Recovering the Eulerian energy spectrum from noisy Lagrangian tracers

Mustafa A. Mohamad and Andrew J. Majda
Courant Institute of Mathematical Sciences, New York University

November 14, 2018

Abstract

We consider the time-sequential state estimation of a flow field given a stream of noisy measurements that are provided by instruments advected by the flow, known as *Lagrangian tracers* or *drifters*. Lagrangian drifters collect real-time data as they move through the velocity field and are an important data collection method for atmospheric and ocean measurements. Here, we quantify the recovery of the Eulerian energy spectra from observations of Lagrangian drifters. This is performed by utilizing special Lagrangian data assimilation algorithms, known as conditionally Gaussian nonlinear filters. Here we address the following questions: how much of the turbulent Eulerian energy spectra can be recovered from assimilation of Lagrangian trajectory data and how accurately are the various energetic scales recovered relative to the truth. These issues are primarily studied in the perfect model scenario, but we quantify recovery due to model error by reduced order models via spectral truncation of the forecast model. We demonstrate high recovery skill of the two-dimensional turbulent energy spectra for both an exact filter and an imperfect filter, based on extreme localization of the covariance matrix, which is vastly cheaper than the exact filter, for both an inverse cascade spectrum with slope $k^{-5/3}$ and a direct cascade spectrum with slope k^{-3}. The dependence of the spectral energy recovery skill on the number of tracers and the spectral truncation grid size is also studied.

Contents

1 Introduction

We consider the time-sequential state estimation of a flow field given a stream of noisy measurements that are provided by instruments that are advected by the flow, referred to as *Lagrangian tracers* or *drifters*. Lagrangian drifters collect real-time observation as they move through the velocity field and are a crucial data collection source for atmospheric and ocean measurements [27, 20].

The setup of Lagrangian data assimilation involves the assimilation of massless particles that are advected by the fluid velocity field for the purpose of estimating the underlying velocity field state. The flow model is assumed incompressible. We utilize the Bayesian formulation, where the posterior probability of the system state is updated using the model prediction (prior information) and the next available measurement data from the Lagrangian observation process.

A central problem in Lagrangian data assimilation is the highly nonlinear nature of the observation process, since the drifter observations are coupled to the fluid velocity. In general, the resulting nonlinear filtering problem would necessitate the use of approximate filtering strategies (e.g. particle filters or ensemble Kalman filters, see [22, 16, 12]). However, if the velocity field is a linear stochastic turbulence model (of the type discussed in [22, chap. 5]), the resulting problem has a special conditionally Gaussian structure. This structure permits the use of a *conditionally Gaussian nonlinear filter* for the Lagrangian data assimilation of linear stochastic turbulence velocity models. The conditionally Gaussian nonlinear filter is exact (in the mean square sense), and thus no filter approximation errors are made, as with e.g. extended Kalman filters or particle filters for non-Gaussian systems. This fact was first noted and used for the nonlinear filtering of tracer observation data in [8] for incompressible flows. The conditionally Gaussian Kalman filter, originally due to Liptser and Shiryaev in [18, chap. 12], was also used in [9] for special types of compressible flows that are relevant in geophysical applications. The analysis in [8, 9] focused on theoretical properties of the filter and the numerical experiments involved a small grid consisting of order 20 points for two-dimensional flows, resolving only the very largest scales of the flow.

1.1 Background

The Lagrangian characterization of diffusion processes and their statistics from observation and experiments has been investigated in various works, such as [15, 3]. A typical approach involves studying the statistics of single particles or multiple (pairs or groups) of particles as they are advected by the flow to understand dispersion and other properties of the flow. The Eulerian characterization of turbulent diffusion is also an important perspective, see for instance [21, 23]. Connecting the statistics derived from Lagrangian data to Eulerian properties of flow field has also been explored in works such as [15, sec. 2.8] and [23]. On the other hand, simplified stochastic models of tracer particle trajectories that reproduce their observed frequency spectrum from a data modeling perspective where considered in [17].

Lagrangian statistics are closely related to data assimilation methods. The assimilation of Lagrangian particles in various contexts has been considered in [25, 26, 13], focusing on incorporating Lagrangian measurements into general ocean circulation models. On the other hand, studies on various Bayesian strategies for Lagrangian data assimilation have been considered in works including [1, 14], which explore various approximate filters based on Kalman type methods and Markov chain Monte Carlo sampling.

Filtering spatially extended systems, of the type considered in this work, has been extensively studied in various settings. The setup addressed here considers forecast models with higher resolution than the number of drifter particles. The filtering prediction problem thus involves 'super-resolution' [10, 7] or recovery

of scales unobserved. Super-resolution and filtering algorithms have been analyzed in different setups for one-dimensional turbulent advection systems [7, 22] and two-dimensional turbulent Navier-Stokes flows [5]. The super-resolution problem is more delicate in the Lagrangian data assimilation context, since the observation operator mixes information from all the scales in the flow as the Lagrangian particles are advected throughout the fluid; as opposed to the more direct study of filter skill when observing the spatial scales of the flow directly (see [5]).

Information theory and criteria has been explored in the literature for model improvement and Bayesian class selection for dynamical systems. For spectra recovery, a Bayesian class selection framework was studied in [28, 2] for input-output measurement systems and output only measurements systems. We refer to [6] for information-theoretic approaches towards model prediction improvement and [4] for information-based criteria for filtering dynamical systems. However, in the twin-experiment or perfect model scenario, studied here, model improvement is not relevant, as we assume the forward model parameters match the true model parameters. We mention these works, since model selection is a crucial aspect for the real-world filtering of turbulent signals. Our goal here is to first study the twin-experiment setting to establish base guidelines and understand the inherent recovery skill of the algorithms. This understanding can be utilized to inform more intricate setups with imperfect models.

1.2 Objectives

One of the central goals in this work is to study the skill in the recovery of the Eulerian kinetic energy spectra of fluid flow models from a finite number of tracers, primarily in the perfect model setting (twin-experiment scenario). In the perfect model setting the forecast and true velocity model are exactly equivalent. We investigate filter performance and accuracy for realistic setups with moderately large grid sizes on the order of 1000 dimensions. We study performance of a conditionally Gaussian nonlinear filter (i.e. the full or exact filter) in addition to a reduced, diagonal approximation of the exact filter, which is is vastly more computationally efficient, that is derived by extreme localization of the filter covariance matrix. We also study recovery skill of the turbulent velocity spectra in the small tracer limit, as well as when the number of tracer observations are large, since both are relevant configuration from a practical standpoint.

The main sources of error in filtering that are considered here include: model error from using forward models with dynamics that have coarser resolution (reduced order models) compared to the (high dimensional) truth and observation errors that result from a nonlinear observation operator that mixes information from all the scales in the flow, i.e. 'sub-grid' effects, (refereed to as representation error [10, 11, 19]) and sparse measurements due to a finite number of tracers (incomplete and noisy measurements). Model error in the forecast or forward dynamics considered here is exclusively due to spectral truncation (in general it would also include misspecified system parameters). Furthermore, we consider the skill of a cheap imperfect filter, which ignores all particle correlations in the filter covariance.

1.3 Contributions and overview

The following summarize the main contributions of this work

- Quantification of the recovery of the Eulerian velocity spectra from observations of Lagrangian tracers (inverse problem).
- Recovery in the context of model error due to reduced-order models that are based on spectrally truncated forecast models.
- How much of the turbulent velocity spectrum can be recovered and just how accurately. Are there energetic scales that are easy or hard to recover? What kind of velocity spectrum shape is recovered relative to the true energy spectrum.
- Assessment of the relevant issues for moderately large dimensional velocity models on the order of 1000 grid points and for various flow regimes. Furthermore, study skill in the small drifter scenario.
- In addition to the assessment of the exact nonlinear filter, demonstration of the skill of an imperfect filter that is derived based on a mean-field theory of the tracer observations, which is vastly cheaper than the exact filter.

1.4 Outline

In section 2, we describe the general stochastic turbulence velocity models and the Lagrangian observation process. An overview overview of the assimilation problem is included in section 2.3 and, in section 3, we provide a full description of the filter problem formulation. In section 4, the exact and an inexact filtering algorithms are described in full detail. In section 5, we discuss and define the Eulerian energy spectra of the stochastic turbulence velocity models and, in section 6, we discuss how the kinetic energy spectra of the velocity models can be recovered through filtering strategies. Performance and accuracy of the filtering algorithms are provided in section 7. Concluding remarks are made in section 8.

2 General velocity models and the tracer observation process

Here we discuss the velocity models and the observation process. The velocity models are formulated based on linear stochastic models, which are described in detail here, along with flow regimes of the model. After the description of the velocity and observation process, we state the filtering problem and also discuss simplified velocity models that serve as instructive test problems.

2.1 The general incompressible stochastic velocity field model

As the velocity model we consider a two-dimensional flow in a periodic domain $D = [0, 2\pi)^d$, $d = 2$, where the velocity field is a superposition of divergence free modes with random amplitudes:

$$\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{w}(t) + \sum_{\boldsymbol{k}\in I_\Lambda} \hat{v}_{\boldsymbol{k}}(t)\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}), \quad \boldsymbol{v} \in \mathbb{R}^d, \quad \boldsymbol{x} \in [0, 2\pi)^d. \tag{1}$$

The term $\boldsymbol{w}(t)$ is a spatially uniform *sweeping* component, in other words the time-dependent background velocity field. The system is resolved on wavenumbers $I_\Lambda := \{\boldsymbol{k} \in \mathbb{Z}^d \setminus \{0\} \colon -\Lambda \leqslant k_i \leqslant \Lambda \text{ for } i = 1, \dots, d\}$, on a total grid of $N^d = (2\Lambda + 1)^d$ points. The basis are divergence free Fourier modes that are given by

$$\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}) = \boldsymbol{r}_{\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \quad \text{where} \quad \boldsymbol{r}_{\boldsymbol{k}} = \frac{i\boldsymbol{k}^\perp}{k}; \tag{2}$$

note $\boldsymbol{\psi}_{-\boldsymbol{k}} = \boldsymbol{\psi}_{\boldsymbol{k}}^*$ and we also require $\hat{v}_{\boldsymbol{k}}^* = \hat{v}_{-\boldsymbol{k}}$ to ensure the velocity field is real valued. The vector $\boldsymbol{r}_{\boldsymbol{k}}$ is orthogonal to the wavevector $\boldsymbol{k}$ in order to enforce the incompressibility constraint $\nabla \cdot \boldsymbol{v} = 0$. The imaginary unit in $\boldsymbol{r}_{\boldsymbol{k}}$ ensures that wavevector $\boldsymbol{k}$ is conjugate to $-\boldsymbol{k}$ and the denominator term ensures the vector $\boldsymbol{r}_{\boldsymbol{k}}$ is non-dimensional. The random coefficients are described by independent Ornstein-Uhlenbeck (OU) processes

$$d\hat{v}_{\boldsymbol{k}}(t) = -d_{\boldsymbol{k}}\hat{v}_{\boldsymbol{k}}(t) + f_{\boldsymbol{k}}\, dt + \sigma_{\boldsymbol{k}}\, dB_{\boldsymbol{k}}(t), \tag{3}$$

where $d_{\boldsymbol{k}}$ is damping, $f_{\boldsymbol{k}}$ is deterministic forcing, and $B_{\boldsymbol{k}} = \frac{1}{\sqrt{2}}(B_{\boldsymbol{k}}^1 + iB_{\boldsymbol{k}}^2)$ is a unit complex Wiener process with real valued amplitude $\sigma_{\boldsymbol{k}}$ and $\frac{1}{\sqrt{2}}(B_{-\boldsymbol{k}}^1 + iB_{-\boldsymbol{k}}^2) = \frac{1}{\sqrt{2}}(B_{\boldsymbol{k}}^1 - iB_{\boldsymbol{k}}^2)$, to ensure conjugate symmetry of the basis coefficients.

The dynamics of the sweeping flow $\boldsymbol{w}(t)$ is given by,

$$d\boldsymbol{w}(t) = -d_0\boldsymbol{\omega}(t)\, dt + \Omega_0\boldsymbol{\omega}(t)\, dt + \boldsymbol{f}_0\, dt + \sigma_0\, d\boldsymbol{W}_0(t), \tag{4}$$

where Ω_0 is a skew-symmetric matrix representing rotation effects, and $\boldsymbol{W}_0$ is a real valued Wiener process in $\mathbb{R}^d$ with independent components. The sweeping flow mimics inhomogeneous flows at large scales in nature, which can strongly impact spectral recovery of the Eulerian velocity.

Eqs. (2) and (4) define the stochastic turbulence velocity models that are utilized in this work. The description of the Eulerian velocity spectra for these stochastic velocity models is provided in section 5.

2.1.1 Terminology of various flow regimes

We define some common terminology here that pertains to various flow states of the velocity model. *Sweeping flows* are flows where the spatially uniform background velocity component is nonzero $\boldsymbol{w}_t = \overline{\boldsymbol{w}}_t + \widetilde{\boldsymbol{w}}_t \neq \boldsymbol{0}$. The sweeping flow may be constant with a nonzero mean (implying nonzero forcing) or may be fluctuating in time with a zero or nonzero mean. For simplicity, we denote a flow with *constant sweeps* as a flow with

a nonzero mean and no fluctuations and *fluctuating sweeps* as a flow with zero forcing $\boldsymbol{f}_0 = \mathbf{0}$ and nonzero fluctuations $\sigma_0 \neq 0$; thus fluctuating and forced sweeps have $\boldsymbol{f}_0 \neq \mathbf{0}$ and $\sigma_0 \neq 0$, fluctuating sweeping flows have $\boldsymbol{f}_0 = \mathbf{0}$ and $\sigma_0 \neq 0$, constant sweeps have $\sigma_0 = 0$ and $\boldsymbol{f}_0 \neq 0$, and zero background flows when $\boldsymbol{w}_t \equiv \mathbf{0}$.

Mean flows are flows where $\boldsymbol{f}_{\boldsymbol{k}} \neq 0$ for at least one wavenumber. In other words, velocity models with a mean flow have spatial variations in the mean velocity field $\overline{\boldsymbol{v}}(\boldsymbol{x}, t) \neq 0$ due to forcing. Note, the mean flow terminology defined here does not imply any structure on the sweeping component of the velocity field. If we do not specify that the velocity field has a mean flow, it is assumed that it is a zero mean flow.

2.2 The Lagrangian tracer observation process

As our observation process, we are given L noisy trajectories of Lagrangian drifters $\boldsymbol{X}_l \in \mathbb{R}^d$ with dynamics given by

$$d\boldsymbol{X}_l(t) = \boldsymbol{v}(\boldsymbol{X}_l(t), t)\, dt + \sigma_l\, d\boldsymbol{W}_{X_l}(t)\, dt, \quad l = 1, \ldots, L, \tag{5}$$

where $\boldsymbol{W}_{X_l}(t)$ is a d-dimensional real valued independent Wiener process. Furthermore, we assume all the tracers have the same instrumental or observational error level so that $\sigma_l = \sigma_x$ for $l = 1, \ldots, L$.

2.3 Overview of filtering the velocity field via Lagrangian tracer observations

In the filtering problem, our aim is to recover the signals $\boldsymbol{w}(t)$ and $\hat{v}_{\boldsymbol{k}}(t)$ from the L noisy observations of $\boldsymbol{X}_l(t)$. Thus in the general setting, the combined forecast model and observation process is given by:

$$\text{forecast model:} \qquad \boldsymbol{v}^f(\boldsymbol{x}, t) = \boldsymbol{w}^f(t) + \sum_{\boldsymbol{k} \in I_{\Lambda_f}} \hat{v}^f_{\boldsymbol{k}}(t) \boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}), \tag{6}$$

$$d\boldsymbol{w}^f(t) = -d^f_0 \boldsymbol{\omega}(t)\, dt + \Omega^f_0 \boldsymbol{\omega}(t)\, dt + \boldsymbol{f}^f_0\, dt + \sigma^f_0\, d\boldsymbol{W}_0(t), \tag{7}$$

$$d\hat{v}^f_{\boldsymbol{k}}(t) = -d^f_{\boldsymbol{k}} \hat{v}_{\boldsymbol{k}}(t)\, dt + f^f_{\boldsymbol{k}}(t)\, dt + \sigma_{\boldsymbol{k}}{}^f\, dB_{\boldsymbol{k}}(t), \tag{8}$$

$$\text{observation process:} \qquad d\boldsymbol{X}_l(t) = \boldsymbol{v}(\boldsymbol{X}_l(t), t)\, dt + \sigma_x\, d\boldsymbol{W}_{X_l}(t), \quad l = 1, \ldots, L. \tag{9}$$

For simplicity, we assume deterministic forcing, although in general the forcing may contain a periodic component, and no rotation in the background velocity field and hence $\Omega_0 = 0$.

As mentioned earlier, in the twin-experiment scenario, the forecast model eq. (6) is identical the truth model in eq. (1). Even if perfect model parameters are specified (i.e. $d^f = d$, $\Omega^f_0 = \Omega_0$, $f^f = f$, $\sigma^f = \sigma$), which we assume, model error may exists due to model truncation. Namely, model error due to truncation exists when the forecast model is resolved on wavenumbers that differ from the true model $I_{\Lambda_f} \neq I_\Lambda$, where I_{Λ_f} are the modes resolved by the forecast velocity field and I_Λ are the modes resolved by the true model. Typically, in realistic applications, the forecast resolution is much lower than the true model resolution, i.e. $I_{\Lambda_f} \ll I_\Lambda$, so this scenario is useful to study from a practical perspective.

The key filtering idea is that although the tracer evolution eq. (5) is nonlinear, given drifter observations $\boldsymbol{X}_l$(t), the signal processes $\hat{v}_{\boldsymbol{k}}$ are conditionally linear on these observations. In other words, although the observations depend nonlinearly on the signal (the velocity field), once we observe the drifter process, the velocity is then conditionally Gaussian given these observations. Hence, the posterior distribution is also Gaussian, assuming the signal is initially Gaussian distributed. This conditionally Gaussian structure is exploited for filtering. The optimal minimum variance filter in this case is given by the conditionally Gaussian nonlinear Kalman filter, as first described in [18], with explicit closed-form differential equations for the posterior mean and covariance.

2.4 Reduced velocity models: the aligned shear modes model

The general velocity model in eq. (1) contains a rich number of interesting representative flow regimes. It is instructive to study reduced velocity models that represent certain general flow regimes. Shear flows represent one such flow regime contained in the general velocity model. An instructive testbed problem consists of a velocity model where all the wavenumbers are aligned in the same direction $\hat{\boldsymbol{k}}$, and the modes

that are not aligned in this direction are neglected. Suppose that all the modes are aligned in the x-axis, i.e. $\hat{\boldsymbol{k}} = \boldsymbol{e}_1 = [1; 0]$, then the *aligned shear modes model* is explicitly given by:

$$\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{w}(t) + \sum_{\substack{k=-\Lambda \\ k\neq 0}}^{\Lambda} \hat{v}_k(t) i \begin{bmatrix} 0 \\ 1 \end{bmatrix} e^{ikx}, \qquad \hat{v}_k^* = -\hat{v}_{-k} \tag{10}$$

$$d\boldsymbol{w}(t) = -d_0 \boldsymbol{\omega}(t)\, dt + \Omega_0 \boldsymbol{\omega}(t)\, dt + \boldsymbol{f}_0\, dt + \sigma_0\, d\boldsymbol{W}_0(t), \tag{11}$$

$$d\hat{v}_n(t) = -d_n \hat{v}_n(t)\, dt + f_n\, dt + \sigma_n\, dB_n(t). \tag{12}$$

The vast reduction in the number of wavenumbers in the aligned shear modes model makes it a practically useful testbed for studying filter skill and related issues in Lagrangian tracer recovery (a total of 2Λ modes compared to $(2\Lambda)^d$ of the full model). Moreover, these are the simplest flows that can be used to test spectral recovery and model error, such as those due to mixing in the observation process and representation error due to model truncation. Although we mention this model, we focus on the full two-dimensional system defined in eq. (1) in this work.

2.5 General remarks on the velocity models

The general velocity model in eq. (1), despite its conceptual simplicity, can generate a rich number of interesting flow regimes (from jet like flows to purely random flows), and thus serves as a practical test model to study the inherent skill of filter approximations and model error. We explicitly included a sweeping flow component $\boldsymbol{w}_t$ in the velocity model and the dynamics of this component also follows general linear, stochastic dynamics with added mixing due to rotation. This background term may represent, for example, seasonal affects and other important inhomogeneous time-dependent, large-scale flow features that are observed in nature, which are prevalent in geophysical applications. It is therefore instructive to understand the impact of this spatially uniform background flow in the observation process and its affects on filter skill.

As an aside, we mention that special cases of the velocity models in eq. (1) are solutions of the two-dimensional quasi-geostrophic equations used in the study of geophysical flows [24]. The velocity models are exact solutions under Kolmogorov shell forcing and general damping and topography. The spatially uniform background component $\boldsymbol{w}_t$ is constant when there is rotation and no topography. When topography and rotation affects are present, the sweeping component is time dependent. We can understand the sweeping component $\boldsymbol{w}_t$ as being generated due to topography, but this interpretation is not necessary. For further details regarding this connection to geophysical flows, we refer to [24, Chap. 1].

We remark that the velocity model can be written in terms of the streamfunction $\psi(\boldsymbol{x}, t)$ as $\boldsymbol{v}(\boldsymbol{x}, t) = \boldsymbol{w}(t) + \nabla^\perp \psi(\boldsymbol{x}, t)$, where $\nabla^\perp = (-\partial_y, \partial_x)$. Explicitly, the streamfunction, in terms of the spectral coefficients $\hat{v}_{\boldsymbol{k}}(t)$, is given by

$$\psi(\boldsymbol{x}, t) = \sum_{\boldsymbol{k}\in I_\Lambda} \hat{\psi}_{\boldsymbol{k}}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} = \sum_{\boldsymbol{k}\in I_\Lambda} \frac{\hat{v}_{\boldsymbol{k}}(t)}{k} e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{13}$$

The vorticity, on the other hand, defined by $\boldsymbol{\omega} = \nabla \times \boldsymbol{v} = \nabla \times \nabla^\perp \psi = \Delta\psi$, is given by

$$\boldsymbol{\omega}(\boldsymbol{x}, t) = \sum_{\boldsymbol{k}\in I_\Lambda} \hat{\omega}_{\boldsymbol{k}}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} = \sum_{\boldsymbol{k}\in I_\Lambda} -k\hat{v}_{\boldsymbol{k}}(t) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}. \tag{14}$$

3 Formulation of the assimilation of Lagrangian trajectories for turbulence velocity models

Here we formulate the filter for the given problem involving a general d-dimensional flow with a background velocity field. In section 4, we use the setup here to describe the exact filter along with the imperfect filter algorithm. The description is adapted from [8], with the additional inclusion of a background velocity term.

First group the signals $\boldsymbol{w}$ and $\hat{v}_{\boldsymbol{k}}$ into a $(2\Lambda)^d + d$ dimensional vector,

$$\boldsymbol{U}(t) = \begin{bmatrix} \boldsymbol{w}(t) \\ \vdots \\ \hat{v}_{\boldsymbol{k}}(t) \\ \vdots \end{bmatrix} \tag{15}$$

so that we can compactly write the velocity field dynamics as

$$\boldsymbol{v}(\boldsymbol{x},t) = \boldsymbol{w}(t) + \sum_{\boldsymbol{k}\in I_\Lambda} \hat{v}_{\boldsymbol{k}}(t)\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}) = P(\boldsymbol{x})\boldsymbol{U}(t), \tag{16}$$

where

$$P(\boldsymbol{x}) = \left[\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_d, \cdots, \boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}), \cdots\right] \tag{17}$$

is a $d \times ((2\Lambda)^d + d)$ dimensional matrix, where the columns correspond to basis vectors and the rows the spatial dimensions. The dynamics of the vector of combined background velocity and random basis coefficients can also be compactly written as

$$d\boldsymbol{U}(t) = -\Gamma\boldsymbol{U}(t)\,dt + \boldsymbol{F}\,dt + \Sigma_u\,d\boldsymbol{B}_u(t). \tag{18}$$

Above, Γ is a diagonal matrix representing the damping terms with entries $[\Gamma]_{\boldsymbol{k},\boldsymbol{k}} = d_{\boldsymbol{k}}$ and $[\Gamma]_{\boldsymbol{0},\boldsymbol{0}} = d_0 I_{d\times d}$ (the notation $[A]_{\boldsymbol{0},\boldsymbol{0}}$ denotes the diagonal entries of matrix A corresponding to the background velocity field $\boldsymbol{w}_t$). The term $\boldsymbol{F}$ is a vector of the deterministic forcing terms where $[\boldsymbol{F}]_{\boldsymbol{k}} = f_{\boldsymbol{k}}$ and $[\boldsymbol{F}]_{\boldsymbol{0}} = \boldsymbol{f}_{\boldsymbol{0}}$. The noise Σ_u is a real diagonal matrix where $\boldsymbol{\Sigma}_u = \Sigma_u\Sigma_u^*$ has entries $[\boldsymbol{\Sigma}_u]_{\boldsymbol{k},\boldsymbol{k}} = \sigma_{\boldsymbol{k}}^2$ and $[\boldsymbol{\Sigma}_u]_{\boldsymbol{0},\boldsymbol{0}} = \sigma_0 I_{2\times 2}$.

Given a realization of the velocity field $\boldsymbol{v}$, the trajectory of one of the L tracers is given by

$$d\boldsymbol{X}_l(t) = \left(\boldsymbol{w}(t) + \sum_{\boldsymbol{k}\in I_\Lambda} \hat{v}_{\boldsymbol{k}}(t)\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{X}_l(t))\right)dt + \sigma_x\,d\boldsymbol{W}_{X_l}(t), \tag{19}$$

$$= P(\boldsymbol{X}_l(t))\boldsymbol{U}(t)\,dt + \sigma_x\,d\boldsymbol{W}_{X_l}(t). \tag{20}$$

We can group all $\boldsymbol{X}_l(t)$ into $d \times L$ dimensional observation vector:

$$\boldsymbol{X}(t) = \begin{bmatrix} \boldsymbol{X}_1(t) \\ \vdots \\ \boldsymbol{X}_L(t) \end{bmatrix}. \tag{21}$$

The dynamics of the observation process can thus be compactly written as

$$d\boldsymbol{X}(t) = P(\boldsymbol{X}(t))\boldsymbol{U}(t)\,dt + \sigma_x\,d\boldsymbol{W}_X(t), \tag{22}$$

with

$$P(\boldsymbol{X}(t)) = \begin{bmatrix} P(\boldsymbol{X}_1(t)) \\ \vdots \\ P(\boldsymbol{X}_L(t)) \end{bmatrix} \quad \text{and} \quad \boldsymbol{W}_X(t) = \begin{bmatrix} \boldsymbol{W}_{X_1}(t) \\ \vdots \\ \boldsymbol{W}_{X_L}(t) \end{bmatrix}, \tag{23}$$

where we abuse the P notation, to keep the notation simple.

In summary, using the notation defined above, we have that filtering $\boldsymbol{v}(\boldsymbol{x},t)$ from the observations $(\boldsymbol{X}_1(s),\ldots,\boldsymbol{X}_L(s))_{s\le t}$ is equivalent to filtering $\boldsymbol{U}(t)$ using $\boldsymbol{X}(s \le t)$, where:

$$\text{forecast model:} \qquad d\boldsymbol{U}(t) = -\Gamma\boldsymbol{U}(t)\,dt + \boldsymbol{F}(t)\,dt + \Sigma_u\,d\boldsymbol{B}(t), \tag{24}$$

$$\text{observation process:} \qquad d\boldsymbol{X}(t) = P(\boldsymbol{X}(t))\boldsymbol{U}(t)\,dt + \sigma_x d\boldsymbol{W}_X(t). \tag{25}$$

3.1 Distribution of the statistical attractor or prior model density

The equilibrium distribution or statistical attractor (the long time limiting distribution) of the velocity field $\boldsymbol{U}$ can be computed from the Fokker-Planck equation, which provides the least biased estimate without any observational drifter data and serves as the prior distribution of the estimate of the velocity field. We denote this limiting equilibrium density by $\pi_{\text{att}} = \mathcal{N}(\boldsymbol{U}_{\text{att}}, R_{\text{att}})$, where the mean and covariance satisfy, respectively,

$$\frac{d\boldsymbol{U}_{\text{att}}}{dt} = -\Gamma\boldsymbol{U}_{\text{att}} + \boldsymbol{F} = 0 \qquad \Longrightarrow \qquad \boldsymbol{U}_{\text{att}} = \Gamma^{-1}\boldsymbol{F}, \tag{26}$$

$$\frac{dR_{\text{att}}}{dt} = -\Gamma R_{\text{att}} - R_{\text{att}}\Gamma^* + \Sigma\Sigma^* = 0 \qquad \Longrightarrow \qquad R_{\text{att}} = \frac{1}{2}\Gamma^{-1}\Sigma\Sigma^*. \tag{27}$$

4 Exact and imperfect filtering algorithms for Lagrangian data assimilation

Here we describe the exact filter for Lagrangian data assimilation for the velocity models discussed in section 2. We then motivate and describe the inexact filter, which is based on a mean-field theory in the large tracer limit. The inexact filter approximation amounts to ignoring the correlations of the non-aligned modes due to the observation process.

4.1 Exact filtering algorithm for Lagrangian tracers

The conditionally Gaussian Kalman nonlinear filter [18] states that conditioned on observations of the tracer, the optimal state of $\boldsymbol{U}_t$ is Gaussian with covariance R_t and mean $\hat{\boldsymbol{U}}_t$ given by the following algorithm.

Algorithm 1 (exact filter for Lagrangian tracers)
The posterior distribution of the state $\boldsymbol{U}_t$ given Lagrangian tracer observations of $\boldsymbol{X}_{s\leqslant t}$, where

$$d\boldsymbol{U}(t) = -\Gamma \boldsymbol{U}(t)\, dt + \boldsymbol{F}(t)\, dt + \Sigma_u\, d\boldsymbol{B}(t), \tag{28}$$
$$d\boldsymbol{X}(t) = P(\boldsymbol{X}(t))\boldsymbol{U}(t)\, dt + \sigma_x d\boldsymbol{W}_X(t), \tag{29}$$

is given by the density $p(\boldsymbol{U}_t \mid \boldsymbol{X}_{s\leqslant t}) = \mathcal{N}(\hat{\boldsymbol{U}}_t, R_t)$, where the filter mean and covariance satisfy the following system of equations

$$d\hat{\boldsymbol{U}}_t = (-\Gamma \hat{\boldsymbol{U}}_t + \boldsymbol{F}_t)\, dt + \sigma_x^{-2} R_t P^*(\boldsymbol{X}_t)\big(d\boldsymbol{X}_t - P(\boldsymbol{X}_t)\hat{\boldsymbol{U}}_t\, dt\big), \tag{30}$$
$$\frac{dR_t}{dt} = -\Gamma R_t - R_t\Gamma^* + \boldsymbol{\Sigma}_u - \sigma_x^{-2} R_t \mathbf{P}_t R_t, \tag{31}$$

where the interaction matrix is defined as

$$\mathbf{P}_t = P^*(\boldsymbol{X}_t)P(\boldsymbol{X}_t) = \sum_{l=1}^{L} P^*(\boldsymbol{X}_l(t))P(\boldsymbol{X}_l(t)). \tag{32}$$

Note, that the covariance R_t satisfies a random matrix Riccati differential equation.

Explicitly the entries of the matrix of interactions $\boldsymbol{P}_t$, corresponding to the wavenumber entries (excluding the background), is given by

$$[\mathbf{P}_t]_{\boldsymbol{n},\boldsymbol{m}} = \sum_{l=1}^{L} e^{i(\boldsymbol{m}-\boldsymbol{n})\cdot \boldsymbol{X}_l(t)} \boldsymbol{r}_{\boldsymbol{n}}^* \boldsymbol{r}_{\boldsymbol{m}}. \tag{33}$$

We see that unless the two wavenumbers in the covariance are aligned, their interaction represented through $\mathbf{P}_t$ is zero. The corresponding entries representing interactions with the background velocity can also be explicitly written out, where a similar conclusion holds.

4.2 A reduced, diagonal imperfect filtering strategy and its properties

Most realistic systems have a high dimensional state space, which prohibits the use of even the most simple filtering algorithms. For a state space of dimension $O(N^d)$ the covariance is of size $O(N^d \times N^d)$, thus propagating the matrix Riccati differential equation in eq. (31) costs $O(N^{d\times 3})$, which is impractical or impossible for all but the smallest problems, and is especially challenging for stiff systems with shallow energy spectra. Inexpensive, imperfect filters are necessary to mitigate the problem associated with propagating a high dimensional covariance equation. We describe a cheap filter here, which avoids propagating the covariance matrix by neglecting all tracer interaction terms: a type of extreme covariance localization. We discuss this cheap algorithm and also describe the filter's theoretical behavior and performance under special limits. As our aim is not to repeat certain theoretical results and their proofs, we include only the essential theorems and restate them in a setting suitable to motivate and provide rigorous intuition for the inexact filter.

The reduced filter strategy for Lagrangian recovery consists of utilizing the mean-field dynamics of the observations in the limit of a large number of tracers L. Recall that the covariance equation in eq. (31) is random since the observations enter through the interaction matrix $\mathbf{P}_t$, which is a function of the tracer

trajectories. In the mean field limit of a large number of tracers, the random affects are averaged out and the covariance equation is deterministic. Furthermore, in this large tracer limit the interaction matrix $\mathbf{P}_t$ is in fact diagonal.

The mean field limit argument relies upon ergodicity of the tracer observations, which have been shown to converge to the uniform distribution. Note, conditioned on $\boldsymbol{v}(\boldsymbol{x},t)$ the tracers $\boldsymbol{X}_l(t)$ for $l=1,\dots,L$ are independent. Ergodicity and convergence of the tracer observations to the uniform distribution given a realization of the velocity field is the first main theoretical result from [8, Theorem 3.1].

Theorem 1 (tracer observations limiting distribution)
The distribution of $\boldsymbol{X}_l(t)$ given an almost sure realization $\boldsymbol{v}(\boldsymbol{x},t)$ converges (geometrically) fast towards the uniform distribution on $[0,2\pi)^d$.

Based on the result above, we have the following result regarding the limiting behavior of the posterior covariance R_t for large L.

Theorem 2 (posterior covariance limit)
Assume that the initial location of each tracer is independently and uniformly distributed in the domain. The fixed point solution of eq. (31) denoted by R_L in the limit of a large number of tracers $L\to\infty$ is a diagonal matrix with diagonal entries corresponding to the wavenumber $\boldsymbol{k}$ being

$$[R_L]_{\boldsymbol{k},\boldsymbol{k}} = \frac{\sigma_{\boldsymbol{k}}^2}{d_{\boldsymbol{k}} + \sqrt{d_{\boldsymbol{k}}^2 + L\sigma_x^{-2}\sigma_{\boldsymbol{k}}^2}}. \tag{34}$$

See [8, Theorem 3.3] and [8, Lemma 4.1] for the proof and further details. Briefly, taking the mean field average of the interaction matrix $\mathbf{P}_t$, when L is large, results in a diagonal matrix with diagonal entries equal to L. Under this condition, it is then possible to show that the posterior matrix equation is diagonal and the fixed point solution of the resulting differential equations for the diagonal entries of the covariance matrix are equal to

$$\frac{dr}{dt} = -2d_{\boldsymbol{k}}r + \sigma_{\boldsymbol{k}}^2 - L\sigma_x^{-2}r^2, \quad \text{where} \quad r \coloneqq [R_L]_{\boldsymbol{k},\boldsymbol{k}}, \tag{35}$$

which has the solution given in eq. (34).

Utilizing this mean field reduced diagonal covariance approximation yields the following imperfect, diagonal filter which is vastly cheaper than the optimal filter.

Algorithm 2 (approximate diagonal filter for Lagrangian tracers)
The diagonally reduced approximate filter based on the mean field limit of the covariance equation of the exact filter is given by the density $p(\boldsymbol{U}_t \mid \boldsymbol{X}_{s\leqslant t}) \approx \mathcal{N}(\hat{\boldsymbol{U}}_t, R_L)$, where the filter mean and covariance satisfy

$$d\hat{\boldsymbol{U}}_t = (-\Gamma\hat{\boldsymbol{U}}_t + \boldsymbol{F}_t)\,dt + \sigma_x^{-2}R_L P^*(\boldsymbol{X}_t)\big(d\boldsymbol{X}_t - P(\boldsymbol{X}_t)\hat{\boldsymbol{U}}_t\,dt\big), \tag{36}$$

$$[R_L]_{\boldsymbol{k},\boldsymbol{k}} = \frac{\sigma_{\boldsymbol{k}}^2}{d_{\boldsymbol{k}} + \sqrt{d_{\boldsymbol{k}}^2 + L\sigma_x^{-2}\sigma_{\boldsymbol{k}}^2}}. \tag{37}$$

The diagonal filter completely avoids propagation of the full covariance matrix and only involves a differential equation involving the filter mean, which brings to cost to $O(N^d)$ differential equation solves, compared to the $O(N^{d\times 3})$ cost of the perfect filter. We study the numerical performance of this approximate filter in section 7, and show high recovery skill of the approximate filter, even when it used in small tracer regimes, away from the asymptotic conditions that the algorithm's approximations are based upon.

We can gain more intuition on the behavior of $\hat{\boldsymbol{U}}_t$ by substituting in the observation process

$$d\hat{\boldsymbol{U}}_t = \big(-\Gamma\hat{\boldsymbol{U}}_t + \boldsymbol{F}_t + \sigma_x^{-2}R_t\mathbf{P}_t(\boldsymbol{U}_t - \hat{\boldsymbol{U}}_t)\big)\,dt + \sigma_x^{-1}R_L P^*(\boldsymbol{X}_t)\,d\boldsymbol{W}_X(t). \tag{38}$$

In the mean-field limit we have that $R_t\to R_L$ and $\mathbf{P}_t\to LI$, where I is a diagonal matrix. If we further define $\boldsymbol{e}_t \equiv \boldsymbol{U}_t - \hat{\boldsymbol{U}}_t$ we obtain an equation for the deviation of filter mean from the truth signal for the imperfect filter:

$$d\boldsymbol{e}_t = -(\Gamma + L\sigma_x^{-2}R_L)\boldsymbol{e}_t\,dt + \Sigma_u d\boldsymbol{W}_u(t) - \sigma_x^{-1}R_L P^*(\boldsymbol{X}_t)\,d\boldsymbol{W}_X(t). \tag{39}$$

We see that the error of the filter mean is a randomly driven system, where the noise is dependent upon tracer observations.

Further, we can rigorously show that this filtering mean error goes to zero and the posterior covariance approaches R_L in the asymptotic limit of a large number of drifters, which justifies alg. 2. We refrain from repeating the derivations, but refer to [8, Theorem 3] for the precise statement and proof.

Theorem 3 (asymptotic convergence of the posterior filter distribution)
In the limit that the number of tracers L goes to infinity, given a fixed, almost sure realization of the velocity field $\boldsymbol{v}(\boldsymbol{x}, t)$, the filtering mean $\hat{\boldsymbol{U}}_t$ in eq. (36) converges to the truth $\boldsymbol{U}_t$ and the posterior covariance R_t approaches R_L. In other words, the error between the truth converges to zero $\boldsymbol{e}_t \to \boldsymbol{0}$ and the posterior covariance approaches a deterministic limit $\|R_L - R_t\| \to 0$.

The above theorem means that the posterior distribution $p(\boldsymbol{U}_t \mid \boldsymbol{X}_{s\leqslant t})$ asymptotically converges to $\mathcal{N}(\hat{\boldsymbol{U}}_t, R_L)$ as the number of tracer observations L increases to infinity.

5 Energy spectra of the turbulence velocity models

Here we define and discuss the energy spectrum of the turbulence velocity models in eq. (1). The general spectral inversion problem involves estimation of the energy of the various scales of the flow from realtime drifter observation. This is performed by first recovering the velocity field and then computing the energy spectrum of the recovered signal through a suitable (online) average. The performance of the recovered spectrum depends on the accuracy of the posterior distribution of the filter, the observation time length, observation process noise, and the forecast model (because of model error).

The total kinetic energy of the velocity field for a fixed realization is given by

$$E(t;\omega) = \frac{1}{2}\frac{1}{A_D}\int_D \|\boldsymbol{v}(\boldsymbol{x},t;\omega)\|^2 = \frac{1}{2}\|\boldsymbol{w}(t;\omega)\|^2 + \frac{1}{2}\sum_{\boldsymbol{k}}|\hat{v}_{\boldsymbol{k}}(t;\omega)|^2, \tag{40}$$

where A_D is the area of the domain $D = [0, 2\pi)^d$, $d = 2$. As the velocity field is random, we are interested in recovering the ensemble averaged energy,

$$E(t) = \langle E(t;\omega)\rangle = \frac{1}{2}\langle\|\boldsymbol{w}(t;\omega)\|^2\rangle + \frac{1}{2}\sum_{\boldsymbol{k}}\langle|\hat{v}_{\boldsymbol{k}}(t;\omega)|^2\rangle = E_0(t) + \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}(t), \tag{41}$$

where E_0 is the energy of the background flow and $E_{\boldsymbol{k}}$ is the energy spectrum of the various spatial scales of the flow.

Now decompose the mode $\hat{v}_{\boldsymbol{k}}(t)$ into its mean $\langle\hat{v}_{\boldsymbol{k}}(t;\omega)\rangle = \overline{\hat{v}_{\boldsymbol{k}}}(t)$ and fluctuations $\hat{v}'_{\boldsymbol{k}}(t;\omega) = \hat{v}_{\boldsymbol{k}}(t;\omega) - \overline{\hat{v}_{\boldsymbol{k}}}(t)$. The energy spectrum in wavenumber space can be written as,

$$E_{\boldsymbol{k}} = \overline{E}_{\boldsymbol{k}} + \widetilde{E}_{\boldsymbol{k}}, \quad \text{where} \quad E_{\boldsymbol{k}} = \frac{1}{2}\langle|\hat{v}_{\boldsymbol{k}}|^2\rangle, \quad \overline{E}_{\boldsymbol{k}} = \frac{1}{2}|\overline{\hat{v}_{\boldsymbol{k}}}|^2, \quad \widetilde{E}_{\boldsymbol{k}} = \frac{1}{2}\langle|\hat{v}'_{\boldsymbol{k}}|^2\rangle, \tag{42}$$

which consists of the mean spectrum $\overline{E}_{\boldsymbol{k}}$ and the variance spectrum $\widetilde{E}_{\boldsymbol{k}}$. The variance spectrum tells us how much energy is distributed across the scales in the system. To compute the variance spectra from a time series requires the time dependent mean, which is a statistical average and is not available from a single realization of the flow. The time averaged squared amplitude or (total) energy spectra $E_{\boldsymbol{k}}$, on the other hand, does not require knowledge of the time dependent values of the mean. As we assume ergodicity of the velocity models, we replace the statistical averages in the quantities above by their long time averages instead, which is why we drop dependence on t in the notation above.

We define the radial summation of the energy spectra (sometimes referred to as the omnidirectional spectra) by the following integral where $\boldsymbol{k} = k(\cos(\theta), \sin(\theta))$:

$$E_k = \int_0^{2\pi} E(\boldsymbol{k}) k \, d\theta. \tag{43}$$

We employ the omnidirectional spectra to quantify spectra recovery skill, since it is often used in applications and since it succinctly summarizes recovery accuracy for various spatial resolutions in the flow.

The velocity model in eq. (1) can be easily shown to have the following steady state energy spectrum [22, chap. 5]

$$E_{\boldsymbol{k}} = \frac{1}{2}\Big(\frac{f_{\boldsymbol{k}}^2}{d_{\boldsymbol{k}}^2} + \frac{\sigma_{\boldsymbol{k}}^2}{2d_{\boldsymbol{k}}}\Big), \quad \text{where} \quad \overline{\hat{v}_{\boldsymbol{k}}} = \frac{f_{\boldsymbol{k}}}{d_{\boldsymbol{k}}} \ \text{ and } \ \langle |\hat{v}'_{\boldsymbol{k}}|^2 \rangle = \frac{\sigma_{\boldsymbol{k}}^2}{2d_{\boldsymbol{k}}}. \tag{44}$$

Typical turbulent spectra involve power laws $\widetilde{E}_k = E_0|k|^{-\alpha}$, but there is interest in more rough energy spectra, such as white noise (i.e. equipartition), where $\widetilde{E}_{\boldsymbol{k}}$ is constant. A spectra with constant energy for the largest scales or lowest frequency modes and a power law for the small, high frequency, scales mimics realistic turbulent processes:

$$\widetilde{E}_k = \begin{cases} kE_0, & k \leqslant k_0, \\ k_0 E_0 \big|\frac{k}{k_0}\big|^{-\alpha}, & k > k_0 \end{cases}, \quad \text{where } \alpha = 5/3 \text{ (inverse cascade) or } \alpha = 3 \text{ (direct cascade).} \tag{45}$$

When the power law exponent is $\alpha = 5/3$, the spectra mimics turbulent systems with inverse cascade of energy, typically refereed as a Kolmogorov spectrum. When the power law exponent is $\alpha = 3$ the spectra mimics direct cascade of energy in turbulent processes. We focus specifically on these two regimes when testing filter performance, later in section 7.

The energy spectrum depends on the noise $\sigma_{\boldsymbol{k}}$ and damping $d_{\boldsymbol{k}}$. The damping processes include viscosity νk^2, which is scale dependent, and uniform damping $d > 0$, i.e. linear drag, which can represent idealized geophysical processes, such as radiative damping:

$$d_{\boldsymbol{k}} = d + \nu k^2. \tag{46}$$

The mean model decorrelation time scale, in other words the decorrelation time of the largest scales, is based on the damping model, and is approximately $1/d$ time units or $1/\nu$ time units if $d = 0$ (ν is typically very small).

6 Eulerian velocity spectra estimation through filtering

Here we describe how spectrum estimates are obtained through filtering. As we will see in the numerical experiments, the filter posterior mean prediction deteriorates at high wavenumbers. For the smallest scales where the filtering mean performance is poor, we can rely on the filter model (i.e. model prior) to improve the estimate of the energy spectrum. In the twin-experiment scenario this should yield exact recovery if the filter is exact. We describe the conditions necessary in order to estimate the spectra through a single (long time) realization of the Lagrangian drifter observation process and how the model covariance can be utilized to enhance spectrum estimation.

For simplicity consider a complex scalar valued stochastic process denoted by $u_t \in \mathbb{C}$. Our aim is to estimate the statistical energy of this process, in other words the average square amplitude of the process $\langle |u_t|^2 \rangle$. Ultimately, we are interested in the equilibrium statistics of the attractor, in the long time limit for $t \to \infty$, after initial conditions are forgotten $\langle |u_\infty|^2 \rangle$. Moreover, we assume the process is ergodic, so we can replace statistical averages by time averages of a single realization in the equilibrium state:

$$\langle |u_\infty|^2 \rangle = \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} |u_t|^2 \, dt \tag{47}$$

for t_0 sufficiently large to ensure we have reached the equilibrium state.

In the context of velocity recovery from Lagrangian particles, the state estimation problem of the velocity field is solved using special filtering algorithms. Consider again the variable u_t, we assume u_t is conditionally Gaussian given observations of the process z_t and that we can utilize an exact filtering algorithm for the state estimation problem. The Bayesian resolution of the hidden state u_t, is the distribution of the underlying process at the current time conditioned on the observations z_t up to the current time:

$$p(u_t \mid z_{s\leqslant t}) = N(\hat{u}_t, r_t). \tag{48}$$

In other words, the filtering problem involves estimating u_t from observations $\{z_s, s \leqslant t\}$. The optimal mean-square filter estimate is the conditional expectation

$$\hat{u}_t = \mathbb{E}(u_t \mid z_{s\leqslant t}) = \int u_t p(u_t \mid z_{s\leqslant t})\, du_t, \tag{49}$$

which minimizes the mean-square error

$$r_t = \mathbb{E}(|u_t - \hat{u}_t|^2 \mid z_{s\leqslant t}) = \int |u_t - \hat{u}_t|^2 p(u_t \mid z_{s\leqslant t})\, du_t. \tag{50}$$

Through manipulation of conditional distributions we can relate the statistical estimate of the energy at time t with the posterior prediction that depends on the observations, through

$$\langle |u_t|^2 \rangle = \int |u_t|^2 p(u_t)\, du_t = \iint |u_t|^2 p(u_t \mid z_{s\leqslant t}) p(z_{s\leqslant t})\, du_t\, d(z_{s\leqslant t}). \tag{51}$$

Next write $|u_t|^2 = -|\hat{u}_t|^2 + 2\,\mathrm{Re}(\hat{u}_t u) + |u_t - \hat{u}_t|^2$. Substituting this into the equation above and recognizing terms from eqs. (49) and (50), we find

$$\langle |u_t|^2 \rangle = \int r_t\, p(z_{s\leqslant t})\, d(z_{s\leqslant t}) + \int |\hat{u}_t|^2 p(z_{s\leqslant t})\, d(z_{s\leqslant t}) \tag{52}$$

$$= \langle r_t \rangle_z + \langle |\hat{u}_t|^2 \rangle_z. \tag{53}$$

The statistical averages on the right hand side of the equation above are with respect to the observation process z_t and involve an ensemble average over multiple realizations of the observations. If the observations are ergodic, we can replace the ensemble averages with a suitable average in time. Under this condition we find

$$\langle |u_t|^2 \rangle = \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} (r_t + |\hat{u}_t|^2)\, dt; \tag{54}$$

which clearly shows how the energy of the truth signal is related to the filtering mean and covariance.

6.1 Spectra recovery from Lagrangian tracers through a single realization

In the general setting of spectrum recovery from Lagrangian tracers, we need to ensure for any fixed number of tracers L that the tracer locations are ergodic, since we would like to estimate the spectrum from a single realization of the observation process. Fortunately, this fact has been established in theorem 1, which proves that the uniform distribution is the attracting equilibrium distribution for the tracers location. This fact was needed to establish the mean-field limit of the filter in section 4.2, as well. Furthermore, the mean field limit of the covariance establishes $\langle R_t \rangle_z = R_L$ and thus the spectrum estimate utilizing the diagonal filter takes a simplified form involving a time average of the squared amplitude of the posterior mean and the asymptotic covariance matrix R_L. We state below the method utilized to estimate the energy spectra from the Lagrangian filtering algorithms.

Algorithm 3 (Eulerian velocity spectra estimation from Lagrangian tracer filtering algorithms) *Assume t_0 is sufficiently large to ensure we have reached the equilibrium state and T is large. The exact filter estimate of the energy spectra is given by*

$$E_{\boldsymbol{k}} = \frac{1}{2} \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} ([R_t]_{\boldsymbol{k},\boldsymbol{k}} + |\hat{v}_{\boldsymbol{k}}(t)|^2)\, dt, \tag{55}$$

where $\hat{v}_{\boldsymbol{k}}$ and R_t are from eqs. (30) and (31), respectively.

The reduced, imperfect filter approximation of the energy spectra is given by

$$E_{\boldsymbol{k}} \approx \frac{1}{2}\Big([R_L]_{\boldsymbol{k},\boldsymbol{k}} + \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} |\hat{v}_{\boldsymbol{k}}(t)|^2\, dt\Big) = \frac{1}{2}\Big(\frac{\sigma_{\boldsymbol{k}}^2}{d_{\boldsymbol{k}} + \sqrt{d_{\boldsymbol{k}}^2 + L\sigma_x^{-2}\sigma_{\boldsymbol{k}}^2}} + \lim_{T\to\infty} \frac{1}{T} \int_{t_0}^{t_0+T} |\hat{v}_{\boldsymbol{k}}(t)|^2\, dt\Big), \tag{56}$$

where $\hat{v}_{\boldsymbol{k}}$ is obtained from eq. (36). The radial summation of the energy spectra is then obtain using eq. (43) to obtain the estimate E_k from $E_{\boldsymbol{k}}$.

7 Numerical experiments of the performance and accuracy of the filters

Here we quantify and discuss the performance and accuracy of the conditionally Gaussian nonlinear filters for Lagrangian data assimilation, as described in section 4, and recovery of the Eulerian energy spectra by the algorithms in section 6. We first discuss the performance metric used to quantify filter skill in section 7.1, and then in section 7.2, include the results and analysis for our numerical experiments testing recovery for velocity models without a mean and background, in two different turbulence regimes.

7.1 Performance measures

Here we define quantitative performance measures that describe the accuracy of the filtering estimates relative to the truth. Recall that the spectrally truncated forecast model is resolved on I_{Λ_f} modes and the truth model on I_Λ modes:

$$\boldsymbol{v}^f(\boldsymbol{x},t) = \boldsymbol{w}^f(t) + \sum_{\boldsymbol{k}\in I_{\Lambda_f}} \hat{v}^f_{\boldsymbol{k}}(t)\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}) \quad \text{and} \quad \boldsymbol{v}(\boldsymbol{x},t) = \boldsymbol{w}(t) + \sum_{\boldsymbol{k}\in I_{\Lambda}} \hat{v}_{\boldsymbol{k}}(t)\boldsymbol{\psi}_{\boldsymbol{k}}(\boldsymbol{x}), \tag{57}$$

respectively, where $\boldsymbol{v}^f \in \mathcal{H}_{\Lambda_f}$ and $\boldsymbol{v} \in \mathcal{H}_\Lambda$, and where the space spanned by the forecast model $\mathcal{H}_{\Lambda_f}$ is a subset of the space spanned by the truth $\mathcal{H}_\Lambda$, which are both finite dimensional subspaces of a Hilbert space (these distinctions are important, since we can define error measures in either space, which have different interpretations).

The *root mean square error* (MSE) between the truth and the filtering estimate is given by the L_2 norm of the residual $\boldsymbol{v}_N - \boldsymbol{v}$ in the spaces $\mathcal{H}_\Lambda$ and $\mathcal{H}_{\Lambda_f}$:

$$\big(\mathrm{RMSE}(\boldsymbol{v}^f,\boldsymbol{v})\big)^2 := \|\boldsymbol{v}^f - \boldsymbol{v}\|^2_{L^2(\mathcal{H}_\Lambda)} = \|\boldsymbol{w}^f - \boldsymbol{w}\|^2 + \sum_{\boldsymbol{k}\in I_\Lambda} |\hat{v}^f_{\boldsymbol{k}}(t) - \hat{v}_{\boldsymbol{k}}(t)|^2, \tag{58}$$

$$\big(\mathrm{RMSE}_f(\boldsymbol{v}^f,\boldsymbol{v})\big)^2 := \|\boldsymbol{v}^f - P_{\Lambda_f}\boldsymbol{v}\|^2_{L^2(\mathcal{H}_{\Lambda_f})} = \|\boldsymbol{w}^f - \boldsymbol{w}\|^2 + \sum_{\boldsymbol{k}\in I_{\Lambda_f}} |\hat{v}^f_{\boldsymbol{k}}(t) - \hat{v}_{\boldsymbol{k}}(t)|^2, \tag{59}$$

where P_{Λ_f} is the projection onto $\mathcal{H}_{\Lambda_f}$, i.e. the modes greater than the Λ_f modes resolved by the forecast model are set to zero. We denote the root mean square error by RMSE. Similarly, we define the *pattern correlation* $0 \leqslant \mathrm{XC} \leqslant 1$ (taking the real component, since the real and imaginary parts are equal):

$$\mathrm{XC}(\boldsymbol{v}_N,\boldsymbol{v}) := \frac{\langle \boldsymbol{v}^f, \boldsymbol{v}\rangle_{L^2(\mathcal{H}_\Lambda)}}{\|\boldsymbol{v}^f\|_{L^2(\mathcal{H}_\Lambda)}\|\boldsymbol{v}\|_{L^2(\mathcal{H}_\Lambda)}}, \qquad \mathrm{XC}_f(\boldsymbol{v}_N,\boldsymbol{v}) := \frac{\langle \boldsymbol{v}^f, P_{\Lambda_f}\boldsymbol{v}\rangle_{L^2(\mathcal{H}_{\Lambda_f})}}{\|\boldsymbol{v}^f\|_{L^2(\mathcal{H}_{\Lambda_f})}\|P_{\Lambda_f}\boldsymbol{v}\|_{L^2(\mathcal{H}_{\Lambda_f})}}., \tag{60}$$

We additionally define the error of the radially averaged spectra, the *root mean square spectrum error* (RMSSE) by

$$\big(\mathrm{RMSSE}(E_k, E^f_k)\big)^2 := \|E_k - E^f_k\|^2_{\ell^2(\mathcal{H}_\Lambda)} = \sum_{k=0}^{\Lambda} |E(k) - E^f(k)|^2, \tag{61}$$

where Λ is the maximum wavenumber resolved; a similar definition holds for RMSSE_f with the summation instead from $k = 0$ to $k = \Lambda_f$. The distinction between the root mean square spectrum error and the root mean square error, is that the spectrum error measures an error that incorporates the filter posterior covariance.

The performance measures described above are averaged in time to obtain average filter errors. The RMSE error quantifies the proximity of the filtered signal relative to the truth signal on the Λ resolved modes by the truth model. The ratio of the RMS error and the average magnitude of the truth signal gives the normalized percentage error, which we denote by NRMSE (normalized root mean square error); the normalized RMS error of the radially averaged spectra is also similarly defined, and is denoted by NRMSSE (normalized root mean square spectrum error). The pattern correlation, quantifies how close we recover the pattern of the truth signal, i.e. how closely the two signals align with each other. Although we include the pattern correlation of the filtering mean in the results below, we note that the pattern correlation does not predict how well the filtering mean performs for spectra recovery, since that is based on the 'closeness' metric described by the RMSE and RMSSE.

The definitions of the error metrics in the space spanned by the true velocity model $\mathcal{H}_\Lambda$, naturally introduces an information barrier represented by the sum of the errors from the modes $\Lambda_f < k_i \leqslant \Lambda$ that

are not resolved by the forecast dynamics. Defining the RMSE, RMSSE, and XC over the Λ_f modes resolved by forecast as in RMSE_f, RMSSE_f , and XC_f, represents the performance of the Λ_f resolved modes to the truth signal. We include these measures, since they tell us how well the *resolved modes* perform relative to the truth, without incorporating errors due to the information barrier represented by the modes that are impossible to resolve by the forecast model.

7.2 Numerical results for the perfect and approximate filters for flows without a mean and no background sweeps

In this section, we quantify the performance of the filtering algorithms for the turbulent velocity models for both a Kolmogorov energy spectrum, representing backwards energy cascade with slope $k^{-5/3}$, and a forward cascade energy spectrum with slope k^{-3}, corresponding to direct cascade of energy. The turbulence models studied here are unforced and do not include sweeping, background velocity component. We describe the exact numerical setup in section 7.2.1 and in section 7.2.2 discuss and analyze the results.

7.2.1 Setup and summary of the filtering performance

Table 1 Parameter and numerical simulation values for the truth model and filter model. For these model cases there is no background velocity term $\boldsymbol{w}(t) = 0$ and model forcing is zero $f_{\boldsymbol{k}} = 0$.

(a) model parameters

parameter		*value*
computational domain	D	$[0, 2\pi)^2$
grid size	N^2	32^2
total simulation time	t_{end}	$105\, t_{\text{decor}}$
burn in period	t_{burn}	$5\, t_{\text{decor}}$
damping $d_{\boldsymbol{k}} = d + \mu k$		
linear drag	d	0
viscosity	μ	0.05
energy spectrum $\widetilde{E}_k$		
energy level	E_0	1.0
constant energy modes	k_0	2
slope	α	5/3, 3

(b) filter model parameters

parameter		*value*
grid size	N_f^2	4^2, 8^2, 16^2, 32^2
observation noise	σ_x	0.25
number of drifters	L	1, 2, 4, 8, 16, 32, 64, 128

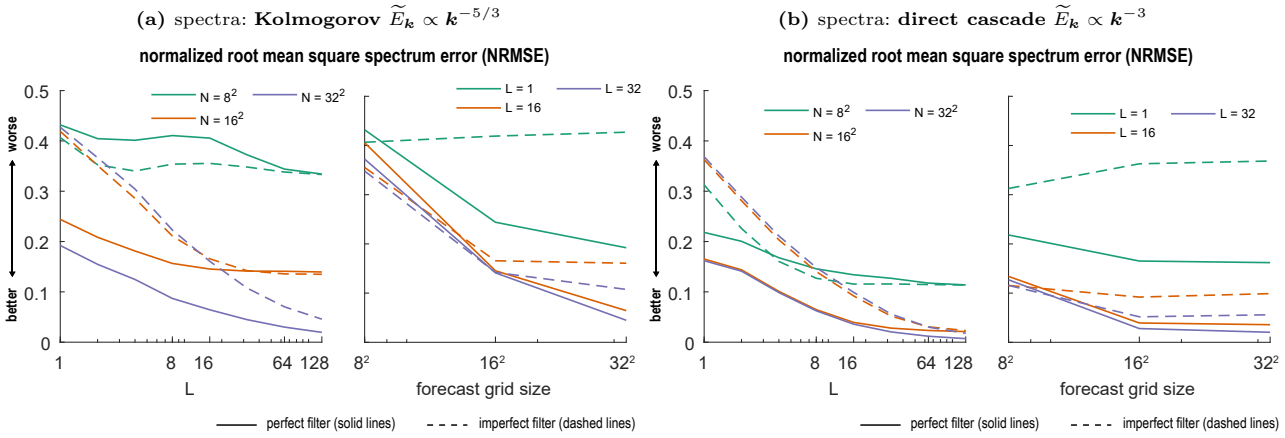


Figure 1 The normalized root-mean-square spectrum error (NRMSSE) for the Kolmogorov (1a) and third power law (1b) models, as a function of the forecast model grid size N_f^2 (true model is $N^2 = 32^2$) and the number of tracers L. The solid lines correspond to the perfect filter and the dashed lines to the imperfect filter; see tab. 1 for the model and filter parameters. For simplicity we do not include all the filter parameters.

The model parameters are provided in tab. 1a and the filter model parameters in tab. 1b. The numerical simulations for both the perfect and diagonal filter equations are simulated according to algs. 1 and 2, respectively. In our implementation, we note that the *perfect filter runs at roughly 10 times the cost of the diagonal filter*, and quickly becomes prohibitive for all but the fewest number of tracer observations and model grid sizes.

In fig. 1, the (normalized root mean) spectra error is included for the Kolmogorov and direct cascade turbulent velocity models. In fig. 2, we include comparisons of the spectrum recovery performance as a function of the number of tracers, when the forecast grid size is $N_f^2 = 16^2$, for the Kolmogorov turbulence model. Fig. 2a, directly compares the filter mean performance to the true spectra; in fig. 2b, the spectra of the filter mean plus the variance component is shown, computed according to eqs. (55) and (56), in order to understand the affect of the variance term of the perfect and approximate filters on spectra recovery. In fig. 4, we include a summary of the performance measures that are based on the filtering mean solution, for the Kolmogorov turbulence model. In fig. 3, we plot the recovered streamfunctions of both the perfect and diagonal filter and compare them relative to the true streamfunction, for $N_f = N = 32$, when $L = 4$ and $L = 128$. Additional supporting information is included in appendix B.1.

7.2.2 Analysis and discussion

We observe that the filter performance results are more marked for velocity models with shallow spectra (intuitively expected, since the system is more energetic), which can be seen in fig. 1 and the supporting data in appendix B.1. As a consequence, in the following discussion, we directly refer to the Kolmogorov velocity model, with the understanding that the conclusions also hold for models with steeper spectra, including the direct cascade turbulence model.

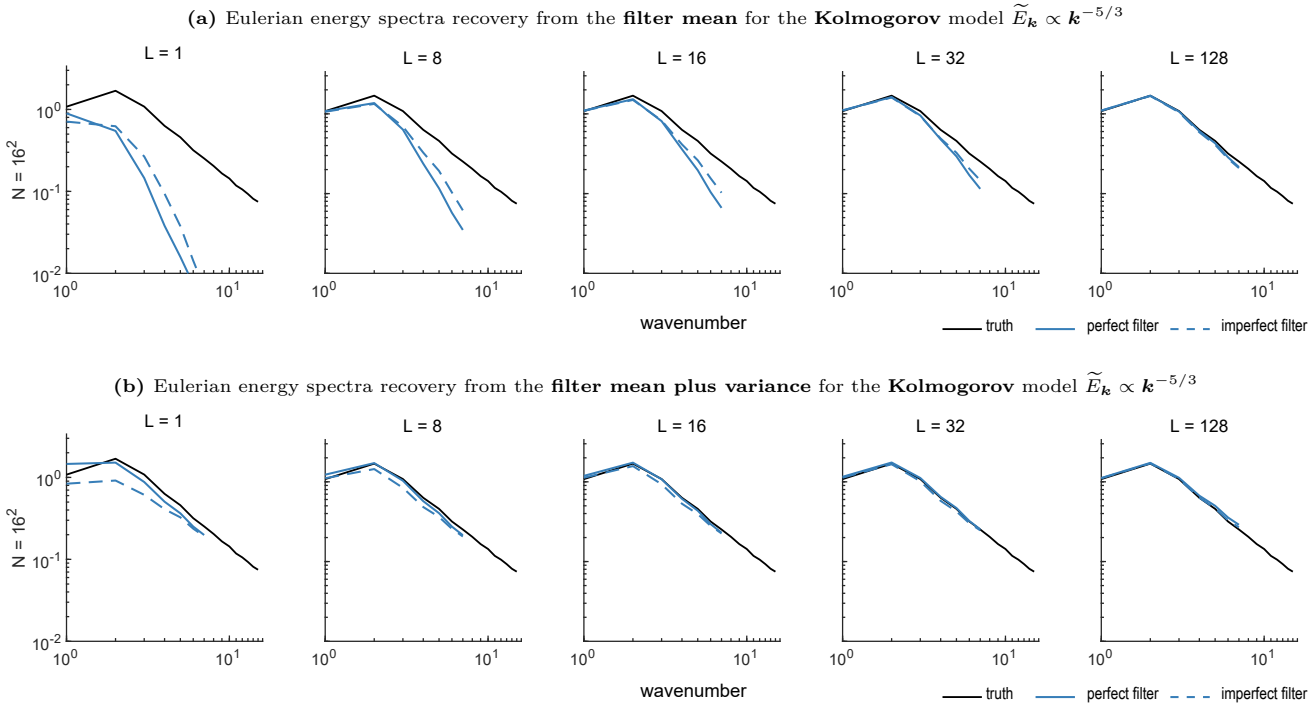


Figure 2 Eulerian energy spectra recovery for the Kolmogorov turbulence model, comparing the full filter in solid blue, the approximate diagonal filter in dashed blue line, and the true spectrum in solid black. The Eulerian spectra are shown for a forecast model of grid size $N_f^2 = 16^2$, where the true model grid size is $N^2 = 32^2$. The columns denote filter results with different numbers of tracer observations L. The approximations involving the variance are obtained from eqs. (55) and (56); see tab. 1 for the model and filter parameters.

From fig. 1, we see that when the forecast grid is coarse relative to the true model grid, there is little difference in spectrum recovery in running the approximate filter compared to the full filter, even when the number of tracer observations is low. On the other hand, when the forecast grid size approaches the true model grid size, and for a very small umber of tracers, we find the largest discrepancy in the spectrum recovery skill of the imperfect filter compared to the exact filter. For a moderate to large number of tracer observations, the approximate filter's performance is comparable to the exact filter. We additionally observe that equivalent skill is achievable by running an imperfect filter with more observations relative to running an exact filter but with fewer observations. This is interesting since the diagonal filter is much more inexpensive compared to the perfect filter; for computational demanding models with plentiful tracer observations, it may be advantageous to simply run the diagonal filter and assimilate more observations, instead of running the perfect filter but limiting the number observations

(due to computational constraints). However, after a large enough number of tracers, both the imperfect filter and perfect filter errors are observed to converge to the same values. We summarize these conclusions below:

- When the forecast grid is coarse relative to the true mesh size, the approximate filter performance is almost as good as the full filter, even for small number of tracer observations.
- If the forecast model fully resolves the mesh scale of the true model, there is a large discrepancy in spectrum recovery skill using the imperfect filter relative to the true filter, when the number of tracer observations are small.
- For a moderate to large number of tracer observations, the approximate filter's performance is comparable to the exact filter.
- Equivalent performance is achievable with an imperfect filter assimilating more observations compared to running the perfect filter with fewer observations.

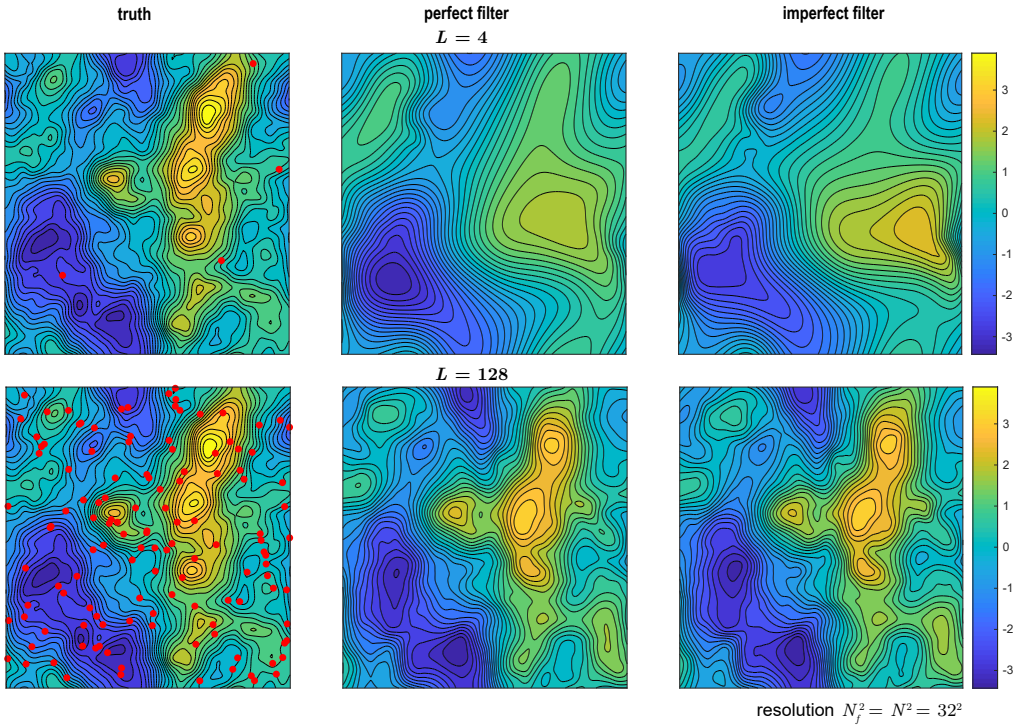


Figure 3 Streamfunction recovery performance for the **Kolmogorov** turbulence model, comparing the truth (left) to the perfect filter (middle) and the diagonal filter (right), when the forecast and model resolution are equivalent $N_f^2 = N^2 = 32^2$. The red dots in the truth streamfunction (left) plots mark the location of the tracer observations. The top row shows the performance with $L = 4$ tracers and the bottom row with $L = 128$ tracers.

From the spectrum recovery figures in fig. 2 and also figs. 5 and 6 in appendix B.1, we observe that the approximate filter's mean tends to over estimate the energy in the smaller scales and under estimate the energy in the larger scales, relative to the perfect filter, regardless of the forecast grid size and number of tracer observations (this can also be observed in fig. 3). These affects are especially pronounced when the number of tracer observations are small. For energy spectra estimation, including the component due to the posterior variance, we also find that the approximate filter systematically underestimates the energy of the largest scales, but here the inclusion of the variance factor appropriately inflates the energy of the prediction of the small scales so that it closely matches the performance of the perfect filter. Again, these effects are most pronounced when the number of tracer observations are small. Additionally, we note that the perfect filter with the variance contribution, systemically underestimates the energy of the smallest scales and over estimates the energy of the largest scales, in cases where there are small tracer observations. When there is severe model truncation and for very shallow spectra, we see that perfect filter overestimates the true spectra; interestingly the under dispersion of the imperfect filter spectrum estimate, in these cases, leads to smaller estimation errors (see also fig. 1). In summary we find that:

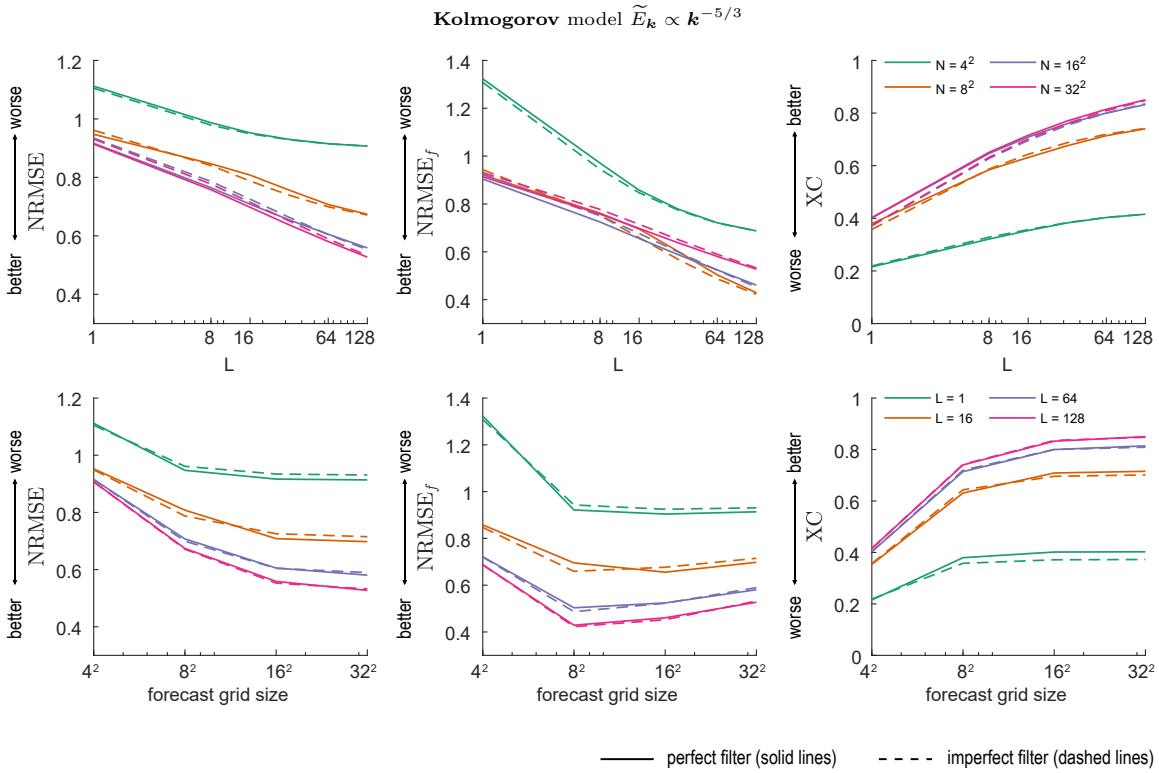


Figure 4 The normalized root-mean-square error (NRMSE) and pattern correlation (XC) for the **Kolmogorov** turbulence model, as a function of the forecast model grid size and the number of tracers. The solid lines correspond to the perfect filter and the dashed lines to the imperfect filter; see tab. 1 for further details on the model and filter parameters.

- The general trend for both the imperfect and perfect filter for spectrum estimation is a relative under dispersion of the small scales and over dispersion of the largest scales, *relative to the true spectrum*.
- The imperfect filter mean tends to underestimate the energy of the largest scales and overestimate the energy of the smallest scales, *relative to the perfect filter*.
- For severely truncated forecast models, lower error is achievable by the imperfect filter, for the spectrum estimate that include the filter variance; this is observed in turbulence models with shallow spectra.

The results in fig. 4 demonstrate robustness of the approximate filter to both the number of tracer observations and model truncation. It is a key point that the imperfect filter performs well even when the number of tracer observations are few, since the derivation of the algorithm is based on a limiting argument as the number of observations L tends to infinity. Robustness of the results in this section have also been shown in numerical experiments for various flow configurations involving forcing and various background velocity terms, see appendix B.2.

8 Conclusions

We studied the time-sequential state estimation of a flow field given noisy measurements provided from Lagrangian tracers that are passively advected by the flow. We discussed special Lagrangian data assimilation algorithms, known as conditionally Gaussian nonlinear filters for two-dimensional linear stochastic turbulence models. Approximate filters are proposed to alleviate the computational costs due to high dimensionality of the perfect nonlinear filter, which are based on extreme localization of the posterior covariance matrix. The proposed diagonal filters are vastly cheaper than the perfect filter and run approximately ten times faster in numerical experiments with reduced storage costs. Through comprehensive numerical experiments, we quantify how accurately such filters can recover the various energetic scales of the true fluid flow model. Although the central questions are addressed in the perfect model scenario, we study model error due to reduced-order forecast models. We find that the diagonal

filter performs comparable to the true filter and is robust to various turbulent flow regimes. Even though the approximation is based on arguments where the number of tracer tends to infinity, we see find that it performs comparable to the perfect filter, even when the number of tracer observations are small. This suggests such filtering algorithms may be useful in more complex situations. Future work aims to investigate issues related to parameter estimation of the forecast model and applying such filtering methods to prototype models used in climate science.

Acknowledgments

A. J. M. acknowledges partial support by the Office of Naval Research through MURI 821 N00014-16-1-2161 and DARPA through W911NF-15-1-0636. M.A.M. is supported as a postdoctoral fellow on both grants.

References

[1] A. Apte, C. K. R. T. Jones, and A. M. Stuart. "A Bayesian approach to Lagrangian data assimilation". In: *Tellus A: Dynamic Meteorology and Oceanography* 60.2 (2008), pp. 336–347.

[2] J. L. Beck and K.-V. Yuen. "Model selection using response measurements: Bayesian probabilistic approach". In: *Journal of Engineering Mechanics* 130.2 (2004), pp. 192–203.

[3] F. J. Beron-Vera and J. H. LaCasce. "Statistics of simulated and observed pair separations in the Gulf of Mexico". In: *Journal of Physical Oceanography* 46.7 (2016), pp. 2183–2199.

[4] M. Branicki, B. Gershgorin, and A. J. Majda. "Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters". In: *Journal of Computational Physics* 231.4 (2012), pp. 1462–1498.

[5] M. Branicki, A. J. Majda, and K. Law. "Accuracy of some approximate gaussian filters for the Navier-Stokes equation in the presence of model error". In: (2018). Submitted.

[6] M. Branicki and A. J. Majda. "An information-theoretic framework for improving imperfect dynamical predictions via multi-model ensemble forecasts". In: *Journal of Nonlinear Science* 25.3 (June 2015), pp. 489–538.

[7] M. Branicki and A. J. Majda. "Dynamic stochastic superresolution of sparsely observed turbulent systems". In: *Journal of Computational Physics* 241 (2013), pp. 333–363.

[8] N. Chen, A. J. Majda, and X. T. Tong. "Information barriers for noisy Lagrangian tracers in filtering random incompressible flows". In: *Nonlinearity* 27.9 (2014), p. 2133.

[9] N. Chen, A. J. Majda, and X. T. Tong. "Noisy Lagrangian tracers for filtering random rotating compressible flows". In: *Journal of Nonlinear Science* 25.3 (June 2015), pp. 451–488.

[10] I. Grooms, Y. Lee, and A. J. Majda. "Ensemble Kalman filters for dynamical systems with unresolved turbulence". In: *Journal of Computational Physics* 273 (2014), pp. 435–452.

[11] T. Janjić and S. E. Cohn. "Treatment of observation error due to unresolved scales in atmospheric data assimilation". In: *Monthly Weather Review* 134.10 (2006), pp. 2900–2915.

[12] A. Jazwinski. *Stochastic Processes and Filtering Theory*. Mathematics in Science and Engineering 64. Academic Press, 1970.

[13] M. Kamachi and J. J. O'Brien. "Continuous data assimilation of drifting buoy trajectory into an equatorial Pacific Ocean model". In: *Journal of Marine Systems* 6.1 (1995). Data Assimilation in Manine Science, pp. 159–178.

[14] L. Kuznetsov, K. Ide, and C. K. R. T. Jones. "A method for assimilation of Lagrangian data". In: *Monthly Weather Review* 131.10 (2003), pp. 2247–2260.

[15] J. H. LaCasce. "Statistics from Lagrangian observations". In: *Progress in Oceanography* 77.1 (2008), pp. 1–29.

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

[17] J. M. Lilly, A. M. Sykulski, J. J. Early, and S. C. Olhede. "Fractional Brownian motion, the Matérn process and stochastic modeling of turbulent dispersion". In: *Nonlinear Processes in Geophysics* 24.3 (2017), pp. 481–514.

[18] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes II. Applications*. 2nd ed. Stochastic Modelling and Applied Probability 6. Springer-Verlag Berlin Heidelberg, 2001.

[19] A. C. Lorenc. "Analysis methods for numerical weather prediction". In: *Quarterly Journal of the Royal Meteorological Society* 112.474 (1986), pp. 1177–1194.

[20] R. Lumpkin and M. Pazos. "Measuring surface currents with Surface Velocity Program drifters: the instrument, its data, and some recent results". In: *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Ed. by A. Griffa, A. D. Kirwan Jr., A. J. Mariano, T. Özgökmen, and H. T. Rossby. Cambridge University Press, 2007, pp. 39–67.

[21] A. J. Majda and B. Gershgorin. "Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum and intermittency". In: *Philosophical Transactions of the Royal Society of London A* 371.1982 (2013).

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

[23] A. J. Majda and P. R. Kramer. "Simplified models for turbulent diffusion: Theory, numerical modelling, and physical phenomena". In: *Physics Reports* 314.4 (1999), pp. 237–574.

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

[25] A. Molcard, T. M. Özgökmen, A. Griffa, L. I. Piterbarg, and T. M. Chin. "Lagrangian data assimilation in ocean general circulation models". In: *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Ed. by A. Griffa, A. D. Kirwan Jr., A. J. Mariano, T. Özgökmen, and H. T. Rossby. Cambridge University Press, 2007, pp. 172–203.

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

[27] T. Rossby. "Evolution of Lagrangian methods in oceanography". In: *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*. Ed. by A. Griffa, A. D. Kirwan Jr., A. J. Mariano, T. Özgökmen, and H. T. Rossby. Cambridge University Press, 2007, pp. 38–1.

[28] K.-V. Yuen, J. L. Beck, and L. S. Katafygiotis. "Probabilistic approach for modal identification using non-stationary noisy response measurements only". In: *Earthquake Engineering & Structural Dynamics* 31.4 (2002), pp. 1007–1023.

A Notation

Below we define the notation used in this work:

- Vector in bold italic notation, such as $\boldsymbol{X}, \boldsymbol{x}$
- Vectors may be denoted by $[a;\, b;\, c]$ or $[a, b, c]^{\mathsf{T}}$
- Matrix in uppercase non-bold notation, e.g. A or bold upright notation, e.g. $\mathbf{A}$
- Entries of matrix A by $[A]_{ij} = a_{ij}$
- Unit vector in the direction $\boldsymbol{a}$ by $\hat{\boldsymbol{a}}$
- ℓ_2 norm of vector $\boldsymbol{a}$ by $\|\boldsymbol{a}\|$ or $\|\boldsymbol{a}\|_{\ell_2}$ or simply a
- L_2 norm of a function f by $\|f\|$ or $\|f\|_{L_2}$
- Hilbert space by $\mathcal{H}$
- Identity matrix by I or more explicitly $I_{n\times n}$ for an $n \times n$ identity matrix.
- Absolute value or modulus by $|\cdot|$
- Hermitian transpose of matrix A by A^*
- Statistical mean of a random variable or stochastic process by the notation $\langle \cdot \rangle$, i.e. for a variable x we write $\overline{x} = \langle x \rangle$.
- Fluctuations of a random variable or stochastic process x by $\tilde{x} = x - \overline{x}$
- Dependence on the probability space of a random variable by $x(\omega)$ or for a stochastic process by $x(t;\omega)$, where $\omega \in \Omega$ is an element of the probability space Ω.
- Time dependence of a stochastic process by $x(t)$ or x_t

B Supplementary information

B.1 Numerical results for the perfect and approximate filters for flows without forcing and no background velocity

In figs. 5 and 6 we include comparisons of the spectral recovery performance for various number of tracers and forecast grid sizes, for the Kolmogorov and direct cascade spectra, respectively. We include two subplots in these figures: (a) compares the filter mean without the covariance; and (b) compares the filter mean plus the covariance, to show the impact of the covariance in the perfect and approximate assimilation methods on their mean. Additionally, in figs. 7 and 8 we include the NRMSE and XC error measures as a function of the grid size and number of tracers, also for both turbulence models.

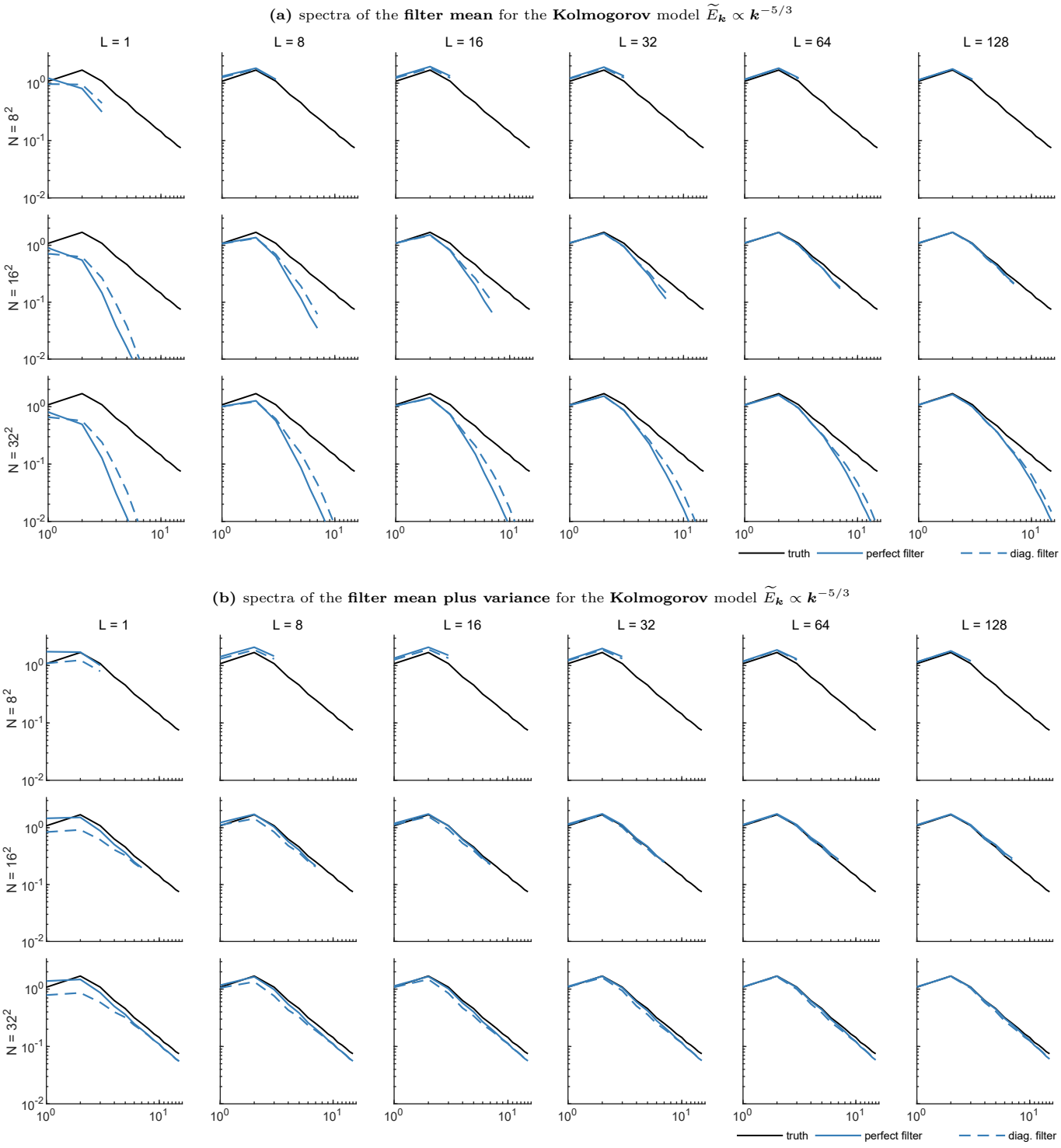
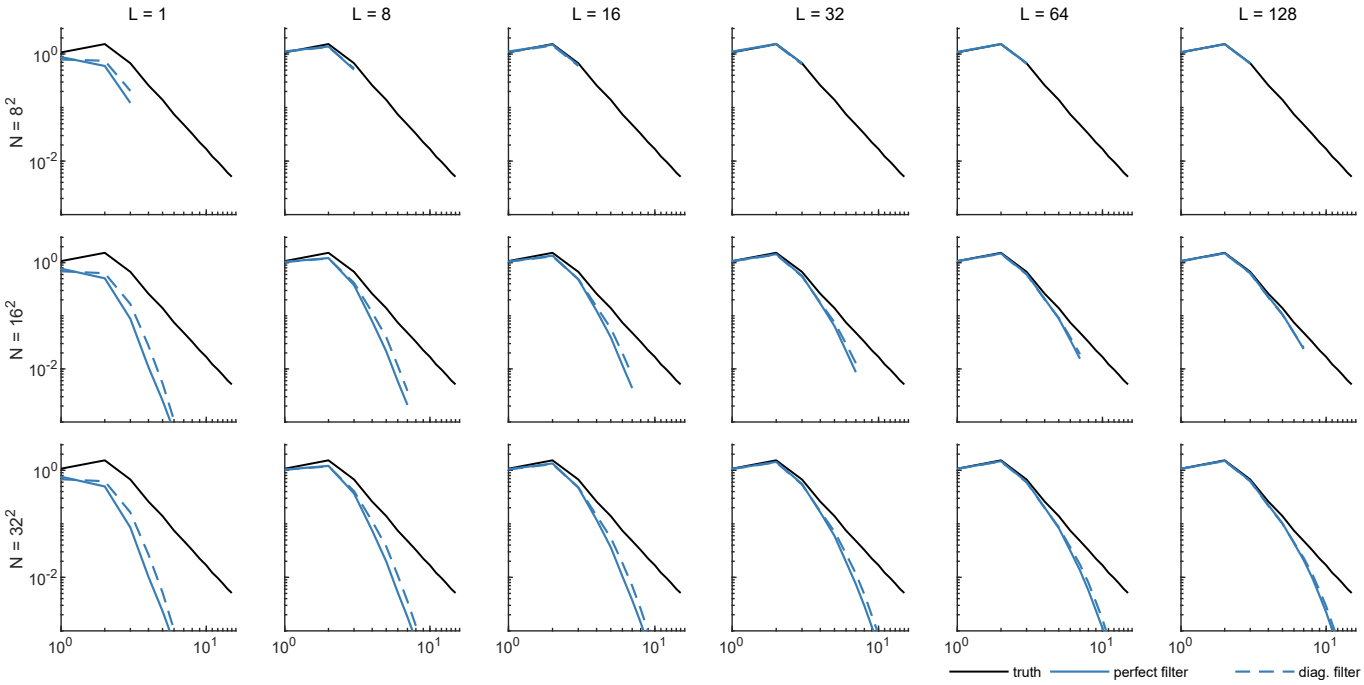


Figure 5 Eulerian energy spectra recovery performance for the **Kolmogorov** turbulence model, comparing the full filter in solid blue and the approximate diagonal filter in dashed blue line, where the true spectrum is shown in solid black. In the matrix of figures above, each columns represents a forecast model when the observations are given by L tracers and rows denote the forecast model grid size; see tab. 1 for the model and filter parameters.

(a) spectra of the **filter mean** for the **direct cascade** model $\widetilde{E}_{\boldsymbol{k}} \propto \boldsymbol{k}^{-3}$



(b) spectra of the **filter mean plus variance** for the **direct cascade** model $\widetilde{E}_{\boldsymbol{k}} \propto \boldsymbol{k}^{-3}$

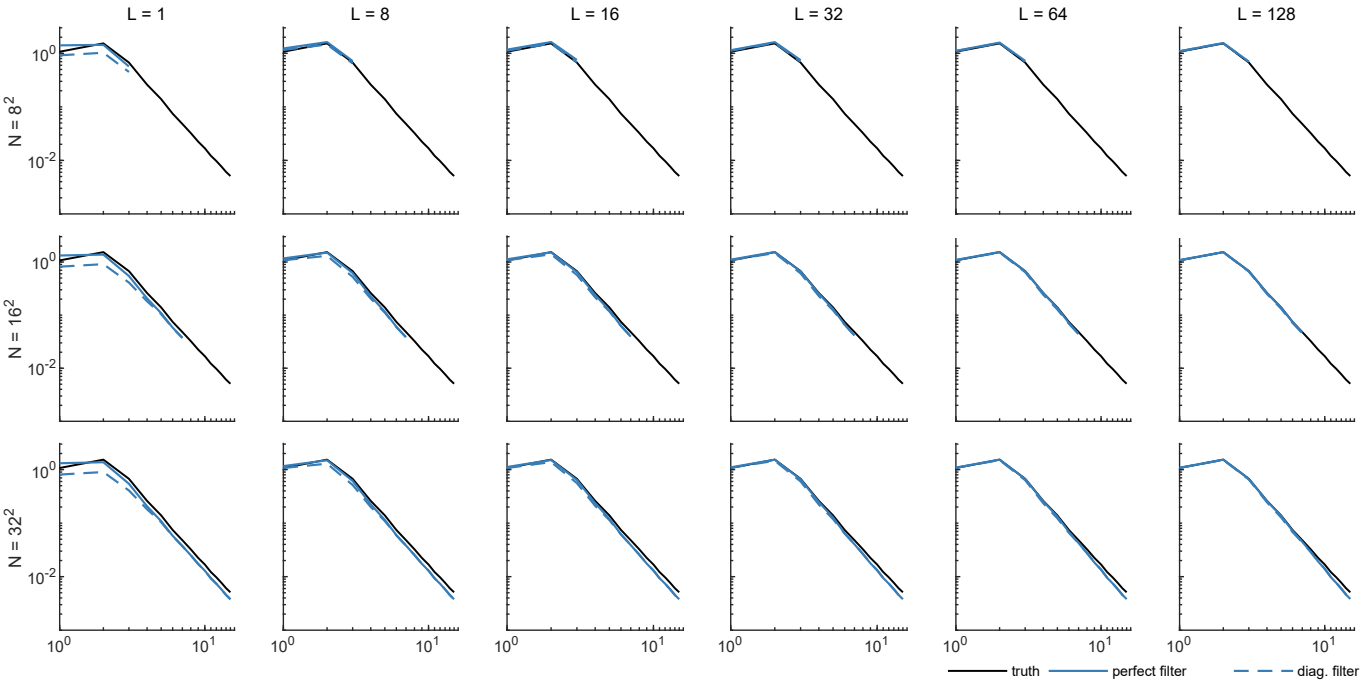


Figure 6 Eulerian energy spectra recovery performance the **direct cascade** turbulence model, comparing the full filter in solid blue and the approximate diagonal filter in dashed blue line, where the true spectrum is shown in solid black. In the matrix of figures above, each columns represents a forecast model when the observations are given by L tracers and rows denote the forecast model grid size; see tab. 1 for the model and filter parameters.

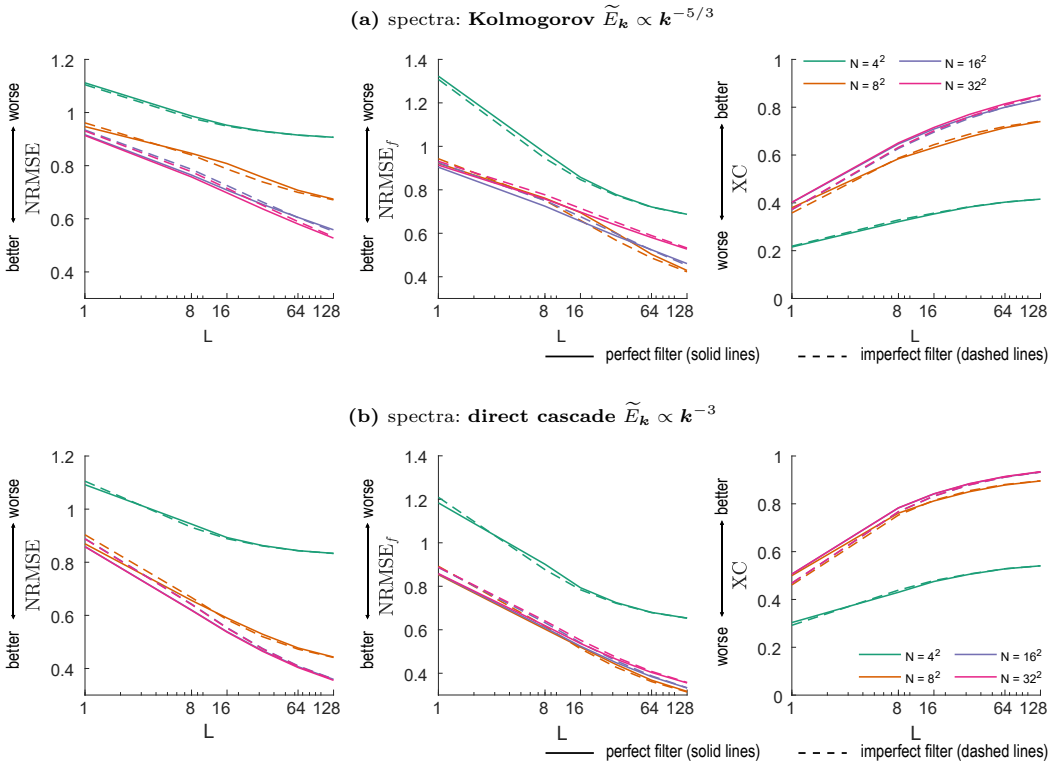


Figure 7 The normalized room-mean-square error (NRMSE) and pattern correlation (XC), as a function of the number of tracers L, where dashed lines denote the diagonal filter and solid lines the perfect filter; see tab. 1 for the model and filter parameters. Fig. 7a includes the results for a Kolmogorov velocity model and in fig. 7b a third power law model.

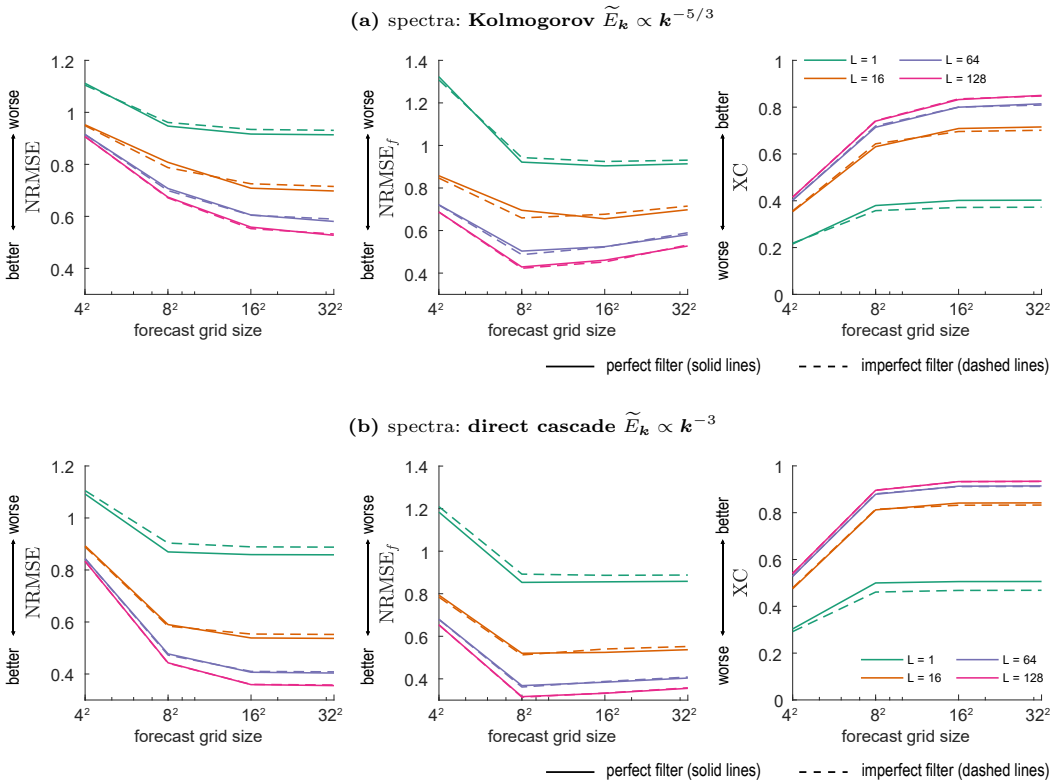


Figure 8 The normalized room-mean-square error (NRMSE) and pattern correlation (XC), as a function of the forecast model grid size, where dashed lines denote the diagonal filter and solid lines the perfect filter; see tab. 1 for further details on the model and filter parameters. Fig. 8a includes the results for a Kolmogorov velocity model and in fig. 8b a third power law model. The dashed lines denote the diagonal filter and solid lines the perfect filter.

B.2 Numerical results for the diagonal filters in various flow regimes

Here we study the diagonal filter in various flow regimes to demonstrate robustness of the diagonal filter, when there is no model truncation error and using a truth model with a Kolmogorov spectrum. We investigate three flow regimes with parameters in tab. 2 (see tab. 1 for the rest of the model parameters). They include a combination of flows with/without forcing and with/without a background velocity term. The forcing is applied to induce a dominant shear structure in the flow.

In fig. 9 we include comparisons of the spectra recovery performance as a function of the number of tracers. We see that model forcing greatly enhances recovery of the total energy spectra, since forcing strongly biases the forced mode. We find that there is no difference in the skill of the filter for recovery the spectra of the various scales in the flow when there is a background flow compared to conditions that include a background term.

Table 2 Different flow regimes including cases with and without a mean velocity and cases with and without a background velocity term; see tab. 1 for the rest of the model parameters

(a) Unforced flow with no background.

background		
damping	d_0	0
noise	σ_0	0
forcing	f_0	0
model forcing		
forcing	$f_{\boldsymbol{k}}$	0

(b) Unforced flow with a background.

background		
damping	d_0	2
noise	σ_0	0.8
forcing	f_0	0
model forcing		
forcing	$f_{\boldsymbol{k}}$	0

(c) Forced flow with a background.

background		
damping	d_0	2
noise	σ_0	0.8
forcing	f_0	0
model forcing on $\boldsymbol{k} = (1, 0)$		
forcing	$f_{\boldsymbol{k}}$	0.5

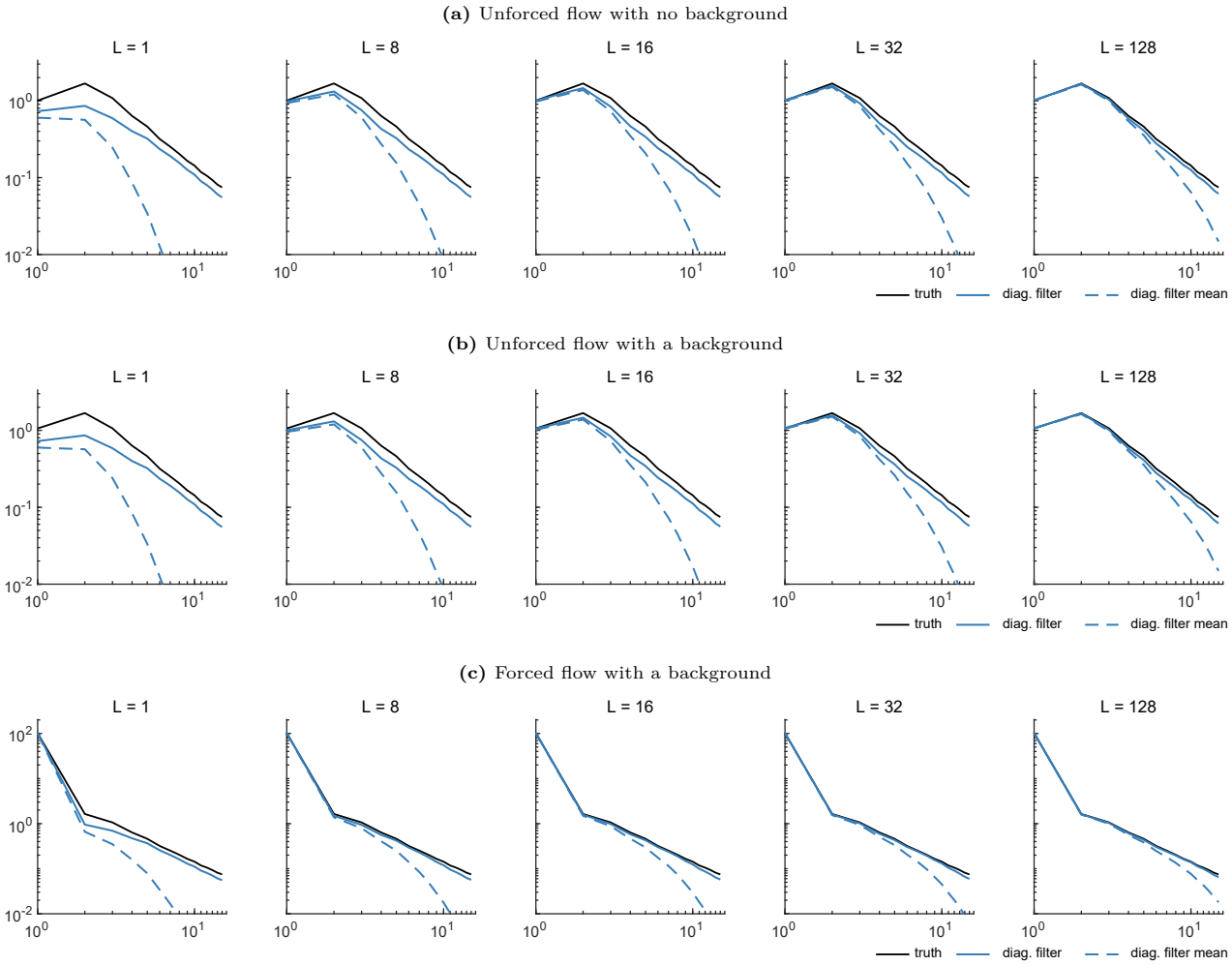


Figure 9 Eulerian energy spectra recovery performance for the **Kolmogorov** turbulence model, comparing the full filter in solid blue and the approximate diagonal filter in dashed blue line, where the true spectrum is shown in solid black; see tab. 2 for the model parameters for the various cases (also tab. 1 for the main model parameters).