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] *School of Mathematics, University of Edinburgh, Edinburgh, UK*

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

[3] *School of Mathematics, University of Manchester, Manchester, UK*

Abstract. Bayesian state estimation of a dynamical system given a stream of noisy measurements is important in many geophysical and engineering applications where high dimensionality of the state space, sparse observations, and model error pose key challenges. Here, three computationally feasible, approximate Gaussian data assimilation/filtering algorithms are considered in various regimes of turbulent 2D Navier-Stokes dynamics in the presence of model error. The first source of error arises from the necessary use of reduced models for the forward dynamics, while a particular and idealised type of so-called representation error arises from the finite resolution of observations which mix up information about resolved and unresolved dynamics. Two stochastically parameterised filtering algorithms, referred to as SPEKF and GCF, are compared with 3DVAR - a prototypical time-sequential algorithm which is known to be accurate for filtering dissipative systems given optimally inflated 'background' covariance. We provide the first evidence that the stochastically parameterised algorithms, which do not require covariance inflation, outperform the optimally tuned 3DVAR algorithm, and they can overcome competing sources of error in a wide range of dynamical scenarios of our turbulent test problem.

1. Introduction

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., [46]) or engineering problems (e.g., [30]) where online predictions are required in the presence of uncertainty in the initial conditions, the observations, and in the dynamics itself. Such problems can be cast within the Bayesian framework which allows for a systematic combination of incoming observations with a dynamical model in order to solve a sequence of inverse problems on the current state of the estimated process. In principle, when an exact dynamics is known subject to uncertain initial conditions, the posterior or filtering distribution on the system state given the observations can be derived. This may be performed exactly for linear systems subject to Gaussian noise, leading to the Kalman filter (e.g., [45, 2, 43, 36]). In nonlinear and non-Gaussian scenarios the particle filter (e.g., [5, 22]) provably approximates the true posterior distribution as the number of particles increases. Standard implementations of this method perform poorly in high-dimensional systems [75] although there is a growing literature on particle filtering in high-dimensional systems [55, 72, 85, 6, 59, 71].

The field of data assimilation has grown out of the necessity to obtain computationally feasible approximations to the filtering distribution when one is faced with a high-dimensional state estimation and/or imperfect knowledge of the underlying dynamics, and a vast amount of incoming data (e.g., weather prediction). In such situations one is typically forced to employ approximations based on physical insight and computational expediency, while the incoming data are used to compensate for the uncertainty in

E-mail address: M.Branicki@ed.ac.uk, kody.law@manchester.ac.uk, jonjon@cims.nyu.edu.

the model and in the initial conditions. The development of practical and robust data assimilation algorithms for high-dimensional dynamics is an active research area, see e.g., [15, 16, 23, 66, 67, 84, 85]. Many such algorithms invoke some form of Gaussian approximation which generally destroys optimality of the mean estimates and might cause filter divergence [58, 34, 49] (optimality of the posterior mean implies minimisation of the error covariance between itself and the true signal over all estimators constructed as L^2 functions of the observation sequence [5, 43]). The study of accuracy and stability of data assimilation algorithms has been a developing area over the last few years. Informally, filter accuracy concerns the closeness of filter estimates to the true signal underlying the data, and stability is concerned with long-time convergence between two sequences of filter estimates driven by the same noisy data but initialised differently (see, e.g., [52] for details). A recent series of papers provides a generalisation of the theory in [76, 81] to dissipative infinite-dimensional dynamical systems which are prototypical of the high-dimensional problems to which filters are applied in practice [12, 9, 48, 4, 8, 37]. In many cases carefully tuned approximate filters can be stable and accurate for estimating mean dynamics but they typically perform poorly when predicting the associated uncertainty (e.g., [54, 32, 68, 56, 86, 87]). Moreover, the tuning procedure relies on retrospective, hind-cast adjustments which require large amounts of historical data. Model error in the forward dynamics is compounded further by sparsity of observations which has adverse effects on the accuracy and stability of data assimilation algorithms [15, 81].

Here, we provide the first evidence that a class of computationally cheap, approximate Gaussian, stochastically parameterised filtering algorithms introduced in [28, 27, 66] is capable of overcoming commonly encountered uncertainties to produce accurate mean estimates in realistic models of turbulent dynamics. We compare the performance of two such algorithms, referred to as cSPEKF and GCF, with a 'reference' algorithm - 3DVAR [60] - which is prototypical of many approximate Gaussian filters used in practice today and has its origin in weather forecasting (e.g., [3, 19, 54]). All three algorithms introduce model error by simplifying the forward dynamics and by employing Gaussian approximations of the posterior (filtering) distributions; amongst the possible approximate Gaussian nonlinear filtering algorithms these are the simplest and least expensive which guides the choice for this study. The 3DVAR method and its generalisations such as the Extended Kalman filter (ExKF, [43]) and the Ensemble Kalman filter (EnKF, [23]), are observed to be accurate 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 [54, 80]. However, the perfect model scenario does not apply in practice and the stochastically parameterised algorithms have been introduced as an efficient way of dealing with this issue. Here, the stochastic parameterisation approach exploits cheap, exactly solvable, conditionally Gaussian forward models in the spectral domain to accommodate model error and to propagate the covariance information via an online learning of certain auxiliary processes that adapt the model dynamics. This method has been shown to be effective on a range of test models [28, 27, 35, 29, 10, 47, 66], and it was extended in [65, 11] to superresolving the state of one-dimensional PDE models from sparse aliased observations. While the numerical evidence for efficacy of SPEKF-type filters is very promising, rigorous analysis poses a number of technical challenges, especially for spatially sparse observations. Analytical results for filtering dissipative PDEs using 3DVAR in an idealised case when noisy observations of individual spectral modes are available may be found in [12, 9] which exploit techniques from [37, 70, 8]; the latter body of work has been extended also to spatial observations [4], following on from earlier results of [26, 44, 25, 50]. The main difficulties in the analysis of SPEKF-type filters stem from the fact that, in contrast to 3DVAR, bounds on error in both the mean and covariance updates need to be considered simultaneously. Mindful of these difficulties we focus for now on a detailed numerical study of performance of SPEKF-type algorithms.

The emphasis is on studying the interplay between different sources of error in a more realistic but still academic setting, rather than on tests including all operational constraints.

Two sources of model error in filtering are considered: The first one stems from the necessary use of finite-resolution approximations for the forward dynamics, while the second source of error arises as a consequence of spatially sparse observations which is manifested by some degree of mode aliasing in the spectral domain. The spectral projection is necessary for dynamical model reduction and efficient state estimation in the SPEKF-type filters (see §4.2.1) but it also necessitates a spectral projection of the sparse observations onto the modes resolved by the approximate forward model; this results in scrambling up the information about resolved/unresolved dynamics and it is used in our setup as a particular incarnation of so-called representation error [60, 41, 31]. Here, this error arises from the fact that the aliased modes are accounted for in the observations but they are not necessarily represented in the dynamics of the forward models of the filters; see [40, 1] for further discussion on representation error. Moreover, as a follow up to [47] and [11], we use a special case of equally-spaced observations to analyse the potential of the three algorithms for a dynamic superresolution of observations. In this setup we find that SPEKF and GCF outperform optimally tuned 3DVAR algorithms in a wide range of dynamical scenarios. SPEKF-type algorithms can learn and to some extent filter-out the unresolved modes from the scrambled observations, thus correcting aspects of representation error on-the-fly. This is important and encouraging given that these stochastically parameterised algorithms are computationally cheap and do not require covariance inflation. A survey of other recent multi-scale approaches to filtering and prediction can be found in [63].

The rest of the paper is structured as follows: In section 2 we outline the main characteristics of the 2D Navier-Stokes dynamics which is used as a test-bed for comparing the performance of data assimilation algorithms in various turbulent regimes. The Navier-Stokes dynamics is 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. Section 3 outlines the main concepts leading to the derivation of approximate Gaussian filters which are used throughout this paper; the three specific algorithms, 3DVAR, cSPEKF, and GCF, are described in section 4. The bulk of numerical tests are discussed in section 5. We conclude in section 6, summarising the main findings and outlining directions for future work.

2. Test Problem

Comparison of data assimilation algorithms requires a tuneable test-bed dynamics for generating the truth signal and the observation data, as well as for constructing the forward dynamics with model error used in the subsequent state estimation. A version of the two-dimensional Navier-Stokes equation provides such a benchmark problem which, as a prime example of a dissipative infinite-dimensional dynamical system with a wide range of dynamical regimes, is prototypical of the high-dimensional state estimation to which data assimilation is applied in practice. Some necessary concepts used in the subsequent sections are introduced below. Further details concerning 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. The above system has a global attractor whose dimensionality grows with the ratio of forcing to dissipation (e.g., [17, 77]); 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 (e.g., [44, 70, 37]), and thus more attractive for testing data assimilation algorithms. It is noted that short-term predictability is governed by eigenfunctions corresponding to finite-time Lyapunov exponents, which may also include algebraically growing modes. Such eigenfunctions have been used to design effective filters for example in the work [53]. Here, the additional linear damping term generates a more suitable attractor dynamics with less energy at the large scales compared to the standard 2D Navier-Stokes dynamics (i.e., $\kappa = 0$ in (1)); this modification takes account of a large-scale dissipation and is commonly used in modelling quasi-2D turbulence encountered in experiments [14, 73, 83, 82]. In the subsequent sections, we consider three dynamical regimes of the system (1)-(3) with 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., [77], but we repeat the main steps and notions since they will be needed in subsequent considerations. First, consider the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}^1$ given by 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-Kondrachov theorem. Then, any real-valued $u \in \mathcal{H}$, including weak solutions of (1)-(3), 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 $k = (k_1,k_2) \in \mathbb{Z}^2 \setminus \{0\}$, and $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$, $\psi_k\colon \mathbb{T}^2 \to \mathbb{C}^2$, is the orthonormal basis in $\mathcal{H}$

$$\psi_k(x) := \frac{k^\perp}{|k|}\exp\left(\frac{2\pi \mathrm{i}\, k\cdot x}{L}\right), \quad k^\perp := (k_2,-k_1)^{\mathrm{T}}, \quad |k| = (k_1^2+k_2^2)^{1/2}. \tag{6}$$

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 $\phi_k = |k|^{-1}(k^\perp\cdot\psi_k)$. We confine attention to time-independent *Kolmogorov forcing* $f \in \mathcal{H}$ (e.g., [61])

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

which acts at a subset of the wavenumbers $\mathfrak{S}_{N_f} := \{k^i \in \mathbb{Z}^2\setminus\{0\}\colon |k^i| = N_f\}$. Such a special forcing has an attractive mathematical theory (see [61, Chapter 2]) and it is sufficient for our purposes.

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., [17]) 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

$$\langle \mathcal{B}(u,u), u\rangle = 0, \qquad \langle \mathcal{B}(u,v), v\rangle \leq C\, \|u\|\|v\|\, |v|, \quad C > 0, \ \ \forall\ u, v \in \mathcal{H}^1. \tag{10}$$

The global upper bound on the norm of the solution of (8) is obtained (see, e.g. [77, 18]) by combining the above properties with the Gronwall lemma, and it is given by

$$|u(t)|^2 \leqslant |u(0)|^2 e^{-\nu\tilde{\lambda}_1 t} + \frac{|f|^2}{\nu^2\tilde{\lambda}_1^2}\Big(1 - e^{-\nu\tilde{\lambda}_1 t}\Big), \qquad t \geqslant 0. \tag{11}$$

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., [17, 77]. The nature of long-time dynamics of (8) depends on the number of positive Lyapunov exponents which are controlled by the forcing f and dissipation ν parameters, and are a proxy for the dimensionality of the unstable manifold of the attractor. 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 effective dimensionality of the long-time dynamics. In the subsequent sections we focus on estimating finite-dimensional dynamics derived from the spectral truncation of (8) and evolving on the attractor $\mathfrak{A}_\Lambda$, $\Lambda < \infty$ (see §3.1 and, e.g., [79, 78]).

3. Approximate Gaussian framework for time-sequential data assimilation

Here, we focus on state estimation of the deterministic dynamics (8) from sparse and noisy observations. The canonical stochastic filtering problem aims at a sequential-in-time, Bayesian estimation of the state $u(x,t)$ by combining the governing dynamics (8) with its observations. Assuming that the observations of the true state are known at discrete times $\{t_n\}_{n\in\mathbb{N}}$, the goal of stochastic filtering is to find a map

$$\mathbb{P}_n(u\,|\,\mathcal{Y}_n) \xrightarrow{\mathfrak{F}} \mathbb{P}_{n+1}(u\,|\,\mathcal{Y}_{n+1}), \tag{12}$$

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. However, simulating the infinite-dimensional dynamical system (8) is not possible even if the truth dynamics and the initial condition are known. Moreover, the probability measures in (12) are defined on the function space $\mathcal{H}$ which points to a further computational intractability of filtering PDE dynamics.

Here, we study approximate Gaussian filtering or data assimilation algorithms which combine noisy observations of the truth with *forward* dynamics obtained from a finite-dimensional approximation of the original dynamics (see, e.g., [52, 66] for details). Despite important differences between various approximate Gaussian algorithms, they all share the same general structure, owing to the same approximations

imposed on the underlying prior and posterior probability densities. Numerical tests of performance of three such algorithms, 3DVAR, SPEKF and GCF, which are described in §4, are presented in §5.

The key to deriving a tractable Bayesian data assimilation framework for nonlinear problems arising from PDE's lies in imposing Gaussian constraints on the prior and posterior probability measures. In what follows we assume that the considered probability measures have a density; in particular, assume that the measure on the initial conditions has a Gaussian density

$$\mathbb{P}_0(u) \simeq \mathcal{N}(m_0, C_0). \tag{13}$$

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

$$y_n = H\Psi_{t_n}(u_0) + \eta_n, \qquad n \in \mathbb{N}, \tag{14}$$

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

$$u_{n+1} = \Psi_\Delta(u_n), \tag{15}$$

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,

$$\mathbb{P}_n(y \,|\, u_n) \simeq \mathcal{N}(Hu_n, \Gamma). \tag{16}$$

The aim of the filtering/data assimilation algorithm is to find the conditional (or filtering) density $\mathbb{P}_n(u \,|\, \mathcal{Y}_n)$ given the observations $\mathcal{Y}_n := \{y_i\}_{i=1}^n$, $n \in \mathbb{N}$ (see, e.g, [5] for a rigorous formulation). In what follows, we consider approximate Gaussian filtering algorithms which enforce

$$\mathbb{P}_n(u \,|\, \mathcal{Y}_n) \simeq \mathcal{N}(m_n, C_n). \tag{17}$$

Consequently, designing an approximate Gaussian filter relies on constructing an update rule (e.g., [52, 66])

$$(m_n, C_n) \to (m_{n+1}, C_{n+1}). \tag{18}$$

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

$$\mathbb{P}_{n+1}(u \,|\, \mathcal{Y}_n) \simeq \mathcal{N}(\widehat{m}_{n+1}, \widehat{C}_{n+1}), \tag{19}$$

and utilising the linear form of the observations in (14) with additive Gaussian noise $\eta_n \sim \mathcal{N}(0,\Gamma)$. In situations when the observation sequence is discrete in time, the update (18) is usually split into two parts

$$(m_n, C_n) \xrightarrow{\mathfrak{P}} (\widehat{m}_{n+1}, \widehat{C}_{n+1}) \xrightarrow{\mathfrak{A}} (m_{n+1}, C_{n+1}). \tag{20}$$

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

$$(m_n, C_n) \to (\widehat{m}_{n+1}, \widehat{C}_{n+1}), \tag{21}$$

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

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

For the prediction step one simply imposes the approximation

$$\widehat{m}_{n+1} = \Phi_\Delta(m_n), \tag{23}$$

where the flow map Φ_Δ induced by the model dynamics is usually different from Ψ_Δ in (15); the choice of $\widehat{C}_{n+1}$ depends on the specific filter (see §4). For the analysis step the assumptions (17) and (19) imply

$$\mathbb{P}_{n+1}(u \,|\, \mathcal{Y}_{n+1}) \simeq \mathcal{N}(m_{n+1}, C_{n+1}), \tag{24}$$

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

$$C_{n+1} = \widehat{C}_{n+1} - \widehat{C}_{n+1}H^*(\Gamma + H\widehat{C}_{n+1}H^*)^{-1}H\widehat{C}_{n+1}, \tag{25}$$

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

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

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

in (25)-(26) is referred to as the (Kalman) gain. Equations (25)-(26) have the same structure as the standard Kalman filter equations (e.g., [52]) and they are linear in the incoming observations y_{n+1}. However, none of the approximate Gaussian filters studied in §4, 5 reduce to the standard Kalman filter. Note that the estimates generated by (25)-(26) 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 the prior covariance $\widehat{C}_{n+1} = \Psi_\Delta C_n \Psi_\Delta^*$; in such a case (25) reduces to the algebraic Riccati equation and the filter estimates are optimal w.r.t. the mean square error (e.g., [24, 2]). 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, (26)-(25), 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 (19); 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. In the subsequent experiments the 'synthetic' truth is generated from a numerical simulation of (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

$$u_\Lambda(x,t) = \sum_{|k_{x,y}| \leqslant \Lambda} u_k(t)\psi_k(x), \qquad u_{-k} = -u_k^*, \quad 1 \ll \Lambda < \infty, \tag{28}$$

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

$$\frac{du_\Lambda}{dt} + \mathcal{L}u_\Lambda + P_\Lambda \mathcal{B}(u_\Lambda, u_\Lambda) = P_\Lambda f, \qquad u_0 \in \mathcal{H}_\Lambda \subset \mathcal{H}, \tag{29}$$

where P_Λ denotes the projection onto $\mathcal{H}_\Lambda$ and $u_\Lambda(t) = \Psi_t^\Lambda(u_0)$. Thus, we focus on estimating finite-dimensional dynamics derived from the spectral truncation of (8) and evolving asymptotically on the attractor $\mathfrak{A}_\Lambda$, $\Lambda < \infty$ (see, e.g., [79, 78]). The numerical simulation of the dynamics in (29) 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 (6) which is combined with a Runge-Kutta time-stepping; here, we use a modification of the fourth-order Runge-Kutta method, ETD4RK [20], in which the heat semigroup is used together with Duhamel's formula to solve exactly for the diffusion term. A spectral Galerkin method [38] 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 has to contend with poor spatial resolution, particularly in the case of the atmosphere-ocean applications. Here, the important resolution consideration is that the unstable modes are resolved, which in the case of (1)-(2) have long spatial scales and support in low wavenumbers (e.g., [17, 25, 26, 77]). Therefore, the objective is to obtain high temporal resolution 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

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \qquad u_{-k} = -u_k^*, \quad N \ll \Lambda < \infty, \tag{30}$$

The choice of the forward dynamics Φ_Δ in the filter update (26)-(25) will depend on the algorithm used.

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

$$\frac{du_N}{dt} + \mathcal{L}u_N + P_N\mathcal{B}(u_N,u_N) = P_N f, \qquad u_0 \in \mathcal{H}_N \subset \mathcal{H}, \tag{31}$$

where $P_N : \mathcal{H} \to \mathcal{H}_N$, $P_N \circ P_N = \mathbb{1}$, 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. 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.

3.3. **Generation of observations.** In line with the setup of §3, the observations (14) are linear in the state variable and corrupted by an additive i.i.d. Gaussian noise. Moreover, we assume throughout that the observation operator H in (14) is trace-class but it need not be diagonal in $\mathcal{H}$. We consider two classes of observations which will have important consequences on the output of the filtering algorithms.

3.3.1. ***Non-aliased observations***. In this idealised case noisy observations of $2M \times 2M$ individual modes of the truth process $u(x,t)$ solving (1)-(2) are available, i.e., we have

$$y^{\mathrm{M}}(x,t_n) = \sum_{0<|k_{1,2}|\leqslant M} y_k(t_n)\psi_k(x) = \sum_{0<|k_{1,2}|\leqslant M} \big(u_k(t_n)+\eta_k\big)\psi_k(x), \qquad n, M \in \mathbb{N}, \tag{32}$$

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

$$Y_n = HU_n + \eta, \quad \eta \sim \mathcal{N}(0,\Gamma), \qquad n \in \mathbb{N}, \tag{33}$$

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 with rank $(2M)^2$ is diagonal in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ of $\mathcal{H}$.

The configuration in (32) is unrealistic from the practical viewpoint since it implies the ability to observe the dynamics of individual modes ψ_k, $0<|k_{1,2}|\leqslant M$, which correspons to spatially continuous observations. However, this setup is amenable to detailed analysis, especially for the 3DVAR filter in [9], which is why we consider it here and compare filter performance given this type of observations in §5.

3.3.2. ***Aliased observations***. Sparse, finite observations, including irregularly spaced ones with missing data, scramble up the observed information about resolved/unresolved dynamics; this is manifested differently in the spatial domain and in the spectral domain but the source of this lack of information is ultimately the same. Thus, in practice one has to contend with spatially coarse observations or with some degree of aliasing in the spectral domain. Dynamical model reduction in the stochastically parameterised filters outlined in §4.2.1 requires spectral projection of the dynamics, allowing for computational efficiency due to systematic approximations. Data assimilation in this framework necessitates a spectral projection of the observations onto the modes resolved by the forward dynamics; thus, in this case the sparsity of observations leads to a particular incarnation of so-called representation error (e.g., [60, 41, 31]). Here, we assume that the state $u(x,t_n)$ solving (1)-(2) is observed at nodes of a finite equi-spaced grid in the spatial domain, $x_{i,j} := (ih, jh)$, $1 \leqslant i,j \leqslant 2M+1$, $(2M+1)h = 2\pi$ (see Figure 1), so that the nodal observations are given by

$$y^{\mathrm{M}}_{ij}(t_n) = u\big(x_{ij},t_n\big) + \zeta_{i,j}\,, \qquad 1 \leqslant i,j \leqslant 2M+1, \quad M,n \in \mathbb{N}, \tag{34}$$

where ζ is an uncorrelated Gaussian field with $\zeta_{i,j} \sim \mathcal{N}(0,\Gamma^o)$. Irregular observations help mitigate the effects of aliasing but they do not remove them (e.g. [13]). The special case of spatially sparse observations in (34) is considered for a number of reasons: (i) the standard aliasing is easier to represent and control in our test problem, (ii) aliasing can be used to study some aspects of the interplay between representation error and the gained computational efficiency (see §5). Moreover, as a follow up to [11], we use the special case of equally-spaced observations to analyse the potential of the three algorithms for a dynamic superresolution of spatially sparse observations which, if need be, can be interpolated on to a regular grid.

As shown in [65, 33], spatially sparse regular measurements will alias the observed information from modes $|k_1| \vee |k_2| > M$ of the truth signal into modes $|k_1| \wedge |k_2| \leqslant M$ (see Figure 1). In fact, that the observation process (34) can be written as

$$y^{\mathcal{A}\{\mathrm{M}\}}_{i,j}(t_n) = \sum_{0<|\ell_{1,2}|\leqslant M} \Bigg(\sum_{k\in\mathcal{A}(\ell)} u_k(t_n) + \eta_{\ell,n} \Bigg) \psi_\ell(x_{i,j}), \qquad n \in \mathbb{N},\ \ 1 \leqslant i,j \leqslant 2M+1,\ \ M \leqslant N, \tag{35}$$

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

$$\mathcal{A}(\ell) = \Big\{ k \in \mathbb{Z}^2 \backslash \{0\} : \quad k_{1,2} = \ell_{1,2} + Mq_{1,2}, \quad q_{1,2} \in \mathbb{Z} \Big\}. \tag{36}$$

The aliasing sets $\mathcal{A}(\ell)$ are indexed by the *primary* modes, $\ell{=}(\ell_1,\ell_2)$, which are determined by the resolution of 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 observations $y^{\mathcal{A}\{\mathrm{M}\}}(x,t_n)$ can be represented in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\backslash\{0\}}$ as

$$Y^{\mathcal{A}\{\mathrm{M}\}}_n = H_{\mathcal{A}\{\mathrm{M}\}} U_n + \eta_n, \qquad n \in \mathbb{N}, \quad \eta_n \sim \mathcal{N}(0,\tilde{\Gamma}), \tag{37}$$

where $H_{\mathcal{A}\{\mathrm{M}\}}$ is a linear rank $(2M)^2$ operator which, in contrast to the setup of §3.3.1, is not diagonal.

In the numerical tests in §5 we consider two distinct filtering configurations with aliased observations which depend on the choice of the spectral resolution, $(2N)^2$, of the forward model (cf. §3.2) relative to the spectral resolution, $(2M)^2$, of the observation grid. These two scenarios correspond to

- **(i)** $\boldsymbol{N = M}$. In this configuration only the primary modes, determined by the resolution of the observation grid, are estimated from the aliased observations (34). Here, in contrast to (32), observations of the primary modes are corrupted by both the aliased 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 to provide estimates on modes beyond the spectral resolution of the observation

grid [11, 65, 35, 67]. Superresolution with SPEKF-type algorithms was studied in [11] for a variety of one-dimensional PDE dynamics; below, we extend this to the case of $2D$ Navier-Stokes dynamics.

4. Filtering algorithms

The general framework for data assimilation exploiting approximate Gaussian filters (cf. §3) admits various algorithms which all utilise the same *analysis* update (25)-(26). Important differences between these algorithms appear at the *prediction* step (21) which is reflected in the approximation Φ_Δ of the flow map Ψ_Δ in (23) and the update of the prior covariance $\widehat{C}$ in (25)-(26). In the numerical tests discussed in §5, we consider three approximate Gaussian filtering algorithms which are described below.

4.1. 3DVAR.

This algorithm has its origin in weather forecasting [60] and it is prototypical of many approximate Gaussian filters used in practice when dealing with high-dimensional estimation problems. Recall that the analysis update in (25)-(26) requires the knowledge of the prior mean and covariance, $\widehat{m}_n$, $\widehat{C}_n$, $n \in \mathbb{N}^+$. In high-dimensional nonlinear problems, such as those arising in filtering nonlinear PDEs, brute-force attempts at updating the prior covariance quickly become computationally intractable. The simplest approximation which drastically reduces the computational cost is to assume $\widehat{C}_n = const.$ in (25)-(26); the most common approach to parameterising the prior covariance is

$$\widehat{C}_n = \widehat{C}_{\alpha,\beta} := \alpha + \beta C_0, \qquad \alpha, \beta > 0, \tag{38}$$

where C_0 is the so-called *background covariance* (e.g., [46]) is estimated empirically from the attractor dynamics 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., $\Phi_\Delta^N = \Psi_\Delta$ in (23), provided that appropriate inflation is used to weigh the observations in favour of the model [54]; theoretical results explaining the numerical evidence in the case of 3DVAR may be found in [12]. $\widehat{C}_{\alpha,\beta}$ 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 (25)-(26), the 3DVAR algorithm [60] for the mean estimate is described by

$$m_{n+1} = (I - K)\Phi_\Delta^N(m_n) + K y_{n+1}, \tag{39}$$

where the gain operator K in (27) simplifies to

$$K = \widehat{C}_{\alpha,\beta} H^* (\Gamma + H \widehat{C}_{\alpha,\beta} H^*)^{-1}. \tag{40}$$

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). The case corresponding to non-aliased observations (§3.3.1), was studied in [12, 9] where the diagonal observation operator $H = P_M$ projecting onto $\{\psi_k\}_{0<|k_{1,2}|\leqslant M}$ was assumed to commute with the diagonal operators

$$\widehat{C}_{0,\beta}\,\psi_k \propto \eta^2\,|k|^{-2\zeta}\psi_k, \qquad \Gamma\,\psi_k \propto |k|^{-2\gamma}\psi_k, \tag{41}$$

so that the gain in (40) is

$$K\psi_k \propto \begin{cases} 0\cdot\psi_k & \text{for } |k_{1,2}| > M, \\ \eta^2|k|^{2(\gamma-\zeta)}\big(1+\eta^2|k|^{2(\gamma-\zeta)}\big)^{-1}\,\psi_k & \text{for } |k_{1,2}| \leqslant M. \end{cases} \tag{42}$$

Thus, increasing η corresponds to variance inflation which results in weighting the estimates in favour of observations on the resolved modes, as can be seen from (39). In the numerical tests of §5 we consider

3DVAR filtering with both the aliased and non-aliased observations, keeping in mind that the case of non-aliased observations provides analytical simplifications at the expense of abandoning realistic constraints. A more general case was considered subsequently in [69].

4.2. **Stochastic parameterisation Kalman filters (SPEKF).** There are two logical steps beyond the 3DVAR algorithm within the approximate Gaussian framework. One is to use approximate dynamics in order to make the update of the prior covariance in (25)-(26). For example, 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 Kalman filter algorithm at the expense of a potential filter divergence [58, 34, 49]. In high-dimensional state estimation problems, covariance updates become computationally expensive even in the case of ExKF. A class of so-called Ensemble Kalman filters (EnKF), which proved popular and reliable in applications (e.g., [46]), reduces this computational complexity by estimating the prior covariance form a finite ensemble of predictions propagated by the (nonlinear) flow map Φ_Δ (e.g., [52]). Another approximation - one we exploit here - relies on constructing a reduced stochastic model of the forward dynamics which is generally non-Gaussian but linear in 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 (25)-(26). This so-called SPEKF approach originated from [33, 35, 67, 66, 28, 27], where it was demonstrated that such a simple modelling of the forward dynamics can be effective for filtering chaotic systems in high dimension. Further extensions in [10, 11] showed the efficacy of this approach for super-resolving one-dimensinoal PDE dynamics from aliased observations. We first provide a formal derivation of a general family of reduced stochastic models which can be used in the SPEKF framework and then describe how this approach may be used in filtering turbulent regimes of the Navier-Stokes dynamics.

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 (36) 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

$$\frac{du_{\mathcal{A}}}{dt} + P_{\mathcal{A}}\mathcal{L}P_{\mathcal{A}}u_{\mathcal{A}} + P_{\mathcal{A}}\mathcal{B}(u,u) = P_{\mathcal{A}}f, \qquad u_{\mathcal{A}} \in \mathcal{H}_{\mathcal{A}} \subseteq \mathcal{H}, \quad u \in \mathcal{H}, \tag{43}$$

where $\mathcal{L}$ and $\mathcal{B}$ were defined in §2.1, and we skip the explicit dependence on the primary wavenumbers ℓ indexing the aliasing sets $\mathcal{A}$ in (36). The nonlinear term $P_{\mathcal{A}}\mathcal{B}$ in (43), 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 [11] is that the projection of the nonlinear interactions on the aliasing set vanish, i.e.,

$$P_{\mathcal{A}}\,\mathcal{B}\big(u_{\mathcal{A}},u_{\mathcal{A}}\big) = 0, \quad \text{or} \quad \langle u_{\mathcal{A}}, \mathcal{B}(u_{\mathcal{A}},u_{\mathcal{A}})\rangle = 0, \tag{44}$$

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 [51], namely

$$-P_{\mathcal{A}}\Big(\mathcal{B}\big(u_{\mathcal{A}},(1-P_{\mathcal{A}})u\big) + \mathcal{B}\big((1-P_{\mathcal{A}})u,u_{\mathcal{A}}\big)\Big) \approx -\big(\Gamma_{\mathcal{A}}(t) + i\,\Omega_{\mathcal{A}}(t)\big)u_{\mathcal{A}}, \tag{45}$$

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

$$P_{\mathcal{A}}\,\mathcal{B}\big((1-P_{\mathcal{A}})u,(1-P_{\mathcal{A}})u\big)dt \approx -B_{\mathcal{A}}(t)dt - \Sigma_{\mathcal{A}}dW_{\mathcal{A}}(t), \tag{46}$$

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 with quadratic nonlinearities, including the Navier-Stokes equation (8). The idea for replacing nonlinear interactions between spectral modes by multiplicative stochastic damping/frequency corrections and additional stochastic forcing arises from stochastic modelling of shear turbulence [74, 21]. The resulting stochastic approximation has the form of a linear SPDE on $\mathcal{H}_{\mathcal{A}}$

$$du_{\mathcal{A}} = \Big(-\big(\mathcal{L}_{\mathcal{A}} + \Gamma_{\mathcal{A}}(t) + i\,\Omega_{\mathcal{A}}(t)\big)u_{\mathcal{A}} + B_{\mathcal{A}}(t) + f_{\mathcal{A}}\Big)dt + \Sigma_{\mathcal{A}} dW_{\mathcal{A}}(t), \qquad u_{\mathcal{A}}(0) \in \mathcal{H}_{\mathcal{A}}, \tag{47}$$

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 [57, 29, 27]. In computations, when finite-dimensional approximations of the dynamics are employed, this strategy allows for propagating the second-order statistics in (25)-(26) based on analytical formulas which can be utilised in a number of different approximate Gaussian filtering algorithms outlined below (see [28, 10, 11] 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

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \tag{48}$$

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 (36) 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 (47) 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:

$$\begin{aligned}
&(a) \quad \mathrm{d}u_k(t) = \Big[-\big(\bar{l}_k + \gamma_k(t) + i\omega_k(t)\big)u_k(t) + b_k(t) + f_k(t)\Big]\mathrm{d}t + \sigma_{u_k}\mathrm{d}W_{u_k}(t),\\
&(b) \quad \mathrm{d}\gamma_k(t) = -d_{\gamma_k}\gamma_k(t)\mathrm{d}t + \sigma_{\gamma_k}\mathrm{d}W_{\gamma_k}(t),\\
&(c) \quad \mathrm{d}\omega_k(t) = -d_{\omega_k}\omega_k(t)\mathrm{d}t + \sigma_{\omega_k}\mathrm{d}W_{\omega_k}(t),\\
&(d) \quad \mathrm{d}b_k(t) = \big[(-d_{b_k} + i\omega_{b_k})b_k(t)\big]\mathrm{d}t + \sigma_{b_k}\mathrm{d}W_{b_k}(t),
\end{aligned} \tag{49}$$

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. Importantly, in contrast to (8) or (31), the statistics of (49) is exactly solvable (see [28, 11, 64]) and the covariance of the augmented state has a block-diagonal structure when grouped into sets $\mathfrak{U}^\ell = \{u_k, \gamma_k, \omega_k, b_k : k \in \mathcal{A}(\ell)\}$.

Remark 4.1.

- The dynamics of each spectral mode u_k in the basis $\{\psi_k\}_{k\in\mathbb{Z}^2\setminus\{0\}}$ is linear in u_k and decoupled from the other modes u_l, $l \neq k$. Thus, when filtering with the reduced forward model (49), distinct modes evolve independently during the forecast step are coupled only at the analysis step when observations are assimilated. The statistics of (49) is exactly solvable (see [28, 11, 64]) which allows for a computationally efficient propagation of the prior mean and covariance in the analysis step (25)-(26).

- Since the aliasing sets $\mathcal{A}(\ell)$ in (36) for each primary mode $|\ell_{1,2}| \leqslant M \leqslant N$ resolved by the sparse observation grid are disjoint (see [65], §3.3 and figure 1), the $(2N)^2$-dimensional filtering problem in the physical space can be converted to $(2M)^2$ independent filtering problems and efficiently parallelised. This fact was already used and validated in [33, 35, 47, 11].
- Filtering within each aliasing set $\mathcal{A}(\ell)$ involves estimation of the state vector $u_{\mathcal{A}}$ in (43) and the associated non-physical processes $(\gamma_k, \omega_k, b_k), k \in \mathcal{A}(\ell)$ which provide means for bias correction. Importantly, the dynamics of these unobserved processes is dynamically adjusted from the assimilated data, allowing the algorithm to 'learn' some aspects of model error on-the-fly from the incoming data [11].

4.2.2. *SPEKF-type algorithms.* The stochastically parameterised filtering algorithms with the forward dynamics (49) fall into the category of approximate Gaussian filters outlined in §3. Therefore, in the discrete-time setting, the analysis step in SPEKF-type algorithms is given by (25)-(26), similar to the 3DVAR filter outlined in §4.1. However, in this case the prior mean and covariance in the forecast step (21) are updated using analytical formulas ([28, 11, 64]) which leads to substantial computational gains (see Remark 4.1). In particular, the path-wise solvability of the forward model (49) yields an explicit expression for the stochastic flow $\Phi_\Delta(\cdot, \omega)$ which is employed to analytically update the prior mean and covariance on each disjoint set of augmented states $\mathfrak{U}^\ell = \{u_k, \gamma_k, \omega_k, b_k : k \in \mathcal{A}(\ell)\}$ via

$$\mathfrak{U}_n^\ell \sim \mathcal{N}\big(m_n^\ell, C_n^\ell\big) \quad \longrightarrow \quad \begin{cases} \widehat{m}_{n+1}^\ell = \mathbb{E}\big[\Phi_\Delta(\mathfrak{U}_n^\ell, \omega)\big], \\ \widehat{C}_{n+1}^\ell = Cov\big[\Phi_\Delta(\mathfrak{U}_n^\ell, \omega), \Phi_\Delta(\mathfrak{U}_n^\ell, \omega)\big]. \end{cases} \tag{50}$$

Unlike other approximate Gaussian filters, the exact statistical solvability of (49) leads to analytical formulas for $\widehat{m}_{n+1}^\ell, \widehat{C}_{n+1}^\ell$ in (50). The dimensionality of $\widehat{m}_{n+1}$ and $\widehat{C}_{n+1}$ depends on the spectral resolution $(2N)^2$ of the forward dynamics (49), and on the spectral resolution $(2M)^2$ of the observations (cf. §3.3). For $N = M$ each aliasing set $\mathcal{A}(\ell)$ contains a single mode; hence, $\widehat{m}_n \in \mathbb{C} \times \mathbb{R} \times \mathbb{R} \times \mathbb{C} \simeq \mathbb{C}^3$ and $\widehat{C}_n \in \mathbb{C}^{3\times 3}$ for each of the $(2N)^2$ estimated modes. 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 of the $(2M)^2$ aliasing sets. Given that $\{u_k, \gamma_k, \omega_k, b_k\}$ and $\{u_j, \gamma_j, \omega_j, b_j\}, k \neq j$, evolve independently in the forward dynamics (49), correlations between different augmented states decay during the forecast step (50) and can only be introduced during the analysis step when $N \neq M$. This fact can be exploited in the SPEKF framework to further reduce the computational cost of the forecast step. Details of various simplified algorithms were derived and discussed in [10, 11]; here, we recapitulate the properties of two most efficient algorithms which are be compared against 3DVAR in the next section:

- **cSPEKF**: This *crude SPEKF* algorithm utilises analytical updates to derive $\widehat{m}_{n+1}^\ell$ and the diagonal entries of $\widehat{C}_{n+1}^\ell$ in (50). The off-diagonal terms in $\widehat{C}_{n+1}^\ell$, 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, assuming rapid mode decorrelation relative to the assimilation time window. It was shown in [11] that, apart from reducing the computational cost, this approximation resulted in increased stability in a wide range of dynamical regimes.
- **GCF** : This *Gaussian Closure Filter* algorithm does not neglect cross-correlations in the prior covariance $\widehat{C}_{n+1}^\ell$ in (50) but it approximates the statistics of the forward model (49) via the simple Gaussian moment closure; this approach is used frequently in the statistical theory of turbulence and was utilised for filtering turbulent signals in [10, 11]. For systems with quadratic nonlinearities, such as (49), this closure correctly accounts for the turbulent backscatter in the mean but it neglects the third order moments of fluctuations in the evolution of the covariance. In [10, 11] this algorithm emerged as the most suitable trade-off between the skill and the computational complexity.

5. Numerical Results

In this section we compare the performance of the three approximate Gaussian filters described in §4 for estimation of the spatially extended system given by the 2D Navier-Stokes dynamics (1)-(2). We assume throughout that the posterior filtering distributions are well-defined, implying that the prior on the initial conditions, and the observation likelihood are well-defined throughout the time interval considered. The comparisons outlined below concern the ability of the filtering algorithms to reconstruct the mean of the true posterior on the state of the truth dynamics. 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 introduced in the considered approximate Gaussian filtering algorithms. Key questions driving the choice of experiments concern the following:

(i) Does updating the prior covariance $\widehat{C}_{n+1}$ in (25) in the approximate Gaussian filtering algorithms via reduced order stochastic dynamics with model error improve the estimation relative to setting $\widehat{C}_{n+1} = const.$ as in 3DVAR?

(ii) Which class of the approximate Gaussian filters - 3DVAR or cSPEKF/GCF - is better suited for superresolving sparse, aliased observations of a complex spatially extended dynamics?

The above issues are studied for the filters in question by means of numerical experiments which currently provides the only validation method, given that analytical results for sparsely observed dynamics do not currently exist. A systematic numerical study of these issues requires dynamics with varying degree of complexity. As highlighted in §2.1, the number of positive Lyapunov exponents in (8) is a proxy for the dimensionality of the unstable manifold of the attractor of (8) and for the complexity of the solutions; the larger the number of positive Lyapunov exponents the less predictable the underlying dynamics. In practice, the spectral resolution of the 'truth' dynamics is finite and given by the truncated version (29) of the infinite-dimensional system (8). However, a similar dependence between the attractor dimensionality and the complexity of long-time solutions is observed in the truncations (e.g., [62]). In the examples discussed below the synthetic truth is computed from (29) with $\Lambda = 232$, and the forward model resolution, and the resolution of the observations are arranged so that $M \leqslant N \ll \Lambda$. The numerical simulation of the dynamics in (29) is carried out in a standard fashion which was outlined in §3.1. 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 various dynamical regimes 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 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 (see figures 6-8 for representative snapshots of the vorticity fields). In these three dynamical regimes we compare the performance of the filtering algorithms introduced in §4 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 (29)). The configuration corresponding to filtering in the idealised configuration with non-aliased observations (cf. §3.3.1) is considered first. Then, we consider the state estimation using the same filtering algorithms with aliased observations; this configuration allows to superresolve the observations and study some effects of representation error (cf. §3.3.2).

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) is given by the L^2 norm of the residual $u-m_N$ in the space $\mathcal{H}_\Lambda \times \mathcal{I}$

$$\text{RMS}(u, m_N) = \|u - m_N\|_{L^2(\mathcal{H}_\Lambda \times \mathcal{I})} := \Lambda^{-1}|\mathcal{I}|^{-1/2}\Big(\sum_{i\in\mathcal{I}}\sum_{j,k=-\Lambda}^{\Lambda}\big(u(x_{j,k},t_i) - m_N(x_{j,k},t_i)\big)^2\Big)^{1/2}, \tag{51}$$

where $m_N \in \mathcal{H}_N \subseteq \mathcal{H}_\Lambda \subseteq \mathcal{H}$[1]. Similarly, the RMS error on $\mathcal{H}_N \times \mathcal{I}$ is defined as

$$\text{RMS}_N(u, m_N) = \|P_N u - m_N\|_{L^2(\mathcal{H}_N \times \mathcal{I})} := N^{-1}|\mathcal{I}|^{-1/2}\Big(\sum_{i\in\mathcal{I}}\sum_{j,k=-N}^{N}\big(u(x_{j,k},t_i) - m_N(x_{j,k},t_i)\big)^2\Big)^{1/2}, \tag{52}$$

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

$$\text{XC}(u, m_N) := \frac{\langle u, m_N\rangle_{\mathcal{H}_\Lambda\times\mathcal{I}}}{\|u\|_{L^2(\mathcal{H}_\Lambda\times\mathcal{I})}\|m_N\|_{L^2(\mathcal{H}_\Lambda\times\mathcal{I})}} \propto \sum_{i\in\mathcal{I}}\sum_{j,k=-\Lambda}^{\Lambda} u(x_{j,k},t_i)\, m_N(x_{j,k},t_i), \tag{53}$$

and

$$\text{XC}_N(u, m_N) := \frac{\langle P_N u, m_N\rangle_{\mathcal{H}_N\times\mathcal{I}}}{\|P_N\, u\|_{L^2(\mathcal{H}_N\times\mathcal{I})}\|m_N\|_{L^2(\mathcal{H}_N\times\mathcal{I})}} \propto \sum_{i\in\mathcal{I}}\sum_{j,k=-N}^{N} u(x_{j,k},t_i)\, m_N(x_{j,k},t_i). \tag{54}$$

Clearly, $\text{RMS}(u, m_N)$ and $\text{XC}(u, m_N)$ quantify the error in filter estimates relative to the truth solution, while $\text{RMS}_N(u, m_N)$, $\text{XC}_N(u, m_N)$ quantify the filter error on the N modes resolved by the forward model.

In order to assure a consistent comparison, the algorithms are tuned using the same data obtained from long runs of the simulated truth dynamics (29). In the context of 3DVAR (cf. §4.1), and in line with [54, 9], the tuning entails estimating the 'background' covariance C_0 in (38) 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 (49) which are roughly estimated from the equilibrium statistics as in [10, 11]; 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 [28, 27, 10, 11]). 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 (38) which requires more than just the estimates of statistics from the training data. The well-known importance of covariance inflation in 3DVAR is highlighted 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 (33) 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, and the spectral resolution of the non-aliased observations is $M \leqslant N$. 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. We assume

[1]See §3.1, 3.2 for the definitions of the Hilbert spaces $\mathcal{H}$, $\mathcal{H}_\Lambda$ and $\mathcal{H}_N$.

that the true dissipation parameters ν and κ, as well as the forcing are known; filtering with uncertain parameters and/or forcing was deemed too technical for this exposition.

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

$$\omega(t,x) = \sum_k u_k(t)\left(\nabla^\perp\cdot\psi_k(x)\right) = \sum_k \hat{\omega}_k(t)\,\phi_k(x), \qquad \hat{\omega}_{-k} = \hat{\omega}_k^*, \tag{55}$$

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 (synthetic) 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 (51) and (53).

(ii) Estimates based on observations of the $M = N$ (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 $\mathcal{H}_M$ in terms of $\textsc{rms}_M(u, y^M)$ and $\textsc{xc}_M(u, y^M)$ in (52) and (54).

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.

Figures 4, 5 show the RMS and the correlation XC measures for filtering the attractor dynamics of (29) with different resolutions, N, of the forward models in three distinct dynamical regimes 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 $\widehat{C}_{0,\beta}$ (38) 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 $\widehat{C}_{\alpha,\beta}$ (38) has a much less pronounced effect; hence, we set $\alpha = 0$ in all examples shown. The black solid lines indicate the accuracy 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 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

$$\omega_N(t,x) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\left(\nabla^\perp\cdot\psi_k(x)\right), \tag{56}$$

while the vorticity field corresponding to the synthetic truth solving (29) is given by (56) with $N = \Lambda$. Results are shown for two spectral resolutions N of the forward models in the algorithms with a fully 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 (29) (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 Appendix A for details. The performance of the filtering algorithms is assessed for the mean estimates and expressed in terms of the RMS error (51) and pattern correlation, XC (53), for non-aliased observations and the resolution $N = 70$ of the forward models in the filtering algorithms. The observation error is set to $\varepsilon = 0.15E$ where E is the energy per mode in steady state; we set $N = M$ as the modes resolved by the models are assumed to be observable directly in the spectral domain. Results for 3DVAR are shown for the optimal value of the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (38).

We summarise the main results below:

- For non-aliased observations the state estimation with cSPEKF and GCF algorithms §4.2 provides results which either outperform or shadow those obtained with optimally tuned 3DVAR (cf. §4.1).
- In the laminar regime when a small number of modes contains significant energy (see Figure 3) cSPEKF and GCF estimation provides better results than 3DVAR with optimally inflated covariance.
- In the turbulent regime 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. For sufficiently low resolution of the forward dynamics SPEKF and GCF estimation is marginally worse than 3DVAR with optimally inflated covariance but the differences are negligible.
- All considered filters beat the corresponding estimates obtained from observations, i.e., $\mathrm{RMS}_N(u, m_N) < \mathrm{RMS}_N(u, y)$ and $\mathrm{XC}_N(u, m_N) > \mathrm{XC}_N(u, y)$ (with $N = M$); moreover, when filtering with resolution sufficiently beyond the forcing scale $\mathrm{RMS}(u, m_N) < \mathrm{RMS}(u, y)$ and $\mathrm{XC}(u, m_N) > \mathrm{XC}(u, y)$. Unsurprisingly, the accuracy of the estimates improves with the number of filtered modes N.

5.2. **Filtering with aliased observations.** In this configuration we consider estimation of $u(x, t)$ solving (29) with the filtering algorithms 3DVAR (§4.1) and cSPEKF, GCF (§4.2) given noisy observations of the state $u(x, t)$ in (37) 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 (37) 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. In all tests we assume that the spectral resolution of the observations is fixed with $M = 10$. The filtering algorithms 3DVAR §4.1, and cSPEKF, GCF in §4.2 are considered at different spectral resolutions $N = PM$, $P \in \mathbb{N}^+$. Here, it is also important to investigate if the superresolution (i.e., $P > 1$ in the forward models) helps improve the estimates of the dynamics of the primary modes.

The results below are presented in terms of the vorticity field (56) and the coefficients $\{u_k\}_{|k_{1,2}|\leqslant N}$ are estimated from noisy observations via the 3DVAR, cSPEKF, and GCF algorithms in the three dynamical regimes of (29) illustrated in Figure 3. As a reference, the accuracy 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 (37). 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 (51) and (53).

(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 $\mathrm{RMS}_M\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$ and $\mathrm{XC}_M\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$ in (52) and (54).

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 the superresolving algorithms (i.e., $N > M$); the corresponding error in the filtering estimates, $\mathrm{RMS}\big(u, m_N\big)$ and $\mathrm{XC}\big(u\,, m_N\big)$, 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 M modes of the truth state from the aliased observations; the corresponding error in the filtering estimates, $\mathrm{RMS}_M\big(u, m_N\big)$ and $\mathrm{XC}_M\big(u\,, m_N\big)$, has to be smaller than the observation error in (iv) for 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., $N = M$).

Figure 10 shows a comparison of the filtering algorithms of §4 in terms of the error in the mean estimates, using $\mathrm{RMS}\big(u, m_N\big)$ in (51) and $\mathrm{XC}\big(u\,, m_N\big)$ in (53) 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 $\widehat{C}_{0,\beta}$ (38) which is indicated on the horizontal axes in the insets. The black dotted lines indicate the accuracy of estimates obtained directly from the aliased noisy observations of the truth (see (iii) above), the black dash-dotted lines indicate the accuracy 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 - $\mathrm{RMS}\big(u, y^M\big)$, $\mathrm{XC}\big(u, y^M\big)$ with y^M in (33) - represent the gold (though unachievable) reference standard, while the observation errors based on the aliased observations of the truth - $\mathrm{RMS}\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$, $\mathrm{XC}\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$ - 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 and 12 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 13 shows a comparison of the accuracy 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 $\mathrm{RMS}_M\big(u, m_N\big)$ in (52) and $\mathrm{XC}_M\big(u\,, m_N\big)$ in (54). Aliased observations of the truth dynamics (29) 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 $\widehat{C}_{0,\beta}$ (38) which is indicated on the horizontal axes in the insets. The black dotted lines indicate the quality of estimates $\mathrm{RMS}_M\big(u, y^{\mathcal{A}\{\mathrm{M}\}}\big)$,

$\text{XC}_M\left(u, y^{\mathcal{A}\{\text{M}\}}\right)$ obtained directly from the aliased noisy observations of the truth (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.

We summarise the main findings of this section below:

- cSPEKF and GCF algorithms significantly outperform the 3DVAR algorithm when filtering aliased observations and mitigating the associated representation error. This is most pronounced in the fully turbulent regime (see Figures 10, 12). Moreover, cSPEKF/GCF do not require covariance inflation.
- Superresolution of aliased observations improves the estimates of the primary modes when filtering with cSPEKF and GCF but it is detrimental for 3DVAR (Figure 13).
- For aliased observations the state estimation with SPEKF-type algorithms (cf. §4.2) benefits from the stochastic parameterisation of model error which mitigates model rigidity present in the 3DVAR with the forward dynamics based on the truncation (31). In contrast to the filtering with non-aliased observations, the aliased observations retain some information about the unobserved modes $|k_{1,2}| > M$ which is then accounted for in the estimates of the superresolved modes due to the fact that $H_{\mathcal{A}\{\text{M}\}}$ in (37) is not diagonal and, consequently, the gain (27) is non-zero on the unobserved spectral modes.
- Superresolution with cSPEKF 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 accuracy based on the RMS and XC in the spatial domain.

6. Conclusions

Data assimilation algorithms are important for improving predictive performance of simulations in many geoscience and engineering applications. However, incorporating sparse noisy data into uncertain computational models in a way which actually improves the overall performance poses major challenges, especially as prediction is pushed to increasingly longer time horizons. In this paper we studied the performance of three approximate Gaussian data assimilation algorithms: the prototypical 3DVAR, and two stochastically parameterised algorithms SPEKF and GCF. The emphasis was on the interplay between different sources of error in a realistic but nevertheless 'academic' setting, rather than on tests including all operational constraints. We provided the first evidence that the computationally cheap stochastically parameterised filtering algorithms are capable of overcoming model error in the forward dynamics and mitigate some representation errors to produce accurate mean estimates in realistic models of turbulent dynamics. In particular, we found that SPEKF and GCF outperform optimally tuned 3DVAR algorithms in various regimes of 2D Navier-Stokes dynamics which was sparsely observed in the physical domain. We summarise the main findings below:

(i) For noisy observations of individual spectral modes (i.e., idealised case with no representation error in our setup) SPEKF/GCF and optimally tuned 3DVAR perform well in reproducing the mean of the posterior filtering distribution in various regimes of the dissipative 2D Navier-Stokes dynamics. However, in contrast to 3DVAR, SPEKF and GCF do not rely on the hind-cast covariance inflation.

(ii) For spatially sparse observations (leading to representation error in our setup via mode aliasing in the spectral domain) the SPEKF-type algorithms perform significantly better than a tuned 3DVAR.

These conclusions are intrinsic to the considered algorithms, and result from the approximations made in order to create tractable online implementations; the basic conclusions are not expected to change by use of different dynamical models. Here, we focused on comparing the accuracy of predictions for the mean

state rather than the underlying posterior probability distribution. Uncertainty of the mean estimates was not considered because of a very high computational cost which would require Markov Chain sampling, or particle filtering of multimodal densities over high-dimensional spectral domains and long time windows.

There are many possible directions for future research in this area which require attention. First, we note that the ability of various data assimilation algorithms to predict uncertainty from a fully Bayesian perspective was considered in [54] in the absence of model error and for non-aliased observations. In that work MCMC sampling was used 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 computations in [54] were carried out in regimes of the 2D Navier-Stokes dynamics which were chaotic but characterised by 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 distributions. An analogous study in the turbulent regimes considered here poses 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 studied in [54]. 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 [54].

We note that comparisons of a number of other approximate Gaussian estimators have been carried out recently for variants of 3DVAR, 4DVAR, ExKF and EnKF. In particular, [68, 86, 87] compare the EnKF forecast with 3DVAR and 4DVAR in real-data experiments based on the Weather Research and Forecasting model (WRF). The conclusions are that EnKF and 4DVAR perform best with respect to the root-mean-square error (RMS), while the EnKF forecast performs better for longer lead times. Two fundamental classes of EnKFs were compared theoretically in the large ensemble limit in [56], where the stochastic version in which observations are randomised is found to be more robust to perturbations in the forecast distribution than the deterministic one. Another interesting comparison was carried out in [32] in which several ensemble filters, alternative to EnKF in operational use, were compared with respect to RMS. A numerical comparison of of the performance of SPEKF algorithms with these filters deserves a separate study and will be soon reported elsewhere. Theoretical results, analogous to those for 3DVAR in [12, 9], explaining the properties of the SPEKF-type algorithms are needed. Such analysis 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 even in the absence of spatially sparse observations. This research is currently ongoing and we hope to report on this in the near future.

It would be interesting to conduct a study, similar to the one undertaken here, for simple models of atmospheric dynamics exhibiting behaviour analogous to atmospheric blocking events, or for more realistic quasigeostrophic models which admit baroclinic instabilities [66, 47]. Also, with recent progress in consistent multilevel Monte Carlo (MLMC) sampling algorithms [39, 7, 42], it may soon be possible 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 or perhaps even Navier-Stokes. Then, a study may be performed along the lines of [54] 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 (29). In the context of 3DVAR (cf. §4.1), and in line with [54, 9], the tuning entails estimating the 'background' covariance C_0 in (38) 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 [54], 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 (38) 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 (49) which are roughly estimated from the equilibrium statistics as in [10, 11]; 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 [28, 27, 10, 11]). 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

$$u_N(x,t) = \sum_{0<|k_{1,2}|\leqslant N} u_k(t)\psi_k(x), \tag{57}$$

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

$$dU = -MUdt + \sqrt{2\Re e[M]\Xi}\, dW_t, \tag{58}$$

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 (58) is a Gaussian process with mean zero and covariance Ξ which are tuned to the truth via

$$\Xi = \lim_{T\to\infty} \frac{1}{T}\int_0^T [u(t)-\bar{u}]\otimes[u(t)-\bar{u}]^* dt, \qquad \bar{u} = \lim_{T\to\infty}\frac{1}{T}\int_0^T u(t)dt. \tag{59}$$

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

$$Corr_k(\tau) = \lim_{T\to\infty} \Xi_{k,k}^{-2}\int_0^T C_{k,k}(t,\tau)dt, \qquad C(t,\tau) = [u(t-\tau)-\bar{u}]\otimes[u(t)-\bar{u}]^*, \tag{60}$$

and

$$T_k + i\Theta_k = \int_0^\infty Corr_k(\tau)d\tau, \qquad \Re e[M_{k,k}] = \frac{T_k}{T_k^2+\Theta_k^2}, \quad \Im m[M_{k,k}] = -\frac{\Theta_k}{T_k^2+\Theta_k^2}. \tag{61}$$

In practice the integrals in (59), (60) 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

(38) which requires more than just the estimates of statistics from the training data. The two parameters estimated from data in the SPEKF forward model (49) are set as $\bar{l}_k = M_{k,k}$, $\sigma^2_{u_k} = 2\Re e[M_{k,k}]\,\Xi_{k,k}$, and the remaining parameters are set as $d_{\gamma_k} = d_{b_k} = 0.1\,\Re e[\bar{l}_k]$, $d_{\omega_k} = \omega_{b_k} = 0.1\,\Im m[\bar{l}_k]$, $\sigma_{\gamma_k} = \sigma_{\omega_k} = \sigma_{b_k} = 0.6\,\sigma_{u_k}$.

In order to systematically estimate the background covariance in 3DVAR we first note that the discrete-time solution of the OU process in (58) is given by the linear stochastic map

$$U_{n+1} = LU_n + \sqrt{Q}\,\xi_n, \tag{62}$$

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 (62) the update map $(m_n, C_n) \to (m_{n+1}, C_{n+1})$ in (21) of §3 yields the Kalman filter with

$$\widehat{m}_{n+1} = Lm_n, \qquad \widehat{C}_{n+1} = LC_nL^* + Q. \tag{63}$$

In the spirit of 3DVAR, the above update can be improved by updating the covariance as in (63) and updating the mean by the nonlinear flow map corresponding to (31), namely

$$\widehat{m}_{n+1} = \Phi^N_\Delta(m_n), \qquad \widehat{C}_{n+1} = LC_nL^* + Q, \tag{64}$$

$$m_{n+1} = (I - K_{n+1}H)\widehat{m}_{n+1}, \qquad C_{n+1} = (I - K_{n+1}H)\widehat{C}_{n+1}, \tag{65}$$

$$K_{n+1} = \widehat{C}_{n+1}H^*(H\widehat{C}_{n+1}H^* + \Gamma)^{-1}. \tag{66}$$

We note that because L is diagonal with $LL^* < 1$, the covariance C_n converges to a limit [52] that can be computed numerically off-line and, asymptotically the algorithm behaves like 3DVAR; thus, in line with [54] this asymptotic covariance is used as the systematic choice of background covariance $\widehat{C}_0$ in (38). Alternatively, one may set $\widehat{C}_0 = \Xi$ which corresponds to the update (64) 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] M. Ades and P.J. Van Leeuwen. The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141(687):484–503, 2015.

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

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

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

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

[6] A. Beskos, D. Crisan, A. Jasra, K. Kamatani, and Y. Zhou. A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, 49(1):24–48, 2017.

[7] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou. Multilevel sequential monte carlo samplers. *Stochastic Processes and their Applications*, 2016.

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

[9] 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.

[10] 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.

[11] M. Branicki and A.J. Majda. Dynamic stochastic superresolution of sparsely observed turbulent systems. *J. Comp. Phys*, 241:333–363, 2013.

[12] Ch. Brett, K. F. Lam, K. Law, D.S. McCormick, M. R Scott, and A.M. Stuart. Accuracy and stability of filters for dissipative pdes. *Physica D*, 245(1):34–45, 2013.

[13] G. L. Bretthorst. *Maximum Entropy and Bayesian Methods*, chapter Nonuniform sampling: Bandwidth and aliasing. Kluwer, 2000.

[14] 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.

[15] 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*, 18:023112, 2008.

[16] A. Chorin, M. Morzfeld, and X. Tu. Implicit particle filters for data assimilation. *Comm. in Applied Math and Comp. Sci*, page 221, 2010.

[17] P. Constantin and C. Foias. *Navier-Stokes Equations*. Chicago Lectures in Math. Chicago, 1988.

[18] P. Constantin, C. Foias, I. Kukavica, and A.J. Majda. Dirichlet quotients and 2D periodic Navier-Stokes equations. *J. Math. Pures Appl.*, 76:125–153, 1997.

[19] 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.

[20] S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *J. Comput. Phys.*, 176(2):430–455, 2002.

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

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

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

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

[25] C. Foias and G. Prodi. Sur le comportement global des solutions non-stationnaires des équations de Navier-Stokes en dimension 2. *Rend. Semin. Mat. U Pad.*, 39:1–34, 1967.

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

[27] B. Gershgorin, J. Harlim, and A.J. Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comp. Phys*, 229(1):32–57, 2010.

[28] B. Gershgorin, J. Harlim, and A.J. Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comp. Phys*, 229(1):1–31, 2010.

[29] B. Gershgorin and A.J. Majda. Filtering a statistically exactly solvable test model for turbulent tracers from partial observations. *J. Comp. Phys*, 230(4):1602–1638, 2011.

[30] R. Ghanem, D. Higdon, and H. Owhadi. *Handbook of Uncertainty Quantification*. Springer International Publishing, 2017.

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

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

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

[34] J. Harlim and A.J. Majda. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm. Math. Sci.*, 8(1):27–43, 2010.

[35] J. Harlim and A.J. Majda. Filtering turbulent sparsely observed geophysical flows. *Mon. Wea. Rev*, 138(4):1050–1083, 2010.

[36] A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman filter*. CUP, 1991.

[37] K. Hayden, E. Olson, and E.S. Titi. Discrete data assimilation in the Lorenz and 2D Navier–Stokes equations. *Physica D*, 240(18):1416–1425, 2011.

[38] J.S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral Methods for Time-Dependent Problems*, volume 21. Cambridge Univ Pr, 2007.

[39] V.H. Hoang, Ch. Schwab, and A.M. Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.

[40] D. Hodyss and N. Nichols. The error of representation: Basic understanding. *Tellus A: Dynamic Meteorology and Oceanography*, 67(1):24822, 2015.

[41] 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.

[42] A. Jasra, K. Kamatani, K.J.H. Law, and Y. Zhou. Multilevel particle filter. *arXiv:1510.04977*.

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

[44] 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.

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

[46] E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. CUP, 2003.

[47] S.R. Keating, A.J. Majda, and K.S. Smith. New methods for estimating poleward eddy heat transport using satellite altimetry. *Mon. Wea. Rev.*, 140:1703–1722, 2012.

[48] D. Kelly, K.J.H. Law, and A.M. Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579, 2014.

[49] D. Kelly, A.J. Majda, and X.T. Tong. Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proc. Natl. Acad. Sci*, 112(34):10589–10594, 2015.

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

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

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

[53] KJH Law, D Sanz-Alonso, A Shukla, and AM Stuart. Filter accuracy for the lorenz 96 model: Fixed versus adaptive observation operators. *Physica D: Nonlinear Phenomena*, 325:1–13, 2016.

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

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

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

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

[58] 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.

[59] F.P. Llopis, N. Kantas, A. Beskos, and A. Jasra. Particle Filtering for Stochastic Navier-Stokes Signal Observed with Linear Additive Noise. *arXiv preprint arXiv:1710.04586*, 2017.

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

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

[62] A. Majda and X. Wang. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press, 2006.

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

[64] A.J. Majda and M. Branicki. Lessons in Uncertainty Quantification for Turbulent Dynamical Systems. *Discrete Contin. Dynam. Systems*, 32(9)(9):3133–3231, 2012.

[65] 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.

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

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

[68] 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.

[69] A. Moodey, A. Lawless, R. Potthast, and P.J. Van Leeuwen. Nonlinear error dynamics for cycled data assimilation methods. *Inverse Problems*, 29(2):025002, 2013.

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

[71] J. Poterjoy, R.A. Sobash, and J.L. Anderson. Convective-scale data assimilation for weather research and forecasting model using the local particle filter. *Mon. Wea. Rev*, 145(5):1897–1918, 2017.

[72] P. Rebeschini, R. Van Handel, et al. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.

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

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

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

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

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

[78] R. Temam. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, volume 68 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1997.

[79] R. Temam. *Navier-Stokes equations: Theory and numerical analysis*. Amer Math. Society, 2001.

[80] X. T Tong, A.J. Majda, and D. Kelly. Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *Comm. Math. Sci.*, 14(5), 2016.

[81] 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.

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

[83] 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.

[84] P.J. Van Leeuwen. Particle filtering in geophysical systems. *Mon. Wea. Rev*, 137:4089–4114, 2009.

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

[86] 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.

[87] M. Zhang and F. Zhang. E4DVAR: Coupling an ensemble Kalman filter with 4D variational data assimilation in a limited-area weather prediction model. *Mon. Wea. Rev.*, 140:587–600, 2012.

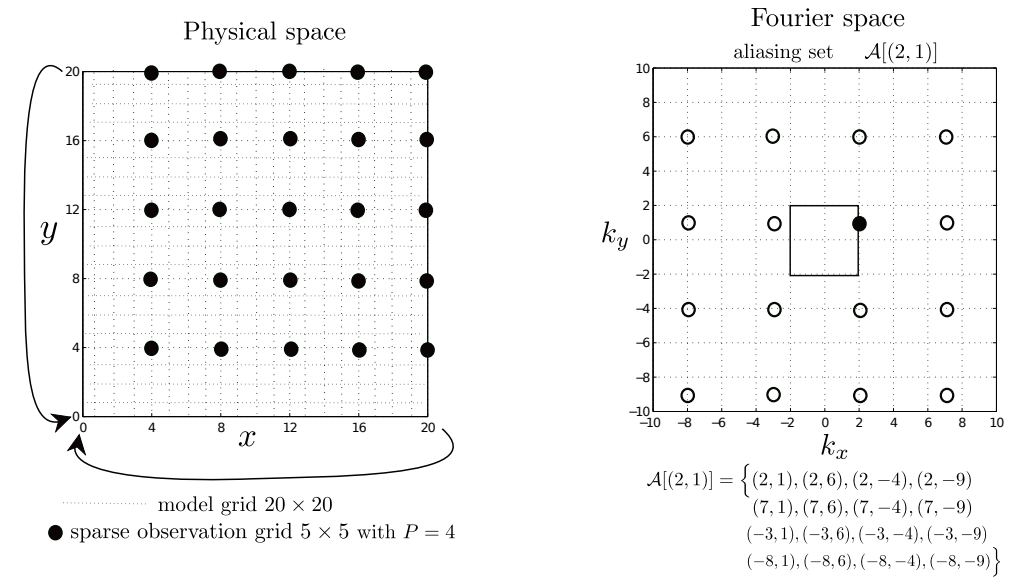

FIGURE 1. Schematics of aliasing in two dimensions (right) due to sparse nodal observations in the spatial domain (left); here, the 5×5 sparse observation grid is a regular subset of the doubly periodic 20×20 model mesh so that every $P = 4$ node is observed. The aliasing set $\mathcal{A}(\ell)$ of wavenumber $\ell = (2, 1)$ is shown in the spectral domain (right). In this case, modes with $|k_{1,2}| > 2$ are aliased into the *primary* modes $|k_{1,2}| \leqslant 2$ which can be resolved by the observation grid.

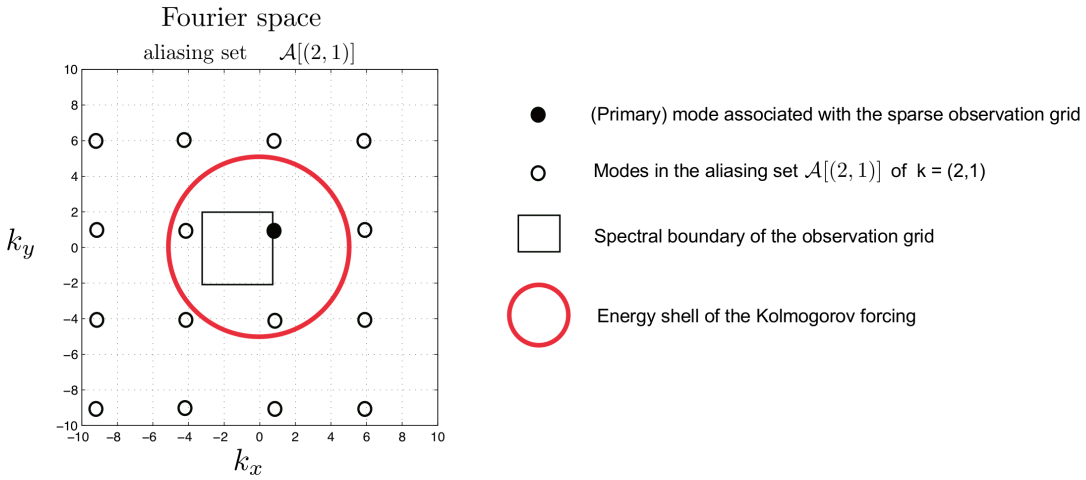

FIGURE 2. Desired test configuration for filtering NSE with Kolmogorov forcing and sparse aliased observations (in physical space). In a dynamical regime with sufficiently large Reynolds number the primary (observed) modes are not always the most energetic ones due to the (possibly intermittent) energy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

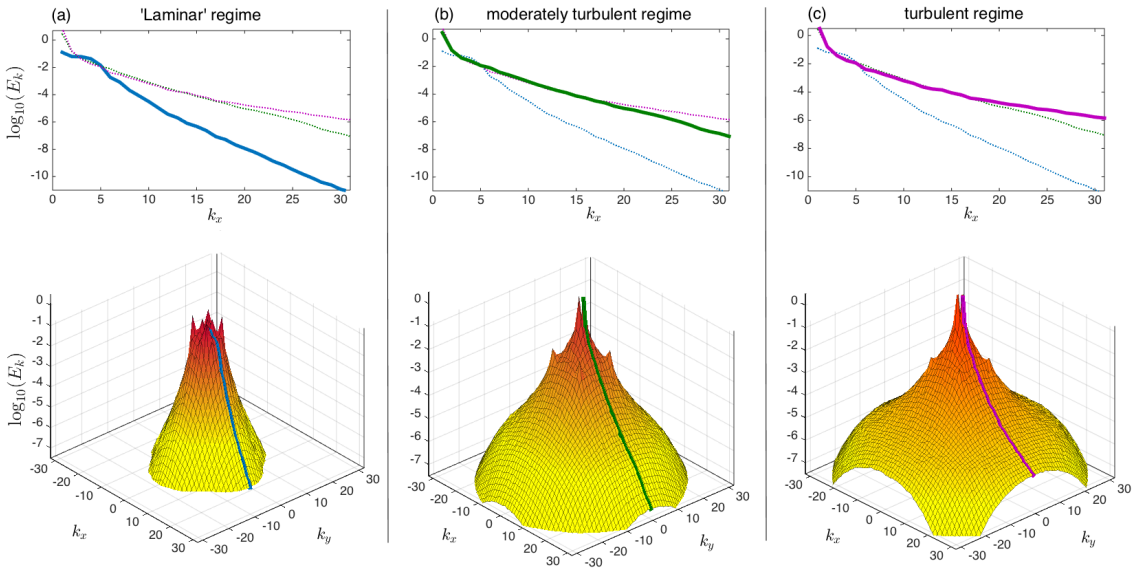

FIGURE 3. Examples of numerically simulated spectra of the truth dynamics for $u_\Lambda(x,t)$ in (29). Top row shows cross-sections for $(k_1, k_2 = 0)$ of the 2D spectra in the bottom row for three different regimes used in the numerical tests of the filtering algorithms introduced in §4. The dynamics (29) with $\Lambda = 64$ is forced at $|k_1| = |k_2| = 8$ and amplitude $|f_k| = 8$. The remaining parameters in the three regimes are: (a) $\nu = 0.1, \kappa = 0.01$, (b) $\nu = 0.01, \kappa =$, (c) $\nu = 0.001, \kappa = 0.01$.

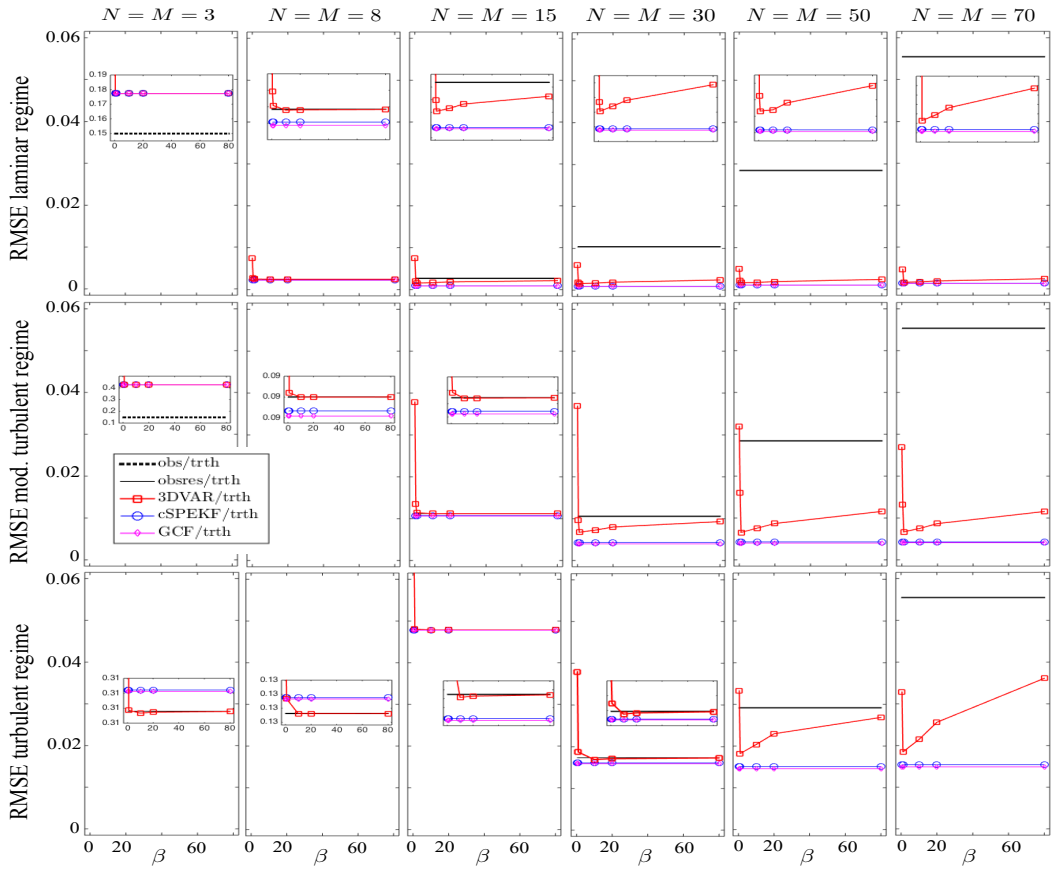

FIGURE 4. NON-ALIASED OBSERVATIONS OF (29). Comparison of performance of the filtering algorithms described in §4 in different dynamical regimes of (29) (cf. Figure 3) in terms of the RMS error (51) in the mean estimates for different resolutions N of the forward models and the observation error $\varepsilon = 0.15E$ (E is the energy per mode in steady state, $N = M$ as the modes resolved by the forward models are assumed to be observable in the spectral domain). Results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (38); this filter diverges for sufficiently small values of β but this effect is not resolved in detail. The insets show the 'performance hierarchy' when difficult to discern in the original axes and their scales differ. See figure 5 for a comparison of in terms of the XC measures (53).

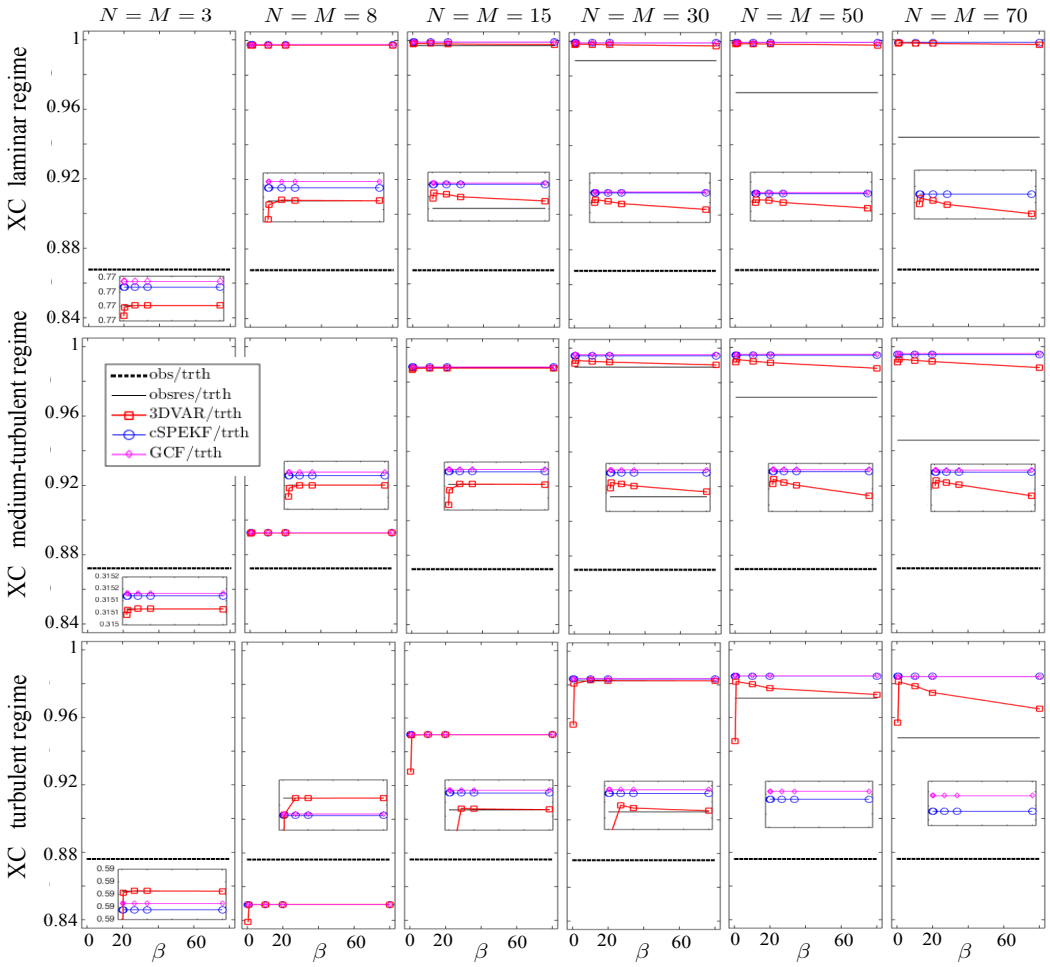

FIGURE 5. NON-ALIASED OBSERVATIONS OF (29). Comparison of performance of the filtering algorithms of §4 in different dynamical regimes of (29) (cf. Figure 3) using XC measures (53) of error in the mean estimates. Results are shown for different resolutions N of the forward models, and the observation error $\varepsilon = 0.15E$; results for 3DVAR depend on the multiplicative covariance inflation parameter β in $\widehat{C}_{0,\beta}$ (38); see figure 4 for the comparison in terms of the RMS measures. The insets show the 'performance hierarchy' when difficult to discern in the original axes and their scales differ.

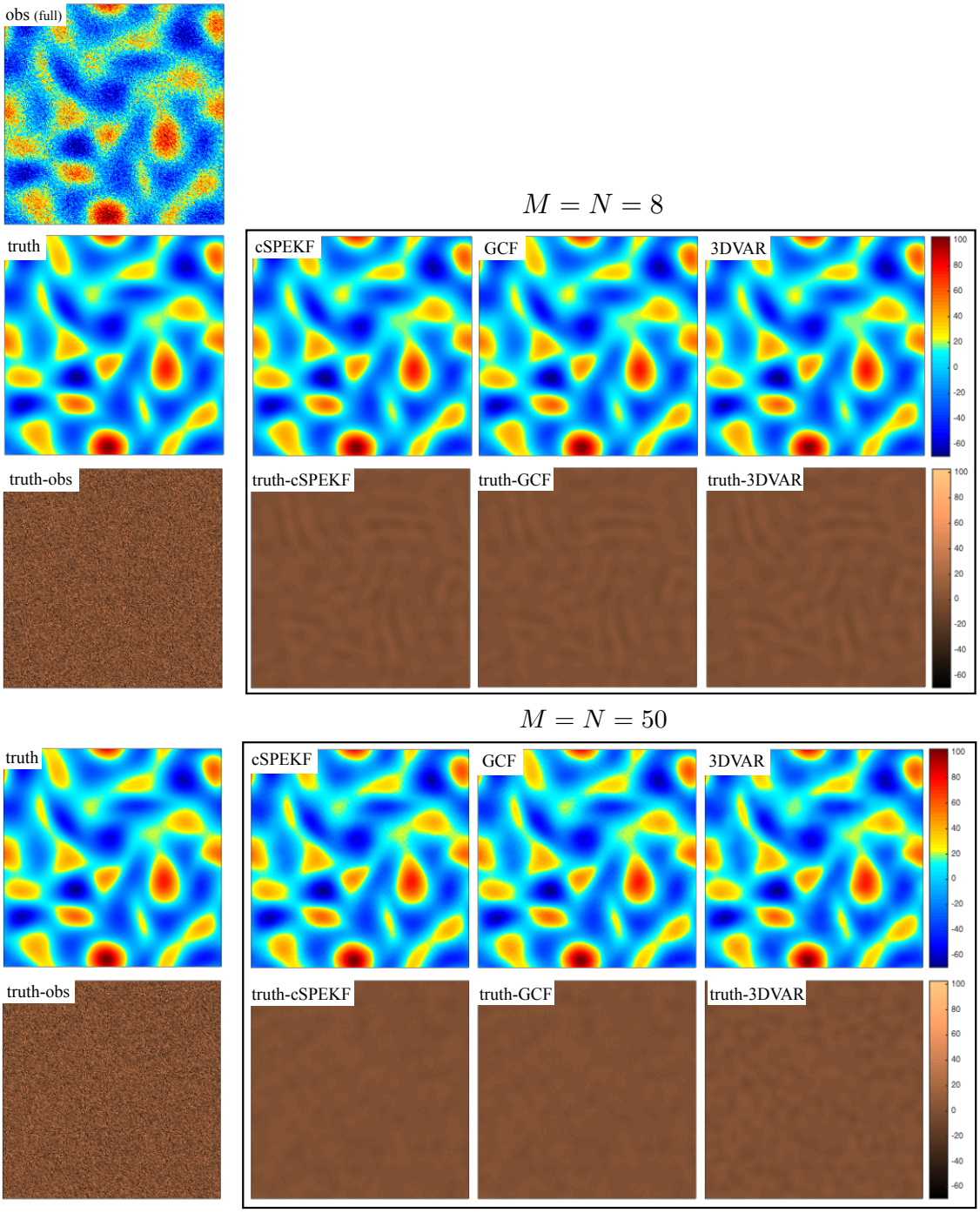

FIGURE 6. FILTERING WITH NON-ALIASED OBSERVATIONS; LAMINAR REGIME OF (8)/(29) (cf. Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the fully 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.

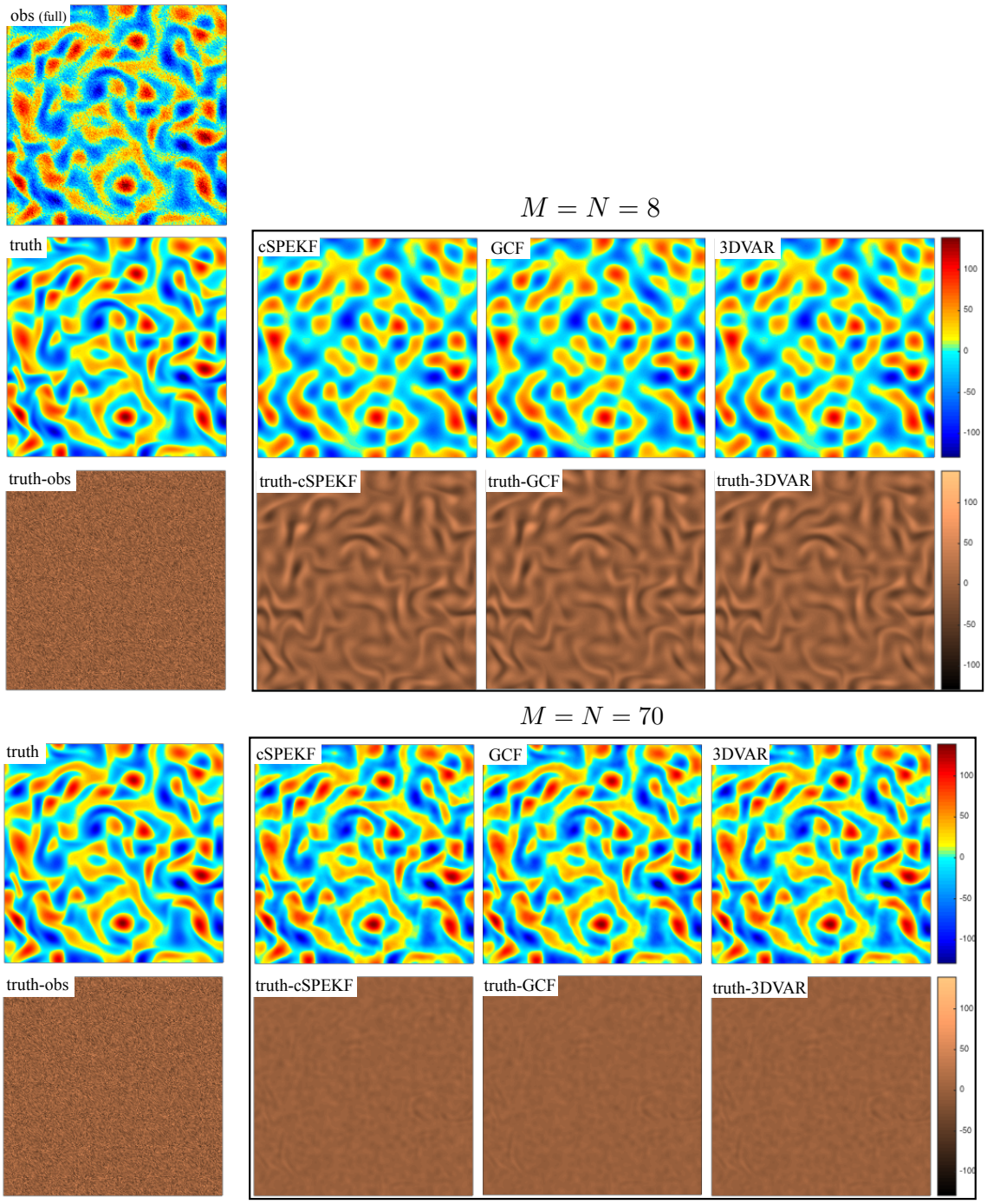

FIGURE 7. FILTERING WITH NON-ALIASED OBSERVATIONS; MODERATELY TURBULENT REGIME OF (8)/(29) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates 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.

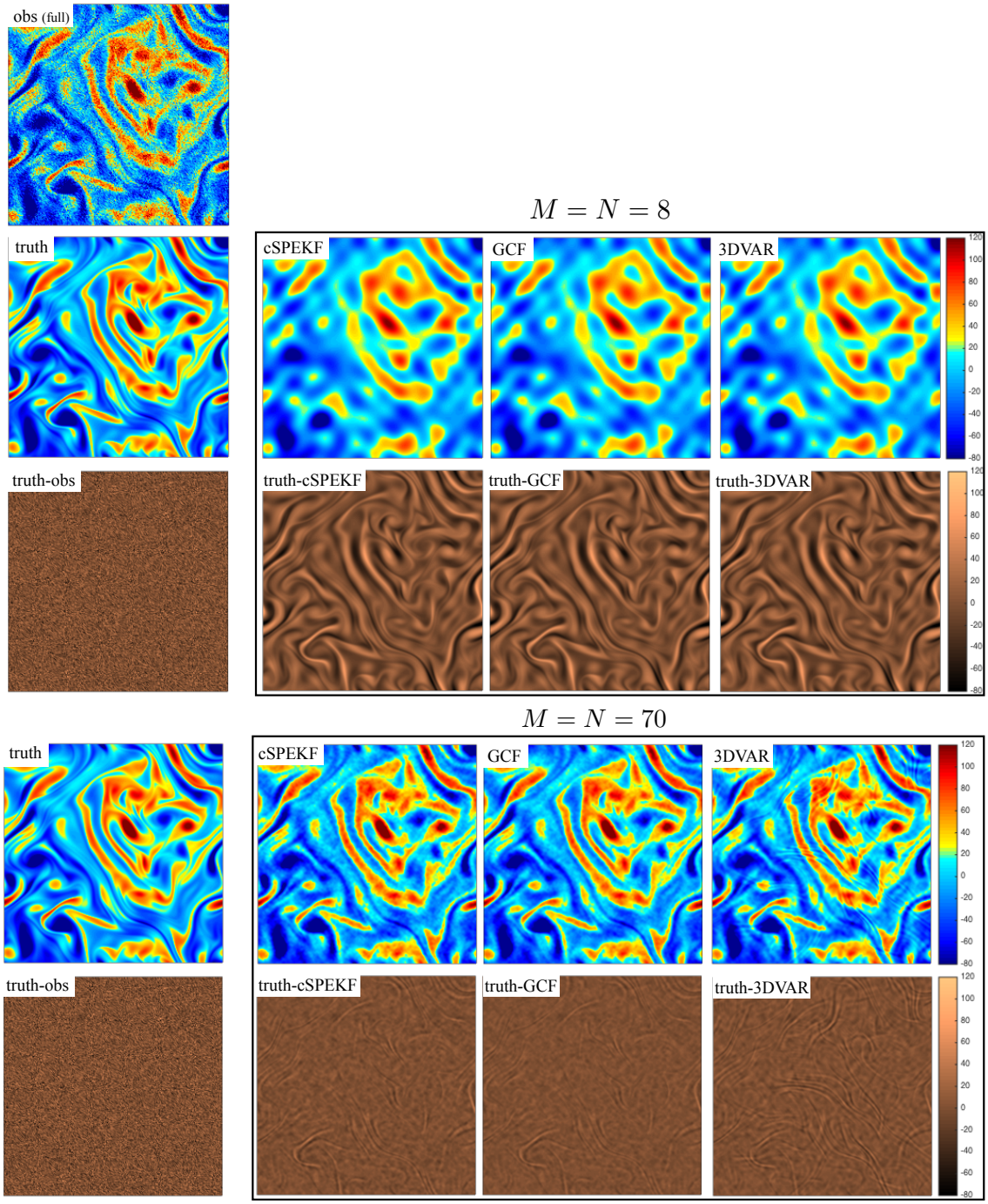

FIGURE 8. FILTERING WITH NON-ALIASED OBSERVATIONS; TURBULENT REGIME OF (8)/(29) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the fully 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.

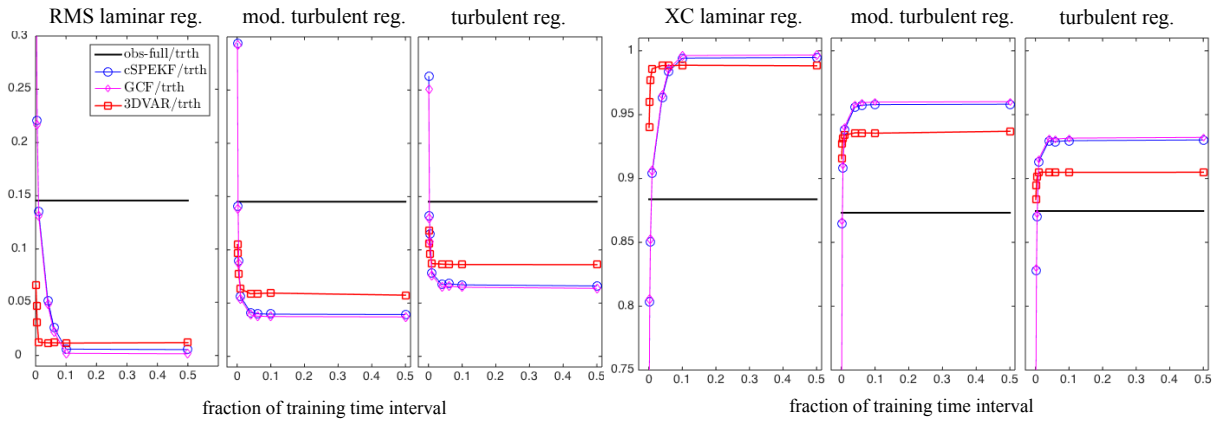

FIGURE 9. FILTER ACURACY AS A FUNCTION OF TRAINING DATA; NON-ALIASED OBSERVATIONS. Comparison of performance of the filtering algorithms of §4 in different dynamical regimes of (29), illustrated in Figure 3, as a function of the length of the training data used to fix the tuneable parameters in the filtering algorithms; see Appendix A. 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) $\sim$ 220 mean decorrelation time units in the laminar regime, (ii) $\sim$ 560 mean decorrelation time units in the moderately turbulent regime, and (iii) $\sim$ 750 mean decorrelation time units in the turbulent regime. Results for 3DVAR are shown for the optimal multiplicative covariance inflation β in $\widehat{C}_{0,\beta}$ (38).

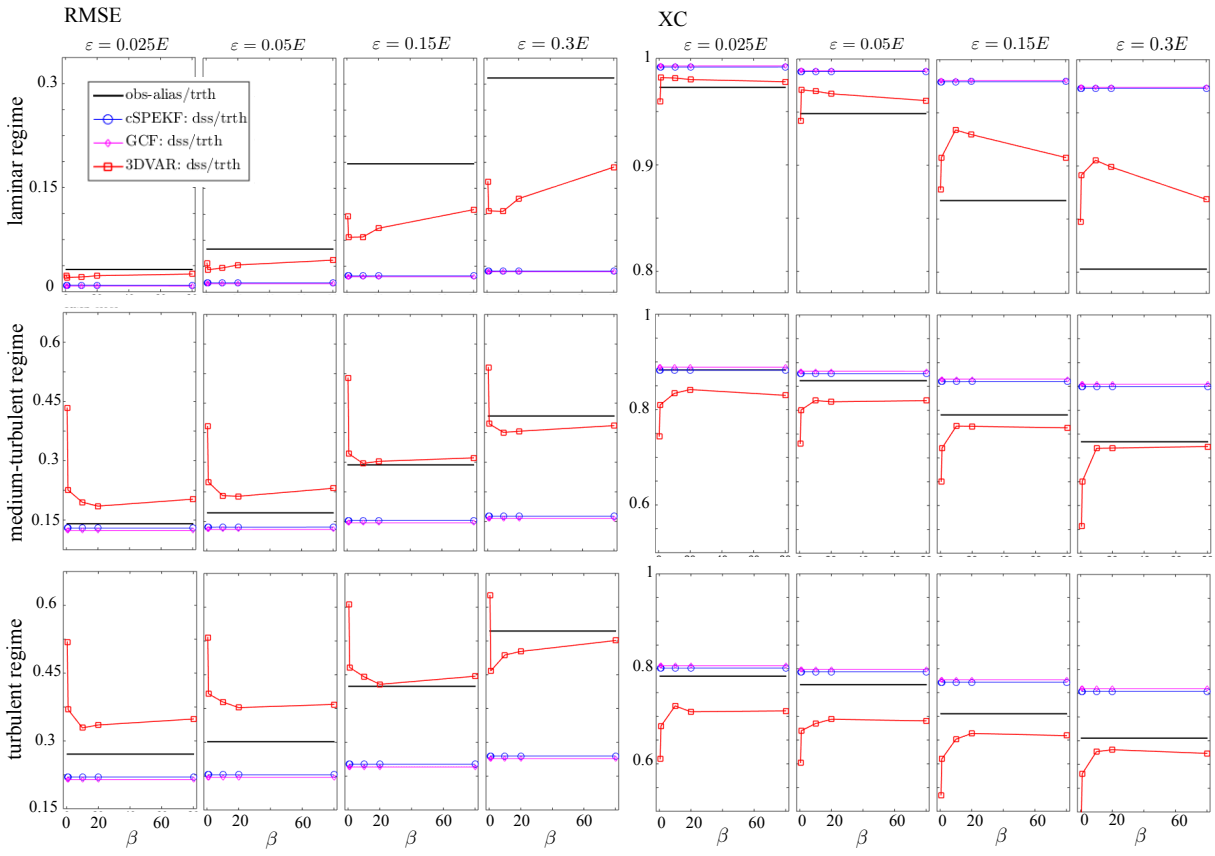

FIGURE 10. FILTERING WITH ALIASED OBSERVATIONS OF (29). Comparison of performance of the filtering algorithms of §4 in different dynamical regimes of (8) (cf Figure 3) in terms of the error in the mean estimates, using RMS (51) and XC measures (53) for different resolutions N of the forward models; here $M = 10$ 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 $\widehat{C}_{0,\beta}$ (38); 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.

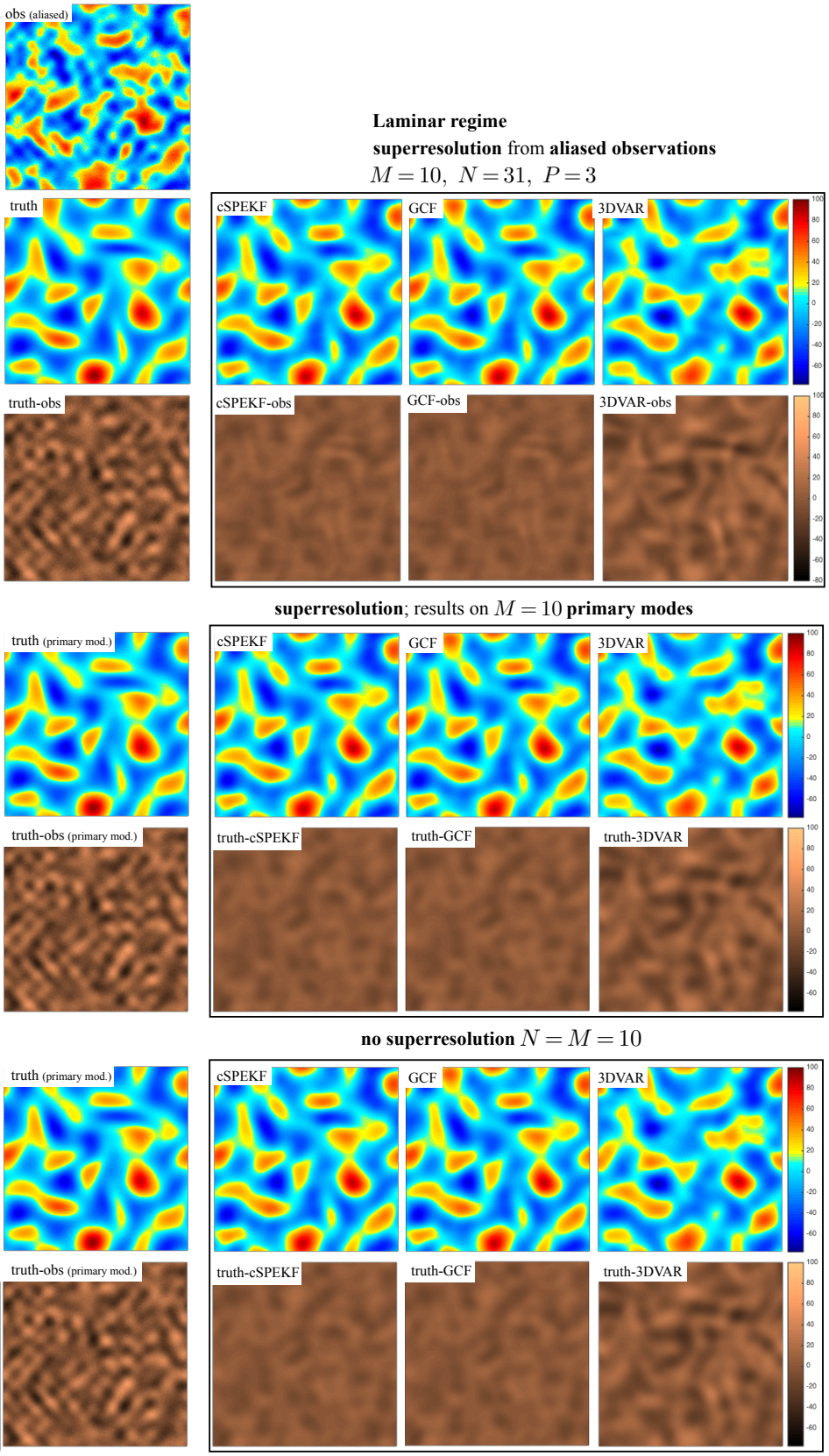

FIGURE 11. FILTERING WITH ALIASED OBSERVATIONS; LAMINAR REGIME OF (8)/(29) (cf Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the truth (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.

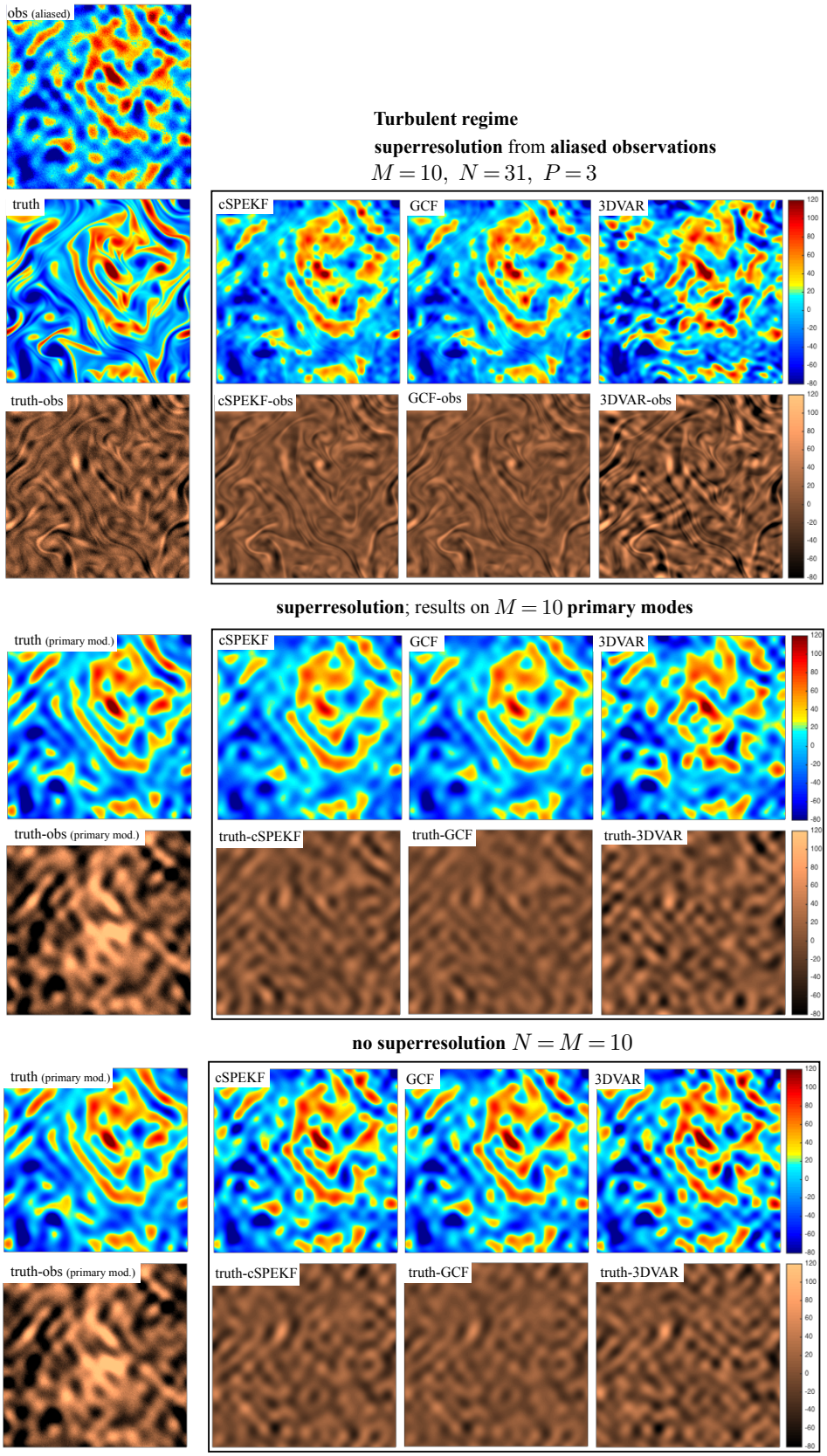

FIGURE 12. FILTERING WITH ALIASED OBSERVATIONS; TURBULENT REGIME OF (8)/(29) (cf. Figure 3). Snapshots of the observed, true and estimated vorticity fields obtained from the filtering algorithms 3DVAR (§4.1), cSPEKF and GCF (§4.2) and the corresponding residuals between the mean estimates and the truth (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.

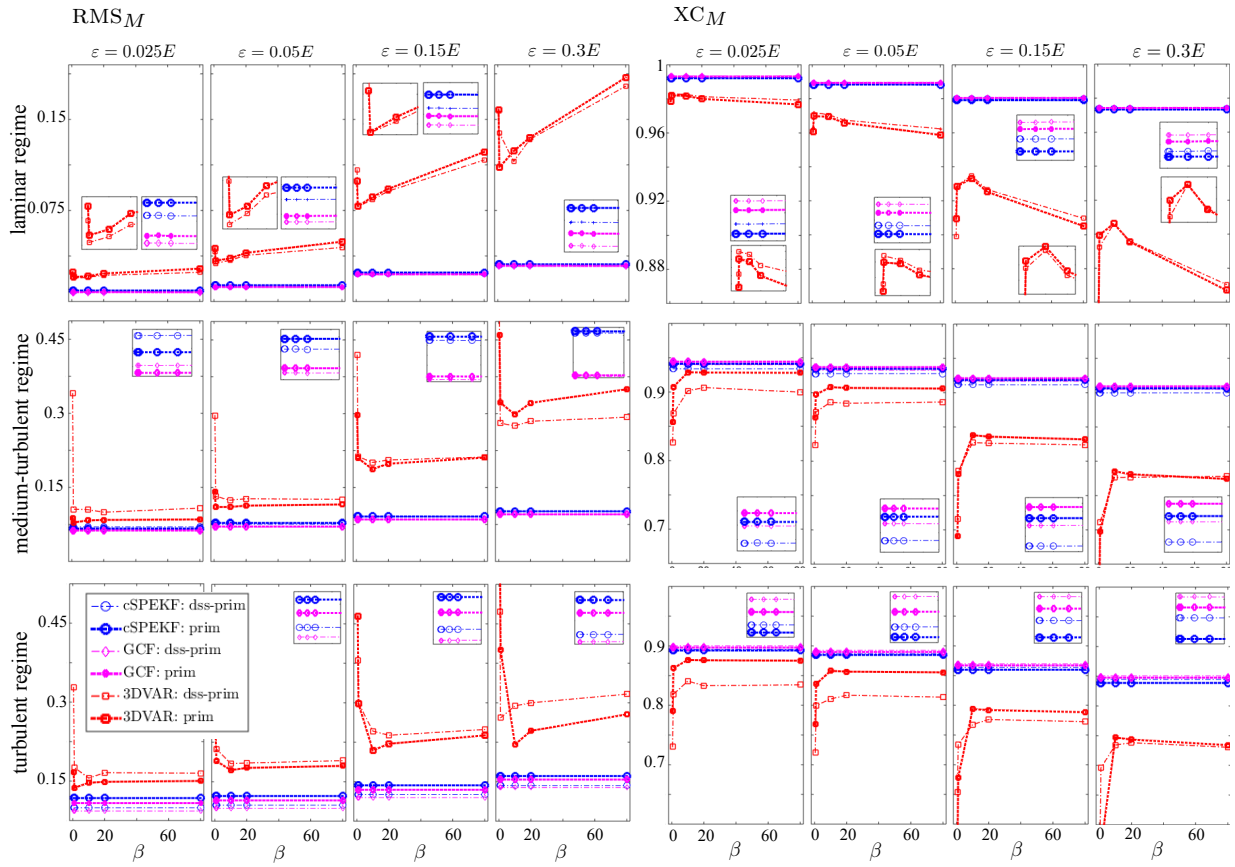

FIGURE 13. FILTERING THE DYNAMICS OF (29) WITH SUPERRESOLVING VS NOT SUPERRESOLVING ALGORITHMS FOR ALIASED OBSERVATIONS. Comparison of the quality of the filtering algorithms of §4 for estimating the M primary modes using superresolving algorithms ($N > M$) and non-superresolving algorithms ($N = M$); the comparison is carried out in terms of the error in the mean estimates using $\mathrm{RMS}_M(u, m_N)$ in (52) and $\mathrm{XC}_M(u, m_N)$ in (54); see (iv), §5.2 in the text. Aliased observations of the truth dynamics (29) 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 $\widehat{C}_{0,\beta}$ (38).