ACCURACY OF SOME APPROXIMATE GAUSSIAN FILTERS FOR THE NAVIER-STOKES EQUATION IN THE PRESENCE OF MODEL ERROR

M. BRANICKI[1*], A. J. MAJDA[2], K.J.H. LAW[3]

[1] *Department of Mathematics, University of Edinburgh, Edinburgh, UK*

[2] *Department of Mathematics, Courant Institute, NYU, New York, USA*

[3] *Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

Abstract. The problem of Bayesian state estimation of a dynamical system given a stream of noisy and incomplete measurements is important in many applications involving real-time prediction and control. Time-sequential data assimilation leads naturally to a Bayesian formulation in which the posterior probability distribution of the system state, given all past observations, is updated from the prior model prediction and the information from next available measurement. High dimensionality of the state space, partial observations, and model error pose key challenges for data assimilation in geosciences and engineering applications. Here, two classes of computationally feasible, approximate Gaussian data assimilation algorithms are compared in the presence of model error and sparse observations. Two main sources of model error in filtering are considered. The first arises from the necessary use of reduced models with coarse resolution for the forward operator, while the second source of error arises in the observation operator which - even for linear observation process - mixes the resolved and unresolved dynamics through aliasing of the truth signal. Non-aliased observations ignore the second source of model error by equating the true observations with those of the reduced dynamics alone. The two-dimensional, incompressible Navier-Stokes equation with a linear drag in a periodic geometry is utilised in this study to generate the "truth" dynamics as it has the necessary features of complex dissipative systems encountered in practice, and it can be tuned to generate many distinct dynamical regimes with sparse observations aliasing fine-scale information into the assimilation space. The 3DVAR filter is prototypical of sequential methods used to combine incoming observations with a dynamical system in order to improve the state estimation. Given optimally inflated covariance, 3DVAR is known to be (provably) accurate for filtering dissipative systems in the absence of model error in the forward dynamics. In contrast, SPEKF (Stochastically Parmeterised Extended Kalman Filter) algorithms do not require covariance inflation or a detailed knowledge of the underlying dynamics, and they have been shown (empirically) to be effective in mitigating model error in state estimation of turbulent dynamical systems. The primary conclusions are that, under the assumption of a well-defined posterior probability distribution: (i) with appropriate tuning and for non-aliased observations the considered approximate Gaussian filters perform well in reproducing the mean of the desired filtering probability distribution; (ii) given sparse aliased observations of turbulent dynamics (complex and high-dimensional in spectral domain), the SPEKF algorithms perform significantly better than a tuned 3DVAR.

1. Introduction

The problem of state estimation of a dynamical process given its noisy and incomplete measurements arriving in a time-sequential manner is of importance in a wide range of applications. Examples include atmosphere-ocean science (e.g., [39]), and engineering problems (e.g., [61]) where online predictions are required in the presence of uncertainty in the initial conditions, the observed data, or in the dynamics

E-mail address: M.Branicki@ed.ac.uk.

itself. A natural framework for approaching such problems is that of Bayesian statistics which allows for a systematic combination of the incoming data with the dynamical model in order to solve a sequence of inverse problems on the current state of the dynamical process. This topic has enjoyed a long-standing symbiosis between the stochastic analysis through the control-theoretic and probabilistic approaches, and a physical reasoning aimed at approximations of the underlying, usually complex, structure of the true dynamics. When an exact dynamics is known subject to uncertain initial conditions, the posterior distribution on the system state given the incoming data can be derived, in principle, in a way that is optimal with respect to the error covariance [4, 36]. The corresponding framework is referred to as the stochastic (Bayesian) filtering, and provably accurate discretisations of the resulting probabilistic estimates provide a gold-standard solution to the filtering problem if they can be efficiently computed. This may be performed exactly for linear systems subject to Gaussian noise, leading to the Kalman filter (e.g., [38, 1, 36, 31]). In nonlinear and non-Gaussian scenarios the particle filter (e.g., [4, 18]) provably approximates the desired probability distribution as the number of particles increases; nevertheless, standard implementations of this method perform poorly in high-dimensional systems [63]. Recently, clustered particle filters were developed [47] to capture capture significant non-Gaussian features with few particles in a suite of stringent numerical tests for filtering and prediction of turbulent systems but a rigorous theory remains to be developed. The field of data assimilation has grown out of the necessity to obtain computationally feasible approximations to the filtering problem when one is faced with a high-dimensional state estimation and/or imperfect knowledge of the underlying dynamics. Indeed, in many real-world applications (e.g., weather prediction), the dimensionality of the underlying dynamics and the vast amount of incoming data makes the analysis of the Bayesian posterior distribution of the state given data formidably complicated and computationally infeasible in online implementations. In such situations practitioners are typically forced to employ approximations based on either physical insight and/or computational expediency, while the incoming data are used to compensate for both the uncertainty in the model and in the initial conditions. Consequently, many operational data assimilation algorithms are ad-hoc and the theoretical understanding of their accuracy and robustness for state estimation in the presence of model error remains limited. Our goal here is to contribute towards understanding the effects of model error and sparse aliased observations on the performance of filtering/data assimilation algorithms.

The development of practical and robust filtering/data assimilation algorithms for high-dimensional dynamical systems is an active research area and we refer the reader to, e.g., [13, 19, 55, 57, 70, 71] for further insight into this subject. The study of accuracy and stability of data assimilation algorithms has been a developing area over the last few years. The term accuracy refers to establishing closeness of the filter to the true signal underlying the data, and stability is concerned with studying the distance between estimates two sequences of filter estimates, initialised differently, but driven by the same noisy data. Theoretical analysis of filter accuracy and stability for control systems has a long history and the paper [64] is a fundamental contribution to the subject, while the paper [67] highlights the importance of observing and assimilating data that constrains the unstable directions in the underlying dynamics for accurate estimates. The paper [12] contains finite-dimensional theory and numerical experiments in a variety of finite and discretised infinite-dimensional systems. Finally, a recent sequence of papers provides a significant generalisation of the theory in [64] to dissipative infinite-dimensional dynamical systems which are prototypical of the high-dimensional problems to which filters are applied in practice [10, 7, 42, 3, 6, 32]. Many of the methods used for constructing data assimilation algorithms invoke some form of ad hoc Gaussian approximation which generally destroys the optimality of the estimates, which might cause filter divergence [50, 30, 41]. In some cases, rigorous theory and algorithms are available to eliminate such

a severe pathology [41]. Moreover, while carefully tuned approximate filters can be stable and perform well in predicting mean behaviour, they typically perform poorly when predicting uncertainty, such as covariance information (e.g., [46, 27, 58, 48, 72, 73]). Importantly, the tuning procedure usually relies on retrospective, hind-cast adjustments which require large amounts of historical data and makes such methods not 'rapidly-deployable'. These issues are compounded further by the presence of an unavoidable model error in the forward dynamics of the filtering algorithms, and the sparsity of observations has adverse effects on the *assimilation in the unstable subspace* which was identified as an important condition for accuracy and stability of data assimilation algorithms [12, 67] (see, however, [55, Chapter 3] for a discussion of limitations of this approach in a finite-dimensional context with non-normal operators.) Here, we focus on the performance of two data assimilation algorithms with model error which both utilise Gaussian approximations when constructing the posterior (filtering) distributions of the system state but they differ in updating the covariance information given sparse observations of the underlying true dynamics. Studying the effects of the sparsity of observations on the resulting estimates plays a crucial role in the analysis. This is driven by the desire to mimic the configuration encountered in realistic problems when the observed data corresponds to finite resolution measurements, and the information about the unresolved dynamics is, at best, scrambled with the information about the resolved components of the underlying dynamics. Mindful of the difficulties that arise in the theoretical treatment and highlighted below we focus in this paper on a detailed numerical study of the above issues; analytical considerations prove challenging (but not hopeless) and are postponed to another publication. One can single out two main sources of model error in filtering which are studied below. The first arises from the necessary use of reduced models with coarse resolution for the forward operator, while the second source of model error occurs in the observation operator which - even for linear observation process - mixes the resolved and unresolved dynamics through aliasing of the truth signal. Non-aliased observations ignore the second source of model error by equating the true observations with those of the reduced dynamics alone. This second source of model error is called the representation error [51, 34, 26] and is important to mitigate when filtering complex systems.

Two classes of data assimilation algorithms are considered. In particular, amongst the possible approximate Gaussian nonlinear filtering algorithms these are by far the simplest and least expensive. The first algorithm, 3DVAR [51], is prototypical of many approximate Gaussian filters used in practice today and has its origin in weather forecasting. The 3DVAR method and its generalisations such as the Extended Kalman filter (ExKF, [36]) and the Ensemble Kalman filter (EnKF, [19]), are observed to be accurate when applied in the absence of model error (i.e., the *perfect model scenario*), provided that appropriate so-called *covariance inflation* is used to weigh the observations in favour of the model [46, 66]. Theoretical results explaining these characteristics in the case of 3DVAR may be found in [10, 7] in the context of filtering dissipative (finite and infinite-dimensional) dynamical systems; however, this important analysis concerns the tractable, idealised case when noisy observations of individual spectral modes of the truth are available and certain operators are assumed to commute. Several technical difficulties arise in the presence of spatially sparse observations - the most important one is associated with non-diagonality of the observation operator in an appropriate spectral basis. It is important to note that the results of [10, 7] rely on the existence of an inertial manifold and are closely related to the works [32, 59, 6]. The latter body of work has been extended also to data assimilation schemes for spatial observations [3], following on from the early results of [22, 37], and all leveraging the concept of finitely many determining degrees of freedom first introduced in the seminal works [21, 43].

Practical data assimilation is typically conducted in the presence of significant model error, and the perfect model scenario does not apply [56]. The second algorithm considered is the SPEKF filter [24, 55], which represents a more general class of approximate Gaussian filters, has recently been introduced as an efficient way of dealing with this issue. This approach exploits cheap, exactly solvable, conditionally Gaussian forward models to accommodate model error and to propagate the covariance information via an online learning of certain auxiliary processes that affect the model dynamics; see §4.2.1 for a new formal derivation of the forward model in SPEKF filters. This method has been shown to be effective on a range of test models and full-scale applications [24, 23, 29, 25, 8, 40, 55], and it was further extended in [8] and in [9] to the case of superresolving the state of one-dimensinoal PDE models from aliased observations; also, see [54] and [55, Chapter 7] for nonlinear forward operators with aliased observations. While the numerical evidence concerning the performance of SPEKF-type filters is very promising, theoretical analysis poses a number of technical challenges which stem from the non-diagonal structure of the observation operator and the fact that bounds on error in both the mean and covariance updates need to be established. Thus, as the first step in the analysis, it is of interest to compare the SPEKF and 3DVAR algorithms on a canonical test problem which allows for introduction and control of model error in the forward model, while assimilating spatially sparse data. The 2D Navier-Stokes equation provides such a test problem as it represents a prime example of a dissipative infinite-dimensional dynamical system prototypical of the high-dimensional state estimation problem to which data assimilation is applied in practice. The model error in the forward model dynamics is introduced by the spectral (Galerkin) truncation of the original dynamics, while the sparsity of observations and their spectral resolution is controlled by the distance between nodes on the observation grid in the spatial domain. The main conclusions of our study are as follows:

(i) With appropriate parameter choices and for non-aliased observations, both approximate Gaussian filters perform well in reproducing the mean of the desired filtering probability distribution in various dynamical regimes of the dissipative 2D Navier-Stokes dynamics. (Uncertainty of the mean estimates, e.g., the covariance information, is not considered here because of a very high computational cost of obtaining these quantities, which in the present setting, requires Markov Chain Monte Carlo sampling of multimodal densities over high-dimensional spectral domains).

(ii) In the presence of sparse aliased observations of turbulent dynamics (complex and high-dimensional in spectral space) the SPEKF-type algorithms, which do not require covariance tuning, perform significantly better than a tuned 3DVAR algorithm.

(iii) The need to modify the background covariance in the 3DVAR algorithm in order to induce stability of estimates and avoid divergence poses a significant drawback to this approach and related methods; the background covariance is estimated from historical data while the optimal tuning via the covariance inflation requires hind-cast adaptation.

(iv) Superresolution of aliased observations forward model with the spectral resolution higher than the resolution of the observations is beneficial for the SPEKF filters but it is detrimental to the 3DVAR filter. This applies to the performance for recovering the full truth signal, as well as when the recovery of the primary modes resolved by the observations is considered. Roughly, superresolution allows one to treat the effect of aliasing on the primary modes as a coloured noise, while some information on the unobserved dynamics can be recovered due to the action of the gain operator which projects the aliased observations onto the unresolved modes. SPEKF algorithms can learn, and to some extent filter-out the additional coloured noise on-the-fly, while the forward model in 3DVAR does not have enough degrees of freedom to account for such effects.

Detailed discussion of the above conclusions is presented below. We also note that comparisons of a number of approximate Gaussian estimators excluding SPEKF have been carried out recently (e.g., [27, 58, 48, 72, 73] for variants of 3DVAR, 4DVAR, ExKF and EnKF). Comparison of SPEKF filters with the other algorithms is needed as a complement of this work and we comment in conclusions on further work related to this issue. A survey of other recent multi-scale approaches to filtering and prediction of complex systems can be found in [53, Chapter 5].

The rest of the paper is structured as follows: In section 2 we describe the dynamics used as a test problem used in the comparison of performance of data assimilation algorithms; some necessary terminology and concepts used later, including definitions of relevant Hilbert spaces, is introduced there. Section 3 outlines the main concepts and procedures which lead to the derivation of a class of approximate Gaussian filters which are used throughout this paper in the context of data assimilation. Notions of forward model dynamics used in the filtering algorithms are discussed subsequently, as well as data generation configurations leading to either aliased or non-aliased observations. Three particular data assimilation algorithms, 3DVAR, SPEKF, and GCF, which we focus on are described in section 4. In particular, we outline there technical differences between superresolving and non-superresolving versions of these algorithms. The bulk of the numerical tests are discussed systematically in section 5; non-superresolving and superresolving algorithms are compared in various regimes of the 2D Navier-Stokes dynamics, and for various spectral resolutions of forward models in the filters. We conclude in section 6 summarising the main findings and outlining the directions for future work.

2. Test Problem

Comparison of data assimilation algorithms requires a canonical test problem which allows to control the dynamical properties of the truth signal and generation of observation data, as well as introduction and control of model error in the forward model used in the subsequent state estimation. A version of two-dimensional (2D) Navier-Stokes equation provides such a test problem as it represents a prime example of a dissipative infinite-dimensional dynamical system with a wide range of dynamical regimes, which is prototypical of the high-dimensional state estimation problem to which data assimilation is applied in practice. Some necessary terminology and concepts used in the subsequent sections, including definitions of relevant Hilbert spaces, are introduced here. Further details concerning the set up necessary for the subsequent state estimation are discussed in §3.

2.1. Incompressible 2D Navier-Stokes equation with a linear drag.

We consider the dissipiative dynamics with a global attractor given by a modified version of the incompressible Navier-Stokes equation on the torus $\mathbb{T}^2 := [0, L) \times [0, L)$, $L > 0$, with an additional linear dissipative term:

$$\partial_t u + \kappa^2 u - \nu \Delta u + u \cdot \nabla u + \nabla p = f, \qquad \text{for all} \quad (x,t) \in \mathbb{T}^2 \times (0,\infty), \tag{1}$$

$$\nabla \cdot u = 0, \qquad \text{for all} \quad (x,t) \in \mathbb{T}^2 \times (0,\infty), \tag{2}$$

$$u(x,0) = u_0(x), \quad \text{for all} \quad x = (x_1, x_2) \in \mathbb{T}^2. \tag{3}$$

Here $u \colon \mathbb{T}^2 \times (0,\infty) \to \mathbb{R}^2$ is a time-dependent vector field representing the velocity, $p \colon \mathbb{T}^2 \times (0,\infty) \to \mathbb{R}$ represents the pressure, $f \colon \mathbb{T}^2 \to \mathbb{R}^2$ is the forcing. We assume throughout that u_0 and f average to zero over $\mathbb{T}^2$ which implies that $u(\cdot, t)$ solving (1)-(3) has zero average over $\mathbb{T}^2$ for $t \geqslant 0$. Both the viscosity $\nu > 0$ and the linear 'drag' coefficient κ induce dissipation in the dynamics (1)-(2) but their effects on the long-time dynamics are very different. As outlined below, the above system has a global attractor whose dimensionality grows with the ratio of forcing to dissipation. Increasing the dimension of the unstable manifold of the attractor and the number of positive Lyapunov exponents results in dynamics that becomes

progressively less predictable, and thus more attractive for testing data assimilation algorithms. The linear damping term, controlled by the parameter κ, provides a more suitable attractor dynamics whose energy is not as concentrated at the large scales as for the standard 2D Navier-Stokes dynamics; this type of modification takes account of a large-scale dissipation and is commonly used in modelling quasi-2D turbulence encountered in experiments [11, 60, 69, 68]. In the subsequent sections, we will consider three different dynamical regimes of the system (1)-(3) with drastically different energy spectra on the attractor (see Figure 3).

The infinite-dimensional dynamical system corresponding (1)-(3) is derived in a standard fashion from the functional representation of the above equations; more details can be found in, e.g., [65], but we repeat the main steps and notions since they will be needed in subsequent considerations. Define the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}^1$ as the closures of the set of divergence-free functions

$$\mathcal{T} := \left\{ u(\,\cdot\,,t) \in L^2(\,\mathbb{T}^2,\mathbb{C}^2) : \quad u(\,\cdot\,,t) \ \text{ trig. polynom.}, \ \ \nabla\cdot u(\,\cdot\,,t) = 0, \ \ \int_{\mathbb{T}^2} u(x,t)dx = 0 \right\}, \tag{4}$$

in, respectively, $L^2(\mathbb{T}^2,\mathbb{C}^2)$ and $H^1(\mathbb{T}^2,\mathbb{C}^2)$. Denote the inner product in $\mathcal{H}^1$ by $\langle\langle\cdot,\cdot\rangle\rangle$ and the induced norm by $\|\cdot\|$. The inner product in $\mathcal{H}$ is denoted by $\langle\cdot,\cdot\rangle$ with the induced norm denoted by $|\cdot|$. The inclusion $\mathcal{H}^1 \hookrightarrow \mathcal{H}$ is compact by the Rellich compactness theorem. Then, any real-valued $u \in \mathcal{H}$, including weak solutions of (1)-(3) discussed below, can be represented as

$$u(x,t) = \sum_{k\in\mathbb{Z}^2\setminus\{0\}} u_k(t)\psi_k(x), \qquad u_{-k} = -u_k^*. \tag{5}$$

where $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$, $\psi_k\colon \mathbb{T}^2 \to \mathbb{C}^2$, is the orthonormal basis in $\mathcal{H}$

$$\psi_k(x) := \frac{k^\perp}{|k|}\exp\left(\frac{2\pi \mathrm{i}\, k\cdot x}{L}\right), \tag{6}$$

with $k = (k_1,k_2) \in \mathbb{Z}^2\setminus\{0\}$ and $k^\perp := (k_2,-k_1)^{\mathrm{T}}$ and $|k| = (k_1^2+k_2^2)^{1/2}$. We will confine our attention to time-independent *Kolmogorov forcing* $f \in \mathcal{H}$ (e.g., [52])

$$f(x) = \sum_{k=\mathfrak{S}_{N_f}} f_k\,\psi_k(x), \qquad f_{-k} = -f_k^*, \tag{7}$$

which acts at a subset of the wavenumbers $\mathfrak{S}_{N_f} := \{k^i \in \mathbb{Z}^2\setminus\{0\}\colon |k^i| = N_f\}$. The basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ is related to the Fourier basis $\{\phi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ via $\phi_k = |k|^{-1}(k^\perp\cdot\psi_k)$. Such a special forcing has an attractive mathematical theory (see [52, Chapter 2]).

The functional form of (1)-(3) is obtained via the orthogonal (Leray) projection $\mathscr{P}_L : L^2(\mathbb{T}^2,\mathbb{C}^2) \to L^2(\mathbb{T}^2,\mathbb{C}^2)$ with the range in $\mathcal{H}$ so that

$$\frac{du}{dt} + \mathcal{L}u + \mathcal{B}(u,u) = f, \qquad u_0 \in \mathcal{H}, \tag{8}$$

which is understood in the dual of $\mathcal{H}^1$. Here, $\mathcal{B}(u,v) = \mathscr{P}_L\big((u\cdot\nabla)v\big)$ is a bilinear form on $\mathcal{H}^1$, and $\mathcal{L} = -\mathscr{P}_L(\Delta - \kappa^2)$ is a closed positive operator in $\mathcal{H}$ with the domain of definition $\mathcal{H}^2(\mathbb{T}^2,\mathbb{C}^2)\cap\mathcal{H}^1$ and eigenvalues $(2\pi/L)^2 + \kappa^2 = \tilde\lambda_1 < \tilde\lambda_2 < \ldots$, which are related to the eigenvalues, $\{\lambda_i\}_{i\in\mathbb{N}}$, of the (closed positive) Stokes operator, $A = -\mathscr{P}_L\Delta$, via $\tilde\lambda_i = \lambda_i + \kappa^2$. Classical theorems (see, e.g., [14]) imply that, for all $u_0 \in \mathcal{H}$, the system (8) has a unique weak solution $u \in C_b(\mathcal{H},\mathbb{R}^+)\cap C_{loc}(\mathcal{H}^1,\mathbb{R}^+)\cap L^2_{loc}(\mathcal{H}^1,\mathbb{R}^+)$, where the one-parameter semigroup $\Psi_t : \mathcal{H}^1 \to \mathcal{H}^1$, $t \geqslant 0$, may be extended to act on $\mathcal{H}$ so that $u(t) = \Psi_t(u_0)$ for $u_0 \in \mathcal{H}$.

The existence of the global attractor for the system (8) stems from the fact that $\mathcal{L} : \mathcal{H} \to \mathcal{H}$ is a coercive linear operator satisfying

$$\langle \mathcal{L}u, u\rangle \geqslant \tilde{\lambda}_1 |u|^2, \tag{9}$$

and the bilinear operator $\mathcal{B} : \mathcal{H}^1 \times \mathcal{H}^1 \to \mathcal{H}$ satisfies

$$\langle \mathcal{B}(u,u), u\rangle = 0, \quad \forall\, u \in \mathcal{H}^1, \tag{10}$$

$$\langle \mathcal{B}(u,v), v\rangle \leq C\, \|u\|\|v\|\, |v|, \quad C > 0, \ \ \forall\ u, v \in \mathcal{H}^1. \tag{11}$$

The global upper bound on the norm of the solution of (8) is obtained (see, e.g. [65, 15]) by combining the above properties with the Gronwall lemma, and it is given by

$$|u(t)|^2 \leqslant |u(0)|^2 e^{-\nu\tilde{\lambda}_1 t} + \frac{|f|^2}{\nu^2\tilde{\lambda}_1^2}\Big(1 - e^{-\nu\tilde{\lambda}_1 t}\Big), \qquad t \geqslant 0. \tag{12}$$

Consequently, the system (8) has a global attractor $\mathfrak{A} \subset \mathcal{H}$ satisfying

$$\mathfrak{A} := \left\{ u_0 \in \bigcap_{t\geqslant 0} \Psi_t \mathcal{H} :\ |\Psi_t u_0| \leqslant \frac{|f|^2}{\nu^2\tilde{\lambda}_1^2},\ t \in \mathbb{R} \right\},$$

which is the smallest compact, connected subset of $\mathcal{H}$ that attracts all the solutions. Global attractors for the Navier-Stokes equations have been studied extensively in, e.g., [14, 65]. The nature of solutions on the attractor depends on the dimensionality of its unstable manifold which, in turn, is controlled by the forcing f and dissipation ν parameters. Thus, the dynamics of (1)-(3) or (8) provides a useful test problem with which to examine some of the issues inherent in data assimilation, given the possibility of generating different dynamical regimes with a controllable number of positive Lyapunov exponents which affect the effective dimensionality of the long-time dynamics. In the subsequent sections we focus on estimating the state of (8) evolving on the attractor $\mathfrak{A}$ based on partial noisy observations.

3. Framework for time-sequential data assimilation

Here, we focus on state estimation in the context of the deterministic dynamics (8) acting on the Hilbert space $\mathcal{H}$. The classical (stochastic) filtering problem aims at a sequential-in-time, probabilistic (i.e., measure-valued) estimation of the state $u(x,t)$ by combining the governing dynamics (8) with its partial noisy observations within a Bayesian framework. Assuming that the observations of the true state are known at discrete times $\{t_n\}_{n\in\mathbb{N}}$, the goal of filtering is to find a map

$$\mathbb{P}_n(u\,|\,\mathcal{Y}_n) \xrightarrow{\mathfrak{F}} \mathbb{P}_{n+1}(u\,|\,\mathcal{Y}_{n+1}), \tag{13}$$

where $\big\{\mathbb{P}_n(u\,|\,\mathcal{Y}_n)\big\}_{n\in\mathbb{N}}$ denotes an ordered sequence of conditional probability measures on the state $u(x,t_n)$ given the sequence of observations $\mathcal{Y}_n$ of that state up to time t_n. This is particularly important in applications where the initial condition $u(x,0)$ in (3) is not known exactly, and the noisy data can be used to compensate for this lack of initial knowledge of the system state. Note, however, that in realistic scenarios simulating the infinite-dimensional dynamical system corresponding to (8) is not possible even if the truth dynamics is known exactly. Moreover, the probability measures in (13) are defined on the function space $\mathcal{H}$, a fact that points to a further computational intractability of general filtering problems arising from PDE's.

Here, our aim is to study filtering algorithms which combine noisy observations of the truth with *forward* dynamics obtained from a finite-dimensional approximation of the truth. For clarity and self-containment, we outline below the basic notions of stochastic filtering in discrete time, and we describe steps required

for deriving a class of approximate Gaussian filters (see, e.g., [45, 55] for more details). Despite important differences between various incarnations of such filters, they all share the same general structure, owing to the same Gaussian approximations imposed on the underlying probability densities. The numerical tests of performance of two particular algorithms, 3DVAR [2, 16, 46] and SPEKF [24, 23, 8, 9] discussed in §4, is examined in §5.

The key to deriving a tractable data assimilation framework for nonlinear problems arising from PDE's lies in imposing Gaussian constraints on the so-called prior and posterior probability measures, as outlined below. In what follows we assume that all the measures have a density and, for notational simplicity, we do not distinguish between the measure and its density. Assume that the measure on the initial conditions is given by

$$\mathbb{P}_0(u) \simeq \mathcal{N}(m_0, C_0), \tag{14}$$

so that the initial state to be estimated is only known statistically. Here, the truth is represented by the solution to (8) and given by $u(x,t) = \Psi_t(u_0(x))$, where Ψ_t is generated by the truth dynamics and given by a one-parameter semigroup on $\mathcal{H}$ (see §2.1). Let H denote a linear operator from $\mathcal{H}$ into some Polish space $\mathscr{Y}$, and assume that one observes the state at equally-spaced time intervals $t_n = n\Delta$, $0 < \Delta < \infty$, and that the observations are of the form

$$y_n = H\Psi_{t_n}(u_0) + \eta_n, \qquad n \in \mathbb{N}, \tag{15}$$

where $\{\eta_n\}_{n\in\mathbb{N}}$ is an i.i.d sequence, independent of u_0, with $\eta_n \sim \mathcal{N}(0,\Gamma)$. The truth process $\{u_n\}_{n\in\mathbb{N}}$ at the sequence of the observation times $\{t_n\}_{n\in\mathbb{N}}$ can be written as

$$u_{n+1} = \Psi_\Delta(u_n), \tag{16}$$

where $u_n = \Psi_{t_n}(u_0) := \Psi_{n\Delta}(u_0)$ so that $u_{n+1} = \Psi_\Delta \circ \Psi_{n\Delta}(u_0) = \Psi_{(n+1)\Delta}(u_0)$ and, consequently,

$$\mathbb{P}_n(y \,|\, u_n) \simeq \mathcal{N}(Hu_n, \Gamma). \tag{17}$$

The aim of the filtering algorithm is to find the conditional (so-called filtering) density $\mathbb{P}_n(u \,|\, \mathcal{Y}_n)$ given the observations $\mathcal{Y}_n := \{y_i\}_{i=1}^n$, $n \in \mathbb{N}$, accumulated up to the time t_n (see, e.g, [4] for a more rigorous formulation). In what follows, we will consider approximate Gaussian filtering algorithms which enforce

$$\mathbb{P}_n(u \,|\, \mathcal{Y}_n) \simeq \mathcal{N}(m_n, C_n), \tag{18}$$

on the filtering density. The key issue in designing an approximate Gaussian filter is to find an update rule of the form (e.g., [45, 55])

$$(m_n, C_n) \to (m_{n+1}, C_{n+1}). \tag{19}$$

This update rule is determined directly by imposing another Gaussian constraint, now on the prior density,

$$\mathbb{P}_{n+1}(u \,|\, \mathcal{Y}_n) \simeq \mathcal{N}(\widehat{m}_{n+1}, \widehat{C}_{n+1}), \tag{20}$$

and utilising the linear form of the observations in (15) with additive Gaussian noise $\eta_n \sim \mathcal{N}(0,\Gamma)$. In situations when the observation sequence is discrete in time, the update (19) is usually split into two parts

$$(m_n, C_n) \xrightarrow{\mathfrak{P}} (\widehat{m}_{n+1}, \widehat{C}_{n+1}) \xrightarrow{\mathfrak{A}} (m_{n+1}, C_{n+1}). \tag{21}$$

The prediction (or forecast) step $\mathfrak{P}$ is the map

$$(m_n, C_n) \to (\widehat{m}_{n+1}, \widehat{C}_{n+1}). \tag{22}$$

The subsequent *analysis* $\mathfrak{A}$ step is given by

$$(\widehat{m}_{n+1}, \widehat{C}_{n+1}) \to (m_{n+1}, C_{n+1}). \tag{23}$$

For the prediction step one simply imposes the approximation

$$\widehat{m}_{n+1} = \Phi_\Delta(m_n), \tag{24}$$

utilising the flow map, Φ_Δ, of an approximate model which is usually different from the truth dynamics, Ψ_Δ, while the choice of $\widehat{C}_{n+1}$ depends on the specific filter (as will be described in §4). For the analysis step the assumptions (18) and (20) imply that

$$\mathbb{P}_{n+1}(u \,|\, \mathcal{Y}_{n+1}) \simeq \mathcal{N}(m_{n+1}, C_{n+1}), \tag{25}$$

and an application of the Bayes' rule yields the following map for the analysis step:

$$C_{n+1} = \widehat{C}_{n+1} - \widehat{C}_{n+1}H^*(\Gamma + H\widehat{C}_{n+1}H^*)^{-1}H\widehat{C}_{n+1}, \tag{26}$$

$$m_{n+1} = \widehat{m}_{n+1} + \widehat{C}_{n+1}H^*(\Gamma + H\widehat{C}_{n+1}H^*)^{-1}(y_{n+1} - H\widehat{m}_{n+1})), \tag{27}$$

where m_{n+1} represents the filter estimate of the state u_{n+1} in (8), and C_{n+1} is a linear symmetric positive-definite operator from $\mathcal{H}$ into itself. The term

$$K_{n+1} = \widehat{C}_{n+1}H^*(\Gamma + H\widehat{C}_{n+1}H^*)^{-1}, \tag{28}$$

in (26)-(27) is referred to as the (Kalman) gain. Note that the equations (26)-(27) have the same structure as the standard Kalman filter equations (e.g., [45]); in particular, they are linear in the incoming observations y_n. However, none of the approximate Gaussian filters studied in §4, 5 reduce to the standard Kalman filter. Note that the estimates generated by (26)-(27) would coincide with the Kalman filter estimates only if Φ_Δ were a linear map which coincides with the truth, i.e. $\Phi_\Delta = \Psi_\Delta$, and with $\widehat{C}_{n+1}$ given by

$$\widehat{C}_{n+1} = \Psi_\Delta C_n \Psi_\Delta^*; \tag{29}$$

in such a case (26) reduces to the algebraic Riccati equation and the filter estimates are optimal w.r.t. the mean square error (e.g., [20, 1]). The above requirement is clearly impossible in the present setting, since the truth dynamics (8) is non-linear. Both filters considered in the subsequent sections will share the same structure of the update map, (27)-(26), but they will employ different forward dynamics with model error, i.e., $\Phi_\Delta \neq \Psi_\Delta$, and they will use different updates of the covariance $\widehat{C}_n$ in the prior density (20); these differences will be shown to have important consequences on the performance of the filters.

3.1. **Generation of the synthetic truth in tests of filtering algorithms.** We follow the standard setup when considering the performance of filtering algorithms from the numerical viewpoint. Thus, in the subsequent experiments the synthetic 'truth' data will be generated from a numerical simulation of the system (1)-(3) on the torus $\mathbb{T}^2 := [0, L) \times [0, L)$, $L > 0$, resolving a large but unavoidably finite number of modes ψ_k (6). That is, the synthetic 'truth' is given by

$$u_\Lambda(x,t) = \sum_{|k_{x,y}| \leqslant \Lambda} u_k(t)\psi_k(x), \qquad u_{-k} = -u_k^*, \quad \Lambda \gg 1, \tag{30}$$

which solves the dynamics on $\mathcal{H}_\Lambda \subseteq \mathcal{H}$

$$\frac{du}{dt} + \mathcal{L}u + P_\Lambda \mathcal{B}(u,u) = P_\Lambda f, \qquad u_0 \in \mathcal{H}_\Lambda \subset \mathcal{H}, \tag{31}$$

where P_Λ denotes the projection onto $\mathcal{H}_\Lambda$ and the synthetic truth solutions are given by $u_\Lambda(t) = \Psi_t^\Lambda(u_0)$. The numerical simulation of the dynamics in (31) is carried out in a standard fashion via Galerkin approximation of the velocity field and solved by a pseudo-spectral method in the divergence-free basis defined through (6) which is combined with a 4th order Runge-Kutta time-stepping. We use a modification of a fourth-order Runge-Kutta method, ETD4RK Cox and Matthews (2002), in which the heat semigroup is used together with Duhamel's principle to solve exactly for the diffusion term. A spectral Galerkin method Hesthaven et al. (2007) is used in which the convolutions arising from products in the nonlinear term are computed via FFTs. A double-sized domain in each dimension is used, buffered with zeros, which results in 642 grid-point FFTs, and only half the modes are retained when transforming back into spectral space in order to prevent dealiasing, which is avoided as long as fewer than 2/3 the modes are retained. Data assimilation in practice always contends with poor spatial resolution, particularly in the case of the atmosphere in which there are many billions of degrees of freedom. Here, the important resolution consideration is that the unstable modes, which usually have long spatial scales and support in low wavenumbers, are resolved. Therefore, the objective is to obtain high temporal resolution in the sense of reproducibility rather than high spatial resolution.

3.2. **Forward model dynamics in filters.** The filtering algorithms we study operate on a finite-dimensional subspace $\mathcal{H}_N \subseteq \mathcal{H}_\Lambda \subseteq \mathcal{H}$; this set-up is dictated by both the computational constraints and the desire to adhere to realistic scenarios. Consequently, the forward dynamics in the filters will be based on spectrally truncated models with solutions spanned by a finite set of modes $\{\psi_k\}_{|k_{1,2}|\leqslant N}$ so that

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \qquad u_{-k} = -u_k^*, \quad N \ll \Lambda, \tag{32}$$

The choice of the forward dynamics Φ_Δ in the filter update (27)-(26) will depend on the algorithm used.

In the 3DVAR algorithm (§4.1) the forward dynamics will be given by the truncated dynamics of (8) which is given by

$$\frac{du}{dt} + \mathcal{L}u + P_N\mathcal{B}(u,u) = P_N f, \qquad u_0 \in \mathcal{H}_N \subset \mathcal{H}, \tag{33}$$

where $P_N : \mathcal{H} \to \mathcal{H}_N$ is the projection onto the finite dimensional subspace of $\mathcal{H}$, and $u_N(t) = \Phi_t^N(u_0)$ with $\Phi_t^N : \mathcal{H}_N \to \mathcal{H}_N$.

In the SPEKF algorithm (§4.2) the forward dynamics on $\mathcal{H}_N$ will be given by a linear stochastic non-Gaussian model which is statistically exactly solvable and thus computationally inexpensive. It is important to note that, as long as $N < \Lambda$, the forward dynamics in both algorithms will contain a model error; this configuration aims at mimicking realistic scenarios in which the true dynamics is not known exactly while the best possible estimates are sought.

3.3. **Generation of observations.** In line with the setup of §3, we assume that the observations (15) are linear in the state variable and corrupted by an additive i.i.d. Gaussian noise with a time-independent covariance. We assume throughout that the observation operator H in (15) is trace class but it need not be diagonal in $\mathcal{H}$. Thus, we consider two distinct classes of observations which will have important consequences on subsequent output of the filtering algorithms.

3.3.1. ***Non-aliased observations***. In this case noisy observations of $M \times M$ individual modes of the truth process are available, i.e.,

$$y^{\text{M}}(x,t_n) = \sum_{0<|k_{1,2}|\leqslant M} \big(u_k(t_n) + \eta_k\big)\psi_k(x), \qquad n \in \mathbb{N}, \tag{34}$$

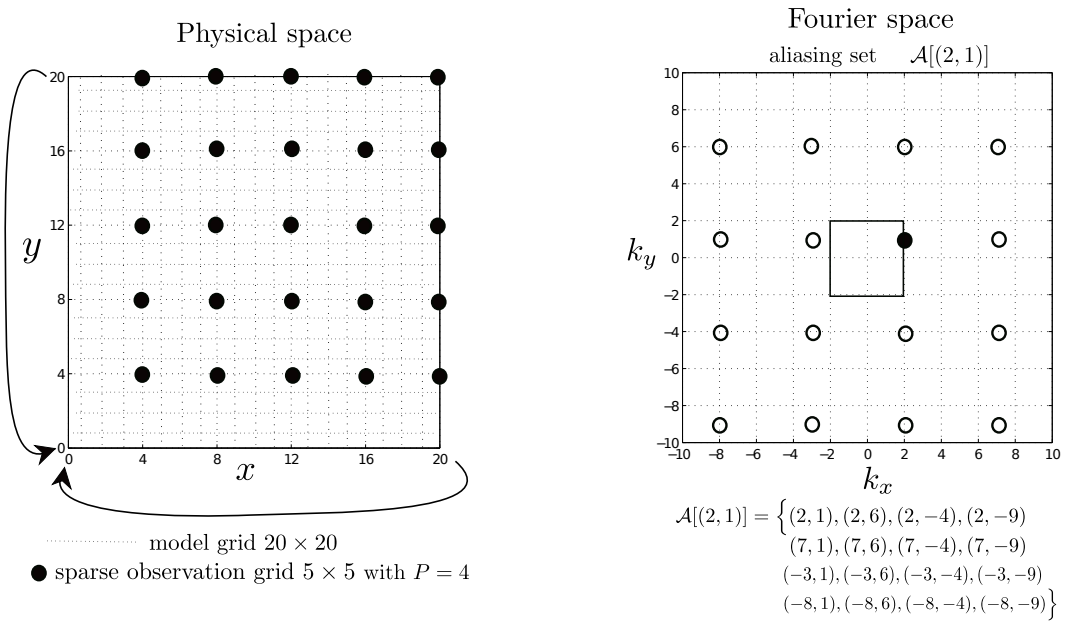

FIGURE 1. Schematic illustration of aliasing in the 2D spectral domain with the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ in (6) due to sparse observations in the 2D spatial domain $(x_1, x_2) \in \mathbb{T}^2$; here, the 5×5 sparse observation grid is a regular subset of the doubly periodic 20×20 model mesh so that every $P = 4$ node is observed (left). The aliasing set $\mathcal{A}(\ell)$ of wavenumber $\ell = (2, 1)$ is shown in the spectral (Fourier) domain (right). In this case, the sparsity of observations aliases modes with $|k_{1,2}| > 2$ into the *primary* modes $|k_{1,2}| \leqslant 2$ which can be resolved by the observation grid. Note that the primary modes need not be the most energetic modes in the respective aliasing sets.

where $\eta_k \sim \mathcal{N}(0, \Gamma)$ for any $0 < |k_{1,2}| \leqslant M$, and the *forward model* (cf. §3.2) resolves $N \times N$ modes ψ_k in (6). The observations $y^{\mathrm{M}}(x, t_n) = \sum_{0<|k_{1,2}|\leqslant M} y_k(t_n)\psi_k(x)$ can be represented in the associated sequence space as

$$Y_n = HU_n + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma), \qquad n \in \mathbb{N}, \tag{35}$$

where $Y_n := \big(y_k(t_n)\big)_{0<|k_{1,2}|\leqslant M}$, $U_n := \big(u_k(t_n)\big)_{k\in\mathbb{Z}^2\setminus\{0\}}$, and the linear operator H is diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and has rank $(2M)^2$.

The configuration in (34) is unrealistic from the practical viewpoint since it ignores representation error [51, 34, 26] and it implies the ability to observe the dynamics of individual modes ψ_k, $0 < |k_{1,2}| \leqslant M$. However, this setup is amenable to detailed analysis, at least for the 3DVAR filter in [7], which is why we consider it here and compare filter performance given this type of observations in §5.

3.3.2. ***Aliased observations***. In this case the state $u(x, t_n)$ is observed at nodes of a finite grid in the spatial domain $x \in \mathbb{T}^2$. In line with realistic configurations (see, e.g., [54]) we assume that the observation nodes lie on a regular grid $x_{i,j} := (x_i, y_j) = (ih, jh)$, $1 \leqslant i, j \leqslant 2M + 1$, such that $(2M + 1)h = 2\pi$. In such a case one has

$$y^{\mathrm{M}}_{i,j}(t_n) = u\big(x_{i,j}, t_n\big) + \zeta_{i,j}\,, \qquad n \in \mathbb{N},\ \ 1 \leqslant i, j \leqslant 2M + 1,\ M \leqslant N, \tag{36}$$

where ζ is an uncorrelated Gaussian field with $\zeta_{i,j} \sim \mathcal{N}(0, \Gamma^o)$. As shown in [54, 28], spatially sparse regular observations on $\mathbb{T}^2$ will alias the information from modes $|k_1| \vee |k_2| > M$ of the truth signal into the observed modes $|k_1| \wedge |k_2| \leqslant M$ (see Figure 1). In fact, a simple calculation similar to that carried out

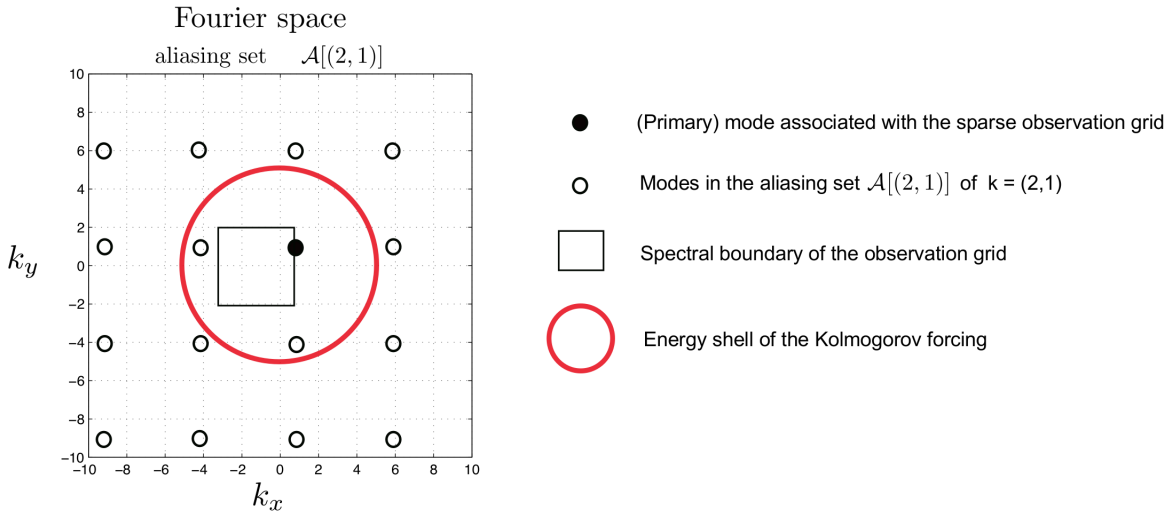

FIGURE 2. Desired test configuration for filtering NSE with Kolmogorov forcing and sparse aliased observations (in physical space). In a dynamical regime with sufficiently large Reynolds number the primary (observed) modes are not always the most energetic ones due to the (possibly intermittent) energy transfer to small scales.

for the discrete Fourier transform in [54, 28] shows that the observation process (36) can be written as

$$y^{\mathrm{M}}_{i,j}(t_n) = \sum_{0<|\ell_{1,2}|\leqslant M} \left(\sum_{k\in\mathcal{A}(\ell)} u_k(t_n) + \eta_{\ell,n} \right) \psi_\ell(x_{i,j}), \qquad n\in\mathbb{N},\ \ 1\leqslant i,j\leqslant 2M+1,\ \ M\leqslant N, \tag{37}$$

where $\eta_{l,n} \sim \mathcal{N}\big(0, \Gamma^0/(2M+1)^2\big)$. The disjoint sets of *aliased* modes $\mathcal{A}(\ell)$, with $\ell=(\ell_1,\ell_2)$, and $0 < |\ell_{1,2}| \leqslant M$, are defined as

$$\mathcal{A}(\ell) = \Big\{ k\in\mathbb{Z}^2\backslash\{0\} : \quad k_{1,2} = \ell_{1,2} + Mq_{1,2}, \quad q_{1,2}\in\mathbb{Z} \Big\}. \tag{38}$$

The aliasing sets $\mathcal{A}(\ell)$ are indexed by the *primary* wavenumbers, $\ell=(\ell_1,\ell_2)$, resolved by the observation grid $\{x_{i,j}\}_{0\leqslant i,j\leqslant 2M+1}$. Similar to the case of non-aliased observations §3.3.1, the observation process $\big(y^{\mathrm{M}}_n\big)_{n\in\mathbb{N}}$ can be written as

$$Y^{\mathrm{M}}_n = H^{\mathcal{A}\{\mathrm{M}\}} U_n + \eta_n, \qquad n\in\mathbb{N}, \quad \eta_n \sim \mathcal{N}(0,\tilde{\Gamma}), \tag{39}$$

where $H^{\mathcal{A}\{\mathrm{M}\}}$ is a linear rank $(2M)^2$ operator which, in the presence of aliased observations, is generally not diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\backslash\{0\}}$.

In summary, aliased observations represent a much more realistic configuration than the non-aliased setup in §3.3.1 since they account for representation model error [55, 51, 34, 26]. In such a case the observation operator is not diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\backslash\{0\}}$. In the numerical tests discusssed in §5 we consider two distinct filtering configurations which depend on the choice of the spectral resolution, $N\times N$, of the forward model (cf. §3.2) relative to the spectral resolution, $M\times M$, of the observation grid:

(i) $\boldsymbol{N = M}$. This configuration corresponds to a filtering procedure aimed at estimating/resolving only the $M\times M$ observed modes (36). In contrast to (34) observations of these modes are corrupted by both the aliased information from modes $|k_{1,2}| > M$ and by the observation noise.

(ii) $\boldsymbol{N = PM}$, $P \in \mathbb{N}^+$. This configuration allows to superresolve the observations within the Bayesian filtering framework and provide estimates on modes in the dynamics beyond spectral resolution $M\times M$ imposed by the grid scale in the physical domain [9, 54, 29, 57]. Similar to the non-superresolving case, $N{=}M$, the observations are corrupted by both the aliased information from the unresolved modes and the observation noise and the observation operator $H^{\mathcal{A}\{\mathrm{M}\}}$ in (39) is not diagonal. However, each of the $(2M)^2$ disjoint aliasing sets $\mathcal{A}$ in (38) contains exactly P^2 modes resolved by the $N\times N$-dimensional forward model. Superresolution algorithms were studied numerically in [9] for a variety of one-dimensional PDE dynamics based on a class of filtering algorithms referred to as SPEKF (see §4.2 and [24, 23, 8]). Below, we extend this study to the case of $2D$ Navier-Stokes dynamics discussed in §2.1, and we compare the performance of two aforementioned algorithms, 3DVAR (cf. §4.1) and SPEKF (cf. §4.2) in this context.

4. Filtering algorithms

The general framework, outlined in §3, for data assimilation exploiting approximate Gaussian filters admits various algorithms which all exploit the same *analysis* update (26)-(27). Important differences between these algorithms appear at the *prediction* step which requires knowledge or approximations of the underlying dynamics which is reflected in the approximation Φ_Δ of the flow map Ψ_Δ in (24) and the update of the prior covariance $\widehat{C}$ in (26)-(27). In the numerical tests discussed in §5, we consider two approximate Gaussian filtering algorithms which are described below. One simplifies covariance computations and the other simplifies the model.

4.1. 3DVAR.

This algorithm has its origin in weather forecasting [51] and it is prototypical of many approximate Gaussian filters used in practice when dealing with high-dimensional estimation problems. Recall that the key update formulas (26)-(27) require a rule for updating the prior mean and the prior covariance, respectively, $\widehat{m}_n$, $\widehat{C}_n$, $n \in \mathbb{N}^+$. In high-dimensional nonlinear problems, such as those arising in filtering nonlinear PDEs, updating the prior covariance quickly becomes computationally intractable even if a finite-dimensional approximation for the forward model dynamics Φ^N_Δ, $1 \ll N < \infty$, is employed on $\mathcal{H}^N \subset \mathcal{H}$ (cf. §3.2). The simplest approximation which drastically reduces the computational cost is to assume $\widehat{C}_n = const.$ in the update equations (26)-(27), and the most common approach to parameterising the prior covariance is

$$\widehat{C}_n = \widehat{C}_{\alpha,\beta} := \alpha + \beta C_0, \qquad \alpha, \beta > 0, \tag{40}$$

where C_0 is the so-called *background covariance* estimated empirically from the attractor dynamics (of (8) in this case), and the constants α, β are commonly referred to as the additive and multiplicative covariance inflation parameters. Covariance inflation is observed to be accurate when applied in the absence of model error, i.e., $\Phi^N_\Delta = \Psi_\Delta$ in (24), provided that appropriate inflation is used to weight the observations in favour of the model [46]; theoretical results explaining the numerical evidence in the case of 3DVAR may be found in [10]. $\widehat{C}_{\alpha,\beta}$ is usually assumed to be diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ which reduces computational complexity; here, we also assume that $\widehat{C}_{\alpha,\beta}$ is diagonal but this is not strictly necessary provided that one can reliably estimate the off-diagonal covariance terms. Given the above ad-hoc simplification of the update (26)-(27), the 3DVAR algorithm [51] for the mean estimate is described by

$$m_{n+1} = (I - K)\Phi^N_\Delta(m_n) + Ky_{n+1}, \tag{41}$$

where the gain operator K in (28) simplifies to

$$K = \widehat{C}_{\alpha,\beta} H^* (\Gamma + H \widehat{C}_{\alpha,\beta} H^*)^{-1}. \tag{42}$$

Note that K is generally not diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ because the observation operator H is generally not diagonal in that basis (cf. §3.3); this fact represents the main technical obstacle in proving accuracy and stability of the 3DVAR algorithm for filtering dissipative PDE's. The tractable case, corresponding to non-aliased observations (§3.3.1), was studied in [7] where the (diagonal) observation operator $H = P_M$ taken to be a projection onto $\{\psi_k\}_{0<|k_{1,2}|\leqslant M}$ was assumed to commute with the diagonal operators $\widehat{C}_{0,\beta}$ and Γ where

$$\widehat{C}_{0,\beta}\,\psi_k \propto \eta^2\,|k|^{-2\zeta}\psi_k, \qquad \Gamma\,\psi_k \propto |k|^{-2\gamma}\psi_k, \tag{43}$$

so that the gain in (42) is

$$K\psi_k \propto \begin{cases} 0\cdot\psi_k & \text{for} |k_{1,2}| > M, \\ \eta^2|k|^{2(\gamma-\zeta)}\big(1+\eta^2|k|^{2(\gamma-\zeta)}\big)^{-1}\,\psi_k & \text{for} |k_{1,2}| \leqslant M. \end{cases} \tag{44}$$

Thus, increasing η corresponds to variance inflation which results in weighting the estimates in favour of observations on the observed modes, as can be seen from (41). In the numerical tests of §5 we will consider 3DVAR filtering with both the aliased and non-aliased observations, while keeping in mind that the case of non-aliased observations provides crucial analytical simplifications at the expense of abandoning realistic constraints.

4.2. **Stochastic parameterisation Kalman filters (SPEKF).** There are two next logical steps beyond the 3DVAR algorithm which employs constant covariances $\widehat{C}_n$ in (26)-(27) when incorporating new data. is to use approximate dynamics in order to make the update. One of the earliest approximations, the Extended Kalman filter (ExKF), utilises a linear tangent approximation of the flow map Φ_Δ so that the prior covariance $\widehat{C}_n$ can be updated in a similar way to the standard Kalman filter algorithm, as in (29), at the expense of a potential filter divergence [50, 30, 41]. In high-dimensional state estimation problems, updating the covariances becomes computationally expensive even in the case of ExKF. A class of so-called Ensemble Kalman filters (EnKF), which proved popular in applications (e.g., [39]), further reduces the computational complexity by estimating the prior covariances $\widehat{C}_n$ via Gaussian approximations of a finite ensemble of predictions propagated by the (nonlinear) flow map Φ_Δ (e.g., [45]). Another approximation, one we exploit here, relies on constructing a reduced stochastic model which is generally non-Gaussian but linear in the modes of the spectral basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and statistically exactly solvable, thus providing an efficient way of updating the prior covariances $\widehat{C}_n$ in the update (26)-(27). It was first demonstrated in the context of linear stochastic models represented in Fourier basis [28, 57, 55] that such a simple modelling of the forward dynamics can be quite effective for filtering systems exhibiting chaotic behaviour in high dimension. The family of SPEKF-type filters, whcih exploits more general non-Gaussian forward models described below, represents a more general class of approximate Gaussian filters whose forward dynamics remains exactly solvable. This approach originated from [28, 29, 55, 24, 23] and was further extended in [8] and in [9] to the case of superresolving the state of one-dimensinoal PDE models from aliased observations. In this subsection we first provide a formal derivation of this reduced stochastic model and then describe how this approach may be used in filtering turbulent regimes of the Navier-Stokes system under consideration.

4.2.1. *The forward model in SPEKF filters.* Consider the truth Navier-Stokes dynamics in the form (8) on the Hilbert space $\mathcal{H}$ and a family of projections $\{P_{\mathcal{A}(\ell)}\}_{\ell\in\mathbb{Z}^2\setminus\{0\}}$ on the disjoint aliasing sets $\mathcal{A}(\ell)$ in (38) so that $u_{\mathcal{A}(\ell)} = P_{\mathcal{A}(\ell)}u \in \mathcal{H}_{\mathcal{A}(\ell)}$, where $\mathcal{H}_{\mathcal{A}(\ell)} \subseteq \mathcal{H}$ is spanned by the modes in the respective aliasing set. Then, the evolution of modes within any aliasing set $\mathcal{A}$ is obtained from (8) as

$$\frac{du_{\mathcal{A}}}{dt} + P_{\mathcal{A}}\mathcal{L}P_{\mathcal{A}}u_{\mathcal{A}} + P_{\mathcal{A}}\mathcal{B}(u,u) = P_{\mathcal{A}}f, \qquad u_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}} \subseteq \mathcal{H}, \quad u \in \mathcal{H}, \tag{45}$$

where $\mathcal{L}$ and $\mathcal{B}$ were defined in §2.1. In order to simplify the notation, we skip the explicit dependence on the primary wavenumbers ℓ indexing the aliasing sets $\mathcal{A}$ in (38). The nonlinear term $P_{\mathcal{A}}\mathcal{B}$ in (45), which couples the evolution of modes in $\mathcal{A}$ to the remaining modes, can be decomposed as

$$P_{\mathcal{A}}\,\mathcal{B}(u,u) = P_{\mathcal{A}}\Big(\mathcal{B}\big(u_{\mathcal{A}},u_{\mathcal{A}}\big) + \mathcal{B}\big(u_{\mathcal{A}},(1-P_{\mathcal{A}})u\big) + \mathcal{B}\big((1-P_{\mathcal{A}})u,u_{\mathcal{A}}\big) + \mathcal{B}\big((1-P_{\mathcal{A}})u,(1-P_{\mathcal{A}})u\big)\Big).$$

Then, the crucial and simple observation [9] is that the projection of the nonlinear interactions on the aliasing set vanish, i.e.,

$$P_{\mathcal{A}}\,\mathcal{B}\big(u_{\mathcal{A}},u_{\mathcal{A}}\big) = 0, \quad \text{or} \quad \langle u_{\mathcal{A}}, \mathcal{B}(u_{\mathcal{A}},u_{\mathcal{A}})\rangle = 0, \tag{46}$$

implying a lack of direct nonlinear interactions between the aliased modes. The remaining nonlinear terms need to be approximated in order to enforce the invariance of $\mathcal{H}_{\mathcal{A}}$ w.r.t. the dynamics of the aliased modes; this is achieved via the Kraichnan's decimated-amplitude scheme [44], namely

$$-P_{\mathcal{A}}\Big(\mathcal{B}\big(u_{\mathcal{A}},(1-P_{\mathcal{A}})u\big) + \mathcal{B}\big((1-P_{\mathcal{A}})u,u_{\mathcal{A}}\big)\Big) \approx -\big(\Gamma_{\mathcal{A}}(t) + i\,\Omega_{\mathcal{A}}(t)\big)u_{\mathcal{A}}, \tag{47}$$

where $\Gamma_{\mathcal{A}},\Omega_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}}$ are real and trace-class, $\Gamma_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}}$ is positive-definite, and

$$P_{\mathcal{A}}\,\mathcal{B}\big((1-P_{\mathcal{A}})u,(1-P_{\mathcal{A}})u\big)dt \approx -B_{\mathcal{A}}(t)dt - \Sigma_{\mathcal{A}}dW_{\mathcal{A}}(t), \tag{48}$$

where $B_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}}$, $\Sigma_{\mathcal{A}}$ is a trace-class operator, and $W_{\mathcal{A}}(t)$ is a cylindrical Wiener process on $\mathcal{H}_{\mathcal{A}}$. The above approximations are not rigorously derived and are based on a physical reasoning in the context of turbulent dynamics of the Navier-Stokes equation. The resulting stochastic approximation has the form of a linear SPDE on $\mathcal{H}_{\mathcal{A}}$

$$du_{\mathcal{A}} = \Big(-\big(\mathcal{L}_{\mathcal{A}} + \Gamma_{\mathcal{A}}(t) + i\,\Omega_{\mathcal{A}}(t)\big)u_{\mathcal{A}} + b_{\mathcal{A}}(t) + f_{\mathcal{A}}\Big)dt + \sigma_{\mathcal{A}}dW_{\mathcal{A}}(t), \qquad u_{\mathcal{A}}(0) \in \mathcal{H}_{\mathcal{A}}, \tag{49}$$

where $\mathcal{L}_{\mathcal{A}} = P_{\mathcal{A}}\mathcal{L}P_{\mathcal{A}}$, and $f_{\mathcal{A}} = P_{\mathcal{A}}f$. The above formal derivation provides an approximate dynamical model which is exactly solvable as long as it remains conditionally Gaussian [49, 25, 23]. In computations, when finite-dimensional approximations of the dynamics are employed, this strategy allows for propagating the second-order statistics in (26)-(27) based on analytical formulas which can be utilised in a number of different approximate Gaussian filtering algorithms outlined below (see [24, 8, 9] for details).

Now, consider now such a computationally realistic situation when the forward model (cf. §3.2) resolves $N < \infty$ spectral modes in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ of $\mathcal{H}$ so that

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \tag{50}$$

and recall that (cf. §3.3) if the observations resolve M spectral modes of the truth, then there exist $(2M)^2$ disjoint aliasing sets $\mathcal{A}(\ell)$, $0 < |\ell_{1,2}| \leqslant M$ defined in (38) into which all the modes $\{u_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ are partitioned. Consequently, for $N < \infty$ the number of modes u_k resolved by the forward model in each aliasing set $\mathcal{A}(\ell)$ is also finite. Then, the stochastic dynamics of the forward model (49) takes a

particularly simple form for $\Gamma_{\mathcal{A}}, \Omega_{\mathcal{A}}, B_{\mathcal{A}}, \Sigma_{\mathcal{A}}$ diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ so that the evolution of modes $\{u_k\}_{k\in\mathcal{A}(\ell)}$ in each aliasing set $\mathcal{A}$ is given by the following system:

$$
\begin{aligned}
&(a)\quad \mathrm{d}u_k(t) = \Big[-\big(\bar{l}_k + \gamma_k(t) + i\omega_k(t)\big)u_k(t) + b_k(t) + f_k(t)\Big]\mathrm{d}t + \sigma_{u_k}\mathrm{d}W_{u_k}(t),\\
&(b)\quad \mathrm{d}\gamma_k(t) = -d_{\gamma_k}\gamma_k(t)\mathrm{d}t + \sigma_{\gamma_k}\mathrm{d}W_{\gamma_k}(t),\\
&(c)\quad \mathrm{d}\omega_k(t) = -d_{\omega_k}\omega_k(t)\mathrm{d}t + \sigma_{\omega_k}\mathrm{d}W_{\omega_k}(t),\\
&(d)\quad \mathrm{d}b_k(t) = \big[(-d_{b_k} + i\omega_{b_k})b_k(t)\big]\mathrm{d}t + \sigma_{b_k}\mathrm{d}W_{b_k}(t),
\end{aligned}
\tag{51}
$$

where γ_k, ω_k, b_k represent the stochastic (Gaussian) multiplicative and additive bias correction terms, which arise from the approximations described above, and $W_{u_k}, W_{\gamma_k}, W_{\omega_k}, W_{b_k}$, are the classical independent Wiener processes. The dynamics of each mode u_k is controlled by a number of tuneable parameters: the stationary mean $\bar{l}_k$, three damping parameters d_{b_k}, $d_{\gamma_k}, d_{\omega_k} > 0$, one phase parameter ω_{b_k}, and noise amplitudes $\sigma_{u_k}, \sigma_{b_k}, \sigma_{\gamma_k}, \sigma_{\omega_k} > 0$; f_k is a deterministic forcing.

A number of relevant remarks is required here concerning the model error judiciously introduced into the forecast model (51):

- We assume that the dynamics of the spectral modes ψ_k in each aliasing set $\mathcal{A}$ is independent of the dynamics of modes in the remaining sets; this approximation was already used in §3.3 and [54], and validated in [28, 29, 40].
- Since the aliasing sets, $\mathcal{A}(\ell)$, for each primary (observed) mode ℓ are disjoint, the N^2-dimensional filtering problem in the physical space can be converted to M^2 independent P^2-dimensional filtering problems for each of the primary modes associated with the sparse observation grid (see figure 1 for an example).
- Unlike the true dynamics, the modes in each aliasing set are coupled only during the assimilation step when (sparse) observations are assimilated. The idea for replacing the complex nonlinear interactions between different spectral modes by multiplicative stochastic damping/frequency corrections and additional stochastic forcing is familiar from stochastic modelling of shear turbulence [62, 17]. Justification of this approximation in the filtering context is given in [9] where it is shown that Fourier modes in geophysical systems with quadratic, advection-type nonlinearity are not directly coupled with the other modes in the same aliasing set.
- Filtering within the aliasing set $\mathcal{A}(\ell)$ involves state vector $\boldsymbol{u}_{\mathcal{A}}$ and the associated non-physical processes $(\gamma_k, \omega_k, b_k), k \in \mathcal{A}(\ell)$ which provide bias correction due to model error. Importantly, the dynamics of these augmented, unobserved processes is adjusted during the data assimilation process, and it allows the algorithm to 'learn' some aspects of model error on-the-fly from the incoming data.

4.2.2. *SPEKF algorithms for superresolving sparsely observed systems.* SPEKF algorithms fall into the category of approximate Gaussian filters. Therefore, in the discrete-time setting, the analysis step in the state estimation process is given by (26)-(27), similar to the 3DVAR filter outlined in §4.1. However, an important difference arises at the forecast step, $(m_n, C_n) \to (\widehat{m}_{n+1}, \widehat{C}_{n+1})$, where the prior mean and covariance are computed from the previous estimates (m_n, C_n) using analytical formulas, due to the aforementioned exact solvability of the forward model (51) for given Gaussian initial conditions. These formulas are complicated and lengthy and an interested reader is referred to [24] and [8, 9]. In short, the exact path-wise solvability of (51) provides an explicit expression for the stochastic flow $\Phi_{\Delta;\omega}$ on each augmented space $\mathfrak{H}_{\mathcal{A}(\ell)}$ indexed by the aliasing sets with $\mathfrak{U} = \{u_k, \gamma_k, \omega_k, b_k : k \in \mathcal{A}(\ell)\} \in \mathfrak{H}_{\mathcal{A}(\ell)}$ which is

then employed to compute the prior mean and covariance as

$$\mathfrak{U}_n \sim \mathcal{N}(m_n, C_n) \quad \longrightarrow \quad \begin{cases} \widehat{m}_{n+1} = \mathbb{E}[\Phi_{\Delta;\omega}(\mathfrak{U}_n)], \\ \widehat{C}_{n+1} = Cov\big(\Phi_{\Delta;\omega}(\mathfrak{U}_n), \Phi_{\Delta;\omega}(\mathfrak{U}_n)\big), \end{cases} \tag{52}$$

independently in each of the $(2M)^2$ disjoint aliasing sets $\mathcal{A}(\ell)$; the expectations are taken w.r.t. a product measure combining the measure on the initial condition, $\mathfrak{U}_n$, and the path measure on the SDE (51). The exact statistical solvability of (51) which leads to analytical formulas for $\widehat{m}_{n+1}, \widehat{C}_{n+1}$ in (52) removes the need for tangent approximations commonly used in the Extended Kalman Filter which is particularly prone to divergence when dealing with turbulent systems [8]. Note that the dimensionality of $\widehat{m}_{n+1}$ and $\widehat{C}_{n+1}$ depends on the spectral resolution N of the forward model, and on the spectral resolution M of the observations (cf §3.3). For $N = M$ (no superresolution of observations) the forward model accounts for only one mode in each aliasing set $\mathcal{A}(\ell)$; hence, $\widehat{m} \in \mathbb{C} \times \mathbb{R} \times \mathbb{R} \times \mathbb{C} \simeq \mathbb{C}^3$ and $\widehat{C} \in \mathbb{C}^{3\times 3}$ for each of the $(2M)^2$ aliasing sets. In the superresolution mode, when $N = P\,M$, $\widehat{m} \in \mathbb{C}^{3P}$ and $\widehat{C} \in \mathbb{C}^{3P\times 3P}$ in each aliasing set. Given that the forward dynamics (51) is linear and diagonal in the augmented state $\mathfrak{U}$, the cross-correlations between different aliased modes $\{u_k, \gamma_k, \omega_k, b_k\}$ and $\{u_j, \gamma_j, \omega_j, b_j\}, k \neq j,\ k, j \in \mathcal{A}(\ell)$, are only possible when $N \neq M$ and they are introduced via the initial conditions obtained at the previous analysis step (26)-(27). Consequently, the off-diagonal terms in the prior covariance $\widehat{C}_{n+1}$ can only decay in the forecast step (52). This fact can be exploited in the context of the model (51) to further reduce the computational cost of the forecast step which might be desirable when superresolving observations. Details of various simplified algorithms were derived and discussed in [8, 9]; here, we recapitulate the properties of the two most efficient algorithms which will be compared against 3DVAR in the next section:

- **cSPEKF**: This *crude SPEKF* algorithm utilises analytical updates to derive $\widehat{m}_{n+1}$ and for the diagonal entries of $\widehat{C}_{n+1}$ in (52). The off-diagonal terms in $\widehat{C}_{n+1}$, corresponding to cross-correlations between $\{u_k, \gamma_k, \omega_k, b_k\}$ and $\{u_j, \gamma_j, \omega_j, b_j\}$, $k \neq j$, $k, j \in A(\ell)$ are neglected. It was shown in [9] that, apart from computational cost reduction, this technique resulted in increased stability in a wide range of dynamical regimes. Neglecting the off-diagonal terms in the prior covariance $\widehat{C}$ can result in reduced filter performance relative to the full covariance version but, as we show later, this seems unlikely when filtering dynamics with a wide range of spatial scales which decorrelate rapidly compared to the assimilation time interval.
- **GCF** : This Gaussian Closure Filter algorithm uses an approximate statistics of the foreward model (51) which is obtained via the simple Gaussian moment closure used frequently in the statistical theory of turbulence; this approach was introduced and used for filtering turbulent signals in [8, 9]. In GCF the covariance matrix $\widehat{C}_{n+1}$ includes all cross-correlations between different modes $\{u_k, \gamma_k, \omega_k, b_k\}$ and $\{u_j, \gamma_j, \omega_j, b_j\}$ for $k, j \in A(\ell)$ within the same aliasing set. For systems with quadratic nonlinearities, such as (51), this closure correctly accounts for the turbulent backscatter in the evolution of the mean but it neglects the third order moments of fluctuations in the evolution of the covariance. In the numerical studies of [8, 9] this algorithm emerged as the most suitable trade-off between the skill and the computational complexity associated with estimating the cross-correlations. The dynamical system for the evolution of the approximate statistics of GCF is given in [9].

5. Numerical Results

In this section we compare the performance of various approximate Gaussian filters described in §4 for state estimation of the spatially extended system given by the 2D Navier-Stokes dynamics (8) outlined

in §2.1. The conclusions discussed below are based on the ability of those Bayesian filtering algorithms to reconstruct the mean of the desired probability measure on the state of the truth dynamics. Thus, we assume throughout that the Bayesian posterior (filtering) distributions are well-defined, implying that the prior on the initial conditions, and observation likelihood are well-defined throughout the time interval considered. The question of sensitivity of the results to choice of prior is not addressed here. Instead, we focus on the issues related to the effects of various approximations at both the prediction and the analysis steps of the considered approximate Gaussian filters.

Key questions driving the choice of the experiments discussed below concern the following:

(i) Does updating the prior covariance $\widehat{C}_{n+1}$ in (26) in the approximate Gaussian filtering algorithms with model error improve the filter performance relative to setting $\widehat{C}_{n+1} = const.$ as in 3DVAR? Moreover, can updating the prior covariance improve the performance of an approximate Gaussian filter relative to 3DVAR with perfect forward dynamics ? (i.e., $N = \Lambda$ in (33) of §3.2)

(ii) Does the performance of the approximate Gaussian filters - 3DVAR or SPEKF- depend on the dynamical regime of (8)?

(iii) How does the performance of the considered filters depend on the spectral resolution N of the forward model, and the spectral resolution M of the observations?

(iv) Does aliasing of observation matter for state estimation?

(v) Which class of the approximate Gaussian filters - 3DVAR or SPEKF - is better suited for super-resolving sparse, aliased observations of a complex spatially extended dynamics?

(vi) Does superresolution of sparse aliased observations help improve the estimates of the spectral modes resolved by the observations?

The above issues are studied for the filters in question by means of numerical experiments which currently provides the only validation method of the studied filtering algorithms, given that analytical results for sparsely observed dynamics do not currently exist.

A systematic numerical investigation of the issues highlighted above requires requires the ability to generate the synthetic truth dynamics with varying degree of complexity. As discussed in §2.1, the dynamics (8) has a global attractor whose dimensionality grows with the ratio of the forcing amplitude to dissipation. In particular, the nature of solutions on the attractor depends on the number of positive Lyapunov exponents which is a proxy for the dimensionality of the unstable manifold of the attractor; the larger the number of positive Lyapunov exponents the more complex and the less predictable the underlying dynamics becomes. Thus, it is obvious that the choice of the forcing and dissipation parameters in the truth dynamics is crucial for generating appropriate dynamical regimes for testing the filtering algorithms. In practice, the spectral resolution of the 'truth' dynamics is finite and given by the truncated version, (31), of the infinite-dimensional system (8). In the examples discussed below the synthetic truth is computed from (31) with $\Lambda = 232$, and the forward model resolution and the resolution of the observations are arranged so that $M \leqslant N \ll \Lambda$. As outlined in §3.1, the numerical simulation of the dynamics in (31) is carried out in a standard fashion by a pseudo-spectral method in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ defined through (6) which is combined with a 4th order Runge-Kutta time-stepping. The choice of the forcing scale N_f in (7) and its amplitude relative to the viscosity ν is relevant in order to set up a chaotic (mildly turbulent) dynamical system with non-negligible energy in the band $N \leqslant |k| \leqslant \Lambda$ so that the model error in the forward model is significant (and at a tuneable level controlled by changing the dissipation ν). We force the dynamics at a single scale $N_f = 8$ and choose three values of the viscosity $\nu = 0.03, 0.003, 0.001$ and $\kappa = 0.001$ in order to obtain three distinct dynamical regimes characterised by different number of active modes on the attractor; these regimes are referred to as laminar ($\nu = 0.03$, $|f_k| = 8$, $N_f = 10$), moderately

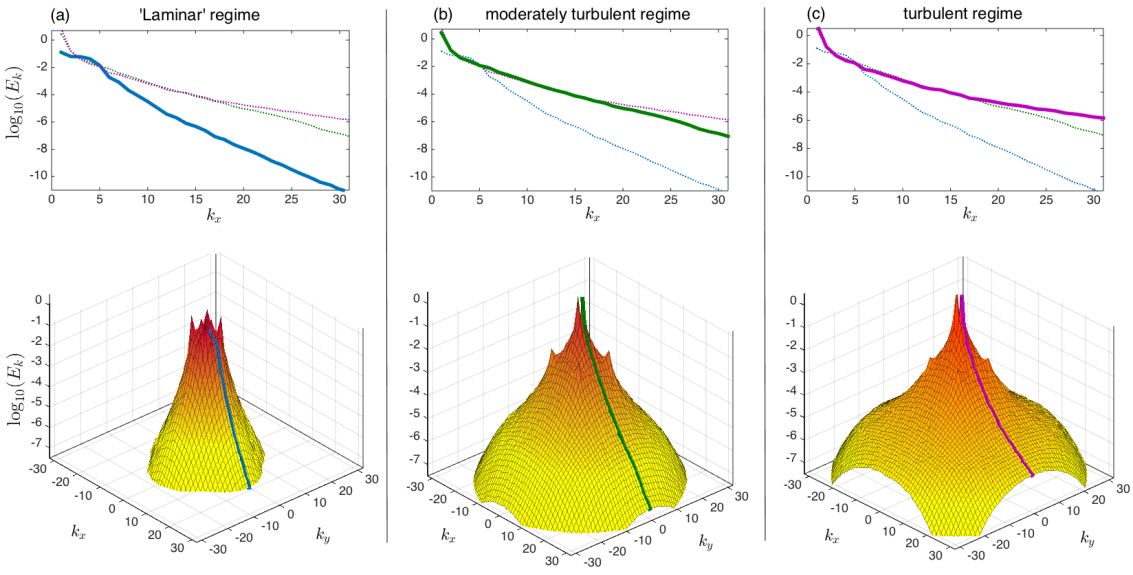

FIGURE 3. Examples of numerically simulated spectra of the truth dynamics for $u_\Lambda(x,t)$ in (31). Top row shows cross-sections for $(k_1, k_2 = 0)$ of the 2D spectra in the bottom row for three different regimes used in the numerical tests of the filtering algorithms introduced in §4. The dynamics (31) with $\Lambda = 64$ is forced at $|k_1| = |k_2| = 8$ and amplitude $|f_k| = 8$ in (8). The remaining parameters in the three regimes are: (a) $\nu = 0.1, \kappa = 0.01$, (b) $\nu = 0.01, \kappa =$, (c) $\nu = 0.001, \kappa = 0.01$.

turbulent ($\nu = 0.003$, $|f_k| = 8$, $N_f = 10$) and turbulent ($\nu = 0.001$, $|f_k| = 8$, $N_f = 10$). Figure 3 shows the relevant spectra associated with these regimes, together with representative snapshots of the vorticity fields. In all these regimes we compare the performance of the approximate Gaussian filtering algorithms introduced in §4, namely: 3DVAR (§4.1), and two algorithms with the stochastic forward model (51), cSPEKF and GCF, described in §4.2.

For each dynamical regime considered below, we study the filter performance for varying spectral resolution N of the forward models such that $N \ll \Lambda$ (i.e., the model resolution is much worse than that of the 'truth' dynamics in (31)). The configuration corresponding to filtering with non-aliased observations (cf. §3.3.1), when the observation operator is diagonal in the spectral basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$, is considered first. Then, we consider the state estimation using the same filtering algorithms in the realistic configuration with aliased observations (cf. §3.3.2) so that the observation operator non-diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$, as described in §3.3. When filtering with aliased observations two cases of interest are considered:

- **No superresolution**. Filtering only the modes resolved by the sparse aliased observations ($N{=}M$ with H in (39) non-diagonal in the spectral basis.)
- **Supperresolution**. Filtering with forward models with spectral resolution higher than that of the aliased observations ($N = PM$, $P \in \mathbb{N}^+$ with H in (39) non-diagonal in the spectral basis). As described in §3.3 and [9] this procedure allows, in principle, to estimate the dynamics of the truth modes which are not resolved by the observations.

The main results are highlighted and summarised as appropriate in the following subsections.

We consider two types of space-time measures to assess the performance of the mean filter estimates. Denote the (conditional) mean estimate obtained from an approximate Gaussian filter resolving N spectral modes by $m_N(x,t)$ and the truth by $u(x,t)$. Then, the respective measures are defined as follows:

(i) The root-mean-square-error (RMS) given by the L^2 norms of the residual $u-m_N$ in the spaces $\mathcal{H}_\Lambda \times \mathcal{I}$ and $\mathcal{H}_N \times \mathcal{I}$, which are defined, respectively as

$$\text{(53)} \qquad \text{RMS}(u,m_N) = \|u-m_N\|_{L^2(\mathcal{H}_\Lambda\times\mathcal{I})} := \Lambda^{-1}|\mathcal{I}|^{-1/2}\left(\sum_{i\in\mathcal{I}}\sum_{j,k=-\Lambda}^{\Lambda}\big(u(x_{j,k},t_i)-m_N(x_{j,k},t_i)\big)^2\right)^{1/2},$$

where $m_N \in \mathcal{H}_N \subseteq \mathcal{H}_\Lambda \subseteq \mathcal{H}$ is naturally embedded into $\mathcal{H}_\Lambda$, and

$$\text{(54)} \quad \text{RMS}_N(u,m_N) = \|P_N u-m_N\|_{L^2(\mathcal{H}_N\times\mathcal{I})} := N^{-1}|\mathcal{I}|^{-1/2}\left(\sum_{i\in\mathcal{I}}\sum_{j,k=-N}^{N}\big(u(x_{j,k},t_i)-m_N(x_{j,k},t_i)\big)^2\right)^{1/2},$$

where P_N is the orthogonal projection onto $\mathcal{H}_N \subseteq \mathcal{H}$.

(ii) Pattern correlation $0 \leqslant \text{XC} \leqslant 1$ defined via the inner product in the spaces $\mathcal{H}_\Lambda \times \mathcal{I}$ and $\mathcal{H}_N \times \mathcal{I}$. These measures are defined, respectively, as

$$\text{(55)} \qquad \text{XC}(u,m_N) := \frac{\langle u, m_N\rangle_{\mathcal{H}_\Lambda\times\mathcal{I}}}{\|u\|_{L^2(\mathcal{H}_\Lambda\times\mathcal{I})}\|m_N\|_{L^2(\mathcal{H}_\Lambda\times\mathcal{I})}} \propto \sum_{i\in\mathcal{I}}\sum_{j,k=-\Lambda}^{\Lambda} u(x_{j,k},t_i)\, m_N(x_{j,k},t_i),$$

and

$$\text{(56)} \qquad \text{XC}_N(u,m_N) := \frac{\langle P_N u, m_N\rangle_{\mathcal{H}_N\times\mathcal{I}}}{\|P_N\, u\|_{L^2(\mathcal{H}_N\times\mathcal{I})}\|m_N\|_{L^2(\mathcal{H}_N\times\mathcal{I})}} \propto \sum_{i\in\mathcal{I}}\sum_{j,k=-N}^{N} u(x_{j,k},t_i)\, m_N(x_{j,k},t_i).$$

Clearly, $\text{RMS}(u,m_N)$ and $\text{XC}(u,m_N)$ quantify the error in the filter estimates relative to the truth solution, while $\text{RMS}_N(u,m_N)$ and $\text{XC}_N(u,m_N)$ quantify the filter performance on the N modes resolved by the forward model. (See §3.1, 3.2 for the definitions of $\mathcal{H}_\Lambda$ and $\mathcal{H}_N$; the 'synthetic' truth should be denoted as u_Λ but we simplify the notation.)

In order to assure a consistent comparison, the algorithms are tuned using the same data obtained from long runs of the simulated truth dynamics (31). In the context of 3DVAR (cf. §4.1), and in line with [46, 7], the tuning entails estimating the 'background' covariance C_0 in (40) which is taken to be diagonal in the spectral basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and estimated as described below. It is important to stress that 3DVAR needs an additional tuning step represented by the inflation of the background covariance in order to prevent filter divergence. On the other hand, the tuning of SPEKF filters requires setting values of the free parameters in the forward model (51) which are roughly estimated from the equilibrium statistics as in [8, 9]; the performance of the SPEKF filters turns out to be not very sensitive to the choice of the tuning parameters, and only the parameters in the equations for the spectral modes $\{u_k\}_{0<|k_{1,2}|\leqslant N}$ resolved by the forward model need to be estimated directly from the data (see [24, 23, 8, 9]). As shown in Figure 9, a satisfactory accuracy is reached relatively quickly in terms of the length of the 'training' time interval; note, however, that the results for 3DVAR are shown for the optimal choice of the multiplicative inflation parameter β in (40) which requires more than just the estimates of statistics from the training data. The well-known importance of covariance inflation in 3DVAR is illustrated in the subsequent results. The first step of tuning procedure is similar for both 3DVAR and SPEKF filters and it utilises an Ornstein-Uhlenbeck (OU) process as a model for the dynamics of the modes $u_k(t)$ in the solution of the forward map; details of this procedure are outlined in Appendix A.

5.1. **Filtering with non-aliased observations.** In this idealised configuration we assume that noisy observations of individual modes are available, as described in §3.3.1, which implies that the observation operator H in (35) is diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$. The filtering algorithms 3DVAR (cf. §4.1), and cSPEKF, GCF in §4.2 utilise forward models with spectral resolution N, given the spectral resolution of

the non-aliased observations $M \leqslant N$. Consequently, the resolution of the forward models in the spatial domain is $(2N+1)\times(2N+1)$, and the spatial resolution of observations is $(2M+1)\times(2M+1)$. The data assimilation time interval Δt_{obs} is chosen to be about 50% of the mean decorrelation time on all resolved modes for each dynamical regime considered. Throughout this paper, we assume that the true dissipation parameters ν and κ, as well as the forcing are known. Filtering with uncertain parameters and/or forcing adds another layer of complexity to the problem and is deemed too technical for this exposition; we comment briefly on these configurations at the end of the section.

The results below are presented in terms of the vorticity field $\omega = \nabla^\perp \cdot u$, where $\nabla^\perp = (\partial_2, -\partial_1)^{\mathrm{T}}$ and $u(t,x) = \sum_{k\in\mathbb{Z}^2\setminus\{0\}} u_k(t)\psi_k(x)$ solves (1)-(3). The spectral representation of the vorticity field in terms of $\{u_k(t)\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ is given by

$$\omega(t,x) = \sum_k u_k(t)\left(\nabla^\perp\cdot\psi_k(x)\right) = \sum_k \hat{\omega}_k(t)\,\phi_k(x), \qquad \hat{\omega}_{-k} = \hat{\omega}_k^*, \tag{57}$$

where $\hat{\omega}_k(t) = (2\pi i/L)|k|\,u_k(t)$ represent the coefficients of $\omega(t,x)$ in the Fourier basis $\{\phi_k(x)\}_{k\in\mathbb{Z}^2\setminus\{0\}}$, with $\phi_k(x) = |k|^{-1}\big(k^\perp\cdot\psi_k(x)\big)$. The coefficients $\{u_k\}_{|k_{1,2}|\leqslant N}$ are estimated from the filtering algorithms 3DVAR (§4.1) and SPEKF (§4.2) in the three dynamical regimes of (8) illustrated in Figure 3, given noisy information about the evolution of the first M non-aliased spectral modes. As a reference, the quality of the filtering estimates is compared against two different estimates obtained purely from the observations:

(i) Estimates based on observations of all (non-aliased) modes of the truth. In this case the error between the truth and the observations is assessed in the space $\mathcal{H}_\Lambda$ based on $\textsc{rms}\big(u, y^M\big)$ and $\textsc{xc}\big(u\,, y^M\big)$ defined, respectively, in (53) and (55).

(ii) Estimates based on observations of N (non-aliased) modes resolved by the forward model. In this case the error between the truth and estimates from observations is assessed in the space $\mathcal{H}_N$ in terms of $\textsc{rms}_N(u,y)$ and $\textsc{xc}_N(u,y^M)$ in (54) and (56), where $N = M$ in the present setting.

Note that the above measures of *observation error* focus on different performance aspects and have to be considered appropriately to the specific goal. The observation error based on the measures in (i) indicates the ability to reconstruct the truth from observations resolving M noisy modes of the truth state. This formulation provides a benchmark for assessing the quality of estimating the truth state from filtering algorithms; the corresponding error in the filtering estimates, $\textsc{rms}\big(u, m_N\big)$ and $\textsc{xc}\big(u\,, m_N\big)$, has to be smaller than the observation error in (i) for the filtering to be beneficial. The observation error based on the measures in (ii) above provides a benchmark for recovering the first N modes of the truth state from the noisy observations; the corresponding error in the filtering estimates, $\textsc{rms}_N\big(u, m_N\big)$ and $\textsc{xc}_N\big(u\,, m_N\big)$, has to be smaller than the observation error in (ii) for the filtering to be beneficial in this context. We illustrate the results on a number of numerical tests below.

Figures 4, 5 show the RMS and the correlation XC measures for filtering the attractor dynamics of (8) with different resolutions, N, of the forward models in three distinct dynamical regimes of the dynamics (8) which are illustrated in Figure 3; in all cases the variance of the observation noise in the spatial domain is $\Gamma^0 = 0.15E$ where E is the energy (i.e., L^2 norm) of the solutions on the attractor. The results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40) which is needed for the stability of the algorithm, as discussed in §4.1; this filter diverges for sufficiently small values of β but this effect is not resolved in detail. Additive covariance inflation obtained by varying α in $\widehat{C}_{\alpha,\beta}$ (40) has a much less pronounced effect and is not shown ($\alpha = 0$ is used in all examples shown). The black solid lines indicate the quality of estimates of the truth obtained from observations of its N spectral modes, while the black dotted lines indicate the quality of estimates obtained without filtering directly from the noisy

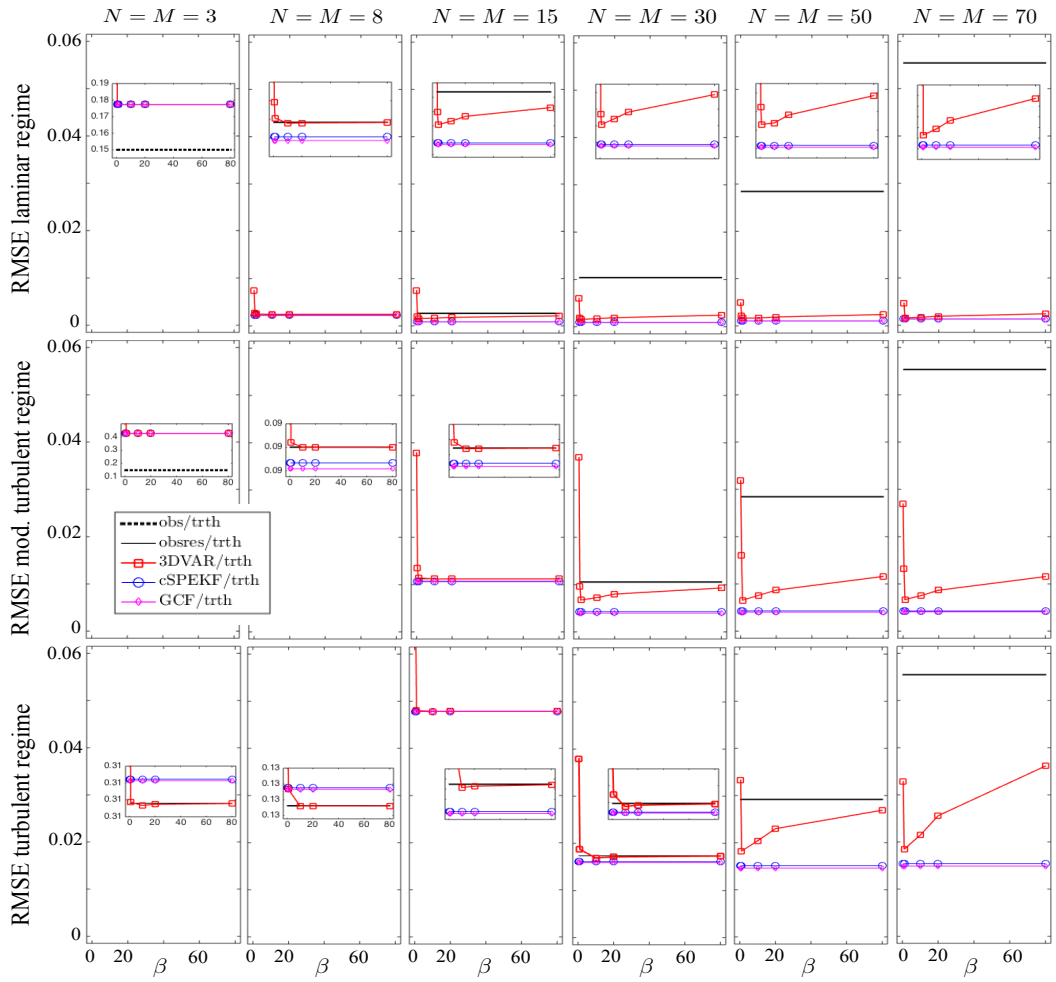

FIGURE 4. NON-ALIASED OBSERVATIONS OF (8). Comparison of performance of the filtering algorithms described in §4 in different dynamical regimes of (8) (cf Figure 3) in terms of the RMS error (53) in the mean estimates for different resolutions N of the forward models and the observation error $\varepsilon = 0.15E$ (here E is the energy per mode in steady state, and $N = M$ as the modes resolved by the forward models are assumed to be observable directly in the spectral domain). Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40); this filter diverges for sufficiently small values of β but this effect is not resolved in detail. Additive covariance inflation obtained by varying α in $\widehat{C}_{\alpha,\beta}$ (40) has a much less pronounced effect and is not shown. See figure 5 for a comparison of in terms of the XC measures (55).

observations of all spectral modes of the truth (see (i) above); the latter case represents the gold standard for pure observation-based estimates given that noisy information about all truth modes is utilised.

Figures 6, 7, 8 show snapshots of the true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1) and cSPEKF (§4.2) and the corresponding spatially resolved residuals between the mean estimates and fully resolved truth. The estimated signal in the spatial domain is recovered from

$$\omega_N(t,x) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\left(\nabla^\perp \cdot \psi_k(x)\right), \tag{58}$$

while the vorticity field corresponding to the synthetic truth solving (31) is given by (58) with $N = \Lambda$. Results are shown for two spectral resolutions N of the forward models in the algorithms with a fully

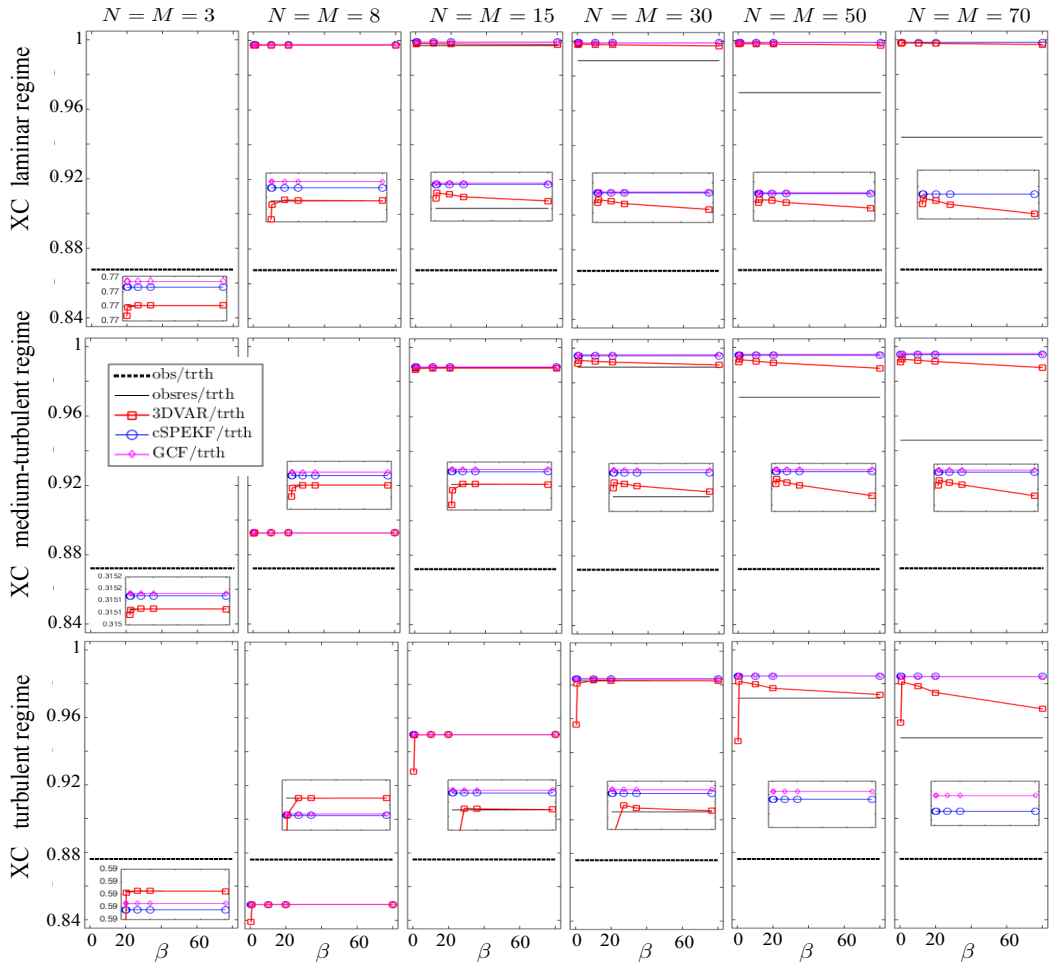

FIGURE 5. NON-ALIASED OBSERVATIONS OF (8). Comparison of performance of the filtering algorithms of §4 in different dynamical regimes of (8) (cf Figure 3) in terms of the error in the mean estimates, using XC measures (55) for different resolutions N of the forward models, and the observation error $\varepsilon = 0.15E$; results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40); see figure 4 for the comparison in terms of the RMS measures and more information.

observed state, $M = N$, in cSPEKF, GCF, and 3DVAR. The colorscale is the same for all examples which enables an easy visual comparison of the estimation errors for different cases.

Finally, figure 9 shows a comparison of performance of cSPEKF, GCF, and 3DVAR in different dynamical regimes of (8) (cf Figure 3) as a function of the length of the training data used to fix the free parameters in cSPEKF/GCF algorithms (§4.2) and to estimate the background covariance in 3DVAR (§4.1); see A for more details. The performance of the filtering algorithms is assessed for the mean estimates and expressed in terms of the RMS error (53) and pattern correlation, XC (55), for non-aliased observations and the resolution $N = 70$ of the forward models in the filtering algorithms. and the observation error $\varepsilon = 0.15E$ where E is the energy per mode in steady state. $N = M$ as the modes resolved by the forward models are assumed to be observable directly in the spectral domain. The total length of the training time interval consists of 4000 simulation time steps which correspond to: (i) $\sim$ 220 mean decorrelation time units in the laminar regime, (ii) $\sim$ 560 mean decorrelation time units in the moderately turbulent regime, and

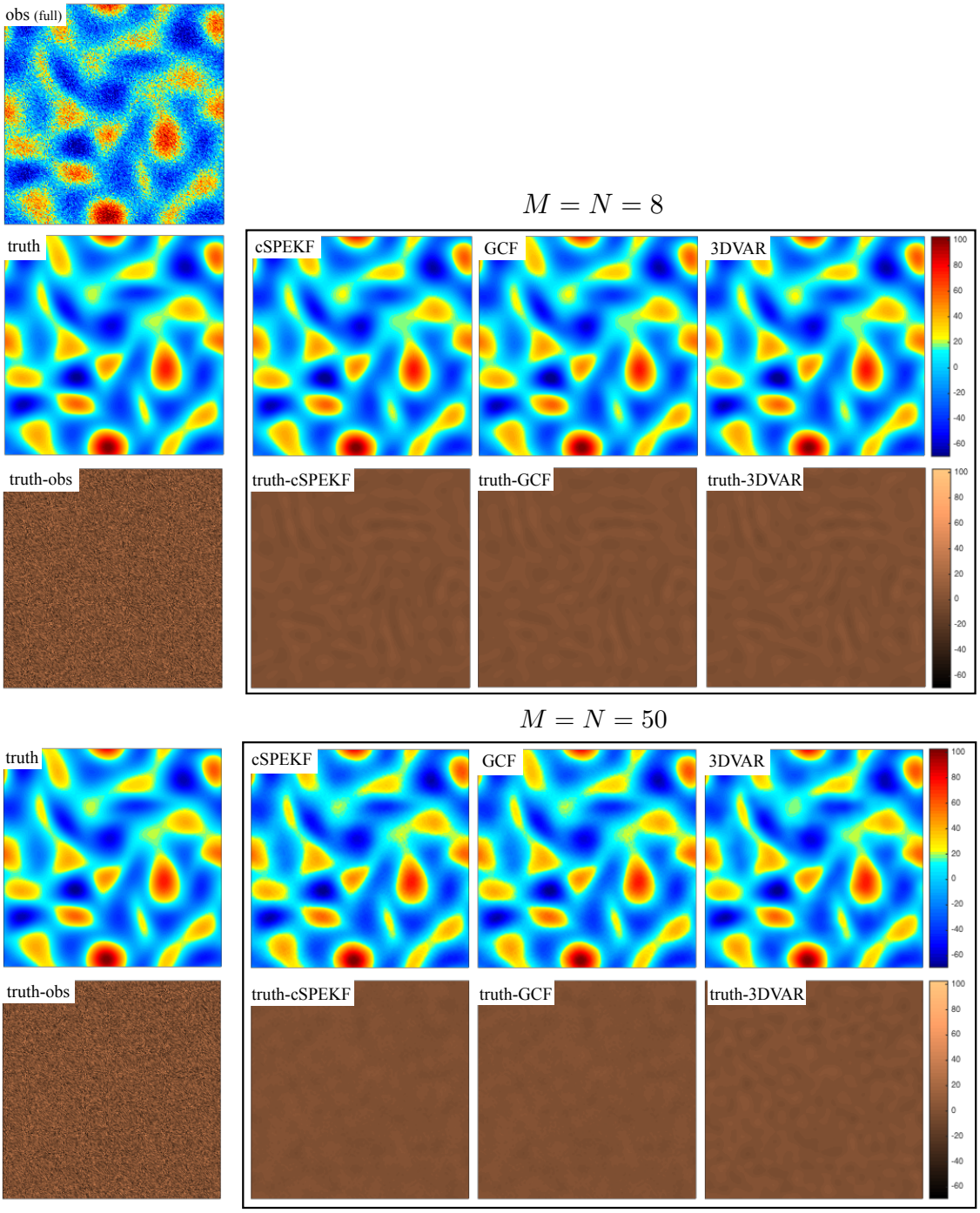

FIGURE 6. FILTERING WITH NON-ALIASED OBSERVATIONS; LAMINAR REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the fully resolved truth. Results are shown for two spectral resolutions N of the forward models in the algorithms with fully observed state, $M = N$, in the forward models of cSPEKF, GCF, and 3DVAR. Observation error $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare these results with those in Figure 4, and 5.

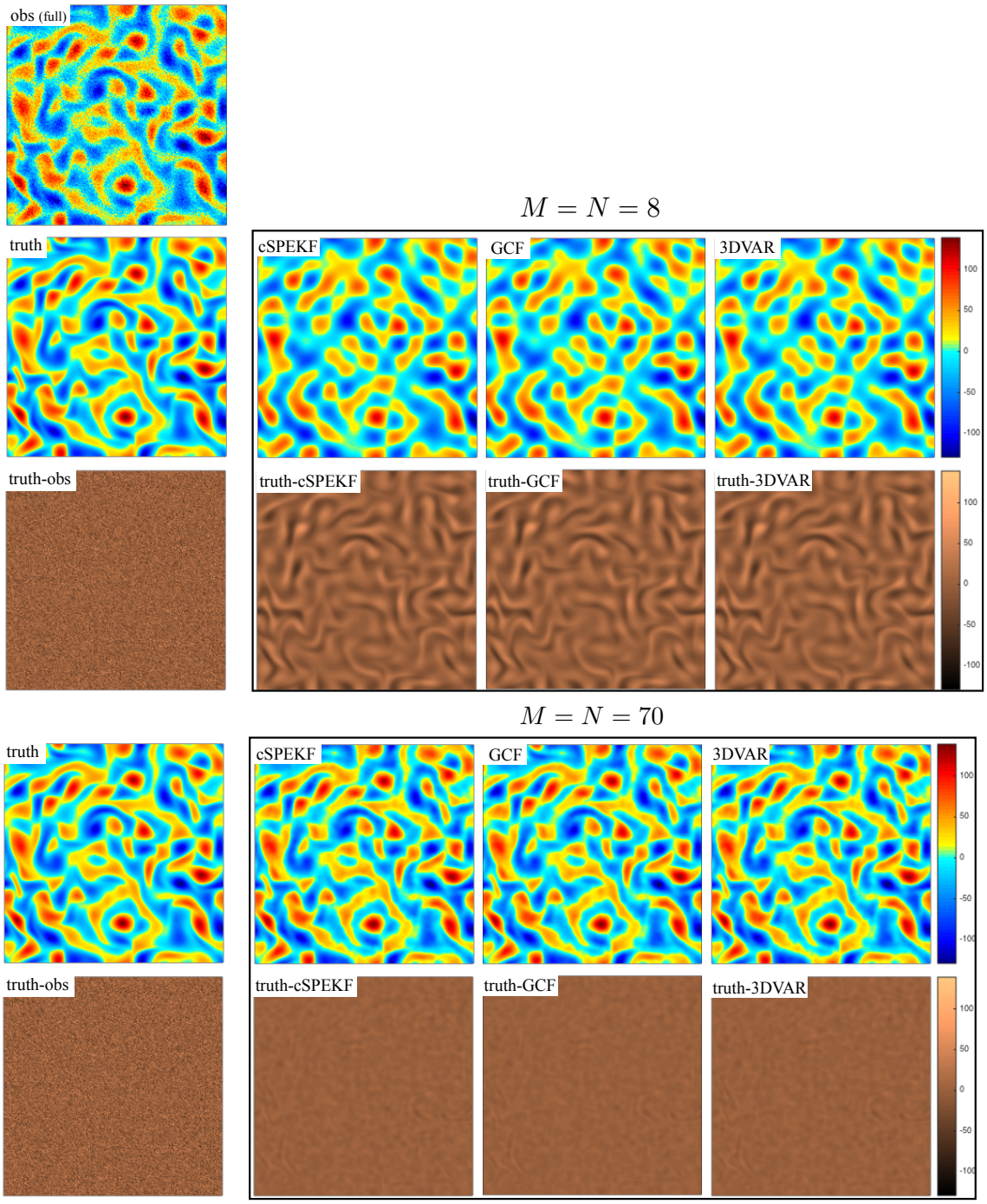

FIGURE 7. FILTERING WITH NON-ALIASED OBSERVATIONS; MODERATELY TURBULENT REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the fully resolved truth. Results are shown for two spectral resolutions N of the forward models in the algorithms with fully observed state, $M = N$, in the forward models of cSPEKF, GCF, and 3DVAR. Observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare these results with those in Figure 4, and 5, and with Figures 6, 8.

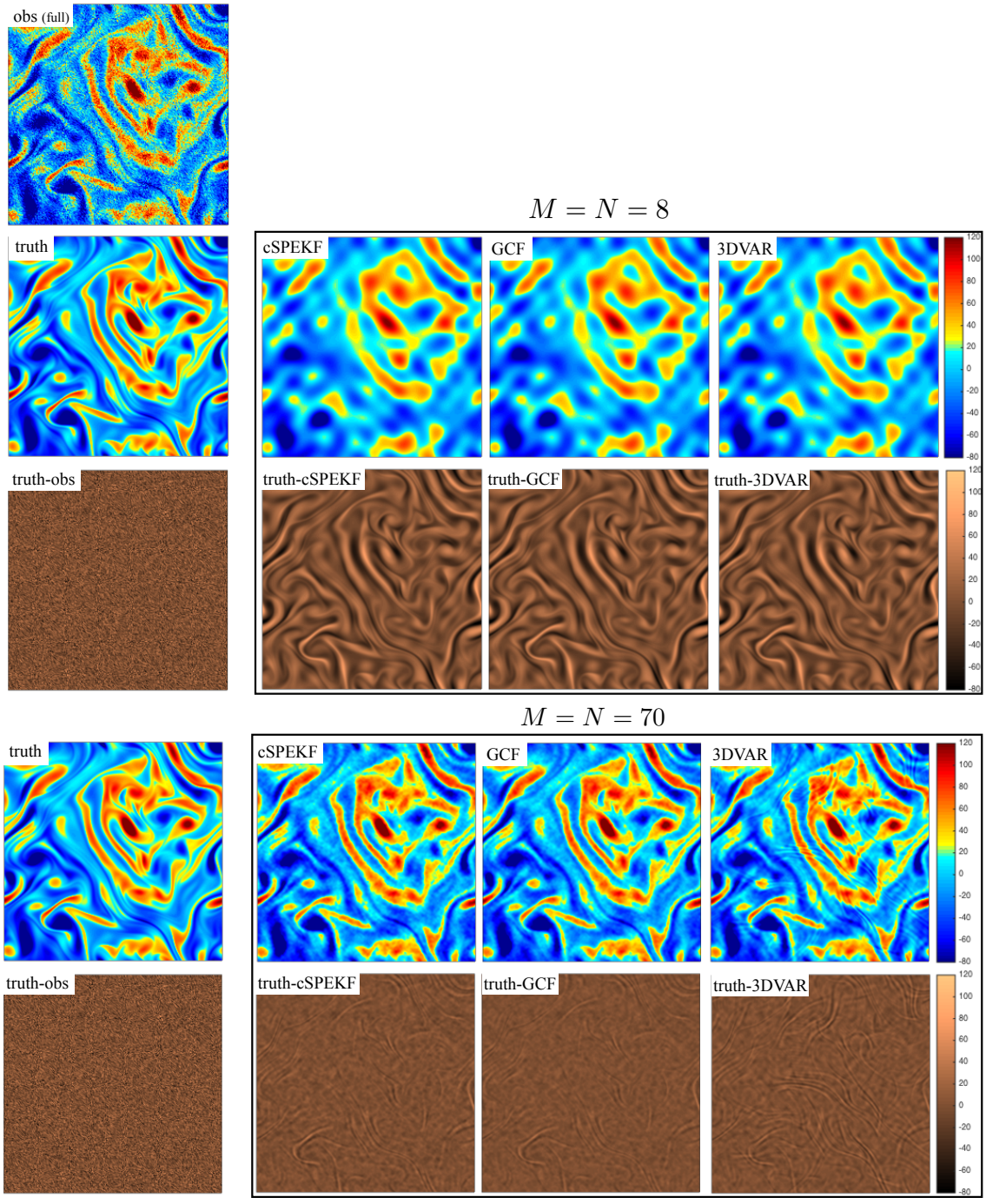

FIGURE 8. FILTERING WITH NON-ALIASED OBSERVATIONS; TURBULENT REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the fully resolved truth. Results are shown for two spectral resolutions N of the forward models in the algorithms with fully observed state, $M = N$, in the forward models of cSPEKF, GCF, and 3DVAR. Observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare with Figures 4, 5 and Figures 6, 7.

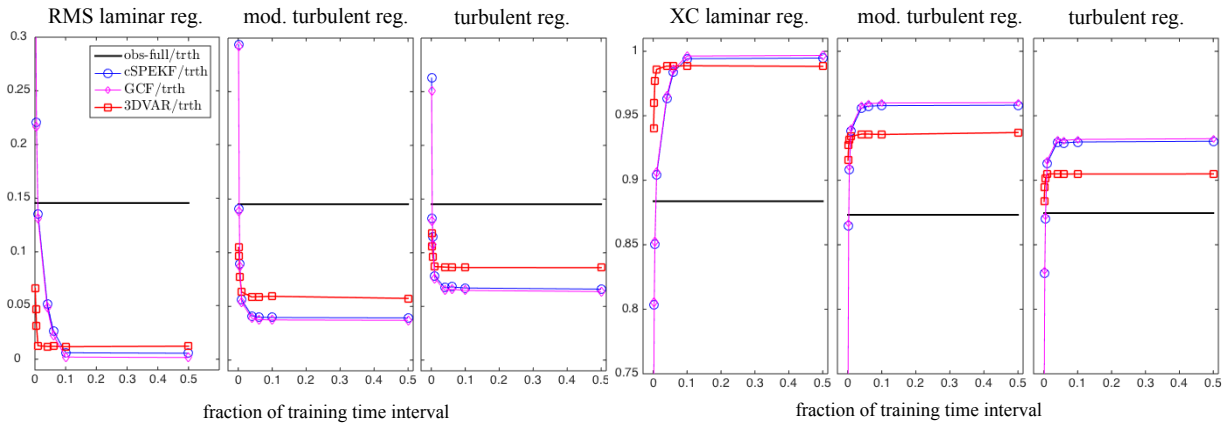

FIGURE 9. FILTERING PERFORMANCE AS A FUNCTION OF TRAINING DATA; NON-ALIASED OBSERVATIONS. Comparison of performance of the filtering algorithms described in §4 in different dynamical regimes of (8), illustrated in Figure 3, as a function of the length of the training data used to fix the tuneable parameters in the filtering algorithms; see Appendix A for more details. Results are shown for non-aliased observations ($M = N$) and the resolution $N = 70$ of the forward models in the filtering algorithms; the observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. The total length of the training time interval consists of 12000 simulation time steps which correspond to: (i) ~ 220 mean decorrelation time units in the laminar regime, (ii) ~ 560 mean decorrelation time units in the moderately turbulent regime, and (iii) ~ 750 mean decorrelation time units in the turbulent regime. Results for 3DVAR are shown for the optimal value of the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40).

(iii) ~ 750 mean decorrelation time units in the turbulent regime. Results for 3DVAR are shown for the optimal value of the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40). Additive covariance inflation obtained by varying α in $\widehat{C}_{\alpha,\beta}$ (40) has a much less pronounced effect and is not shown.

We summarise the results below:

- Both classes of algorithms provide better estimates than those obtained only from the observations of $M = N$ non-aliased modes of the truth (compare $\text{RMS}\big(u, y^N\big)$ and $\text{XC}\big(u\,, y^N\big)$ with $\text{RMS}\big(u, m_N\big)$ and $\text{XC}\big(u\,, m_N\big)$). This indicates that if one can observe M modes of the truth, filtering $N = M$ modes with the considered filtering algorithms provides better estimates than those obtained from the observations.
- Unsurprisingly, the forward models in both classes of filters have to resolve sufficiently large number of spectral models in order to outperform the quality of estimates obtained from noisy observations of all the non-aliased modes of the truth (cf. $\text{RMS}\big(u, y^\Lambda\big)$ and $\text{XC}\big(u\,, y^\Lambda\big)$ in (i) above).
- For non-aliased observations the state estimation with SPEKF algorithms §4.2 which contain model error due to significant simplifications in the forward model provides results which either outperform or shadow those obtained with 3DVAR (cf. §4.1). SPEKF algorithms diverge when the forward model has a higher spectral resolution than the resolution of the observations; this is due to the model error in the dynamics (51) which is not constrained by the observations on the modes $M \leqslant |k_{1,2}| \leqslant N$.
- In the laminar regime when a small number of modes contains significant energy (see Figure 3) filtering with relatively low resolution provides estimates which are more accurate than those obtained from the **full** noisy observations of all the spectral modes resolved by the truth dynamics (31). SPEKF estimation provides better results than 3DVAR with optimally inflated covariance.

- In the turbulent regime (and in the moderately turbulent regime not shown) a sufficiently large number of modes needs to be observed and resolved by the forward models in order to provide good estimates of the system state to beat the fully resolved observations. In these regimes there is a wide range of modes with significant energy and the number of the resolved modes has to be larger then the forcing scale for good estimation relative to the estimates obtained from the noisy observations of all the spectral modes resolved by the truth dynamics (31). For sufficiently low resolution of the forward dynamics SPEKF and GCF estimation provides is marginally worse than 3DVAR with optimally inflated covariance but the differences are negligible.
- For non-aliased observations the quality of the filtering estimates in terms of their ability to estimate the first N modes of the truth in terms of $\textsc{rms}_N\big(u, m_N\big)$ and $\textsc{xc}_N\big(u\,, m_N\big)$ are qualitatively similar to the case of estimating the truth in terms of $\textsc{rms}\big(u, m_N\big)$ and $\textsc{xc}\big(u\,, m_N\big)$. Both classes of filters beat the corresponding estimates obtained from observations (in terms of $\textsc{rms}_N\big(u, y_N\big)$ and $\textsc{xc}_N\big(u\,, y_N\big)$ described in (ii) above), and the quality of the estimates improves with the number of filtered modes N provided that the model state is fully observed $M = N$. Distinction between these metrics becomes relevant for aliased observations discussed in the next section.
- Considering the performance of filtering the dynamics (8) with incorrect dissipation and forcing parameters requires a great deal of additional numerical tests and is not discussed here in detail. However, the results are largely in line with those presented here. In particular, SPEKF type filters are significantly less affected by the incorrect forcing than 3DVAR. The performance of 3DVAR with incorrect parameters and forcing can be improved and stabilised by appropriate covariance inflation but, even for optimally chosen inflation, this filter remains inferior to SPEKF/GCF filters.

5.2. **Filtering with aliased observations.** In this more realistic configuration we consider the state estimation with the filtering algorithms 3DVAR (§4.1) and SPEKF (§4.2) given noisy observations of the state $u(x,t)$ in (39) on a $(2M{+}1){\times}(2M{+}1)$ grid in the spatial domain; consequently, these observations alias the modes of the truth with u_k, $|k_{1,2}| > M$ into the modes resolved by the observations with $u_k, 0 < |k_{1,2}| \leqslant M$. As discussed in §3.3.2, this implies that the observation operator H in (39) is not diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and the information about the modes resolved by the forward model in the filtering algorithms is corrupted by both the observation noise and the aliased modes. Similar to the configuration with non-aliased observations in §5.1, the data assimilation time interval Δt_{obs} is chosen to be about 50% of the mean decorrelation time on all resolved modes for each dynamical regime considered; moreover we consider filtering with correct dissipation parameters ν and κ, as well as the correct forcing. We comment briefly on filtering with uncertain parameters and/or forcing at the end of the section.

In the tests considered in this section we assume that the spectral resolution of the observation is fixed with $M = 10$ and the filtering algorithms 3DVAR §4.1, and cSPEKF, GCF in §4.2 are considered at different spectral resolutions N with the resolution of the observations $N = PM$, $P \in \mathbb{N}^+$. Apart from the main question concerned with the ability of the filters to estimate the spatially extended truth, it is also interesting to investigate if the superresolution (i.e., $P > 1$ in the forward models) helps improve the estimates of the dynamics of the primary modes. Note that superresolution corresponds to filtering with $N > M$ which caused poor performance of the filters in the case of non-aliased observations in §5.1 - especially the SPEKF filters which contain a significant model error which was very detrimental for the estimates of the unobserved modes with $|k_{1,2}| > M$. The case of aliased observations ($H^{\mathcal{A}\{\textsc{m}\}}$ in (39) not diagonal) is different the since the information about the unobserved modes is not lost but is instead aliased into the observed modes; the consequences of this fact on the filtering are illustrated and discussed below.

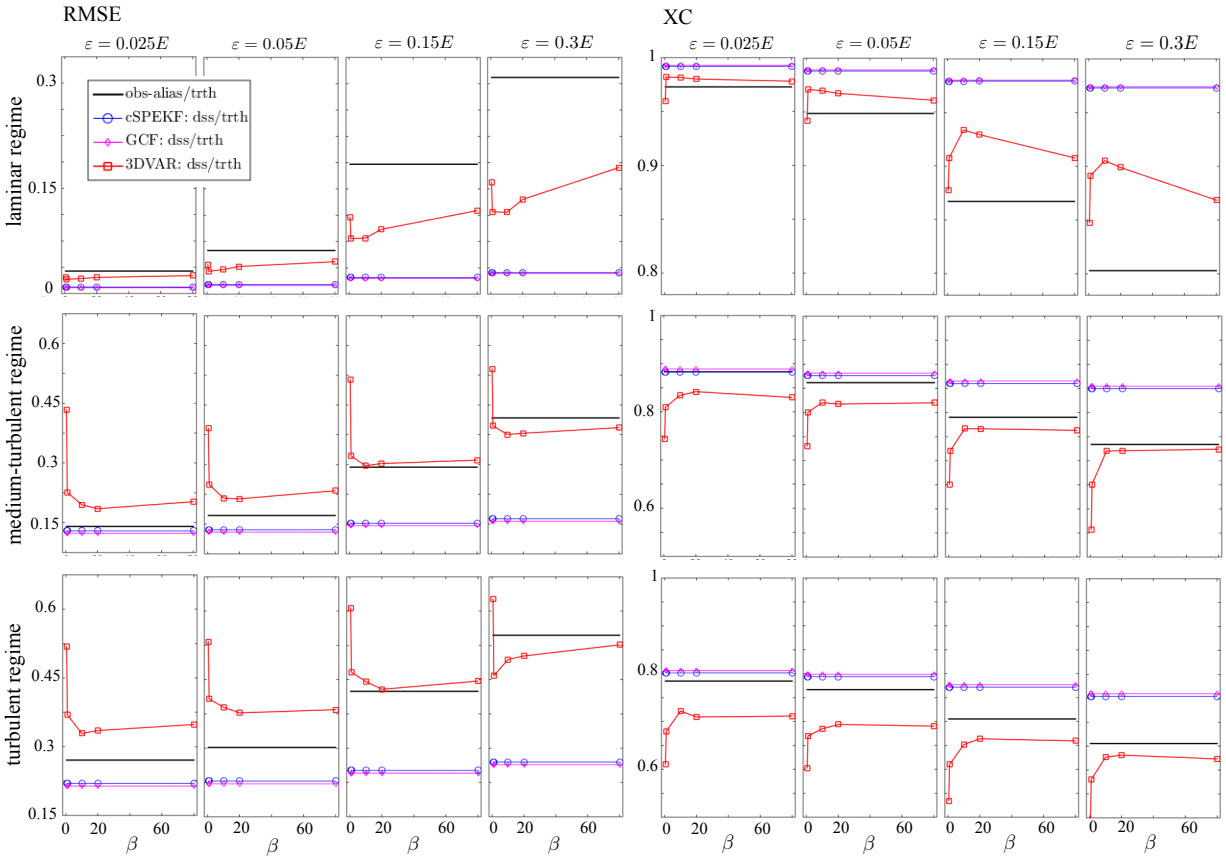

FIGURE 10. FILTERING WITH ALIASED OBSERVATIONS OF (8). Comparison of performance of the filtering algorithms of §4 in different dynamical regimes of (8) (cf Figure 3) in terms of the error in the mean estimates, using RMS (53) and XC measures (55) for different resolutions N of the forward models; here $M = 10$ and $P = 3$ (see §5.2, and §3.3.2). Observation error is indicated in terms of E - the energy per mode in steady state. Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40); dependence on the additive inflation parameter in $\widehat{C}_{0,\beta}$ is much less pronounced and not shown. 3dVAR diverges for sufficiently small values of β but this effect is not resolved in detail.

The results below are presented in terms of the vorticity field (58) and the coefficients $\{u_k\}_{|k_{1,2}|\leqslant N}$ are estimated from the filtering algorithms 3DVAR (§4.1) and SPEKF (§4.2) in the three dynamical regimes of (8) illustrated in Figure 3, given noisy aliased observations of the truth. As a reference, the quality of the filtering estimates is compared against two different estimates obtained purely from the observations:

(iii) Estimates based on observations of all aliased modes of the truth $y^{\mathcal{A}\{\mathrm{M}\}}$ in (39). In this case the error between the truth and the observations is assessed in the space $\mathcal{H}_\Lambda$ based on $\mathrm{RMS}\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$ and $\mathrm{XC}\big(u\,, y^{\mathcal{A}\{\mathrm{M}\}}\big)$ defined, respectively, in (53) and (55).

(iv) Estimates based on the M primary modes resolved by the observations. In this case the error between the truth and estimates from observations is assessed in the space $\mathcal{H}_M$ in terms of $\mathrm{RMS}_M\left(u, y^{\mathcal{A}\{\mathrm{M}\}}\right)$ and $\mathrm{XC}_M\left(u, y^{\mathcal{A}\{\mathrm{M}\}}\right)$ in (54) and (56).

In addition, we compare the filtering results with the estimates based on observations of all non-aliased modes of the truth, as described in (i) in §5.1. The observation error based on the measures in (iii) indicates the ability to reconstruct the truth from aliased observations resolving M primary modes of the truth state. This formulation provides a benchmark for assessing the quality of estimating the truth state

from filtering/superresolving algorithms; the corresponding error in the filtering estimates, $\textsc{rms}\big(u, m_N\big)$ and $\textsc{xc}\big(u\,, m_N\big)$, has to be smaller than the observation error in (iii) for the filtering to be beneficial. The observation error based on the measures in (iv) provides a benchmark for recovering the first N modes of the truth state from the aliased observations; the corresponding error in the filtering estimates, $\textsc{rms}_N\big(u, m_N\big)$ and $\textsc{xc}_N\big(u\,, m_N\big)$, has to be smaller than the observation error in (iv) for the filtering to be beneficial in this context. In particular, we use the measure in (iv) to investigate the utility of superresolution (i.e., filtering with $N = PM$, $P > 1$) for estimating the M primary modes from aliased observations compared to filtering with no superresolution (i.e., $N = M$). We illustrate the results on a number of numerical tests below.

Figure 10 shows a comparison of the filtering algorithms of §4 in terms of the error in the mean estimates, using $\textsc{rms}\big(u, m_N\big)$ in (53) and $\textsc{xc}\big(u\,, m_N\big)$ in (55) for aliased observations with $M = 10$ with different observation noise and the resolution of the forward models $N = 3M$. Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40) which is indicated on the horizontal axes in the insets. The black dotted lines indicate the quality of estimates obtained directly from the aliased noisy observations of the truth (without filtering; see (iii) above), the black dash-dotted lines indicate the quality of estimates of the truth obtained from observations of M primary modes of the truth, while the error obtained from non-aliased observations of all the truth modes is indicated above the insets in terms of the energy of the truth solution. The errors based on the non-aliased observations of the truth - $\textsc{rms}\big(u, y^M\big)$, $\textsc{xc}\big(u, y^M\big)$ with y^M in (35) - represent the gold (though unachievable) reference standard, while the observation errors based on the aliased observations of the truth - $\textsc{rms}\big(u, y^{\mathcal{A}\{\textsc{m}\}}\big)$, $\textsc{xc}\big(u, y^{\mathcal{A}\{\textsc{m}\}}\big)$ - provide a more realistic target against which to compare the performance of various filters. The performance the considered filtering algorithms can be inferred from the curves described in the legend.

Figures 11, 12 and 13 show snapshots of the true, observed, and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF algorithms (§4.2) and the corresponding spatially resolved RMS errors between the mean estimates and the fully resolved truth. Results are shown for the laminar and fully turbulent regimes (see Figure 3) for the spectral resolution $N = 31$ of the forward models; the resolution of the aliased observations is $M = 10$.

Finally, figure 14 shows a comparison of the quality of the filtering algorithms of §4 for estimating the M primary modes using superresolving algorithms ($N > M$) and non-superresolving algorithms ($N = M$); the comparison is carried out in terms of the error in the mean estimates (iv), using $\textsc{rms}_M\big(u, m_N\big)$ in (54) and $\textsc{xc}_M\big(u\,, m_N\big)$ in (56). Aliased observations of the truth dynamics (31) in the fully turbulent regime are used with $M = 10$ at different levels of the observation noise; the resolution of the forward models in the superresolving mode is $N = 3M$ and in the non-superresolving mode $N = M$. Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40) which is indicated on the horizontal axes in the insets. The black dotted lines indicate the quality of estimates $\textsc{rms}_M\big(u, y^{\mathcal{A}\{\textsc{m}\}}\big)$, $\textsc{xc}_M\big(u, y^{\mathcal{A}\{\textsc{m}\}}\big)$ obtained directly from the aliased noisy observations of the truth (without filtering; see (iv) above), while the error obtained from non-aliased observations of all the truth modes is indicated above the insets in terms of the energy of the truth solution. The performance the considered filtering algorithms can be inferred from the curves described in the legend.

We summarise the results below:

- When filtering with aliased observations the superresolving SPEKF and GCF algorithms significantly outperform the superresolving 3DVAR algorithm. This is most pronounced in the fully turbulent regime (see Figures 10, 13) but is also present to a lesser extent in other dynamical regimes. Moreover, SPEKF/GCF algorithms do not require the covariance inflation step.

- Superresolution of aliased observations (i.e., $N > M$) improves the estimates of the M primary modes of the truth dynamics when filtering with SPEKF and GCF but it is detrimental on the estimates with the 3DVAR filter (Figure 14).
- For aliased observations the state estimation with SPEKF algorithms §4.2 benefits from the model error in the prediction step which avoids model rigidity present in the 3DVAR predictions based on the truncated dynamics (33). In contrast to the filtering with non-aliased observations when these filters diverged for $N > M$, the aliased observations retain the aliased information about the unobserved modes $|k_{1,2}| > M$ which is then propagated into the estimates of the corresponding modes in the superresolving filters due to the fact that $H^{\mathcal{A}\{\mathrm{M}\}}$ in (39) is not diagonal and, consequently, the gain (28) is non-zero on the unobserved spectral modes (unlike the gain (44) for the non-aliased observations).
- Superresolution with SPEKF and GCF algorithms provides similar results over a wide range of assimilation times. This fact seems to be a consequence of the distribution of decorrelation times across the modes in the dynamics. The small scale modes $|k_{1,2}| \gg 1$ decorrelate very fast compared to practically conciveable assimilation times, and the estimation error in these modes dominates the overall results of the RMS and XC in the spatial domain.
- For aliased observations the quality of the filtering estimates in terms of their ability to estimate the M primary modes of the truth dynamics in terms of $\mathrm{RMS}_N\big(u, m_N\big)$ and $\mathrm{XC}_N\big(u\,, m_N\big)$ are qualitatively similar to the case of estimating the truth in terms of $\mathrm{RMS}\big(u, m_N\big)$ and $\mathrm{XC}\big(u\,, m_N\big)$. Both classes of filters beat the corresponding estimates obtained from observations (in terms of $\mathrm{RMS}_M\left(u, y^{\mathcal{A}\{\mathrm{M}\}}\right)$ and $\mathrm{XC}_M\left(u\,, y^{\mathcal{A}\{\mathrm{M}\}}\right)$ described in (ii) above), and the quality of the estimates improves with the number of filtered modes N provided that the model state is fully observed $M = N$. The estimation with superresolving SPEKF filters is comparable to 3DVAR in the laminar regime and significantly better than 3DVAR estimates in the turbulent regime; these results are not shown due to their similarity with the above results.
- The performance of 3DVAR, and the SPEKF algorithms for filtering the dynamics (8) with aliased observations and incorrect dissipation and forcing parameters is not discussed here in detail as it requires a great deal of additional numerical tests. However, based on our numerical evidence, the results in this more general setting are largely in line with those presented here. In particular, SPEKF type filters are significantly less affected by the incorrect forcing than 3DVAR with optimal covariance inflation.

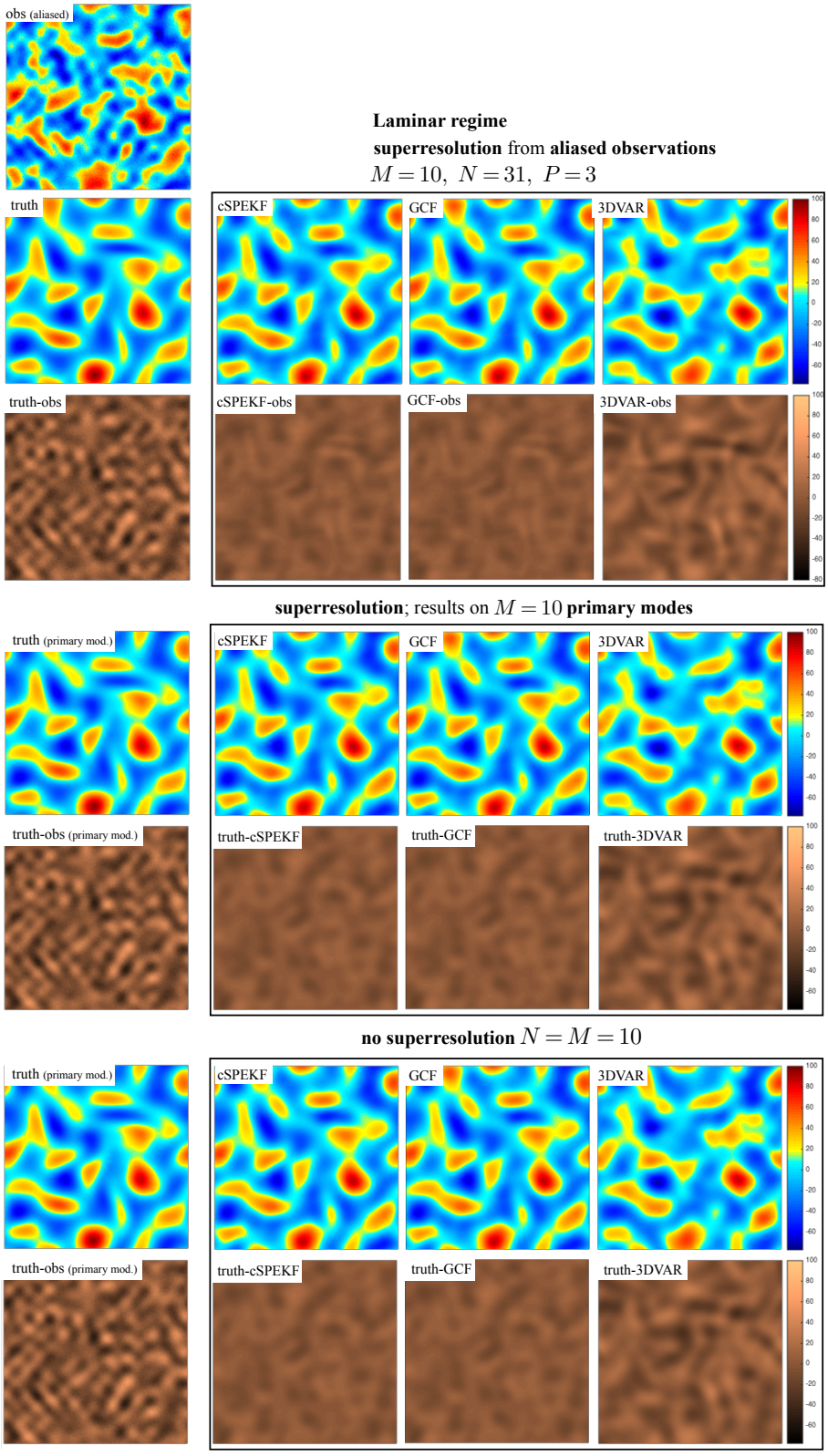

FIGURE 11. FILTERING WITH ALIASED OBSERVATIONS; LAMINAR REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the truth (full or primary modes). Results are shown for filtering with superresolving algorithms ($N > M$) and in the absence of superresolution $M = N$ in the forward dynamics of cSPEKF, GCF, and 3DVAR. Observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare with Figure 10 and see §5.2 for more information.

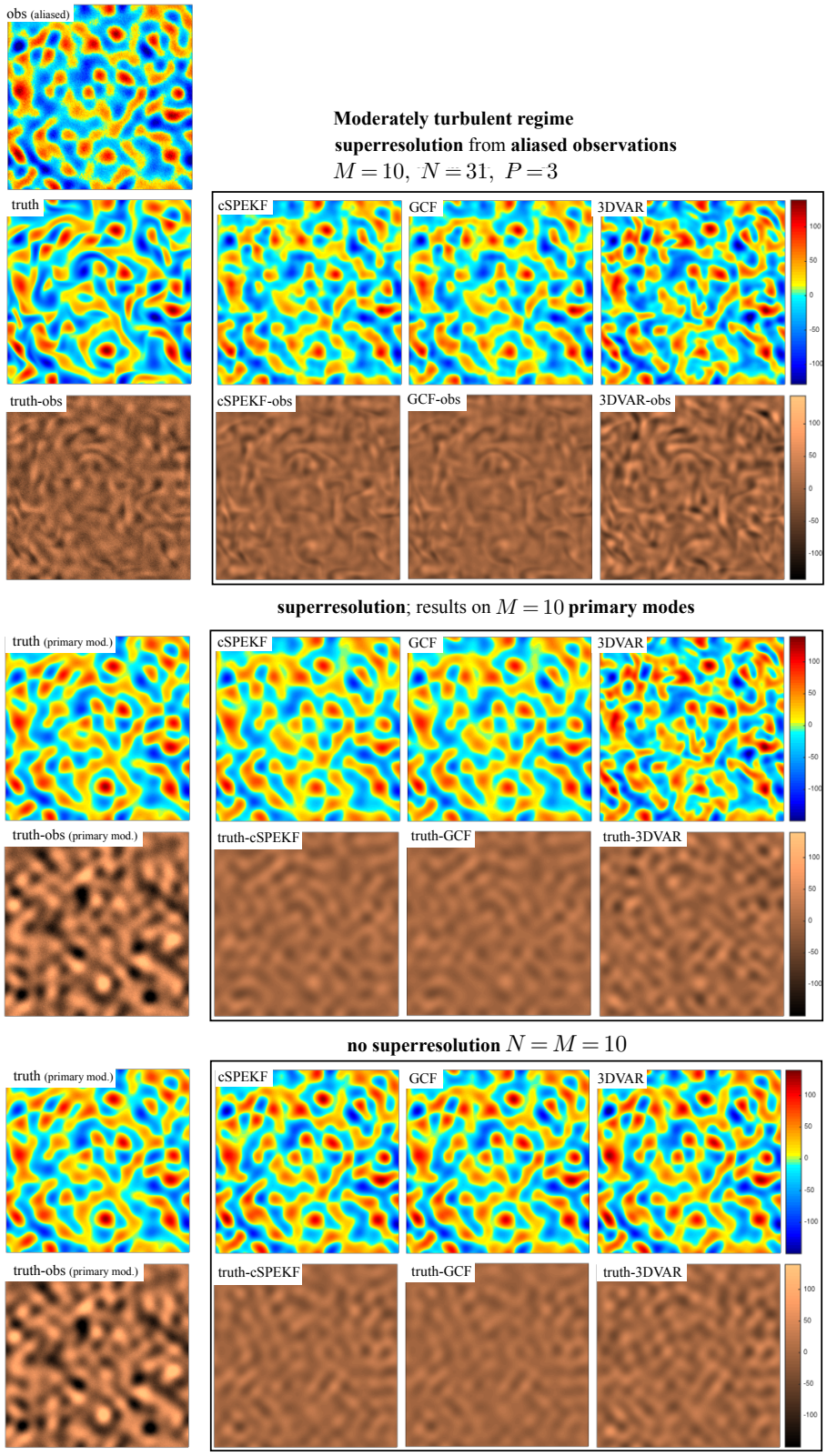

FIGURE 12. FILTERING WITH ALIASED OBSERVATIONS; MODERATELY TURBULENT REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the truth (full or primary modes). Results are shown for filtering with superresolving algorithms ($N > M$) and in the absence of superresolution $M = N$ in the forward dynamics of cSPEKF, GCF, and 3DVAR. Observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare with Figure 10 and see §5.2 for more information.

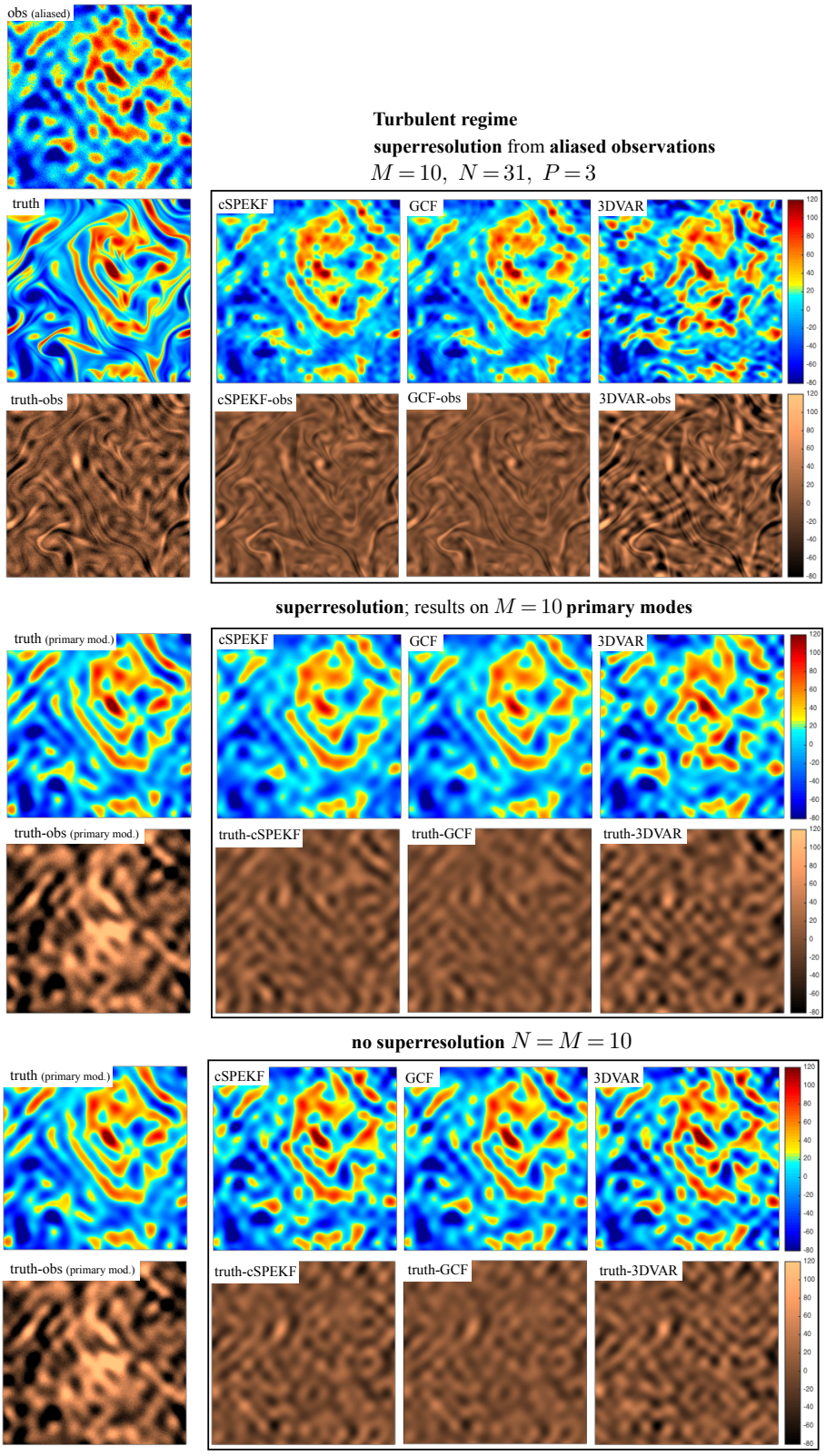

FIGURE 13. FILTERING WITH ALIASED OBSERVATIONS; TURBULENT REGIME OF (8) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the truth (full or primary modes). Results are shown for filtering with superresolving algorithms ($N > M$) and in the absence of superresolution $M = N$ in the forward dynamics of cSPEKF, GCF, and 3DVAR. Observation error is $\varepsilon = 0.15E$ where E is the energy per mode in steady state. Compare with Figure 10 and see §5.2 for more information.

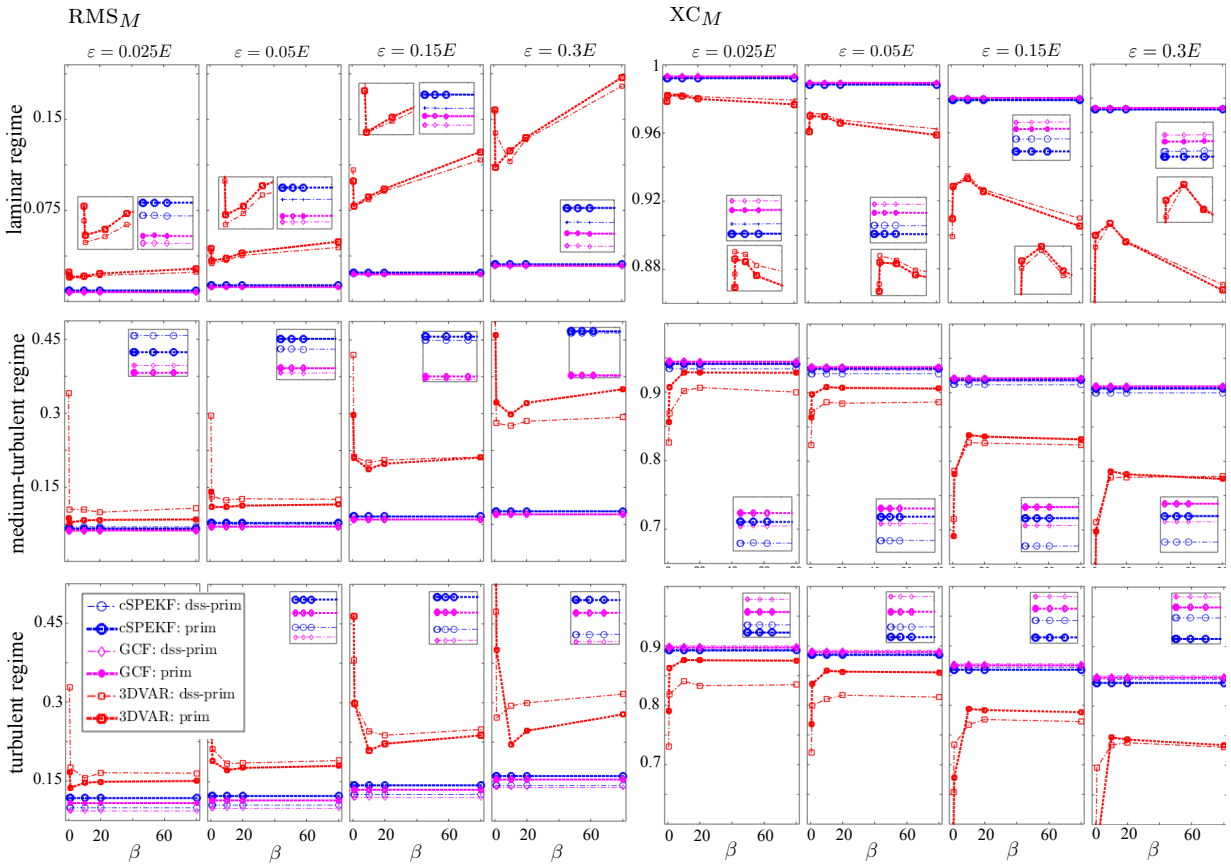

FIGURE 14. FILTERING THE DYNAMICS OF (8) WITH SUPERRESOLVING VS NOT SUPERRESOLVING ALGORITHMS FOR ALIASED OBSERVATIONS. Comparison of the quality of the filtering algorithms of §4 for estimating the M primary modes using superresolving algorithms ($N > M$) and non-superresolving algorithms ($N = M$); the comparison is carried out in terms of the error in the mean estimates using $\text{RMS}_M(u, m_N)$ in (54) and $\text{XC}_M(u, m_N)$ in (56); see (iv), §5.2 in the text. Aliased observations of the truth dynamics (31) in the fully turbulent regime are used with $M = 10$ at different levels of the observation noise; the resolution of the forward models in the superresolving mode is $N = 3M$ and in the non-superresolving mode $N = M$. Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (40).

6. Conclusions

Data assimilation algorithms have had an important impact on improving predictive performance of simulations in many geoscience and engineering applications. However, incorporating noisy data into uncertain computational models presents a major challenge in terms of assessment of accuracy of the procedure due to the potential accumulation of model error, loss of observability due to sparse, partial observations. The practical necessity to obtain computationally feasible approximations to the state estimation problem with a high-dimensional state space and imperfect knowledge of the underlying dynamics results in ad-hoc algorithms whose accuracy and robustness in the presence of model error remains limited. Nonetheless, the analysis of performance of these algorithms is important, especially as prediction is pushed to increasingly longer time horizons, or when the subtlety of physical processes modelled is increased; this situation is common in numerical weather prediction. In this paper we focused on the performance of two classes of approximate Gaussian data assimilation algorithms. Both of these classes of filters, i.e., 3DVAR and SPEKF/GCF, utilise Gaussian approximations to construct the posterior (filtering) distributions of the system state given sparse observations of the underlying true dynamics but they differently update the error covariance information; in both cases the model error is introduced in the forward dynamics determining the prior estimates of the state. The 3DVAR filter is prototypical of sequential methods used to combine incoming observations with a dynamical system in order to improve the state estimation. Given optimally inflated covariance, 3DVAR is known to be (provably) accurate for filtering dissipative systems in the absence of model error in the forward dynamics. In contrast, SPEKF/GCF algorithms do not require covariance inflation or a detailed knowledge of the underlying dynamics, and they have been shown (empirically) to be effective in mitigating model error in state estimation of turbulent dynamical systems. Here, we focused on comparing the accuracy of predictions for the mean state rather then the underlying posterior probability distribution. Uncertainty of the mean estimates, e.g., the error covariance information, is not considered here because of a very high computational cost of obtaining these quantities, which in the present setting, would require Markov Chain Monte Carlo (MCMC) or sequential Monte Carlo (SMC) sampling, or particle filtering of multimodal densities over high-dimensional spectral domains and long time windows. Particular attention was paid to the effects of the sparsity of observations on the resulting estimates which is commonplace but rarely considered. This is driven by the desire to mimic realistic problems when the observed data corresponds to finite-resolution measurements, and the information about the unresolved dynamics is, at best, scrambled with the information about the resolved components of the underlying dynamics.

As the first step in the analysis, we compared the numerical performance of 3DVAR and SPEKF algorithms on a canonical test problem given by the 2D Navier-Stokes dynamics with a linear dissipation which allows for introduction of a controllable model error in the forward model, while assimilating spatially sparse data. Moreover, this dynamics provides a very useful test problem as it represents a prime example of a dissipative infinite-dimensional dynamical system prototypical of the high-dimensional state estimation problem to which data assimilation is applied in practice. The model error in the forward model dynamics is introduced by the spectral (Galerkin) truncation of the original dynamics, while the sparsity of observations and their spectral resolution is controlled by the distance between nodes on the observation grid in the spatial domain. These studies, under the assumption of a well-defined posterior probability distribution, lead to four main conclusions:

(i) With appropriate parameter choices and for non-aliased observations, both approximate Gaussian filters - 3DVAR and SPEKF/GCF - perform well in reproducing the mean of the desired filtering

probability distribution. This applies to both the perfect and imperfect model scenarios in various dynamical regimes of the dissipative 2D Navier-Stokes dynamics.

(ii) In the presence of sparse aliased observations of turbulent dynamics (complex and high-dimensional in spectral space) the SPEKF-type algorithms, which do not require covariance inflation, perform significantly better than a tuned 3DVAR algorithm.

(iii) The need to modify the background covariance in the 3DVAR algorithm in order to induce stability of estimates and avoid divergence poses a significant drawback to this approach and related methods; the background covariance can be estimated from historical data while the optimal tuning via the covariance inflation requires hind-cast adaptation. Filter stabilisation via covariance inflation ameliorates the instability inherent in 3DVAR and can improve long-term accuracy of the algorithm in predicting the mean of the distribution, but it makes it impossible to predict the covariance.

(iv) Superresolution of aliased observations by means of utilising a forward model with the spectral resolution higher than the resolution of the observations is beneficial for the SPEKF filters but it is detrimental to the 3DVAR filter. This applies to the ability to recover the mean of the full truth signal, as well to the recovery of the primary modes resolved by the observations. SPEKF algorithms can 'learn', and to some extent filter-out the additional coloured noise due to the aliased observations on-the-fly, while the forward model in 3DVAR does not have enough degrees of freedom to account for such effects.

These conclusions are intrinsic to the considered algorithms, and result from the nature of the approximations made in order to create tractable online algorithms; the basic conclusions are not expected to change by use of different dynamical models or by modifying the parameters of those algorithms.

There are many possible directions for extension of this work and future research in this area which require attention. We outline the most important issues in the context of already existing and related research themes. First, we note that the ability of various data assimilation algorithms to predict uncertainty from a fully Bayesian perspective was considered in [46] in the absence of model error and for non-aliased observations. In that work the authors relied on MCMC sampling to compare the true posterior (filtering) distribution over the system state with the distributions obtained from approximate sequential or variational data assimilation algorithms (including 3DVAR, 4DVAR, ExKF, and EnKF, but not SPEKF). Although, in principle, consistent statistical sampling algorithms such as MCMC and SMC samplers can recover any distribution, this becomes prohibitively expensive for multimodal distributions with rare transitions between modes. Consequently, the necessary computations in [46] were carried out in regimes of the 2D Navier-Stokes dynamics which were chaotic but characterised by unimodal, nearly Gaussian distributions with a sufficiently small number of 'energetic' modes to allow state-of-the-art, fully resolved MCMC computation of the Bayesian posterior distribution. The authors of [46] concluded that, while many of the considered algorithms could reliably reproduce the mean state provided that they were appropriately tuned, the uncertainty of the estimates was unreliable in most cases. An analogous study and computations in the turbulent regimes considered here pose a significant computational and algorithmic challenge which is yet to be performed for both 3DVAR and SPEKF algorithms; it would be very interesting to see if the SPEKF algorithms provide better uncertainty estimates than the older but well established filters. It would also be preferable to look at long time intervals and turbulent regimes, rather than short time intervals and chaotic regimes such as in [46]. Theoretical results explaining these characteristics in the case of 3DVAR may be found in [10, 7] in the context of filtering dissipative dynamical

systems; however, this important analysis focuses on the more tractable, idealised case when noisy observations of individual spectral modes of the truth are available (i.e., the case of non-aliased observations) and under the assumption that the observation and covariance operators are diagonal and commute. A similar study in the context of SPEKF algorithms, even in the absence of aliased observations, poses a number of technical challenges due to the fact that the mean and covariance evolution of the posterior distribution are coupled in a non-trivial fashion.

Moreover, we note that many comparisons of various data assimilation and variational algorithms, with the exception of SPEKF, have been carried out recently. For example, [58, 72, 73] compare the EnKF forecast with 3DVAR and 4DVAR (without updated covariance) in real-data experiments based on the Weather Research and Forecasting model (WRF). The conclusion are that EnKF and 4DVAR perform best with respect to the root-mean-square error (the pattern correlation is not considered), while the EnKF forecast performs better for longer lead times. This result is consistent with that of [46], although it could be explained by an improved approximation of the posterior distribution at each update time. Our results indicate 4DVAR could perform better here, as long as the approximate filtering distribution of 4DVAR with the propagated Hessian is used. Of course this is too expensive in practice and often a constant covariance is used; this will limit performance in reproducing the statistical variation of the posterior filtering distribution for prior in the next cycle. This issue is addressed partially in [58, 73] where EnKF is coupled to 4DVAR and the covariance comes from the former, while the mean is updated by the latter, and the resulting algorithm outperforms either of the individual ones in the RMS sense. Two fundamental classes of EnKFs were compared theoretically in the large ensemble limit in [48], and it was found that the stochastic version (also considered in [46]) in which observations are perturbed is more robust to perturbations in the forecast distribution than the deterministic one. Another interesting comparison was carried out in [27] in which several ensemble filters, alternative to EnKF in operational use, were compared with respect to RMS, as well as other diagnostics such as rank histograms (Anderson 1996). A numerical comparison of of the performance of SPEKF algorithms with these filters deserves a separate study and will be soon reported elsewhere.

Finally, it would be interesting to conduct a study, similar to the one undertaken here, for simple models of atmospheric dynamics such as Lorenz-96, models exhibiting behaviour analogous to atmospheric blocking events, or for more realistic quasigeostrophic models which admit baroclinic instabilities [55, 40]. With recent progress in consistent multilevel Monte Carlo (MLMC) sampling algorithms [33, 5, 35], it may be possible in the foreseeable future to obtain reliable estimates of the full posterior filtering distribution over long-time windows for low-dimensional yet suitably complex systems with turbulent dynamics, such as Lorenz-96. Then a study may be performed along the lines of [46] to follow up this work.

Acknowledgements. M.B. acknowledges the support of Office of Naval Research grant ONR N00014-15-1-2351. A. J. M. acknowledges the support of the Office of Naval Research MURI N00014-16-1-2161 and DARPA through W91NF-15-1-0636.

Appendix A. Tuning the forward models in the filtering algorithms

In order to assure a consistent comparison, the algorithms are tuned using the same data obtained from long runs of the simulated truth dynamics (31). In the context of 3DVAR (cf. §4.1), and in line with [46, 7], the tuning entails estimating the 'background' covariance C_0 in (40) which is taken to be diagonal in the spectral basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and estimated as described below. Alternatively, in line with one of the implementations used in [46], the background covariance could be chosen as $C_0 \propto \mathcal{L}^{-2}$, where $\mathcal{L}$ defined in (8) is a closed positive operator on $\mathcal{H}$. This parameterisation reflects the empirical fact that

the ratio of the prior to the observational covariance (assumed constant and diagonal) is larger for smaller wavenumbers; this choice gives qualitatively similar but quantitavely worse results for both algorithms. The second tuning step required in filtering with 3DVAR to prevent filter divergence relies on empirical inflation of the background covariance $\widehat{C}_{\alpha,\beta}$ in (40) through the multiplicative and additive parameters α and β; the importance of the covariance inflation was illustrated in various numerical tests in this paper and is well-known in the data assimilation litarature. On the other hand, the tuning of SPEKF/GCF filters requires setting values of the free parameters in the forward model (51) which are roughly estimated from the equilibrium statistics as in [8, 9]; the performance of the SPEKF filters turns out to be not very sensitive to the choice of the tuning parameters, and only the parameters in the equations for the spectral modes $\{u_k\}_{0<|k_{1,2}|\leqslant N}$ resolved by the forward model need to be estimated directly from the data (see [24, 23, 8, 9]). Thus, SPEKF-type filters require a single tuning step while 3DVAR requires two tuning steps. The first step of the tuning procedure is similar for both 3DVAR and SPEKF filters and it utilises an Ornstein-Uhlenbeck (OU) process as a model for the dynamics of the modes $u_k(t)$ in the solution of the forward map

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \tag{59}$$

in such a way that the second-order statistics of the modes u_k of the OU process on the attractor coincides with that of the truth on the attractor. Due to the exact solvability of the OU process, this is done as follows: The OU dynamics is given by

$$dU = -MUdt + \sqrt{2\Re e[M]\Xi}\, dW_t, \tag{60}$$

where U represents a vector of all the resolved spectral coefficients, $M, \Xi > 0$ are diagonal and positive definite, and W_t is the standard Wiener process in an appropriate dimension. The stationary solution of (60) is a Gaussian process with mean zero and covariance Ξ which are tuned to the truth via

$$\Xi = \lim_{T\to\infty} \frac{1}{T}\int_0^T [u(t)-\bar{u}]\otimes[u(t)-\bar{u}]^* dt, \qquad \bar{u} = \lim_{T\to\infty}\frac{1}{T}\int_0^T u(t)dt. \tag{61}$$

The diagonal entries $M_{m,m}$ are set based on the attractor statistics of the truth using the formulas

$$Corr_k(\tau) = \lim_{T\to\infty} \Xi_{k,k}^{-2}\int_0^T C_{k,k}(t,\tau)dt, \qquad C(t,\tau) = [u(t-\tau)-\bar{u}]\otimes[u(t)-\bar{u}]^*, \tag{62}$$

and

$$T_k + i\Theta_k = \int_0^\infty Corr_k(\tau)d\tau, \tag{63}$$

so that

$$\Re e[M_{k,k}] = \frac{T_k}{T_k^2+\Theta_k^2}, \qquad \Im m[M_{k,k}] = -\frac{\Theta_k}{T_k^2+\Theta_k^2}. \tag{64}$$

In practice the integrals in (61), (62) are approximated by finite discrete sums; furthermore, we set the off-diagonal entries of Ξ to zero to obtain a diagonal model. As shown in Figure 9, a satisfactory accuracy is reached relatively quickly in terms of the length of the 'training' time interval; note, however, that the results for 3DVAR are shown for the optimal choice of the multiplicative inflation parameter β in (40) which requires more than just the estimates of statistics from the training data. The two parameters estimated from data in the SPEKF forward model (51) are set as

$$\bar{l}_k = M_{k,k}, \quad \sigma_{u_k}^2 = 2\Re e[M_{k,k}]\,\Xi_{k,k},$$

and the remaining parameters are set as

$$d_{\gamma_k} = d_{b_k} = 0.1\,\Re e[\bar{l}_k], \quad d_{\omega_k} = \omega_{b_k} = 0.1\,\Im m[\bar{l}_k], \quad \sigma_{\gamma_k} = \sigma_{\omega_k} = \sigma_{b_k} = 0.6\,\sigma_{u_k}. \tag{65}$$

In order to systematically estimate the background covariance in 3DVAR we first note that the discrete-time solution of the OU process in (60) is given by the linear stochastic map

$$U_{n+1} = LU_n + \sqrt{Q}\,\xi_n, \tag{66}$$

where $L = \exp(-M\Delta)$ and $Q = \big(I - \exp(-2\Re e[M]\Delta)\big)\Xi$ are both diagonal in the spectral basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ and $\{\xi_n\}$ is i.i.d. with $\xi_n \sim \mathcal{N}(0, I)$. For the forward model in (66) the update map

$$(m_n, C_n) \to (m_{n+1}, C_{n+1}), \tag{67}$$

in (22) of §3 yields the Kalman filter with

$$\widehat{m}_{n+1} = Lm_n, \qquad \widehat{C}_{n+1} = LC_nL^* + Q. \tag{68}$$

In the spirit of 3DVAR, the above update can be improved by updating the covariance as in (68) and updating the mean by the nonlinear flow map corresponding to (33), namely

$$\widehat{m}_{n+1} = \Phi^N_\Delta(m_n), \qquad \widehat{C}_{n+1} = LC_nL^* + Q, \tag{69}$$

$$m_{n+1} = (I - K_{n+1}H)\widehat{m}_{n+1}, \qquad C_{n+1} = (I - K_{n+1}H)\widehat{C}_{n+1}, \tag{70}$$

$$K_{n+1} = \widehat{C}_{n+1}H^*(H\widehat{C}_{n+1}H^* + \Gamma)^{-1}. \tag{71}$$

We note that because L is diagonal with $LL^* < 1$, the covariance C_n converges to a limit [45] that can be computed numerically off-line and, asymptotically the algorithm behaves like 3DVAR; thus, in line with [46] this asymptotic covariance is used as the systematic choice of background covariance $\widehat{C}_0$ in (40). Alternatively, one may set $\widehat{C}_0 = \Xi$ which corresponds to the update (69) with $\Delta \to 0$ in L and Q; both choices of $\widehat{C}_0$ give very similar results in our tests due to the fact that that Δ is small relative to the correlation times for a large fraction of modes in the forward models.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Dover Books on Electrical Engineering, 1979.

[2] E. Andersson and H. Järvinen. Variational Quality Control. *Quarterly Journal Royal Met. Society*, 125(554):697, 1998.

[3] A. Azouani, E. Olson, and E.S. Titi. Continuous data assimilation using general interpolant observables. *Journal of Nonlinear Science*, 24(2):277–304, 2014.

[4] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*, volume 60. Springer-Verlag New York, 2009.

[5] Alexandros Beskos, Ajay Jasra, Kody Law, Raul Tempone, and Yan Zhou. Multilevel sequential monte carlo samplers. *Stochastic Processes and their Applications*, 2016.

[6] H. Bessaih, E. Olson, and E.S. Titi. Continuous data assimilation with stochastically noisy data. *Nonlinearity*, 28(3):729, 2015.

[7] D. Blömker, K. Law, A. M. Stuart, and K. C. Zygalakis. Accuracy and stability of the continuous-time 3DVAR filter for the Navier–Stokes equation. *Nonlinearity*, 26(8):2193–2219, 2013.

[8] M. Branicki, B. Gershgorin, and A.J. Majda. Filtering skill for turbulent signals for a suite of nonlinear and linear kalman filters. *J. Comp. Phys*, 231:1462–1498, 2012.

[9] Michal Branicki and Andrew J Majda. Dynamic stochastic superresolution of sparsely observed turbulent systems. *Journal of Computational Physics*, 241:333–363, 2013.

[10] Charles EA Brett, Kei Fong Lam, KJH Law, DS McCormick, Michael R Scott, and AM Stuart. Accuracy and stability of filters for dissipative pdes. *Physica D: Nonlinear Phenomena*, 245(1):34–45, 2013.

[11] J M Burgess, C. Bizon, W. D. McCormick, J. B. Swift, and H L Swinney. Instability of the Kolmogorov flow in a soap film. *Phys. Rev. E*, 60(1):715–721, 1999.

[12] A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi. Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18:023112, 2008.

[13] A. Chorin, M. Morzfeld, and X. Tu. Implicit particle filters for data assimilation. *Communications in Applied Mathematics and Computational Science*, page 221, 2010.

[14] P. Constantin and C. Fojas. *Navier-Stokes Equations*. Chicago Lectures in Mathematics. Chicago/London, 1988.

[15] P. Constantin, C. Fojas, I. Kukavica, and A.J. Majda. Dirichlet quotients and 2d periodic navier-stokes equations. *J. Math. Pures Appl.*, 76:125–153, 1997.

[16] P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q.J.R. Meteorol. Soc.*, 124:1783–1807, 1998.

[17] T. DelSole. Stochastic Models of Quasigeostrophic Turbulence. *Surveys in Geophys.*, 25(2):107–149, 2004.

[18] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice*. Springer Verlag, 2001.

[19] G. Evensen. *Data Assimilation: the Ensemble Kalman Filter*. Springer Verlag, 2009.

[20] W. H. Fleming and R. W. Rishel. *Deterministic and Stochastic Optimal Control. New York: Springer*. Springer New York, 1975.

[21] C. Foias and G. Prodi. Sur le comportement global des solutions non-stationnaires des équations de Navier-Stokes en dimension 2. *Rendiconti del Seminario Matematico della Università di Padova*, 39:1–34, 1967.

[22] C. Foias and E.S. Titi. Determining nodes, finite difference schemes and inertial manifolds. *Nonlinearity*, 4(1):135, 1991.

[23] B. Gershgorin, J. Harlim, and AJ Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *Journal of Computational Physics*, 229(1):32–57, 2010.

[24] B. Gershgorin, J. Harlim, and AJ Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *Journal of Computational Physics*, 229(1):1–31, 2010.

[25] B. Gershgorin and AJ Majda. Filtering a statistically exactly solvable test model for turbulent tracers from partial observations. *Journal of Computational Physics*, 230(4):1602–1638, 2011.

[26] I. Grooms, Y. Lee, and A.J. Majda. Ensemble kalman filters for dynamical systems with unresolved turbulence. *J. Comput. Phys.*, 273:435–452, 2014.

[27] C. Snyder Hamill, T. and 2000: R. Morss. A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. *Mon. Wea. Rev.*, 128:1835–1851, 2000.

[28] J. Harlim and AJ Majda. Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity*, 21:1281, 2008.

[29] J. Harlim and A.J. Majda. Filtering turbulent sparsely observed geophysical flows. *Monthly Weather Review*, 138(4):1050–1083, 2010.

[30] John Harlim, Andrew J Majda, et al. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, 8(1):27–43, 2010.

[31] A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman filter*. Cambridge Univ Pr, 1991.

[32] Kevin Hayden, Eric Olson, and Edriss S Titi. Discrete data assimilation in the Lorenz and 2D Navier–Stokes equations. *Physica D: Nonlinear Phenomena*, 240(18):1416–1425, 2011.

[33] Viet Ha Hoang, Christoph Schwab, and Andrew M Stuart. Complexity analysis of accelerated mcmc methods for bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[34] T. Jancic and S.E. Cohn. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev*, 134:2900–2915, 2006.

[35] Ajay Jasra, Kengo Kamatani, Kody JH Law, and Yan Zhou. Multilevel particle filter. *arXiv preprint arXiv:1510.04977*, 2015.

[36] A.H. Jazwinski. *Stochastic processes and filtering theory*. Academic Pr, 1970.

[37] D.A. Jones and E.S. Titi. On the number of determining nodes for the 2D Navier-Stokes equations. *Journal of Math. Anal. Appl.*, 168(1):72–88, 1992.

[38] Rudolph Emil Kalman et al. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[39] Eugenia Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge university press, 2003.

[40] S.R. Keating, A.J. Majda, and K.S. Smith. New methods for estimating poleward eddy heat transport using satellite altimetry. *submitted to Monthly Weather Review*, 2011.

[41] David Kelly, Andrew J Majda, and Xin T Tong. Concrete ensemble kalman filters with rigorous catastrophic filter divergence. *Proceedings of the National Academy of Sciences*, 112(34):10589–10594,

2015.

[42] DTB Kelly, KJH Law, and Andrew M Stuart. Well-posedness and accuracy of the ensemble kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.

[43] R. H. Kraichnan. Inertial ranges in two-dimensional turbulence. *The Physics of Fluids*, 10(7):1417–1423, 1967.

[44] Robert H Kraichnan. Decimated amplitude equations in turbulence dynamics. pages 91–135, 1985.

[45] K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, 2015.

[46] Kody JH Law and Andrew M Stuart. Evaluating data assimilation algorithms. *Monthly Weather Review*, 140(11):3757–3782, 2012.

[47] Y. Lee and A. J. Majda. State estimation and prediction using clustered particle filters. *Proc. Nat. Acad. Sci.*, 113(51):14609–14614, 2016.

[48] J. Lei, P. Bickel, and C. Snyder. Comparison of ensemble Kalman filters under non-Gaussianity. *Mon. Wea. Rev.*, 138:1293–1306, 2010.

[49] R.S. Liptser and A.N. Shiryaev. *Statistics of Random Processes*. Springer New York, 1978.

[50] Lennart Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24(1):36–50, 1979.

[51] A.C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[52] A. Majda and X. Wang. *Non-linear dynamics and statistical theories for basic geophysical flows*. Cambridge Univ Pr, 2006.

[53] A.J. Majda. *Introduction to Turbulent Dynamical Systems for Complex Systems*. Springer, 2016.

[54] A.J. Majda and M.J. Grote. Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems. *PNAS*, 104:1124–1129, 2007.

[55] A.J. Majda and J. Harlim. *Filtering Complex Turbulent Systems*. ISBN-13:9781107016668. Cambridge University Press, 2012.

[56] A.J. Majda and J. Harlim. *Filtering Complex Turbulent Systems*. CUP, 2012.

[57] A.J. Majda, J. Harlim, and B. Gershgorin. Mathematical strategies for filtering turbulent dynamical systems. *Dynamical Systems*, 27(2):441–486, 2010.

[58] Z. Meng and F. Zhang. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a month-long experiment. *Mon. Wea. Rev.*, 136:3671–3682, 2008.

[59] E. Olson and E.S. Titi. Determining modes for continuous data assimilation in 2D turbulence. *Journal Statist. Phys*, 113(5):799–840, 2003.

[60] M. Rivera and X. L. Wu. External Dissipation in Driven Two-Dimensional Turbulence. *Phys. Rev. Lett.*, 85:976, 2000.

[61] David Higdon Roger Ghanem and Houman Owhadi. *Handbook of Uncertainty Quantification*. Springer International Publishing, 2017.

[62] R. Salmon. *Lectures on Geophysical Fluid Dynamics*. Oxford University Press, 1998.

[63] T. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review.*, 136:4629–4640, 2008.

[64] T.J. Tarn and Y. Rasis. Observers for nonlinear stochastic systems. *Automatic Control, IEEE Transactions on*, 21(4):441–448, 1976.

[65] R. Temam. *Navier-Stokes Equations and Nonlinear Functional Analysis*. Number 66. Society for Industrial Mathematics, 1995.

[66] Xin T Tong, Andrew J Majda, and David Kelly. Nonlinear stability of the ensemble kalman filter with adaptive covariance inflation. *Communications in Mathematical Sciences*, 14(5), 2016.

[67] A. Trevisan and L. Palatella. Chaos and weather forecasting: the role of the unstable subspace in predictability and state estimation problems. *Int. J. Bifurcation Chaos*, 21:3389–415, 2011.

[68] Y-K Tsang. Nonuniversal velocity probability densities in two-dimensional turbulence: The effect of large-scale dissipation. *Phys Fluids*, 22:115102, 2010.

[69] Y-K Tsang and W. R. Young. Forced-dissipative two-dimensional turbulence: A scaling regime controlled by drag. *Phys. Rev. E*, 79,:045308 R, 2009.

[70] P.J. Van Leeuwen. Particle filtering in geophysical systems. *Monthly Weather Review*, 137:4089–4114, 2009.

[71] PJ van Leeuwen. Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quarterly Journal of the Royal Meteorological Society*, 136(653):1991–1999, 2010.

[72] M. Zhang, X. Huang, and X. Zhang. Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. *Mon. Wea. Rev.*, 139:566–572, 2010.

[73] M. Zhang and F. Zhang. E4DVAR: Coupling an ensemble Kalman filter with four-dimensional variational data assimi- lation in a limited-area weather prediction model. *Mon. Wea. Rev.*, 140:587–600, 2012.