Small-scale reconstruction in three-dimensional Kolmogorov flows using four-dimensional variational data assimilation

Yi Li[1,†], Jianlei Zhang[2], Gang Dong[2] and Naseer S. Abdullah[1]

[1]School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK

[2]National Key Laboratory of Transient Physics, Nanjing University of Science and Technology, Nanjing 210094, PR China

Significant insights in computational fluid dynamics have been obtained in recent years by adopting the data assimilation methods developed in the meteorology community. We apply the four-dimensional variational method to reconstruct the small scales of three-dimensional turbulent velocity fields with a moderate Reynolds number, given a time sequence of measurement data on a coarse set of grid points. The problem presents new challenges because the evolution of the flow is dominated by the nonlinear vortex stretching and the energy cascade process, which are absent from two-dimensional flows. The results show that, reconstruction is successful when the resolution of the measurement data, given in terms of the wavenumber, is of the order of the threshold value $k_c = 0.2\eta_K^{-1}$, where η_K is the Kolmogorov length scale of the flow. When the data are available over a period of one large eddy turnover time scale, the filtered enstrophy and other small-scale quantities are reconstructed with a 30 % or smaller normalized pointwise error, and a 90 % pointwise correlation. The spectral correlation between the reconstructed and target fields is higher than 80 % for all wavenumbers. Minimum volume enclosing ellipsoids (MVEEs) and MVEE trees are introduced to quantitatively compare the geometry of non-local structures. Results show that, for the majority samples, errors in the locations and the sizes of the reconstructed structures are within 15 %, and those in the orientations are within 15°. Overall, for this flow, satisfactory reconstruction of the scales two or more octaves smaller is possible if data at large scales are available for at least one large eddy turnover time. In comparison, a direct substitution scheme results in three times bigger pointwise discrepancy in enstrophy. The spectral difference between the reconstructed and target velocity fields is more than ten times higher than what is obtained with the four-dimensional variational method. The results show that further investigation is warranted to verify the efficacy of the method in flows with higher Reynolds numbers.

Key words: computational methods, turbulence simulation, turbulence modelling

† Email address for correspondence: yili@sheffield.ac.uk

1. Introduction

Data assimilation (DA) uses available experimental or observational data to improve computational model predictions. It has had a long history in meteorology, in particular numerical weather prediction (NWP) research (see, e.g. Courtier *et al.* (1993) and Kalnay (2003) for reviews). Many DA methods have been developed. Four-dimensional variational (4DVAR) methods and the ensemble Kalman filter (EnKF) are among the most popular methods (Kalman 1960; Kalnay 2003; Evensen 2009). The 4DVAR methods are based on optimal control (Lions 1971), where a constrained optimization problem is solved. Spatial as well as temporal data are assimilated into the model by minimizing the difference between the model prediction and the data. In recent applications, time sequences of three-dimensional spatial data have been assimilated, hence the name 4DVAR. EnKF is a sequential method where the error of the prediction is estimated when the prediction (known as the *a priori* estimate (Brown & Hwang 2012) or the background (Kalnay 2003)) is made. The observational data are assimilated by combining the *a priori* estimate with the measurement data, resulting in an updated prediction called the *a posteriori* estimate or the analysis. The weight for the measurement is chosen to minimize the error of the *a posteriori* estimate.

In recent years, these two methods have been applied to fluid dynamic problems (Hayase 2015). Gronskis, Heitz & Memin (2013) apply the variational method to reconstruct the inflow and initial conditions for two-dimensional (2-D) mixing layers and wake flows behind a cylinder. Mons *et al.* (2016) consider 2-D wake flows, where a comprehensive comparison between different DA schemes is presented. Kato & Obayashi (2013), Kato *et al.* (2015) and Li *et al.* (2017) use the EnKF method to improve Reynolds-averaged Navier–Stokes based prediction of turbulent flows. Meldi & Poux (2017) applies EnKF to large eddy simulations and detached eddy simulation. Similar problems are also investigated with variational methods (Foures *et al.* 2014). In D'adamo *et al.* (2007), Artana *et al.* (2012) and Protas, Noack & Osth (2015), variational methods are coupled with reduced-order models based on Galerkin truncation. Variational methods are also used in state estimation in the context of flow control (Bewley & Protas 2004; Chevalier *et al.* 2006; Colburn, Cessna & Bewley 2011), to extrapolate experimental data in a dynamically consistent way (Heitz, Memin & Schnorr 2010; Combes *et al.* 2015) and to obtain optimal sensor placements (Mons, Chassaing & Sagaut 2017). Mons *et al.* (2014) apply a variational method to investigate decaying isotropic turbulence, where the eddy-damped quasi-normal Markovian model is used.

The aforementioned research has demonstrated the potential of DA in turbulent simulations as well as in understanding the physics of turbulent flows. As having been demonstrated in NWP research, DA is unique in its ability to improve our modelling or prediction of instantaneous flow fields, in addition to their statistics. However, few studies have explored this aspect in the simulation of 3-D turbulent fields (see, e.g. the list of recent works on DA tabulated in Mons *et al.* (2017)). In Yoshida, Yamaguchi & Kaneda (2005), the ability of a DA scheme to recover the instantaneous small scales in a 3-D isotropic turbulent field is investigated. In this study, the data, given as Fourier modes, directly replace model predictions given by the Navier–Stokes equations at every time step. It is found that the small-scale instantaneous velocity field can be asymptotically recovered exactly when Fourier modes with wavenumber up to a threshold value approximately equal $k_c \equiv 0.2\eta_K^{-1}$ are provided, where η_K is the Kolmogorov length scale. When the amount of data decreases towards the threshold, the time needed to recover the small scales tends to

infinity (hence an infinitely long time sequence of measurement data is needed). The problem is investigated in Lalescu, Meneveau & Eyink (2013) from the perspective of chaos synchronization and a similar conclusion is reached. The DA method used in these works can be termed 'direct substitution' (DS). As far as we know, the ability of 4DVAR or EnKF to reconstruct the small scales has not been investigated. In this paper, we present an analysis based on the 4DVAR method. We consider a Kolmogorov flow in a 3-D periodic box. It is assumed that a time sequence of velocity data is given on a set of grid points. The 4DVAR method is employed to reconstruct the initial velocity field such that the velocity at later times matches given measurement data. The time sequence of velocity fields computed from the initial field are compared with the 'true' velocity fields (the target fields). The objective is to ascertain how well the instantaneous small-scale velocity fields can be reconstructed with 4DVAR for a given set of parameters. The 3-D turbulence in a periodic box is the simplest turbulent flow where the nonlinear inter-scale interaction plays the dominant role in the dynamics. The vortex stretching process, e.g. is absent in 2-D flows. Therefore, although the mathematics is similar for 2-D and 3-D problems, the latter does present significant new challenges.

Using the 4DVAR method to reconstruct a 3-D fully developed (i.e. statistically stationary) turbulent field is the first contribution of this investigation. To evaluate the reconstruction, it is important to quantify the instantaneous difference between the reconstructed and the target velocity fields. Much information can be learned from pointwise correlations and/or the statistics of pointwise difference. However, these statistics are not best suited to capturing the geometry of non-local structures populating the small scales of turbulent fields (such as the vortex filaments). This deficiency becomes a significant obstacle when the distribution of the non-local structures becomes one of our main interests.

The morphology of non-local structures has long been described in qualitative terms, assisted by visualization. Great efforts have been made in recent years to develop methods for quantitative description. Bermejo-Moreno & Pullin (2008) use the probability density function (PDF) of the curvatures on the surface of a structure as its signature. Yang & Pullin (2011) describe the structures in a channel flow in terms of curvelets and angular spectra. Indices, named shapefinders, defined in terms of the Minkowski functionals, are used in Leung, Swaminathan & Davidson (2012) to classify the structures. These methods provide very detailed quantitative descriptions of the structures. However, they focus on the intrinsic geometry; the information about locations, orientations and sizes of the structures sometimes is missing, which happens to be important when we compare the geometry in two different fields. Therefore, as the second contribution of this investigation, we propose to use minimum volume enclosing ellipsoids (MVEEs) to describe the geometry of a non-local structure, and use MVEE trees where the structures are highly non-convex. MVEE is used widely in areas such as statistical estimate, cluster analysis and image processing (Todds 2016). Its application in turbulence research, however, has not been reported. The results demonstrate that MVEEs and MVEE trees are useful tool for the analysis of the non-local geometry in turbulence.

The paper is organized as follows. In § 2, the 4DVAR formulation of the problem is introduced; the description of the small scales of homogeneous turbulence is reviewed, and then the calculation of MVEEs and the MVEE tree is explained. In § 3, the simulations and the results are presented and analysed. The conclusions are summarized in § 4.

2. The governing equations and the description of small scales of turbulent velocity fields

2.1. *The problem set-up and the optimality system*

Let B denote the three-dimensional periodic box $[0, 2\pi]^3$, and $\boldsymbol{v}(\boldsymbol{x}, t)$ be a time sequence of velocity fields in B for $t \in [0, T]$ with $T > 0$; T is called the optimization horizon hereafter. It is assumed that only part of $\boldsymbol{v}(\boldsymbol{x}, t)$ is known from measurement, and this part is denoted by $\mathcal{F}\boldsymbol{v}$, where $\mathcal{F}$ is a filter to be defined later; $\mathcal{F}$ plays the same role as the measurement operator used in the meteorology community. The question is, from given $\mathcal{F}\boldsymbol{v}(\boldsymbol{x}, t)$, how do we reconstruct the full field $\boldsymbol{v}(\boldsymbol{x}, t)$ for $t \in [0, T]$?

To do so, an underlying 'model' for the data $\boldsymbol{v}(\boldsymbol{x}, t)$ has to be assumed. We assume $\boldsymbol{v}(\boldsymbol{x}, t)$ is the solution of the Navier–Stokes equations (NSEs), from some unknown initial condition (IC). Let $\boldsymbol{u}(\boldsymbol{x}, t)$ be a solution of the NSEs in B. If one finds a velocity field $\boldsymbol{\varphi}(\boldsymbol{x})$ such that, if $\boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{\varphi}(\boldsymbol{x})$ then $\boldsymbol{u}(\boldsymbol{x}, t)$ agrees with $\boldsymbol{v}(\boldsymbol{x}, t)$ for $(\boldsymbol{x}, t) \in B \times [0, T]$, then $\boldsymbol{u}(\boldsymbol{x}, t)$ is a reconstruction of $\boldsymbol{v}(\boldsymbol{x}, t)$. It is unlikely to find $\boldsymbol{\varphi}(\boldsymbol{x})$ such that $\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{v}(\boldsymbol{x}, t)$ exactly. In 4DVAR, one attempts to approximate $\boldsymbol{\varphi}(\boldsymbol{x})$ by the solution of a constrained optimization problem. The detail is described in what follows.

We define the inner product between two scalar or vector fields $\boldsymbol{a}(\boldsymbol{x}, t)$ and $\boldsymbol{b}(\boldsymbol{x}, t)$ as

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle \equiv \frac{1}{(2\pi)^3} \int_{\boldsymbol{x} \in B} \boldsymbol{a} \cdot \boldsymbol{b} \, \mathrm{d}\boldsymbol{x}. \tag{2.1}$$

The cost function J for the optimization problem is defined as

$$J = \frac{1}{2} \int_0^T \langle \mathcal{F}\boldsymbol{u} - \mathcal{F}\boldsymbol{v}, \mathcal{F}\boldsymbol{u} - \mathcal{F}\boldsymbol{v} \rangle \, \mathrm{d}t. \tag{2.2}$$

Qualitatively speaking, J is the total difference between the measurements based on $\boldsymbol{u}$ and $\boldsymbol{v}$ over the space–time domain $B \times [0, T]$. The objective of the optimization problem is to find $\boldsymbol{\varphi}(\boldsymbol{x})$ such that J is minimized while $\boldsymbol{u}(\boldsymbol{x}, t)$ satisfies

$$\boldsymbol{N}(\boldsymbol{u}, p) \equiv -\partial_t \boldsymbol{u} - \boldsymbol{u} \cdot \nabla \boldsymbol{u} - \nabla p + \nu \nabla^2 \boldsymbol{u} + \boldsymbol{f} = 0, \quad \nabla \cdot \boldsymbol{u} = 0, \quad \boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{\varphi}(\boldsymbol{x}), \tag{2.3a–c}$$

where p is the pressure, ν is the kinematic viscosity and $\boldsymbol{f}(\boldsymbol{x}, t)$ is the forcing term. The flow has been assumed to be incompressible.

The optimal solution for the problem will be found with the adjoint method. We introduce the Lagrangian functional,

$$L(\boldsymbol{u}, \boldsymbol{\xi}, \sigma, \lambda, \boldsymbol{\varphi}) = J(\boldsymbol{u}) - \int_0^T \langle \boldsymbol{\xi}, \boldsymbol{N} \rangle \, \mathrm{d}t - \int_0^T \langle \sigma, \nabla \cdot \boldsymbol{u} \rangle \, \mathrm{d}t - \langle \lambda, \boldsymbol{u}(\boldsymbol{x}, 0) - \boldsymbol{\varphi}(\boldsymbol{x}) \rangle, \tag{2.4}$$

where $\boldsymbol{\xi}(\boldsymbol{x}, t)$, $\sigma(\boldsymbol{x}, t)$ and $\lambda(\boldsymbol{x})$ are the adjoint variables corresponding to the NSEs, the continuity equation and the IC; $\boldsymbol{\xi}$ and σ are also called the adjoint velocity and the adjoint pressure. The minima of J are found at the stationary points for the Lagrangian, where the functional derivatives of L with respect to $\boldsymbol{u}$, $\boldsymbol{\xi}$, σ, λ and $\boldsymbol{\varphi}$ are zero. These conditions give rise to equations for the adjoint variables (i.e. the adjoint equations) in addition to (2.3). The equations for $\boldsymbol{\xi}$ read,

$$-\partial_t \boldsymbol{\xi} - \boldsymbol{u} \cdot \nabla \boldsymbol{\xi} + \nabla \boldsymbol{u} \cdot \boldsymbol{\xi} + \nabla \sigma - \nu \nabla^2 \boldsymbol{\xi} - \boldsymbol{F} = 0, \quad \nabla \cdot \boldsymbol{\xi} = 0, \tag{2.5a,b}$$

with the endpoint condition

$$\boldsymbol{\xi}(\boldsymbol{x}, T) = 0. \tag{2.6}$$

The forcing term $\boldsymbol{F}$ in (2.5) is given by

$$\boldsymbol{F}(\boldsymbol{x}, t) \equiv -\frac{\partial J}{\partial \boldsymbol{u}} = -\mathcal{F}^{+} \mathcal{F}[\boldsymbol{u}(\boldsymbol{x}, t) - \boldsymbol{v}(\boldsymbol{x}, t)], \tag{2.7}$$

where $\mathcal{F}^{+}$ is the adjoint operator of $\mathcal{F}$. The forcing term $\boldsymbol{f}(\boldsymbol{x}, t)$ in (2.3) has been assumed to be independent of $\boldsymbol{u}$ and, as a consequence, does not have a counterpart in the adjoint equations. When $\boldsymbol{u}$, p, $\boldsymbol{\xi}$ and σ satisfy (2.3) and (2.5)–(2.7), the gradient of J with respect to $\boldsymbol{\varphi}$, denoted by $\mathrm{D}J/\mathrm{D}\boldsymbol{\varphi}$, is given by

$$\frac{\mathrm{D}J}{\mathrm{D}\boldsymbol{\varphi}} = \frac{\partial L}{\partial \boldsymbol{\varphi}} = -\boldsymbol{\xi}(\boldsymbol{x}, 0). \tag{2.8}$$

Equations (2.3), (2.5)–(2.7) and the condition $\mathrm{D}J/\mathrm{D}\boldsymbol{\varphi} = 0$ constitute the optimality system of the optimization problem. The solution of the optimization problem is a solution of the optimality system.

The solution method of the optimality system will be explained later. The optimal solution for $\boldsymbol{\varphi}(\boldsymbol{x})$ will be denoted as $\boldsymbol{\varphi}^{u}(\boldsymbol{x})$. The solution $\boldsymbol{u}(\boldsymbol{x}, t)$ of (2.3) with $\boldsymbol{\varphi}(\boldsymbol{x}) = \boldsymbol{\varphi}^{u}(\boldsymbol{x})$ is the reconstruction of $\boldsymbol{v}(\boldsymbol{x}, t)$ from $\mathcal{F}\boldsymbol{v}(\boldsymbol{x}, t)$. In this investigation, a sequence of known fully developed turbulent velocity fields in the 3-D periodic box B is used as $\boldsymbol{v}(\boldsymbol{x}, t)$; $\mathcal{F}$ is chosen as a cutoff filter in the Fourier space with a cutoff wavenumber k_m, such that only the low wavenumber Fourier modes are used in the optimization problem; k_m is an indicator of the spatial resolution of the measurement data. As a result, the cost function J can be written as

$$J = \frac{1}{2} \int_0^T \int_{k \leqslant k_m} (\hat{\boldsymbol{u}}(\boldsymbol{k}, t) - \hat{\boldsymbol{v}}(\boldsymbol{k}, t)) \cdot (\hat{\boldsymbol{u}}^{*}(\boldsymbol{k}, t) - \hat{\boldsymbol{v}}^{*}(\boldsymbol{k}, t)) \, \mathrm{d}\boldsymbol{k}, \tag{2.9}$$

where $k = |\boldsymbol{k}|$, and $\hat{\boldsymbol{v}}(\boldsymbol{k}, t)$ and $\hat{\boldsymbol{u}}(\boldsymbol{k}, t)$ denote the Fourier modes of $\boldsymbol{v}(\boldsymbol{x}, t)$ and $\boldsymbol{u}(\boldsymbol{x}, t)$, respectively. Letting $\hat{\boldsymbol{F}}(\boldsymbol{k}, t)$ be the Fourier transform of $\boldsymbol{F}(\boldsymbol{x}, t)$, we obtain

$$\hat{\boldsymbol{F}}(\boldsymbol{k}, t) = \begin{cases} -[\hat{\boldsymbol{u}}(\boldsymbol{k}, t) - \hat{\boldsymbol{v}}(\boldsymbol{k}, t)] & \text{for } k \leqslant k_m, \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

2.2. *The questions to be answered*

We call $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{v}(\boldsymbol{x}, t)$ the 'reconstructed' and the 'target' fields, respectively. The target field at $t = 0$, $\boldsymbol{v}(\boldsymbol{x}, 0)$, will also be denoted by notation $\boldsymbol{\varphi}^{v}(\boldsymbol{x})$, corresponding to its reconstruction $\boldsymbol{u}(\boldsymbol{x}, 0) \equiv \boldsymbol{\varphi}^{u}(\boldsymbol{x})$. Since $\boldsymbol{u}(\boldsymbol{x}, t)$ is the solution of the NSEs with $\boldsymbol{\varphi}^{u}(\boldsymbol{x})$ as the IC, it is expected $\hat{\boldsymbol{u}}(\boldsymbol{k}, t) \approx \hat{\boldsymbol{v}}(\boldsymbol{k}, t)$ when $k \leqslant k_m$. The interesting question is how closely $\hat{\boldsymbol{u}}(\boldsymbol{k}, t)$ matches $\hat{\boldsymbol{v}}(\boldsymbol{k}, t)$ when $k > k_m$; in other words, to what extent small scales in $\boldsymbol{v}(\boldsymbol{x}, t)$ can be reconstructed.

As $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{v}(\boldsymbol{x}, t)$ are both solutions of the forced NSEs, one expects their statistics to be the same for $t \to T$ when $\boldsymbol{u}$ has fully developed. Therefore, for $t \to T$, we are interested in the pointwise, instantaneous difference or correlation between $\boldsymbol{u}$ and $\boldsymbol{v}$. However, good reconstruction at $t \to T$ is possible only when the IC $\boldsymbol{\varphi}^{u}(\boldsymbol{x})$ captures, to some extent, the features of real turbulence. Therefore, we will also examine selected statistics of $\boldsymbol{\varphi}^{u}(\boldsymbol{x})$. The statistics provide information on how much can be reconstructed at $t = 0$ given the measurement data $\mathcal{F}\boldsymbol{v}(\boldsymbol{x}, t)$, hence are also relevant to the observability of the system. Nevertheless, the full discussion of this aspect is left for the future.

The efficacy of 4DVAR can be assessed by comparison with other methods. As we will show later, the straightforward Lagrangian interpolation fails badly. In Yoshida *et al.* (2005), an assimilation scheme we call 'direct substitution' is used. The authors integrated the NSEs starting from a Gaussian random field. The low wavenumber modes with $k \leqslant k_m$ in the solution $\boldsymbol{u}(\boldsymbol{x}, t)$ are replaced by those in the target field $\boldsymbol{v}(\boldsymbol{x}, t)$ at every time step. Yoshida *et al.* (2005) show that, if $k_m \geqslant 0.2\eta_K^{-1}$, the small scales in $\boldsymbol{u}(\boldsymbol{x}, t)$ approach those in $\boldsymbol{v}(\boldsymbol{x}, t)$ asymptotically as $t \to \infty$ (see also Lalescu *et al.* (2013)). In most cases we are investigating, the resolution k_m is slightly below this threshold (cf. the parameters in § 3.1), and the measurement data are available only for $t \in [0, T]$ where T is of the order of one large eddy turnover time scale. As a consequence, the DS scheme will only achieve partial reconstruction. The DS scheme will be compared with the 4DVAR method, to demonstrate the improvement provided by the latter.

2.3. *Description of the small scales of turbulent velocity fields*

To compare the small scales of the reconstructed field $\boldsymbol{u}$ and the target field $\boldsymbol{v}$, the filtering approach is used to separate different scales and the analysis is conducted mainly in the physical space. The filtered velocity field is defined by

$$\tilde{\boldsymbol{u}}(\boldsymbol{x}, t) = \int G_\Delta(\boldsymbol{x} - \boldsymbol{y})\boldsymbol{u}(\boldsymbol{y}, t)\, \mathrm{d}\boldsymbol{y}, \tag{2.11}$$

where $\tilde{\boldsymbol{u}}$ denotes the filtered velocity and $G_\Delta(\boldsymbol{x})$ is a filter with length scale Δ; G_Δ separates $\boldsymbol{u}$ into $\tilde{\boldsymbol{u}}$ and the subgrid-scale (SGS) velocity. The parameters used to describe the structures of $\tilde{\boldsymbol{u}}$ and their interactions with the SGS scales are now summarized. In the analysis of the velocity field $\boldsymbol{u}$, we always use the Gaussian filter (Pope 2000).

The local structures of $\tilde{\boldsymbol{u}}$ are described by the filtered velocity gradient $\tilde{A}_{ij} \equiv \partial_j \tilde{u}_i$, the filtered strain rate tensor $\tilde{s}_{ij} \equiv (\tilde{A}_{ij} + \tilde{A}_{ji})/2$ and the filtered vorticity $\tilde{\omega}_i \equiv \varepsilon_{ijk}\tilde{A}_{kj}$. For $\tilde{A}_{ij}$, much insight has been gained from its tensor invariants

$$Q \equiv -\tfrac{1}{2}\tilde{A}_{ij}\tilde{A}_{ji}, \quad R \equiv -\tfrac{1}{3}\tilde{A}_{ij}\tilde{A}_{jk}\tilde{A}_{ki}. \tag{2.12a,b}$$

One of the key features of turbulence is that the joint PDF of R and Q displays a skewed teardrop shape, which captures the prevalence of straining motion in turbulent flows; see, e.g. Cantwell (1992).

The eigenvectors of $\tilde{s}_{ij}$ are denoted by $\boldsymbol{e}_\alpha^s$, $\boldsymbol{e}_\beta^s$ and $\boldsymbol{e}_\gamma^s$, with corresponding eigenvalues $\lambda_\alpha^s \geqslant \lambda_\beta^s \geqslant \lambda_\gamma^s$; $\lambda_\alpha^s + \lambda_\beta^s + \lambda_\gamma^s = 0$ due to the incompressibility of the filtered velocity field. The enstrophy of the resolved velocity field, defined by $\tilde{\omega}^2 \equiv \tilde{\omega}_i\tilde{\omega}_i$, is a measure of the magnitude of the resolved vorticity. The main source of growth for $\tilde{\omega}^2$ is the vortex stretching term $P_\omega \equiv \tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j$. Evidently,

$$P_\omega = \tilde{\omega}^2(\lambda_\alpha^s \cos^2 \theta_\alpha^s + \lambda_\beta^s \cos^2 \theta_\beta^s + \lambda_\gamma^s \cos^2 \theta_\gamma^s), \tag{2.13}$$

where θ_α^s is the angle between $\tilde{\boldsymbol{\omega}}$ and $\boldsymbol{e}_\alpha^s$ with θ_β^s and θ_γ^s defined in a similar way. Equation (2.13) shows that P_ω are determined by the eigenvalues of $\tilde{s}_{ij}$ as well as the alignment between $\tilde{\omega}_i$ and the eigenvectors of $\tilde{s}_{ij}$. The eigenvectors of $\tilde{s}_{ij}$ form an orthogonal coordinate frame. The orientation of $\tilde{\omega}_i$ can be described by the polar and azimuthal angles relative to $\boldsymbol{e}_\alpha^s$ in this frame, and they are denoted by θ_{sp}^ω and ϕ_{sp}^ω.
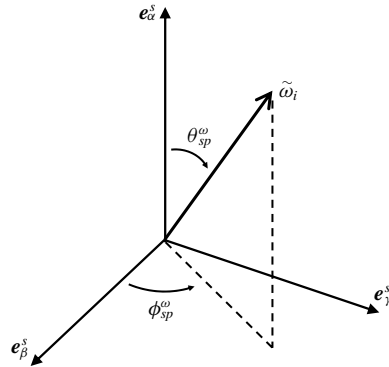
FIGURE 1. Definitions of the polar and azimuthal angles used to describe the orientation of $\tilde{\omega}_i$ in the eigenframe of $\tilde{s}_{ij}$. Angles defined in a similar way for other vectors and eigenframes are also used in the paper.

Figure 1 gives an illustration of the angles. In turbulence, by examining the joint PDF of $\cos \theta_{sp}^{\omega}$ and ϕ_{sp}^{ω} or the PDFs of θ_i^s $(i = \alpha, \beta, \gamma)$, it has been shown that $\tilde{\omega}_i$ prefers to align with e_{β}^s.

The dynamics of $\tilde{u}_i$ is governed by the filtered NSEs, which is unclosed due to the SGS stress tensor $\tau_{ij} = \widetilde{u_i u_j} - \tilde{u}_i \tilde{u}_j$ (see, e.g. Meneveau & Katz (2000)). The tensor τ_{ij} represents the nonlinear interactions between the resolved and the SGS scales. An important parameter related to τ_{ij} is the SGS energy dissipation rate

$$\Pi_\Delta = -\tau_{ij}^d \tilde{s}_{ij}, \tag{2.14}$$

where $\tau_{ij}^d \equiv \tau_{ij} - \delta_{ij}\tau_{kk}/3$ is the deviatoric part of τ_{ij}; Π_Δ is the turbulent kinetic energy flux cascading from resolved scales to the SGS scales. Correctly reproducing the statistics of Π_Δ is one of the main objectives in SGS modelling. As a consequence, the statistics of τ_{ij}^d and Π_Δ and their correlations with $\tilde{s}_{ij}$, $\tilde{\omega}_i$, P_ω and $\tilde{A}_{ij}$ have been extensively documented; see, e.g. Meneveau & Katz (2000) and Sagaut (2002) for reviews on SGS modelling and large eddy simulations.

The value of Π_Δ is determined by the relative alignment between the eigenvectors of τ_{ij}^d and $\tilde{s}_{ij}$ as well as their eigenvalues. Specific preferential alignment has also been observed in turbulent flows (Tao, Katz & Meneveau 2002). The eigenvectors of $-\tau_{ij}^d$ are denoted by e_{α}^{τ}, e_{β}^{τ} and e_{γ}^{τ}, with corresponding eigenvalues $\lambda_{\alpha}^{\tau} \geqslant \lambda_{\beta}^{\tau} \geqslant \lambda_{\gamma}^{\tau}$ where $\lambda_{\alpha}^{\tau} + \lambda_{\beta}^{\tau} + \lambda_{\gamma}^{\tau} = 0$ by definition. The vorticity $\tilde{\omega}_i$ also displays preferential alignment with $-\tau_{ij}^d$ (Horiuti 2003); specifically $\tilde{\omega}_i$ tends to be perpendicular to e_{γ}^{τ} while aligned with e_{α}^{τ} or e_{β}^{τ}.

2.4. *Description of the non-local structures*

The filtered velocity gradient $\tilde{A}_{ij}$ and related quantities describe the local structure of the filtered velocity field in an infinitesimal neighbourhood of a spatial location. Non-local structures of finite sizes are also of great interest. The most well-known example of such structures is that of the vortex filaments that are characterized by strong enstrophy. The ability of the reconstructed velocity field to capture the locations, the dimensions and the orientations of these structures is not always clear
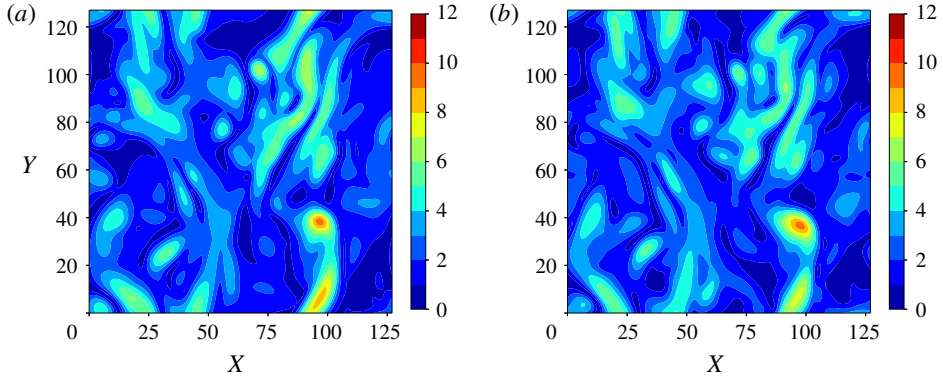
FIGURE 2. Instantaneous distributions of $\tilde{\omega} \equiv (\tilde{\omega}_i \tilde{\omega}_i)^{1/2}$ on a plane perpendicular to the z axis. (*a*) From a target field; (*b*) from the corresponding reconstructed field. X and Y are the labels of the grid points in the x and y directions, respectively.

from the pointwise statistics. To illustrate this point, figure 2 plots selected contours of $\tilde{\omega} \equiv (\tilde{\omega}_i \tilde{\omega}_i)^{1/2}$ in one target field and the corresponding reconstructed field. Detailed comparison is given later. However, it is already clear that the locations, orientations and sizes of the contours display remarkable similarity, especially in high $\tilde{\omega}$ regions. Usual statistics, including joint correlations between multiple points, do not always capture this similarity. Additional diagnostics are needed.

Note that, although an intuitive description of the structures visualized by the contours is straightforward, the precise description of their shapes, sizes and even locations is difficult, and actually there are no unique definitions for these concepts (also see discussions in, e.g. Bermejo-Moreno & Pullin (2008), Yang & Pullin (2011) and Leung *et al.* (2012)). As explained in § 1, we use the MVEEs of a structure to define and describe its geometry. By definition, of all the ellipsoids that contain a structure, the MVEE is the one with the smallest volume. The location, dimensions and orientation of an MVEE are well defined, and are given by its centre, lengths of the three axes and the orientations of the axes. We thus define the location, dimensions and orientation of a structure using those of its MVEE. The structures in the reconstructed and the target fields are compared on this basis. The comparison includes four steps of calculations, which are explained below and illustrated in figure 3 using a 2-D slice.

In the first step, the set $\mathcal{P}$ of the grid points where a certain threshold condition is satisfied are extracted; $\mathcal{P}$ usually consists of a number of disjoint regions. The 'structures' we will be discussing refer to this type of region; each structure corresponds to a disjoint region (cf. figure 3*b*).

In the second step, individual structures in $\mathcal{P}$ are extracted and distinguished from others. The programmatic realization of this task is non-trivial despite its innocuous appearance. It is accomplished by recasting it as a clustering problem, and using a density based clustering algorithm called DBSCAN (Ester *et al.* 1996; Schubert *et al.* 2017). In this algorithm, the 'neighbours' of a point $\boldsymbol{p}$ in $\mathcal{P}$ are first identified, where a neighbour is defined as a point whose distance to $\boldsymbol{p}$ is less than a given value ϵ_c. A point with fewer than n_c neighbours is considered 'noise'. A cluster is defined according to the following rule: the neighbours of a non-noise point $\boldsymbol{p}$ are in the same cluster to which point $\boldsymbol{p}$ belongs. Algorithmically, the points in the cluster are
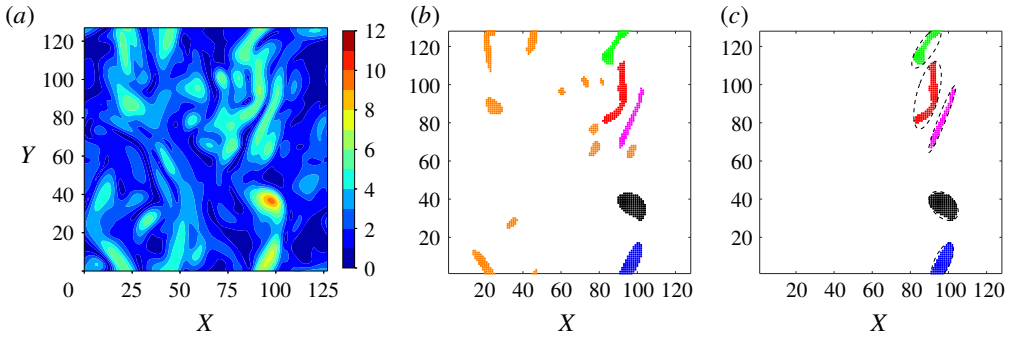
FIGURE 3. The calculation of the MVEEs. (*a*) The contour plot of $\tilde{\omega}$ on a 2-D slice (same as figure 2*b*). (*b*) The 17 structures with $\tilde{\omega} \geqslant 2\langle\tilde{\omega}_i\tilde{\omega}_i\rangle^{1/2}$ shown with the grid points, which are identified by the DBSCAN algorithm (see text for details). The 5 largest structures are highlighted with black, red, blue, green and magenta, respectively. (*c*) The MVEEs for the 5 largest structures. *X* and *Y* are the labels of the grid points in the *x* and *y* directions, respectively.

identified by recursively applying this rule until all the points identified are noise. The same process is repeated for other non-noise points if they do not belong to any clusters that have been found.

For more details of the DBSCAN algorithm, the readers are referred to Ester *et al.* (1996). The outcome of applying the algorithm is that all the points in a structure are found and stored separately from those in other structures, as illustrated in figure 3(*b*). We normalize the distance in such a way that the distance between two consecutive grid points is 1. As a consequence, $\epsilon_c = D^{1/2}$ and $n_c = D + 1$ are used, where D is the dimension of the embedding space (i.e. $D = 2$ for structures on a 2-D slice whereas $D = 3$ for those in the 3-D space).

In the third step, the MVEE for each structure is calculated. An ellipsoid is defined by a symmetric positive definite matrix $\boldsymbol{E}$, which specifies the orientations and lengths of the axes, and a vector $\boldsymbol{c}$, which specifies its centre. Its volume is proportional to the determinant of $\boldsymbol{E}^{1/2}$, det $\boldsymbol{E}^{1/2}$. Let P be the set of points $\boldsymbol{p}_i$ ($i = 1, \dots, N$) in the structure. The MVEE is given by the optimal $\boldsymbol{E}$ and $\boldsymbol{c}$ that minimize det $\boldsymbol{E}^{1/2}$, subject to the constraints

$$(\boldsymbol{p}_i^T - \boldsymbol{c}^T)\boldsymbol{E}^{-1}(\boldsymbol{p}_i - \boldsymbol{c}) \leqslant 1, \quad \forall \boldsymbol{p}_i \in P. \tag{2.15}$$

The constraints ensure $\boldsymbol{p}_i$ is inside the ellipsoid. This minimization problem can be solved by the Khachiyan algorithm (Khachiyan 1996; Todds 2016). We use the MATLAB implementation by Moshtagh (2009).

In the fourth (and last) step, the correspondence between a structure in the target field and its reconstruction in the reconstructed field is established. The above three steps are applied to both the reconstructed and the target fields. If the reconstructed field mimics the target field perfectly, one would then obtain two groups of MVEEs, with a one-to-one correspondence between the members. The correspondence can be trivially established. In reality, however, this naive method breaks down because the difference between the two fields sometimes is big enough to destroy the trivial one-to-one correspondence. The following procedure thus has been used, where, in essence, the structures are identified by their locations and sizes. The structures in the target field are first arranged into a list in the descending order according to their sizes,
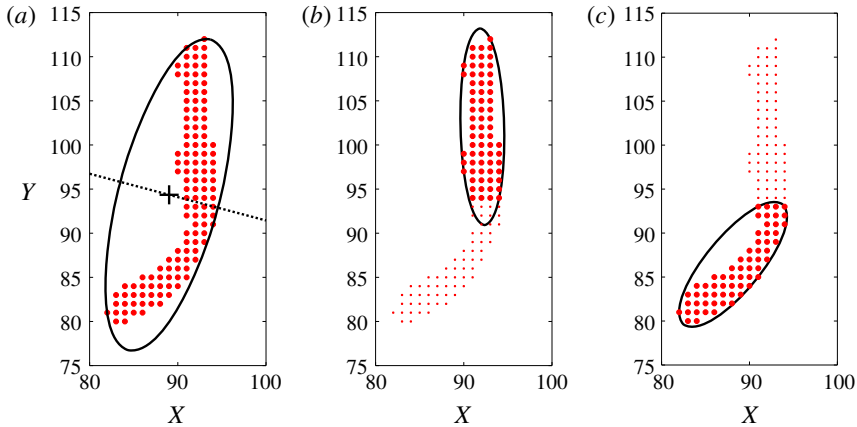
FIGURE 4. Calculation of sub-MVEEs and the MVEE tree. (*a*) One of the vortical structures shown in figure 3, which is enclosed by the level 0 sub-MVEE and is split into two groups along the dotted line. The red dots are the grid points in the structure. (*b*) The level 1 sub-MVEE for the upper half. (*c*) The level 1 sub-MVEE for the lower half. *X* and *Y* are the labels of the grid points in the *x* and *y* directions, respectively.

where the size of a structure is defined as the number of grid points in the structure. Let $\mathcal{S}_T$ be a structure in the list. The structure in the reconstructed field whose distance to $\mathcal{S}_T$ is the minimum is then identified, where the distance between two structures is defined as the shortest distance from one to the other. Let $\mathcal{S}_O$ be this structure. If $\mathcal{S}_O$ is not unique, the one with the largest size is chosen. This $\mathcal{S}_O$ is taken as the reconstruction of $\mathcal{S}_T$; $\mathcal{S}_T$ and $\mathcal{S}_O$ are called the matching structures. The above procedure is repeated for each structure in the target field, starting from the one largest in size, until all matching structures are identified.

The matching structures and corresponding MVEEs obtained after these four steps are then compared and statistics are calculated, including relative displacement, alignment and relative sizes, among others. The analysis is applied to several different physical quantities, including high strain rate structures for which the magnitude of $\tilde{s}_{ij}$ is large, vortical structures with strong $\tilde{\omega}$ and structures with high SGS energy dissipation rate Π_Δ or negative Π_Δ. In all cases, the analysis is performed directly on 3-D structures.

MVEEs provide a good description of the enclosed structure when the latter is convex. A more complete description is desirable, however, if the structure is non-convex, such as the second structure from the top in figure 3(*c*). In this case, we propose to use what we call the MVEE trees. The construction of MVEE trees is illustrated with figure 4. It is supposed that the MVEE for the structure has been found. In the context of MVEE trees, this MVEE is called the level 0 (L0) MVEE. The L0 MVEE allows one to split the grid points in the structure into two groups, using the symmetric plane of the MVEE perpendicular to its major axis (see of figure 4*a*). The MVEEs for the two parts can then be found. These are called the L1 sub-MVEEs. The splitting can be repeated for each structure in this level, and those in each new level. The procedure thus forms a tree-like hierarchy of MVEEs, hence the name MVEE trees.

The MVEE trees are used to analyse strongly non-convex structures. For this purpose, the degree of non-convexity of a structure is defined as the ratio of the

volume of the structure to the volume of its convex hull (Boyd & Vandenberghe 2004). The ratio is denoted by R_V. Smaller R_V indicates stronger non-convexity. Only structures with small R_V are selected when MVEE trees are calculated.

Moment of inertia ellipsoids (MoIEs) have also been used in the past to characterize non-local structures. However, MoIEs do not capture the extremal points in a structure, and hence tend to miss the details of highly non-convex structures. In comparison, MVEE and MVEE trees are better in this regard.

3. Simulations and results

3.1. *Parameters and the computation of the target flow fields*

The target fields $v(x, t)$ are obtained by numerically integrating the NSEs with a fully dealiased pseudo-spectral code in a $[0, 2\pi]^3$ box with periodic boundary conditions. A second-order explicit improved Euler method based on the trapezoidal rule is used in time stepping; $v(x, t)$ is made statistically stationary by a constant forcing term f, where

$$f = (0, f_a \cos(k_f x_1), 0), \tag{3.1}$$

with $f_a = 0.15$ and $k_f = 1$ (these and the following parameters are all given in code units). Therefore, the flow is a 3-D Kolmogorov flow with a sinusoidal mean velocity profile (Borue & Orszag 1996; Kang & Meneveau 2005). The Kolmogorov flow has been chosen because the forcing term for this flow is particularly simple.

In all cases 128^3 grid points have been used. The viscosity is $\nu = 0.006$. Time step is $\delta t = 0.00575$. The time and length scales of v are estimated from f_a, k_f, ν and the root-mean-square (r.m.s.) velocity v_{rms} which is found numerically to be approximately 0.65. The large eddy turnover time scale is thus $\tau_L \equiv v_{rms}/f_a \approx 4.35$. The energy dissipation rate ϵ is calculated numerically from the energy spectrum and ν, which gives $\epsilon \approx 0.08$. The Kolmogorov time scale is $\tau_K \equiv (\nu/\epsilon)^{1/2} \approx 0.28$. The Kolmogorov length scale is $\eta_K \equiv (\nu^3/\epsilon)^{1/4} \approx 0.04$. Therefore, $\delta x/\eta_K \approx 1.25$, where δx is the grid size of the simulations. The Reynolds number based on the Taylor micro-scale is $Re_\lambda \approx 75$, whereas the Reynolds number based on the forcing length scale is $Re_f \approx 340$.

3.2. *The solution of the optimality system*

The optimality system ((2.3) and (2.5)–(2.8)) is solved iteratively. Given an initial guess for φ, the NSEs (2.3) are numerically integrated from $t = 0$ to $t = T$ to find $u(x, t)$. The cost J is calculated from u. Unless J is already smaller than a given tolerance, φ has to be updated to reduce J. To do so, equation (2.5) is integrated backward in time from $t = T$ to $t = 0$, starting from the endpoint condition given by (2.6). The time sequence of the forward solution $u(x, t)$ is used in the backward integration. The solution for ξ provides $DJ/D\varphi$ according to (2.8), which is then used to update φ. The updating step uses the nonlinear conjugate gradient method using the Polak–Ribière formula (Nocedal & Wright 1999). Brent's method in Numerical Recipe (Press *et al.* 1992) is used in the line search step.

The above several steps constitute the main loop of the solution algorithm. Due to the chaotic nature of the solution of the NSEs, our experience is that the iterations may become difficult to converge when T is large. We thus solve the optimality system first with a small T, using the iterations outlined above. The optimal φ for this problem is used as the initial guess for the problem with a larger T. The procedure is repeated until the intended optimization horizon is reached. As the

computational time is usually very long on the computer we used (see below for more information), we save the optimal solutions for the intermediate T values on the hard disk, which safeguards against unexpected interruptions of the computation. Therefore, this implementation is also useful from a practical point of view. The choices for the initial small T and the increment in T are mostly based on trial and error. An initial $T = 0.25\tau_L \approx 4\tau_K$ with an increment equal $0.025\tau_L \approx 0.4\tau_K$ has been used in most cases, but we expect other similar choices will also work.

We now explain briefly the solution of the individual equations in the main loop; $u(x, t)$ is solved from the NSEs using the same numerical methods and parameters as those for $v(x, t)$. The initial guess for $u(x, 0) \equiv \varphi$ is usually a divergence free Gaussian random velocity field, with the Fourier modes with $k \leqslant k_m$ replaced by those in the target data $\varphi^v(x) \equiv v(x, 0)$ (cf. § 2.1). Our tests show that the initial guess has no effects on the statistics of optimal solutions. It is possible to use an initial guess with an energy spectrum matching the initial target field. However, our tests showed that it did not improve the results nor speed up convergence. A possible reason is that matching the energy spectra imposes only a weak constraint on the control φ since the latter has a very large number (3×128^3) of components.

As will be explained below, we run several related cases with different parameters. Sometimes the optimal solution φ^u for a solved case is then used as the initial guess for φ in a related case. For example, the optimal solution obtained with a larger tolerance has been used as the initial guess for a case with a smaller tolerance; the one with a smaller T used as the initial guess for a case with a larger T.

The adjoint equations (2.5) are solved with the same numerical schemes used for v and u. The divergence free condition for ξ is used to eliminate the adjoint pressure σ, in the same way as p is eliminated from the NSEs when they are solved numerically. However, equation (2.5) has to be integrated backwards in time, as it is supplemented with the endpoint condition in (2.6). The time sequence of $u(x, t)$ ($t \in [0, T]$) is needed for the solution of ξ. Therefore, $u(x, t)$ has to be saved in the forward integration of the NSEs. In our computation, typically the available RAM is not large enough to hold the whole time sequence. We thus use checkpointing, where $u(x, t)$ is saved on the hard disk at selected t values (the checkpoints) in the forward integration. The sequence of $u(x, t)$ between two checkpoints, which will be short enough to store in the RAM, are computed (using the data on a checkpoint as the initial condition) when the backward integration of ξ reaches the time interval between the two checkpoints. Checkpointing thus incurs several passes of forward integrations in one single backward integration. But it avoids having to read data at every time step from the hard disk, which typically is a time consuming operation. The technique can be optimized by choosing the checkpoints judiciously, although we do not attempt to do so in this investigation. More details about checkpointing can be found in, e.g. Wang, Moin & Iaccarino (2009).

3.3. *The choices of computational parameters*

We assume that the data $\mathcal{F}v$ are available continuously, so that the forcing term in (2.7) is applied at each time step. We assume the optimal solution is found when $J/J(u = 0)$ is less than a tolerance e_{tot}, where, according to (2.2), $J(u = 0)$ is the total kinetic energy of the measurement data $\mathcal{F}v(x, t)$ ($0 \leqslant t \leqslant T$). Due to our limited computational resources, this investigation has used only 128^3 grid points. However, several values of T, k_m and e_{tot} are examined, and up to five realizations are run for each set of parameters. The target data $v(x, t)$ in different realizations are different segments of the same time series of DNS data.
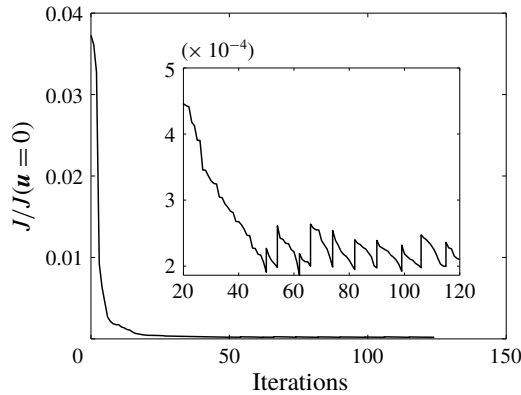
FIGURE 5. The value of $J/J(\boldsymbol{u}=0)$ as a function of iterations in one of the runs. Inset: a zoom-in between the 20th and the 120th iterations. The jumps occur at the points where the iterations have converged and T is extended.

Case	T/τ_L	k_m	$10^4 e_{tot}$	N_R	Case	T/τ_L	k_m	$10^4 e_{tot}$	N_R
C1	**1**	**4**	**2**	**5**	C7	0.5	4	1	2
C2	1	4	1.5	2	C8	0.5	4	0.5	2
C3	0.5	2	2	2	C9	0.5	8	2	2
C4	0.5	2	1	2	C10	0.5	8	1	2
C5	0.5	2	0.5	2	C11	0.5	8	0.5	2
C6	0.5	4	2	5	—	—	—	—	—

TABLE 1. The parameters for the cases being investigated. N_R is the number of realizations. Case C1 is the baseline case with the parameters highlighted in bold.

The parameters are summarized in table 1. We consider T up to one large eddy turnover time scale τ_L and resolutions of $k_m = 2$, 4 and 8. However, for $T = \tau_L$, only cases with $k_m = 4$ are reported for the following reasons. Firstly, for $(T, k_m) = (\tau_L, 2)$, we failed to find convergent results (see the discussion of the results below for further comments). Secondly, for $(T, k_m) = (\tau_L, 8)$, it turns out that the optimal solutions are the same as those with $(T, k_m) = (0.5\tau_L, 8)$ with $e_{tot} = 0.5 \times 10^{-4}$ (i.e. the cases with a shorter T but more stringent tolerances). Therefore, the results for these two k_m values have been omitted.

For $(T, k_m) = (\tau_L, 4)$, optimal solutions are found for $e_{tot} = 2 \times 10^{-4}$ in five realizations (case C1). In each of these realizations, it took around 15 days to find the optimal solution, using up to 48 G RAM and 4 cores on an Intel Ivybridge or Westmere based CPU, with the optimal solutions from case C6 as the initial guess. In two of the realizations, the optimal solutions are further improved by additional iterations to reduce $J/J(\boldsymbol{u}=0)$ below 1.5×10^{-4} (case C2). The improvement required approximately 15 more days of computer time. Given the significant computational cost, we did not attempt to reduce the cost function further. On the other hand, for $T = 0.5\tau_L$, optimal solutions were obtained for $e_{tot} = 10^{-4}$ and 0.5×10^{-4} in two realizations for all three k_m values. The cost decreases rapidly initially, but further improvement becomes very slow. Figure 5 plots the change of $J/J(\boldsymbol{u} = 0)$ in the iterations for a run in case C6, which depicts this behaviour. The inset zooms into the curve to show the change at later iterations. Abrupt jumps in J are observed
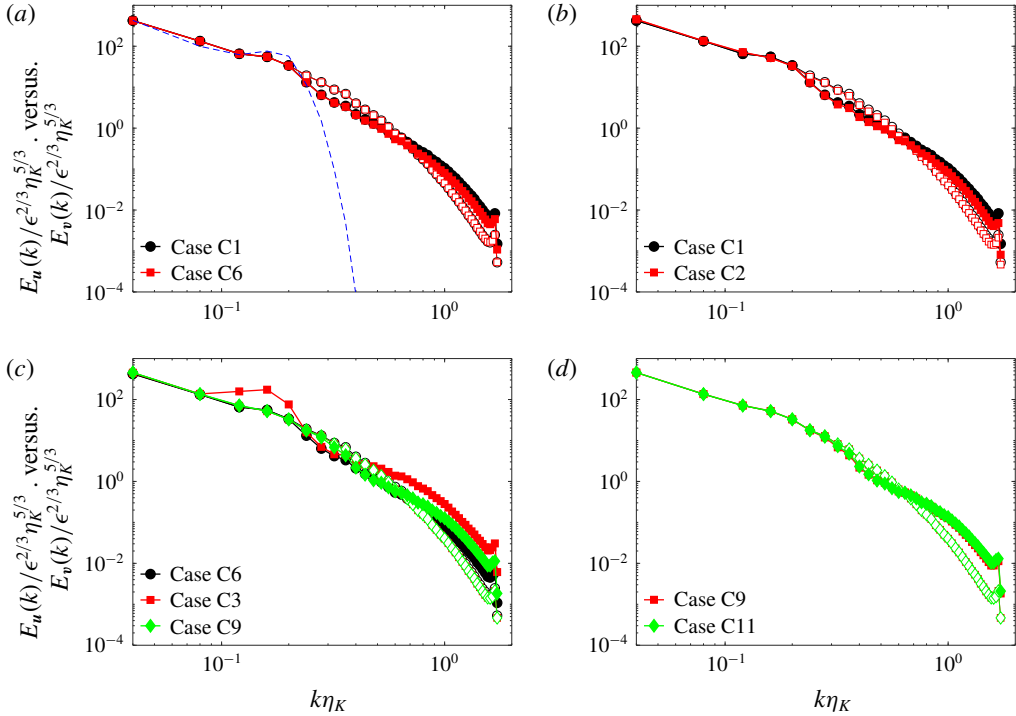
FIGURE 6. The energy spectra for $\varphi^u(x)$ (solid symbols) and $\varphi^v(x)$ (empty symbols of same shapes). (*a*) With different optimization horizons T (the dashed line shows the energy spectrum of the initial guess for $\varphi^u(x)$ in one of the realizations); (*b*) with different tolerances e_{tot}; (*c*) with different measurement resolutions k_m; (*d*) with different tolerances for $k_m = 8$.

when T is extended. In this run, the optimal solution for $T = 0.5\tau_L$ is found after 124 iterations. In each iteration, around 5 forward integrations are performed in order to find the conjugate gradient direction. Taking backward integration into account, 700–800 passes of integration are used, although we note that part of these integrations are over shorter optimization horizons.

We consider case C1 (shown in bold in table 1) the baseline case. The comparison between this and other cases will illustrate the effects of the resolution of the measurement data k_m, the optimization horizon T and the tolerance e_{tot}. However, the majority of results are presented only for the baseline case. The statistics are averaged over all the realizations in each case.

3.4. *The energy spectra and spectral correlation for different computational parameters*

The spectral distributions of u and v obtained with different computational parameters are compared in this subsection. The purpose is to give an overview of the quality of the reconstruction. More detailed investigation, focusing on the baseline case, will be presented later. Symbols u and v in the subscripts or superscripts are used to denote results from u (the reconstruction) and v (the target), respectively.

Figure 6 shows the energy spectra for φ^u and φ^v, which are plotted with solid and empty symbols, respectively. The figure shows that a perfect match between φ^u and

$\boldsymbol{\varphi}^v$ is obtained for $k \leqslant k_m$, as expected. Figures 6(a) and 6(b) show that, for $k_m = 4$, $\boldsymbol{\varphi}^u$ under-predicts the spectrum for intermediate k and over-predicts it for large k. Nevertheless, the spectra of $\boldsymbol{\varphi}^u$ broadly follow those of $\boldsymbol{\varphi}^v$. The agreement does not change much with T and e_{tot} for the ranges of the parameters we are considering. Note that it is non-trivial to observe the fairly good agreement. The dashed line in figure 6(a) plots the energy spectrum for the initial guess for $\boldsymbol{\varphi}^u(\boldsymbol{x})$, which shows that the small scales in $\boldsymbol{\varphi}^u$ are not present in the initial guess and are generated by the data assimilation process.

The resolution of the measurement data k_m has significant effects on $E_u(k)$, as is demonstrated in figure 6(c); $E_u(k)$ for $k_m = 2$ essentially fails to capture the features of $E_v(k)$. On the other hand, the spectrum for $k_m = 8$ shows a similar level of discrepancy as for $k_m = 4$ at the high wavenumber end, which is somewhat surprising. Reducing the tolerance to $e_{tot} = 0.5 \times 10^{-4}$ does not improve the agreement, as shown in figure 6(d).

Since both $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{v}(\boldsymbol{x}, t)$ are solutions of the forced NSEs, their statistics are expected to be the same for $t \to T$. It is interesting, however, to investigate the instantaneous pointwise difference between $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{v}(\boldsymbol{x}, t)$ at a later time t for $t \in [0, T]$. Two quantities are used to quantify the difference. The first one is the energy spectrum for the velocity difference $\delta\boldsymbol{u}(\boldsymbol{x}, t) = \boldsymbol{u}(\boldsymbol{x}, t) - \boldsymbol{v}(\boldsymbol{x}, t)$, denoted by $E_D(k)$, and called the spectral difference. Obviously, a smaller $E_D(k)$ indicates better match between $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{v}(\boldsymbol{x}, t)$, hence a better reconstruction. The second one is the spectral correlation (i.e. normalized co-spectrum) between $\boldsymbol{u}$ and $\boldsymbol{v}$, defined by

$$\hat{\rho}_{\boldsymbol{uv}}(k) \equiv \frac{\oint \langle \hat{\boldsymbol{u}}^*(\boldsymbol{k}, t) \cdot \hat{\boldsymbol{v}}(\boldsymbol{k}, t)\rangle \, \mathrm{d}S_k}{2\sqrt{E_u(k)E_v(k)}}. \tag{3.2}$$

The pointed brackets in the numerator denote ensemble average whereas the integral is a surface integral over the sphere $|\boldsymbol{k}| = k$. By definition, $0 \leqslant \hat{\rho}_{\boldsymbol{uv}} \leqslant 1$, and a larger $\hat{\rho}_{\boldsymbol{uv}}$ indicates better pointwise agreement between $\boldsymbol{u}$ and $\boldsymbol{v}$. Note that $E_D(k)$ and $\hat{\rho}_{\boldsymbol{uv}}(k)$ both depend on time t, although the dependence is not given explicitly in the expressions.

Figure 7 compares $E_D(k)$ and $\hat{\rho}_{\boldsymbol{uv}}(k)$ from different cases in various ways. Figures 7(a) and 7(b) show that, in cases C1 and C2, the pointwise agreement between $\boldsymbol{u}$ and $\boldsymbol{v}$ improves over time within the optimization horizon; $\hat{\rho}_{\boldsymbol{uv}}$ reaches 0.8 or higher at $t = \tau_L$ for all wavenumbers, whereas $E_D(k)/E_v(k)$ can be reduced down to 30 % even for $k = 40$. Note that the measurement data contain approximately $8^3/2 = 256$ Fourier modes, constituting less than 0.1 % of the Fourier modes in the whole flow field.

Figures 7(a) and 7(b) also show that imposing a smaller tolerance e_{tot} can lead to significantly better agreement at later times, even if it is only marginally smaller (i.e. 1.5×10^{-4} versus 2×10^{-4}). On the other hand, the effect is quite limited at earlier times (e.g. at $t = 0$), where significant discrepancy between $\boldsymbol{u}$ and $\boldsymbol{v}$ is observed for both e_{tot}. This observation is interesting, suggesting that the improvements at later times are due to effects that are not completely revealed by $E_D(k)$ and $\hat{\rho}_{\boldsymbol{uv}}$ at earlier times.

The effects of the optimization horizon T are investigated in figure 7(c), where $\hat{\rho}_{\boldsymbol{uv}}$ for $T = 0.5\tau_L$ and $T = \tau_L$ is shown with empty and filled symbols, respectively. The results for $T = 0.5\tau_L$ are shown for $t = 0.25\tau_L$ (squares) and $0.5\tau_L$ (triangles). Those for $T = \tau_L$ are shown for the same times and $t = \tau_L$ too. It is observed that larger T improves the pointwise agreement, and the improvement increases over time. The agreement, again, can be improved by imposing a smaller tolerance. Figure 7(d) compares $\hat{\rho}_{\boldsymbol{uv}}$ at $t = T$ for three e_{tot} values. For $e_{tot} = 0.5 \times 10^{-4}$, the correlation
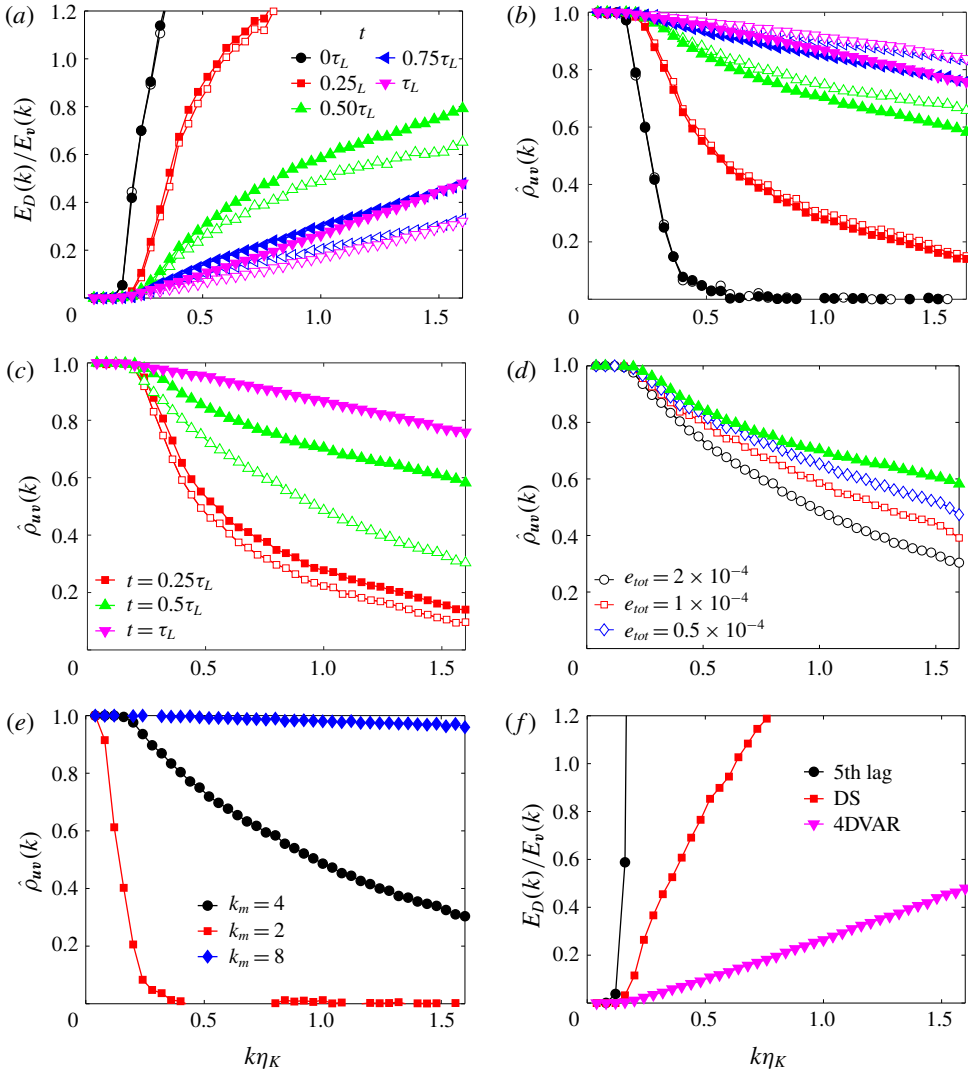
FIGURE 7. (a,b) Values of $E_D(k)/E_v(k)$ and $\hat{\rho}_{uv}(k)$ at different times t for case C1 (solid symbols) and case C2 (empty symbols), with $(T, k_m) = (\tau_L, 4)$ and $e_{tot} = 2 \times 10^{-4}$ or 1.5×10^{-4}. (c) Value of $\hat{\rho}_{uv}(k)$ for case C1 where $T = \tau_L$ (solid symbols) and case C6 where $T = 0.5\tau_L$ (empty symbols). (d) Value of $\hat{\rho}_{uv}(k)$ for different e_{tot} (cases C6, C7, C8; empty symbols) at $t = 0.5\tau_L$. Solid symbols: the result for case C1 as a comparison. (e) Value of $\hat{\rho}_{uv}(k)$ at $t = 0.5\tau_L$ for different k_m (cases C3, C6, C9). (f) Value of $E_D(k)/E_v(k)$ at $t = \tau_L$ obtained from 4DVAR (case C1; triangles), from DS (squares) and from fifth-order Lagrange interpolation (circles).

can be improved to approximately the same level obtained with a longer horizon $T = \tau_L$. Nevertheless, in order to obtain satisfactory agreement over the whole range of wavenumbers, even smaller e_{tot} or $T > 0.5\tau_L$ is needed.

The parameter k_m affects the results strongly. Figure 7(e) plots $\hat{\rho}_{uv}$ at $t = T$ for $T = 0.5\tau_L$ with $k_m = 2, 4, 8$. When $k_m = 2$, $\boldsymbol{u}$ shows little correlation with $\boldsymbol{v}$. For $k_m = 8$, almost perfect correlation is obtained. To understand these results, it is instructive to

compare the 4DVAR method with the DS method (Yoshida *et al.* 2005; Lalescu *et al.* 2013) (cf. § 1), which is shown in figure 7(*f*). In Yoshida *et al.* (2005), the authors integrate the NSEs with a Gaussian random field as the IC. The low wavenumber modes with $k \leqslant k_m$ in the solution $\boldsymbol{u}(\boldsymbol{x}, t)$ are replaced by those in the target field $\boldsymbol{v}(\boldsymbol{x}, t)$ at each time step. They show that, if $k_m \geqslant k_c \equiv 0.2\eta_K^{-1}$, the small scales in $\boldsymbol{u}(\boldsymbol{x}, t)$ approach those in $\boldsymbol{v}(\boldsymbol{x}, t)$ asymptotically as $t \to \infty$, assuming an infinite sequence of data is available. In our investigation, the threshold wavenumber is $k_c = 5$ so $k_m = 4$ is slightly below k_c, and the measurement data are available only for $t \in [0, T]$. As a consequence, the DS scheme will only achieve partial reconstruction. The reconstruction at $t = \tau_L$ using the DS scheme with $k_m = 4$ is compared with the 4DVAR method in figure 7(*f*). It shows that the 4DVAR method improves the reconstruction significantly. A more simplistic reconstruction, obtained with a fifth-order Lagrange interpolation, is also examined in figure 7(*f*). The reconstruction has essentially no correlation with the target field for $k > k_m$.

Figures 7(*e*) and 7(*f*) together thus show that, for k_m approximately equal to or greater than k_c, 4DVAR can exploit a short time sequence of data to obtain better (for $k_m = 4$) or almost perfect (for $k_m = 8$) reconstruction. For k_m much smaller than k_c (e.g. for $k_m = 2$), 4DVAR will fail. These results suggest that the threshold k_c found with the DS method is also valid for 4DVAR. Nevertheless, a more comprehensive investigation is needed to confirm this preliminary observation. Simulations with other k_m values around k_c, as well as other Reynolds numbers, are needed. We leave this investigation for the future.

3.5. *Statistics of the reconstructed fields at $t = 0$*

In this subsection, we examine other dynamically important statistics of the reconstructed initial fields (i.e. $\boldsymbol{\varphi^u}(\boldsymbol{x})$). We consider the cases with $k_m = 4$ only and focus mostly on the baseline case C1 where $T = \tau_L$, $k_m = 4$ and $e_{tot} = 2 \times 10^{-4}$. We conduct the analysis in the physical space, using the filtering approach where necessary (cf. § 2.3). The Gaussian filter is used, and the filter scale Δ is usually taken as $\Delta_m/2$ or $\Delta_m/4$, where $\Delta_m \equiv \pi/k_m$ is the equivalent filter scale associated with k_m.

The joint PDFs of normalized Q and R are plotted in figure 8. The contours from the target fields depict the well-known skewed teardrop shape with the Vieillefosse tail extended in the fourth quadrant (Vieillefosse 1984; Cantwell 1992). Events in the second and the fourth quadrants occur with higher probabilities. These features are qualitatively reproduced by the contours of the reconstructed fields (dashed lines). However, the Vieillefosse tail is shorter, and the probabilities for large excursions are lower by a rather significant amount. The lowest contours for the reconstructions with larger T and smaller e_{tot} ((*a,b*) and (*c,d*)) are closer to those for the target, although the improvement is very small. Panel (*b,d,f*) shows that the discrepancy is bigger when the filter scale is smaller.

The mean strain rate, enstrophy and vortex stretching term, normalized by the Kolmogorov time scale τ_K, are given in figure 9 as functions of the filter scale. Not surprisingly, the differences between the results in the two fields are larger for smaller filter scales. For $\langle \tilde{s}_{ij}\tilde{s}_{ij} \rangle$ and $\langle \tilde{\omega}_i\tilde{\omega}_i \rangle$ the discrepancy is below 20 % for all filter scales. The discrepancy for the stretching term however is much bigger. The latter can be partially explained by the misalignment between $\tilde{\omega}_i$ and $\tilde{s}_{ij}$ in $\boldsymbol{\varphi^u}$, which is shown in figure 10 with the PDFs for the cosines of the angles between $\tilde{\omega}_i$ and the eigenvectors of $\tilde{s}_{ij}$. The results for $\boldsymbol{\varphi^u}$ display qualitatively correct features, i.e. a
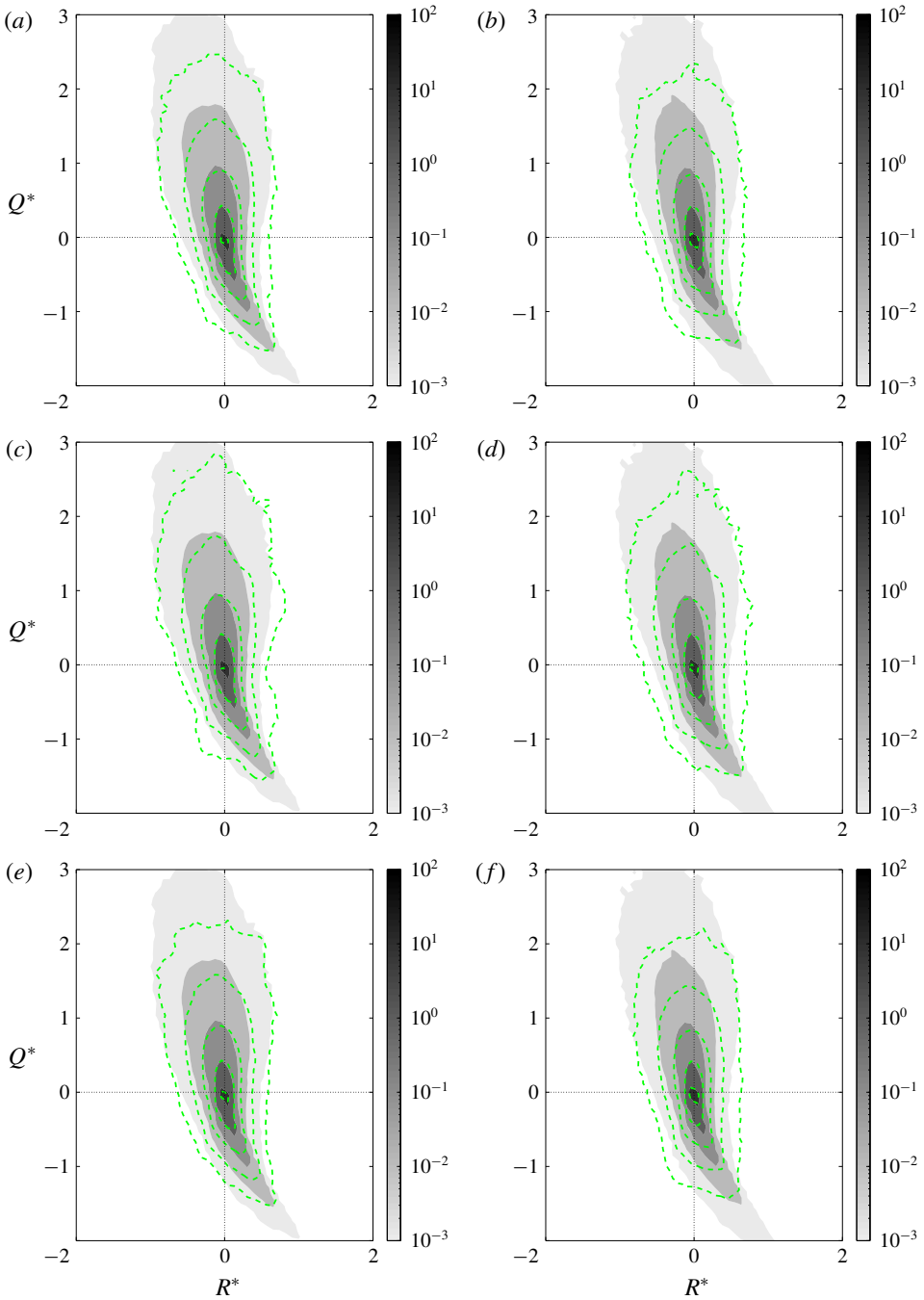
FIGURE 8. The joint PDF of $Q^* \equiv Q/\langle \tilde{A}_{ij}\tilde{A}_{ij}\rangle$ and $R^* \equiv R/\langle \tilde{A}_{ij}\tilde{A}_{ij}\rangle^{3/2}$: (a,c,e) $\Delta = \Delta_m/2$; (b,d,f) $\Delta = \Delta_m/4$. (a,b) Case C1; (c,d) case C2; (e,f) case C6. Grey scales: from $\varphi^v(\boldsymbol{x})$; dashed lines: from $\varphi^u(\boldsymbol{x})$.
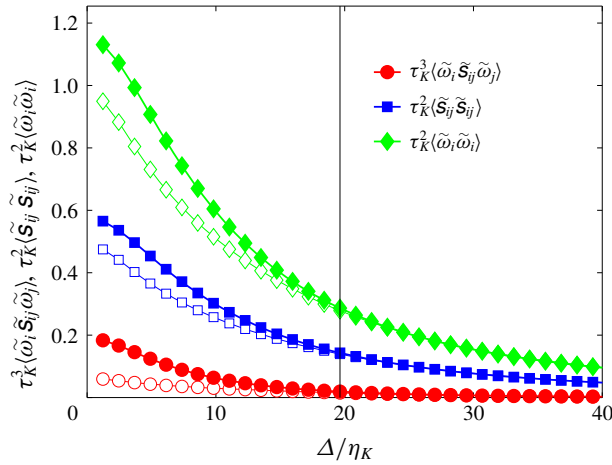
FIGURE 9. Values of $\langle \tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j \rangle$ (circles), $\langle \tilde{s}_{ji}\tilde{s}_{ij} \rangle$ (squares) and $\langle \tilde{\omega}_i\tilde{\omega}_i \rangle$ (diamonds) as functions of filter scale Δ from case C1. Filled and empty symbols: from $\boldsymbol{\varphi^v}(\boldsymbol{x})$ and $\boldsymbol{\varphi^u}(\boldsymbol{x})$, respectively. The vertical line marks the scale where $\Delta = \Delta_m$.
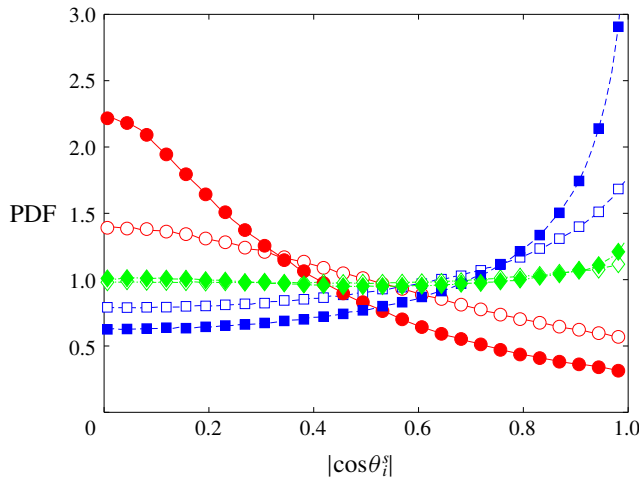
FIGURE 10. PDFs of $|\cos\theta_i^s|$ with θ_i^s being the angle between $\tilde{\boldsymbol{\omega}}$ and $\boldsymbol{e}_\alpha^s$ (diamonds), $\boldsymbol{e}_\beta^s$ (squares) and $\boldsymbol{e}_\gamma^s$ (circles). Filled symbols: $\boldsymbol{\varphi^v}(\boldsymbol{x})$; empty symbols: $\boldsymbol{\varphi^u}(\boldsymbol{x})$. From case C1 and $\Delta = \Delta_m/2$.

preferred alignment between $\tilde{\omega}_i$ and $\boldsymbol{e}_\beta^s$, and the tendency for $\tilde{\omega}_i$ to be perpendicular to $\boldsymbol{e}_\gamma^s$. However, the peaks of the PDFs are lower to a significant degree. Therefore, the preferable alignment in $\boldsymbol{\varphi^u}$ is weaker, and this misalignment contributes to the observed discrepancy in the vortex stretching term.

The mean SGS energy dissipation $\langle \Pi_\Delta \rangle$ is shown in figure 11(a). The results from $\boldsymbol{\varphi^u}$ display the same trends as those from $\boldsymbol{\varphi^v}$. This result is a direct evidence that the nonlinear interactions between different scales are partially reconstructed. At $\Delta = \Delta_m/2$ ($\Delta/\eta_K \approx 10$), the value calculated from $\boldsymbol{\varphi^u}$ is smaller by approximately 25 %. The results for the two different T values show only very slight differences. The PDFs for the normalized Π_Δ are shown in figure 11(b). The well-known
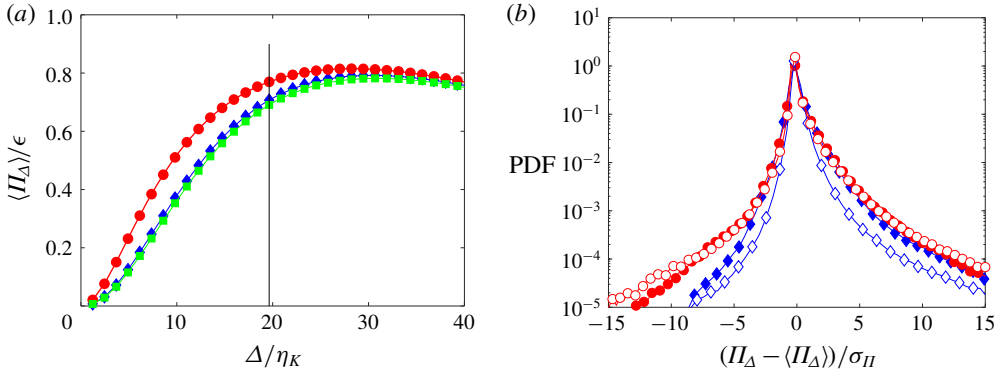
FIGURE 11. (*a*) The mean SGS energy dissipation $\langle \Pi_\Delta \rangle / \epsilon$ as a function of Δ / η_K for $\boldsymbol{\varphi}^v(\boldsymbol{x})$ (circles) and $\boldsymbol{\varphi}^u(\boldsymbol{x})$ from cases C1 (diamonds) and C6 (squares). The vertical line corresponds to the $\Delta = \Delta_m$. (*b*) PDF of normalized Π_Δ for $\Delta = \Delta_m/2$ (filled symbols) and $\Delta = \Delta_m/4$ (empty symbols), from $\boldsymbol{\varphi}^v$ (circles) and $\boldsymbol{\varphi}^u$ (diamonds) for case C1.

positive skewness is captured by $\boldsymbol{\varphi}^u(\boldsymbol{x})$. The agreement on the positive tail is very good, although the negative tail is under-predicted, meaning the reconstructed fields under-predict the probability for instantaneous energy backscattering. Only the PDFs for $T = \tau_L$ are shown, as the results with different T values have no discernible differences.

The conditionally averaged SGS energy dissipation $\langle \Pi_\Delta | Q^*, R^* \rangle$ has also been calculated, where $Q^* \equiv Q / \langle \tilde{A}_{ij} \tilde{A}_{ij} \rangle$ and $R^* \equiv R / \langle \tilde{A}_{ij} \tilde{A}_{ij} \rangle^{3/2}$. Shown in figure 12 is $\langle \Pi_\Delta | Q^*, R^* \rangle$ weighted by the joint PDF $P(R^*, Q^*)$ and normalized by the mean energy dissipation rate ϵ. The results from $\boldsymbol{\varphi}^v$ (*a,d*) show that higher SGS dissipation is observed around the Vieillefosse tail where high straining motion happens. The behaviour is reproduced qualitatively by $\boldsymbol{\varphi}^u$, shown in (*b,e*) and in (*c,f*) for $T = \tau_L$ and $0.5\tau_L$, respectively. Negative SGS dissipation, representing energy backscattering, is found mainly in the first quadrant, a feature also captured by $\boldsymbol{\varphi}^u$. However, overall, $\boldsymbol{\varphi}^u$ tends to underestimate the magnitude of the weighted SGS energy dissipation. The results at $\Delta = \Delta_m/2$ are closer than those at $\Delta = \Delta_m/4$. There are some small improvements with $T = \tau_L$ compared with $T = 0.5\tau_L$.

To summarize, the results in this subsection demonstrate that $\boldsymbol{\varphi}^u(\boldsymbol{x})$ captures the main statistical features of real turbulence with reasonable agreement. This is the basis for the reconstruction of the instantaneous velocity fields at later times. The discrepancy between $\boldsymbol{\varphi}^v(\boldsymbol{x})$ and $\boldsymbol{\varphi}^u(\boldsymbol{x})$, manifested by the several statistics we considered, can be reduced slightly by extending the optimization horizon from $T = 0.5\tau_L$ to $T = \tau_L$ or by using a smaller e_{tot}. Given the chaotic nature of the flow, these small differences might be the explanation for the significant improvement found at later time $t \to T$, as shown in § 3.4. This conjecture may be quantified by calculating the sensitivities of the statistics although this is beyond the scope of this paper.

3.6. *Reconstruction of the instantaneous local structures for $t \to T$*

For the reconstructed field $\boldsymbol{u}(\boldsymbol{x}, t)$ at later times, it is expected that its statistics would agree well with those in the target field $\boldsymbol{v}(\boldsymbol{x}, t)$, for the simple reason that both are the stationary solutions of the NSEs. However, there is no *a priori* reason to assert that
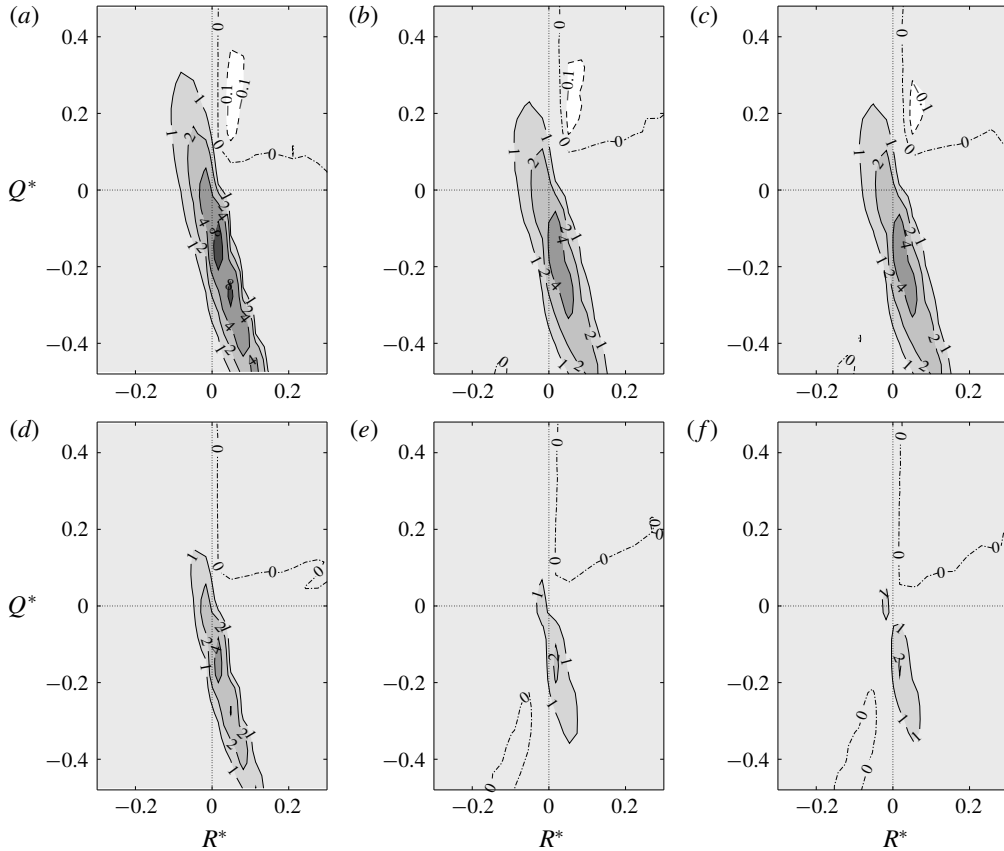
FIGURE 12. The normalized weighted conditional average of the SGS energy dissipation $\epsilon^{-1}P(R^*, Q^*)\langle\Pi_\Delta|R^*, Q^*\rangle$. (a,d) from φ^v; (b,e) from φ^u in case C1; (c,f) from φ^u in case C6. $(a–c)$ $\Delta = \Delta_m/2$; $(d–f)$ $\Delta = \Delta_m/4$.

the $\boldsymbol{u}(\boldsymbol{x}, t)$ would have pointwise agreement with $\boldsymbol{v}(\boldsymbol{x}, t)$. The pointwise differences have been examined briefly using $E_D(k)$ and $\hat{\rho}_{uv}(k)$ in § 3.4 to demonstrate the effects of the choices of computational parameters. In this subsection, the baseline case C1 (cf. table 1) is examined in further detail. Note that time t is now another parameter. We usually choose $t = T$. However, when comparing results with different T values, other values of t are also used.

3.6.1. *Geometry of the local structures*

The instantaneous difference in the geometrical structures of the reconstructed fields $\boldsymbol{u}$ and the target fields $\boldsymbol{v}$ is first investigated. The first quantity we examine is the 'cross-alignment' between the vorticity in $\boldsymbol{v}$ ($\tilde{\omega}_i^v$) and the eigenvectors of the strain rate tensor in $\boldsymbol{u}$ ($\tilde{s}_{ij}^{\boldsymbol{u}}$). The alignment is measured by the joint PDF of $\cos\theta_{sp}^\omega$ and ϕ_{sp}^ω, where the angles define the orientation of $\tilde{\omega}_i^v$ in the eigenframe of $\tilde{s}_{ij}^{\boldsymbol{u}}$ (cf. figure 1). The joint PDF is shown with grey scale contours in figure 13. The joint PDF for the usual alignment between $\tilde{\omega}_i^v$ and $\tilde{s}_{ij}^v$, both from $\boldsymbol{v}$, is plotted with dashed lines as a comparison. Figure 13(a,d) shows that the joint PDF for $\tilde{\omega}_i^v$ and $\tilde{s}_{ij}^{\boldsymbol{u}}$ is very close to the one for $\tilde{\omega}_i^v$ and $\tilde{s}_{ij}^v$ for both $\Delta = \Delta_m/2$ and $\Delta_m/4$. Both joint PDFs display a high
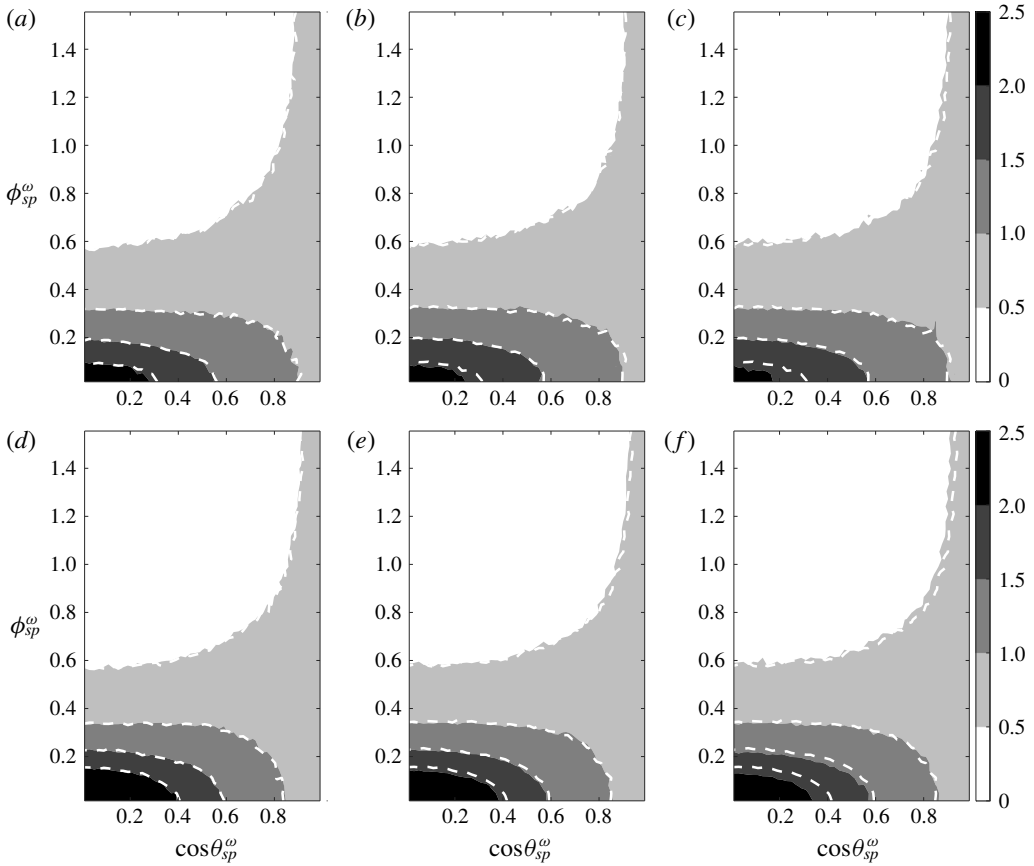
FIGURE 13. The joint PDFs of $(\cos\theta^{\omega}_{sp}, \phi^{\omega}_{sp})$ for angles between $\tilde{\omega}^{v}_{i}$ and $\tilde{s}^{u}_{ij}$ (grey scales) and those between $\tilde{\omega}^{v}_{i}$ and $\tilde{s}^{v}_{ij}$ (dashed lines). $(a\text{–}c)$ $\Delta = \Delta_{m}/2$; $(d\text{–}f)$ $\Delta = \Delta_{m}/4$. (a,d) $t = T$ for case C1; (b,e) $t = 0.5T$ for case C1; (c,f) $t = T$ for case C6.

peak around $(0, 0)$, which shows the well-known preferred alignment of the vorticity vector with $\boldsymbol{e}^{s}_{\beta}$ in turbulence. The peak probability density for the cross-alignment is only slightly lower. However, the fact that replacing $\tilde{s}^{v}_{ij}$ by $\tilde{s}^{u}_{ij}$ still yields a close joint PDF proves that the instantaneous orientations of the eigenvectors of $\tilde{s}^{u}_{i}$ and $\tilde{s}^{v}_{ij}$ are very close to each other. Same behaviours are observed for the results in (b,e), where the results at an earlier time $t = T/2$ for the same $T = \tau_{L}$ are shown. The joint PDFs display only slightly larger discrepancy. The result obtained with a shorter horizon $T = 0.5\tau_{L}$, given in (c,f), shows much larger discrepancy, as the highest level contour is much smaller even though the location of the peak is correctly found at $(0, 0)$.

Figure 14 considers the cross-alignment between $\tilde{\omega}^{v}_{i}$ and the eigenvectors of $-\tau^{d,u}_{ij}$. The dashed lines show the joint PDF for the usual alignment in $\boldsymbol{v}$. The two peaks of the joint PDF depict the known feature in turbulence, where $\tilde{\omega}_{i}$ tends to align with $\boldsymbol{e}^{\tau}_{\alpha}$ and $\boldsymbol{e}^{\tau}_{\beta}$. The cross-alignment between $\tilde{\omega}^{v}_{i}$ and $-\tau^{d,u}_{ij}$ qualitatively reproduces these features. The results at $\Delta = \Delta_{m}/2$ show excellent quantitative agreement, although the discrepancy becomes larger for $\Delta = \Delta_{m}/4$. The comparison at $t = 0.5\tau_{L}$ for $T = 0.5\tau_{L}$ and τ_{L} is not shown, because it simply confirms our expectation that the results with a smaller optimization horizon T display larger discrepancy.
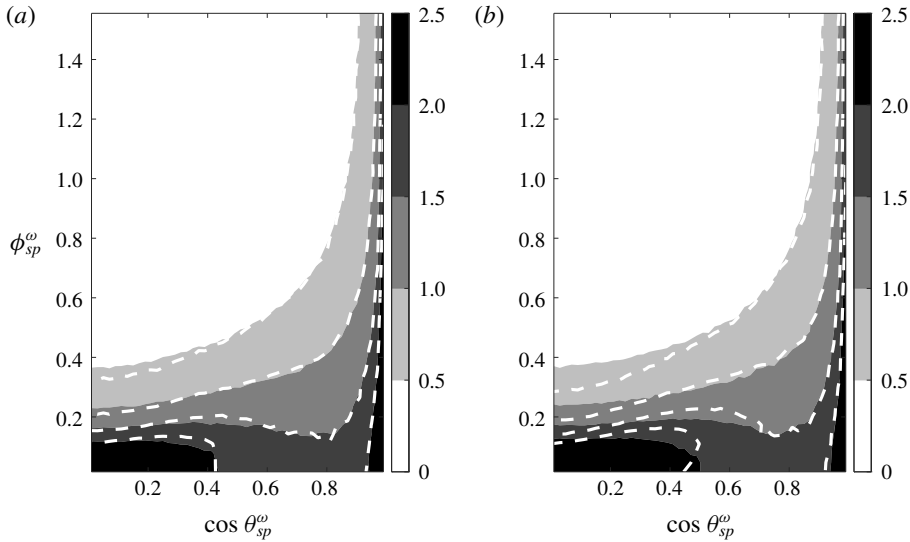
FIGURE 14. The joint PDFs of $(\cos\theta_{sp}^{\omega}, \phi_{sp}^{\omega})$ for angles between $\tilde{\omega}_i^v$ and $-\tau_{ij}^{d,u}$ (grey scales) and those between $\tilde{\omega}_i^v$ and $-\tau_{ij}^{d,v}$ (dashed lines), for $\Delta = \Delta_m/2$ (a) and $\Delta_m/4$ (b) with $t = T$ in case C1.
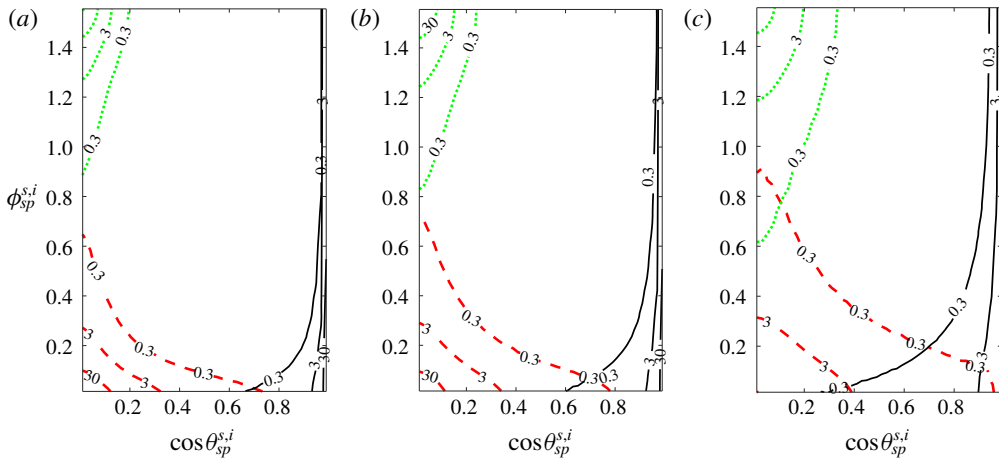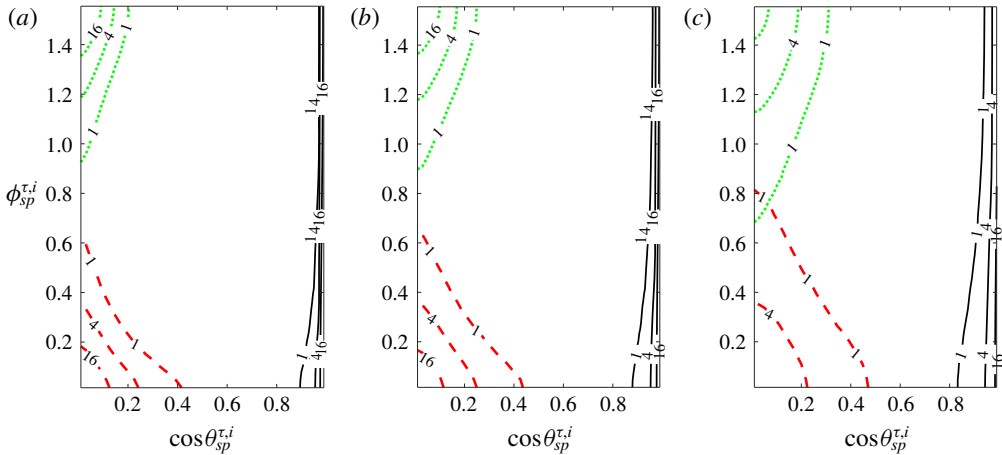

FIGURE 15. The joint PDF of $(\cos\theta_{sp}^{s,i}, \phi_{sp}^{s,i})$ for angles made by $\boldsymbol{e}_i^{s,u}$ in the eigenframe of $\tilde{s}_{ij}^v$. (a) $\Delta = \Delta_m/2$ and $t = T$ from case C1; (b) $\Delta = \Delta_m/4$ and $t = T$ from case C1; (c) $\Delta = \Delta_m/2$ and $t = T$ for case C6. Solid lines: $i = \alpha$; dashed lines: $i = \beta$; dotted lines: $i = \gamma$. The contour levels are 0.3, 3 and 30.

The tensorial structures of $\tilde{s}_{ij}^u$ and $\tilde{s}_{ij}^v$ are compared using the orientation of the eigenvectors of the former in the eigenframe of the latter. The alignment is given in figure 15(a) for $\Delta = \Delta_m/2$ and $T = \tau_L$, in term of the joint PDFs of $(\cos\theta_{sp}^{s,i}, \phi_{sp}^{s,i})$, where $\theta_{sp}^{s,i}$ and $\phi_{sp}^{s,i}$ are the angles made by $\boldsymbol{e}_i^{s,u}$ in the eigenframe of $\tilde{s}_{ij}^v$ (cf. figure 1). According to the definitions of the angles, perfect alignment between $\boldsymbol{e}_\alpha^{s,v}$ and $\boldsymbol{e}_\alpha^{s,u}$ corresponds to $\cos\theta_{sp}^{s,\alpha} = 1$ and an undetermined $\phi_{sp}^{s,\alpha}$. The solid lines in figure 15

FIGURE 16. The joint PDF of $(\cos\theta_{sp}^{\tau,i}, \phi_{sp}^{\tau,i})$ for angles made by $e_i^{\tau,u}$ in the eigenframe of $-\tau_{ij}^{d,v}$. (a) $\Delta = \Delta_m/2$ and $t = T$ from case C1; (b) $\Delta = \Delta_m/4$ and $t = T$ from case C1; (c) $\Delta = \Delta_m/2$ and $t = T$ from case C6. Solid lines: $i = \alpha$; dashed lines: $i = \beta$; dotted lines: $i = \gamma$. The contour levels are 1, 4 and 16.

show a sharp ridge of peak values at $\cos\theta_{sp}^{s,i} = 1$, which indicates that $e_\alpha^{s,v}$ and $e_\alpha^{s,u}$ are strongly aligned. Perfect alignment between $e_\beta^{s,v}$ and $e_\beta^{s,u}$ corresponds to a peak at $(0, 0)$, whereas the perfect alignment between $e_\gamma^{s,v}$ and $e_\gamma^{s,u}$ corresponds to a peak at $(0, \pi/2)$. The joint PDFs in these two cases (shown with dashed and dotted lines, respectively) indeed display high peak probabilities around these points. Therefore, these joint PDFs demonstrate excellent agreement between the orientations of $\tilde{s}_{ij}^u$ and $\tilde{s}_{ij}^v$. The results at $\Delta = \Delta_m/4$ given in (b) show that strong alignment persists to smaller filter scales. The joint PDFs obtained with $T = 0.5\tau_L$, shown in (c), have much weaker peaks, indicating a larger discrepancy in the orientation of the two tensors.

Figure 16 plots the same statistics as those in figure 15, but for the SGS stress tensors $-\tau_{ij}^{d,v}$ and $-\tau_{ij}^{d,u}$. The observations made from figure 15 are also observed in figure 16, except that the alignment shown in the latter is in general somewhat weaker.

3.6.2. *Correlations*

We will look into the covariance and the correlation coefficients in this subsection. For two random variables X and Y, they are denoted by $C(X, Y)$ and $\rho(X, Y)$, respectively. By standard definitions,

$$C(X, Y) \equiv \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle, \quad \rho(X, Y) \equiv \frac{C(X, Y)}{\sigma_X \sigma_Y}, \quad (3.3a,b)$$

where $\sigma_X \equiv \sqrt{C(X, X)}$ is the r.m.s. value of X and σ_Y is defined similarly. The physical quantities in $u(x, t)$ and $v(x, t)$ will be denoted with superscripts u and v, respectively. The difference in X is quantified by the r.m.s. value of $X^u - X^v$, $\sigma_{X^u-X^v}$, which will be normalized by the r.m.s. value of X in the target fields σ_{X^v}.

Figure 17(a) shows $\rho(X^u, X^v)$, where $X \equiv \tilde{s}_{ij}\tilde{s}_{ij}$. The correlation deteriorates with decreasing filter scale, but improves with t/T for t within the optimization horizon. Excellent correlation across all scales is achieved at $t/T = 1$ for $T = \tau_L$. The difference in $X = \tilde{s}_{ij}\tilde{s}_{ij}$ is also examined in the right panel in terms of the r.m.s. value of
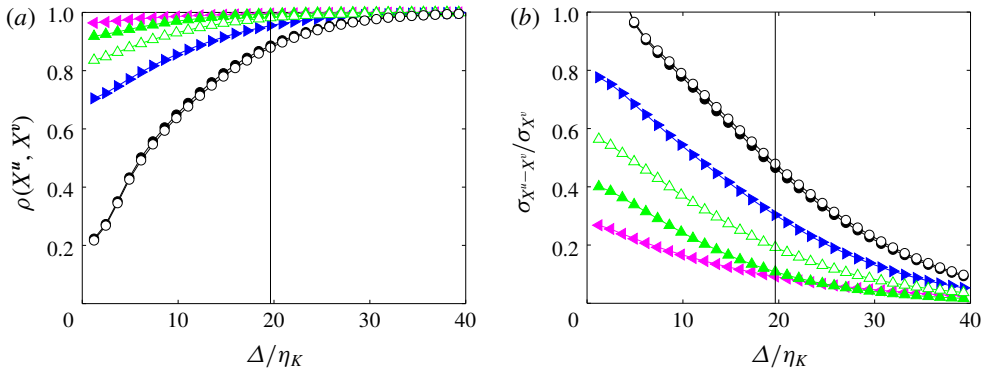
FIGURE 17. The correlation coefficient $\rho(X^u, X^v)$ (a) and the normalized r.m.s. value of $X^u - X^v$ (b) as a function of Δ/η_K, where $X = \tilde{s}_{ij}\tilde{s}_{ij}$. Solid symbols: from case C1; open symbols: from case C6. Circles: $t = 0$; upright triangles: $t = 0.5\tau_L$; left-pointing triangles: $t = \tau_L$ (for case C1 only). Right-pointing solid triangles: from the DS method at $t = \tau_L$ for $k_m = 4$. The vertical line marks the scale $\Delta = \Delta_m$.
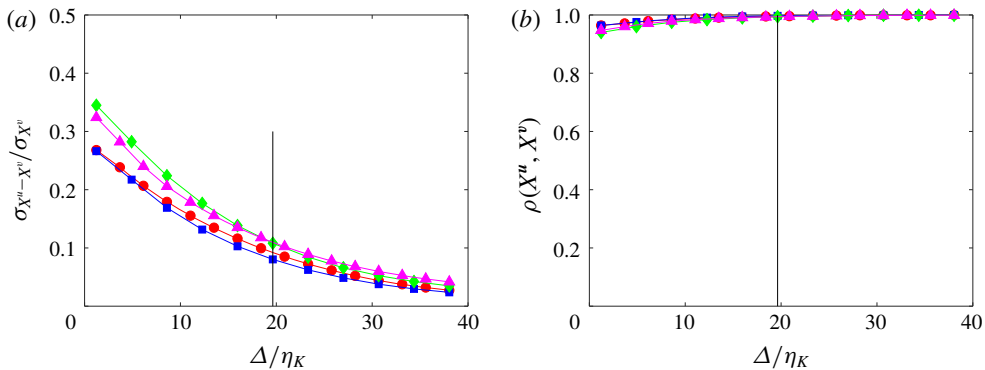

FIGURE 18. (a) The normalized r.m.s. values of the differences $X^u - X^v$ from case C1, where $X \equiv \tilde{s}_{ij}\tilde{s}_{ij}$ (circles), $\tilde{\omega}_i\tilde{\omega}_i$ (squares), $P_\omega(\equiv \tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j)$ (diamonds) and Π_Δ (triangles), respectively. (b) The correlation coefficients $\rho(X^u, X^v)$ from case C1. The vertical line marks the scale $\Delta = \Delta_m$.

$X^u - X^v$ normalized by the r.m.s. value of X^v. For $t = T = \tau_L$, the value can be reduced to approximately 27 % at the Kolmogorov scale where $\Delta/\eta_K \approx 1$. For results from the DS method (right-pointing solid triangles), the correlation ρ at the smallest Δ/η_K is approximately 70 % and the r.m.s. difference is approximately 80 %, indicating significantly higher discrepancy. This contrast is consistently observed in other statistics. Therefore, we will not present more results from the DS method.

The values of $\sigma_{X^u-X^v}/\sigma_{X^v}$ and $\rho(X^u, X^v)$ for $X = \tilde{\omega}_i\tilde{\omega}_i$, P_ω and Π_Δ are plotted in figure 18, where the results for $\tilde{s}_{ij}\tilde{s}_{ij}$ are also given as comparison. For all quantities, almost perfect correlation is observed. The r.m.s. difference for $\tilde{\omega}_i\tilde{\omega}_i$ is similar to $\tilde{s}_{ij}\tilde{s}_{ij}$, both reaching approximately 27 % at the Kolmogorov scale where $\Delta/\eta_K \approx 1$. The discrepancy for Π_Δ and P_ω is somewhat larger with the normalized r.m.s. difference found at approximately 35 % at the smallest filter scale.
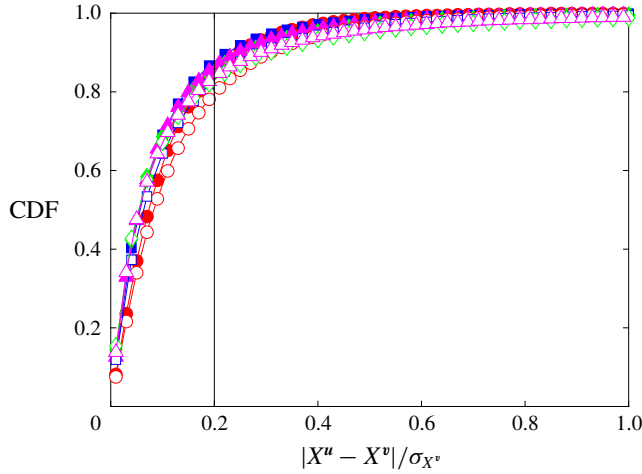
FIGURE 19. The cumulative distribution functions for the normalized differences $|X^u - X^v|$ from case C1, where $X = \tilde{s}_{ij}\tilde{s}_{ij}$ (circles), $\tilde{\omega}_i\tilde{\omega}_i$ (squares), $P_\omega \equiv \tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j$ (diamonds) and Π_Δ (triangles). Filled symbols: $\Delta = \Delta_m/2$; empty symbols: $\Delta = \Delta_m/4$. The vertical line marks the 20 % threshold.

3.6.3. *Joint PDFs and conditional statistics*

To examine the large fluctuations in the differences between the target and reconstructed parameters, we look into the joint PDFs and conditional statistics in this subsection. Figure 19 plots the cumulative distribution functions for the normalized differences. The results for both $\Delta = \Delta_m/2$ and $\Delta_m/4$ are given. The vertical line indicates an arbitrarily chosen 20 % threshold. The probabilities at $\Delta = \Delta_m/4$ are slightly lower than those at $\Delta_m/2$, and those for $\tilde{s}_{ij}\tilde{s}_{ij}$ are slightly lower than those for the other quantities. The probability of observing a less than 20 % difference ranges from 75 % to 90 % for different quantities.

The conditional and joint statistics, e.g. the average of X^u conditioned on X^v and the joint PDF of the two, are presented next. Figure 20(*a*) shows the results for $X \equiv \Pi_\Delta$ with $\Delta = \Delta_m/4$. For a perfectly reconstructed field, the conditional average will fall on the diagonal (dotted line). Therefore, the figure shows that Π_Δ^u slightly underestimates the magnitude of Π_Δ^v, and the underestimate increases slightly with the magnitude of Π_Δ^v. Nevertheless, the agreement is very good. The joint PDF does concentrate around the diagonal, although there is significant scattering. The scattering is indicated by the error bars, which represent $\pm\sigma_{\Pi_\Delta^u|\Pi_\Delta^v}/\langle\Pi_\Delta^v\rangle$ with $\sigma_{\Pi_\Delta^u|\Pi_\Delta^v} \equiv [\langle(\Pi_\Delta^u)^2|\Pi_\Delta^v\rangle - \langle\Pi_\Delta^u|\Pi_\Delta^v\rangle^2]^{1/2}$ being the conditional r.m.s. value of Π_Δ^u. The error bar increases mildly with the magnitude of Π_Δ^v. Figure 20(*b*) plots the same results for $P_\omega \equiv \tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j$. The results for enstrophy and strain rate are shown in figure 21. Good agreement is also observed for these quantities. The error bars are slightly larger for P_ω^u and $\tilde{\omega}_i^u\tilde{\omega}_i^u$, compared with the other two quantities.

3.7. *Reconstruction of non-local structures*

Figure 2 has shown that the shapes, orientations and locations of the strong vortices in the reconstructed field can be remarkably similar to those in the target field. We quantify the similarity of the vortical and other structures using MVEEs and MVEE trees in this subsection.
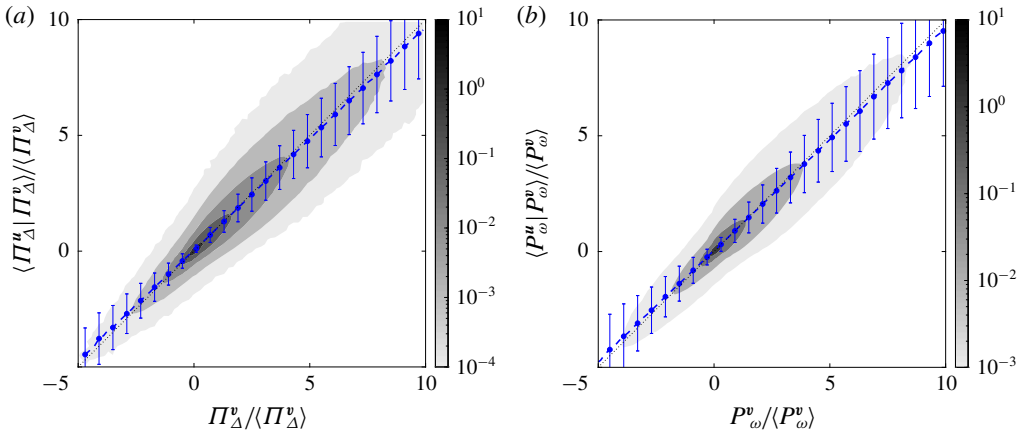
FIGURE 20. (*a*) The averaged Π_Δ in $\boldsymbol{u}(\boldsymbol{x}, t)$ conditioned on Π_Δ in $\boldsymbol{v}(\boldsymbol{x}, t)$ (circles) and the joint PDF of the two (contours). The error bars are the conditional r.m.s. values of $\Pi_\Delta^u / \langle \Pi_\Delta^v \rangle$. (*b*) The same for P_ω. For case C1 with $\Delta = \Delta_m/4$ and $t = T$.
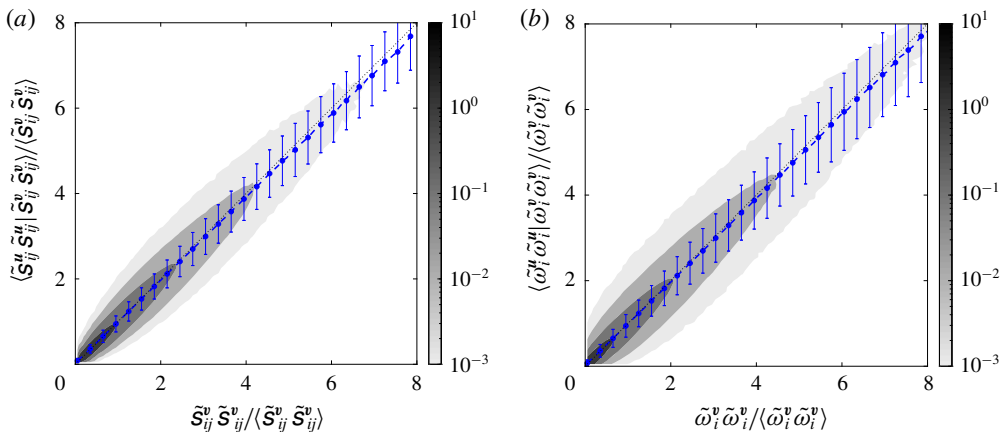

FIGURE 21. Same as figure 20 but for $\tilde{s}_{ij}\tilde{s}_{ij}$ (*a*) and $\tilde{\omega}_i\tilde{\omega}_i$ (*b*).

3.7.1. *Vortical structures*

The calculation of MVEEs have been explained in § 2.4. Let the directions of the axes of an MVEE be denoted by $\boldsymbol{e}_i$ ($i = \alpha, \beta, \gamma$), and the length of axis i be denoted by ℓ_i. The indices α, β and γ denote the major, intermediate and the minor axes, respectively, so that $\ell_\alpha \geqslant \ell_\beta \geqslant \ell_\gamma$. As before, the parameters in the target fields are indicated by superscript/subscript $\boldsymbol{v}$ and those in the reconstructed fields by $\boldsymbol{u}$.

We now consider strong vortical structures where $\tilde{\omega} \geqslant 3\langle \tilde{\omega}_i \tilde{\omega}_i \rangle^{1/2}$ with filter scale $\Delta = \Delta_m/2$ and $\Delta_m/4$. As it turns out, the number of isolated structures in a field may be too large to process. If this happens, only the largest $N_s \equiv 10$ structures in the target $\boldsymbol{v}$ and their matching structures in $\boldsymbol{u}$ are used. Only the data in the five realizations of case C1 are analysed. There are thus in total $N_e = 5N_s = 50$ pairs of matching MVEEs at each time step. As the number of samples is modest, medians instead of means are used to delineate the average behaviours, because the former is
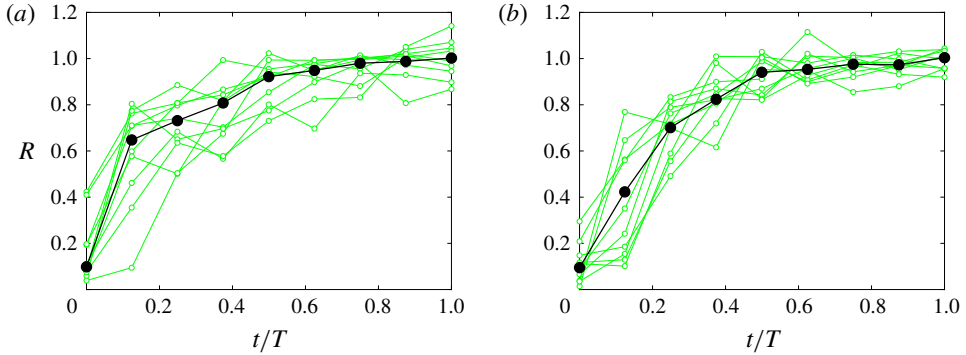
FIGURE 22. The median R of the ratio between the numbers of grid points in two matching structures (circles with thick line). The green thin lines show the medians in different size groups. (a) $\Delta = \Delta_m/2$; (b) $\Delta = \Delta_m/4$. From case C1.
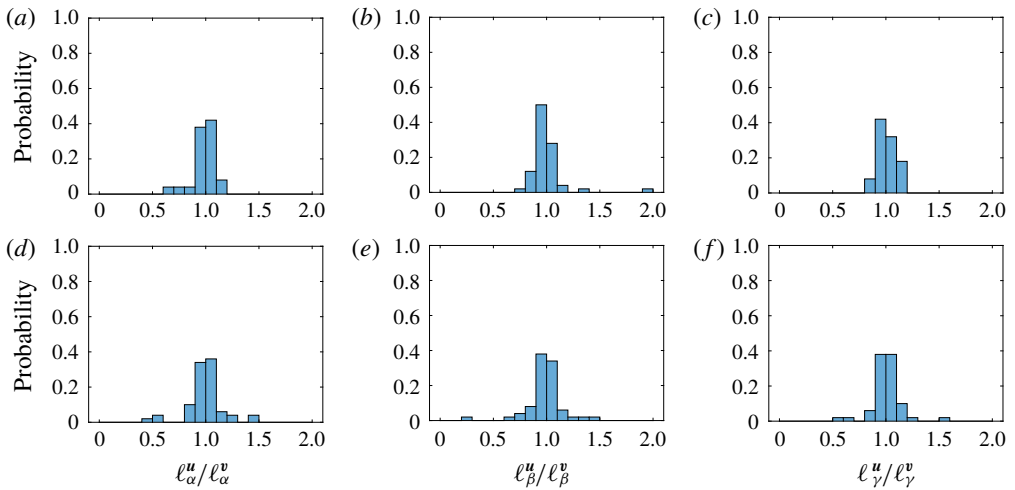

FIGURE 23. The normalized histograms of ℓ_i^u/ℓ_i^v for $i = \alpha$ (a,d), $i = \beta$ (b,e) and $i = \gamma$ (c,f), with $\Delta = \Delta_m/2$ (a–c) and $\Delta = \Delta_m/4$ (d–f). From case C1 with $t = T$.

less sensitive to extraneous fluctuations. Additional details are given by percentiles and histograms.

Figure 22 compares the sizes of the vortical structures in u and v in terms of R, which is the median of the ratio between the numbers of grid points in the two matching vortices. There are $N_s = 10$ groups of vortices of different sizes. The thick line with filled circles shows R calculated over all $N_s = 10$ groups, whereas the thin lines with empty circles show R in each group separately. The figure shows that the vortices in u tend to be smaller initially, the median ratio being only approximately 10 %. However, it increases quickly, reaching 90 % at $t/T \approx 0.5$ and approximately 1 at the end of the optimization horizon. The same behaviour is observed at two filter scales. The thin lines show that there is significant statistical variation between different groups.

The sizes can also be measured with the lengths of the axes of the MVEEs. Figure 23 plots the histograms of the ratios of the axes in u to those in v at $t/T = 1$
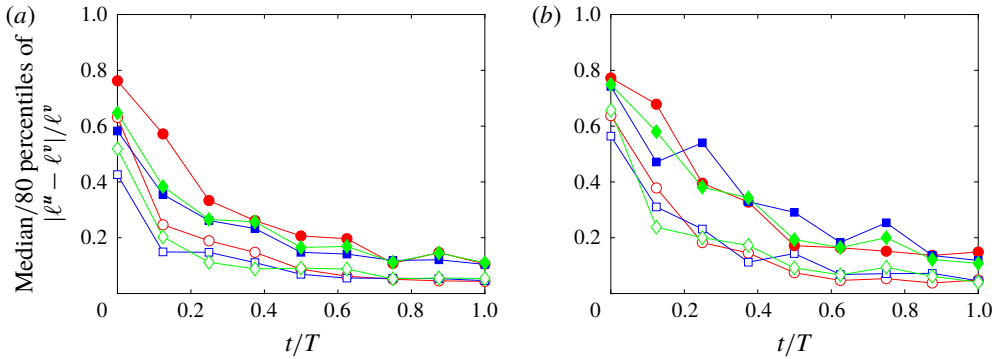
FIGURE 24. The medians (empty symbols) and the 80th percentiles (solid symbols) of $|\ell_i^u - \ell_i^v|/\ell_i^v$ $(i = \alpha, \beta, \gamma)$ for the MVEEs of matching vortices. Circles: $i = \alpha$; squares: $i = \beta$; diamonds: $i = \gamma$. (a) $\Delta = \Delta_m/2$; (b) $\Delta = \Delta_m/4$. From case C1.

for case C1. The results are consistent with figure 22. The peaks of the histograms show that 80 % of the axes of the MVEEs in $\boldsymbol{u}$ fall within ± 10 % of those in $\boldsymbol{v}$.

The median and 80th percentile for $|\ell_i^u - \ell_i^v|/\ell_i^v$ (i.e. the relative error in the lengths of the axes) are shown in figure 24 for two filter scales. Both values decrease over time. At $t/T = 1$, the median is reduced down to somewhere between 5 % and 10 %, i.e. for approximately half of the MVEEs, the relative error in the length of the axes is less than 5 %–10 %. The result for the 80th percentile shows that, for 80 % of the MVEEs the relative error is below 10 %–15 %. The difference between different axes is very small.

The alignment between $\boldsymbol{e}_i^v$ and $\boldsymbol{e}_i^u$ is examined in figure 25, where the histograms of $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)|$ are plotted $(t/T = 0, 1/2, 1$ from left to right, respectively). The ith axes of the matching MVEEs would align perfectly if $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)| = 1$. Initially, the correct alignment between the axes is visible, but the peaks are not strong. Perfect alignment develops quickly. The histograms at $t/T = 1/2$ and 1 basically concentrate around 1. In figure 26, the medians and the 20th percentiles of the cosines are plotted for two filter scales. In both cases, $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)| \geqslant 0.95$ is observed at $t/T = 1$ for the 20th percentiles for all axes, which corresponds to angles smaller than $18°$. Therefore, only 20 % of samples have a misalignment angle bigger than $18°$.

Figures 27 and 28 compare the locations of the MVEEs. The quantity being examined is $d_{c,i}/\ell_i^v$ $(i = \alpha, \beta, \gamma)$, where $d_{c,i} \equiv |(\boldsymbol{c}^v - \boldsymbol{c}^u) \cdot \boldsymbol{e}_i^v|$ is the displacement between the centres of the MVEEs in the direction of $\boldsymbol{e}_i^v$. Thus, $d_{c,i}/\ell_i^v$ is the displacement relative to the lengths of the axes of the MVEE in $\boldsymbol{v}$. The median and the 80th percentile of $d_{c,i}/\ell_i^v$ are given in figure 27. The 80th percentile is quite large initially, but decreases quickly over time. For $\Delta = \Delta_m/2$, the relative displacement at $t/T = 1$ is less than 10 % for the 80th percentiles, and less than 5 % for the medians. For $\Delta = \Delta_m/4$, the displacement is slightly larger, but is still less than 15 % for the 80th percentile.

3.7.2. *Straining structures and SGS energy dissipation structures*

The MVEEs are used in this subsection to investigate the strain rate and the SGS energy dissipation rate. As the results show behaviours similar to those for the vortical structures, only selected results are presented. For the strain rate, the structures with $(\tilde{s}_{ij}\tilde{s}_{ij})^{1/2} \geqslant 2.5\langle\tilde{s}_{ij}\tilde{s}_{ij}\rangle^{1/2}$ are considered. For Π_Δ, we examine both structures with large
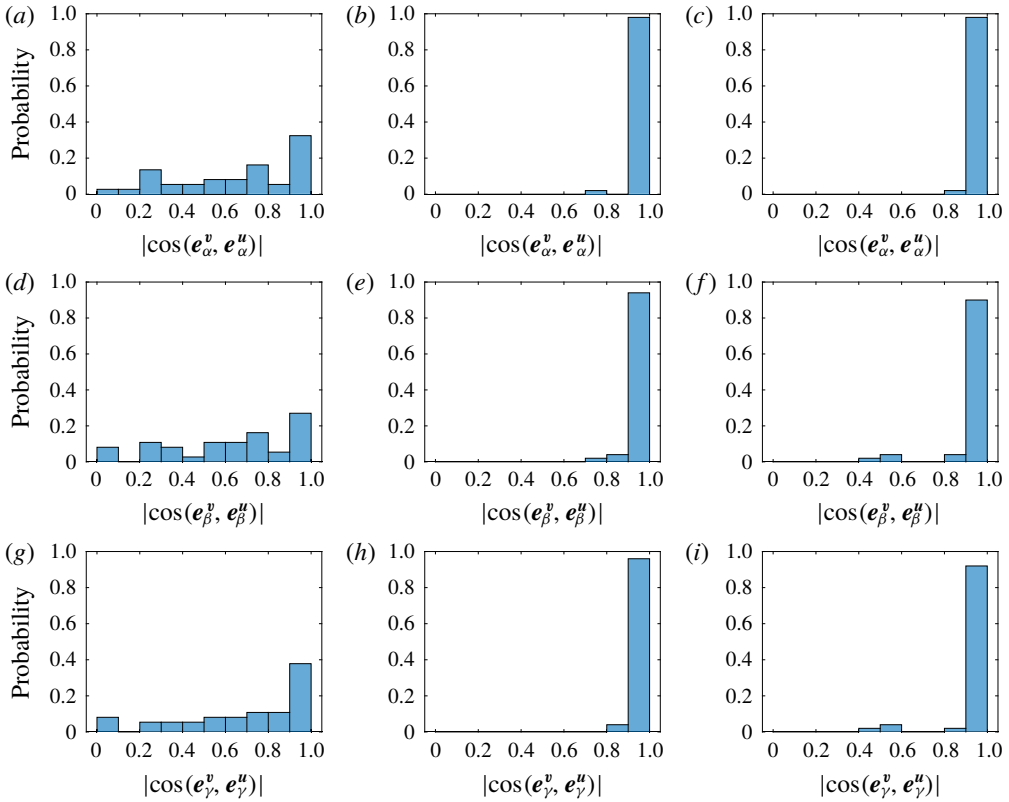
FIGURE 25. The histograms of $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)|$ for $i = \alpha$ (*a–c*), $i = \beta$ (*d–f*) and $i = \gamma$ (*g–i*), at $t/T = 0$ (*a,d,g*), $1/2$ (*b,e,h*), and 1 (*c,f,i*). From case C1 with $\Delta = \Delta_m/2$.
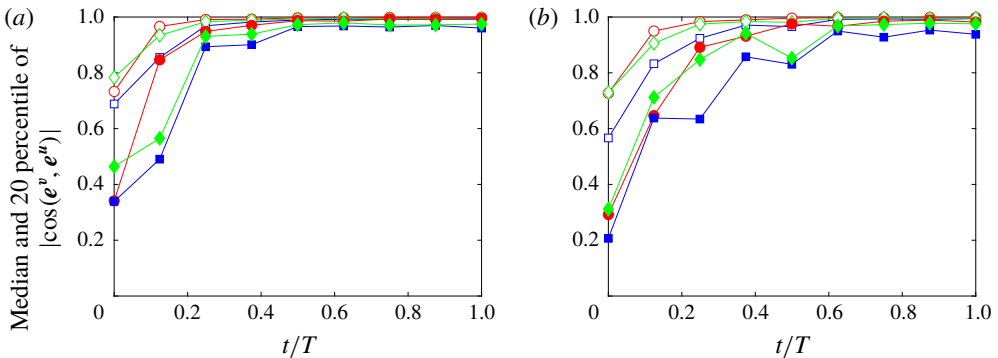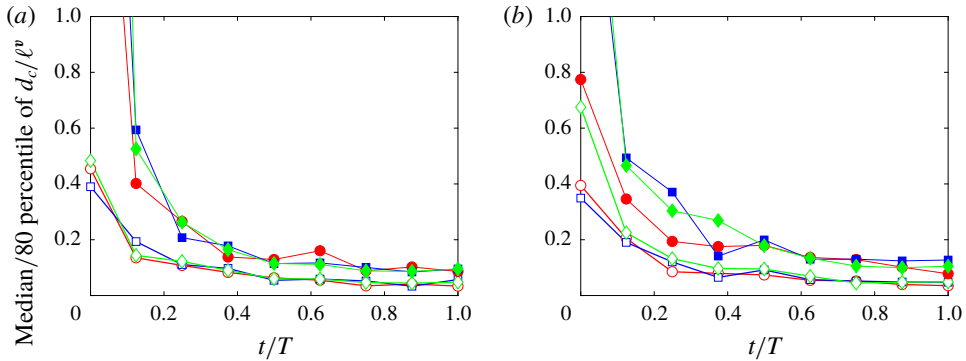

FIGURE 26. The medians (empty symbols) and the 20th percentiles (solid symbols) of $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)|$ ($i = \alpha, \beta, \gamma$) for the MVEEs of matching vortices. Circles: $i = \alpha$; squares: $i = \beta$; diamonds: $i = \gamma$. (*a*) $\Delta = \Delta_m/2$; (*b*) $\Delta = \Delta_m/4$. From case C1.

positive values and those with large negative values. The latter represent the regions where backscattering happens. These structures are defined by $\Pi_\Delta \geqslant 10\langle\Pi_\Delta\rangle$ and $\Pi_\Delta \leqslant -2.5\langle\Pi_\Delta\rangle$, and are referred to as P^+ and P^- structures, respectively.

FIGURE 27. The medians (empty symbols) and the 80th percentiles (solid symbols) for $d_{c,i}/\ell_i^v$ ($i = \alpha, \beta, \gamma$) for the MVEEs of matching vortices. Circles: $i = \alpha$; squares: $i = \beta$; diamonds: $i = \gamma$. (a) $\Delta = \Delta_m/2$; (b) $\Delta = \Delta_m/4$. From case C1.
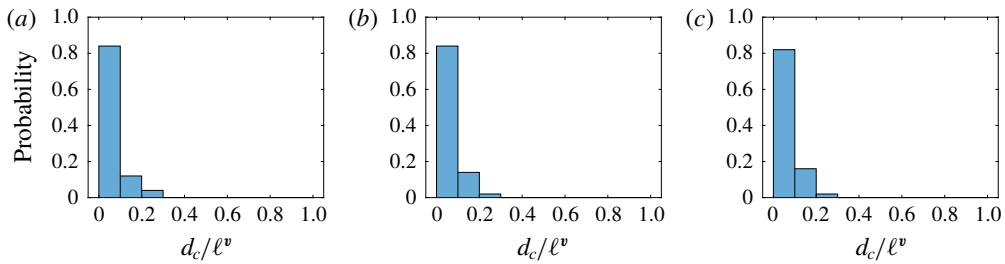

FIGURE 28. The normalized histograms of $d_{c,i}/\ell_i^v$ for $i = \alpha$ (a), $i = \beta$ (b) and $i = \gamma$ (c), from case C1 with $t/T = 1$ and $\Delta = \Delta_m/2$.
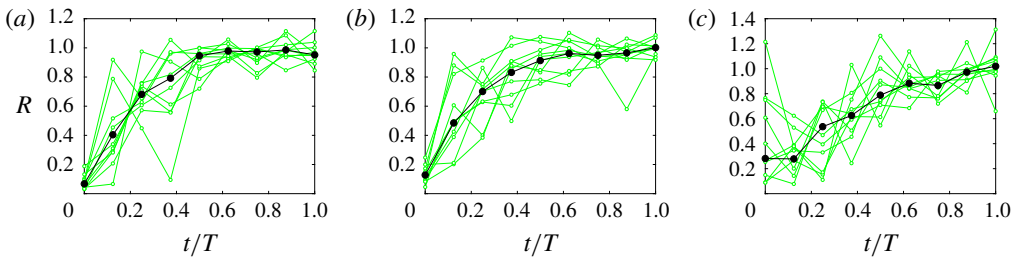

FIGURE 29. The median R of the ratio between the numbers of grid points in two matching structures (circles with thick line). The green thin lines show the medians in different size groups. (a) For straining structures; (b) for structures with large positive Π_Δ; (c) for structures with large negative Π_Δ. From case C1 with $\Delta = \Delta_m/4$.

Figure 29 shows the sizes of the reconstructed structures using the median ratio R. The reconstructed structures tend to be quite small initially, but the discrepancy decreases over time quickly, similar to the results for vortical structures in figure 22. At $t/T = 1$, all curves level off at approximately 1. The sizes for the P^- structures need a longer time to improve.
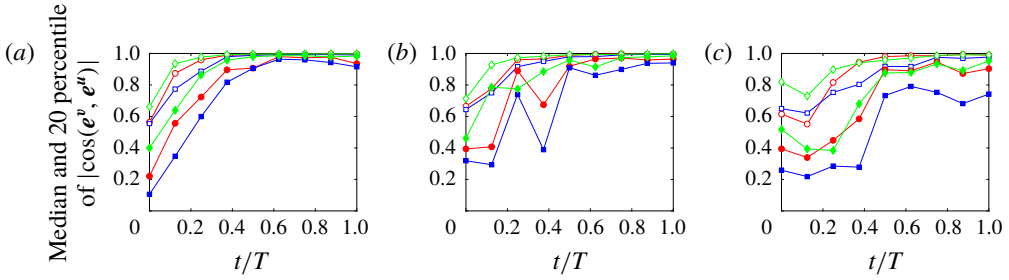
FIGURE 30. The medians (empty symbols) and the 20th percentiles (solid symbols) of $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)|$ for straining structures (*a*), structures with large positive Π_Δ (*b*) and structures with large negative Π_Δ (*c*). Circles: $i = \alpha$; squares: $i = \beta$; diamonds: $i = \gamma$. From case C1 with $\Delta = \Delta_m/4$.
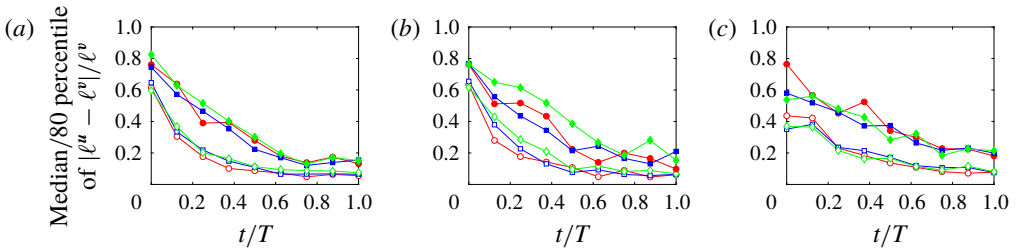

FIGURE 31. The medians (empty symbols) and the 80th percentiles (solid symbols) of $|\ell_i^v - \ell_i^u|/\ell_i^v$ for straining structures (*a*), structures with large positive Π_Δ (*b*) and structures with large negative Π_Δ (*c*). Circles: $i = \alpha$; squares: $i = \beta$; diamonds: $i = \gamma$. From case C1 with $\Delta = \Delta_m/4$.
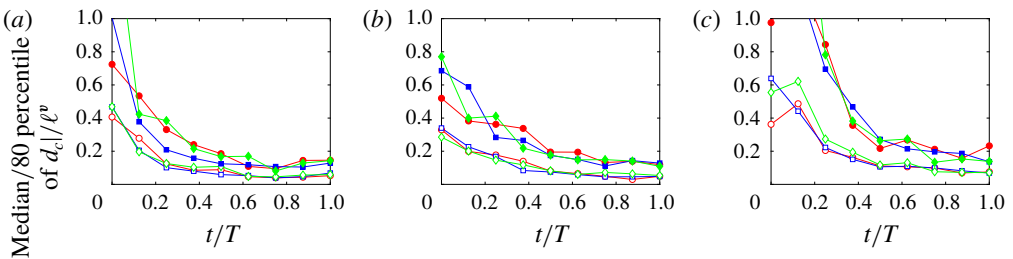

FIGURE 32. Same as figure 31, but for the relative displacement $d_{c,i}/\ell_i^v$.

The medians and selected percentiles for alignment cosine $|\cos(\boldsymbol{e}^v, \boldsymbol{e}^u)|$, relative difference in axis length $|\ell^u - \ell^v|/\ell^v$ and relative displacement d_c/ℓ^v are plotted in figures 30–32. For both straining and P^+ structures, excellent alignment is observed at $t/T = 1$, with the medians and the 20th percentile being mostly above 0.9 (figure 30). The medians and 80th percentiles for $|\ell^u - \ell^v|/\ell^v$ and d_c/ℓ^v are found to be around 5 % and 15 %, respectively, at $t/T = 1$, similar to those for the vortical structures (figures 31 and 32). For P^- structures, however, the results show stronger discrepancies between $\boldsymbol{v}$ and $\boldsymbol{u}$. For the alignment at $t/T = 1$, the medians still reach above 0.9. However, the 20th percentile result for the intermediate axis is only approximately 0.7 (cf. figure 30). The medians and 80th percentiles for $|\ell^u - \ell^v|/\ell^v$
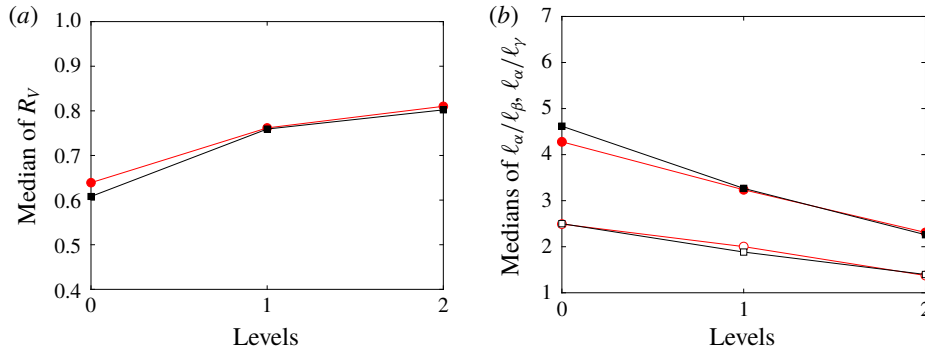
FIGURE 33. (*a*) Median of R_V, the ratio of the volume of a structure to the volume of its convex hull at different levels. Circle: for $\boldsymbol{u}$; squares: for $\boldsymbol{v}$. (*b*) The median of ℓ_α/ℓ_γ (solid symbols) and ℓ_α/ℓ_β (empty symbols) at different levels. Circle: for $\boldsymbol{u}$; squares: for $\boldsymbol{v}$. For strongly non-convex vortical structures only. For case C1 with $t/T = 1$ and $\Delta = \Delta_m/2$.

and d_c/ℓ^v are found at 10 % and 20 %, respectively. These values are slightly higher than those for the other two structures (see figures 31 and 32).

3.7.3. *MVEE trees for vortical structures*

The analyses so far have examined the overall sizes, orientations and locations of the non-local structures as defined by their MVEEs. For strongly non-convex structures, the MVEEs do not provide full descriptions. These structures are further examined with the MVEE trees in this sub-section (cf. § 2.4). Note that, if two MVEEs are significantly different (e.g. seriously misaligned), the difference between their sub-MVEEs will only be bigger. Therefore, this analysis is meaningful mainly for matching structures whose MVEEs are in good agreement; it provides a more detailed comparison for these structures. We will limit our discussion to vortical structures, as the results for other structures are similar, and only the structures at $t = T$ for the baseline case C1 are investigated. The analysis is applied to the non-convex vortical structures with $R_V \leqslant 0.8$, where R_V is the ratio of the volume of a structure to the volume of its convex hull (cf. § 2.4). Five pairs of matching vortical structures are found in each realization, and 25 pairs in total from 5 realizations. Two splittings are used to find the sub-MVEEs (cf. § 2.4) so that 3 levels of sub-MVEEs are calculated: levels 0, 1 and 2, which will be referred to as the L0, L1 and L2 sub-MVEEs; L0 sub-MVEEs are the same as the MVEEs already presented.

The geometries of the sub-structures at different levels are different. Figure 33 displays the difference in two aspects. The median of R_V (panel *a*) shows that R_V is larger at higher levels. Therefore, the sub-structures at higher levels are closer to their convex hulls. Panel (*b*) shows the median of the ratio of the lengths of the axes for the sub-MVEEs. The ratio is smaller for a higher level. The implication is that the sub-structures are rounder than their parent structures.

The alignment results for the sub-MVEEs are given in figure 34. The five curves in the figure (the spaces between the curves are filled with different shades of grey) are the 10th, 20th, 30th, 40th and 50th percentiles. The values for the L0 sub-MVEEs are close to those in figure 26, although they are not exactly the same since, here, only strongly non-convex structures are considered. For the major axes (shown in (*a*)), the alignment becomes weaker at higher levels, since a given percentile decreases
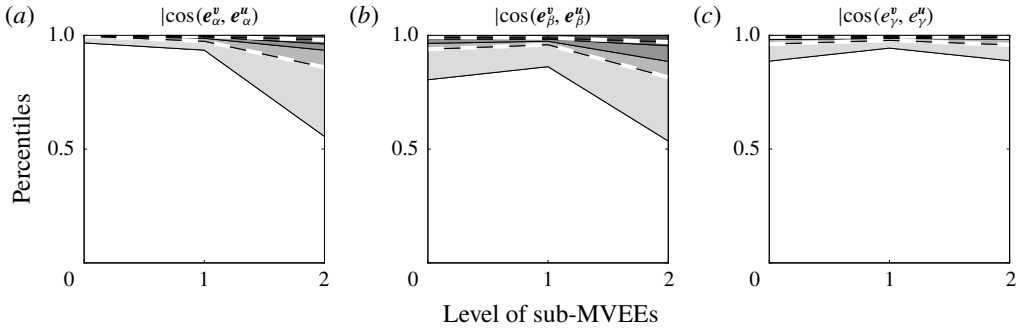
FIGURE 34. The percentiles for $|\cos(\boldsymbol{e}_i^v, \boldsymbol{e}_i^u)|$ $(i = \alpha, \beta, \gamma)$ for MVEEs of matching vortices and their sub-MVEEs. From bottom: the 10th, 20th, 30th, 40th and 50th percentiles. The dashed lines highlight the 20th and 50th percentiles. From case C1 with $t/T = 1$ and $\Delta = \Delta_m/2$.

as the level number increases. For example, the 20th percentile is nearly 1 at L0 and approximately 0.85 at L2. This means that 80 % matching L0 sub-MVEEs align almost perfectly, while we are only able to say the major axes of 80 % matching L2 sub-MVEEs are aligned to within 32°, corresponding to $|\cos(\boldsymbol{e}_\alpha^u, \boldsymbol{e}_\alpha^v)| \geqslant 0.85$. The 50th percentile for L2 is also nearly 1, indicating a near perfect alignment is still observed for more than a half of matching sub-MVEEs. On the other hand, the 10th percentile drops quite significantly at L2. This shows that, although the MVEEs for these matching structures indicate very good agreement in their overall shapes, the compositional parts of these structures do differ quite substantially. This happens for approximately 10 % of the matching structures. For the intermediate (panel *b*) and minor axes (panel *c*), similar behaviours can be observed, although the variation with respect to the levels is somewhat more complicated. However, even at the highest level (L2), the 20th percentile is still almost 1 for the minor axes, and 0.8 for the intermediate axes. Thus, very good alignment is observed for more than 80 % of samples.

There are several competing factors that affect the alignment of the sub-MVEEs. The higher level sub-structures are smaller and are closer to being convex (cf. figure 33*a*). As such, the orientations of the sub-MVEEs are less sensitive to the details of the shapes, so that it is easier to obtain better alignment. On the other hand, the misalignment has a cumulative effect. That is, the misalignment between the sub-MVEEs at a lower level naturally affects the alignment between those at higher levels. Furthermore, higher level structures tend to be rounder (cf. figure 33*b*). This too increases the chance of misalignment at higher levels, because small perturbations to a sphere could lead to ellipsoids with very different axis directions. The competition of these mechanisms may lead to the observed non-monotonic behaviour. For the same reason, misalignment between the MVEEs does not necessarily imply a large deviation in the shapes of the structures.

The relative displacement of the sub-MVEEs is presented in figure 35, which shows higher percentiles increase quicker with the level numbers (see, e.g. the 80th percentiles). This is likely due to the cumulative effect of displacement. The curves for smaller percentiles (which correspond to cases with small relative displacements) are rather flat. Therefore, if the matching MVEEs are located close to each other, so are their sub-MVEEs. Quantitatively, the 50th percentiles at L2 are approximately
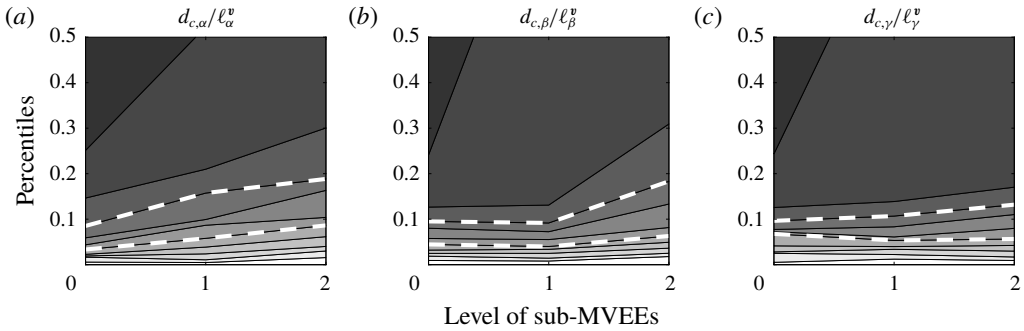
FIGURE 35. The percentiles for $d_{c,i}/\ell_i^v$ ($i = \alpha, \beta, \gamma$) for MVEEs of matching vortices and their sub-MVEEs. From bottom: the 10th to 100th percentiles with increments of 10 %. The dashed lines highlight the 50th and 80th percentiles. For case C1 with $t/T = 1$ and $\Delta = \Delta_m/2$.
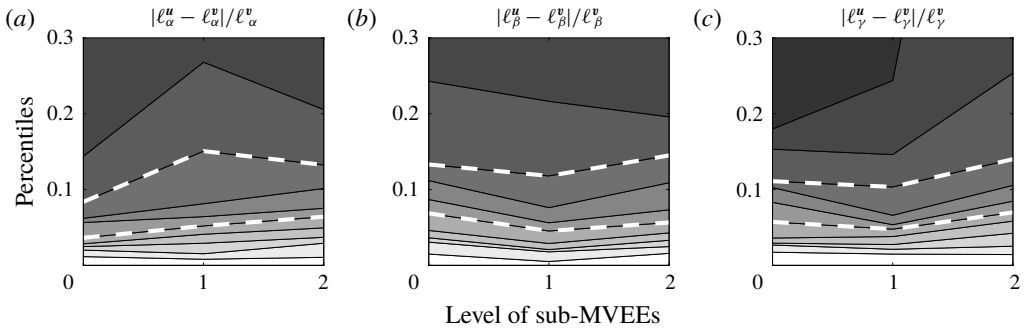

FIGURE 36. Same as figure 35 but for the percentiles of $|\ell_i^v - \ell_i^u|/\ell_i^v$ ($i = \alpha, \beta, \gamma$).

0.08, 0.07 and 0.06 for the major, intermediate and minor axes, respectively. Thus, in approximately 50 % of cases, the relative displacement is less than 8 %, 7 % and 6 % in the major, intermediate and minor axis directions, respectively. Figure 36 plots the relative difference in the lengths of the axes. For this quantity, the relative differences at L2 and L0 are not much different; the median value is found to be approximately 5 %. It is probably not surprising since there is no cumulative effect for the lengths of the axes.

To summarize, we observe that, if the matching MVEEs are already close to each other, their sub-MVEEs also show good agreement in orientations, sizes and locations. Therefore, the description provided by the MVEEs might already be enough in some applications.

4. Discussion and conclusions

The four-dimensional variational method has been used to reconstruct the small scales of a sequence of 3-D Kolmogorov turbulent velocity fields with a moderate Reynolds number. Velocity on a coarse grid from a time sequence of direct numerical simulation velocity fields is used as the measurement data. We focus on the reconstruction of instantaneous distributions of the small scales, and present detailed assessment at scales one and two octaves smaller than the measurement grid. In

terms of Fourier modes, the number of modes contained down to these scales are eight and sixty-four times of those in the measurement grid, respectively. The filtered vorticity $\tilde{\omega}_i$, filtered strain rate tensor $\tilde{s}_{ij}$, the vortex stretching term $\tilde{s}_{ij}\tilde{\omega}_i\tilde{\omega}_j$ and the subgrid-scale energy dissipation $P_\Delta = -\tau_{ij}\tilde{s}_{ij}$ are used to characterize the small scales. In order to quantitatively assess the reconstruction of non-local structures, minimum volume enclosing ellipsoids and MVEE trees are introduced. A method to apply these objects in the analysis is proposed and applied to the 4DVAR data.

Different parameters are tested, including the resolution of the measurement data k_m, the tolerance for the convergence test e_{tot} and the optimization horizon T; k_m has the most significant effects on the reconstruction. In reference to the threshold value $k_c = 0.2\eta_K^{-1}$ found previously in the literature with a direct substitution method, the reconstruction failed when k_m is significantly smaller than k_c. Accurate reconstruction is found for k_m slightly smaller than k_c. Although more research is needed to prove that the threshold value is the same in 4DVAR, these results demonstrate that reconstruction from a finite time sequence of data using 4DVAR can be accomplished for k_m at or above the threshold value k_c.

The reconstructions at the initial time $t = 0$ and at the end of the optimization horizon $t = T$ are examined separately for the successful cases. Larger T and smaller e_{tot} improve the reconstruction at $t = T$. For T equal to the large eddy turnover time scale, accurate reconstruction at scales at least two octaves smaller than the measurement grids is obtained for a range of quantities. The main findings are summarized as follows, where approximate numbers are quoted to give an estimate of the efficacy of the method:

(i) The reconstruction achieves higher than 95 % pointwise correlation for the filtered enstrophy and the other parameters. The spectral correlation for velocity is higher than 80 % across the spectrum. The normalized r.m.s. pointwise difference and the normalized spectral difference increase towards the small scales, and reaches approximately 30 % at the Kolmogorov scale.

(ii) The non-local structures are reconstructed accurately. For most quantities, the misalignment is less than 15° for the majority of the samples, the sizes differing within ±10 %, and the locations displaced by less than 15 % of the lengths of the axes. The structures with strong negative SGS energy dissipation have slightly larger errors.

(iii) A direct substitution scheme provides much poorer reconstruction, where the r.m.s. pointwise difference is approximately three times as big for the filtered enstrophy, and is almost ten times as big in terms of the spectral difference of the velocity fields. Meanwhile, Lagrangian interpolation is simply not feasible.

The reconstruction at $t = 0$ can capture the qualitative features of the small scales in turbulence, including a qualitatively correct energy spectrum and mean SGS energy dissipation. However, significant discrepancy exists and the agreement is improved very little by tuning the optimization horizon or the tolerance. A plausible explanation is that our optimal solutions have not truly converged. Due to the high computational costs, we have not been able to further improve the solution by using even smaller e_{tot}. Given the chaotic nature of the flow, it is challenging to obtain convergent results if T is significantly larger or e_{tot} is much smaller.

This study confirms the efficacy of the 4DVAR scheme, and demonstrates that MVEEs and MVEE trees can provide useful information about non-local structures for computational fluid dynamic simulations with data assimilation when the prediction of the instantaneous velocity field is one of the main interests. As the application of data

assimilation to computational fluid dynamics is still in an early stage, many questions remain open. Most importantly, the Reynolds number for the flow considered here is relatively low. The natural next step is to investigate the quality of reconstruction when coarse-grained models (such as large eddy simulations) are used to model the data. This will allow us to investigate flows with higher Reynolds numbers. To obtain better converged results, a better understanding of the behaviour of the adjoint fields is useful. These questions are the focus of our ongoing research.

Acknowledgements

Declaration of interests

The authors report no conflict of interest.

REFERENCES

ARTANA, G., CAMMILLERI, A., CARLIER, J. & MEMIN, E. 2012 Strong and weak constraint variational assimilations for reduced order fluid flow modeling. *J. Comput. Phys.* **231**, 3264–3288.

BERMEJO-MORENO, I. & PULLIN, D. I. 2008 On the non-local geometry of turbulence. *J. Fluid Mech.* **603**, 101–135.

BEWLEY, T. R. & PROTAS, B. 2004 Skin friction and pressure: the 'footprints' of turbulence. *Physica* D **196**, 28–44.

BORUE, V. & ORSZAG, S. A. 1996 Numerical study of three-dimensional Kolmogorov flow at high Reynolds numbers. *J. Fluid Mech.* **306**, 293–323.

BOYD, S. & VANDENBERGHE, L. 2004 *Convex Optimization*. Cambridge University Press.

BROWN, R. G. & HWANG, P. Y. C. 2012 *Introduction to Random Signals and Applied Kalman Filtering: with MATLAB Exercises*. John Wiley & Sons.

CANTWELL, B. J. 1992 Exact solution of a restricted Euler equation for the velocity gradient tensor. *Phys. Fluids* A **4**, 782–793.

CHEVALIER, M., HOEPFFNER, J., BEWLEY, T. R. & HENNINGSON, D. S. 2006 State estimation in wall-bounded flow systems. Part 2. Turbulent flows. *J. Fluid Mech.* **552**, 167–187.

COLBURN, C. H., CESSNA, J. B. & BEWLEY, T. R. 2011 State estimation in wall-bounded flow systems. Part 3. The ensemble Kalman filter. *J. Fluid Mech.* **682**, 289–303.

COMBES, B., HEITZ, D., GUIBERT, A. & MEMIN, E. 2015 A particle filter to reconstruct a free-surface flow from a depth camera. *Fluid Dyn. Res.* **47**, 051404.

COURTIER, P., DERBER, J., ERRICO, R., LOUIS, J.-F. & VUKICEVIC, T. 1993 Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus* **45A**, 342–357.

D'ADAMO, J., PAPADAKIS, N., MEMIN, E. & ARTANA, G. 2007 Variational assimilation of pod low-order dynamical systems. *J. Turbul.* **8**, N9.

ESTER, M., KRIEGEL, H.-P., SANDER, J. & XU, X. 1996 A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proc.* **1**, 226–231.

EVENSEN, G. 2009 The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst. Magazine* **29**, 83–104.

FOURES, D. P. G., DOVETTA, N., SIPP, D. & SCHMID, P. J. 2014 A data-assimilation method for Reynolds-averaged Navier–Stokes-driven mean flow reconstruction. *J. Fluid Mech.* **759**, 404–431.

GRONSKIS, A., HEITZ, D. & MEMIN, E. 2013 Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation. *J. Comput. Phys.* **242**, 480–497.

HAYASE, T. 2015 Numerical simulation of real-world flows. *Fluid Dyn. Res.* **47**, 051201.

HEITZ, D., MEMIN, E. & SCHNORR, C. 2010 Variational fluid flow measurements from image sequences: synopsis and perspectives. *Exp. Fluids* **48**, 369–393.

HORIUTI, K. 2003 Roles of non-aligned eigenvectors of strain-rate and subgrid-scale stress tensors in turbulence generation. *J. Fluid Mech.* **491**, 65–100.

KALMAN, R. E. 1960 A new approach to linear filtering and prediction problems. *J. Basic Engng* **82** (1), 35–45.

KALNAY, E. 2003 *Atmospheric Modelling, Data Assimilation and Predictability*. Cambridge University Press.

KANG, H. S. & MENEVEAU, C. 2005 Effect of large-scale coherent structures on subgrid-scale stress and strain-rate eigenvector alignments in turbulent shear flow. *Phys. Fluids* **17**, 055103.

KATO, H. & OBAYASHI, S. 2013 Approach for uncertainty of turbulence modeling based on data assimilation technique. *Comput. Fluids* **85**, 2–7.

KATO, H., YOSHIZAWA, A., UENO, C. & OBAYASHI, S. 2015 A data assimilation methodology for reconstructing turbulent flows around aircraft. *J. Comput. Phys.* **283**, 559–581.

KHACHIYAN, L. G. 1996 Rounding of polytopes in the real number model of computation. *Math. Oper. Res.* **21**, 307–320.

LALESCU, C. C., MENEVEAU, C. & EYINK, G. L. 2013 Synchronization of chaos in fully developed turbulence. *Phys. Rev. Lett.* **110**, 084102.

LEUNG, T., SWAMINATHAN, N. & DAVIDSON, P. A. 2012 Geometry and interaction of structures in homogeneous isotropic turbulence. *J. Fluid Mech.* **710**, 453–481.

LI, Z., ZHANG, H., BAILEY, S. C. C., HOAGG, J. B. & MARTIN, A. 2017 A data-driven adaptive Reynolds-averaged Navier–Stokes k–ω model for turbulent flow. *J. Comput. Phys.* **345**, 111–131.

LIONS, J. L. 1971 *Optimal Control of Systems Governed by Partial Differential Equations*. Springer.

MELDI, M. & POUX, A. 2017 A reduced order model based on Kalman filtering for sequential data assimilation of turbulent flows. *J. Comput. Phys.* **347**, 207–234.

MENEVEAU, C. & KATZ, J. 2000 Scale-invariance and turbulence models for large-eddy simulation. *Annu. Rev. Fluid Mech.* **32**, 1–32.

MONS, V., CHASSAING, J.-C., GOMEZ, T. & SAGAUT, P. 2014 Is isotropic turbulence decay governed by asymptotic behavior of large scales? An eddy-damped quasi-normal Markvian-based data assimilation study. *Phys. Fluids* **26**, 115105.

MONS, V., CHASSAING, J.-C., GOMEZ, T. & SAGAUT, P. 2016 Reconstruction of unsteady viscous flows using data assimilation schemes. *J. Comput. Phys.* **316**, 255–280.

MONS, V., CHASSAING, J.-C. & SAGAUT, P. 2017 Optimal sensor placement for variational data assimilation of unsteady flows past a rotationally oscillating cylinder. *J. Fluid Mech.* **823**, 230–277.

MOSHTAGH, N. 2009 Minimum volume enclosing ellipsoid. *MATLAB Central File Exchange. Retrieved June 10, 2018*.

NOCEDAL, J. & WRIGHT, S. J. 1999 *Numerical Optimization*. Springer.

POPE, S. B. 2000 *Turbulent Flows*. Cambridge University Press.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. 1992 *Numerical Recipe in Fortran 77: The Art of Scientific Computing*. Cambridge Universtiy Press.

PROTAS, B., NOACK, B. R. & OSTH, J. 2015 Optimal nonlinear eddy viscosity in Galerkin models of turbulent flows. *J. Fluid Mech.* **766**, 337–367.

SAGAUT, P. 2002 *Large Eddy Simulation for Incompressible Flows: An Introduction*, 2nd edn. Springer.

SCHUBERT, E., SANDER, J., ESTER, M., KRIEGEL, H. P. & XU, X. 2017 Dbscan revisited: why and how you should (still) use dbscan. *ACM Trans. Database Syst.* **42** (3), 19:1–19:21.

TAO, B., KATZ, J. & MENEVEAU, C. 2002 Statisticl geometry of subgrid-scale stresses determined from holographic particle image velocimetry measurements. *J. Fluid Mech.* **457**, 35–78.

TODDS, M. J. 2016 *Minimum-Volume Ellipsoids: Theory and Algorithms*. SIAM.

VIEILLEFOSSE, P. 1984 Internal motion of a small element of fluid in an inviscid flow. *Physica* A **125**, 150–162.

WANG, Q., MOIN, P. & IACCARINO, G. 2009 Minimal repetition dynamic checkpointing algorithm for unsteady adjoint calculation. *SIAM J. Sci. Comput.* **31**, 2549–2567.

YANG, Y. & PULLIN, D. I. 2011 Geometric study of Eulerian and Lagrangian structures in turbulent channel flows. *J. Fluid Mech.* **674**, 67–92.

YOSHIDA, K., YAMAGUCHI, J. & KANEDA, Y. 2005 Regeneration of small eddies by data assimilation in turbulence. *Phys. Rev. Lett.* **94**, 014501.