m \\ \bar{\mathbf{u}}_{m|m} \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{\sigma}_{m+1} \\ K_{m+1}(G\boldsymbol{\sigma}_{m+1} + \boldsymbol{\sigma}_{m+1}^o) \end{bmatrix} + \begin{bmatrix} \mathcal{F}_{m+1} \\ (\mathcal{I} - K_{m+1}G)\mathcal{F}_{m+1}^M + K_{m+1}G\mathcal{F}_{m+1} \end{bmatrix}. \end{aligned} \tag{2.6}$$

The important difference between the dynamics in (2.6) and the augmented Kalman filter systems studied in [2, 60] is that the present framework focuses on the error in the filter estimate rather than on the model error in the filtered forecast; i.e., we focus on the coupled dynamics of  $\mathbf{y}_m = (\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})^T$  rather than the dynamics of  $\boldsymbol{\chi}_m = (\mathbf{u}_m, F^M \bar{\mathbf{u}}_{m-1|m-1})^T$ . Despite the deceptive similarity of these two formulations, there are significant differences between these two problems in the presence of model error in the forecast (2.3); we postpone the discussion of model error in the filter forecast to a separate publication.

The statistics of the linear Gaussian system (2.6) is fully characterized by its mean  $\mathbb{E}[\mathbf{y}_m]$  and covariance  $\mathcal{C}_m \equiv \text{Cov}(\mathbf{y}_m, \mathbf{y}_m) = \mathbb{E}[\mathbf{y}_m \otimes \mathbf{y}_m^*] - \mathbb{E}[\mathbf{y}_m] \otimes \mathbb{E}[\mathbf{y}_m^*]$ , which evolves according to

$$\mathcal{C}_{m+1} = \mathcal{F}_m \mathcal{C}_m \mathcal{F}_m^* + \mathcal{R}_m, \tag{2.7}$$

where the operators  $\mathcal{F}$  and  $\mathcal{R}$  in (2.7) are given respectively by

$$\mathcal{F}_m = \begin{bmatrix} F & 0 \\ K_{m+1}GF & (\mathcal{I} - K_{m+1}G)F^M \end{bmatrix} \tag{2.8}$$

and

$$\mathcal{R}_m = \begin{bmatrix} r & rG^*K_{m+1}^* \\ K_{m+1}Gr & K_{m+1}(r^o + GrG^*)K_{m+1}^* \end{bmatrix}. \tag{2.9}$$

Given the augmented state,  $\mathbf{y}_m = (\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})^T$ , the covariance  $\mathcal{C}_m$  in (2.7) has a natural block decomposition

$$\mathcal{C}_m = \begin{bmatrix} \text{Cov}(\mathbf{u}_m, \mathbf{u}_m) & \text{Cov}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) \\ \text{Cov}(\bar{\mathbf{u}}_{m|m}, \mathbf{u}_m) & \text{Cov}(\bar{\mathbf{u}}_{m|m}, \bar{\mathbf{u}}_{m|m}) \end{bmatrix} \equiv \begin{bmatrix} \mathcal{C}_{(11)m} & \mathcal{C}_{(12)m} \\ \mathcal{C}_{(12)m}^* & \mathcal{C}_{(22)m} \end{bmatrix}. \tag{2.10}$$

The block  $\mathcal{C}_{(11)}$  of (2.10) denotes the covariance of the truth dynamics,  $\mathcal{C}_{(22)}$  is the covariance of the filter estimate, and  $\mathcal{C}_{(12)}$  quantifies the correlations between the truth  $\mathbf{u}_m$  and the filter estimate  $\bar{\mathbf{u}}_{m|m}$ . It will prove crucial in the following analysis of filter skill to account for both the covariance of the filter estimates  $\mathcal{C}_{(22)}$  and the cross-correlations between the truth and the filter estimates contained in  $\mathcal{C}_{(12)}$ .

One important issue in the present analysis of error in imperfect Kalman filtering concerns the necessary approximations of the Kalman gain,  $K_{m+1}$ , in (2.6) when filtering with imperfect forecast model (2.3). It can be easily shown (see (2.12) below) that when filtering with the forecast dynamics (2.3) the correct update of the Kalman gain in (2.5) is practically inaccessible since it requires the knowledge of both the perfect (2.1) and the imperfect dynamics (2.3); thus, practical implementations of Kalman filtering with model error imply introducing errors in both the forecast dynamics and in the Kalman gain through the necessary approximations.

In the absence of model error, i.e., when  $(F^M = F, \mathcal{F}^M = \mathcal{F}, r^M = r)$  in (2.3), the Kalman gain (2.5) can be computed off-line according to the following coupled dynamics:

$$\begin{aligned} a) \quad & K_{m+1} = \mathcal{E}_{m+1|m} G^* \left( G \mathcal{E}_{m+1|m} G^* + r_{m+1}^o \right)^{-1}, \\ b) \quad & \mathcal{E}_{m+1|m} = F \mathcal{E}_{m|m} F^* + r_{m+1}, \\ c) \quad & \mathcal{E}_{m|m} = (\mathcal{I} - K_m G) \mathcal{E}_{m|m-1}, \end{aligned} \tag{2.11}$$

which can be solved independently of the augmented dynamics in (2.7); this can be easily verified by noticing that the evolution of the forecast error covariance  $\mathcal{E}_{m+1|m} \equiv \mathbb{E}[(\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m}) \otimes (\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m})^*]$  and the filter error covariance  $\mathcal{E}_{m+1|m+1} \equiv \mathbb{E}[(\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m+1}) \otimes (\mathbf{u}_{m+1} - \bar{\mathbf{u}}_{m+1|m+1})^*]$  do not require the knowledge of covariance components in (2.10) and can be computed directly from the system (2.11). The presence of model error in (2.3) and (2.6) drastically changes this situation and, instead of the dynamics in (2.11b), the forecast error covariance,  $\mathcal{E}_{m+1|m}$ , evolves according to

$$\begin{aligned} \mathcal{E}_{m+1|m} = & F \mathcal{C}_{(11)m} F^* + F^M \mathcal{C}_{(22)m} (F^M)^* - F \mathcal{C}_{(12)m} (F^M)^* - F^M \mathcal{C}_{(21)m} F^* + r_{m+1} \\ & + \left( F \langle \mathbf{u}_m \rangle - F^M \langle \mathbf{u}_m \rangle + \mathcal{F}_{m+1} - \mathcal{F}_{m+1}^M \right) \otimes \\ & \otimes \left( F \langle \mathbf{u}_m \rangle - F^M \langle \mathbf{u}_m \rangle + \mathcal{F}_{m+1} - \mathcal{F}_{m+1}^M \right)^*, \end{aligned} \tag{2.12}$$

where the components of the covariance of the augmented system in (2.7) are defined in (2.10); clearly for  $F^M = F$ ,  $\mathcal{F}_{m+1}^M = \mathcal{F}_{m+1}$ , and  $\langle \mathbf{u}_0 \rangle = \langle \mathbf{u}_{0|0} \rangle$  the dynamics in (2.12) reduce to those in (2.11b) since then the filter error covariance becomes  $\mathcal{E}_{m|m} = \mathcal{C}_{(11)m} + \mathcal{C}_{(22)m} - \mathcal{C}_{(12)m} - \mathcal{C}_{(21)m}$ . An immediate complication arising in practical applications when filtering with model error is that the true dynamics in (2.1) are not known and the correct forecast covariance,  $\mathcal{E}_{m+1|m}$ , in (2.12) cannot be propagated. A common approximation aimed at circumventing this problem is to approximate the evolution of the forecast error covariance using the imperfect model dynamics (2.3) with  $F^M$  and  $r^M$  and, instead of (2.11), determine the approximate



Kalman gain by solving

$$\begin{aligned}
 a) \quad K_{m+1}^M &= \mathcal{E}_{m+1|m}^M G^* \left( G \mathcal{E}_{m+1|m}^M G^* + r_{m+1}^o \right)^{-1}, \\
 b) \quad \mathcal{E}_{m+1|m}^M &= F^M \mathcal{E}_{m|m}^M (F^M)^* + r_{m+1}^M, \\
 c) \quad \mathcal{E}_{m|m}^M &= (\mathcal{I} - K_m^M G) \mathcal{E}_{m|m-1}^M.
 \end{aligned} \tag{2.13}$$

The point-wise in time error of the approximation in (2.13b) relative to (2.12) can be formally identified by rewriting the true forecast error covariance,  $\mathcal{E}_{m+1|m}$ , as

$$\mathcal{E}_{m+1|m} = F^M \mathcal{E}_{m|m} (F^M)^* + r_{m+1}^M + \Delta \mathcal{E}_{m+1}, \tag{2.14}$$

where

$$\begin{aligned}
 \Delta \mathcal{E}_{m+1} &= F \mathcal{C}_{(11)m} F^* - F^M \mathcal{C}_{(11)m} (F^M)^* \\
 &+ F^M \mathcal{C}_{(21)m} \left( (F^M)^* - F^* \right) + \left( F^M - F \right) \mathcal{C}_{(12)m} (F^M)^* \\
 &+ r_{m+1} - r_{m+1}^M + \left( F \langle \mathbf{u}_m \rangle - F^M \langle \mathbf{u}_m | m \rangle + \mathcal{F}_{m+1} - \mathcal{F}_{m+1}^M \right) \otimes \\
 &\quad \otimes \left( F \langle \mathbf{u}_m \rangle - F^M \langle \mathbf{u}_m | m \rangle + \mathcal{F}_{m+1} - \mathcal{F}_{m+1}^M \right)^* \\
 &- F^M \langle \mathbf{u}_m - \bar{\mathbf{u}}_m | m \rangle \otimes \langle \mathbf{u}_m^* - \bar{\mathbf{u}}_m^* | m \rangle (F^M)^*.
 \end{aligned} \tag{2.15}$$

Thus, Kalman filtering with imperfect forecast dynamics is associated with two coupled sources of model error: (i) the error introduced into the forecast mean estimate  $\bar{\mathbf{u}}_{m+1|m} = F^M \bar{\mathbf{u}}_m | m$  through the incorrect model dynamics when  $F^M \neq F$ , and (ii) the error introduced into the forecast error covariance  $\mathcal{E}_{m+1|m}$  through the use of the update in (2.13b) instead of the correct but practically inaccessible update in (2.12). The analysis presented in the subsequent sections is aimed at elucidating various effects of the model error in imperfect Kalman filtering via the information-theoretic tools described in the next section. An interesting future research direction concerns developing techniques for ‘on the fly’ minimization of the additive error,  $\Delta \mathcal{E}$ , in the Kalman gain in (2.14) via augmenting the imperfect forecast dynamics in (2.3) with a stochastic model for  $\Delta \mathcal{E}$  in (2.15) in a framework analogous to that used in SPEKF filtering [31, 30, 60] and in earlier works [26, 27, 21, 22].

**2.2. Quantifying ensemble error in filtering through information theory.** The discrepancy between the true signal and its filter estimate can be measured in various ways. The standard measures of filter error are based on path-wise time averages of biases and on correlations between the truth and its filter estimate (e.g., [60, 41, 2]). Here, we introduce a complementary approach to quantifying the filter performance within the statistical superensemble framework via methods of information theory (e.g., [20]); this approach provides systematic off-line ensemble-averaged estimates of filter skill and it allows for identifying lower bounds on the filter error which correspond to information barriers [53, 50, 58] in imperfect filters. The information measures of filter error introduced below are not restricted to the Gaussian framework but they naturally incorporate the Gaussian indicators of filter skill given by the filter error covariance, mean biases and correlations between the truth and the filter estimates. As shown in the following sections, these features are



especially important and desirable when quantifying the error in filtering spatially extended systems from sparse observations which leads to estimating high-dimensional state vectors.

In this section we first introduce some concepts from information theory which are used to quantify different aspects of the performance of imperfect Kalman filters; in Section 2.4 these different measures are used to construct a new information criterion for optimizing imperfect Kalman filters. We then show in Section 2.2.2 that the well-known measures of the asymptotic filter error based on the time-averaged, path-wise estimates of biases and correlations between the truth and the filter estimate have natural information-theoretic analogues; importantly, these can be derived exactly from the attractor statistics of the augmented Kalman filter system (2.6). This extended framework allows, in particular, for assessing the skill for recovering the autocorrelation functions of the filtered signal. Accurate estimates of the autocorrelation functions from measurements are crucial for data-driven design of reduced stochastic models of nonlinear systems from experimental data [54, 57] with the property of having the right forced response to external perturbations, as well as for tuning the sensitivity of imperfect models for accurate prediction of forced response via the fluctuation-dissipation theorem (e.g., [52, 51, 1, 58, 13]). The framework introduced below is subsequently used in Section 3 and Section 4 to study the performance of reduced order filters for spatially extended turbulent systems.

**2.2.1. Information-theoretic measures of filter error.** Here, we consider three distinct and complementary information-theoretic measures of filter fidelity. As shown below, a combination of all these measures is necessary to quantify the accuracy of the imperfect filter estimates in the statistical superensemble framework. When filtering with imperfect models it is generally not possible to optimize all of these measures, as illustrated in figure 3.1; rough guidelines for choosing the most relevant information measure in imperfect filtering, as well as a new information criterion combining the measures discussed below are discussed in Section 2.4. Although, the following analysis is restricted to the Gaussian framework of Kalman filtering, these measures naturally quantify various facets of information content in non-Gaussian filter estimates which will be discussed in subsequent publications. For a more general formulation and additional details on the information measures listed below see, for example, [20].

**THE SHANNON ENTROPY**,  $\mathcal{S}(\mathbf{u}_m)$ , of the residual  $\mathbf{u}_m := \mathbf{u}_m - \bar{\mathbf{u}}_{m|m}$ ,  $\mathbf{u}_m \sim p$ , is given by

$$\mathcal{S}(\mathbf{u}_m) := -\mathbb{E}^p[\ln p(\mathbf{u}_m)] \equiv -\int p(\mathbf{u}_m) \ln p(\mathbf{u}_m) d\mathbf{u}_m, \tag{2.16}$$

and it expresses the uncertainty in the filter estimate  $\bar{\mathbf{u}}_{m|m}$  about the true state  $\mathbf{u}_m$  at time  $t_m$ ; the process  $\mathbf{u}_m$  is in general time-dependent but we skip the explicit time dependence in the joint probability density  $p$  for clarity. While it is well known (e.g., [20]) that the entropy of the residual  $\mathbf{u}_m = (\mathbf{u}_m^{(1)}, \dots, \mathbf{u}_m^{(N)})$  in (2.16) can be negative, it is bounded from above by the entropy of its components

$$\mathcal{S}(\mathbf{u}_m) \leq \sum_i \mathcal{S}(\mathbf{u}_m^{(i)}), \tag{2.17}$$

with the equality in (2.17) only when the components  $\mathbf{u}_m^{(i)}$  are statistically independent. We will show below that the entropy of the filter error in (2.16) depends on

the correlations between the truth and the filter estimates but is insensitive to mean biases in the filter estimates.

**THE RELATIVE ENTROPY**,  $\mathcal{P}(\pi, \pi^f)$ , quantifies the lack of information in the statistics of the filter estimate with density  $\pi^f$  relative to the statistics of the truth with density  $\pi$  as

$$\mathcal{P}(\pi, \pi^f) := \int \pi(\mathbf{u}) \ln \frac{\pi(\mathbf{u})}{\pi^f(\mathbf{u})} d\mathbf{u}, \tag{2.18}$$

where, given the joint density of the truth and the filter estimate  $p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$ , the truth density is expressed via the marginal  $\pi(\mathbf{u}_m) \equiv \int p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) d\bar{\mathbf{u}}_{m|m}$ , and the imperfect density of the filter estimate is given by the marginal  $\pi^f(\bar{\mathbf{u}}_{m|m}) \equiv \int p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) d\mathbf{u}_m$  (for simplicity, we skip the explicit conditioning on the observations in the densities above). The relative entropy is often interpreted as a ‘distance’ between the two probability densities but it is not a true metric. Nevertheless, it has many desirable properties for characterizing model error [46, 47, 55, 52, 53, 58]; in particular, (i)  $\mathcal{P} \geq 0$  with  $\mathcal{P} = 0$  only when  $\pi = \pi^f$ , and (ii) it is invariant under non-linear changes of variables. We show below that when  $\pi, \pi^f$  represent the marginal densities of time-lagged state of the augmented system (2.6), the relative entropy captures both the statistical biases in the mean filter estimates and the correlations between  $\mathbf{u}_m$  and  $\bar{\mathbf{u}}_{m|m}$  which are equally important when assessing the filter error.

**THE MUTUAL INFORMATION** between the truth  $\mathbf{u}_m$  and the filter estimate  $\bar{\mathbf{u}}_{m|m}$  measures the dependence between these two processes and is given by the symmetric formula

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) := \iint p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) \ln \frac{p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})}{\pi(\mathbf{u}_m) \pi^f(\bar{\mathbf{u}}_{m|m})} d\mathbf{u}_m d\bar{\mathbf{u}}_{m|m}. \tag{2.19}$$

The mutual information in (2.19) is useful in the context of filtering because it can be interpreted as the measure of uncertainty reduction in the knowledge about the true state  $\mathbf{u}_m$  given the filter estimate  $\bar{\mathbf{u}}_{m|m}$ ; this interpretation becomes obvious upon rewriting (2.19) as

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = \mathcal{S}(\mathbf{u}_m) - \int \mathcal{S}(\mathbf{u}_m | \bar{\mathbf{u}}_{m|m}) \pi^f(\bar{\mathbf{u}}_{m|m}) = \mathcal{S}(\bar{\mathbf{u}}_{m|m}) - \int \mathcal{S}(\bar{\mathbf{u}}_{m|m} | \mathbf{u}_m) \pi(\mathbf{u}_m). \tag{2.20}$$

Another useful interpretation of the mutual information is as a measure of lack of information in the factorized density  $\pi(\mathbf{u}) \pi^f(\bar{\mathbf{u}}_{m|m})$  relative to the joint density  $p(\mathbf{u}, \bar{\mathbf{u}}_{m|m})$  which follows from the identity

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = \mathcal{P}\left(p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}), \pi(\mathbf{u}_m) \pi^f(\bar{\mathbf{u}}_{m|m})\right). \tag{2.21}$$

Hence, similar to the relative entropy in (2.18), the mutual information is nonnegative and it is invariant under nonlinear changes of variables. We show below that, while the mutual information measures the statistical correlations between the truth and its filter estimate, it is insensitive to the mean biases in the filter estimates which are particularly important when filtering turbulent dynamics with model error.

The above three measures of information content in the filter estimates quantify different aspects of filter error. For Kalman filtering with linear Gaussian dynamics (2.6) the differences between these measures are best expressed through the explicit formulas whose derivation is simplified by the following useful facts:

**FACT 1.** The Shannon entropy (2.16) for the filter bias,  $\mathbf{U}_m = \mathbf{u}_m - \bar{\mathbf{u}}_{m|m}$ , where the truth  $\mathbf{u}_m$  and the filter estimate  $\bar{\mathbf{u}}_{m|m}$  satisfy (2.6) is given by

$$\mathcal{S}(\mathbf{U}_m) - \frac{1}{2}N(1 + \ln 2\pi) = \frac{1}{2} \ln \det [\mathbf{U}_m \otimes \mathbf{U}_m] = \frac{1}{2} \ln \det [\mathcal{C}_{(11)m} + \mathcal{C}_{(22)m} - 2\Re[\mathcal{C}_{(12)m}]], \tag{2.22}$$

where  $N$  is the dimension of  $\mathbf{u}_m$  and  $\mathcal{C}_{(11)}$ ,  $\mathcal{C}_{(12)}$ , and  $\mathcal{C}_{(22)}$  are the respective blocks of the covariance in (2.7). This simple result is obtained from the the expression for Shannon entropy for Gaussian processes (see, e.g., [20]) and the fact that  $Cov(\mathbf{U}, \mathbf{U}) = \mathcal{C}_{(11)} + \mathcal{C}_{(22)} - 2\Re[\mathcal{C}_{(12)}]$ . It is clear from (2.22) that the entropy for the continuous valued random process can be negative; in the information measures exploited below we will use the monotonic mapping of  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$  into the positive half-line using  $\exp(\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}))$  which, for scalar Gaussian fields, becomes a multiple of the ensemble averaged root mean square error, as discussed in Section 2.2.2.

**FACT 2.** The relative entropy between the Gaussian truth process  $\mathbf{u}_m$  and its Kalman filter estimate  $\bar{\mathbf{u}}_{m|m}$  can be written as a sum of *signal* and *dispersion* terms

$$\begin{aligned} \mathcal{P}(\pi(\mathbf{u}_m), \pi^{\dagger}(\bar{\mathbf{u}}_{m|m})) &= \frac{1}{2} \mathbb{E}[\mathbf{u}_m^* - \bar{\mathbf{u}}_{m|m}^*] \mathcal{C}_{(22)m}^{-1} \mathbb{E}[\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}] \\ &+ \frac{1}{2} \left[ \text{tr}(\mathcal{C}_{(11)m}(\mathcal{C}_{(22)m})^{-1}) - \ln \det(\mathcal{C}_{(11)m}(\mathcal{C}_{(22)m})^{-1}) - N \right], \end{aligned} \tag{2.23}$$

where  $N$  is the dimension of  $\mathbf{u}_m$ , the first term is the signal and the second term is the dispersion (e.g. [46, 47, 55, 52]) and the truth  $\mathcal{C}_{(11)}$  and model  $\mathcal{C}_{(22)}$  covariance blocks are defined in (2.10); the signal part in (2.23),

$$sig_{\mathcal{P}} = \frac{1}{2} \mathbb{E}[\mathbf{u}_m^* - \bar{\mathbf{u}}_{m|m}^*] \mathcal{C}_{(22)m}^{-1} \mathbb{E}[\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}], \tag{2.24}$$

quantifies the bias in the mean filter estimate while the dispersion part,

$$dis_{\mathcal{P}} = \frac{1}{2} \left[ \text{tr}(\mathcal{C}_{(11)m}(\mathcal{C}_{(22)m})^{-1}) - \ln \det(\mathcal{C}_{(11)m}(\mathcal{C}_{(22)m})^{-1}) - N \right], \tag{2.25}$$

quantifies the lack of information in the covariance of the filter estimate  $\mathcal{C}_{(22)}$  relative to the true signal covariance  $\mathcal{C}_{(11)}$ . The relative entropy operates on the marginal densities,  $\pi(\mathbf{u}_m)$ ,  $\pi^{\dagger}(\bar{\mathbf{u}}_{m|m})$ ; consequently, it does not take into account the cross-correlations between the truth and the filter estimate but it accounts for the biases in the ensemble mean through the signal term in (2.23).

**FACT 3.** The mutual information (2.19) between the Gaussian truth  $\mathbf{u}_m$  and its Kalman filter estimate  $\bar{\mathbf{u}}_{m|m}$  satisfying (2.6) is given by

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = -\frac{1}{2} \ln \det (\mathcal{I} - \mathcal{C}_{(22)m}^{-1} \mathcal{C}_{(12)m}^* \mathcal{C}_{(11)m}^{-1} \mathcal{C}_{(12)m}), \tag{2.26}$$

where the  $\mathcal{C}_{(11)}$  is the covariance of the truth,  $\mathcal{C}_{(22)}$  is the covariance of the filter estimate, and  $\mathcal{C}_{(12)}$  is the cross-covariance between the truth and the filter estimate in (2.7). The mutual information in (2.26) is insensitive to biases in the mean filter estimate  $\mathbb{E}[\bar{\mathbf{u}}_{m|m}]$  relative to the true ensemble mean  $\mathbb{E}[\mathbf{u}_m]$ .

The identity in (2.26) can be obtained from the representation of the mutual entropy in (2.20) which for Gaussian  $\mathbf{u}_m$  and  $\bar{\mathbf{u}}_{m|m}$  which satisfy (2.6) can be expressed

via the respective covariances in (2.10) as follows:

$$\mathcal{S}(\mathbf{u}_m) = \frac{1}{2} \ln \det \mathcal{C}_{(11)m} + \frac{1}{2} N(1 + \log 2\pi), \quad \mathcal{S}(\bar{\mathbf{u}}_{m|m}) = \frac{1}{2} \ln \det \mathcal{C}_{(22)m} + \frac{1}{2} N(1 + \log 2\pi), \quad (2.27)$$

and

$$\int \mathcal{S}(\bar{\mathbf{u}}_{m|m} | \mathbf{u}_m) \pi(\mathbf{u}_m) = \frac{1}{2} \ln \det \left( \mathcal{C}_{(22)m} - \mathcal{C}_{(12)m}^* \mathcal{C}_{(11)m}^{-1} \mathcal{C}_{(12)m} \right) + \frac{1}{2} N(1 + \log 2\pi), \quad (2.28)$$

where  $\mathcal{C}_{(11)}$ ,  $\mathcal{C}_{(22)}$  are blocks of the covariance matrix (2.10), and  $N$  is the dimension of  $\mathbf{u}_m$  and  $\mathbf{u}_{m|m}$ . Substitution of (2.27) and (2.28) into (2.20) leads to (2.26).

The above information measures quantify the filter error point-wise in time. Often, a more comprehensive, nonlocal-in-time, diagnostic is necessary for understanding the effects of errors in the correlation functions in the estimated signals. This is particularly important in the development and validation of physics constrained multi-level nonlinear regression models [54, 57] which simultaneously reflect the causality and energy conserving principles of the underlying nonlinear dynamics. Particularly important uses of the autocorrelation estimates in the data-driven design of reduced stochastic models of nonlinear systems from experimental data [54, 57] involve ensuring the right forced response to external perturbations, as well as tuning the sensitivity of imperfect models for accurate prediction of forced response via the fluctuation-dissipation theorem (e.g., [52, 51, 1, 58, 13]). The following fact extends our framework to this much more general setting:

**FACT 4** [*Lag-embedded measures of filter error.*] The pointwise-in-time measures of filter error introduced in facts 1-3 above have direct lag-embedded counterparts given by the entropy,  $\mathcal{S}(\mathbf{u}_m^L)$ , mutual information,  $\mathcal{M}(\mathbf{u}_m^L, \bar{\mathbf{u}}_{m|m}^L)$ , and the relative entropy,  $\mathcal{P}(\pi(\mathbf{u}^L), \pi^\dagger(\bar{\mathbf{u}}_{m|m}^L))$ , of the lag-embedded states given by

$$\mathbf{u}_m^L \equiv (\mathbf{u}_m, \mathbf{u}_{m-1}, \dots, \mathbf{u}_{m-L})^T, \quad \bar{\mathbf{u}}_{m|m}^L \equiv (\bar{\mathbf{u}}_{m|m}, \bar{\mathbf{u}}_{m-1|m-1}, \dots, \bar{\mathbf{u}}_{m-L|m-L})^T. \quad (2.29)$$

It is trivial to observe that the above lag-embeddings of Gaussian processes yield Gaussian processes in the extended phase space which makes facts 1-3 immediately applicable to the lag-embedded case. The nonlocal-in-time error in the lag-embedded filter estimates  $\bar{\mathbf{u}}_{m|m}^L$  provides additional information about the filter performance beyond the one-point statistics of the filter estimates. In particular, note that in contrast to the time-point-wise framework, the relative entropy of lag-embedded states depends on the cross-correlation between the truth and the filter estimates; explicit formulas for the covariance of the lag embedded state of the discretized spatially extended canonical system (3.1) are presented in Appendix A.

**2.2.2. Ensemble counterparts of asymptotic path-wise measures of filter error.** Here, we establish some important links between two asymptotic path-wise measures of filter error and two information measures which exploit the statistical ensemble framework and avoid the need for path-wise sampling. The path-wise error measures given below are matrix generalizations of the standard root-mean-square error and pattern correlation; these generalized notions are better suited for higher-dimensional correlated inputs generated in filtering spatially extended systems. The path-wise measures considered here are the following:

**ASYMPTOTIC PATH-WISE RMS ERROR**

$$\left(\text{RMSE}(\mathbf{u}_{\{\infty\}})\right)^2 := \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{m=\tilde{m}}^K (\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}) \otimes (\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})^*, \quad (2.30)$$

where  $t_{\tilde{m}}$  is some suitably large time instant and  $\mathbf{u}_m = \mathbf{u}_m - \bar{\mathbf{u}}_{m|m}$  is the path-wise filter error at  $t_m$ . The asymptotic RMS error in (2.30) measures average path-wise distance between the truth and the filter estimate.

In order to define the second path-wise measure of filter skill which is sensitive to correlations in the path-wise filter estimates and the truth signal, we first define the path-wise empirical approximations for the covariances of the truth and the filter estimate given respectively by

$$\mathcal{C}_K = \frac{1}{K} \sum_{k=\tilde{m}}^K (\mathbf{u}_k - \tilde{\mathbf{u}}_K) \otimes (\mathbf{u}_k - \tilde{\mathbf{u}}_K)^*, \quad \mathcal{C}_{K|K} = \frac{1}{K} \sum_{k=\tilde{m}}^K (\bar{\mathbf{u}}_{k|k} - \tilde{\mathbf{u}}_{K|K}) \otimes (\bar{\mathbf{u}}_{k|k} - \tilde{\mathbf{u}}_{K|K})^*, \quad (2.31)$$

where  $\tilde{\mathbf{u}}^K = \frac{1}{K} \sum_{k=\tilde{m}}^K \mathbf{u}_k$  and  $\tilde{\mathbf{u}}_{K|K} = \frac{1}{K} \sum_{k=\tilde{m}}^K \bar{\mathbf{u}}_{k|k}$  are the respective path-wise sample average with some sufficiently large  $\tilde{m}$  guaranteeing the attractor dynamics. Given the path-wise covariance estimates in (2.31), we make the following definition.

**ASYMPTOTIC PATH-WISE CORRELATION MATRIX** between the truth,  $\mathbf{u}$ , and the filter estimate,  $\bar{\mathbf{u}}_{\cdot| \cdot}$ ,

$$\text{Corr}(\mathbf{u}_{\{\infty\}}, \bar{\mathbf{u}}_{\{\infty|\infty\}}) := \lim_{K \rightarrow \infty} \mathcal{C}_{K|K}^{-1} \mathcal{C}_K^* \mathcal{C}_K^{-1} \mathcal{C}_K, \quad (2.32)$$

where the path-wise cross-correlation matrix in (2.32) is given by

$$\mathcal{C}_K = \frac{1}{K} \sum_{k=\tilde{m}}^K (\mathbf{u}_k - \tilde{\mathbf{u}}_K) \otimes (\bar{\mathbf{u}}_{k|k} - \tilde{\mathbf{u}}_{K|K})^*. \quad (2.33)$$

By construction the correlation matrix in (2.32) between the truth and the filter estimate is insensitive to biases in the mean dynamics of the filter estimates.

REMARK 2.1.

- Note that the asymptotic RMS error and pattern correlation in (2.30) and (2.32) depend on the whole history of the solutions  $\mathbf{u}_m$  and  $\bar{\mathbf{u}}_{m|m}$ . The abbreviated notation on the left-hand sides of (2.30) and (2.32) should not be confused with evaluating the respective quantities at the asymptotic limit point-wise in time.
- The above definitions can be reduced to the standard definitions of RMS error and pattern correlation by taking the trace of (2.30) and (2.32) in the case when the components of  $\mathbf{u}_m$  and  $\bar{\mathbf{u}}_{m|m}$  are uncorrelated; this situation is discussed in Section 3 when dealing with plentiful observations of spatially extended systems.

The links between the path-wise asymptotic measures of filter error introduced above and the information-theoretic measures of the attractor statistics of the filter error are given by the following:

PROPOSITION 2.1. *The entropy of the asymptotic filter error,  $\mathbf{u}_\infty := \mathbf{u}_\infty - \bar{\mathbf{u}}_{\infty|\infty}$ , on the attractor of (2.6) can be expressed through the asymptotic path-wise RMS error as*

$$\mathcal{S}(\mathbf{u}_\infty) = \frac{1}{2} N (1 + \ln 2\pi) + \frac{1}{2} \ln \det \left[ (\text{RMSE}(\mathbf{u}_{\{\infty\}}))^2 - \mathbb{E}[\mathbf{u}_\infty] \otimes \mathbb{E}[\mathbf{u}_\infty^*] \right], \quad (2.34)$$

where the asymptotic path-wise RMS error is given by (2.30).

PROPOSITION 2.2. *The asymptotic mutual information,  $\mathcal{M}(\mathbf{u}_\infty, \bar{\mathbf{u}}_{\infty|\infty})$ , between the truth  $\mathbf{u}_\infty$  and the filter estimate  $\bar{\mathbf{u}}_{\infty|\infty}$  on the attractor of the system (2.6) can be expressed through the correlation matrix as*

$$\mathcal{M}(\mathbf{u}_\infty, \bar{\mathbf{u}}_{\infty|\infty}) = -\frac{1}{2} \ln \det \left[ \mathcal{I} - \text{Corr}(\mathbf{u}_{\{\infty\}}, \bar{\mathbf{u}}_{\{\infty|\infty\}}) \right], \tag{2.35}$$

where  $\text{Corr}(\mathbf{u}_{\{\infty\}}, \bar{\mathbf{u}}_{\{\infty|\infty\}})$  is given by (2.32) and  $\mathcal{I}$  is the identity matrix.

REMARK 2.2. In the above propositions the information measures on the left-hand-sides of (2.34) and (2.35) are evaluated point-wise in time based on the on the attractor statistics of the filter error of (2.6) while the expressions on the right-hand-sides of (2.34) and (2.35) involve the nonlocal-in-time attractor solutions of (2.6) as described in (2.30) and (2.32).

The proofs of the above propositions are straightforward provided that the dynamics in (2.6) are controllable ( $\mathbb{E}[\sigma^M \otimes (\sigma^M)^*] = r^M \neq 0$  in (2.6)) and observable ( $G \neq 0$  in (2.6)); these two conditions imply stability of the filtering procedure (e.g., [2, 17]) and the existence of an ergodic Gaussian attractor in the dynamics of (2.6). First, we justify the statement in Proposition 2.1. If the filtering described by the system (2.6) is stable with an ergodic attractor, we have

$$\mathbb{E}[\mathbf{u}_\infty \otimes \mathbf{u}_\infty^*] = \lim_{K \rightarrow \infty} \frac{1}{K} \sum_{k=\tilde{m}}^K \mathbf{u}_k \otimes \mathbf{u}_k^* \equiv \left( \text{RMSE}(\mathbf{u}_{\{\infty\}}) \right)^2, \tag{2.36}$$

where  $\tilde{m} \gg 1$  is large enough so that the system (2.6) transitions onto the attractor dynamics. Note that since the dynamics in (2.6) are Gaussian, the following holds for any time  $t_m$ :

$$\mathbb{E}[\mathbf{u}_\infty \otimes \mathbf{u}_\infty^*] - \mathbb{E}[\mathbf{u}_m] \otimes \mathbb{E}[\mathbf{u}_m^*] = \text{Cov}(\mathbf{u}_m, \mathbf{u}_m) = \frac{1}{(2\pi)^N e^N} e^{\mathcal{S}(\mathbf{u}_m)}, \tag{2.37}$$

where the second equality in (2.37) is due to Fact 1 in Section 2.2.1. The statement in Proposition 2.1 follows by taking the limit  $m \rightarrow \infty$  in (2.37) and combining it with (2.36) to obtain (2.34).

The proof of Proposition 2.2 follows similar reasoning; assuming that the filtering procedure described by the system (2.6) is stable so that it has a Gaussian ergodic attractor, we have

$$\text{Corr}(\mathbf{u}_{\{\infty\}}, \bar{\mathbf{u}}_{\{\infty|\infty\}}) = \mathcal{C}_{(22)\infty}^{-1} \mathcal{C}_{(12)\infty}^* \mathcal{C}_{(11)\infty}^{-1} \mathcal{C}_{(12)\infty}, \tag{2.38}$$

where the path-wise correlation matrix is given in (2.32) and  $\mathcal{C}_{(11)}$ ,  $\mathcal{C}_{(22)}$ , and  $\mathcal{C}_{(12)}$  are the block matrices in the covariance of the augmented system (2.6) defined in (2.10). The statement in Proposition 2.2 follows from (2.38) and Fact 2 in Section 2.2.1.

**2.3. Bounds on filter error statistics.** The quantification of the filter skill in the present superensemble framework relies on the entropy of the filter error  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$ , the mutual information  $\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$ , and the relative entropy  $\mathcal{P}(\pi(\mathbf{u}_m), \pi^\dagger(\bar{\mathbf{u}}_{m|m}))$ , which provide measures of an ‘information distance’ between the truth  $\mathbf{u}_m$  and its filter estimate  $\bar{\mathbf{u}}_{m|m}$ . The relative entropy,  $\mathcal{P}$ , plays a special role in this framework since it measures the lack of information in the statistics of the

estimated signal, as opposed to the filter error entropy  $\mathcal{S}$  and the mutual information  $\mathcal{M}$  which focus on the correlations between the truth and the filter estimates. As already pointed out in Section 2.2.1, all of the above ‘measures’ are not proper metrics but they have a natural interpretation in terms of bits of information (e.g., [20]) and are easily computable quantifiers of biases and correlations between the truth and the filter estimates based on the statistical superensemble averages. However, it is also desirable to derive bounds on the error in the filter estimates for the statistics of the truth signal in terms of proper metrics in a suitable Hilbert space of probability densities; this is particularly important when assessing bounds on filter error statistics from the information measures, as shown below.

The starting point here is to consider a general class of *f-divergences* between probability measures with densities  $p$  and  $q$  and defined as

$$D_f(p, q) := \int f\left(\frac{p(\mathbf{u})}{q(\mathbf{u})}\right) q(\mathbf{u}) d\mathbf{u}, \tag{2.39}$$

where  $f: [0, \infty) \rightarrow \mathbb{R}$ ,  $f(1) = 0$ , is a convex function (for a more general definition in any measurable space and with respect to an arbitrary dominating measure see, e.g., [35]). Clearly, the above class of *f-divergences* includes the relative entropy (or Kullback-Leibler divergence) defined in (2.18) which is recovered from (2.39) by setting  $f(x) = x \ln x$ . As already pointed out earlier, the relative entropy,  $\mathcal{P}$  in (2.23), is not a proper metric but this useful measure of ‘lack of information’ can be bounded by metrics obtained from (2.39) which, in contrast to  $\mathcal{P}$ , do not have an information-theoretic interpretation.

The Hellinger distance,  $d_H$ , between two probability densities  $p$  and  $q$  is obtained from (2.39) by setting  $f = \frac{1}{2}(\sqrt{x} - 1)^2$ , which leads to

$$d_H^2(p, q) = \frac{1}{2} \int \left(\sqrt{p(\mathbf{u})} - \sqrt{q(\mathbf{u})}\right)^2 d\mathbf{u}, \quad 0 \leq d_H \leq 1. \tag{2.40}$$

The usefulness of the Hellinger distance for our purposes is due to the existence of the following bounds on the relative entropy  $\mathcal{P}(p, q)$  (see, e.g., [67, 35]):

$$d_H^2(p, q) \leq \mathcal{P}(p, q) \leq d_H(p, q) + \frac{1}{2} \chi^2(p, q), \tag{2.41}$$

where the  $\chi^2$ -divergence is defined by

$$\chi^2(p, q) = \int \frac{(p(\mathbf{u}) - q(\mathbf{u}))^2}{q(\mathbf{u})} d\mathbf{u}, \tag{2.42}$$

and can be obtained from (2.39) for  $f = (x - 1)^2$ . Here, we employ the bounds on  $d_H$  in (2.41) to constrain the bias in the filter estimates for the mean dynamics and biases in the covariance of the estimated dynamics. In order to determine bounds on the filter error statistics, we exploit the general fact [19] that for any two second-order processes,  $x_p \sim p$  and  $x_q \sim q$ , with values in some separable Banach space  $(X, \|\cdot\|)$  and densities  $p$  and  $q$ , the Hellinger distance  $d_H(p, q)$  bounds the mean residual between the two processes, i.e,

$$\left\| \mathbb{E}^p[x_p] - \mathbb{E}^q[x_q] \right\| \leq 2 \left( \mathbb{E}^p[\|x_p\|^2] + \mathbb{E}^q[\|x_q\|^2] \right)^{1/2} d_H(p, q). \tag{2.43}$$



Moreover, if the processes  $x_p$  and  $x_q$  are fourth-order stable, the bias in the covariance of the filter estimates is bounded by

$$\left\| \mathbb{E}^P[x_p \otimes x_p] - \mathbb{E}^q[x_q \otimes x_q] \right\| \leq 2 \left( \mathbb{E}^P[\|x_p\|^4] + \mathbb{E}^q[\|x_q\|^4] \right)^{1/2} d_H(p, q). \quad (2.44)$$

Hence, assuming  $x_p = \mathbf{u}_m \sim \pi$ ,  $x_q = \bar{\mathbf{u}}_{m|m} \sim \pi^f$ , in (2.43)-(2.44) we obtain the following bounds on the second-order statistics of the filter estimates based on the Hellinger distance:

$$\left\| \mathbb{E}^\pi[\mathbf{u}_m] - \mathbb{E}^{\pi^f}[\bar{\mathbf{u}}_{m|m}] \right\| \leq 2 \left( \mathbb{E}^\pi[\|\mathbf{u}_m\|^2] + \mathbb{E}^{\pi^f}[\|\bar{\mathbf{u}}_{m|m}\|^2] \right)^{1/2} d_H(\pi, \pi^f), \quad (2.45)$$

and

$$\left\| \mathbb{E}^\pi[\mathbf{u}_m \otimes \mathbf{u}_m] - \mathbb{E}^{\pi^f}[\bar{\mathbf{u}}_{m|m} \otimes \bar{\mathbf{u}}_{m|m}] \right\| \leq 2 \left( \mathbb{E}^\pi[\|\mathbf{u}_m\|^4] + \mathbb{E}^{\pi^f}[\|\bar{\mathbf{u}}_{m|m}\|^4] \right)^{1/2} d_H(\pi, \pi^f). \quad (2.46)$$

Combining (2.45) and (2.46) with the first inequality in (2.41) provides an upper bound on the biases in the filter estimates for the mean dynamics and its covariance in terms of the relative entropy,  $\mathcal{P}(\pi, \pi^f)$ , which can be easily computed via (2.18) or (2.23).

Now, recall that the mutual information (2.19) can be interpreted as the lack of information in the product of the marginal densities  $\pi(\mathbf{u}_m)\pi^f(\bar{\mathbf{u}}_{m|m})$  relative to the joint density  $p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$ , which is expressed via the relative entropy as (see Section 2.2.1)

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = \mathcal{P}\left(p(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}), \pi(\mathbf{u}_m)\pi(\bar{\mathbf{u}}_{m|m})\right). \quad (2.47)$$

Thus, the bounds in (2.41) provide the lower and upper bounds on the mutual information in (2.47) and, consequently, on the pattern correlation (see Proposition 2.2). The above relationships allow for a flexible comparison of the present framework with other statistical estimates and provide a means of deriving more suitable bounds in particular applications.

**2.3.1. Some desirable properties of the Hellinger distance.** In spite of the lack of information-theoretic interpretation akin to that associated with the relative entropy (2.18), the Hellinger distance in (2.40) possesses some desirable properties which are worth mentioning here in order to foreshadow some later results associated with filtering sparsely observed systems in Section 4.2.2.

Similar to the relative entropy  $\mathcal{P}(\pi, \pi^f)$  in (2.18), the Hellinger distance  $d_H(\pi, \pi^f)$  in (2.40) is based on the marginal densities,  $\pi(\mathbf{u}_m)$ ,  $\pi^f(\bar{\mathbf{u}}_{m|m})$ , and it does not account for cross-correlations between the truth and the filter estimate. However, while the relative entropy is undefined in the limit when the filter densities  $\pi^f(\bar{\mathbf{u}}_{m|m})$  become singular, the Hellinger distance remains well-defined and bounded; this is a desirable property since it turns out that some reduced filtering strategies for sparsely observed spatially extended systems lead to singular filter densities,  $\pi^f$ , as discussed later in Section 4.2.2.

In particular, the Hellinger distance (2.40) between the Gaussian truth  $\mathbf{u}_m$  with mean  $\langle \mathbf{u}_m \rangle$  and covariance  $\mathcal{C}_{(11)}$ , and the filter estimate  $\bar{\mathbf{u}}_{m|m}$  with mean  $\langle \mathbf{u}_m \rangle$  and covariance  $\mathcal{C}_{(22)}$ , is given by

$$d_H^2\left(\pi(\mathbf{u}_m), \pi^\dagger(\bar{\mathbf{u}}_{m|m})\right) = 1 - \frac{\sqrt{2}(\det \mathcal{C}_{(11)m} \det \mathcal{C}_{(22)m})^{1/4}}{(\det [\mathcal{C}_{(11)m} + \mathcal{C}_{(22)m}])^{1/2}} \quad (2.48)$$

$$\times \exp\left\{-\frac{1}{4}(\langle \mathbf{u}_m \rangle - \langle \bar{\mathbf{u}}_{m|m} \rangle)^* (\mathcal{C}_{(11)m} + \mathcal{C}_{(22)m})^{-1} (\langle \mathbf{u}_m \rangle - \langle \bar{\mathbf{u}}_{m|m} \rangle)\right\},$$

where the coupled evolution of the truth and the filter estimate covariances,  $\mathcal{C}_{(11)}$  and  $\mathcal{C}_{(22)}$ , is given by (2.7). This standard result is obtained by direct calculation following a similar procedure that leads to the expression for the relative entropy (2.23) between two Gaussian densities. Note that if the filter estimates are correlated, i.e.,  $\det \mathcal{C}_{(22)} = 0$ , the Hellinger distance in (2.48) attains its upper bound,  $d_H = 1$ , while the relative entropy in (2.23) is undefined.

**2.4. Information optimization of imperfect Kalman filters.** The present information-theoretic framework allows for a systematic optimization of the information content in the filter estimates, as well as identification of information barriers [50, 58, 13, 10] in the imperfect Kalman filters. As discussed in Section 2.2.1, the quantification of different facets of filter skill in the present superensemble framework relies on the entropy of the filter error  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$ , the mutual information  $\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$ , and the relative entropy  $\mathcal{P}(\pi(\mathbf{u}_m), \pi^\dagger(\bar{\mathbf{u}}_{m|m}))$ , which provide measures of an ‘information distance’ between the truth  $\mathbf{u}_m$  and its filter estimate  $\bar{\mathbf{u}}_{m|m}$ . The rough guidelines for optimizing these measures in filtering are as follows:

- (i) Minimizing the entropy  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$  in (2.16) of the filter residual leads to the classical definition of filter optimality ([41, 2]) which requires minimization of path-wise filter error variance (see Section 2.2.2).
- (ii) Filters that maximize the mutual information  $\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$  in (2.19) provide the best estimates of the truth signal in terms of pattern correlation (see Section 2.2.2); this property is useful, for example, when stochastic filtering is used for deriving flux estimates from partial noisy observations [45, 12].
- (iii) The relative entropy  $\mathcal{P}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$  in (2.23) is useful for assessing the error in the statistics of the estimated signal (see also Section 2.3); filters with the smallest  $\mathcal{P}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$  yield estimates with the smallest lack of information relative to the statistics of the truth dynamics.
- (iv) Seeking the extremal values of the above information measures for the lag embedded state (2.29) is desirable when it is necessary to estimate the two-point lagged correlations between the truth and the filter estimates (e.g., in the context of forced response estimation via fluctuation-dissipation theory [52, 51, 1, 58, 13]).

It is known [14, 29, 70] that for Kalman filtering with no model error the minimum error variance filters, as in (i), also maximize the pattern correlation between the truth and the estimates, as the filters in (ii); analogous lower bounds on the filter error were derived for scalar fields from imperfect Kalman filters in [28]. However, additional barriers to achieving optimal filtering with model error arise from structural constraints in the forward dynamics of the imperfect filters. Consequently, tuning the imperfect filters to minimize the entropy of the filter error does not necessarily extremize the mutual information (or pattern correlation) and/or the relative entropy (mean biases) between the truth and the filter estimate. Thus, the practically achievable filter skill requires trade-offs between the information measures of filter fidelity,

including the RMS error in (2.34) and pattern correlation in (2.35), which are often equally important in applications (e.g., [32, 33, 61, 37, 38]).

The construction of the best imperfect filter with a given structure of the forecast model can be formulated as a constrained optimization problem over the filter parameters with the constraints given by the desired bounds on the three information measures of filter fidelity. For linear Gaussian filtering the general problem can be cast in terms of the above information measures as follows.

**DEFINITION 2.3 (Information Criterion).** *Consider the information criterion,  $\mathfrak{M}$ , of the augmented Kalman filter system (3.17) which depends on the true dynamics  $F, r, \mathcal{F}$  in (2.6), the observation noise variance  $r^o$ , and on the imperfect model  $F^M, r^M, \mathcal{F}^M$  in (2.6) so that  $\mathfrak{M}_{(F,r,\mathcal{F})}(F^M(\boldsymbol{\xi}), r^M(\boldsymbol{\xi}), \mathcal{F}^M(\boldsymbol{\xi}); r^o)$  has the following properties:*

- (i)  $\mathfrak{M}$  is nonnegative,
- (ii)  $\mathfrak{M}$  decreases monotonically with increasing  $\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$  in (2.26),
- (iii)  $\mathfrak{M}$  decays monotonically for decreasing entropy,  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$  in (2.22),
- (iv)  $\mathfrak{M}$  decays monotonically for decreasing relative entropy in (2.23),  $\mathcal{P}(\pi(\mathbb{P}^\ell \mathbf{u}_m), \pi^f(\mathbb{P}^\ell \bar{\mathbf{u}}_{m|m}))$ , on the observed modes;  $\mathbb{P}^\ell$  is the projection of the state vector onto the subset of spatial modes which are resolved by the observations.

The optimal imperfect model  $F^M(\boldsymbol{\xi}_*), r^M(\boldsymbol{\xi}_*), \mathcal{F}^M(\boldsymbol{\xi}_*)$  in (2.6) according to the information criterion  $\mathfrak{M}$  is given by

$$\mathfrak{M}_{(F,r,\mathcal{F})}(F^M(\boldsymbol{\xi}_*), r^M(\boldsymbol{\xi}_*), \mathcal{F}^M(\boldsymbol{\xi}_*); r^o) = \min_{\boldsymbol{\xi}} \mathfrak{M}_{(F,r,\mathcal{F})}(F^M(\boldsymbol{\xi}), r^M(\boldsymbol{\xi}), \mathcal{F}^M(\boldsymbol{\xi}); r^o). \tag{2.49}$$

**REMARK 2.3.**

- The above constraints on the information criterion are sufficient only in the linear Gaussian filtering problem. While the individual information measures can be applied to non-Gaussian processes with sufficiently smooth densities, the possibility of multiple extrema occurring in the information measures used in (i)-(iii) above requires much more detailed considerations.
- The use of the relative entropy in (iv) on the marginal densities associated with the observed modes stems from some technical problems associated with filtering with reduced filters through sparse observations. As shown later in Section 4.2.2 in such cases the relative entropy might be undefined for some unresolved modes, or it might be singular for the joint densities involving the resolved and unresolved modes due to correlations in filter estimates. An alternative strategy (see Section 4.4) is to use the Hellinger distance (2.40), which is well defined in such cases, instead of the relative entropy (2.23) in the criterion  $\mathfrak{M}$ ; however, in contrast to the relative entropy, the Hellinger distance between two probability densities does not have an information-theoretic interpretation.

One particular choice of the information criterion in Definition 2.3 for assessing the skill of imperfect linear Gaussian filters is given by

$$\mathfrak{M}_{(F,\mathcal{F},r)}(F^M, \mathcal{F}^M, r^M; r^o) = \frac{\exp(\mathcal{S}(\mathbf{u}_\infty - \bar{\mathbf{u}}_{\infty|\infty})) + \mathcal{P}(\pi(\mathbb{P}^\ell \mathbf{u}_\infty), \pi^f(\mathbb{P}^\ell \bar{\mathbf{u}}_{\infty|\infty}))}{\mathcal{M}(\mathbf{u}_\infty, \bar{\mathbf{u}}_{\infty|\infty})}, \tag{2.50}$$

where  $\mathbf{u}_\infty$  is the asymptotic truth state,  $\bar{\mathbf{u}}_{\infty|\infty}$  is the asymptotic the filter estimate, and the respective information measures can be expressed explicitly through  $F^M, \mathcal{F}^M, r^M, F, \mathcal{F}, r, r^o$  based on (2.6) and the definitions in Section 2.2.1 (see also Section 4.4 for some other possible choices of  $\mathfrak{M}$ ). While more suitable choices of the information criterion in (2.50) might be derived in the future, it is worth stressing that the utility of the information-based approach stems from the fact that it is not restricted to the Gaussian case and hence applicable to a much wider class of filtering problems than those considered here. This is particularly important when filtering non-Gaussian turbulent systems with sparse observations and unresolved modes contributing to intermittent energy transfers across multiple scales; relevant extensions will be reported elsewhere.

As shown in [28] for scalar fields, the lower bound on the error in imperfect Kalman estimates when filtering linear Gaussian systems is given by the error of the perfect filter (i.e.,  $F^M = F, \mathcal{F}^M = \mathcal{F}, r^M = r$  in (2.6)). Thus, the information barrier in an imperfect Kalman filter can be defined in the present Gaussian setting as follows.

**DEFINITION 2.4 (Information Barrier).** *Consider the information criterion  $\mathfrak{M}$  with the properties as in Definition 2.3 and the  $\xi$ -parametrized family of imperfect Kalman filters  $\mathcal{K} := \{F^M(\xi), \mathcal{F}^M(\xi), r^M(\xi)\}$  with the optimal imperfect filter in (2.6) according to  $\mathfrak{M}$  given by  $F^M(\xi_*), \mathcal{F}^M(\xi_*), r^M(\xi_*)$ . The information barrier for the family  $\mathcal{K}$  is defined as*

$$\mathfrak{B} = \mathfrak{M}_{(F, \mathcal{F}, r)}(F, \mathcal{F}, r; r^o) - \mathfrak{M}_{(F, \mathcal{F}, r)}(F^M(\xi_*), \mathcal{F}^M(\xi_*), r^M(\xi_*); r^o). \tag{2.51}$$

**REMARK 2.4.** Note that while the above definition depends on the choice of the information criterion  $\mathfrak{M}$ , the existence of an information barrier, in the sense that  $\mathfrak{B} \neq 0$ , does not depend on  $\mathfrak{M}$  in the Gaussian framework provided that it satisfies the conditions in Definition 2.3. However, depending on the particular application and objectives reflected in the choice of the information criterion in (2.51), the information barrier present in the problem may be negligible.

Simple examples of filter optimization are discussed in Section 3.4.2 for the case of a single Fourier mode; this basic setup illustrates a number of important issues arising in filtering with imperfect models.

### 3. Ensemble filter error for Gaussian spatially extended systems: Plentiful observations

Addressing the issue of filter stability and accuracy for a general discretized non-linear spatially extended turbulent dynamics is clearly the ‘holy grail’ of this line of research. However, it is also a formidable one and the initial step in developing a systematic framework for quantifying filter error in such systems is to consider an appropriate constant coefficient stochastic PDE test problem incorporating suitable observations [60]. This approach is justified by the fact that in many practical geophysical or engineering applications the equilibrium energy spectrum of the turbulent modes is reasonably well known and a suitable linear stochastic approximation can be considered [59, 60, 61]. Similar to the classical numerical analysis of the stability of finite difference schemes, the immediate advantages of this setup lie in the possibility of using discrete Fourier series to separate variables and reduce the analysis for a PDE system to the stability properties of matrix amplification factors for each spatial wave number separately. We begin the analysis by studying the case when the observations of the true signal are available at every node of the discretized forward

model in the filter. As already discussed in [59, 15, 60] and summarized below, in such a configuration all the Fourier modes of the truth signal that are resolved by the filter forecast model are directly observed and, for independent initial conditions, the filtering problem in the Fourier domain splits into  $N + 1$  independent filtering problems for the individual Fourier modes.

Even in this relatively simple linear configuration, there exist non-trivial effects associated with filtering the spatially extended turbulent dynamics with imperfect models which commonly arise in applications [15, 60]. It was shown earlier in [15, 37, 61, 31, 30, 11] that in this idealized setting, cheaper reduced filters with judicious model errors often have much higher skill than standard ensemble Kalman methods for filtering signals with rough turbulent spectra in physical space. Moreover, numerous later studies showed agreement of these guidelines in much more complex scenarios ranging from filtering sparsely observed atmospheric and oceanic flows in [38, 61, 60, 12], to estimating the poleward eddy heat flux in the oceans from sparse satellite altimetry data in [45]. Below, we first briefly summarize the canonical test model of [59, 15, 37, 60] and introduce a suite of imperfect models which are obtained via the finite-difference approximations of the true dynamics in (3.1). The effects of the filter error are then studied and quantified using the information measures introduced in Section 2.2. The more complicated and important case of sparse regular observations in physical space is studied in Section 4.

It is important to note that the present framework exploiting measures of information content in the filter estimates is not restricted to the Gaussian case and hence it is applicable to a much wider class of problems than those considered here.

**3.1. The canonical test model for filtering spatially extended systems in the Fourier domain.** The simplest models for representing turbulent fluctuations replace nonlinear interactions by additional linear damping and stochastic white noise forcing in time which are tuned to reproduce the equilibrium energy spectrum and turnover time for the turbulent field (e.g., [52, 56, 23, 59, 60]). Thus, based on the earlier developments in [59, 15, 37, 60], the first non-trivial test model for quantifying the Kalman filter error in turbulent spatially extended dynamical systems is to consider the following canonical linear PDE in a one-dimensional periodic domain

$$\frac{\partial u(x,t)}{\partial t} = P \left( \frac{\partial}{\partial x} \right) u(x,t) - \gamma \left( \frac{\partial}{\partial x} \right) u(x,t) + \sigma(x)\dot{W}(t) + \mathcal{F}(x,t), \quad (3.1)$$

where the operators  $P$ , and  $\gamma$ , in (3.1) are defined in the Fourier domain below. The scalar random field solutions  $u(x,t)$  of (3.1) with initial condition  $u(x,0) = u_0(x)$  are driven by the deterministic forcing  $\mathcal{F}$  and the spatially correlated, Gaussian white noise field which we write in the ‘physicist’ notation as  $\sigma(x)\dot{W}(t)$ . The minimum requirement for a global existence of stochastic solutions to (3.1), apart from constraints on  $P$ ,  $\gamma$ , and  $\mathcal{F}$ , is that the energy of the initial condition and of the solution are finite, i.e.,  $\int u_0^2 < \infty$  and  $\int u^2 < \infty$ , which impose restrictions on the noise spectrum in (3.1). However, there is a genuine physical interest in situations with a rough turbulent spectrum [59, 60, 61] where the energy spectrum is approximately constant up to a certain cut-off value in the Fourier domain; thus, similar to [30], the correlated ‘white’ noise forcing in (3.1) is understood in Fourier space as

$$\sigma(x)\dot{W}(t) := \sum_{k=-K}^K \hat{\sigma}_k \dot{W}_k(t), \quad \hat{\sigma}_k = \sigma_{-k}^*, \quad W_k = W_{-k}^*, \quad (3.2)$$

for some suitably large cut-off value  $K$  and with  $W_k(t)$  in (3.2) representing independent Wiener processes with values in the complex plane and independent components.

Here, similar to [59, 37, 60, 61], we consider a discretization of the dynamics in (3.1) on a regular mesh of points  $\{x_j = jh, j = 0, \dots, 2N + 1\}$  such that  $(2N + 1)h = 2\pi$ , so that applying the Fourier transform to (3.1) leads to the Ornstein-Uhlenbeck (OU) dynamics for each mode:

$$d\hat{u}_k(t) = [-\lambda_k \hat{u}_k(t) + \hat{\mathcal{F}}_k] dt + \hat{\sigma}_k dW_k(t), \quad \langle \hat{\sigma}_k \otimes \hat{\sigma}_{k'}^* \rangle = \delta_{k-k'} \hat{r}_k, \quad (3.3)$$

where  $\lambda_k = \hat{\gamma}_k - \hat{P}_k$  and the scalars  $\hat{\gamma}_k, \hat{P}_k$  define the operators  $\mathcal{P}, \gamma$ , of (3.1) in the Fourier domain and  $\hat{\mathcal{F}}_k$  the  $k$ -th Fourier mode of the deterministic forcing. The energy spectrum associated with the dynamics in (3.1) is given by (see, e.g., [37, 30, 60])

$$E_k = \frac{\hat{r}_k}{2\Re[\lambda_k]}. \quad (3.4)$$

Note that the energy spectrum in (3.4) for statistical equilibrium solutions of (3.1) can be tuned to mimic various turbulent spectra, ranging from the ‘equipartition’ spectrum,  $E_k \propto 1, |k| \leq K$  and  $E_k = 0, |k| \geq K$ , to steep energy spectra  $E_k \propto k^{-\beta}, \beta > 1$ . The imaginary part of  $\lambda_k$ , given by  $\Im m[\hat{P}_k]$ , can be tuned to match any physical dispersion relation [60]. Both these cases will be exploited in the tests discussed in the subsequent sections.

**3.2. Filter forecast models.** Given the linear Gaussian dynamics of the test model (3.1), the Kalman filter in (2.6) yields the optimal procedure for estimating the state of the discretized system provided that the perfect dynamics for the discretized forecast model is used and when the observability and controllability conditions are satisfied (e.g., [15, 60]). However, when filtering with imperfect models, significant errors can be introduced through the choice of discretization and the spatial and temporal time step. These issues were already studied in detail in [15, 60] and we use the same setup here in order to illustrate the utility of the information-theoretic framework for gaining new insight into the performance of imperfect Kalman filters when filtering turbulent spatially extended systems.

**PERFECT DISCRETE-TIME FORECAST MODEL**

The perfect dynamics of (3.1) in Fourier space is given by the independent Ornstein-Uhlenbeck processes (3.3). Thus, the discrete-time forward operator consistent with the dynamics in (3.3) is given by

$$\hat{u}_{k,m+1} = F_k \hat{u}_{k,m} + \hat{\mathcal{F}}_{k,m} + \hat{\sigma}_{k,m}, \quad (3.5)$$

where

$$F_k = e^{-\lambda_k \Delta t}, \quad \lambda_k = \hat{\gamma}_k - \hat{P}_k, \quad r_k = \langle \hat{\sigma}_{k,m} \otimes \hat{\sigma}_{k,m}^* \rangle = \frac{\hat{\sigma}_k^2}{2\Re[\lambda_k]} (1 - e^{-2\Re[\lambda_k] \Delta t}), \quad (3.6)$$

with  $\hat{\gamma}_k, \hat{P}_k$  representing the symbols of the pseudo-differential operators in (3.1).

**IMPERFECT DISCRETE-TIME FORECAST MODELS**

Here, the imperfect forward models introduce errors at the forecast step of the filtering procedure either through incorrect parameters and/or through finite discretizations of the correct dynamics in (3.1). As described below, the time discretizations of

(3.1) are obtained via the forward Euler, backward Euler, and the trapezoidal methods [60]. The spatial discretization of the true dynamics (3.1) are carried out using the upwind finite differences, which leads to the imperfect representation for the differential operators  $\gamma, P$ ; these imperfect operators are represented in the Fourier domain by combining the finite difference approximations with the discrete Fourier transform of the solution,  $u(x_j, t_m) = \sum_{|k| \leq N} \hat{u}_k(t_m) e^{ikx_j}$ , which leads to

$$\gamma_k^M = \sum_j \alpha_j c_k^{(j)}, \quad P_k^M = \sum_j \beta_j c_k^{(j)}. \quad (3.7)$$

where the coefficients  $c_k^{(j)}$  are defined via

$$\frac{\partial u(x, t)}{\partial x} \rightarrow \sum_{|k| \leq N} \frac{1 - e^{-ikh}}{h} \hat{u}_k = \sum_{|k| \leq N} c_k^{(1)} \hat{u}_k, \quad (3.8)$$

$$\frac{\partial^2 u(x, t)}{\partial x^2} \rightarrow \sum_{|k| \leq N} \frac{1 - 2e^{-ikh} + e^{-2ikh}}{h^2} \hat{u}_k = \sum_{|k| \leq N} c_k^{(2)} \hat{u}_k, \quad (3.9)$$

⋮

$$\frac{\partial^n u(x, t)}{\partial x^n} \rightarrow \sum_{|k| \leq N} c_k^{(n)} \hat{u}_k. \quad (3.10)$$

In particular, we consider the following four imperfect discrete-time forward models in the Fourier domain:

- Forward model with **correct dynamics for each Fourier mode**. This forecast model does not introduce discretization errors and has the same structure as the perfect system (3.6) but it allows for incorrect parameters  $\lambda_k^M, \sigma_k^M$ , in the forward operator and the model covariance

$$F_k^M = e^{-\lambda_k^M \Delta t}, \quad r_k^M = \frac{\sigma_k^{M2}}{2\Re e[\lambda_k^M]} (1 - e^{-2\Re e[\lambda_k^M] \Delta t}). \quad (3.11)$$

- Forward model obtained via **forward Euler time discretization** of (3.1) and upwind finite difference discretizations in space. This model is specified by

$$F_k^M = 1 - \lambda_k^M \Delta t, \quad R_k^M = \sigma_k^{M2} \Delta t, \quad \lambda_k^M = \hat{\gamma}_k^M - \hat{P}_k^M, \quad (3.12)$$

where  $\hat{\gamma}_k^M$  and  $\hat{P}_k^M$  have the form as in (3.7).

- Forward model obtained via **backward Euler time discretization** of (3.1) and upwind finite difference discretizations in space. This model is specified by

$$F_k^M = (1 + \lambda_k^M \Delta t)^{-1}, \quad r_k^M = |1 + \lambda_k^M \Delta t|^{-2} \sigma_k^{M2} \Delta t, \quad \lambda_k^M = \hat{\gamma}_k^M - \hat{P}_k^M, \quad (3.13)$$

where  $\hat{\gamma}_k^M$  and  $\hat{P}_k^M$  have the form as in (3.7).

- Model obtained via **trapezoidal time discretization** of (3.1) and upwind finite difference discretizations in space. This model is specified by

$$F_k^M = \frac{1 - \frac{1}{2} \lambda_k^M \Delta t}{1 + \frac{1}{2} \lambda_k^M \Delta t}, \quad r_k^M = |1 + \frac{1}{2} \lambda_k^M \Delta t|^{-2} \sigma_k^{M2} \Delta t, \quad \lambda_k^M = \hat{\gamma}_k^M - \hat{P}_k^M, \quad (3.14)$$

where  $\hat{\gamma}_k^M$  and  $\hat{P}_k^M$  have the form as in (3.7).



It is worth pointing out here that the choice of the noise amplitude,  $\sigma_k^M$ , in the imperfect forward models is not restricted to  $\sigma_k^M = \sigma_k$  and this fact was exploited in [15, 60] to optimize the imperfect filters by minimizing the lack of information in the model statistics relative to the truth. One of the advantages of the framework developed here is that it allows one to show that this method of optimization cannot improve the pattern correlation between the truth and the filter estimate.

**3.3. Augmented Kalman filter system for plentiful observations.** In most practical applications the spatially extended dynamics is observed at a regular subset of the model grid  $\{x_n = n\tilde{h}, n = 0, \dots, 2M + 1\}$  such that  $\tilde{h}/h = P \geq 1$  and the observations are given by

$$v(x_n, t_m) = Gu(x_n, t_m) + \sigma_{n,m}^o, \quad \langle \sigma_{n,m}^o \otimes (\sigma_{n',m'}^o)^* \rangle = \delta_{n-n'} \delta_{m-m'} r^o, \quad (3.15)$$

where  $G$  is a linear, possibly rank-deficient operator. The case  $P = 1$  considered in this section corresponds to *full observations* when every grid point and every resolved Fourier mode of the model is observed, i.e.,  $M = N$  and the discrete Fourier transform of (3.15) on the  $2N + 1$  node mesh leads to [59, 15, 60]

$$\hat{v}_{k,m+1} = G \hat{u}_k(t_{m+1}) + \hat{\sigma}_k^o, \quad \langle \hat{\sigma}_k^o \otimes (\hat{\sigma}_{k'}^o)^* \rangle = \delta_{k-k'} r^o / (2N + 1) \equiv \delta_{k-k'} \hat{r}_k^o. \quad (3.16)$$

Filtering the solutions of (3.1) with plentiful observations (3.16) and uncorrelated initial conditions results in independent Kalman filter systems for  $N + 1$  independent Fourier modes described by the augmented system analogous to (2.6) and given by

$$\begin{aligned} \begin{bmatrix} \hat{u}_{k,m+1} \\ \tilde{\hat{u}}_{k,m+1|m+1} \end{bmatrix} &= \begin{bmatrix} F_k & 0 \\ K_{k,m+1}^M G F_k & (\mathcal{I} - K_{k,m+1}^M G) F_k^M \end{bmatrix} \begin{bmatrix} \hat{u}_{k,m} \\ \tilde{\hat{u}}_{k,m|m} \end{bmatrix} \\ &+ \begin{bmatrix} \sigma_{k,m+1} \\ K_{k,m+1}^M (G \sigma_{k,m+1} + \sigma_{k,m+1}^o) \end{bmatrix} + \begin{bmatrix} \mathcal{F}_{k,m+1} \\ \mathcal{F}_{k,m+1}^M + K_{k,m+1}^M G (\mathcal{F}_{k,m+1} - \mathcal{F}_{k,m+1}^M) \end{bmatrix}, \end{aligned} \quad (3.17)$$

where  $\hat{u}_{k,m} \in \mathbb{C}$ ,  $|k| < N$ , represent the true Fourier modes at  $t_m$ , and  $\tilde{\hat{u}}_{k,m|m} \in \mathbb{C}$  are the respective filter estimates. The more complicated and important case of *sparse regular observations* at every  $P$ -th model grid point so that  $P > 1$  and  $M < N$  is discussed in Section 4.

Given the state vector  $\mathbf{u}_m \equiv (\hat{u}_{-N}, \dots, \hat{u}_N)$  of the truth in the Fourier domain and its filter estimates  $\tilde{\mathbf{u}}_m \equiv (\tilde{\hat{u}}_{-N}, \dots, \tilde{\hat{u}}_N)$ , we denote the augmented state of the filtered spatially extended system in the Fourier domain as  $\hat{\mathbf{y}}_m \equiv (\hat{\mathbf{u}}_m, \tilde{\hat{\mathbf{u}}}_m)^T$ . The Gaussian statistics of the augmented state  $\hat{\mathbf{y}}_m$  is fully characterized by its mean  $\langle \mathbf{y}_m \rangle$  and covariance  $\mathcal{C}_m \equiv Cov(\mathbf{y}_m, \mathbf{y}_m)$ , with the evolution of the latter given by

$$\mathcal{C}_{m+1} = \mathcal{F}_m \mathcal{C}_m \mathcal{F}_m^* + \mathcal{R}_m, \quad (3.18)$$

where the operators  $\mathcal{F}$  and  $\mathcal{R}$  in (3.18) are defined analogously to (2.8) and (2.9). For plentiful observations only the  $k$ -th Fourier mode  $\hat{u}_k$  and its estimate  $\tilde{\hat{u}}_k$  are correlated through the dynamics of (3.17) so that all the blocks in the covariance in (3.18),

$$\mathcal{C}_m = \begin{bmatrix} Cov(\mathbf{u}_m, \mathbf{u}_m) & Cov(\mathbf{u}_m, \tilde{\mathbf{u}}_m) \\ Cov(\tilde{\mathbf{u}}_m, \mathbf{u}_m) & Cov(\tilde{\mathbf{u}}_m, \tilde{\mathbf{u}}_m) \end{bmatrix} \equiv \begin{bmatrix} \mathcal{C}_{(11)m} & \mathcal{C}_{(12)m} \\ \mathcal{C}_{(12)m}^* & \mathcal{C}_{(22)m} \end{bmatrix}, \quad (3.19)$$

are diagonal, i.e.,

$$\mathcal{C}_{(11)m} = \text{diag}[\mathcal{C}_{(11)-N,m}, \dots, \mathcal{C}_{(11)N,m}],$$

$$\mathcal{C}_{(22)m} = \text{diag}[\mathcal{C}_{(22)-N,m}, \dots, \mathcal{C}_{(22)N,m}],$$

$$\mathcal{C}_{(12)m} = \text{diag}[\mathcal{C}_{(12)-N,m}, \dots, \mathcal{C}_{(12)N,m}],$$

where  $\mathcal{C}_{(11)k,m} \equiv \text{Cov}(\hat{u}_{k,m}, \hat{u}_{k,m})$ ,  $\mathcal{C}_{(22)k,m} \equiv \text{Cov}(\tilde{u}_{k,m|m}, \tilde{u}_{k,m|m})$ , and  $\mathcal{C}_{(12)k,m} \equiv \text{Cov}(\hat{u}_{k,m}, \tilde{u}_{k,m|m})$ . The diagonality of the cross correlation matrix  $\mathcal{C}_{(12)}$  and the filter estimates covariance  $\mathcal{C}_{(22)}$  stems from the plentiful observations while the diagonality of  $\mathcal{C}_{(11)}$  arises due to the linearity of the truth dynamics.

Consequently, the mutual information in (2.19) which measures the reduction of uncertainty due to the filtering can be expressed as a sum of components involving covariances between the individual modes:

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = -\frac{1}{2} \sum_{|k| \leq N} \ln [1 - \lambda_{k,m}^2], \quad \lambda_{k,m} = \sqrt{\frac{|\mathcal{C}_{(12)k,m}|^2}{\mathcal{C}_{(22)k,m} \mathcal{C}_{(11)k,m}}}, \quad 0 \leq \lambda_{k,m} \leq 1, \quad (3.20)$$

with the coefficients  $\mathcal{C}_{(11)k,m} \equiv \text{Cov}(\hat{u}_{k,m}, \hat{u}_{k,m})$ ,  $\mathcal{C}_{(22)k,m} \equiv \text{Cov}(\tilde{u}_{k,m}, \tilde{u}_{k,m})$ , and  $\mathcal{C}_{(12)k,m} \equiv \text{Cov}(\hat{u}_{k,m}, \tilde{u}_{k,m})$ . Note that in the asymptotic limit  $t_m \rightarrow \infty$ ,  $\lambda_{k,\infty}$  in (3.20) becomes the pattern correlation (2.32) between the true and filtered  $k$ -th Fourier mode, assuming filter stability. Similarly, the entropy of the filter error (2.16) simplifies to

$$\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}) = (1 + \ln 2\pi) + \sum_{|k| \leq N} \ln [\mathcal{C}_{(11)k,m} + \mathcal{C}_{(22)k,m} - 2\Re e[\mathcal{C}_{(12)k,m}]], \quad (3.21)$$

and the relative entropy (2.18), measuring the lack of information in the probability density of the filter estimates relative to the statistics of the truth, can be decomposed as

$$\begin{aligned} \mathcal{P}(\pi(\mathbf{u}_m), \pi^{\dagger}(\bar{\mathbf{u}}_{m|m})) &= \frac{1}{2} \sum_{|k| \leq N} |\langle u_{k,m} \rangle - \langle \bar{u}_{k,m|m} \rangle|^2 \mathcal{C}_{(22)k,m}^{-1} - (2N+1) \\ &\quad + \frac{1}{2} \sum_{|k| \leq N} \left[ \left( \mathcal{C}_{(11)k,m} \mathcal{C}_{(22)k,m}^{-1} \right) - \ln \left( \mathcal{C}_{(11)k,m} \mathcal{C}_{(22)k,m}^{-1} \right) \right]. \end{aligned} \quad (3.22)$$

We use (3.20) and (3.21) to quantify the superensemble model error in filtering in the following sections.

**3.4. Filtering a single Fourier mode.** Here, we utilize the plentiful observation setup outlined in Section 3.1 and study the ensemble error in filtering a single Fourier mode using the imperfect forecast models (3.11)-(3.14) in (3.17). Recall that the filtering of (3.1) discretized on a spatial grid with  $2N+1$  nodes with plentiful observations and uncorrelated initial conditions splits into  $N+1$  independent filtering complex scalar problems for the individual Fourier modes given by (3.17). First, we focus on the model error due to the temporal discretizations in the imperfect models in the absence of errors due to the spatial discretization (i.e.,  $\hat{\gamma}_k^M = \hat{\gamma}_k$  and  $\hat{P}_k^M = \hat{P}_k$  in (3.11)-(3.14)). The case of filtering the full spatially extended system with plentiful observations is discussed subsequently in Section 3.5 and it draws from the single mode analysis. We use this simple configuration to illustrate the importance of applying more general error measures in imperfect filtering than the mean square error estimates (2.30) which are often insensitive to the information barriers for imperfect filtering improvement. Below we discuss simple examples of such information barriers

which show unambiguously that for imperfect filtering it is often much more advantageous to judiciously modify the filter forecast model than trying to optimize the filter parameters.

**3.4.1. Asymptotic filter error.** Filtering the linear Gaussian test model (3.1) with plentiful observations is equivalent to independent filtering of the dynamics of each Fourier mode in (3.3). In this case the information measures of filter skill, the error entropy (2.16), mutual information (2.19), and relative entropy (2.23), assume a particularly simple and revealing form. This complex scalar configuration, exploited earlier in [60] in the path-wise framework, allows for direct analysis of filter stability and accuracy based on explicit formulas for the asymptotic covariance of the augmented system (3.17). In particular, this framework is very useful for illustrating the issues of optimality for imperfect filters discussed earlier in Section 2.4 and subsequently in Section 3.4.2.

In the present complex Gaussian scalar case the filter error entropy (2.16) becomes

$$\mathcal{S}(\mathcal{U}_m) = \frac{1}{2}(1 + \ln 2\pi) + \frac{1}{2} \ln [\mathcal{C}_{(11)m} + \mathcal{C}_{(22)m} - 2\Re[\mathcal{C}_{(12)m}]], \tag{3.23}$$

where the evolution of the covariance  $\mathcal{C}_m$  is given by (3.18) with  $\mathcal{C}_{(11)m} + \mathcal{C}_{(22)m} \geq 2\Re[\mathcal{C}_{(12)m}]$ . The mutual information in (2.26) simplifies to

$$\mathcal{M}(u_m, \bar{u}_{m|m}) = \frac{1}{2} \ln \frac{1}{1 - \lambda_m^2}, \quad \lambda_m = \sqrt{\frac{|\mathcal{C}_{(12)m}|^2}{\mathcal{C}_{(11)m}\mathcal{C}_{(22)m}}}, \quad 0 \leq \lambda_m \leq 1, \tag{3.24}$$

and the relative entropy in (2.23) simplifies to

$$\mathcal{P}(\pi(u_m), \pi^\dagger(\bar{u}_{m|m})) = \frac{|\bar{u}_m - \bar{u}_{m|m}|^2}{\mathcal{C}_{(22)m}} + \frac{\mathcal{C}_{(11)m}}{\mathcal{C}_{(22)m}} - 1 - \ln \frac{\mathcal{C}_{(11)m}}{\mathcal{C}_{(22)m}}. \tag{3.25}$$

Provided that the filtering problem in (3.17) is stable (i.e., the system (3.17) has a stable fixed point), exact formulas for the above information measures can be obtained in the limit  $m \rightarrow \infty$  since it can be easily shown by direct calculation that the components of the asymptotic covariance  $\mathcal{C}_\infty$  in (3.18) satisfying

$$\mathcal{C}_\infty = \mathcal{F}\mathcal{C}_\infty\mathcal{F}^T + \mathcal{R} \tag{3.26}$$

are given by

$$\mathcal{C}_{(11)\infty} = \frac{r}{1 - |F|^2}, \quad \mathcal{C}_{(12)\infty} = \frac{GK_\infty^M (|F|^2\mathcal{C}_{(11)\infty} + r)}{1 - FF^{M*}(1 - GK_\infty^M)}, \tag{3.27}$$

$$\begin{aligned} \mathcal{C}_{(22)\infty} = & (1 - |F^M|^2(1 - GK_\infty^M)^2)^{-1} \left( |F|^2G^2(K_\infty^M)^2\mathcal{C}_{11,\infty} \right. \\ & \left. + GK_\infty^M(1 - GK_\infty^M)(FF^{M*}\mathcal{C}_{(12)\infty} + F^MF^*\mathcal{C}_{(12)\infty}^*) + (K_\infty^M)^2(r^o + G^2r) \right), \end{aligned} \tag{3.28}$$

The component  $\mathcal{C}_{(11)\infty}$  is the asymptotic covariance of the truth,  $\mathcal{C}_{(22)\infty}$  is the asymptotic covariance of the filter estimates, and  $\mathcal{C}_{(12)\infty}$  contains the asymptotic cross-correlations between the truth and the filter estimates. The asymptotic imperfect

Kalman gain in (3.27)-(3.28) obtained via the approximation in (2.13) is given by

$$K_\infty^M = \frac{1}{2G} \left( 1 - \frac{G^2 r^M}{|F^M|^2 r^o} - \frac{1}{|F^M|^2} + \left[ \left( 1 - \frac{G^2 r^M}{|F^M|^2 r^o} - \frac{1}{|F^M|^2} \right)^2 + \frac{4G^2 r^M}{|F^M|^2 r^o} \right]^{1/2} \right), \quad (3.29)$$

which was derived in detail in [60].

The optimal values of the filter error entropy in (3.23), mutual information in (3.24), and the relative entropy (3.25) in imperfect filtering are not necessarily achieved for the same imperfect forecast model. In particular, note that high uncertainty reduction in terms of mutual information does not imply achieving accurate estimate of the model covariance; indeed, good pattern correlation,  $\lambda_\infty \sim 1$  in (3.24), can be achieved for imperfect filters with erroneous forecast model variance,  $\mathcal{C}_{(22)\infty} \gg \mathcal{C}_{(11)\infty}$  or  $\mathcal{C}_{(22)\infty} \ll \mathcal{C}_{(11)\infty}$ , provided that  $|\mathcal{C}_{(12)\infty}|^2 \sim \mathcal{C}_{(11)\infty} \mathcal{C}_{(22)\infty}$ . Moreover, it is clear from inspecting the dynamics of (3.17) that large biases in the ensemble mean,  $|\langle u_m \rangle - \langle \bar{u}_m | m \rangle|^2$ , are possible even for statistically consistent models,  $\mathcal{C}_{(11)\infty} \sim \mathcal{C}_{(22)\infty}$ , minimizing the dispersion part of the relative entropy in (3.25). In the next section we use the simple exactly solvable complex scalar configuration to present a few concrete examples of the issues discussed in Section 2.4; namely the lack of a single imperfect model optimizing all the different measures of filter skill which leads to the existence of information barriers in families of imperfect filters.

### 3.4.2. Optimizing imperfect Kalman filters for single Fourier mode.

The general idea of information optimization of imperfect Kalman filter was discussed in Section 2.4. Here, we exploit the simple complex scalar configuration associated with plentiful observations which allow for independent filtering of individual Fourier modes. We use this simple setup and the exact formulas for the asymptotic covariance in filtering problem (3.18) in order to illustrate the lack of a single imperfect filter optimizing all the information measures discussed in Section 2.2.1. The possible lack of a single best filter when confined to a class of imperfect forecast models drives the need for synergistic use of the entropy, mutual information, and the relative entropy in optimizing imperfect filtering with model-induced constraints.

The imperfect forecast models used in the filters for estimating the true dynamics of the individual Fourier modes  $\hat{u}_k$  in (3.17) are described in Section 3.2. The optimal values of model noise,  $r^{M*}$ , in the imperfect filters are computed according to the criterion (2.50) which for the scalar case becomes

$$\begin{aligned} \mathfrak{M} \left( F^M(\boldsymbol{\xi}), r^M(\boldsymbol{\xi}); F, r, r^o \right) &= \frac{1}{2} \ln \frac{\mathcal{C}_{(11)\infty} \mathcal{C}_{(22)\infty}}{\mathcal{C}_{(11)\infty} \mathcal{C}_{(22)\infty} - |\mathcal{C}_{(12)\infty}|^2} \\ &\times \left( \mathcal{C}_{(11)\infty} + \mathcal{C}_{(22)\infty} - 2\Re[\mathcal{C}_{(12)\infty}] + \frac{|\langle u_\infty \rangle - \langle \bar{u}_\infty | \infty \rangle|^2}{\mathcal{C}_{(22)\infty}} + \frac{\mathcal{C}_{(11)\infty}}{\mathcal{C}_{(22)\infty}} - \log \frac{\mathcal{C}_{(11)\infty}}{\mathcal{C}_{(22)\infty}} - 1 \right), \end{aligned} \quad (3.30)$$

with the expressions for the asymptotic covariance components of  $\mathcal{C}_\infty$  (3.18) given by (3.27)-(3.28).

Figure 3.1 illustrates optimization of imperfect Kalman filters for a single Fourier mode in the spatially extended dynamics of (3.1) with no deterministic forcing and zero mean; the top four panels show the time-asymptotic entropy of the filter error,  $\mathcal{S}(u_\infty - \bar{u}_\infty | \infty)$  in (3.23), and mutual information,  $\mathcal{M}(u_\infty, \bar{u}_\infty | \infty)$  in (3.24), for the imperfect filters discussed in Section 3.2 with model error due to various time discretizations. Unsurprisingly, in the case of the filter in (3.11) without time-discretization

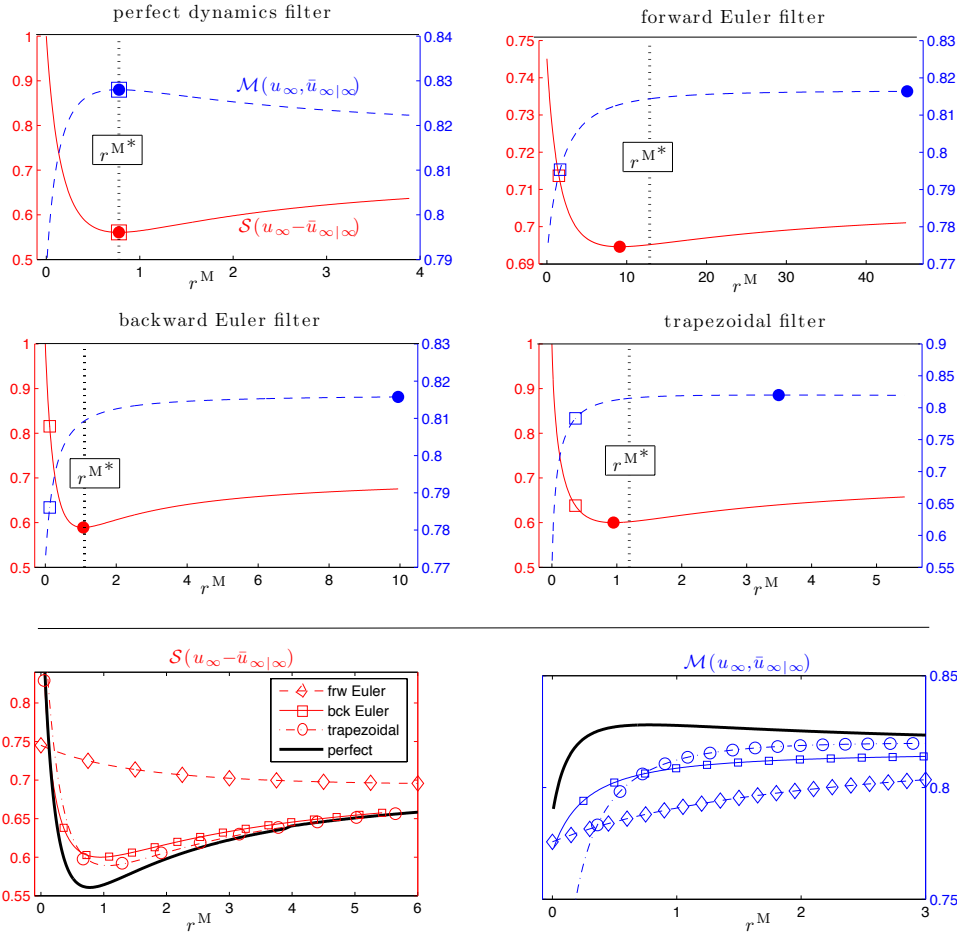


FIG. 3.1. Entropy of filter error  $S$  ( $\sim$  RMS error) and mutual information  $M$  ( $\sim$  pattern correlation) for filtering the complex scalar with imperfect filters (3.12)-(3.14) as a function of the forecast model variance  $r^M$  in the non-stiff regime with  $\lambda = \gamma - i\omega$ ,  $\gamma = 0.5, \omega = 2, \sigma = 1$ , large observation time  $\Delta t^{obs} = 1/\tau_{corr} = 2$ , and observation noise  $r^o = 0.25$ . Unoptimized values of error entropy  $S$  and mutual information  $M$  for the imperfect filters are denoted by the squares; the optimal values of the individual measures are denoted by the dots. Note that for the imperfect filters the optimal error entropy/RMSE does not coincide with optimal mutual information/pattern correlation. The dashed vertical lines indicates the optimal model variance according to the information criterion  $\mathfrak{M}$  in (2.50).

errors the minimum of  $\mathfrak{M}$  in (3.30) coincides with the parameter values as in perfect dynamics, i.e.,  $\lambda^M = \lambda$ ,  $\sigma^M = \sigma$ . Importantly, the optimal values of the error entropy/RMSE (2.22) and the mutual information/pattern correlation (3.24) coincide with the parameter values of the perfect dynamics. For the remaining imperfect filters the extrema of the filter error entropy, mutual information, and the information criterion  $\mathfrak{M}$  generally do not coincide. Note that for parameter values as in the perfect dynamics (squares) the skill of the imperfect filters with unoptimized noise (squares) is significantly reduced relative to the optimal values derived from (3.30). The bottom two panels in figure 3.1 show, respectively the discrepancies between the

entropy of the filter error for the perfect and imperfect models (left), and the mutual information discrepancies between the perfect and imperfect filters as a function of the system noise variance. Clearly, none of the imperfect models considered here can achieve the accuracy of the perfect model which gives rise to the information barriers (2.51) within the  $r^M$ -parametrized families of the imperfect time-discretized models. Note, in particular, that it is possible to have a filter with very poor statistical fidelity of the estimated state but with good performance according to the filter error entropy and the mutual information.

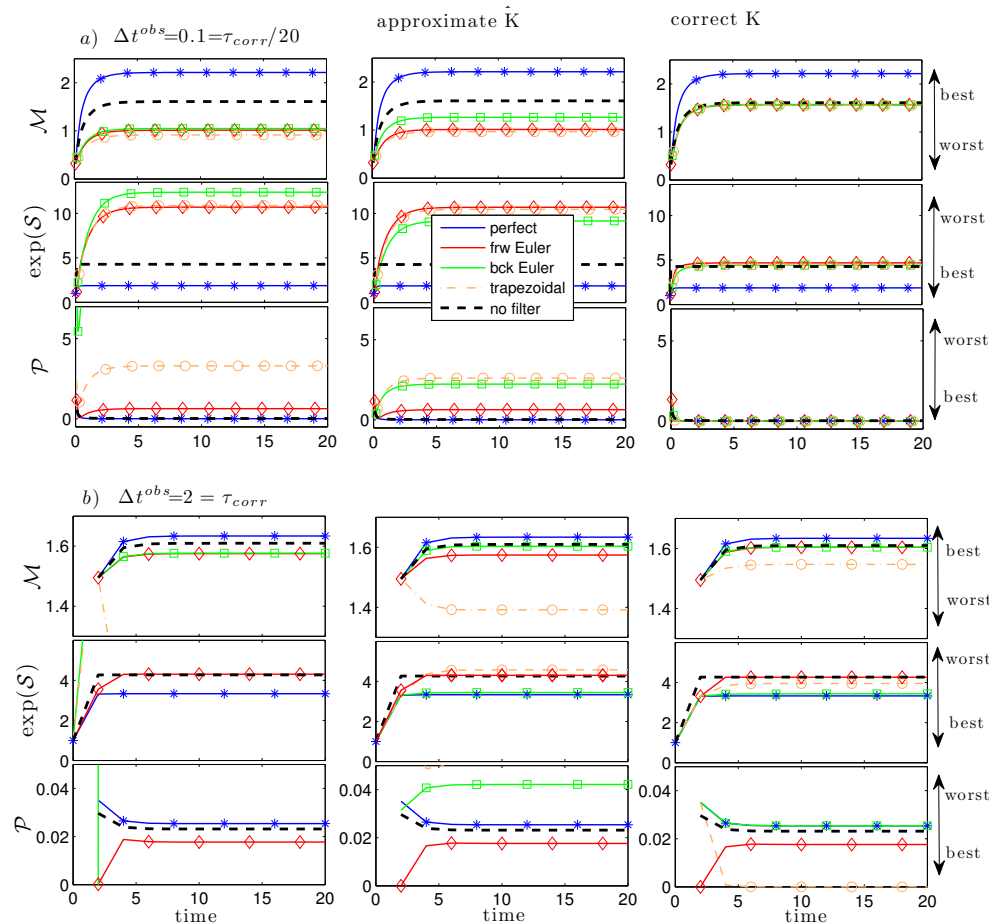


FIG. 3.2. Time evolution of the information measures introduced in Section 2.2.1 for assessing the skill of imperfect filtering; here the skill for filtering a single Fourier mode (i.e., complex scalar) in (3.3) with model error due to different time discretizations (see legend) and Section 3.2 is shown for stiff discrete-time dynamics (3.5) with  $\gamma=0.5, \omega=10, \sigma=1, r^o=0.25$ . The dashed black lines correspond to the information measures computed directly from the noisy input, i.e., they are computed without any filter. Columns corresponds to, respectively, filtering with unoptimized imperfect models (left), imperfect models with approximate Kalman gain (2.13) which are optimized via the simple noise inflation (middle), and imperfect models with correct Kalman gain (2.11); results in the last two columns cannot be achieved without the knowledge of the truth but they illustrate the potential performance gains assuming that suitable approximations are implemented.

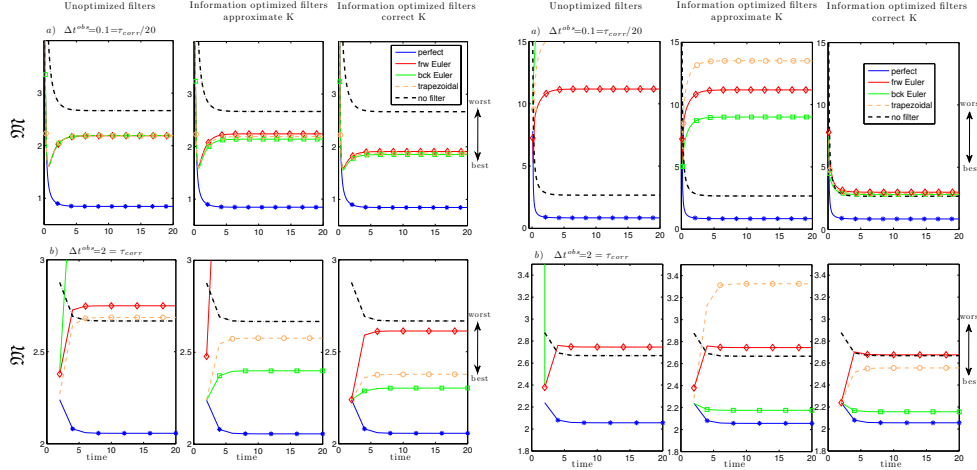


FIG. 3.3. Time evolution of the skill of imperfect filters according to the information criterion  $\mathfrak{M}$  in (2.50) for filtering the dynamics of a single Fourier mode (complex scalar) in (3.3) with model error due to various time discretizations; see figure (3.2) for the individual information measures. (Left) non-stiff dynamics with  $\gamma = 0.5, \omega = 0.7, \sigma = 1, r^o = 0.25$ , (right) stiff dynamics with  $\gamma = 0.5, \omega = 10, \sigma = 1, r^o = 0.25$ .

**3.4.3. Numerical tests in the one-mode configuration.** Here, we present numerical examples of information-theoretic quantification of filter error within the superensemble framework for a single Fourier mode (3.3) in the spatially extended discretized dynamics of (3.1). Sections 3.5.1 and 4.3 provide numerical examples of quantifying filter error in more complex configurations from filtering the spatially extended turbulent dynamics with plentiful observations in Section 3.5.1 to filtering such systems when only sparse observations are available in Section 4.3.

Filtering the single Fourier mode in the dynamics of (3.1) is equivalent to filtering a complex scalar discussed earlier in [60]. We assume for simplicity that the imperfect forecast models for filtering the single Fourier mode introduce model error only through the temporal discretization of the dynamics in (3.1) with the perfect spectral decomposition as in (3.3). The superensemble filter error is quantified via the mutual information (3.24), the filter error entropy (3.23), and the relative entropy (2.23). In order to allow for easy comparison with the path-wise analysis of the same problem discussed in [60], we focus on two distinct dynamical regimes of (3.3) with parameters

$$a) \text{ Stiff reg.: } \gamma = 0.5, \omega = 10, \sigma = 1, \quad b) \text{ Non-stiff reg.: } \gamma = 0.5, \omega = 0.7, \sigma = 1. \quad (3.31)$$

Both these regimes are characterized by the same decorrelation time  $\tau_{corr} = 1/\gamma = 2$  and equilibrium variance  $\sigma^2/(2\gamma) = 1$ , but the stiff regime contains fast oscillatory dynamics. In figures 3.2-3.5 different facets of filter error are examined in the above two dynamical regimes of (3.3) for various observation times  $\Delta t^o$  and observation noise variances  $r^o$ . By analogy to [60], in all these cases we compare the imperfect filter skill for (i) unoptimized filters with parameter values in (3.11)-(3.14) set to the perfect dynamics parameters in (3.31a) or (3.31b), or (ii) imperfect filters with the Kalman gain approximated by (2.13) and model noise  $r^m$  in (3.12)-(3.14) optimized to minimize the asymptotic filter error covariance in terms of the relative entropy in (2.23) (see [60, Section 2.3.4] for details); such an optimization provides the best possible imperfect



filter skill in the absence of mean biases in the filter estimates. We also consider the skill of imperfect filters (3.12)-(3.14) with correct Kalman gain obtained from (2.11). Similar to the information optimization of the filter error covariance this last strategy is not a practical solution since it requires the knowledge of the truth; however, such tests provide guidelines for improving the imperfect filters.

In figure 3.2 we show the time evolution of the error for filtering the dynamics of the single Fourier mode (3.3) in the stiff regime with parameters (3.31a). The filter error is quantified in terms of the three individual information measures  $\mathcal{M}(u_m, \bar{u}_{m|m})$ ,  $\mathcal{S}(u_m - \bar{u}_{m|m})$ ,  $\mathcal{P}(\pi(u_m), \pi^f(\bar{u}_{m|m}))$  discussed in Section 2.2.1; recall that the filter error entropy in (2.22) is linked to the RMS error (2.30) in the absence of mean biases (Proposition 2.1) and the mutual information in (2.26) is linked to the asymptotic pattern correlation (Proposition 2.2). The model error in the imperfect filters (3.12)-(3.14) is introduced via the finite difference approximations of the temporal evolution in (3.1). Figure 3.3 shows the filter skill for the same filters and filter parameters as in figure 3.2 but in terms of the information criterion (2.50) which combines the three information measures.

The most comprehensive insight into the imperfect filter skill is obtained via the time-lagged information criterion  $\mathfrak{M}(\pi(u_\infty^L), \pi(\bar{u}_{\infty|\infty}^L))$  in (2.50) based on the lag-embedded truth and its lag-embedded filter estimate discussed in Section 2.2.1 and defined in (2.29); recall that the lag-embedded state with  $L > 1$  accounts for correlations in the truth and the filter estimate and the information measures applied to the lag-embedded states provide a nonlocal-in-time characterization of filter error. This is particularly important in development and validation of physics constrained multi-level nonlinear regression models [54, 57] which simultaneously reflect the causality and energy conserving principles of the underlying nonlinear physics and, by design, mitigate the non-physical finite-time blow up or pathology present in the invariant measure of ad-hoc quadratic regression strategies (e.g., [48]). Figures 3.4 and 3.5 illustrate different aspects of the the nonlocal-in-time filter error for the imperfect filters with forward dynamics (3.11)-(3.14). In figure 3.4 the two representative stiff and non-stiff cases with parameters (3.31a), (3.31b) are examined. In both regimes the simple information optimization achieved through minimization of the relative entropy between the truth and the filter estimates does not improve the filter skill for short observation time ( $\Delta t^{obs} \ll \tau_{corr}$ ). On the other hand, for long observation times ( $\Delta t^{obs} \sim \tau_{corr}$ ) large improvements can be achieved for both the backward Euler and the trapezoidal filters; the perfect dynamics and the forward Euler filters are unaffected by the simple information optimization. The lag-embedded measures of filter skill are very sensitive to the observation noise levels and quickly deteriorate for  $L > 1$  with increasing values of  $\hat{r}^o$ .

Finally, in figure 3.5 we focus on the recovery of the autocorrelation functions by the imperfect filters; these important statistical properties can be easily extracted from the lagged covariance matrices computed from (2.6) using the relationships obtained in Appendix A. Accurate recovery of the correlation functions by the filter is crucial in many geophysical and engineering applications and it allows for constructing robust reduced stochastic models from data of nonlinear systems [54, 57], and for accurate prediction of forced response of the truth system via the fluctuation-dissipation theorem (e.g., [52, 51, 1, 58, 13]). The four panels in figure 3.5 show the correlation functions estimated from the perfect filter (3.11) and the imperfect filters (3.12)-(3.14). Note that the perfect filter and the forward and backward Euler filters recover the normalized autocorrelation comparably well; however, the backward

Filtering complex scalar dynamics		
	ensemble avg. RMSE	path-wise RMSE
perfect filter	0.4418	0.4450
frw. Euler filter	0.5018	0.5164
bckw. Euler filter	0.9802	0.9609
trapezoidal filter	0.8450	0.8464
frw. Euler filter/info optim.	0.5018	0.4997
bckw. Euler filter /info optim.	0.4464	0.4564
trapezoidal filter /info optim.	0.4945	0.4859
Filtering spatially extended dynamics		
	ensemble avg. RMSE	path-wise RMSE
perfect filter	27.2709	27.0783
frw Euler filter	31.5690	31.3023
bck Euler filter	63.1765	64.0741
trapezoidal filter	62.7162	62.8249
frw Euler filter /info optim.	31.5634	31.6274
bck Euler filter /info optim.	28.7110	28.4654
trapezoidal filter /info optim.	31.7556	31.1732

TABLE 3.1. Numerical evidence in support of the attractor ergodicity assumption which allows for connecting the asymptotic path-wise averages with their ensemble counterparts in Section 2.2.2. The tables above compare the asymptotic RMS error estimated from the path-wise averages (1000 assimilation cycles) and the ensemble RMS error for the complex scalar (top) and the spatially extended system (3.32) with  $N=20$  and plentiful observations and equipartition energy spectrum  $E_k \propto 1$ .

Euler filter grossly underestimates the variance of the truth. The filter based on the trapezoidal discretization underestimates the true variance and introduces spurious oscillations in the estimated autocorrelation function.

In summary, we make the following points in regard to filtering a single Fourier mode which further reinforce the conclusions of the path-wise analysis of [60]:

- The ensemble error in the filter estimates can be computed via the ergodic averages on the attractor of the augmented system (3.17); good agreement between the ensemble averages and time-averages was verified by direct computations (see table 3.1).
- When filtering with imperfect models the ‘minimum error variance’ filter (the filter with minimum error entropy) does not necessarily provide optimal estimates in terms of mutual information/pattern correlation (figure 3.1).
- The lag-embedded information measures of filter skill are very sensitive to the observation noise values and the filter skill quickly deteriorates for  $L > 1$  and for increasing  $\hat{r}^\circ$  (see figure 3.4).
- The performance of the imperfect filters with forecast models (3.12)-(3.14) depends on the interplay between the model stability and the filter stability and controllability properties (e.g., [2, 60]); the main features of filter performance are loosely summarized as follows:
  - The stability of the forecast model in the forward Euler filter is usually violated, except the case of non-stiff dynamics and sufficiently small

observation times  $\Delta t^{obs}$  so that  $|F_k^M| < 1$  in (3.12). When the forecast model is unstable ( $|F_k^M| > 1$ , most of the cases considered in figures 3.2 and 3.3) the unoptimized filter trusts the observations; thus, the forward Euler filter is insensitive to the system noise inflation (stiff dynamics) or loses practical controllability (non-stiff dynamics), as shown in figure 3.3. The performance of this filter does improve in the idealized scenario with correct Kalman gain (2.11), as shown in figure 3.3.

- The unoptimized filter with the backward Euler discretization method in the forecast model (3.13) weighs the filtered solution almost fully to the dynamics with model error, except the case of non-stiff dynamics and sufficiently small observation times  $\Delta t^{obs}$ . The model variance  $r_k^M$  and  $|F_k^M|$  in (3.13) decrease quickly for increasingly stiff dynamics, making this filter lose practical controllability (e.g., [2]). This filter strongly benefits from the noise inflation which improves its controllability; it also performs well in the idealized case with correct Kalman gain (2.11); see figures 3.3, 3.4.
- The unoptimized filter utilizing the unconditionally stable trapezoidal discretization method in the forecast model (3.14) also performs poorly in the stiff dynamics regime due to the loss of practical controllability since the increased accuracy of the discretization in (3.14) leads to very small system model variance  $r_k^M$ ; moreover, the asymptotic stability factor for the mean  $|F^M(1 - K_\infty^M g)|$  (see [2, 60]) is near one in the stiff regime. This filter strongly benefits from the noise inflation which improves its controllability; it also performs well in the idealized case with correct Kalman gain (2.11), as shown in figures 3.3, 3.4.
- The unoptimized forward Euler filter and the backward Euler filter have a similarly good skill for recovering the normalized correlation functions (figure 3.5); however, the backward Euler filter vastly underestimates the truth variance.

More detailed hierarchy of filter skill according to the criterion  $\mathfrak{M}$  in (3.30) is summarized below (figure 3.3).

• **UNOPTIMIZED FILTER SKILL**

*Non-stiff regime:* For short observation times  $\Delta t^{obs} \ll \tau_{corr}$  all imperfect filters are similar and beat raw observations but are significantly worse than the perfect filter. For long observation times  $\Delta t^{obs} \sim \tau_{corr}$  the unoptimized imperfect filters perform poorly.

*Stiff regime:* only the perfect filter beats direct observations.

• **FILTER SKILL AFTER THE SIMPLE INFORMATION OPTIMIZATION** (i.e., filter (3.17) with noise-inflated  $K_m^M$  in (2.13); knowledge of truth required [37, 60]).

*Non-stiff regime:* For short observation times  $\Delta t^{obs} \ll \tau_{corr}$  all imperfect filters are similar and beat observations but are significantly worse than the perfect filter. For long observation times,

perfect filter (3.11) > backward Euler (3.13) > trapezoidal (3.14) > raw obs.,

while the forward Euler in (3.12) loses practical controllability.

*Stiff regime:* For  $\Delta t^{obs} \ll \tau_{corr}$  only the perfect filter outperforms observations; for  $\Delta t^{obs} \sim \tau_{corr}$  the backward Euler filter performs comparably to the

perfect filter while the forward Euler and the trapezoidal filters do not beat the observations.

- **FILTER SKILL WITH CORRECT KALMAN GAIN** (i.e., filter (3.17) with  $K_m$  in (2.11); knowledge of truth required)

*Non-stiff regime:* All filters beat raw observations. For  $\Delta t^{obs} \ll \tau_{corr}$  all imperfect filters are similar but they perform worse than the perfect filter; for  $\Delta t^{obs} \sim \tau_{corr}$ , the filter skill satisfies

perfect filter > backward Euler > trapezoidal > forward Euler > raw obs.

*Stiff regime:* For  $\Delta t^{obs} \ll \tau_{corr}$  only the perfect filter performs better than raw observations; for  $\Delta t^{obs} \sim \tau_{corr}$ , the filter skill satisfies

perfect filter > backward Euler > trapezoidal > raw obs. > forward Euler.

Next, we extend the the off-line assessments of the filtering skill for a single Fourier mode to the spatially extended turbulent dynamics case which factorizes to independent filtering problems of each Fourier mode due to the plentiful observations.

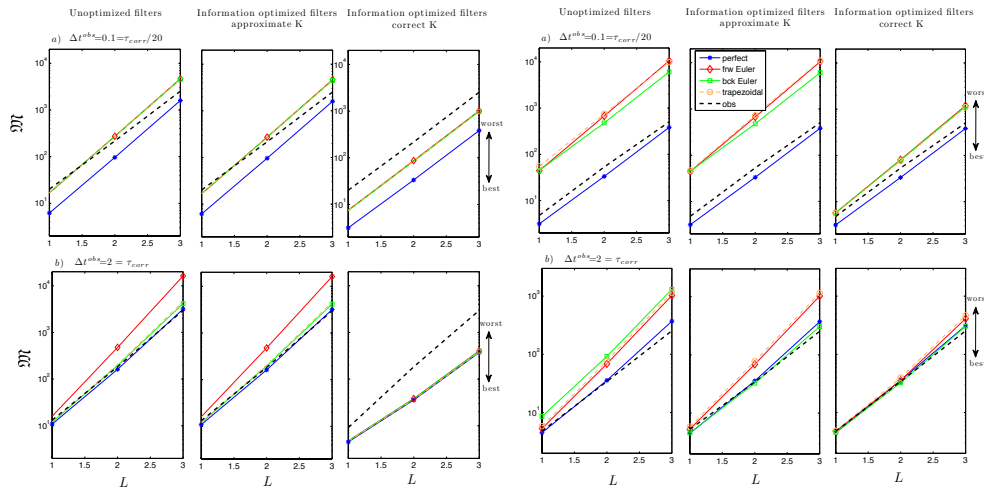


FIG. 3.4. Imperfect filter skill for estimating the dynamics of a single Fourier mode in (3.3) computed via the lag-embedded information criterion in (2.50) (see also (2.29)); the skill is shown as a function of the time lag  $L$ . (Left) non-stiff dynamics with  $\gamma=0.5, \omega=0.7, \sigma=1, r^o=0.05$ , (right) stiff dynamics with  $\gamma=0.5, \omega=10, \sigma=1, r^o=0.05$ . The lag-embedded skill is very sensitive to the observation noise values and quickly deteriorates for  $L > 1$  for increasing  $r^o$ .

### 3.5. Filtering spatially extended systems with plentiful observations.

In order to illustrate the present information-theoretic framework for assessing the superensemble filter error on a concrete spatially extended example, we consider filtering the dynamics of the one-dimensional stochastically forced advection-diffusion equation

$$\frac{\partial u(x,t)}{\partial t} = -du(x,t) - c \frac{\partial u(x,t)}{\partial x} + \mu \frac{\partial^2 u(x,t)}{\partial x^2} + \mathcal{F}(x,t) + \sigma(x)\dot{W}(t), \quad (3.32)$$

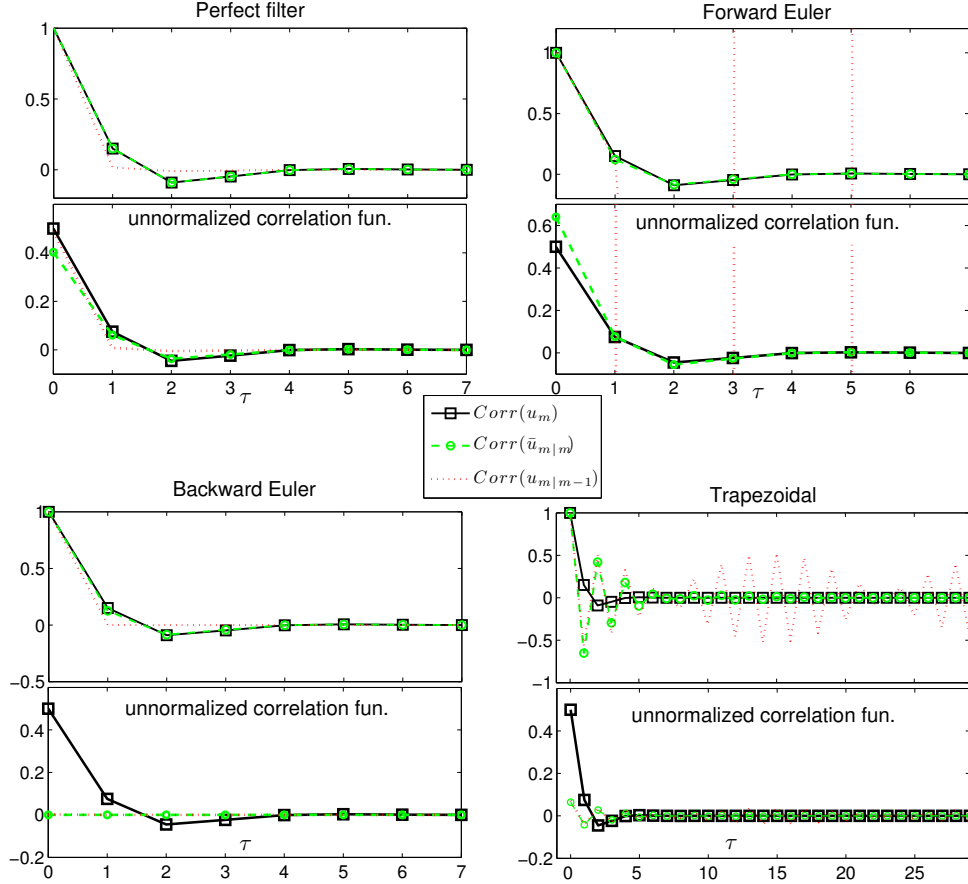


FIG. 3.5. Recovery of correlation functions for a single Fourier mode (3.3) through discrete-time filtering (3.5) from the perfect filter (3.11) and unoptimized imperfect filter estimates (3.12)-(3.14) with long observation time  $\Delta t = 2 = \tau_{\text{corr}}$  in the stiff dynamics regime of (3.3) with  $\gamma = 0.5, \omega = 10, \sigma = 1, r^\circ = 0.25$ .

which is a particular case of the general test model (3.1) with

$$\gamma \left( \frac{\partial}{\partial x} \right) = d + \mu \frac{\partial^2}{\partial x^2}, \quad P \left( \frac{\partial}{\partial x} \right) = -c \frac{\partial}{\partial x}, \quad c, d, \mu \geq 0. \quad (3.33)$$

Following [15, 37, 60], we refer to the dynamics with  $\mu > 0$  as the selective damping case and the dynamics with  $d \neq 0, \mu = 0$  as the uniform damping case; this distinction is easily understood in the Fourier domain where the dynamics of each Fourier mode of (3.32) is given by (3.3) with

$$\lambda_k = d + \mu k^2 - ick, \quad (3.34)$$

so that for  $\mu = 0$  the damping,  $\Re[\lambda_k] = d$ , is the same for all Fourier modes. The deterministic forcing  $\mathcal{F}$  in (3.32) acts in the large-scale waveband, which in the Fourier

domain is realized by

$$\hat{\mathcal{F}}_k(t) = \begin{cases} A_k e^{\omega_0(k)t}, & \text{for } k \leq K, \\ 0, & \text{for } k > K, \end{cases} \tag{3.35}$$

for some fixed  $K \leq N$ . We write the discretized dynamics of (3.32) as

$$\hat{u}_{k,m+1} = F_k \hat{u}_{k,m} + \hat{\mathcal{F}}_{k,m+1} + \hat{\sigma}_{k,m+1}, \tag{3.36}$$

where  $F_k = \exp(-\lambda_k \Delta t)$ ,  $\langle \hat{\sigma}_{k,m+1} \otimes \hat{\sigma}_{k,m+1} \rangle = \hat{r}_k$ , and the deterministic forcing in (3.36) consistent with the continuous dynamics in (3.3) is given by

$$\hat{\mathcal{F}}_{k,m+1} = \begin{cases} \frac{A_k e^{i\omega_0(k)m\Delta t}}{i\omega_0(k) - \lambda_k} \left( e^{i\omega_0(k)\Delta t} - e^{-\lambda_k \Delta t} \right), & \text{for } k \leq K, \\ 0, & \text{for } k > K. \end{cases} \tag{3.37}$$

The dynamics of the forecast models in the imperfect filters is given by (3.12)-(3.14) with

$$\lambda_k^M = d - \mu \frac{1 - 2e^{-ikh} + e^{-2ikh}}{h^2} - c \frac{1 - e^{-ikh}}{h}, \tag{3.38}$$

and  $F_k^M, R_k^M$  specified in Section 3.2 so that all the imperfect forecast models converge to the perfect model dynamics when  $h \rightarrow 0, \Delta t \rightarrow 0$ .

The advection-diffusion dynamics (3.32) will be also used in Section 4 to analyze filter performance with sparse observations when the failure of the classical observability condition [2] for the discretized PDE in (3.32) gives rise to subtle and important effects which affect the filter performance.

**3.5.1. Numerical examples.** Here, similar to the analysis of [15, 60], we consider the problem of filtering the stochastically forced advection-diffusion dynamics in (3.32) with uniform damping  $d \neq 0, \mu = 0$  and plentiful observations  $P = 1$ . The uniform damping setup is the most challenging one. The present superensemble framework for assessing the filter skill further reinforces the previous results obtained in the path-wise framework [15, 60]. Moreover, the superensemble framework allows for assessing the statistical accuracy of the filtered signal via the relative entropy measure which adds important new insight about the imperfect filtering of the turbulent spatially extended dynamics.

The truth in the studied tests is generated by solving for  $2N + 1 = 41$  Fourier modes (3.3) in the discretized advection-diffusion dynamics of (3.32) with  $c = 1, d = 0.01$ , and  $\Delta t = 50 = \tau_{corr}/2$ , where the decorrelation time  $\tau_{corr} = 1/d = 100$  is the same at each wavenumber. In this uniformly damped setting the amplification factors in the perfect system satisfy  $|F_k| = e^{-d\Delta t} = e^{1/2} < 1$ , implying strong asymptotic stability in the truth dynamics and the perfect filter (3.11). We consider the turbulent signals generated from two distinct turbulent spectra: an equipartition spectrum  $E_k = 100$  and a “ $-5/3$ ” spectrum with  $E_k = 1000k^{-5/3}$ . The observation noise variance in the Fourier domain is  $\hat{r}^o = r^o / (2N + 1) \approx 24$  with  $r^o = 1000$ . In total we consider six different imperfect filters including three different temporal discretizations (3.12)-(3.14) and two different ways of choosing the noise variance. Similar to the one-mode study in Section 3.4.3, the unoptimized filters use the same system noise variance as in the perfect

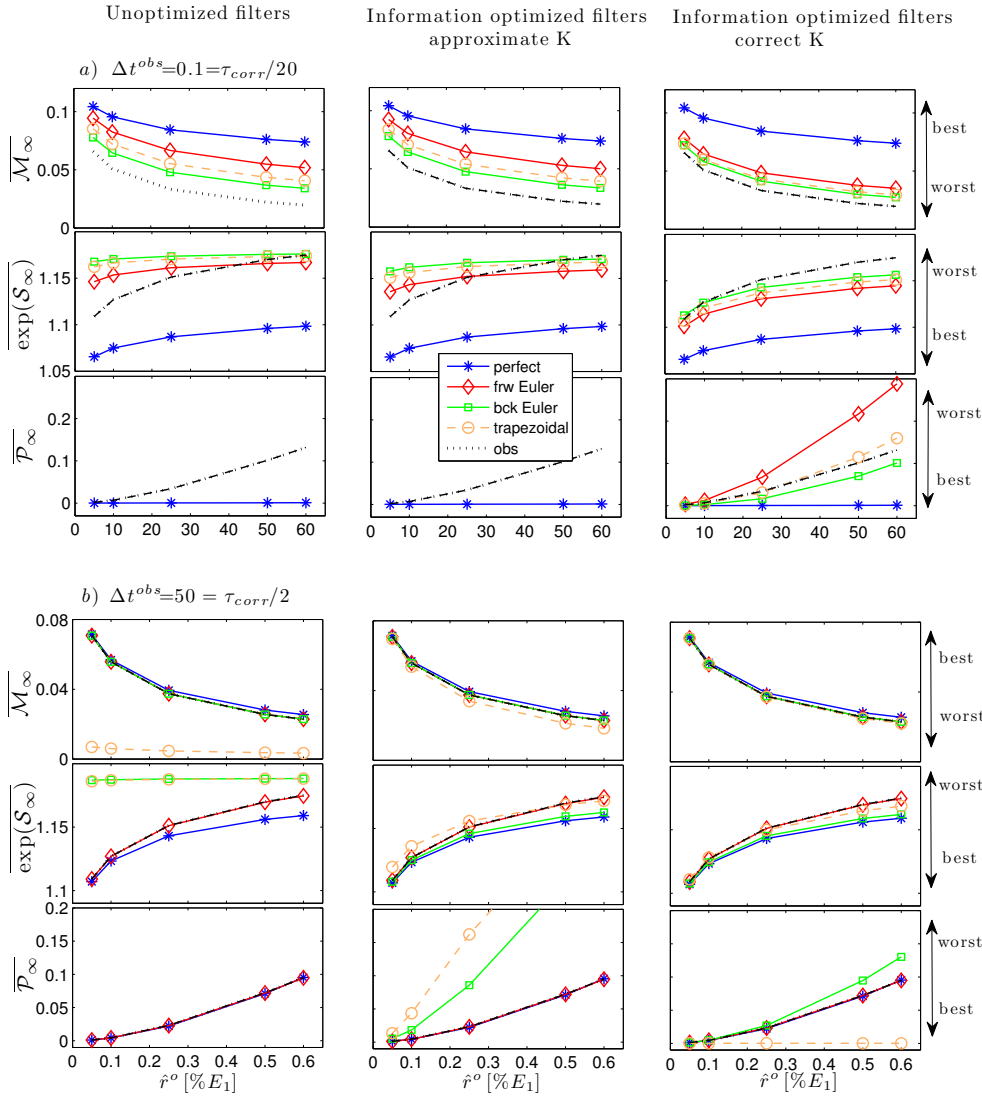


FIG. 3.6. Filtering the turbulent advection-diffusion dynamics (3.32) with **uniform damping and decaying mean through plentiful observations** ( $\mu=0, d=0.01, \mu=0, N=21, M=21, P=1$ ). The three information measures discussed in Section 2.2.1 and averaged over all the resolved modes are shown as a function of the observation noise,  $\hat{r}^{obs}$  in the difficult dynamical regime with equipartition energy spectrum  $E_k=1$  and for two different observation times. The dashed black line corresponds to the signal estimated directly from the observations. The information criterion  $\mathfrak{M}$  in (2.50) combining these measures to rank the filter performance is shown in figure 3.7.

dynamics, i.e.,  $\sigma^M = \sigma$  in (3.12)-(3.14), and the optimized filters have inflated model noise variance,  $r^M$ , according to the information criterion used previously in [60]; the optimal noise variance in these filters corresponds to the minimum lack of information in terms of relative entropy (2.18) between the statistics of the perfect filter error and the imperfect filter error in the absence of mean biases in the filter estimates. For the filters using the backward Euler (3.13) and trapezoidal (3.14) schemes, the condition



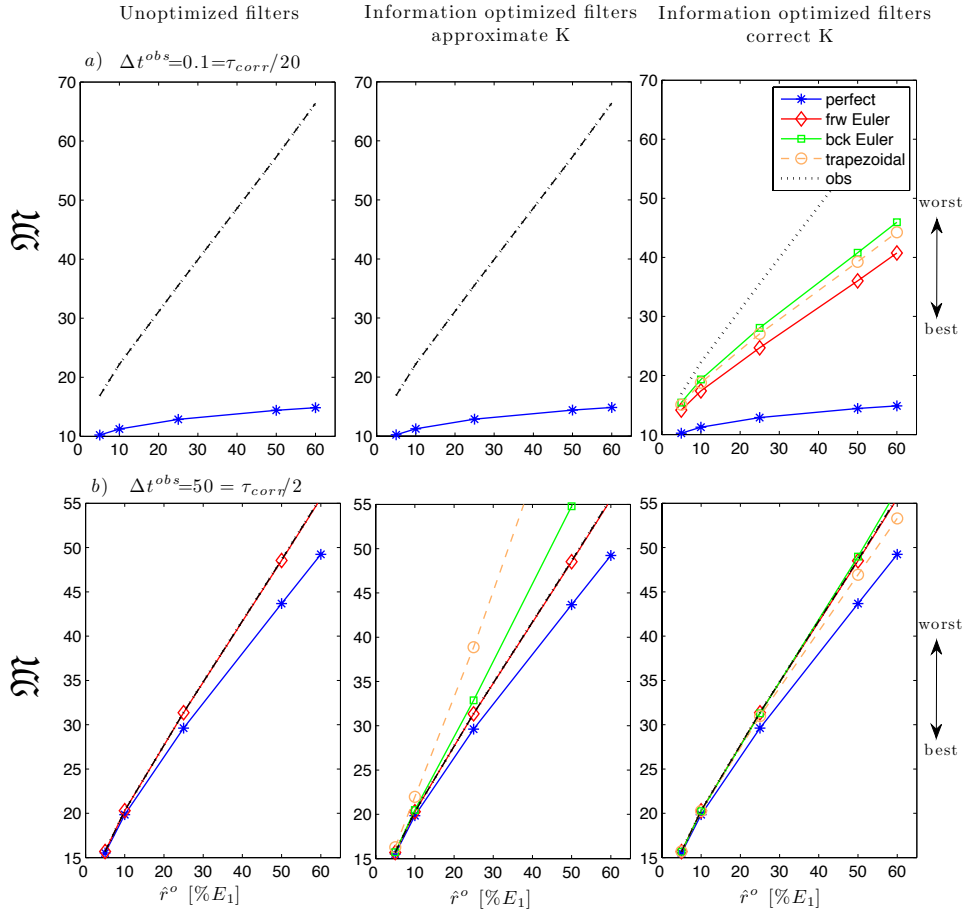


FIG. 3.7. Illustration of potential benefits of imperfect filter optimization. The information criterion  $\mathfrak{M}$  in (2.50) for filtering the turbulent advection-diffusion dynamics (3.32) with **uniform damping and decaying mean through plentiful observations** ( $\mu=0, d=0.01, \mu=0, N=21, M=21, P=1$ ); the individual information measures combining the information measures are shown in figure 3.6. The perfect filter is given by (3.11) and the imperfect filters (3.12)-(3.14) introduce model error due to time-space discretizations (see Section 3.2). The dashed black line corresponds to the signal estimated directly from the observations.

$|F_k^M| < 1$  is always satisfied for any discretization step  $\Delta t$ . In the trapezoidal scheme, however, almost every mode is marginally stable so that  $F_k^M \simeq 1$  for most modes. For the unstable forward Euler scheme in (3.12), the amplitude satisfies  $|F_k^M| > 1$  for every considered time step and it increases sharply with the resolution.

The performance of the imperfect filters (3.11)-(3.14) for recovering the turbulent advection-diffusion dynamics of (3.32) with plentiful observations is summarized in figures 3.6 and 3.7. Figure 3.6 shows the three information measures: the mode-averaged mutual information (3.20), the mode-averaged filter error entropy (3.21), and the mode-averaged relative entropy (3.22) between the truth and the filter estimate as a function of the observation noise variance. Figure 3.7 shows the information

criterion (2.50) which combines the individual information measures from figure 3.6; the three columns in figure 3.7 illustrate the filter skill for unoptimized imperfect filters (left), imperfect filters with approximate Kalman gain (2.13) after the simple noise inflation (middle), and for the imperfect filters with correct Kalman gain (2.11) (right).

In summary, we make the following points in regard to filtering the spatially extended dynamics of the test model (3.32) with plentiful observations:

- The ensemble error in the filter estimates can be computed via the ergodic averages on the attractor of the augmented system (3.17); good agreement between the ensemble averages and time-averages was verified by direct computations (see table 3.1).
- When filtering with imperfect models the ‘minimum error variance’ filter with the smallest error entropy does not necessarily provide optimal estimates in terms of mutual information/pattern correlation or relative entropy (figure 3.7).
- The imperfect filters (3.12)-(3.14) generally fail to recover the variance of the truth signal except in the practically unrealistic case when the correct Kalman gain (2.11) in the filters is implemented (see figure 3.6). For correct forcing the mean of the truth is recovered accurately (see figure 4.11 for a different case when incorrect forcing causes large biases in the mean).
- For long observation times the backward Euler and the trapezoidal filters gain significant skill after the simple system noise inflation via the information optimization of the filter error, or if the correct Kalman gain is used (figure 3.3, 3.4). Both these cases require the knowledge of the truth dynamics but these results point to the potential gains in filter improvement when appropriate approximations are implemented.
- The forward Euler filter trusts the observations and it is largely insensitive to the system noise inflation (see figures 3.6, 3.7).
- The hierarchy of unoptimized filter skill:
  - For short observation times,  $\Delta t^{obs} \ll \tau_{corr}$ , all imperfect filters have good pattern correlation/mutual information (see figure 3.6) but perform worse than the direct observations in terms of filter error entropy. All imperfect filters vastly misrepresent the variance of the truth signal in terms of the relative entropy.
  - For long observation times,  $\Delta t^{obs} \sim \tau_{corr}$ , all filters except the trapezoidal filter perform similarly and slightly better than direct observations in terms of mutual information/pattern correlation; the trapezoidal filter performs worse than the observations. In terms of error entropy and relative entropy the imperfect filters perform worse than observations.
- The hierarchy of filter skill after the simple information optimization (knowledge of truth required).
  - For short observation times  $\Delta t^{obs} \ll \tau_{corr}$  the conclusions are the same as above (see figure 3.6).
  - For long observation times,  $\Delta t^{obs} \sim \tau_{corr}$ , all filters except the trapezoidal filter perform similarly in terms of mutual information and slightly better than direct observations; the trapezoidal filters performs worse than the observations. The information optimization improves the backward Euler so that it is comparable to the perfect filter in terms of filter

error entropy. The trapezoidal and backward Euler vastly misrepresent the variance of the signal while the forward Euler, which trusts the observations, has similar relative entropy to the observations.

- The hierarchy of imperfect filter skill with correct Kalman gain (knowledge of truth required)
  - All filters beat the raw observations estimates in terms of mutual information and pattern correlation (see figure 3.6).
  - For short observation times,  $\Delta t^{obs} \ll \tau_{corr}$ , all imperfect filters beat the raw observations estimates in terms of mutual information and filter error entropy with the hierarchy

perfect filter (3.11) > forward Euler (3.12) > trapezoidal (3.14) >  
> backward Euler (3.13) > raw obs.,

The backward Euler also has good relative entropy.

- For long observation times,  $\Delta t^{obs} \sim \tau_{corr}$ , all filters perform similarly to the perfect filter and raw observations in terms of mutual information.
- The backward Euler performs well and similar to the perfect filter in terms of the filter error entropy. The hierarchy of filter skill is the same as for the short observation times.
- The trapezoidal filter has a high skill for estimating the variance of the truth in terms of the relative entropy (see figures 3.6, 3.7).

#### 4. Ensemble filter error for Gaussian spatially extended systems: sparse regular observations

Robust and accurate filtering of turbulent systems with rough energy spectra near the mesh scale of the discretized models and intermittent energy transfers from the unresolved scales is both challenging and important in practical applications. In low-dimensional filtering problems the standard requirements on the filtering algorithm involve bounds on the error mean and variance. However, the development of appropriate filtering techniques in this high-dimensional setting with sparse observations requires a framework for estimating the filter error and correlations across multiple scales which might be difficult to capture in the path-wise framework. Here, we consider the superensemble error in imperfect, reduced-order filtering techniques proposed earlier in [37, 60] for sparsely observed spatially extended turbulent systems with sufficiently steep energy spectra,  $E_k \propto k^{-\beta}$ ,  $\beta > 1/2$ . In such dynamical regimes the path-wise tests carried out in [37] indicated that these cheap reduced algorithms had high filtering skill which was superior to the full filter skill in the presence of non-trivial mean state and resonant forcing, especially when the classical observability conditions for the filtered dynamics were violated (e.g., [2, 60]). The present superensemble framework for assessing the filter error provides further unambiguous evidence supporting those earlier results. Below, we first recapitulate the most important consequences of sparse observations in physical space on the filtering problem in the Fourier domain. Sparse observations are commonplace in most geophysical applications and they induce correlations between Fourier modes, as described in [59, 37, 60] and outlined below. We then outline the most important features of the reduced filters and study their performance in the superensemble framework using information-theoretic skill measures developed in Section 2. In Section 4.3.1 the filter performance is studied for sparsely observed stochastically forced advection-diffusion

dynamics (3.32) with decaying mean dynamics; the ultimate tests of reduced filter performance are studied in Section 4.3.2 based for the sparsely observed turbulent dynamics of (3.32) with non-trivial mean dynamics due to resonant forcing.

As shown earlier in [59, 37, 60], for sparse observations (3.15) on a grid with  $2M+1$  points ( $M < N$ ) and uncorrelated initial conditions the problem of filtering linear Gaussian dynamics of (3.1) in the Fourier domain splits into independent filtering problems for the  $P$ -dimensional state vectors ( $P = (2N+1)/(2M+1)$ )

$$\mathbf{u}_{\{\ell\}} \equiv \begin{pmatrix} \hat{u}_\ell \\ \hat{u}_{\ell+(2M+1)q_2} \\ \vdots \\ \hat{u}_{\ell+(2M+1)q_P} \end{pmatrix} \equiv \begin{pmatrix} \hat{u}_{k_1} \\ \hat{u}_{k_2} \\ \vdots \\ \hat{u}_{k_P} \end{pmatrix}, \quad (4.1)$$

which contain all the Fourier modes from the aliasing set of wavenumber  $|\ell| \leq M$  given by

$$\mathcal{A}(\ell) = \left\{ k: |k| \leq N, \quad k = \ell + (2M+1)q, \quad k, q \in \mathbb{Z} \right\}. \quad (4.2)$$

We refer to the  $2M+1$  wavenumbers  $\ell$  resolved by the observation grid as the *primary* wavenumbers and note that these primary wavenumbers are not necessarily the most energetic modes in their aliasing sets (see e.g., [37]). Transformation of the sparse observations (3.15) into the Fourier domain produces  $2M+1$  independent sets,  $v_{\{\ell\}}$ , which scramble together the wavenumbers resolved by the discretized model and contained in the aliasing sets  $\mathcal{A}(\ell)$  (see, e.g., [59, 37, 60]); specifically, the aliased observations in the Fourier domain can be written as

$$v_{\{\ell\}, m+1} = \sum_{k \in \mathcal{A}(\ell)} \hat{u}_k(t_{m+1}) + \sigma^o \equiv G_P \mathbf{u}_{\{\ell\}}(t_{m+1}) + \sigma^o, \quad \langle \hat{\sigma}^o \otimes (\hat{\sigma}^o)^* \rangle = r_{\{\ell\}}^o, \quad (4.3)$$

where  $G_P: \mathbb{C}^N \rightarrow \mathbb{C}$  is a linear, rank-deficient observation operator and  $r_{\{\ell\}}^o \equiv \delta_{\ell-\ell'} r^o / (2M+1)$ . Similar to the simple case discussed in Section 2.1, the action of the discrete-time Kalman filter within each aliasing set  $\mathcal{A}(\ell)$  is described by the augmented system

$$\begin{bmatrix} \mathbf{u}_{\{\ell\}, m+1} \\ \bar{\mathbf{u}}_{\{\ell\}, m+1|m+1} \end{bmatrix} = \begin{bmatrix} F_{\{\ell\}} & 0 \\ K_{\{\ell\}, m+1}^M G_P F_{\{\ell\}} & (\mathcal{I} - K_{\{\ell\}, m+1}^M G_P) F_{\{\ell\}}^M \end{bmatrix} \begin{bmatrix} \mathbf{u}_{\{\ell\}, m} \\ \bar{\mathbf{u}}_{\{\ell\}, m|m} \end{bmatrix} + \mathfrak{F} + \begin{bmatrix} \boldsymbol{\sigma}_{\{\ell\}, m+1} \\ K_{\{\ell\}, m+1}^M (G_P \boldsymbol{\sigma}_{\{\ell\}, m+1} + \boldsymbol{\sigma}_{\{\ell\}, m+1}^o) \end{bmatrix}, \quad (4.4)$$

where  $\mathbf{u}_{\{\ell\}}, \bar{\mathbf{u}}_{\{\ell\}, m|m} \in \mathbb{C}^P$  are defined by (4.1) and contain the aliasing Fourier modes in the set  $\mathcal{A}(\ell)$  and

$$\mathfrak{F} := \begin{bmatrix} \mathcal{F}_{\{\ell\}, m+1} \\ (\mathcal{I} - K_{\{\ell\}, m+1}^M G_P) \mathcal{F}_{\{\ell\}, m+1}^M + K_{\{\ell\}, m+1}^M G_P \mathcal{F}_{\{\ell\}, m+1} \end{bmatrix}. \quad (4.5)$$

The diagonal operators  $F_{\{\ell\}} = \text{diag}[\hat{F}_{k_1}, \dots, \hat{F}_{k_P}]$ ,  $F_{\{\ell\}}^M = \text{diag}[\hat{F}_{k_1}^M, \dots, \hat{F}_{k_P}^M]$ ,  $k_i \in \mathcal{A}(\ell)$ , propagate the dynamics of the Fourier modes within each aliasing set. The forward

dynamics of the truth is driven by both the deterministic forcing  $\mathcal{F}_{\{\ell\}}$  and white noise  $\sigma_{\{\ell\}}$  denoted by

$$\mathcal{F}_{\{\ell\}} := [\hat{\mathcal{F}}_{k_1}, \dots, \hat{\mathcal{F}}_{k_P}]^T, \quad \sigma_{\{\ell\}} := [\hat{\sigma}_{k_1}, \dots, \hat{\sigma}_{k_P}]^T, \quad k_i \in \mathcal{A}(\ell).$$

Analogous notation is adopted for the remaining vectors in the imperfect dynamics of the forecast model involving  $\mathcal{F}_{\{\ell\}}^M$  and  $\sigma_{\{\ell\}}^M$ . Here and below, we assume for simplicity that the Fourier modes in the initial data from disjoint aliasing sets are uncorrelated [59, 60].

Note that the filter estimates for the modes in each aliasing set  $\mathcal{A}(\ell)$  are correlated due to the spatially sparse observations; this is true even in the simple canonical linear test problem (3.1) where the true Fourier modes of the augmented system (4.4) evolve independently between the observations. Thus, for sparse observations with  $P > 1$ , the covariance of the augmented state  $\mathbf{y}_m \equiv (\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})^T$  has a block-diagonal structure with the individual blocks containing the covariances of the aliasing modes which evolve according to

$$\mathcal{C}_{\{\ell\}m+1} = \mathcal{F}_{\{\ell\}m} \mathcal{C}_{\{\ell\}m} \mathcal{F}_{\{\ell\}m}^* + \mathcal{R}_{\{\ell\}m}, \tag{4.6}$$

where the operators in (4.6) are given by

$$\mathcal{F}_{\{\ell\}m} = \begin{bmatrix} F_{\{\ell\}m} & 0 \\ K_{\{\ell\}m+1}^M G_P F_{\{\ell\}m} & (\mathcal{I} - K_{\{\ell\}m+1}^M G_P) F_{\{\ell\}m}^M \end{bmatrix}, \tag{4.7}$$

$$\mathcal{R}_{\{\ell\}m} = \begin{bmatrix} r_{\{\ell\}} & r_{\{\ell\}} G_P^* K_{\{\ell\}m+1}^{M*} \\ K_{\{\ell\}m+1}^M G_P r_{\{\ell\}} & K_{\{\ell\}m+1}^M (r^o + G_P r_{\{\ell\}} G_P^*) K_{\{\ell\}m+1}^{M*} \end{bmatrix}, \tag{4.8}$$

and the covariance of the aliasing Fourier modes has a natural block decomposition

$$\mathcal{C}_{\{\ell\}m} = \begin{bmatrix} Cov(\mathbf{u}_{\{\ell\}m}, \mathbf{u}_{\{\ell\}m}) & Cov(\mathbf{u}_{\{\ell\}m}, \bar{\mathbf{u}}_{\{\ell\}m|m}) \\ Cov(\bar{\mathbf{u}}_{\{\ell\}m|m}, \mathbf{u}_{\{\ell\}m}) & Cov(\bar{\mathbf{u}}_{\{\ell\}m|m}, \bar{\mathbf{u}}_{\{\ell\}m|m}) \end{bmatrix} \equiv \begin{bmatrix} \mathcal{C}_{\{\ell\}(11)m} & \mathcal{C}_{\{\ell\}(12)m} \\ \mathcal{C}_{\{\ell\}(21)m}^* & \mathcal{C}_{\{\ell\}(22)m} \end{bmatrix}. \tag{4.9}$$

The block  $\mathcal{C}_{\{\ell\}(11)}$  of (4.9) denotes the covariance of the truth modes  $k \in \mathcal{A}(\ell)$  in the aliasing set (4.2) of the primary mode  $|\ell| \leq M$ ,  $\mathcal{C}_{\{\ell\}(22)}$  is the covariance of the filter estimate, and  $\mathcal{C}_{\{\ell\}(12)}$  quantifies the correlations between the truth  $\mathbf{u}_{\{\ell\}m}$  and the filter estimate  $\bar{\mathbf{u}}_{\{\ell\}m|m}$  within each disjoint aliasing set  $\mathcal{A}(\ell)$ . It will prove crucial in the following analysis to account for both the covariance of the filter estimates  $\mathcal{C}_{(22)}$  and the cross-correlations between the truth and the filter estimates  $\mathcal{C}_{(12)}$ .

The block-diagonal structure of the full covariance with the blocks indexed by the primary modes  $|\ell| \leq M$  allows for simplifications in the general expressions for the mutual information  $\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m})$  in (2.26), the entropy of the filter error  $\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m})$  in (2.22), and the relative entropy projected onto the individual modes,  $\sum_k \mathcal{P}(\pi(u_{k,m}), \pi^\dagger(\bar{u}_{k,m|m}))$  in (2.23); the specific formulas are as follows:

- The mutual information (2.19) for the full sparsely observed system can be expressed through the sum of the covariances within the disjoint aliasing sets  $\mathcal{A}(\ell)$  as

$$\mathcal{M}(\mathbf{u}_m, \bar{\mathbf{u}}_{m|m}) = -\frac{1}{2} \sum_{|\ell| \leq M} \ln \det [\mathcal{I} - \mathcal{C}_{\{\ell\}(22)}^{-1} \mathcal{C}_{\{\ell\}(12)}^* \mathcal{C}_{\{\ell\}(11)}^{-1} \mathcal{C}_{\{\ell\}(12)}], \tag{4.10}$$

where  $|\ell| \leq M$  indexes the aliasing sets  $\mathcal{A}(\ell)$  in (4.2) and the covariance blocks for each aliasing set are defined in (4.9); for clarity we skipped the explicit time dependence on the right hand side of (4.10).

- The entropy of the filter error (2.16) for the full sparsely observed system decomposes as

$$\mathcal{S}(\mathbf{u}_m - \bar{\mathbf{u}}_{m|m}) = \mathfrak{S} + \frac{1}{2} \sum_{|\ell| \leq M} \ln \det \left[ \mathcal{C}_{\{\ell\}(11)} + \mathcal{C}_{\{\ell\}(22)} - 2\Re e[\mathcal{C}_{\{\ell\}(12)}] \right], \quad (4.11)$$

where  $\mathfrak{S} := \frac{1}{2}(2M+1)(1 + \ln 2\pi)$  and the covariance blocks are defined in (4.9).

- The relative entropy (2.23) for the full sparsely observed system projected onto the individual Fourier modes decomposes as

$$\begin{aligned} \mathcal{P}(\pi(\mathbf{u}_m), \pi^\dagger(\bar{\mathbf{u}}_{m|m})) &= \frac{1}{2} \sum_{|\ell| \leq M} [\bar{\mathbf{u}}_{\{\ell\}m}^* - \bar{\mathbf{u}}_{\{\ell\}m|m}^*] (\mathcal{C}_{\{\ell\}(22)m})^{-1} [\mathbf{u}_{\{\ell\}m} - \bar{\mathbf{u}}_{\{\ell\}m|m}] \\ &\quad + \frac{1}{2} \sum_{|\ell| \leq M} \left[ \text{tr} \Xi_{\{\ell\}m} - \ln \det \Xi_{\{\ell\}m} \right] - (2N+1), \end{aligned} \quad (4.12)$$

where  $|\ell| \leq M$  indexes the aliasing sets  $\mathcal{A}(\ell)$  of Fourier modes defined in (4.2),  $\Xi_{\{\ell\}m} := \mathcal{C}_{\{\ell\}(11)m} (\mathcal{C}_{\{\ell\}(22)m})^{-1}$ , and we again skipped the explicit time dependence in the vectors of the aliased Fourier modes and their covariances on the right hand side of (4.12) for clarity. Due to the possibility of correlations in the filter estimates of the Fourier modes it will prove useful to consider the projection of the relative entropy onto the individual Fourier modes which is given by (3.22) considered earlier for plentiful observations.

We use these three information measures to quantify the filter performance through the information criterion in (2.50).

**4.1. Reduced filters for sparsely observed turbulent systems.** The reduced-order techniques for real time filtering of sparsely observed turbulent systems, which are briefly outlined in this section, were originally introduced in [37, 60] as cheap and robust filters for dynamical regimes with sufficiently steep energy spectra. The advantages of these reduced algorithms stem from a vastly improved computational efficiency, and from improved stability, observability, and accuracy when estimating the resolved modes  $\hat{u}_\ell$  in (4.1) with  $|\ell| \leq M \leq N$  (see [37, 60] for details). These reduced algorithms judiciously neglect covariances between the Fourier modes in different aliasing sets and retain the covariances only in a small subset of the most energetic aliasing modes; such approximations guarantee observability on the reduced subspace of the most energetic Fourier modes while providing reasonably consistent estimates for the energy in the remaining modes in the truth signal. Further confirmation of the earlier findings of [37, 60] based on the path-wise analysis is presented in Section 4.2.1, and Section 4.3 based on the information-theoretic superensemble framework introduced in Section 2.

The filters considered here are (see [37, 60] for details):

**FDKF:** The Fourier Domain Kalman Filter (FDKF) is the standard Kalman filter applied to the disjoint sets of aliasing Fourier modes  $\mathcal{A}(\ell)$ ,  $|\ell| \leq M \leq N$ . FDKF represents the perfect filter when filtering the linear Gaussian dynamics of (3.1) with independent initial conditions.

**SDAF** and **VSDAF**: The Strongly Damped Approximate Filter in the Fourier domain (SDAF) is motivated by the following situation which often occurs in practice in dynamics with dominant selective damping and steep energy spectra. These filters introduce a judicious model error by assuming that there are two disjoint sets of Fourier modes in the filtered signal with

- Moderate damping:  $|F_{k_i}^M| \sim \mathcal{O}(1)$ ,  $1 \leq i \leq P_0$ ,  $k_i \in \mathcal{A}(\ell)$ ,
  - Strong damping:  $|F_{k_i}^M| \ll 1$ ,  $P_0 + 1 \leq i \leq P$ ,  $k_i \in \mathcal{A}(\ell)$ .
- (4.13)

The above configuration implies strong damping in a subset of the aliased modes which always contains the primary modes  $\hat{u}_\ell$ ,  $|\ell| \leq M \leq N$ . In this situation the dynamic covariance update in FDKF can be inaccurate because in each aliasing set the cross-covariances between the first  $P_0$  modes and the strongly damped  $P - P_0$  modes involve multiplications by large and small numbers to get order one quantities. Thus, in order to mitigate the potential numerical artifacts, SDAF assumes memoryless dynamics in the filter forecast model for modes  $\hat{u}_{k_i}$  with  $i > P_0$ ,

$$F_{k_i}^M = 0 \quad \text{for } P_0 + 1 \leq i \leq P, P_0 \geq 0, \quad k_i \in \mathcal{A}(\ell), \tag{4.14}$$

so that

$$\hat{u}_{k_i, m+1|m}^M = \mathcal{F}_{k_i, m+1}^M + \sigma_{k_i}^M, \quad i > P_0, \tag{4.15}$$

and the model covariance for the aliasing modes,  $k_i \in \mathcal{A}(\ell)$ , within each aliasing set  $\mathcal{A}(\ell)$ ,  $|\ell| \leq M \leq N$ , takes the form

$$R_{\{\ell\}m+1|m}^M = \left( \begin{array}{c|c} [R_{\{\ell\}, m+1|m}^M]_{(1:P_0, 1:P_0)} & 0 \\ \hline 0 & \text{diag}[r_{k_{P_0+1}}^M, \dots, r_{k_P}^M] \end{array} \right), \tag{4.16}$$

where the model noise,  $r_{k_i}^M, i > P_0$ , for the memoryless modes does not need to be updated dynamically [37, 60]. In contrast to SDAF, the Variance Strong Damping Approximate Filter (**VSDAF**) utilizes the memoryless approximation only in the computations of the variance (4.16) while the mean is updated using the full dynamics rather than (4.15) (see [37, 60] for details).

**RFDKF**: The approximation implemented in the Reduced Fourier Domain Kalman Filter (RFDKF) is based on the idea that for sufficiently rapid decay in the turbulent spectrum of the true signal the primary mode in each aliasing set  $\mathcal{A}(\ell)$  contains the most energy, so only this mode should be actively filtered. Thus, RFDKF always trusts the dynamics for all the aliased modes  $k_i \in \mathcal{A}(\ell)$ ,  $2 \leq i \leq P$ , so that the Kalman gain in (2.6) is approximated as

$$\tilde{K}_{\{\ell\}m+1}^M = \left( K_{k_1, m+1}^M, 0, \dots, 0 \right)^T, \tag{4.17}$$

while the primary mode  $\hat{u}_{k_1}^M$  in each aliasing set is filtered using

$$\hat{u}_{k_1, m+1|m}^M = F_{k_1}^M \hat{u}_{k_1, m|m}^M + \mathcal{F}_{k_1, m+1}^M + \hat{\sigma}_{k_1, m+1}^M, \tag{4.18}$$

$$\hat{v}_{k_1, m+1} = G_P \mathbf{u}_{\{\ell\}m+1} - \tilde{G}_P \mathbf{u}_{\{\ell\}m+1}^M + \sigma_{\{\ell\}m+1}^o, \tag{4.19}$$



where  $G_P = (G \ G \dots G)$ ,  $\tilde{G}_P = (0 \ G \dots G)$ , and the observations of the primary mode  $\hat{u}_{k_1}$  are estimated from the sparse observations  $\hat{v}_{\{\ell\}}$  of the aliased modes  $\mathbf{u}_{\{\ell\}}$  in (4.1) based on the model dynamics for the remaining aliasing modes

$$\hat{u}_{k_i, m+1|m}^M = F_{k_i}^M \hat{u}_{k_i, m|m-1}^M + \mathcal{F}_{k_i, m+1}^M + \hat{\sigma}_{k_i, m+1}^M, \quad k_i \in \mathcal{A}(\ell), \quad 2 \leq i \leq P. \quad (4.20)$$

The augmented system for RFDKF is derived in Appendix B. Compared with SDAF, which uses the memoryless dynamics on the unresolved modes but provides nontrivial estimates for their mean, the RFDKF fully trusts the dynamics on the unresolved modes. Both VSDAF with  $P_0=1$  and RFDKF require detailed Kalman filtering on the primary mode alone in each aliasing set and they both satisfy observability [37, 60]. Furthermore, VSDAF provides a natural blending between the SDAF and RFDKF filters by using the explicit Kalman gain vector to weight the unresolved dynamics and the effect of sparse observations in the state estimates.

**4.2. Characteristics of filter error in the reduced filtering algorithms.**

Here, we discuss various aspects of filter error in the reduced algorithms presented in the previous section. Two important aspects are (i) the effects of the mean dynamics on the filter error in the presence of resonant/non-resonant forcing and violation of the observability condition [2, 60] and (ii) the filter-induced correlations in the estimates for the Fourier modes within the same aliasing set. The framework developed in Section 2.1 allows for a semi-analytical treatment of the above issues which are discussed below.

**4.2.1. Filter bias due to incorrect mean dynamics and lack of observability for sparsely observed turbulent systems.** Here, we focus on the effect of errors in filtering which arise due to errors in the estimates of the mean dynamics, i.e., we consider situations when

$$\lim_{m \rightarrow \infty} \mathbb{E}^P[\mathbf{U}_{\{\ell\}m}] \equiv \mathbb{E}^\pi[\mathbf{u}_{\{\ell\}m}] - \mathbb{E}^{\pi^\dagger}[\bar{\mathbf{u}}_{\{\ell\}m|m}] \neq 0, \quad (4.21)$$

in the filtering problem (4.4) for some aliasing sets  $\mathcal{A}(\ell)$  where  $p(\mathbf{u}_{\{\ell\}m}, \bar{\mathbf{u}}_{\{\ell\}m|m})$  is the joint density of the truth and filter estimate conditioned on the observations, and  $\pi(\mathbf{u}_{\{\ell\}m})$ ,  $\pi^\dagger(\bar{\mathbf{u}}_{\{\ell\}m|m})$  are the respective marginal densities. The importance of the detrimental effects of such biases on the filter performance was recognized and discussed in [37, 60] where it was shown that the skill of FDKF (see Section 4.1) in terms of the asymptotic RMS error (2.36) was significantly degraded in the presence of non-trivial mean dynamics and resonant forcing. In particular, violation of the classical observability condition when filtering spatially extended systems can grossly amplify the filter error, as noted earlier in [18] and discussed in [37, 60]. In this section, we use the statistical superensemble framework in order to further elucidate these issues. In particular, we show that time-dependent errors in the forcing of the filter forecast model induce significant errors in the mean filter estimates,  $\mathbb{E}[\mathbf{U}_{\{\ell\}}]$ , which degrades the filter skill in terms of the relative entropy (2.18) as well as the RMS error (2.36); on the other hand, the remaining information measures of filter error, the entropy in (2.22) and the mutual information in (2.26), are insensitive to the biases on the mean filter estimates. These issues are investigated further for the reduced filters in Section 4.3 and Section 4.3.2.

Recall that, according to the classical control theory literature (e.g., [2, 17]), the filtering problem

$$a) \frac{d\mathbf{u}}{dt} = F\mathbf{u}, \quad b) \mathbf{v} = G\mathbf{u}, \quad \mathbf{u} \in \mathbb{C}^N, \quad F \in \mathbb{C}^{N \times N}, \quad G \in \mathbb{C}^{M \times N}, \quad M \leq N, \quad (4.22)$$

is observable if and only if the initial conditions  $\mathbf{u}_0$  in (4.22a) can be reconstructed from the observations  $\mathbf{v}(t)$ ,  $t \in [0, t_*]$ ,  $t_* < \infty$  in (4.22b). It turns out (e.g., [2, 17]) that the system (4.22) is observable if the following algebraic condition holds:

$$\det [G^T, F^T G^T, (F^T)^2 G^T, \dots, (F^T)^{N-1} G^T] \neq 0. \tag{4.23}$$

The observability condition (4.23) often fails for wave-like systems with uniform damping and, as shown in [37, 60] for the discretization (2.1) of the canonical system (3.1), the above condition reduces to

$$\prod_{i < j} (F_{k_i} - F_{k_j}) \neq 0, \tag{4.24}$$

so that the observability is violated for the discretized dynamics of (2.1) when

$$\hat{P}_{k_i} - \hat{P}_{k_j} = 2\pi i Q / \Delta t^{obs}, \quad Q \in \mathbb{Z}^+, \tag{4.25}$$

where  $\Delta t^{obs}$  denotes the time separation between the data assimilation cycles. Note that for the uniformly damped advection-diffusion dynamics in (3.32) the observability fails simultaneously for all wave numbers which poses a severe computational test problem [60].

In order to study the effects of deterministic forcing on the performance of the reduced filters discussed in Section 4.1, consider the dynamics of the augmented system (2.6) with the error in the forcing of the filter forecast model, i.e.,

$$\mathcal{F}_{\{\ell\}m}^M = \mathcal{F}_{\{\ell\}m} + \Delta \mathcal{F}_{\{\ell\}m}, \tag{4.26}$$

for each aliasing set  $\mathcal{A}(\ell)$ ,  $|\ell| \leq M \leq N$ . We assume that the deterministic forcing acts only on the primary modes associated with the sparse observation grid in each aliasing set so that, given  $\mathcal{A}(\ell) = \{k_1 = \ell, k_2, \dots, k_P\}$ , we have

$$\mathcal{F}_{\{\ell\}} = (\mathcal{F}_{k_1}, 0, \dots)^T, \quad \mathcal{F}_{\{\ell\}}^M = (\mathcal{F}_{k_1}^M, 0, \dots)^T, \quad \Delta \mathcal{F}_{\{\ell\}m} = (\Delta \mathcal{F}_{k_1}, 0, \dots)^T.$$

This formulation combined with the dynamics of the filtering problem in (4.4) leads to the following straightforward results:

**FACT 1.** *The ensemble-averaged performance for the Kalman filtering of linear Gaussian dynamics with model error described by (2.6) does not depend on the details of the deterministic forcing and it is only sensitive to the forcing error  $\Delta \mathcal{F}_{\{\ell\}}$ .*

The mutual information (2.19) and filter error entropy (2.16) for Kalman filtering with linear Gaussian dynamics do not depend on the mean dynamics by construction. The relative entropy in (2.23) and the ensemble RMS error in (2.36) depend on the mean dynamics through the residual,  $\mathbf{U}_{\{\ell\}m+1} = \mathbf{u}_{\{\ell\}m+1} - \bar{\mathbf{u}}_{\{\ell\}m+1|m+1}$ , in (4.27) which can be written explicitly in each disjoint aliasing set  $\mathcal{A}(\ell)$  as

$$\begin{aligned} \mathbf{U}_{\{\ell\}m+1} = & (1 - K_{\{\ell\}m+1}^M \mathbf{G}_P) \left[ F_{\{\ell\}} \mathbf{u}_{\{\ell\}m} - F_{\{\ell\}}^M \bar{\mathbf{u}}_{\{\ell\}m|m} \right] - K_{\{\ell\}m+1}^M \boldsymbol{\sigma}_{\{\ell\}m+1}^o \tag{4.27} \\ & + (1 - K_{\{\ell\}m+1}^M \mathbf{G}_P) \boldsymbol{\sigma}_{\{\ell\}m+1} + (1 - K_{\{\ell\}m+1}^M \mathbf{G}_P) \Delta \mathcal{F}_{\{\ell\}m+1}, \end{aligned}$$

using the augmented system dynamics (2.6); clearly, the residual in (4.27) involves only the forcing error  $\Delta \mathcal{F}_{\{\ell\}m+1}$ .

**FACT 2.** *Given the evolution of the Kalman filter in (2.6) with the Gaussian joint probability density  $p(\mathbf{u}_{\{\ell\}m}, \bar{\mathbf{u}}_{\{\ell\}m|m})$  for the truth and the filter estimate in each disjoint aliasing set  $\mathcal{A}(\ell)$ , the mean filter error  $\mathbb{E}^P[\mathbf{U}_{\{\ell\}m}]$  depends on the error  $\Delta \mathcal{F}_{\{\ell\}}$  in the deterministic forcing but not on the forcing itself.*

This statement is obvious based on (4.27); the mean filter bias is given by

$$\begin{aligned} \mathbb{E}^P [\mathbf{u}_{\{\ell\}m+1} - \bar{\mathbf{u}}_{\{\ell\}m+1|m+1}] &= (1 - K_{\{\ell\}m+1}^M \mathbf{G}_P) \left[ F_{\{\ell\}} \mathbb{E}^\pi [\mathbf{u}_{\{\ell\}m}] - F_{\{\ell\}}^M \mathbb{E}^{\pi^f} [\bar{\mathbf{u}}_{\{\ell\}m|m}] \right] \\ &\quad + (1 - K_{\{\ell\}m+1}^M \mathbf{G}_P) \Delta \mathcal{F}_{\{\ell\}m+1}, \end{aligned} \tag{4.28}$$

where the perfect and imperfect forward dynamics operators and the Kalman gain are deterministic, which is independent of the forcings  $\mathcal{F}_{\{\ell\}m}^M$  and  $\mathcal{F}_{\{\ell\}m}$ .

**FACT 3.** *Finite-dimensional Kalman filtering in (2.6) with the reduced filters in Section 4.1 and correct forcing is stable in the mean provided that the perfect filter is stable in the mean.*

This sufficient condition for the stability of the imperfect filters in Section 4.1 can be derived as follows. Consider first the evolution of the mean filter error in (4.28) with incorrect forcing (4.26) which can be rewritten in terms of the initial values as

$$\begin{aligned} \mathbb{E}^P [\mathbf{u}_{\{\ell\}m}] &= A_m \left( F_{\{\ell\}}^m \mathbb{E}^\pi [\mathbf{u}_{\{\ell\}0}] - (F_{\{\ell\}}^M)^m \mathbb{E}^{\pi^f} [\bar{\mathbf{u}}_{\{\ell\}0|0}] \right) \\ &\quad + \sum_{k=1}^m \left( \prod_{i=1}^{m+1-k} (1 - K_{\{\ell\}m+1-i}^M \mathbf{G}_P) F_{\{\ell\}} \right) F_{\{\ell\}}^{-1} \Delta \mathcal{F}_{\{\ell\}m}^M. \end{aligned} \tag{4.29}$$

where  $A_m = \prod_{i=1}^m (1 - K_{\{\ell\}i}^M \mathbf{G}_P)$ . For correct forcing  $\Delta \mathcal{F}_{\{\ell\}m}^M = 0$  and the last term in (4.29) drops out. Fact 3 follows from the simple estimate of the first term in (4.29)

$$\begin{aligned} \left\| A_m \left( F_{\{\ell\}}^m \mathbb{E}^\pi [\mathbf{u}_{\{\ell\}0}] - (F_{\{\ell\}}^M)^m \mathbb{E}^{\pi^f} [\bar{\mathbf{u}}_{\{\ell\}0|0}] \right) \right\|_\infty &\leq \left\| A_m F_{\{\ell\}}^m \right\|_\infty \left| \mathbb{E}^{\pi^f} [\bar{\mathbf{u}}_{\{\ell\}0|0}] \right|_\infty \\ &\quad + \left\| A_m F_{\{\ell\}}^m \right\|_\infty \left| \mathbb{E}^\pi [\mathbf{u}_{\{\ell\}0}] - \mathbb{E}^{\pi^f} [\bar{\mathbf{u}}_{\{\ell\}0|0}] \right|_\infty, \end{aligned} \tag{4.30}$$

where we used the fact that  $\|F_{\{\ell\}}^M\|_\infty \leq \|F_{\{\ell\}}\|_\infty$  for the reduced filters in Section 4.1. A sufficient condition for the stability of the reduced filters with correct forcing, without imposing restrictions on  $F_{\{\ell\}}$  and  $F_{\{\ell\}}^M$ , requires that  $\|(1 - K_{\{\ell\}\infty}^M \mathbf{G}_P) F_{\{\ell\}}\| < 1$  and  $\|(1 - K_{\{\ell\}\infty}^M \mathbf{G}_P)(F_{\{\ell\}} - F_{\{\ell\}}^M)\| < 1$ ; this condition can be weakened in particular cases of the reduced filters discussed in Section 4.1, but we avoid the unnecessary technicalities here.

**FACT 4.** *If the Kalman filter in (2.6) with correct forcing is stable, its asymptotic skill in terms of the information measures of Section 2.2.1 does not depend on the mean dynamics.*

The mutual information (2.19), entropy (2.16), and the pattern correlation (2.32) for Kalman filtering with linear Gaussian dynamics do not depend on the mean dynamics by construction. The asymptotic RMS error in (2.30) is also independent of the mean dynamics, since for stable Kalman filter in (2.6) (see Fact 3) we have

$$\mathbb{E}^P [\mathbf{u}_{\{\ell\}\infty}] = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m \left( \prod_{i=1}^{m+1-k} (1 - K_{\{\ell\}m+1-i}^M \mathbf{G}_P) F_{\{\ell\}} \right) F_{\{\ell\}}^{-1} \Delta \mathcal{F}_{\{\ell\}m}^M. \tag{4.31}$$

It is clear by inspecting (4.31) that for correct forcing,  $\Delta \mathcal{F}_{\{\ell\}}^M = 0$ , the asymptotic mean filter error is zero so that the asymptotic relative entropy in (2.23) and the

$$A^M = A, \quad \omega_0(k) = \omega_k$$

$$\Delta t = 0.1532, \quad \omega_0(k) = \omega_k$$

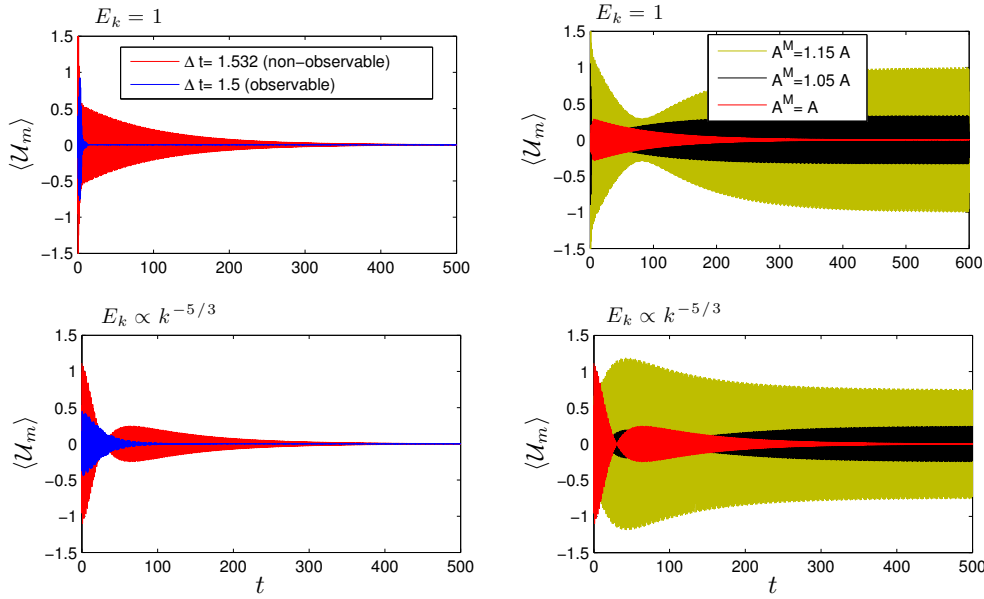


FIG. 4.1. Evolution of the mean dynamics bias  $\langle \mathcal{U}_{\{1\}m} \rangle = \langle \hat{\mathbf{u}}_{\{1\}m} - \bar{\mathbf{u}}_{\{1\}m|m} \rangle$  between the true aliasing Fourier modes and their FDKF estimates for the aliasing set  $\mathcal{A}(1)$ ; the initial conditions are not the same in the different insets. (Left) The evolution of the mean bias for correct resonant forcing  $\mathcal{F}$  in (3.37) with  $\omega_0(k) = ck$  for non-observable assimilation times  $\Delta t^{obs}$  (red) and observable times (blue); see (4.23) for the definition of observability. The mean bias decays to zero in agreement with (4.29) for two different energy spectra. (Right) The evolution of the mean bias for incorrect resonant forcing with non-observable assimilation time for different incorrect amplitudes  $A^M$  of the time-periodic forcing  $\mathcal{F}_{\{1\}}^M$ ; note that small errors in the forcing lead to significant errors in the estimates for the mean dynamics.

RMS error in (2.30) are independent of the mean dynamics of the augmented system (4.4) as claimed. For incorrect forcing, i.e., when  $\Delta \mathcal{F}_{\{\ell\}m}^M \neq 0$ , the issue of the filter error sensitivity to the mean dynamics depends on the limit in (4.31).

The above facts indicate that for correct forcing the filter error in the mean decays asymptotically to zero for both observable and unobservable times. However, the decay rates may vary substantially depending on the observation time and the filtering algorithm which determines  $F_{\{\ell\}}$  and  $K_{\{\ell\}}^M$  in (4.28). For incorrect forcing the asymptotic filter error is given by (4.31) whose asymptotic behavior depends on the nature of the forcing error and on the observability of the filtering problem in (4.4). As discussed in [37, 60], the problem of filtering the sparsely observed dynamics in (3.32) with the constant advection speed  $c$  and  $2M+1$  observed nodes is unobservable when

$$\Delta t^{obs} = \frac{2\pi Q}{c(2M+1)}, \quad Q \in \mathbb{Z}^+. \tag{4.32}$$

On the other hand, the observability of (3.32) is always satisfied in the selective damping case  $\mu \neq 0$ . The effects non-observability when filtering (3.32) are particularly important when filtering in the presence of resonant forcing in the truth signal in (3.36)

so that the  $\omega_0(k) = \omega_k = ck$  for all  $|k| \leq M$  in (3.37). In figure 4.1 we illustrate these issues for filtering the sparsely observed advection-diffusion dynamics (3.32) in the uniform damping case,  $\mu = 0, d \neq 0$  with FDKF (cf. Section 4.1) for both observable and unobservable assimilation times  $\Delta t^{obs}$ . The two columns in figure 4.1 correspond to filtering with correct forcing (left) and incorrect forcing (right) with time-dependent error in the forcing amplitude (cf. (3.37))

$$\Delta \mathcal{F}_{\{1\}^m}^M = \begin{cases} \frac{\Delta A_k e^{ick(m+1)\Delta t}}{d} (1 - e^{-d\Delta t}), & \text{for } k = 1, \\ 0, & \text{for } k > 1. \end{cases} \quad (4.33)$$

For correct forcing the mean residual decays to zero, as predicted in the considerations above and summarized in Facts 1-4, but the rate of decay of the mean dynamics strongly depends on the observability of the system. For time-dependent error in the forcing and unobservable times satisfying (4.32) the mean residual does not vanish and may cause substantial contribution to the filter error from the mean, as shown in figure 4.1 for  $A^M \neq A$  for two different energy spectra. We also note without providing detailed numerical evidence that when the forcing error is independent of time, the asymptotic mean (4.31) tends to zero in a wide variety of cases.

**4.2.2. Correlated filter estimates for modes within the same aliasing set.** Filtering spatially extended turbulent systems with reduced filters and sparse observations can lead to artificial correlations in the filter estimates. Depending on applications this artifact might be either irrelevant or it may lead to spurious biases in the filter estimates. For example, these correlations are irrelevant for the skill of the cheap reduced algorithms with improved observability, introduced in [37, 60] and described in Section 4.1, which were designed for recovering the primary modes  $|\ell| \leq M$  in each aliasing set  $\mathcal{A}(\ell)$  (4.2) in regimes with steep energy spectra and rapidly decorrelating aliasing modes. On the other hand, such correlations need to be avoided in situations when robust estimates of the aliasing modes are required.

Below we show that the artificial correlations in imperfect filtering algorithms arise through the combined effect of the model reduction and Kalman filtering which results in the same information flow into the unobserved modes and degeneracy of the covariance matrix for the filter estimates  $\mathcal{C}_{(22)}$  in (4.6). The analytical treatment of this issue seems difficult. However, the mechanism for the occurrence of the unwanted correlations between the aliasing modes can be intuitively understood by considering the Kalman gain in the reduced algorithms of Section 4.1 in dynamical regimes where the design assumptions for these filters are deliberately and explicitly violated. Specifically, we consider the extreme case of filtering the dynamics with equipartition energy spectrum with SDAF and RFDKF. Recall that the assumption of memoryless dynamics in SDAF for the unresolved modes  $\hat{u}_{k_i}$ , with  $i > P_0$  in each aliasing set  $k_i \in \mathcal{A}(\ell)$ , implies that both the forecast model covariance  $R_{m+1|m}^M$  and the posterior covariance  $R_{m+1|m+1}^M$  have a diagonal structure for the modes with  $i > P_0$  (cf. (4.16)). Thus, in the simple configuration with  $P = 3$  and  $P_0 = 1$  and the observation operator  $G_P = (1 \ 1 \ 1)$  the model covariance for the primary mode  $\ell = k_1$  and the aliasing modes  $k_2, k_3$  is given by

$$R_{\{k_1\}} = \begin{pmatrix} R_{k_1} & 0 & 0 \\ 0 & \hat{r}_{k_2} & 0 \\ 0 & 0 & \hat{r}_{k_3} \end{pmatrix},$$

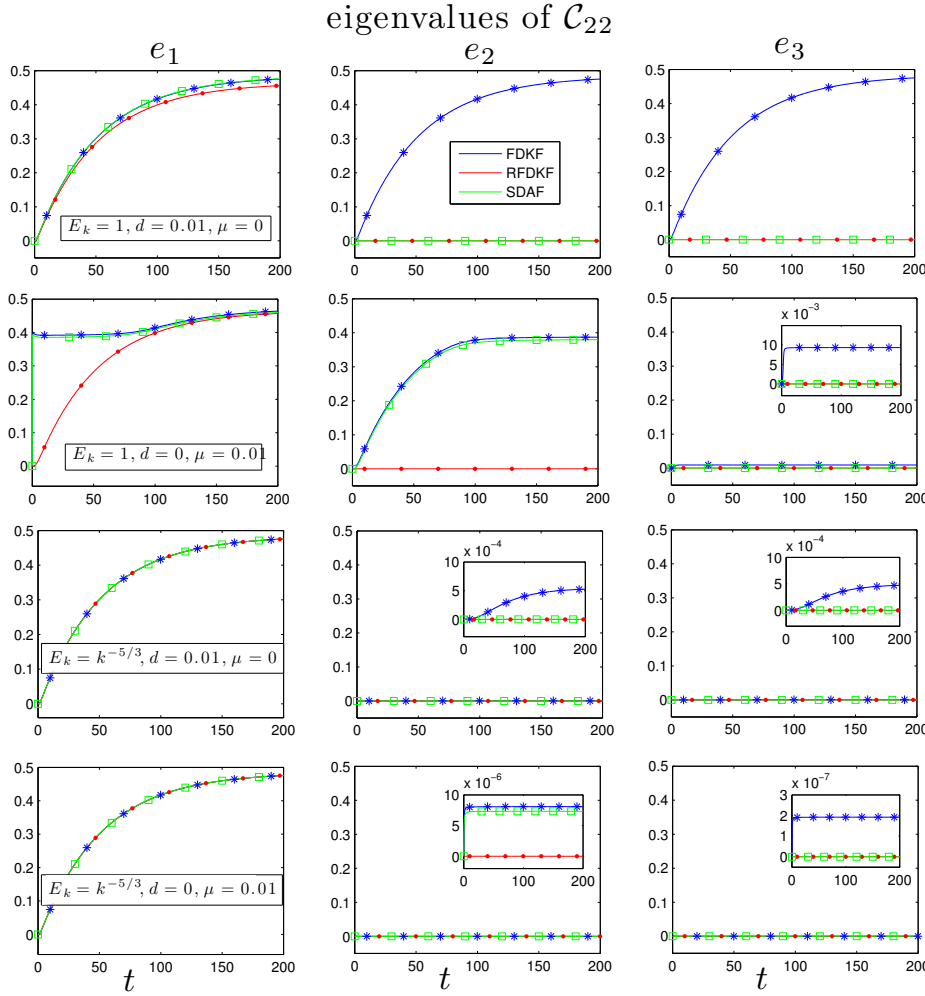


FIG. 4.2. Artificial correlations in the reduced filter estimates based on sparse observations (indicated by  $e_2, e_3 \sim 0$ ) which cause the lack of information for recovery of the aliased modes according to the relative entropy measure (2.23). Time evolution of the eigenvalues of the covariance  $C_{\{1\}22}$  in (4.6) associated with the filter estimates for the aliasing modes in  $\mathcal{A}(1)$  in the turbulent advection-diffusion dynamics (3.32) with decaying mean for different energy spectra with either uniform or selective damping. The first eigenvalue is aligned with the primary mode in  $\mathcal{A}(1)$ . The advection diffusion dynamics in (3.32) computed with  $N=61, M=20, P=3, c=1$ .

where the variance  $R_{k_1}$  of the primary mode is updated via the Kalman filter and the time-independent model noise variances,  $\hat{r}_{k_2}, \hat{r}_{k_3}$ , are tuned to the equilibrium energy of (3.32) via  $\hat{r}_{k_i} = \hat{\sigma}_{k_i}^2 / (2\Re[\lambda_{k_i}])$ ,  $i = 2, 3$ . Then, the Kalman gains (2.11) for SDAF and RFDKF (see Section 4.1) become, respectively,

$$K_{\{k_1\}}^{\text{SDAF}} = R_{\{k_1\}} G_P^* \left( G_P R_{\{k_1\}} G_P^* + \hat{r}^o \right)^{-1} = \begin{pmatrix} R_{k_1}/R_\Sigma \\ \hat{r}_{k_2}/R_\Sigma \\ \hat{r}_{k_3}/R_\Sigma \end{pmatrix}, \quad K_{\{k_1\}}^{\text{RFDKF}} = \begin{pmatrix} R_{k_1}/R_\Sigma \\ 0 \\ 0 \end{pmatrix}, \tag{4.34}$$

where  $R_\Sigma = R_{k_1} + \hat{r}_{k_2} + \hat{r}_{k_3} + \hat{r}^o$ . When filtering the spatially extended system in (3.32) with uniform damping,  $d \neq 0, \mu = 0$ , and the equipartition energy spectrum,  $\hat{r}_{k_1} = \hat{r}_{k_2} = \hat{r}_{k_3} = \hat{r}$ , the Kalman gains for SDAF and RFDKF become

$$K^{\text{SDAF}} = \begin{pmatrix} R_{k_1}/R_\Sigma \\ \hat{r}/R_\Sigma \\ \hat{r}/R_\Sigma \end{pmatrix}, \quad K^{\text{RFDKF}} = \begin{pmatrix} R_{k_1}/R_\Sigma \\ 0 \\ 0 \end{pmatrix}, \quad (4.35)$$

so that the information flow into the unresolved aliasing modes is either the same or zero, suggesting correlations in the filter estimates for the unresolved aliasing modes.

Further evidence of correlations in the unresolved aliasing modes is shown in figure 4.2 in more complicated configurations based on numerical experiments; these examples are computed using the reduced filters for different energy spectra in the sparsely observed truth dynamics (3.32) with either the uniform or selective damping, and  $P=3$  Fourier modes in a single aliasing set  $\mathcal{A}(1)$ . The presence of spurious correlations in the reduced filter estimates can be observed by computing the eigenvalues of the covariance matrix,  $\mathcal{C}_{(22)}$ , of the filter estimates in (4.6). It is important to stress here that the use of the reduced filters from Section 4.1 for filtering the dynamics in (3.32) with equipartition energy spectrum is technically unjustified, since for  $E_k = 1$  the conditions (4.13) allowing for derivation of the reduced filters are violated; nevertheless, these examples serve as a useful illustration of artifacts which might occur in reduced-order Kalman filtering of spatially extended systems based on sparse observations and imperfect forecast models. The tough dynamical regimes imposed on the reduced filters help reveal important differences between skill of the reduced filters outlined in Section 4.1 for recovering the statistics of the unresolved modes. Regardless of the energy spectrum in the dynamics of (3.32), there are strong correlations in the RFDKF estimates for the unresolved aliasing modes, i.e., the only non-zero eigenvalue is aligned with the primary mode and  $\text{rank}(\mathcal{C}_{(22)}) = 1$ . Unsurprisingly, when filtering with SDAF in the extreme case of uniform damping,  $d \neq 0, \mu = 0$ , the correlations in the filter estimates for the unresolved aliasing modes are also significant. However, when filtering the dynamics in (3.32) with selective damping,  $\mu \neq 0$ , SDAF performs comparably to FDKF even for the equipartition energy spectrum (second row in figure 4.2). The reduced filters, SDAF and RFDKF, were designed for recovering only the primary modes in the aliasing sets in dynamical regimes with steep energy spectra. Therefore, the overall good performance of SDAF, even in the dynamical regimes for which it was not designed, is particularly encouraging.

The main technical drawback arising from the presence of such correlations is that the relative entropy in (2.23) is undefined on the marginal densities of the reduced filter estimates within each aliasing set; this is the main reason for projecting the relative entropy measure on the primary mode in the filter skill estimates. An alternative strategy is to use the Hellinger distance (2.40) between two probability densities which is well defined and bounded even in such cases; however the drawback here is that, in contrast to the relative entropy, the Hellinger distance does not have the information-theoretic interpretation.

**4.3. Filtering sparsely observed advection-diffusion equation.** Here, similar to the case with plentiful observations studied in Section 3.5.1 we consider discrete-time filtering of the stochastically forced advection-diffusion dynamics (3.32); the advection-diffusion dynamics represents one particular but rich example of the canonical dynamics in (3.1). In contrast to the case considered in Section 3.5.1, we focus on a much tougher case of filtering the turbulent spatially extended dynamics



of (3.32) with sparse observations. As already indicated in Section 4 and [59, 37, 60], the problem of filtering the  $2N+1$  Fourier modes of the spatially extended system (3.32) with uncorrelated initial conditions and  $2M+1$  spatially sparse observations (4.3) splits into  $M+1$  independent filtering problems for the  $P$ -dimensional sets of aliasing Fourier modes, such that  $(2M+1)P=2N+1$ .

The truth dynamics in the examples studied in this section is generated by solving for the  $2N+1=123$  Fourier modes in (3.3) associated with the advection-diffusion dynamics (3.32) in two distinct dynamical regimes: (i) uniform damping with  $d \neq 0, \mu=0$ , and (ii) selective damping with  $d=0, \mu \neq 0$ ; the default parameter values are  $c=1$ ,  $d=0.01$ , and  $\mu=0.01$  so that even in the uniformly damped setting the amplification factors in the perfect system satisfy  $|F_k|=e^{-d\Delta t} < 1$  for a wide range of observation times, implying strong asymptotic stability. The truth signal is observed at every third node of the model grid, implying  $M=20$  observation nodes and  $P=3$  Fourier modes in each aliasing set  $\mathcal{A}(\ell)$ , associated with the  $2M+1=41$  primary wavenumbers ( $|\ell| \leq M$ ) resolved by the observation grid. Similar to the analysis of [37, 60] and to the plentiful observations case studied in Section 3.5.1, two different turbulent spectra are considered: (i) the equipartition energy spectrum,  $E_k \propto 1$ , and (ii) the steep energy spectrum,  $E_k \propto k^{-5/3}$ ; these spectra were chosen in order to highlight properties of the reduced filters in a wide range of dynamical scenarios.

The error in filtering the sparsely observed dynamics of the canonical turbulent system (3.32) is quantified using the information measures introduced in Section 2.2.1 in the form (4.10), (4.11), (4.12) which is more suitable for configurations with possible correlations in the filter estimates (see Section 4.2.2).

**4.3.1. Filter performance for estimating the turbulent advection-diffusion dynamics via reduced filters based on perfect dynamics.** Here, we utilize the superensemble framework introduced in Section 2 in order to examine the skill of FDKF and the reduced filtering algorithms, introduced in [37, 60] and outlined in Section 4.1. First, the superensemble filter error is examined in a single aliasing set of the turbulent advection-diffusion dynamics (3.32). Subsequently, the performance of FDKF and the reduced strategies is discussed for filtering the full spatially extended dynamics in (3.32) when contributions to the filter error from all the aliasing sets,  $\mathcal{A}(\ell)$ ,  $|\ell| \leq M < N$ , play a role. The analysis of this section involves dynamics with correct forcing in the filters; consequently, the effects of the biases in the estimates for the mean dynamics discussed in Section 4.2.1 are not present in this configuration; the effects biases in the mean dynamics induced by uncertainties in the forcing terms of the imperfect filters are discussed in Section 4.3.2.

#### REDUCED FILTER SKILL FOR DIFFERENT ENERGY SPECTRA

In figures 4.3 - 4.6 we compare the filter error in terms of the individual information measures (4.10), (4.11), (4.12) or based on the information criterion  $\mathcal{M}$  in (2.50); two different turbulent spectra,  $E_k=1$  and  $E_k \propto k^{-5/3}$ , in the dynamics of (3.32) with either the uniform damping or the selective damping are examined. In both cases we compare the filter error for the aliasing set of the large scale wavenumber  $\mathcal{A}(1)=\{1-40\ 42\}$  which contains two aliasing modes with comparable spatial scales; for steeper energy spectra, e.g.,  $k^{-5/3}$ , the aliasing modes contain little energy. The results are qualitatively similar for other aliasing sets and we do not show them here. In the tough configuration with uniform damping and equipartition energy spectrum the reduced filters have significantly worse skill than FDKF; however, even the worst performing reduced filter, RFDKF, beats the least-squares estimates obtained from the raw sparse data for sufficiently large observation noise. For the steep energy

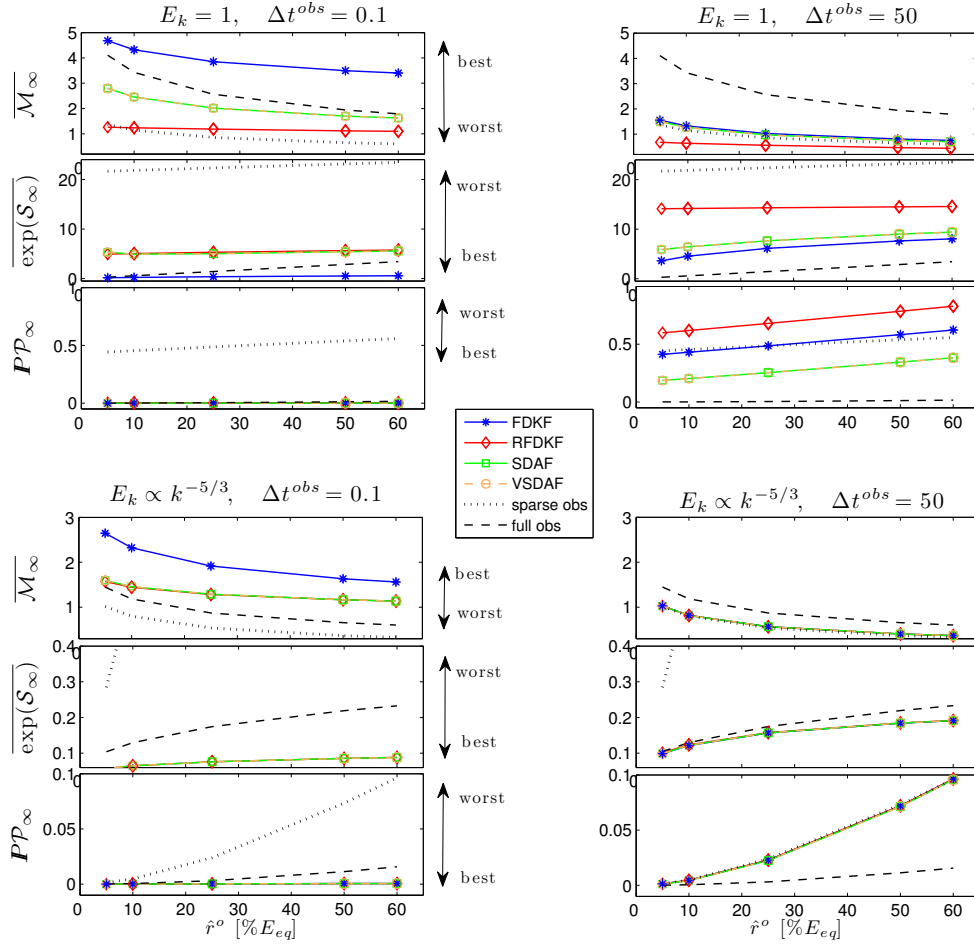


FIG. 4.3. Filtering the turbulent advection-diffusion dynamics (3.32) with uniform damping and decaying mean through sparse observations in the aliasing set  $\mathcal{A}(1)$  (parameters  $d=0.01, \mu=0, c=1, N=61, M=20, P=3$ ). The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  for two different data assimilation times,  $\Delta t^{obs}$ , and for two energy spectra  $E_k=1, E_k \sim k^{-5/3}$ ; the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and in figure 4.2. The reduced filters are described in Section 4.1. The dotted black lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.4 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.

spectrum,  $E_k \propto k^{-5/3}$ , all the examined filters perform very similarly for both the uniform damping (figures 4.3, 4.4) and the selective damping (figures 4.5, 4.6) which further reinforces the path-wise results of [37, 60] and validates the design principles for these filters. For the equipartition energy spectrum SDAF and VSDAF are essentially as good as the perfect filter while RFDKF performs poorly, especially for long observation times.

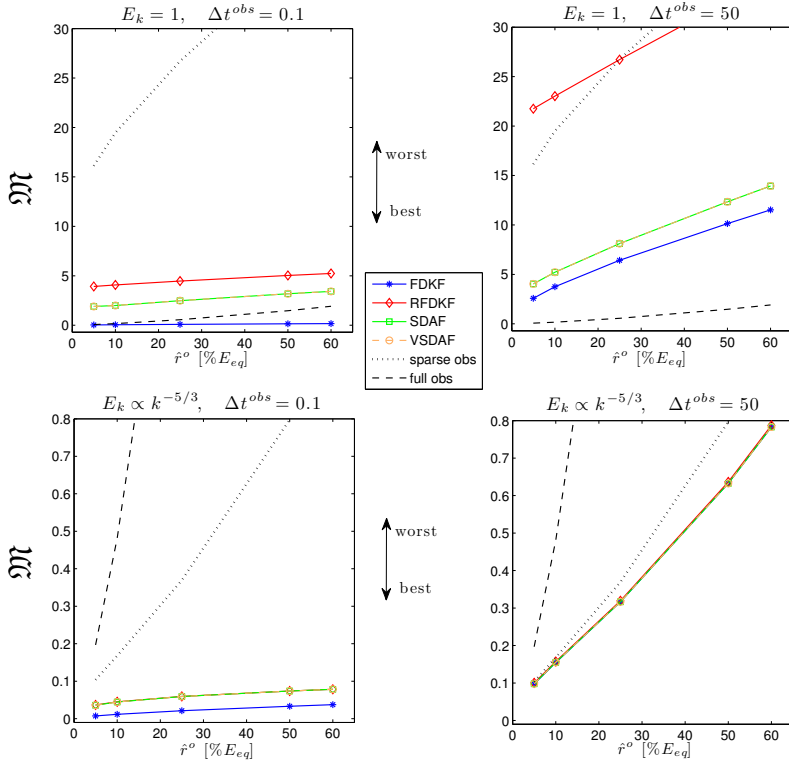


FIG. 4.4. The filter skill according to the information criterion  $\mathfrak{M}$  in (2.50) for the reduced filters (see Section 4.1) applied to weakly damped turbulent advection-diffusion dynamics (3.32) with uniform damping and decaying mean through sparse observations in the aliasing set  $\mathcal{A}(1)$  ( $d=0.01, \mu=0, c=1, N=61, M=20, P=3$ ). The dotted black lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.3 shows the individual information measures combined in (2.50).

In figures 4.7 and 4.8 we illustrate the dependence of the filter skill on the sparseness of the observation grid which is associated with the number,  $P$ , of Fourier modes in each aliasing set  $\mathcal{A}(\ell)$  of the primary modes  $|\ell| \leq M < N$  resolved by the observations. Here, we consider the aliasing set of the large-scale primary mode,  $\mathcal{A}(1)$ , in the tough dynamical regime of uniformly damped ( $\mu=0, d \neq 0$ ) advection diffusion dynamics in (3.32) in the absence of deterministic forcing and identical decorrelation time,  $\tau_{corr} = 1/d$ , for all modes; the effects of filter error due the non-decaying mean dynamics are considered in Section 4.3.2.

For the equipartition spectrum,  $E_k = 1$ , the overall filter skill deteriorates with the number of aliasing modes; this is not surprising since all modes in the aliasing set have the same energy and decorrelation time and filtering through the sparse observations of the primary modes is bound to be affected by the number of the unresolved aliased modes. Recall that the reduced filters outlined in Section 4.1 and introduced in [37] were designed for filtering turbulent dynamics with steep energy spectra. However, even in this tough regime with equipartition spectrum SDAF and

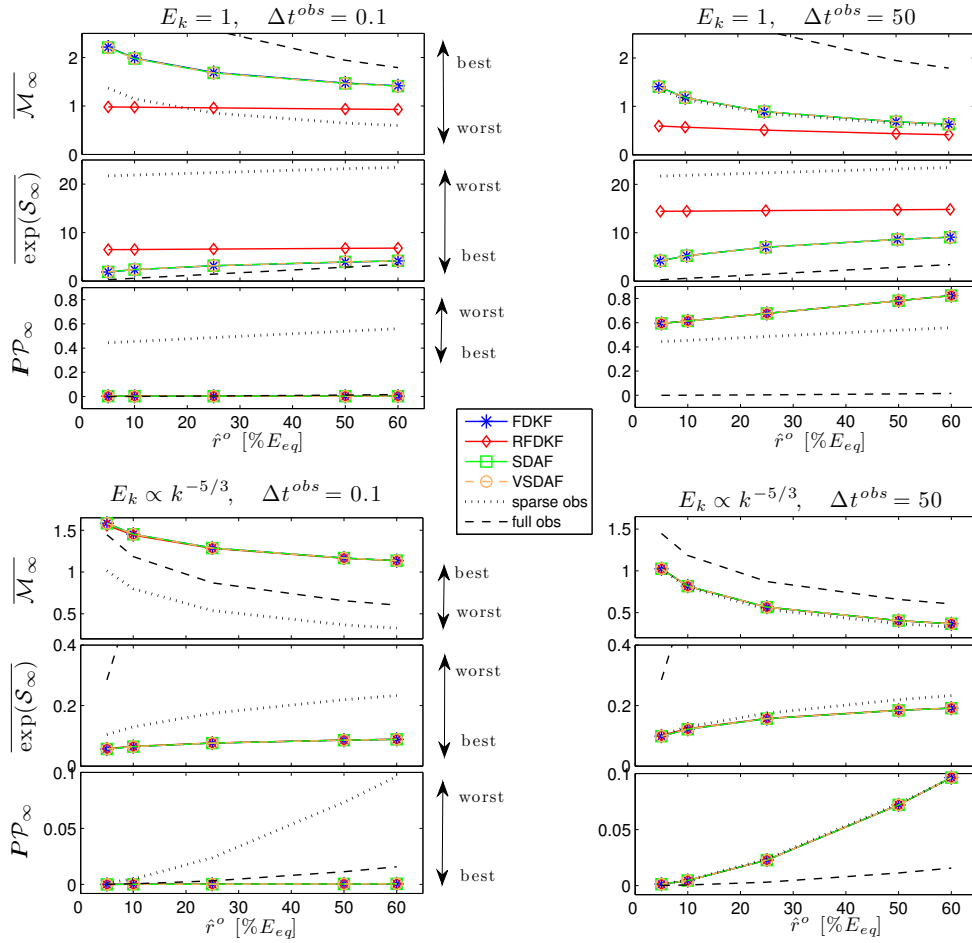


FIG. 4.5. Filtering the turbulent advection-diffusion dynamics (3.32) with selective damping and decaying mean through sparse observations in the aliasing set  $\mathcal{A}(1)$  (parameters  $d=0, \mu=0.01, c=1, N=61, M=20, P=3$ ); the unoptimized reduced filters used for state estimation are described in Section 4.1. The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  for two different data assimilation times,  $\Delta t^{obs}$ , and for two energy spectra  $E_k=1$  (top row), and  $E_k \propto k^{-5/3}$  (bottom row); the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.6 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.

VSDAF perform comparably to the perfect filter for large observation times time  $\Delta t^{obs} \sim \tau_{corr}$  while RFDKF performs well only in terms of the filter error entropy. For sufficiently short observation times  $\Delta t^{obs} \ll \tau_{corr}$  all imperfect filters outperform the least-squares estimates obtained from the raw sparse observations, but the skill of RFDKF deteriorates the fastest with the increasing number of aliasing modes  $P$ .

In the steep energy regime  $E_k \propto k^{-5/3}$  all imperfect filters are effectively indistinguishable from the perfect filter for long observation times, i.e.,  $\Delta t^{obs} \sim \tau_{corr}$ . The skill of all filters does not vary significantly with increasing sparsity; the lack

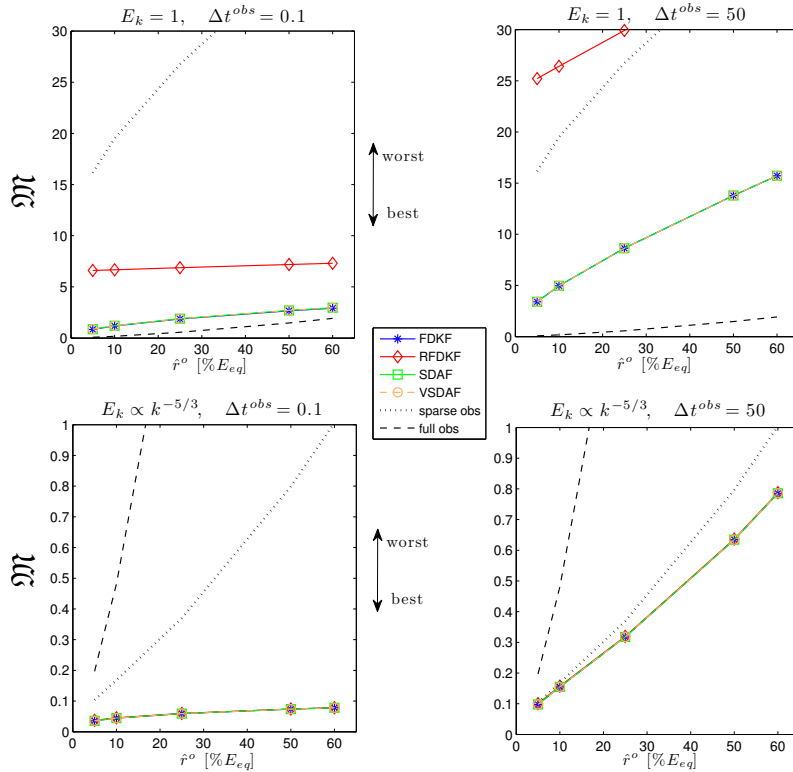


FIG. 4.6. The filter skill according to the information criterion  $\mathfrak{M}$  in (2.50) for the reduced filters (see Section 4.1) applied to weakly damped turbulent advection-diffusion dynamics (3.32) with selective damping and decaying mean through sparse observations in the aliasing set  $\mathcal{A}(1)$  ( $d=0, \mu=0.01, c=1, N=61, M=20, P=3$ ). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.5 shows the individual information measures combined in (2.50).

of skill sensitivity to the number of aliasing modes is a direct consequence of the steep energy spectrum with little energy in the aliasing modes. For short observation times  $\Delta t^{obs} \ll \tau_{corr}$  all the imperfect, reduced filters perform similarly and worse than FDKF; however, all the reduced filters beat the least-squares estimates obtained from the raw sparse observations.

ENSEMBLE FILTER ERROR FOR THE FULL SPATIALLY EXTENDED  
ADVECTION-DIFFUSION DYNAMICS

Finally, we utilize the superensemble framework and the information measures of filter skill introduced in Section 2.2 to examine the skill of the reduced filters algorithms for filtering the full sparsely observed turbulent dynamics of the stochastically forced advection-diffusion system (3.32). Here, the contributions from all the aliasing sets to filter error are taken into account when assessing the filter performance. The information measures of filter skill introduced in Section 2.2.1 are particularly useful and revealing in this complex filtering setting where the sparse observations induce correlations between multiple spatial scales of the turbulent dynamics.

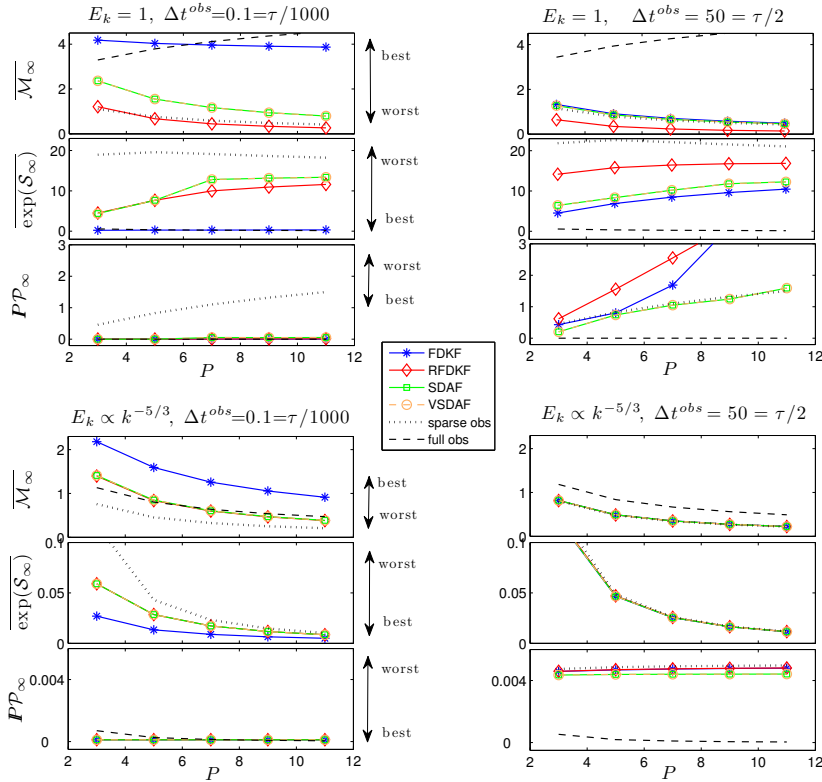


FIG. 4.7. Dependence of filter skill on sparsity of observations  $P$  in filtering the turbulent advection-diffusion dynamics (3.32) with uniform damping and decaying mean is filtered through sparse observations ( $\hat{r}^{obs} = 1\%E_1, d = 0.01, \mu = 0, c = 1, M = 20$ ) in the aliasing set  $\mathcal{A}(1)$ ; the unoptimized reduced filters used for state estimation are described in Section 4.1. The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^\circ$  for two different assimilation times,  $\Delta t^{obs}$ , and for two energy spectra  $E_k = 1$  (top row), and  $E_k \sim k^{-5/3}$  (bottom row); the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.8 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example which is used to rank the filters.

Figures 4.9 and 4.10 illustrate the performance of the reduced filters for estimating the full turbulent advection-diffusion dynamics of (3.32) for two different energy spectra as a function of the observation noise amplitude. In this setting SDAF and VSDF perform well in dynamical regimes ranging from equipartition to steep energy spectra and greatly outperform the least-squares estimates obtained from the raw sparse observations; recall again that the reduced filters were originally designed to cope with steep energy spectra [37]. RFDKF performs well only in the steep energy spectrum regime for which it was originally designed (see Section 4.1 and [37, 60]).

In summary, we make the following points regarding the skill of the reduced filtering algorithms for estimating the turbulent dynamics with no significant mean dynamics component:

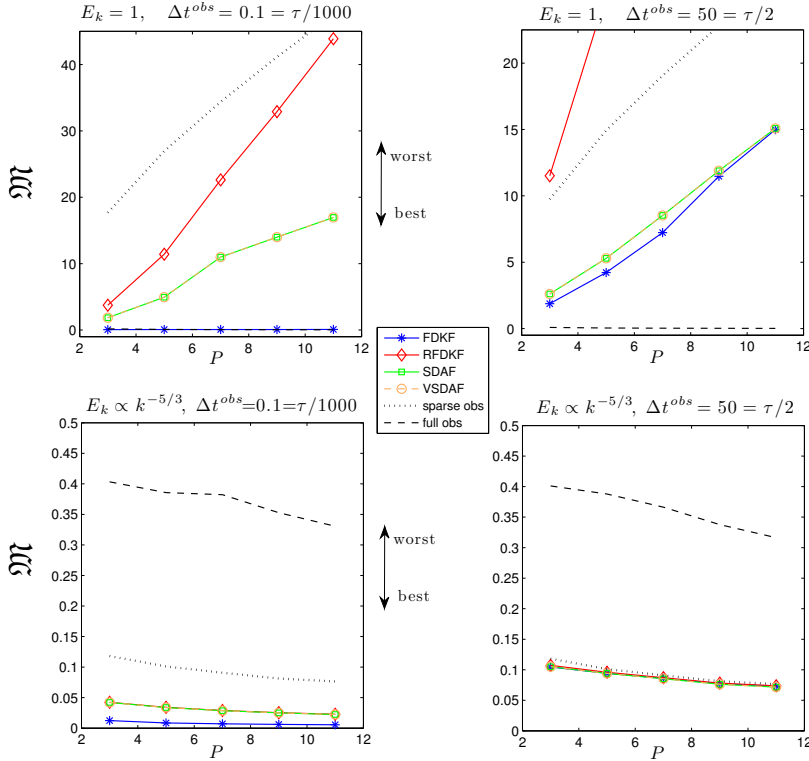


FIG. 4.8. Dependence of filter skill on sparsity of observations  $P$  in filtering the turbulent advection-diffusion dynamics (3.32) with the reduced filters described in Section 4.1 in the aliasing set  $\mathcal{A}(1)$ . The filter skill of the reduced filters is assessed according to the information criterion  $\mathfrak{M}$  in (2.50) for filtering a weakly damped turbulent advection-diffusion dynamics (3.32) with selective damping and decaying mean ( $d=0, \mu=0.01, c=1, N=61, M=20, P=3$ ). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.7 shows the individual information measures combined in  $\mathfrak{M}$ .

- When filtering the spatially extended dynamics with imperfect filters the skill based on the RMS error (2.30) or the entropy of the filter error (2.16) is an insufficient measure of the filter performance; filters with comparable error entropy/RMS error can have very different pattern correlation properties (figures 4.3 - 4.6). The results for the RMS error/ filter error entropy agree with those of [37].
- Performance of the reduced filters for estimating sparsely observed turbulent dynamics (3.32) with steep energy spectra and decaying mean:
  - For selective damping FDKF and the reduced filters, SDAF, VSDAF, and RFDKF are practically indistinguishable in terms of the information measures introduced in Section 2.2. These statistical, superensemble-based results further validate the path-wise results of [37, 60] and the design principles for constructing these reduced filters.
  - For uniform damping the reduced filters are comparable to FDKF for



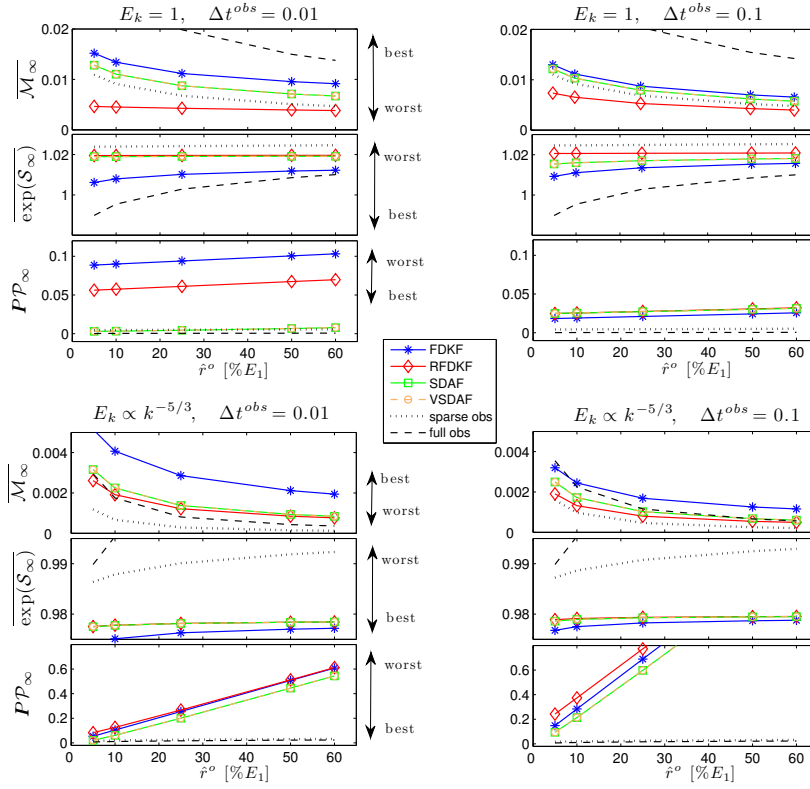


FIG. 4.9. **Filtering the full turbulent advection-diffusion dynamics (3.32) with uniform damping ( $\mu=0, d=0.01, c=1$ ) and decaying mean through sparse observations ( $P=3, M=20, N=61$ ).** The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  for two different assimilation times,  $\Delta t^{obs}$ , and for two energy spectra  $E_k=1$  (top row), and  $E_k \sim k^{-5/3}$  (bottom row); the relative entropy measure  $\mathcal{P}$  in (2.23) is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.10 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.

filtering aliasing sets of the primary modes with sufficiently short decorrelation times relative to the observation time  $\Delta t^{obs}$  (see figure 4.3); these cheap reduced filters have a worse skill than FDKF on the aliasing sets of primary modes with long decorrelation times  $\Delta t^{obs} \sim \tau_{corr}$ .

- The skill of all considered filters does not vary significantly with increasing sparsity of the observations  $P$ ; the lack of skill sensitivity to the number of aliasing modes is a direct consequence of the steep energy spectrum with little energy in the aliasing modes.
- Performance of the reduced filters when estimating sparsely observed turbulent dynamics (3.32) with decaying mean in the difficult setting with equipartition energy spectrum:
  - For selective damping the reduced filters, SDAF and VSDAF, are essentially as good as the perfect filter while RFDKF performs poorly especially for long observation times; these are very positive results be-

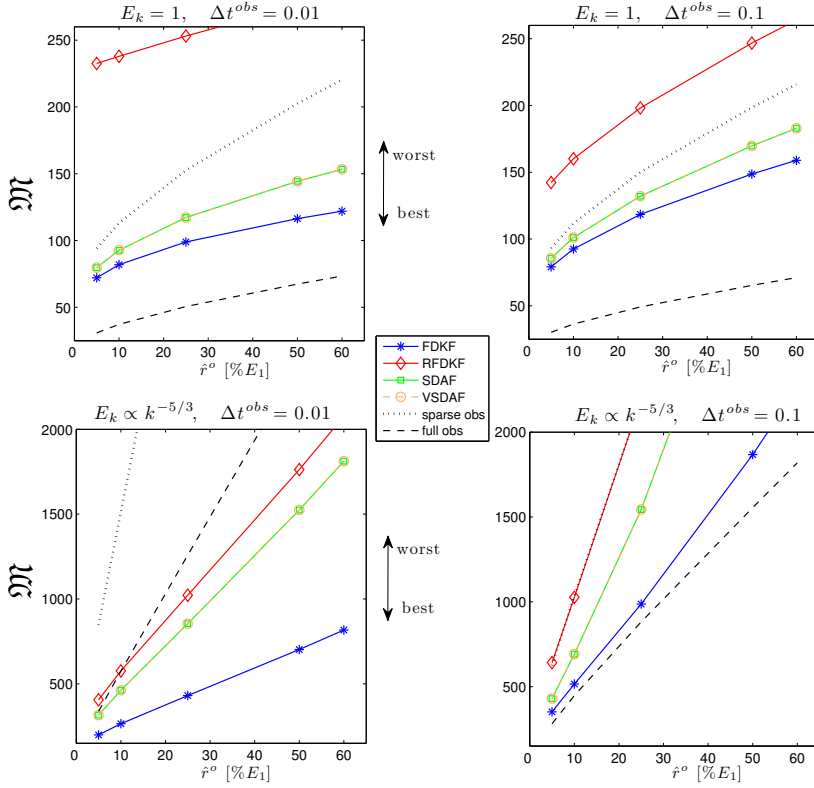


FIG. 4.10. The filter skill according to the information criterion  $\mathfrak{M}$  in (2.50) for the **reduced filters** (see Section 4.1) applied to filter the **full weakly damped turbulent advection-diffusion dynamics (3.32) with uniform damping** ( $\mu=0, d=0.01, c=1$ ) and **decaying mean through sparse observations** ( $P=3, M=20, N=61$ ). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.9 shows the individual information measures combined in  $\mathfrak{M}$ .

cause none of the reduced filters were designed to cope with rough and shallow energy spectra, like the equipartition spectrum.

- For uniform damping all the reduced filters have good and comparable skill for aliasing sets with short decorrelation times ( $\Delta t^{obs} \ll \tau_{corr}$ ) (figure 4.3); for aliasing sets with long decorrelation times ( $\Delta t^{obs} \sim \tau_{corr}$ ) SDAF and VSDAF approach the skill of FDKF while RFDKF has significantly worse mutual information/pattern correlation and relative entropy (see figures 4.3, 4.5, 4.7).
- Unsurprisingly, the skill of all filters deteriorates with the number  $P$  of the aliasing modes in the aliasing set. The skill of RFDKF deteriorates at the fastest rate, especially for long observation times ( $\Delta t^o \sim \tau_{corr}$ ). SDAF performs almost as well as FDKF for any  $P$  especially for long observation times (see figures 4.7, 4.8).
- For sufficiently large observation noise filtering turbulent dynamics through sparse observations even with the poorest performing reduced

filter, RFDKF, generally outperforms the estimates obtained by the least squares estimates obtained from the raw sparse observations (dotted lines in figures 4.3, 4.5, 4.7).

- The superensemble estimates of performance of FDKF and the reduced filters based on the information theoretic measures further reinforce the path-wise estimates of [37, 60].

**4.3.2. Filter performance for estimating the turbulent advection-diffusion dynamics via reduced filters with model error.** Here, we revisit and verify the earlier path-wise results of [37] associated with filter skill in the presence of non-trivial mean dynamics in the truth signal using the information-theoretic framework introduced in Section 2.2. One important feature associated with filtering turbulent signals with non-trivial mean dynamics is that for some observations times,  $\Delta t^{obs}$ , filtering with the perfect model might be non-observable (see Section 4.2.1 and [37, 60]). This effect is particularly severe for the advection-diffusion dynamics in (3.32) where the observability condition (4.23) fails simultaneously for all wave numbers when  $\Delta t^{obs}$  satisfies (4.32) (see Section 4.2.1 and [37, 60]). Moreover, when filtering the turbulent dynamics (3.32) with weak uniform damping ( $\mu = 0, d \ll 1$ ), the observation times  $\Delta t^{obs}$  might be below the decorrelation time  $\tau_{corr} = 1/d$  of all spatial wavenumbers. These features combine to make this problem an extremely difficult testbed for filtering sparsely observed turbulent systems especially with equipartition energy spectrum and resonant forcing,  $\mathcal{F}_m$ , with  $\omega_0(k) = ck$  in (3.37) which often leads to the emergence of large-scale coherent structures in the turbulent field. Some aspects related to this issue were already discussed in Section 4.2.1 in the superensemble framework and in [37, 60] in the path-wise framework where it was shown that incorrect forcing in the filters can lead to large biases in the estimates for the mean dynamics.

On the other hand, the observability condition (4.23) which is crucial for filter accuracy is always satisfied for the canonical model (3.32) in the presence of selective damping  $\mu \neq 0$  when the dissipation,  $\hat{\gamma}_k$ , in (3.32) strictly increases with  $|k|$  so that the determinant in (4.23) is always non-zero. This observation, pointed out in [37, 60], provides one possible computational filtering strategy that avoids the failure of observability when filtering the uniformly damped dynamics in (3.32); this strategy relies on filtering the uniformly damped turbulent truth signal generated from (3.32) with the uniform damping  $\hat{\gamma}_k = d$  using a modified forecast model in the filter with additional selective damping, i.e.,  $\tilde{\gamma}_k = d + \tilde{\mu}k^2$ ,  $0 < \tilde{\mu} \ll 1$  in the forward operator  $F^M = \exp(-(\tilde{\gamma}_k - \hat{P}_k)\Delta t)$  of the discrete time filter forecast, similar to that in (3.37). This strategy was already analyzed within the path-wise framework in [37] where it was shown that the performance of this approach depends on the signal-to-noise ratio of the true mean signal being filtered; those earlier results can be summarized as follows:

- Small signal-to-noise ratio (e.g., with decaying mean or non-resonant periodic forcing). In this regime, the use of modified reduced filters (see Section 4.1) with additional weak selective damping leads to some improvements in estimating the dynamics although the improvement in filter skill is not significant. FDKF was shown to be the best method for equipartition spectrum while SDAF had a significant but worse skill. For smoother spectra all the cheap reduced filters had comparable performance to FDKF because the signal-to-noise ratio for smoother spectra is slightly larger than the one for equipartitioned spectra, especially in the high wave numbers.

- Large signal-to-noise ratio (e.g., dynamics with resonant periodic forcing) when the errors in the filter forcing term induce significant errors in the mean dynamics, as shown in Section 4.2.1. In this situation, SDAF and RFDKF produced the best filter performance since they both satisfy the observability condition and do not introduce model errors into the dynamics of the primary (observed) modes [37, 60]. The modified FDKF filter with the additional selective damping term,  $\tilde{\gamma}_k$ , fails even when the observability condition (4.23) is reinstated; this is because the model errors introduced through the additional selective damping are too large in the presence of large amplitude coherent events in the truth dynamics. The unmodified FDKF algorithm fails completely in this configuration and with these forcing errors because the observability condition (4.23) is practically violated.

Below, we consider the same filtering setup and strategy as the one described in [37, 60], but we assess the performance of the reduced filters with model error within the information-theoretic framework. By analogy to the setup used in [37, 60], we assume no discretization errors in the reduced filters but the model error is purposefully introduced by adding a selective damping through a diffusive coefficient  $\tilde{\mu}=0.01$  in the filter forecast model (3.37), so that  $\tilde{\gamma}_k = d + \tilde{\mu}k^2$  in the discrete-time forecast dynamics  $F^M = \exp(-(\tilde{\gamma}_k - \hat{P}_k)\Delta t)$ . The idea of modifying the perfect forward dynamics is motivated by the aforementioned fact that the dynamics generated by  $F^M$  with weak selective damping in  $\tilde{\gamma}_k$  leads to a filtering problem which is always observable. However, as already shown in [37, 60] within the path-wise framework and confirmed below in the superensemble framework, this strategy leads to large biases in the estimates for the mean dynamics and is generally not suitable for improving the performance of imperfect filters.

#### PERFORMANCE OF REDUCED FILTERS IN THE PRESENCE OF ERRORS IN THE RESONANT FORCING

Here, we consider filtering the sparsely observed turbulent dynamics of (3.32) with resonant forcing  $\mathcal{F}_m$ , in (3.37) with  $\omega_0(k) = ck$  using the reduced filters outlined in Section 4.1. The reduced filters are based on the perfect dynamics but use either the correct resonant forcing or introduce time-dependent forcing errors  $\Delta\mathcal{F}_m$  via (4.33), similar to the setup discussed in Section 4.2.1. The filter skill is assessed via the information measures outlined in Section 2.2.1. Figures 4.11 and 4.12 illustrate the interplay between the non-observability (4.23) and the errors in the resonant periodic forcing used by FDKF and the reduced filters. Figure 4.11 shows the individual information measures for the filter estimates within the aliasing set  $\mathcal{A}(17)$ ; the filter error entropy and the mutual information are averaged over the aliasing modes while the relative entropy is projected onto the primary mode for reasons discussed in Section 4.2.2. The effects of the mean biases in the filter estimates of FDKF and RFDKF are evident in the relative entropy measure which, for non-observable assimilation times  $\Delta t^{obs}$  (see (4.32)) and in the presence of the forcing errors (bottom right), is dominated by the signal component  $sig_{\mathcal{P}_\infty}$  (see Section 2.2.1 and (2.24), (2.25) for definitions of the signal and dispersion in the relative entropy). Note that the deterioration of the skill of FDKF due to the mean biases leaves the error entropy and mutual information for FDKF largely unaffected. While RFDKF has no skill for both the correct and incorrect resonant forcing and non-observable times, the performance of FDKF is significantly reduced by the mean biases, as shown in figure 4.12. Similar skill deterioration in FDKF is observed for the smoother energy spectrum  $E_k \propto k^{-5/3}$ ; we skip the illustration here for brevity. For correct resonant forcing the effects of

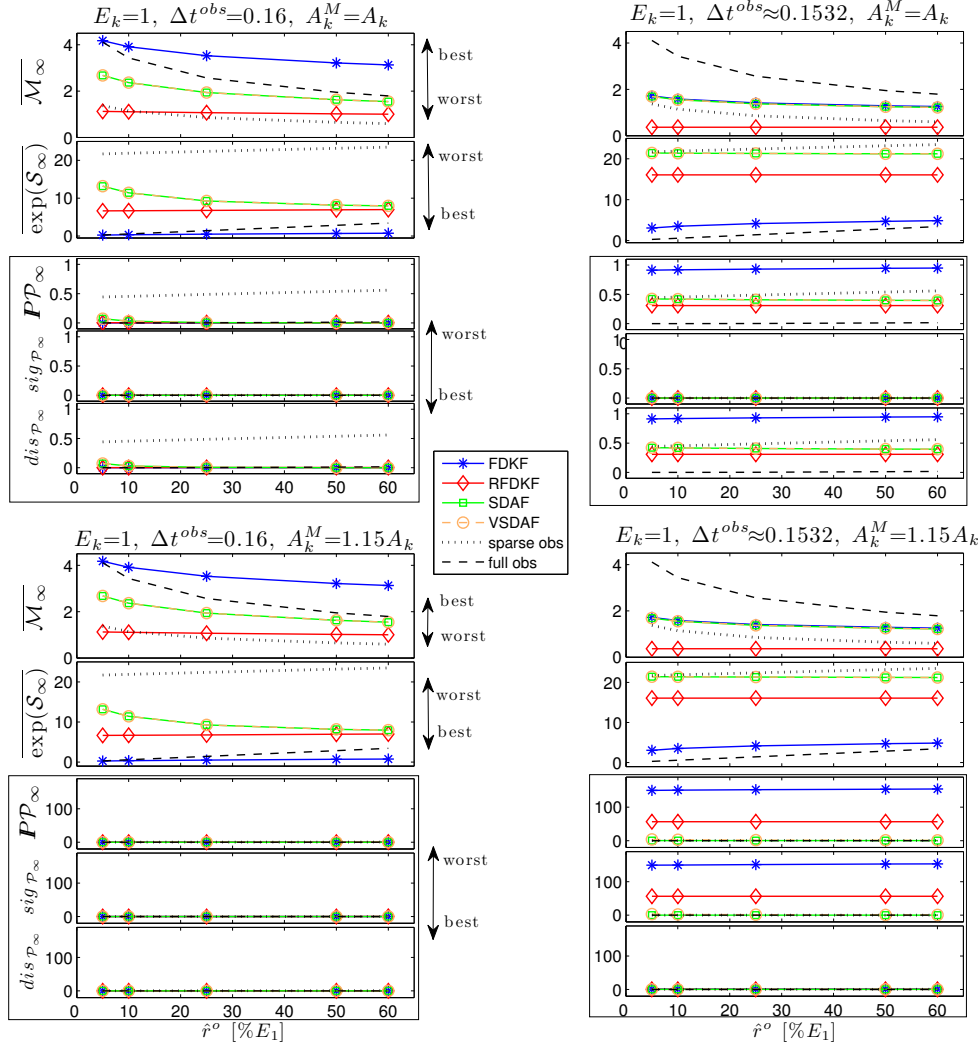


FIG. 4.11. Filtering the aliasing modes in  $\mathcal{A}(17)$  in the turbulent advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k=1$ , non-zero mean, resonant forcing in (3.37) and uniform damping ( $d=0.01, \mu=0, c=1$ ) through sparse observations ( $P=3, M=20, N=61$ ); the reduced filters used for estimating the dynamics are described in Section 4.1. The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  in cases when the dynamics (3.32) is observable (left column) and non-observable (right column); see (4.23) for the observability condition. The top row shows results for correct forcing in the filters and the bottom row corresponds to incorrect forcing with error  $\Delta\mathcal{F}$  given in (4.33); the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.12 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.

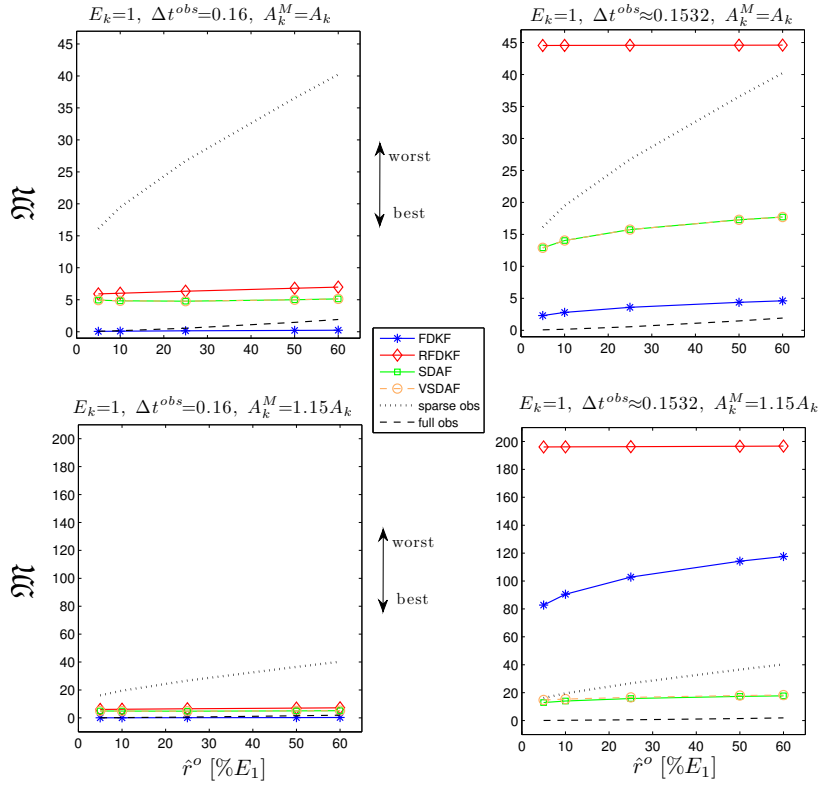


FIG. 4.12. The filter skill according to the information criterion  $\mathfrak{M}$  in (2.50) for filtering the aliasing modes in  $\mathcal{A}(17)$  in the turbulent advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k=1$ , non-zero mean dynamics, resonant forcing (3.37) and uniform damping ( $d=0.01, \mu=0, c=1$ ) through sparse observations ( $P=3, M=20, N=61$ ); the reduced filters used for state estimation are described in Section 4.1. The cases of observable dynamics (left column) and non-observable dynamics (right column) are shown for both correct forcing (top row) and incorrect forcing (bottom row). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.9 shows the individual information measures combined in (2.50).

non-observability are much less pronounced and not shown.

PERFORMANCE OF REDUCED FILTERS WITH ADDITIONAL SELECTIVE DAMPING

Here, we consider the skill of FDKF and the reduced filters described in Section 4.1 with additional selective damping in the forecast dynamics in order to filter the turbulent, uniformly damped, and sparsely observed advection-diffusion dynamics (3.32) with resonant forcing  $\mathcal{F}_m$ , in (3.37) with  $\omega_0(k) = ck$ . In figures 4.13-4.14, we illustrate the performance of FDKF and the reduced filters, SDAF, VSDAF, RFDKF, for filtering the uniformly damped turbulent dynamics in (3.32) after modifying the filters by adding weak selective damping  $0 < \tilde{\mu} \ll 1$ . Recall that while the reduced filters are designed to satisfy the observability condition (4.23) on the primary modes,  $\hat{u}_\ell, |\ell| \leq M \leq N$ , even in the uniformly damped case, FDKF satisfies the observability condition only after adding the selective damping. In this configuration with trivial mean dynamics in the truth signal the addition of the weak selective damping severely

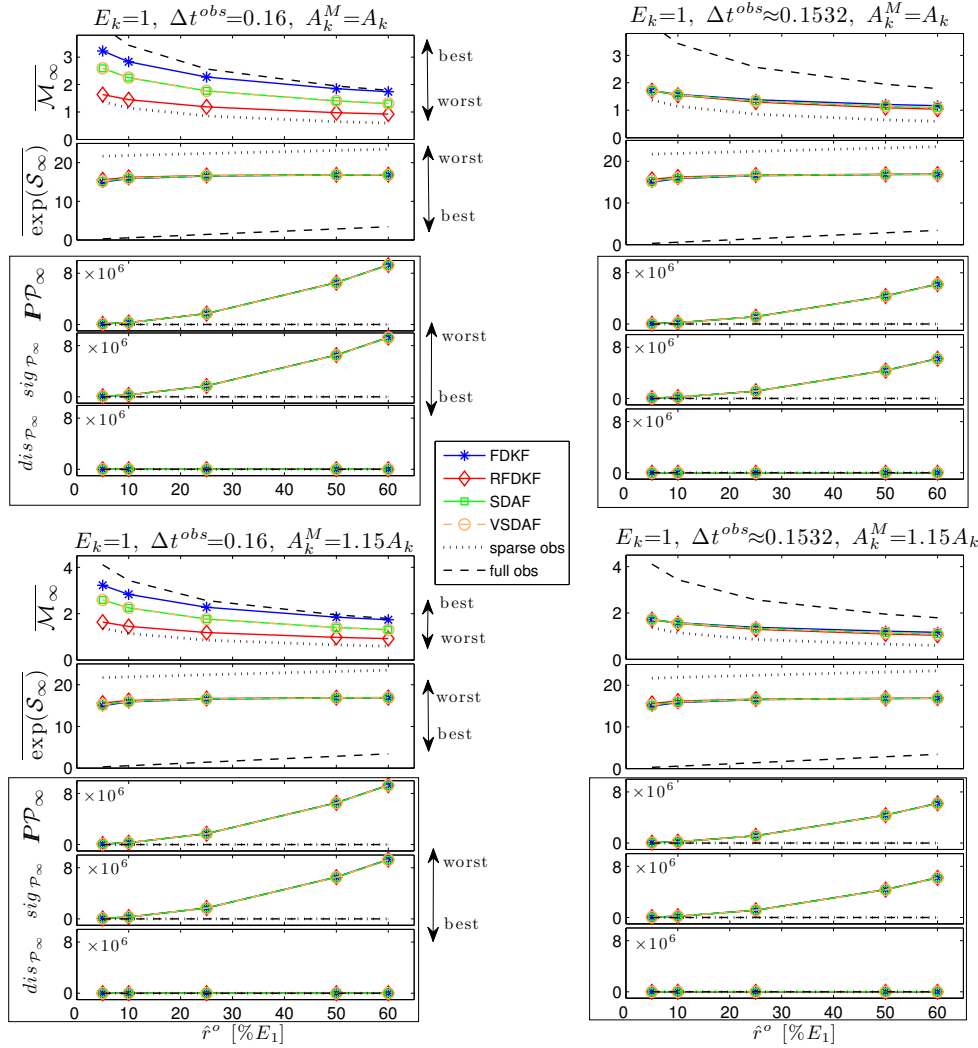


FIG. 4.13. Viscosity-modified filtering of the aliasing modes in  $\mathcal{A}(17)$  in the turbulent advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k = 1$ , non-zero mean, resonant forcing (3.37) and uniform damping ( $d = 0.01, \mu = 0, c = 1$ ) through sparse observations ( $P = 3, M = 20, N = 61$ ); the reduced filters used for state estimation are described in Section 4.1. The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  in cases when the dynamics (3.32) is observable (left column) and non-observable (right column); see (4.23) for the observability condition. The top row shows results for correct forcing in the filters and the bottom row corresponds to incorrect forcing with error  $\Delta\mathcal{F}$  given in (4.33); the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.14 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.



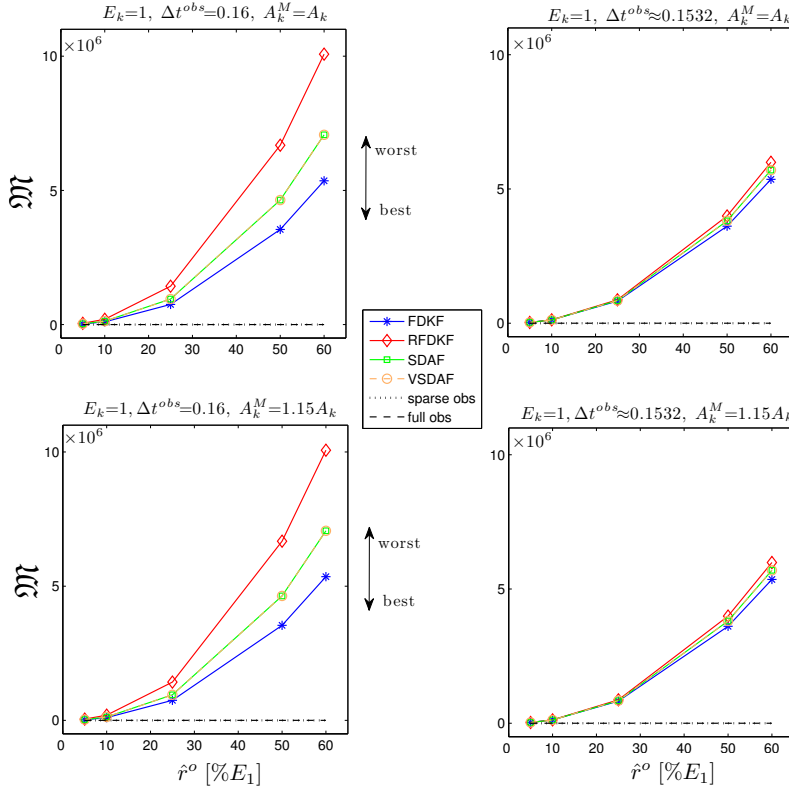


FIG. 4.14. The filter skill according to the information criterion  $\mathfrak{M}$  in (2.50) for viscosity-modified filtering of the aliasing modes in  $\mathcal{A}(17)$  in the turbulent advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k = 1$ , non-zero mean, resonant forcing (3.37) and uniform damping ( $d = 0.01, \mu = 0, c = 1$ ) through sparse observations ( $P = 3, M = 20, N = 61$ ); the reduced filters used for state estimation are described in Section 4.1. The cases of observable dynamics (left column) and non-observable dynamics (right column) are shown for both correct forcing (top row) and incorrect forcing (bottom row). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.13 shows the individual information measures combined in  $\mathfrak{M}$ .

affects the estimates of the mean dynamics, as can be seen in the relative entropy measure in figure 4.13. Similar results hold for the smoother spectrum  $E_k \propto k^{-5/3}$  which are not shown for the sake of brevity. When the mean signal is either decaying (no deterministic forcing in (3.32)) or arises due to a non-resonant periodic forcing, the modified FDKF is comparable to the unmodified FDKF for equipartition energy spectrum and has a slight but insignificant advantage over the unmodified FDKF for  $E_k \propto k^{-5/3}$ . These superensemble results based on the information theoretic measures further reinforce the path-wise estimates of [37, 60].

In order to summarize the results illustrated in figures 4.11-4.14 for filtering turbulent signals with non-zero mean dynamics, we make the following points regarding the skill of the modified reduced filtering algorithms:

- When filtering the turbulent advection-diffusion dynamics (3.32) with uni-

form damping ( $\mu=0$ ,  $d\neq 0$ ) and resonant forcing,  $\mathcal{F}_m$ , so that  $\omega_0(k)=ck$  in (3.37), the performance of FDKF depends largely on the observability (4.23) of the truth dynamics; the following facts are important:

- For observable assimilation times  $\Delta t^{obs}$ , such that (4.23) is satisfied, the performance of FDKF and the reduced filters is essentially the same as for filtering the dynamics with decaying mean discussed in Section 4.3.1.
- For non-observable assimilation times  $\Delta t^{obs}$  such that (4.23) is not satisfied, FDKF has no skill in the presence of forcing errors in the forecast model (right column in figure 4.11); the lack of skill is due to large biases in the estimates for the mean dynamics of the signal (figures 4.11, 4.12).
- High skill in terms of the mutual information and filter error entropy is possible despite very large relative entropy due to large biases in the estimated mean dynamics (see figures 4.11, 4.13).
- When filtering the turbulent advection-diffusion dynamics (3.32) with non-resonant forcing,  $\mathcal{F}_m$ , so that  $\omega_0(k)\neq ck$  in (3.37), the performance of FDKF and the reduced filters is similar to that discussed in Section 4.3.1.
- When filtering the turbulent dynamics in (3.32) with uniform damping ( $\mu=0$ ,  $d\neq 0$ ), the introduction of model error via the addition of weak selective damping,  $0<\tilde{\mu}\ll 1$ , to the forecast dynamics reinstates the observability of filtering with FDKF but the utility of this approach depends on the signal-to-noise ratio of the true mean signal; the following facts are important here:
  - For small signal-to-noise ratio (e.g., dynamics with decaying mean or with non-resonant periodic forcing) the use of modified FDKF with additional weak selective damping leads to some improvements over the reduced filters in estimating the dynamics although the improvement is not significant.
  - For large signal-to-noise ratio (e.g., dynamics with resonant periodic forcing) and uniform damping in the truth dynamics the following hold:
    - \* Unmodified SDAF has the best skill, since it satisfies the observability condition on the primary/observed modes and does not introduce model errors into the dynamics of the primary modes (figures 4.12, 4.14).
    - \* The unmodified FDKF algorithm fails in this configuration for non-observable assimilation times  $\Delta t^{obs}$  in the presence of errors in the forcing in the forecast model (bottom right in figure 4.11).
    - \* The modified FDKF filter with the additional selective damping term  $\tilde{\gamma}_k$  fails even when the observability condition (4.23) is reinstated (figures 4.13, 4.14); this is because the model errors introduced through the additional selective damping are too large in the presence of large amplitude coherent events in the truth dynamics which occur due to the resonant forcing.
    - \* Unmodified RFDKF has acceptable skill only in terms of the error entropy/RMS error (figure 4.11). The skill of modified RFDKF is substantially improved in terms of the mutual information/pattern correlation and error entropy/ RMS error; however, the additional

selective damping introduces large biases in the mean, similar to the other filters.

- The superensemble estimates of performance of FDKF and the reduced filters based on the information theoretic measures further reinforce the path-wise estimates of [37, 60].

**4.4. Measures of filter error on aliased modes for sparsely observed systems.** Throughout the preceding sections the performance of the imperfect, reduced-order filters (cf. Section 4.1) for recovering the statistics of the truth signal was assessed through the relative entropy,  $\mathcal{P}(\pi, \pi^f)$  in (2.18), which measures the lack of information in the statistics of filter estimates,  $\pi^f$ , relative to the truth statistics  $\pi$ . When filtering the sparsely observed turbulent dynamics in Section 4, this approach to quantifying the statistical filter skill required restricting the relative entropy measure to the marginal densities on the modes resolved by the sparse  $(2M+1)$ -node observation grid, i.e., we considered  $\mathcal{P}(\pi(\mathbb{P}^\ell \mathbf{u}_{\{\ell\}m}), \pi^f(\mathbb{P}^\ell \bar{\mathbf{u}}_{\{\ell\}m|m}))$  where  $\mathbb{P}^\ell$  denotes a projection onto the primary/resolved modes,  $|\ell| \leq M \leq N$ , in each aliasing set  $\mathcal{A}(\ell)$  in (4.2) of the truth signal with  $N$  Fourier modes; this projection was necessary (see Section 4.2.2) due to the presence of correlations in the reduced filter estimates for modes within the same aliasing set which lead to singular densities  $\pi^f(\bar{\mathbf{u}}_{\{\ell\}m|m})$  and singularities in  $\mathcal{P}(\pi, \pi^f)$ . While the correlations in the reduced filter estimates for the unresolved modes are to be expected due to the order-reduction approximations for the filtered dynamics, it is nevertheless desirable to investigate the reduced filter skill for recovering the statistics of all the unresolved modes; such assessments are important in developing reduced-order techniques for stochastic superresolution of sparsely observed turbulent systems [12]. An alternative strategy that allows for assessing the filter skill for recovering the statistics of the aliasing modes is to use skill measures which remain well-defined in such cases, e.g., the Hellinger distance (2.40) or the RMS error (2.30). However, the drawback of using these measures, in contrast to the relative entropy (2.18), is that neither the Hellinger distance (2.40) nor the RMS error have a suitable information-theoretic interpretation.

Here, we are primarily concerned with the suitability and ‘equivalence’ of various measures of the statistical fidelity of filter estimates and not with the details of performance of the particular reduced filters. To this end, we compare the information measure proposed in (2.50) with various other measures of statistical filter skill. First, we consider the following generalization of the information criterion in (2.50):

$$\mathfrak{M}_{\mathbb{P}^k \mathcal{P}} := \frac{\exp(\mathcal{S}(\mathbf{u}_{\{\ell\}\infty} - \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})) + \mathcal{P}(\pi(\mathbb{P}^k \mathbf{u}_{\{\ell\}\infty}), \pi^f(\mathbb{P}^k \bar{\mathbf{u}}_{\{\ell\}\infty|\infty}))}{\mathcal{M}(\mathbf{u}_{\{\ell\}\infty}, \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})}, \quad (4.36)$$

where the three information measures,  $\mathcal{S}$ ,  $\mathcal{P}$ ,  $\mathcal{M}$ , of different aspects of the statistical filter skill are defined, respectively, in (4.11), (4.12), and (4.10); the operator  $\mathbb{P}^k$  is a projection onto the  $k$ -th mode in the aliasing set  $\mathcal{A}(\ell) = \{\ell \equiv k_1, k_2, \dots, k_M\}$  in (4.2). In particular, the criterion in (4.36) reduces to that defined in (2.50) when  $\mathbb{P}^k \equiv \mathbb{P}^\ell$  is a projection onto the primary mode; we will denote this criterion as  $\mathfrak{M}_{\mathbb{P}^\ell \mathcal{P}}$  to avoid confusion with the other criteria. For  $\mathbb{P}^k \equiv \mathbb{I}$  in (4.36) the relative entropy is based on the joint densities for all the aliasing modes in each aliasing set; we will denote this criterion as  $\mathfrak{M}_{\mathcal{P}}$ . Note that, in contrast to the information criterion  $\mathfrak{M}_{\mathbb{P}^\ell \mathcal{P}}$ , the relative entropy term in  $\mathfrak{M}_{\mathcal{P}}$  is singular when the filter estimates for the aliasing modes are correlated (see, e.g., (2.23) for  $\det \mathcal{C}_{(22)} = 0$ ), as in the case of reduced-order filtering with sparse observations (Section 4.2.2).

The measure  $\mathfrak{M}_{d_H}$  incorporating the Hellinger distance instead of the relative entropy is defined in a similar fashion to the information criterion in (2.50), and it is given by

$$\mathfrak{M}_{d_H} := \frac{\exp(\mathcal{S}(\mathbf{u}_{\{\ell\}\infty} - \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})) + d_H(\pi(\mathbf{u}_{\{\ell\}\infty}), \pi^f(\bar{\mathbf{u}}_{\{\ell\}\infty|\infty}))}{\mathcal{M}(\mathbf{u}_{\{\ell\}\infty}, \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})}, \quad (4.37)$$

where the Hellinger distance  $d_H(\pi, \pi^f)$  in (2.40) measures the discrepancy between the statistics of the true signal density,  $\pi(\mathbf{u}_m)$ , and the statistics of its filter estimates,  $\pi^f(\bar{\mathbf{u}}_{m|m})$ . Similar to the relative entropy  $\mathcal{P}(\pi, \pi^f)$  in (2.18), the Hellinger distance  $d_H(\pi, \pi^f)$  does not account for correlations between the truth and the filter estimates; however, unlike the relative entropy,  $d_H$  remains well-defined for singular densities of the filter estimates  $\pi^f$ .

Another potential choice of skill measure in the filter estimates of the truth signal statistics is to use the asymptotic RMS error (2.30) on the attractor of the Kalman filter system (4.4) together with the relationship established in (2.34); the corresponding skill measure can be defined as

$$\mathfrak{M}_{\text{RMSE}} := \frac{\exp(\mathcal{S}(\mathbf{u}_{\{\ell\}\infty} - \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})) + \sqrt{\text{tr}[\text{RMSE}^2(\mathbf{u}_{\{\ell\}\infty} - \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})]}}{\mathcal{M}(\mathbf{u}_{\{\ell\}\infty}, \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})}, \quad (4.38)$$

where the ensemble RMS error matrix,  $\text{RMSE}^2(\mathbf{u}_{\{\ell\}\infty} - \bar{\mathbf{u}}_{\{\ell\}\infty|\infty})$  in (2.30), is computed through the statistical ensemble averages on the attractor due to the relation in (2.36). Note that the RMSE term in (4.38) contains cross-correlations between the truth modes and their filter estimates within the same aliasing set; thus, the drawback of the criterion  $\mathfrak{M}_{\text{RMSE}}$  in (4.38) is that it does not contain a term which accounts solely for biases in the statistics of the estimated signal.

In the numerical examples shown in figures 4.15-4.12 we use the same tough dynamical regime as the one considered in figures 4.11-4.12 of Section 4.3.2; i.e., we consider filtering the sparsely observed, stochastically forced advection-diffusion dynamics (3.32) with equipartition energy spectrum  $E_k = 1$ , uniform damping ( $\mu = 0, d = 0.01$ ) and non-trivial mean dynamics induced by a time-periodic, resonant forcing (3.37); here, the filter error is induced by the order reduction in the filter forecast models (see 4.1) and the incorrect forcing in the filters when  $A^M \neq A$  in (3.37). Similar to Section 4.2.2, considering this tough dynamical regime helps reveal differences between the different measures of filter skill for recovering the statistics of the unresolved modes. However, it is important to stress again that the use of the reduced filters from Section 4.1 for filtering the dynamics in (3.32) with equipartition energy spectrum is technically unjustified since the conditions (4.13) allowing for derivation of the reduced filters are violated. Consequently, the discussion below aims at elucidating differences and similarities between the different measures of filter skill rather than the performance of the reduced filters.

In figure 4.15 we show the superensemble error in the filter estimates for the individual aliasing modes in the aliasing set  $\mathcal{A}(17)$  based on the relative entropy (2.23), Hellinger distance (2.40), and RMS error (2.30); the example shown corresponds to filtering in the tough regime with equipartition energy spectrum and uniform damping but with correct resonant forcing and observable assimilation time  $\Delta t^{obs}$  satisfying (4.23). Figure 4.16 shows a comparison between the different measures of the filter skill for recovering the statistics of the truth signal as a function of the observation

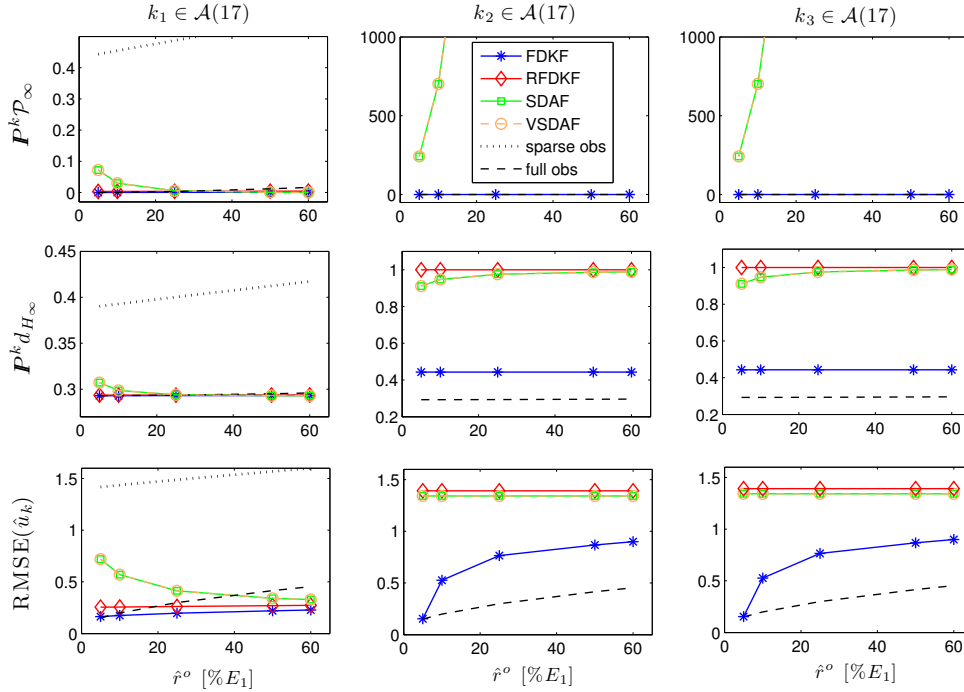


FIG. 4.15. Error in the estimated statistics of individual aliasing modes in set  $\mathcal{A}(17)$  of sparsely observed truth signal from the advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k = 1$ , non-zero mean, resonant forcing in (3.37) and uniform damping ( $d = 0.01, \mu = 0, c = 1$ ),  $P = 3, M = 20, N = 61$ ; the reduced filters used for estimating the dynamics are described in Section 4.1. Three different error measures are shown as a function of the observation noise,  $\hat{r}^\circ$  in case when the dynamics (3.32) is observable (4.23) and with correct forcing in the filters. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figures 4.16, 4.17 show the reduced filter skill based on the estimates of all the aliasing modes.

noise variance  $\hat{r}^\circ$ . Four different cases are considered: the left column filtering with assimilation times,  $\Delta t^{obs}$ , satisfying the observability condition (4.23), while the right column corresponds to filtering with  $\Delta t^{obs}$  violating the observability condition; the case of filtering with correct forcing is shown in the top row, while the incorrect forcing case is shown in the bottom row. Figure 4.17 shows the skill measures defined in (2.50) and (4.36)-(4.38) in the regimes corresponding to those considered in figure 4.16. Figures 4.16 and 4.17 illustrate the main differences between the skill measures  $\mathfrak{M}_{\mathcal{P}^\ell \mathcal{P}}$ ,  $\mathfrak{M}_{\mathcal{P}}$ ,  $\mathfrak{M}_{d_H}$ ,  $\mathfrak{M}_{RMSE}$ , defined above.

We summarize the results illustrated in figures 4.16 and 4.17 by pointing out the following:

- Despite the lack of information-theoretic interpretation, the skill measure  $\mathfrak{M}_{d_H}$  in (4.37), incorporating the Hellinger distance (2.40), yields qualitatively similar results to those obtained using the information criterion  $\mathfrak{M}_{\mathcal{P}^\ell \mathcal{P}}$  in (2.50) except in the cases with large biases in the filter-estimated statistics (e.g., bottom right panel in figure 4.17). Analogous conclusions hold (but are not shown) for filtering in other dynamical regimes, including the dynamics with steep energy spectrum  $E_k \propto k^{-5/3}$  for which the reduced filters in Section

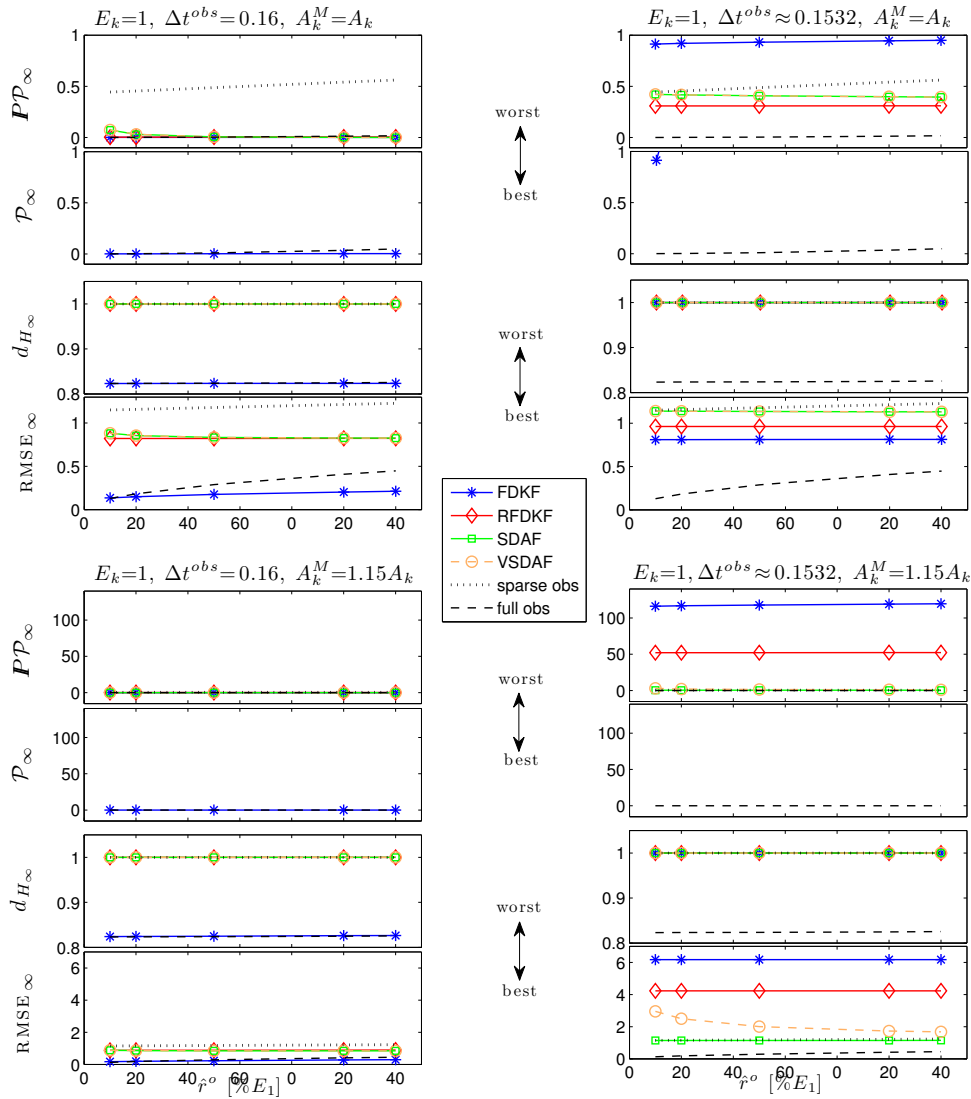


FIG. 4.16. Error in the estimated joint statistics of all aliasing modes in the aliasing set  $\mathcal{A}(17)$  of sparsely observed truth signal from the advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k=1$ , non-zero mean, resonant forcing in (3.37) and uniform damping ( $d=0.01, \mu=0, c=1$ ),  $P=3, M=20, N=61$ ; the reduced filters used for estimating the dynamics are described in Section 4.1. The three information measures discussed in Section 2.2.1 are shown as a function of the observation noise,  $\hat{r}^o$  in cases when the dynamics (3.32) is observable (left column) and non-observable (right column); see (4.23) for the observability condition. The top row shows results for correct forcing in the filters and the bottom row corresponds to incorrect forcing with error  $\Delta\mathcal{F}$  given in (4.33); the relative entropy measure is projected on the primary mode for reasons discussed in Section 4.2.2 and illustrated in figure 4.2. The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. See figure 4.12 for the information criterion  $\mathfrak{M}$  in (2.50) corresponding to this example.

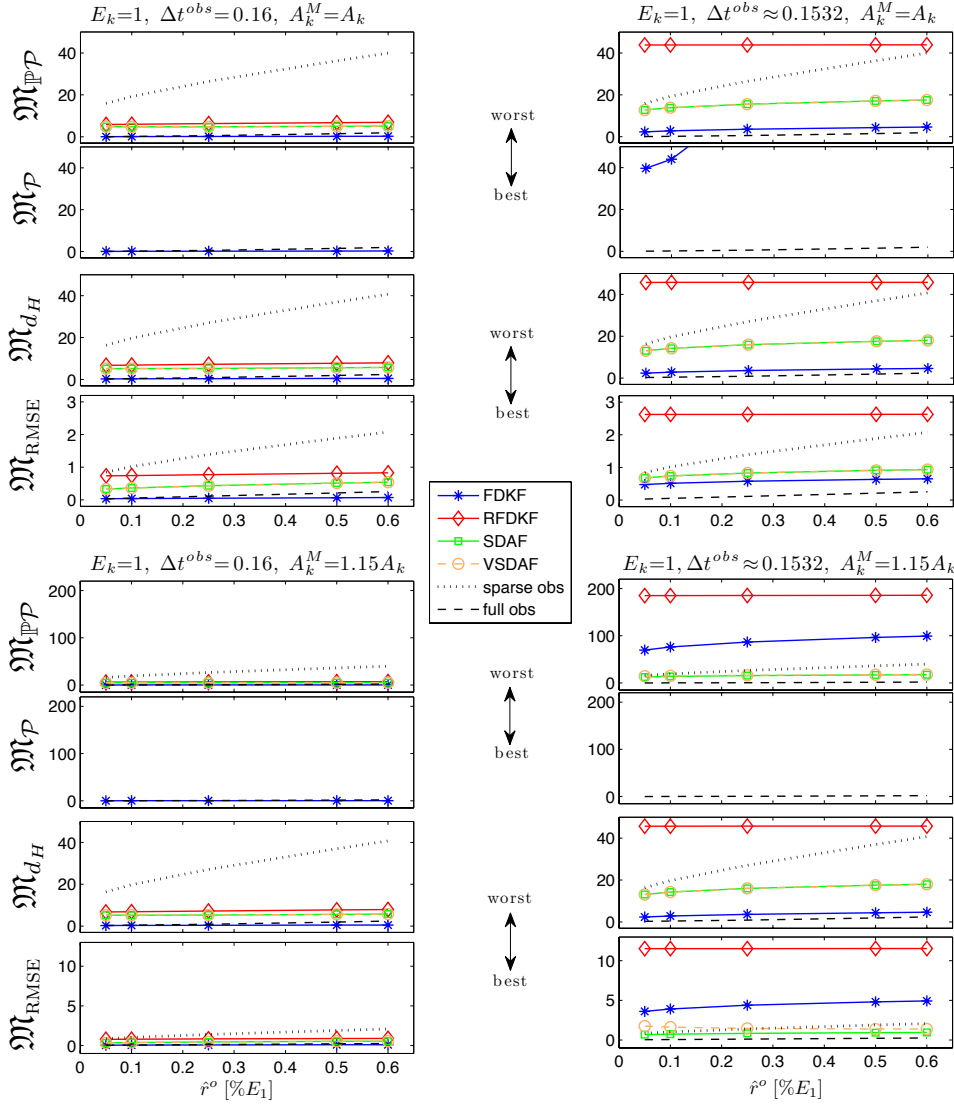


FIG. 4.17. The filter skill according to different criteria defined in (2.50) and (4.36)-(4.38) for filtering the aliasing modes in  $\mathcal{A}(17)$  in the turbulent advection-diffusion dynamics (3.32) with equipartition energy spectrum,  $E_k=1$ , non-zero mean dynamics, resonant forcing (3.37) and uniform damping ( $d=0.01, \mu=0, c=1$ ) through sparse observations ( $P=3, M=20, N=61$ ); the reduced filters used for state estimation are described in Section 4.1. The cases of observable dynamics (left column) and non-observable dynamics (right column) are shown for both correct forcing (top row) and incorrect forcing (bottom row). The black dotted lines correspond to least-squares estimates based on the raw sparse observations and the dashed black lines correspond to estimates from full observations with the same observation noise levels. Figure 4.9 shows the individual information measures combined in (2.50).

4.1 were designed.

- The information criterion  $\mathfrak{M}_{\mathcal{P}}$  in (4.36) is very sensitive to the filter-induced correlations in the estimates for the aliasing modes (see also Section 4.2.2);



thus, this criterion is not suitable for estimating the overall filter skill in sparsely observed turbulent systems. The effects of these spurious correlations are particularly pronounced in the tough regime with equipartition energy spectrum illustrated in figures 4.16 and 4.17 where FDKF is the only filter with skill according to the information criterion  $\mathfrak{M}_{\mathcal{P}}$ .

- The use of RMS error can be misleading as a measure of filter skill for recovery of the statistics of the aliasing modes (figures 4.15 and 4.16); this is because RMSE accounts not only for the statistics of the filter estimates but also for the cross-correlations between the truth and the estimates.
- Unsurprisingly, reduced-order filters which aim to recover the primary modes resolved by the observations do not contain relevant information on the aliasing modes; this is evident in terms of  $\mathfrak{M}_{\mathbb{P}^k\mathcal{P}}, k \neq \ell$  or  $\mathfrak{M}_{\mathcal{P}}$  in the tough dynamical regime with equipartition energy spectrum and in the presence of non-trivial mean (see figure 4.15 and  $\mathfrak{M}_{\mathcal{P}}$  in figure 4.17). Statistically reliable recovery of the aliasing modes has to be carried out using more sophisticated stochastic superresolution techniques [12]. However, the skill measures  $\mathfrak{M}_{\mathbb{P}^{\ell}\mathcal{P}}$  and  $\mathfrak{M}_{d_H}$  are relatively insensitive to these filtering artifacts even in the equipartition energy regime for which the tested reduced filters were not even designed (figure 4.17).

## 5. Conclusions

In this paper an information-theoretic approach was developed to quantify the statistical accuracy of Kalman filters with model error for the discretized dynamics of spatially extended, partially observed turbulent systems. The significance of information theory in filtering problems has been established and systematic information criteria for filter design have been developed. The use of methods from information theory in the context of stochastic filtering was shown to be particularly useful in the assessment of imperfect filter performance for high-dimensional problems associated with state estimation in spatially extended turbulent systems. In the present framework appropriate measures of information content in the filter estimates were used to naturally extend the common path-wise measures of filter accuracy, like the mean-square error or pattern correlation, to the superensemble setting involving all possible initial conditions of the true dynamics. In contrast to the path-wise approach, the information-theoretic framework allows for establishing lower bounds on the imperfect filter performance; these bounds are useful when assessing the information barriers in imperfect filters. The information-theoretic approach proposed here is complementary to the standard path-wise approach to assessing the filter error. The statistical superensemble framework incorporates the ensemble-averaged effects of intermittent interactions between the mean state and fluctuations which is important in accurate assessment of filter error in high-dimensional turbulent systems. On the other hand the path-wise approach is more suitable for studying the filter skill in the presence of ‘rare event’-type phenomena in the turbulent dynamics which are marginalized in the ensemble averaged skill measures. The framework presented here has natural generalizations to Kalman filtering with non-Gaussian but statistically exactly solvable forecast models such as the stochastic parametrized extended Kalman filters (SPEKF) used for filtering [31, 30, 61, 34] and Dynamic Stochastic Superresolution [12].

An important theme present throughout this paper was that of optimality of the imperfect Kalman filters in terms of entropy of the filter error, the mutual information, and the relative entropy between the truth and the filter estimate. We showed that, for a given class of imperfect filters the practically achievable filter skill requires trade-

offs between the relevant information measures of filter accuracy which led to the construction of a new information criterion. In contrast to filtering with the perfect forecast model, imperfect Kalman filters tuned to minimize the RMS error of the filter estimates do not necessarily extremize the mutual information/pattern correlation and the relative entropy between the truth and the filter estimates. The key issue here is that finding the best imperfect filter can be formulated as a constrained optimization problem with the constraints given by the desired bounds on the three information measures to account for the statistics of the filter error, including the mean biases, and the correlations between the truth and the filter estimates.

The information-theoretic approach was illustrated for a suite of imperfect reduced-order filters in the Fourier domain to filter the turbulent advection-diffusion dynamics; these filters derived and analyzed in the path-wise framework in [37, 38] with the goal to avoid the ‘curse of ensemble size’ and improve the filter stability and observability. The developed information criteria were used to analyze the filter stability and accuracy in the presence of model error arising from various spatio-temporal discretization schemes in approximations to the stochastically forced advection-diffusion equation. As already indicated in [37, 38, 60], many subtle issues associated with violation of the classical observability and controllability criteria occur in Kalman filtering with model error even in the relatively simple setting with imperfect linear Gaussian forecast models. Based on the developed information criteria we further strengthened the path-wise results of [37, 38] and showed that the reduced-order filters for turbulent systems with steep energy spectra have high skill and improved stability despite being computationally much cheaper to implement. Another important aspect concerned the quantification of filter error due to biases in the mean filter estimates arising from errors in the filter forcing; these biases are naturally captured in the present framework through the relative entropy between the truth and the filter estimates. Moreover, the present information-theoretic framework allowed for identification of potential correlations in the reduced filter estimates for sparsely observed turbulent systems. While these artificial correlations are not relevant in filtering of the primary Fourier modes associated with the scales resolved by the observation grid [37, 38, 60], their mitigation is important in the Dynamic Stochastic Superresolution [12] and superparametrization algorithms [39]; these effects are particularly difficult to capture in the traditional path-wise framework commonly used in applications.

This work only initiates the information-theoretic analysis of the performance and stability of Kalman filters; appropriate generalizations of the present framework to analyze Bayesian filtering with non-Gaussian but statistically exactly solvable forecast models such as the Stochastically Parametrized Extended Kalman filters (SPEKF) [31, 30, 61, 34] will be soon reported elsewhere.

**Acknowledgments.** M.B. is supported as a postdoctoral fellow on the ONR DRI grant of A.J.M: N0014-10-1-0554. The research of A.J.M. is partially supported by National Science Foundation CMG grant DMS-1025468 and the Office of Naval Research grants ONR DRI N0014-10-1-0554, N00014-11-1-0306, and the MURI award ONR-MURI N00014-12-1-0912.

**Appendix A. Quantifying filter error with time-correlations through lag-embedded state analysis.** Here, we provide additional details on computation of the covariances for the lag embedded state of the augmented system for the Kalman filter in (2.6). Recall that in Fact 4 of Section 2.2.1 the lag-embedded state for the space-time discretized evolution of the truth (2.1) and the filter estimate (2.4) with

$L$  finite-time lags are denoted in the Fourier domain as

$$\mathbf{u}_m^L = (\mathbf{u}_m, \mathbf{u}_{m-1}, \dots, \mathbf{u}_{m-L})^T, \quad \bar{\mathbf{u}}_{m|m}^L = (\bar{\mathbf{u}}_{m|m}, \bar{\mathbf{u}}_{m-1|m-1}, \dots, \bar{\mathbf{u}}_{m-L|m-L})^T. \quad (\text{A.1})$$

Similarly to the formulation adopted in (2.7) and (2.10), the covariance of the lag-embedded, augmented state  $\mathbf{y}^L = (\mathbf{u}_m^L, \bar{\mathbf{u}}_{m|m}^L)^T$  can be written in block form as

$$\mathcal{C}^L := \text{Cov}(\mathbf{y}^L, \mathbf{y}^L) = \begin{pmatrix} \text{Cov}(\mathbf{u}_m^L, \mathbf{u}_m^L) & \text{Cov}(\mathbf{u}_m^L, \bar{\mathbf{u}}_{m|m}^L) \\ \text{Cov}(\bar{\mathbf{u}}_{m|m}^L, \mathbf{u}_m^L) & \text{Cov}(\bar{\mathbf{u}}_{m|m}^L, \bar{\mathbf{u}}_{m|m}^L) \end{pmatrix}. \quad (\text{A.2})$$

The structure of the covariance  $\mathcal{C}^L$  is simplified by the fact that the Fourier modes in different aliasing sets  $\mathcal{A}(\ell)$  of the primary wavenumbers  $|\ell| \leq M$  are uncorrelated (see Section 4) so that  $\mathcal{C}^L$  is block diagonal when the Fourier modes are grouped into the disjoint aliasing sets  $\mathcal{A}(\ell)$ ; namely,  $\mathcal{C}^L$  can be represented as

$$\mathcal{C}^L = \mathcal{A}^{-1} \text{diag}[\mathcal{C}_{\{\ell_1\}}^L, \dots, \mathcal{C}_{\{\ell_{2M+1}\}}^L] \mathcal{A}, \quad |\ell| \leq M, \quad (\text{A.3})$$

where  $\mathcal{A}$  represents the transformation ordering the Fourier modes into the aliasing sets, and the covariances of the lag-embedded state within each disjoint aliasing set  $\mathcal{A}(\ell)$  are given by (in the block tensor product notation)

$$\mathcal{C}_{\{\ell\}}^L = \mathcal{C}_{\{\ell\}} \otimes Q(\mathcal{F}_{\{\ell\}}), \quad (\text{A.4})$$

where  $\mathcal{F}_{\{\ell\}} \in \mathbb{C}^{2P \times 2P}$  is the forward dynamics operator of the augmented Kalman system (4.4), and  $\mathcal{C}_{\{\ell\}}$  is the  $2P \times 2P$  covariance in (4.9) within each disjoint aliasing set; the  $L \cdot 2P \times L \cdot 2P$  block matrix  $Q$  is given by

$$Q(q) = \begin{pmatrix} 1 & q & q^2 & \dots & q^{L-1} \\ q^* & 1 & q & \dots & q^{L-2} \\ (q^*)^2 & q^* & 1 & \dots & q^{L-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (q^*)^{L-1} & (q^*)^{L-2} & (q^*)^{L-3} & \dots & 1 \end{pmatrix}, \quad (\text{A.5})$$

where each entry in (A.5) is a  $2P \times 2P$  block.

When computing the entropy  $\mathcal{S}$ , the relative entropy  $\mathcal{P}$  and the mutual information  $\mathcal{M}$  between the truth and the filter estimate in Section 2.2.1, the following facts are useful.

**PROPOSITION A.1.** *For the matrix  $Q[q]$  with the structure as in (A.5) the following relationships hold:*

- $\det[A] = (1 - |q|^2)^{L-1}$ ,
- For  $L=2$  the inverse matrix  $A^{-1}$  is given by

$$A^{-1} = \frac{1}{1 - |q|^2} \begin{pmatrix} 1 & -q \\ -q^* & 1 \end{pmatrix}, \quad (\text{A.6})$$

- For  $L > 2$  the inverse matrix  $A^{-1}$  a banded tri-diagonal matrix given by

$$A^{-1} = \frac{1}{1 - |q|^2} \begin{pmatrix} 1 & -q & 0 & \dots & 0 \\ -q^* & 1 + |q|^2 & -q & \dots & 0 \\ 0 & -q^* & 1 + |q|^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 + |q|^2 & -q \\ 0 & 0 & \dots & -q^* & 1 \end{pmatrix}. \tag{A.7}$$

The proofs of the above facts are straightforward and follow by direct computation based on (A.5).

**Appendix B. The augmented system for the Reduced Fourier Domain Kalman Filter (RFDKF).** Here, we present the explicit form of the augmented system for the Reduced Fourier Kalman Domain Filter (RFDKF) discussed in Section 4.1. The dynamical system combining the evolution of the truth and its RFDKF estimates is given by

$$\begin{bmatrix} \mathbf{u}_{\{\ell\}m+1} \\ \bar{\mathbf{u}}_{\{\ell\}m+1|m+1} \end{bmatrix} = \mathfrak{A}_{\{\ell\}m} \begin{bmatrix} \mathbf{u}_{\{\ell\}m} \\ \bar{\mathbf{u}}_{\{\ell\}m|m} \end{bmatrix} + \mathfrak{F}_{\{\ell\}m} + \mathfrak{S}_{\{\ell\}m}, \tag{B.1}$$

where

$$\mathfrak{A} = \begin{bmatrix} F_{\{\ell\}} & 0 \\ \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \mathbf{G}_P F_{\{\ell\}} & ((\mathcal{I} - \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \mathbf{G}_P) - \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \tilde{\mathbf{G}}_P) F_{\{\ell\}}^M \end{bmatrix}, \tag{B.2}$$

$$\mathfrak{F} = \begin{bmatrix} \mathcal{F}_{\{\ell\}m+1} \\ ((\mathcal{I} - \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \mathbf{G}_P) - \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \tilde{\mathbf{G}}_P) \mathcal{F}_{\{\ell\}m+1}^M + \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \mathbf{G}_P \mathcal{F}_{\{\ell\}m+1} \end{bmatrix}, \tag{B.3}$$

$$\mathfrak{S} = \begin{bmatrix} \sigma_{\{\ell\}m+1} \\ \tilde{\mathbf{K}}_{\{\ell\}m+1}^M (\mathbf{G}_P \sigma_{\{\ell\}m+1} + \sigma_{\{\ell\}m+1}^o) - \tilde{\mathbf{K}}_{\{\ell\}m+1}^M \tilde{\mathbf{G}}_P \sigma_{\{\ell\}m+1}^M \end{bmatrix}, \tag{B.4}$$

and with the modified Kalman gain given by

$$\tilde{\mathbf{K}}_{\{\ell\}m+1}^M = (K_{k_1, m+1}^M, 0, \dots, 0)^T, \quad K_{k_1, m+1}^M \in \mathbb{C}, \tag{B.5}$$

and  $\tilde{\mathbf{G}}_P$  is a  $P \times P$  matrix given by

$$\tilde{\mathbf{G}}_P = \begin{pmatrix} 0 & G & \dots & G \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}. \tag{B.6}$$

Due to the fact that the Fourier modes in different aliasing sets  $\mathcal{A}(\ell)$  are uncorrelated, the full covariance matrix is block-diagonal with  $P \times P$  blocks evolving independently according to

$$\mathcal{C}_{\{\ell\}m+1} = \mathfrak{F}_{\{\ell\}m} \mathcal{C}_{\{\ell\}m} \mathfrak{F}_{\{\ell\}m}^* + \mathcal{R}_{\{\ell\}m}, \tag{B.7}$$

where

$$\mathfrak{F}_{\{\ell\}m} = \begin{bmatrix} F_{\{\ell\}} & 0 \\ \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}} \mathbf{G}_P F_{\{\ell\}} & \left( (\mathcal{I} - \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}} \mathbf{G}_P) - \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}} \tilde{\mathbf{G}}_P \right) F_{\{\ell\}}^{\mathbf{M}} \end{bmatrix}, \quad (\text{B.8})$$

$$\mathcal{R}_{\{\ell\}m} = \begin{bmatrix} r_{\{\ell\}} & r_{\{\ell\}} \mathbf{G}_P^* \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}*} \\ \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}} \mathbf{G}_P r_{\{\ell\}} & \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}} (r_{\{\ell\}}^{\circ} + \mathbf{G}_P r_{\{\ell\}} \mathbf{G}_P^*) \tilde{\mathbf{K}}_{\{\ell\}m+1}^{\mathbf{M}*} \end{bmatrix}. \quad (\text{B.9})$$

#### REFERENCES

- [1] R.V. Abramov and A.J. Majda, *Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems*, *Nonlin.*, 20(12), 2793–2821, 2007.
- [2] B. Anderson and J. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [3] J.L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, *Mon. Wea. Rev.*, 129, 2884–2903, 2001.
- [4] J.L. Anderson, *A local least squares framework for ensemble filtering*, *Mon. Wea. Rev.*, 131(4), 634–642, 2003.
- [5] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*, *Stochastic Modelling and Applied Probability*, Springer, 60, 2009.
- [6] T. Bengtsson, P. Bickel, and B. Li, *Essays in Honor of David A. Freedman*, *IMS Lecture Notes - Monograph Series in Probability and Statistics*, Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems, *Inst. Math. Sci.*, 2, 316–334, 2008.
- [7] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, NJ: Prentice Hall, 1971.
- [8] L.M. Berliner, R.F. Milliff, and C.K. Wikle, *Bayesian hierarchical modeling of air-sea interaction*, *J. Geophys. Res.*, 108(C4), 3104, 2003.
- [9] P. Bickel, B. Li, and T. Bengtsson, *Essays in Honor of J. K. Gosh*, *IMS Lecture Notes - Monograph Series*, ch. Sharp Failure Rates for the Bootstrap Filter in High Dimensions, *Inst. Math. Sci.*, 318–329, 2008.
- [10] M. Branicki, N. Chen, and A.J. Majda, *Non-Gaussian Test Models for Prediction and State Estimation with Model Errors*, *Chinese Ann. Math.*, 34B(1), 29–64, 2013.
- [11] M. Branicki, B. Gershgorin, and A.J. Majda, *Filtering skill for turbulent signals for a suite of nonlinear and linear Kalman filters*, *J. Comput. Phys.*, 231, 1462–1498, 2012.
- [12] M. Branicki and A.J. Majda, *Dynamic stochastic superresolution of sparsely observed dynamical systems*, *J. Comput. Phys.*, 241, 333–363, 2013.
- [13] M. Branicki and A.J. Majda, *Quantifying uncertainty for statistical predictions with model errors in non-Gaussian models with intermittency*, *Nonlin.*, 25, 2543–2578, 2012.
- [14] R.S. Bucy and P.D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*, New York: Wiley, 1968.
- [15] E. Castronovo, J. Harlim, and A. Majda, *Mathematical criteria for filtering complex systems: Plentiful observations*, *J. Comput. Phys.*, 227, 3678–3714, 2008.
- [16] A.J. Chorin and P. Krause, *Dimensional reduction for a bayesian filter*, *Proc. Natl. Acad. Sci. USA*, 101, 15013–15017, 2004.
- [17] C. Chui and G. Chen, *Kalman Filtering with Real-time Applications*, Springer, New York, 1999.
- [18] S. Cohn and D. Dee, *Observability of discretized partial differential equations*, *SIAM J. Numer. Anal.*, 25(3), 586–617, 1988.
- [19] S.L. Cotter, M. Dashti, and A.M. Stuart, *Approximation of Bayesian inverse problems*, *SIAM J. Numer. Anal.*, 48, 322–345, 2010.
- [20] T.A. Cover and J.A. Thomas, *Elements of Information Theory*, Second Edition, Wiley-Interscience, Hoboken, 2006.
- [21] D.P. Dee and A.M. Da Silva, *Data assimilation in the presence of forecast bias*, *Quart. J. Roy. Meteor. Soc.*, 124, 269–295, 1998.
- [22] D.P. Dee and R. Todling, *Data assimilation in the presence of forecast bias: The GEOS moisture analysis*, *Mon. Wea. Rev.*, 128(9), 3268–3282, 2000.
- [23] T. DelSole, *Stochastic Models of Quasigeostrophic Turbulence*, *Surveys in Geophys.*, 25(2), 107–149, 2004.
- [24] T.E. Duncan, *On the calculation of mutual information*, *SIAM J. Appl. Math.*, 19(1), 215–220, 1970.

- [25] B. Farrell and P. Ioannou, *Distributed forcing of forecast and assimilation error systems*, J. Atmos. Sci., 62, 460–475, 2005.
- [26] B. Friedland, *Treatment of bias in recursive filtering*, IEEE Trans. Automatic Control, AC-14, 359–367, 1969.
- [27] B. Friedland, *Estimating sudden changes of biases in linear dynamical systems*, IEEE Trans. Automatic Control, AC-27, 237–240, 1982.
- [28] J.I. Galdos and D.E. Gustafson, *Information and distortion in reduced-order filter design*, IEEE Trans. Informat. Theory, 23(2), 183–194, 1977.
- [29] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [30] B. Gershgorin, J. Harlim, and A.J. Majda, *Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation*, J. Comput. Phys., 229, 32–57, 2010.
- [31] B. Gershgorin, J. Harlim, and A.J. Majda, *Test models for improving filtering with model errors through stochastic parameter estimation*, J. Comput. Phys., 229, 1–31, 2010.
- [32] B. Gershgorin and A.J. Majda, *A nonlinear test model for filtering slow-fast systems*, Comm. Math. Sci., 6, 611–649, 2008.
- [33] B. Gershgorin and A.J. Majda, *Filtering a nonlinear slow-fast system with strong fast forcing*, Comm. Math. Sci., 8, 67–92, 2009.
- [34] B. Gershgorin and A.J. Majda, *Filtering a statistically exactly solvable test model for turbulent tracers from partial observations*, J. Comput. Phys., 230, 1602–1638, 2011.
- [35] A.L. Gibbs and F.E. Su, *On choosing and bounding probability metrics*, Intern. Stat. Rev., 70(3), 419–435, 2002.
- [36] J. Harlim and A.J. Majda, *Test models for filtering and prediction of moisture-coupled tropical waves*, Q.J.R. Meteorol. Soc., 139, 119–136, 2013.
- [37] J. Harlim and A.J. Majda, *Mathematical strategies for filtering complex systems: Mathematical strategies for filtering complex systems: Regularly spaced sparse observations*, J. Comput. Phys., 227, 5304–5341, 2008.
- [38] J. Harlim and A.J. Majda, *Filtering turbulent sparsely observed geophysical flows*, Mon. Wea. Rev., 138(4), 1050–1083, 2010.
- [39] J. Harlim and A.J. Majda, *Test models for filtering with superparameterization*, Multiscale Model. Simul., 11(1), 282–308, 2013.
- [40] K. Haven, A.J. Majda, and R. Abramov, *Quantifying predictability through information theory: Small sample estimation in a non-gaussian framework*, J. Comput. Phys., 334–362, 2005.
- [41] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- [42] T. Kailath, *A general likelihood ratio formula for random signals in gaussian noise*, IEEE Trans. Informat. Theory, 15, 350–361, 1969.
- [43] R.E. Kalman, *A new approach to linear filtering and prediction problems*, J. Basic Eng. (Series D), 82, 35–45, 1960.
- [44] R.E. Kalman and R.S. Bucy, *New results in linear filtering and prediction problems*, J. Basic Eng. (Series D), 83, 95–107, 1961.
- [45] S.R. Keating, A.J. Majda, and K.S. Smith, *New methods for estimating poleward eddy heat transport using satellite altimetry*, Mon. Wea. Rev., 140, 1703–1722, 2012.
- [46] R. Kleeman, *Measuring dynamical prediction utility using relative entropy*, J. Atmos. Sci., 59, 13, 2057–2072, 2002.
- [47] R. Kleeman, A.J. Majda, and I. Timofeyev, *Quantifying predictability in a model with statistical features of the atmosphere*, Proc. Natl. Acad. Sci. USA, 99, 15291–15296, 2002.
- [48] S. Kravtsov, D. Kondrashov, and M. Ghil, *Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability*, J. Climate, 18, 21, 4404–4424, 2005.
- [49] R.S. Liptser and A.N. Shiryaev, *Statistics of Random Process*, Second, Springer-Verlag, New York, 2001.
- [50] A.J. Majda, *Challenges in climate science and contemporary applied mathematics*, Comm. Pure Appl. Math., 65(7), 920–948, 2012.
- [51] A.J. Majda, R. Abramov, and B. Gershgorin, *High skill in low frequency climate response through fluctuation dissipation theorems despite structural instability*, Proc. Natl. Acad. Sci. USA, 107(2), 581–586, 2010.
- [52] A.J. Majda, R.V. Abramov, and M.J. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems*, CRM Monograph Series, Americal Mathematical Society, Providence, 25, 2005.
- [53] A.J. Majda and B. Gershgorin, *Quantifying uncertainty in climate change science through empirical information theory*, Proc. Natl. Acad. Sci., 107(34), 14958–14963, 2010.
- [54] A. J. Majda and J. Harlim, *Physics constrained nonlinear regression models for time series*,

- Nonlin., 26(1), 201–217, 2013.
- [55] A.J. Majda, R. Kleeman, and D. Cai, *A mathematical framework for predictability through relative entropy*, Meth. Appl. Anal., 9(3), 425–444, 2002.
- [56] A.J. Majda and X. Wang, *Linear response theory for statistical ensembles in complex systems with time-periodic forcing*, Comm. Math. Sci., 8(1), 145–172, 2010.
- [57] A.J. Majda and Y. Yuan, *Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems*, Disc. Cont. Dyn. Sys., 4, 1333–1363, 2012.
- [58] A.J. Majda and M. Branicki, *Lessons in uncertainty quantification for turbulent dynamical systems*, Disc. Cont. Dyn. Sys., 32(9), 3133–3231, 2012.
- [59] A.J. Majda and M.J. Grote, *Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems*, Proc. Natl. Acad. Sci. USA, 104, 1124–1129, 2007.
- [60] A.J. Majda and J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, 2012.
- [61] A.J. Majda, J. Harlim, and B. Gershgorin, *Mathematical strategies for filtering turbulent dynamical systems*, Disc. Cont. Dyn. Sys., 27, 441–486, 2010.
- [62] R. Miller, E. Carter, and S. Blue, *Data assimilation into nonlinear stochastic models*, Tellus A, 51, 167–194, 1999.
- [63] S.K. Mitter and N.J. Newton, *Information and entropy flow in the Kalman–Bucy filter*, J. Stat. Phys., 18, 145–176, 2005.
- [64] N. Newton, *Dual nonlinear filters and entropy production*, SIAM J. Control Optim., 46(5), 1637–1663, 2007.
- [65] N. Newton, *Interactive statistical mechanics and nonlinear filtering*, J. Stat. Phys., 133, 711–737, 2008.
- [66] E. Ott, B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, and J.A. Yorke, *Estimating the state of large spatio-temporally chaotic systems*, Phys. Lett. A, 330, 365–370, 2004.
- [67] R.D. Reiss, *Approximate Distributions of Order Statistics*, New York, Springer-Verlag, 1989.
- [68] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, U. Illinois Press, 1949.
- [69] Ch. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, *Obstacles to high-dimensional particle filtering*, Mon. Wea. Rev., 136, 4629–4640, 2008.
- [70] Y. Tomita, S. Omatu, and T. Soeda, *An application of the information theory to filtering problem*, Inf. Sci., 11, 13–27, 1976.
- [71] H.L. Widemann and E.B. Stear, *Entropy analysis of estimating systems*, IEEE Trans. Informat. Theory, 16, 264–270, 1970.
- [72] M. Zakai and J. Ziv, *Lower and upper bounds on the optimal filtering error of certain diffusion processes*, IEEE Trans. Informat. Theory, 18(3), 325–331, 1972.