



Four-dimensional variational data assimilation of a turbulent jet for super-temporal-resolution reconstruction

Chuangxin He^{1,2}, Xin Zeng^{1,2}, Peng Wang^{1,2}, Xin Wen^{1,2} and Yingzheng Liu^{1,2,†}

¹Key Laboratory of Education Ministry for Power Machinery and Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, PR China

²Gas Turbine Research Institute, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, PR China

(Received 8 April 2023; revised 11 November 2023; accepted 11 November 2023)

The super-temporal-resolution (STR) reconstruction of turbulent flows is an important data augmentation application for increasing the data reach in measurement techniques and understanding turbulence dynamics. This paper proposes a data assimilation (DA) strategy based on weak-constraint four-dimensional variation to conduct an STR reconstruction in a turbulent jet beyond the Nyquist limit from given low-sampling-rate observations. Highly resolved large-eddy simulation (LES) data are used to produce synthetic measurements, which are used as observations and for validation. A segregated assimilation procedure is realised to assimilate the initial condition, inflow boundary condition and model error separately. Different types of observational data are tested. The first type is down-sampled LES data containing many small-scale turbulence structures with or without synthetic noise. The DA results show that the temporal variation of the small-scale structures is well recovered even with noise in the observations. The spectra are resolved to a frequency approximately one order of magnitude higher than what can be captured within the Nyquist limit. The second type of observation is low-sampling-rate tomographic particle image velocimetry (tomo-PIV) data with or without the injection of small-scale structures. The modulation between the large-scale structures contained in the tomo-PIV fields and the small scales injected from the observations is improved. The resultant small scales in the STR reconstruction have the characteristics of authentic turbulence to a considerable extent. Additionally, DA yields much smaller errors in the prediction of particle positions when compared with the Wiener filter, demonstrating the great potential for Lagrangian particle tracking in measurement techniques.

Key words: variational methods, jets, computational methods

† Email address for correspondence: yzliu@sjtu.edu.cn

1. Introduction

The high-frequency sampling of flow data has increasingly drawn attention in the turbulence research community as it has become an important means for understanding the underlying fluid dynamics within the turbulence kinetic energy cascade. However, there remains a great challenge in developing hardware for acquiring highly time-resolved turbulence data, compromising the balancing of the spatial resolution and temporal sampling rate. Although data-driven reconstruction techniques, such as linear stochastic estimation (LSE) and machine learning (ML), have brought new impetus to super-temporal-resolution (STR) reconstruction using time-sparse spatially resolved flow fields, the resultant evolution at small scales is commonly incorrect owing to the reduced-order reconstruction, the lack of physical constraint or insufficient temporal information in the training database. This leads to high demand for the development of model-assisted data-driven techniques such as data assimilation (DA).

As one of the earliest approaches for STR reconstruction, LSE (Adrian & Moin 1988) attempts to integrate the advantages of a high spatial resolution in field-based measurements and a high sampling rate of probe signals. It has been established on the foundation of non-local flow properties and was widely used in early studies for time-resolved reduced-order reconstruction. Relatively inexpensive discrete probes provide high-sampling-rate measurement data while capturing not only the temporal information but also some degree of the spatial information in flow fields. Linear stochastic estimation approximates the conditional averages of the fields in terms of the measurement data and has been sequentially modified to couple with proper orthogonal decomposition (POD) in the estimation of the time series of mode amplitudes used to reconstruct reduced-order flow fields (Hudy, Naguib & Humphreys 2007; Tinney, Ukeiley & Glauser 2008). However, the implementation of LSE is mainly suited to the reduced-order reconstruction of flows having periodic or quasi-periodic behaviours, and the resultant realisations are built in preference to energetic large-scale flow structures represented by several leading POD modes. Moreover, LSE suffers from the problem of overfitting through the amplification of singular eigenvalues of small amplitude (Podvin *et al.* 2018). Although such a process has been improved using the multi-time-delay approach (Durgesh & Naughton 2010; Kerherve, Roux & Mathis 2017), the Kalman smoother (Tu *et al.* 2013), a new variant of the formulation (Podvin *et al.* 2018) and multichannel singular spectrum analysis (Hosseini, Martinuzzi & Noack 2015), there remains a connatural limitation when there are far fewer probes than degrees of freedom of the flow fields. This limits the number of POD modes that can be used in the LSE reconstruction and induces error in amplitude determination. The limitation becomes important when the flow is broad-band and little kinetic energy is contained in the leading POD modes.

Due to the rapid development of computer science in recent years, ML techniques have been widely used in fluid dynamic research including flow and heat transfer prediction (Lee & You 2019; Kim & Lee 2020), turbulence modelling (Duraisamy, Iaccarino & Xiao 2019; Sirignano & MacArt 2023) and active control (Park & Choi 2020; Pino *et al.* 2023). Comprehensive reviews on ML-augmented fluid mechanics were presented by Brenner, Eldredge & Freund (2019) and Brunton, Noack & Koumoutsakos (2019). As a representative application using ML techniques, super-resolution reconstruction can be achieved by constructing a convolutional neural network that directly maps the low-resolution fields to the high-resolution fields once trained using both a low- and a high-resolution database (Fukami, Fukagata & Taira 2019; Liu *et al.* 2020). In addition, high-sampling-rate training data can be used for STR reconstruction (Fukami, Fukagata & Taira 2021). While these supervised deep-learning models require labelled low- and

high-resolution data pairs for training, unpaired data are more practical to use as noted by Kim *et al.* (2021), who proposed an unsupervised deep-learning model for super-resolution reconstruction adopting a cyclic-consistent generative adversarial network. However, purely ML methods often yield non-physical features when the resolution ratio between the target and input fields is large due to the lack of physical constraint in the data-driven optimisation process. The approach based on POD or extended POD, similar to that used in LSE, is usually adopted to avoid irregular prediction. It has been demonstrated that ML techniques account for nonlinearities in flows in the interpolation of the model coefficients and thus perform better in reconstructions using more POD modes than the conventional LSE approach. Recent work has demonstrated the advantages of ML approaches over LSE in reduced-order STR reconstruction (Deng *et al.* 2019; Guemes, Discetti & Ianiro 2019; Giannopoulos & Aider 2020; Jin *et al.* 2020). Although POD-based approaches enable us to recover the main dynamic patterns of the turbulence fluctuations in STR reconstruction, the high-order modes with temporally varying frequencies higher than the sampling rate pose a challenge. The aforementioned works largely aimed at understanding either large-scale motion in the flows or flow feedback control. Additionally, the approaches are only applicable to statistically stable flows as a large database is required in the training process. In the fields of computer vision and video processing, STR reconstruction is commonly known as frame interpolation (Niklaus, Mai & Liu 2017; Chi *et al.* 2020; Bao *et al.* 2021) and is widely used to artificially increase the frame rate of a video by creating fake frames containing sufficient image details between real frames. However, similar studies have rarely been conducted for turbulent flows with the recovery of small-scale structures having evolution frequencies much higher than the sampling rate.

Being different from the purely data-driven techniques, DA (Evensen, Vossepoel & Leeuwen 2022) integrates measurement data (observations) with physical or semi-physical equations (predictive models). Thus, DA avoids irregular motions of the reproduced vortical structures and substantially reduces the training database requirement, and it is undoubtedly an important approach to increasing the data reach and augmenting methodology interpretability. Data assimilation has been applied in fluid mechanics for acoustic state and model parameter predictions using the Bayesian ensemble method (Nóvoa & Magri 2022), ocean wave forecasting using the high-order spectral method coupled with the ensemble Kalman filter (Wang & Pan 2021) and data-driven numerical simulation using a nudging strategy (Zauner *et al.* 2022). Among various DA algorithms, variational DA benefits from the driving of adjoint equations and is able to determine the optimal field with extremely large dimensions and a lower data requirement. Variational DA includes the three-dimensional type for steady-state processes (Foures *et al.* 2014; Mons & Marquet 2021) and four-dimensional type (4D-Var) for unsteady-state prediction (Chandramouli, Memin & Heitz 2020). In addition, variational DA can be implemented in either discrete or continuous form. The former implementation discretises the system before the derivation of the adjoint equation, which has a large memory requirement for expensive matrix computation (Papoutsis-Kiachagias & Giannakoglou 2016), and the latter implementation is thus preferred for complex flow configurations (Foures *et al.* 2014; He *et al.* 2018a; Li *et al.* 2019; Chandramouli *et al.* 2020; He, Wang & Liu 2021). We focus on 4D-Var as the unsteady state is the current topic of interest. Four-dimensional variation is a generalisation of three-dimensional variation that handles observations distributed in time. It thus optimises the constraint problem via an objective function presenting the deviation of the model forecast from the observations in time integration. The classical strong-constraint 4D-Var (Evensen *et al.* 2022; Wang, Wang & Zaki 2022) seeks an initial condition such that the forecast best fits the observations within the assimilation interval,

under the assumption that a precise predictive model entirely determines the true state once initialised. Its applications in fluid mechanics include the determination of the inflow and initial conditions for direct numerical simulation (Gronskis, Heitz & Mémin 2013), the de-noising of particle image velocimetry (PIV) data (Gillissen, Bouffanais & Yue 2019) and the reconstruction of small-scale structures in Kolmogorov flows (Li *et al.* 2019). However, in the case of turbulent flows, 4D-Var is prone to be a weak constraint (Bennett 2002; Tremolet 2007) owing to the error in the predictive models induced by either the subgrid stress or the inappropriate boundary conditions. Weak-constraint 4D-Var has not yet been extensively considered for turbulent flows, except in a few studies involving the simple example of the reactive–diffusive equation (Vidard *et al.* 2000), the development of POD-based reduced-order models in turbulence dynamics (Artana *et al.* 2012; Stefanescu, Sandu & Navon 2015) and stochastic model-error assimilation in a turbulent wake (Chandramouli *et al.* 2020). In addition, STR reconstruction using weak-constraint 4D-Var is suited to turbulence research, yet this has not been considered in previous work.

This paper concentrates on STR reconstruction beyond the Nyquist limit with the recovery of vortex temporal evolution with frequencies much higher than the sampling rate of the measurements. The only observations are the two flow realisations at the start and end instants of the assimilation window, without any training database. A segregated weak-constraint 4D-Var DA procedure is proposed to determine the initial condition, the inflow boundary condition and the model error term separately. In the comprehensive verification of the DA approach for the reconstruction of intermediate instantaneous fields, time-sparse large-eddy simulation (LES) data of a turbulent round jet containing many small-scale structures are used to produce observations. Additionally, tomographic PIV (tomo-PIV) observations generated from synthetic particle images are used in the DA approach, with the discussion largely focusing on the authenticity and enhancement of dynamical features within the evolution of the injected small-scale structures. Furthermore, the accuracy of Lagrangian particle prediction based on the reconstructed STR flows is evaluated for potential application in particle tracking velocimetry (PTV).

2. The DA fundamentals

2.1. General framework

The objective of the present DA is to reconstruct detailed turbulent flow fields between two given three-dimensional realisations with a large time interval. The flow is governed by the incompressible Navier–Stokes (NS) equations subject to initial and boundary conditions. When adopting the DA process in practice, the system predictive (physical or primary) model is subject to errors due to the initial condition, turbulent subgrid stress and boundary condition effects, leading to the weak-constraint problem when using 4D-Var. In the present work, the velocity \mathbf{u} and kinematic pressure p (the static pressure divided by the fluid density) constitute the state variables in the DA system. The predictive model reads

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p - \nu \Delta \mathbf{u} - \boldsymbol{\xi} = 0, \quad (2.1)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.2)$$

where ξ is the model error varying in space and time. The objective function of the constrained Lagrangian optimisation problem is expressed as

$$\begin{aligned} \mathcal{L}(\mathbf{u}, p, \xi, s) = & \int_{t_0}^{t_0+\Delta T} [\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})]^T C_{\epsilon\epsilon}^{-1} [\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})] \delta(t - t^{obs}) dt \\ & + \int_{t_0}^{t_0+\Delta T} [\mathbf{v}, q]^T \mathcal{R} dt + s^{*T} C_{\eta\eta}^{-1} [\mathbf{u}(t_0) - s] + \int_{t_0}^{t_0+\Delta T} \frac{\alpha}{2} \xi^T C_{\xi\xi}^{-1} \xi dt. \end{aligned} \tag{2.3}$$

In (2.3), \mathbf{u}^{obs} is the observation; \mathbb{H} represents the interpolation from the observational grid to the DA grid; $C_{\xi\xi}$, $C_{\eta\eta}$ and $C_{\epsilon\epsilon}$ are the error covariances, which affect the convergence speed of the computation and are cumbersome to construct as noted by Chandramouli *et al.* (2020); δ is the Dirac delta function; and t^{obs} is the instant when the observations are available. This formulation enables the treatment of 4D-Var problems when the time interval of the observations is much longer than the time step of the DA computations. Here \mathcal{R} denotes the governing equations as given in (2.1) and (2.2). The variables of the Lagrangian multiplier $[\mathbf{v}, q]$ are referred to as adjoint variables in the following sections. Terms s and s^* are the initial condition to be assimilated and its adjoint variable, respectively. The last term is the regularisation term which avoids ill-conditioned results of the system and removes noise, with a user-specified coefficient α . Length ΔT is the temporal length of the DA window in which the assimilation is performed with the start time t_0 . Accordingly, the initial condition s , model error ξ and inflow boundary condition can be assimilated using the adjoint formulations.

2.2. Continuous adjoint formulations

Continuous adjoint formulations are used to solve the large-scale optimisation problem in the present DA work. The error covariances can be interpreted as the relative importance of the observational data in the minimising procedure and are set to unity in this study, which means that the observational data at different spatial locations are equally important. This simplification is found to have no appreciable effect on the computational convergence or the DA results in this study. The objective function is thus expressed in a spatial integral as

$$\begin{aligned} \mathcal{L}(\mathbf{u}, p, \xi, s) = & \int_{t_0}^{t_0+\Delta T} \int_{\Omega} [\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})]^2 \delta(t - t^{obs}) dx dt + \int_{t_0}^{t_0+\Delta T} \int_{\Omega} (\mathbf{v}, q) \mathcal{R} dx dt \\ & + \int_{\Omega} s^{*T} [\mathbf{u}(t_0) - s] dx + \frac{\alpha}{2} \int_{t_0}^{t_0+\Delta T} \int_{\Omega} \xi^2 dx dt. \end{aligned} \tag{2.4}$$

Data assimilation looks for appropriate ξ and s that minimise \mathcal{L} , giving rise to the extremum problem of the total variation:

$$\delta \mathcal{L}(\mathbf{u}, p, \xi, s) = \delta_{\mathbf{u}} \mathcal{L} + \delta_p \mathcal{L} + \delta_{\xi} \mathcal{L} + \delta_s \mathcal{L} = 0. \tag{2.5}$$

This problem can be simplified by choosing appropriate adjoint variables \mathbf{v} and q that reduce the variation with respect to the state variables:

$$\delta_{\mathbf{u}} \mathcal{L} + \delta_p \mathcal{L} = 0. \tag{2.6}$$

The adjoint NS equations are thus derived from the spatial integral of (2.6) according to He, Liu & Gan (2020):

$$-\frac{\partial \mathbf{v}}{\partial t} - (\mathbf{u} \cdot \nabla) \mathbf{v} - \nu \Delta \mathbf{v} = -\nabla \mathbf{u} \cdot \mathbf{v} - \nabla q - \frac{2\gamma}{U_0^2} [\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})] \delta(t - t^{obs}), \quad (2.7)$$

$$\nabla \cdot \mathbf{v} = 0. \quad (2.8)$$

The last term in (2.7) is the source for adjoint flow production (i.e. adjoint source) stemming from the difference between prediction and observation. The adjoint flow approaches zero with the convergence of the iterations. A referential velocity U_0 is introduced in the source term for normalisation; it is also important for the scaling of the adjoint velocity amplitudes to retain a low adjoint Courant–Friedrichs–Lewy (CFL) number using the same time step as that used in solving the primary equations. Term γ is the dimension converter of value unity, which addresses the dimensional inconsistency. The tensor form of the adjoint transpose convection term is given for clarification as

$$\nabla \mathbf{u} \cdot \mathbf{v} = v_j \frac{\partial u_j}{\partial x_i}. \quad (2.9)$$

The adjoint boundary conditions can be derived from the surface integral of (2.6) by deduction (He *et al.* 2018) and are reproduced here. On the inflow, wall and far-field boundaries where the primary state variable \mathbf{u} is specified, the boundary conditions for the adjoint velocity \mathbf{v} are

$$v_\tau = 0, \quad v_n = 0, \quad (2.10a,b)$$

$$\mathbf{n} \cdot \nabla q = 0. \quad (2.11)$$

For the outflow boundaries, where the zero-gradient condition is used for the primary state variables \mathbf{u} , the adjoint velocity conditions are

$$u_n \cdot v_\tau + \nu(\mathbf{n} \cdot \nabla) u_\tau = 0, \quad (2.12)$$

$$u_n \cdot v_n + \nu(\mathbf{n} \cdot \nabla) u_n - q = 0. \quad (2.13)$$

Here, the subscripts n and τ denote the normal and tangential components of the variables, respectively. Vector \mathbf{n} is the unit normal vector at the boundaries. The outflow boundary conditions are usually simplified as zero-gradient conditions to improve the computational stability (He *et al.* 2018). According to Othmer (2008), the boundary conditions for the adjoint pressure q are the same as those for the primary pressure p , since q enters the adjoint NS equations in a manner similar to how p enters the primary equations. Practical considerations of all the boundary conditions in the present study are presented in § 2.3. The adjoint terminal and initial conditions are derived from the time-dependent term of (2.6) as

$$\mathbf{v}(t_0 + \Delta T) = 0, \quad (2.14)$$

$$\mathbf{s}^* = \mathbf{v}(t_0). \quad (2.15)$$

The total variation thus reduces to

$$\begin{aligned} \delta \mathcal{L} &= \delta_\xi \mathcal{L} + \delta_s \mathcal{L} = \int_{t_0}^{t_0 + \Delta T} \int_\Omega (\mathbf{v}, q) \delta_\xi \mathcal{R} \, dx \, dt + \alpha \int_{t_0}^{t_0 + \Delta T} \int_\Omega \xi \, dx \, dt - \int_\Omega \mathbf{s}^* \, dx \\ &= - \int_{t_0}^{t_0 + \Delta T} \int_\Omega \mathbf{v} \, dx \, dt + \alpha \int_{t_0}^{t_0 + \Delta T} \int_\Omega \xi \, dx \, dt - \int_\Omega \mathbf{s}^* \, dx. \end{aligned} \quad (2.16)$$

Therefore, the model error and the initial condition are solved iteratively according to

$$\xi = \xi - \lambda_\xi \frac{\partial \mathcal{L}}{\partial \xi} = (1 - \alpha \lambda_\xi) \xi + \lambda_\xi \mathbf{v}, \quad (2.17)$$

$$\mathbf{s} = \mathbf{s} - \lambda_s \frac{\partial \mathcal{L}}{\partial \mathbf{s}} = \mathbf{s} + \lambda_s \mathbf{s}^* = \mathbf{s} + \lambda_s \mathbf{v}(t_0), \quad (2.18)$$

in a finite-volume code using the steepest descent algorithm with an appropriate choice of the step lengths λ_ξ and λ_s (see § 2.3). Equation (2.17) shows that the necessary condition of α to ensure that the iteration procedure converges is $0 \leq \alpha \lambda_\xi \leq 1$. Parameter α can be set at $1/\lambda_\xi$ to remove noise, and should be smaller in cases of problematic instability or observations with a high signal-to-noise ratio.

The inflow boundary condition \mathbf{u}_{in} was assimilated in a manner similar to that for the initial condition in the work of Lemke (2015). However, considering the observations of two snapshots at the start and end instants ($\Delta T \gg \Delta t$, where Δt is the time step of the computation) of the assimilation window throughout the three-dimensional domain, this inflow assimilation strategy exhibits strong instability and high sensitivity to the step size of the steepest descent algorithm. Moreover, cross-talk exists in the DA procedure between the behaviour of the forcing and the initial and inflow conditions. This cross-talk results from the fact that the DA residual is still reduced by the updating of ξ even if incorrect initial and inflow conditions are used. Assimilating the model error and initial and inflow conditions all at once does not achieve the desired results even though the objective function has decreased to a minimum. A segregated approach is thus proposed to assimilate the desired quantities separately. Detailed algorithms are presented in Appendix A.

2.3. Numerical and practical considerations

As the computational domain considered in this study is a small cuboid, consistent with a typical tomo-PIV or 4D-PTV measurement, it is far from sufficient for conventional numerical simulations. The first aspect that should be carefully considered is the boundary conditions for both the primary and adjoint equations. Once the inflow field has been assimilated, the Dirichlet condition for the primary velocity \mathbf{u} can be applied. The convective velocity condition can be applied at the outflow boundary for \mathbf{u} . The Neumann condition for the primary pressure p is used for the inflow boundary; this also includes the correcting pressure δp in the algorithm of the semi-implicit method for pressure-linked equations (SIMPLE). To improve the computational stability, the Neumann condition is used on the outflow boundary for p , but the Dirichlet condition is used for δp . This gives rise to a relatively steady and spatially smooth pressure distribution near the outflow boundary, which avoids divergence in the pressure correction step in the SIMPLE algorithm. The free-slip condition for velocity and Neumann condition for pressure are applied on the side surfaces for both primary and adjoint equations. The above treatment indeed induces error in the primary flows, but this error is included in the model error ξ and is thus assimilated in the DA procedure. Equation (2.10) presents the no-slip wall condition for the adjoint velocity \mathbf{v} at the inflow boundary, which reverses the adjoint flows to the downstream direction. This condition is applicable for large DA domains including a considerable portion of a free-stream region without observational data (He *et al.* 2018). However, the present application with full-field observations inevitably suffers from the conservation problem as all the adjoint flows are directed upstream, and an outflow for \mathbf{v} is required to maintain the flow continuity. Therefore, the equivalent Neumann condition

is applied to both the inflow and outflow boundaries by extrapolating v and q from the inner grid onto the boundary (Ferziger, Peric & Street 2020). For the sidewalls of the computational domain, which have no appreciable effects on the computational stability and are almost parallel to the flow mainstream, the free-slip condition for both u and v and the Neumann condition for both p and q are used.

The second aspect that should be addressed is the discretisation scheme. The second-order backward scheme is used for the instantaneous terms of both the primary and adjoint equations. The central differencing scheme is used to discretise the pressure and viscous terms. For the primary convection terms, the total-variation diminishing scheme with the SuperBee limiter (Waterson & Deconinck 2007) is used; the scheme is implemented through deferred correction (Ferziger *et al.* 2020) to improve the numerical stability. For the adjoint convection terms, the central differencing scheme is used in most parts of the computational domain, whereas it blends to the downwind scheme near the inflow and outflow boundaries. The adjoint transpose convection and adjoint source terms are implemented explicitly.

The numerical stability and convergence speed are sensitive to the step lengths λ_ξ and λ_s in (2.17) and (2.18). A larger step length accelerates the convergence but may induce severe instability. An appropriate step length can be estimated according to the flow sensitivity with respect to ξ or s and is then fixed throughout the computation (He *et al.* 2020). In the present application, however, a constant step length results in extremely slow convergence after dozens of iterative loops. The linear search algorithm can be used to improve the convergence behaviour in combination with using nonlinear conjugate gradient methods to determine the optimal search direction (Lemke 2015; Li *et al.* 2019) rather than using the steepest descent method. Chandramouli *et al.* (2020) used a second-order limited-memory Broyden–Fletcher–Goldfarb–Shanno method for their 4D-Var problem. However, such optimisation algorithms rely on additional computations of the direct problem, which treble the computational cost in each loop. Moreover, these approaches do not give converging results in the present study as the residual is assessed only at the terminal time of the DA window. Therefore, a new algorithm based on the steepest descent method with an adaptive step length (ALSD) is proposed. The algorithm is based on the estimation of λ_ξ^0 and λ_s^0 according to He *et al.* (2020), followed by an increase in λ_ξ^k when the residual ε_ξ decays or a decrease in λ_ξ^k (or λ_s^k) when the residual ε_ξ (or ε_s) increases. The superscript k denotes the loop number of the iteration. The algorithm is formulated as

$$\lambda_s^k = \begin{cases} \lambda_s^{k-1} & (\varepsilon_s^k \leq \varepsilon_s^{k-1}) \\ 0.2\lambda_s^{k-1} & (\varepsilon_s^k > \varepsilon_s^{k-1}), \end{cases} \quad (2.19a,b)$$

$$\lambda_\xi^k = \begin{cases} \sqrt{\frac{\varepsilon_\xi^0}{\varepsilon_\xi^{k-1}}} \lambda_\xi^{k-1} & (\varepsilon_\xi^k \leq \varepsilon_\xi^{k-1}) \\ 0.5\lambda_\xi^{k-1} & (\varepsilon_\xi^k > \varepsilon_\xi^{k-1}). \end{cases} \quad (2.20a,b)$$

The determination of λ_ξ^0 and λ_s^0 depends on a rough estimation. These initial step lengths can be initialised through trial and error using the first DA window. The resultant cost increase relative to the overall DA is negligible (<1%). The use of the ALSM method greatly reduces the number of iterations required to reach the convergence criterion but does not increase the computational cost for each iteration.

The last aspect of concern is storage, which is a well-known issue for 4D-Var. It is necessary to store the primary and adjoint flow fields at all time steps, resulting in great

demands for storage space and a long reading and writing time. A checkpointing strategy has been widely used in 4D-Var, where the primary or adjoint simulation is checked at selected points with flow data stored in memory. All the intermediary data between the checkpoints in the adjoint or primary simulation are obtained through re-computation. However, in the present application, all the flow data of the primary and adjoint equations in the previous iteration are stored in memory owing to the small length of the DA window. This substantially reduces the time required for re-computation and that required for reading and writing.

All the computations are conducted using an in-house finite-volume code written in Fortran 90. The code solves the primary and adjoint NS equations on a structured Cartesian grid with a collocated arrangement using the SIMPLE algorithm for velocity–pressure coupling. The primary NS solver is adapted from the work of Ferziger *et al.* (2020) and has considerable computational efficiency and stability. Although only the serial version has been developed presently, this code has high computational efficiency and is capable of large-scale computations with millions of cells.

3. Computational and synthesis tomo-PIV set-ups

3.1. Description of the raw LES data

The observational data required in the DA computation are obtained from the LES results of a circular jet with Reynolds number $Re = 6000$ based on a nozzle diameter $D = 0.014$ mm and jet bulk velocity at the nozzle exit $U_0 = 0.42$ m s⁻¹ (He *et al.* 2018*b*). This LES was performed on a multiblock grid with adaptive refinement based on the velocity gradients. This resulted in a grid of approximately 9 million nodes in the computational domain. The length scale of the grid cells was approximately 300 μm in the region of DA computation. The dissipation rate of the jet flow can be estimated as (Panchapakesan & Lumley 1993)

$$\varepsilon \approx 0.015 \frac{U_c^3}{r_{1/2}}, \quad (3.1)$$

where U_c is the centreline velocity and $r_{1/2}$ is the jet half-width. Adopting the above approximation, the Taylor microscale and Kolmogorov scale are estimated to be 1 mm and 60 μm, respectively. This suggests that the computational grid has extremely high spatial resolution for the LES. The time step is fixed at $\Delta t_{LES} = 0.0002$ s, resulting in a maximum local CFL number of approximately 1. The mean velocity and the fluctuation are validated using planar PIV data by He *et al.* (2018). Data are written with a time interval of $\Delta t = 0.001$ s, and 1500 instantaneous fields are saved for the following DA computation and validation.

3.2. Synthetic tomo-PIV measurement

Tomographic PIV is applied to synthetic particle images produced from the LES data. Four virtual cameras with a resolution of 1000×2000 pixels are installed azimuthally with an included angle of approximately 20° around the jet using the layout adopted by He *et al.* (2022). The virtual measurement domain has fixed dimensions of 0.106 m (7.57*D*), 0.053 m (3.79*D*) and 0.053 m (3.79*D*) in the x , y and z directions, respectively, and is placed 0.037 m (2.64*D*) downstream of the nozzle exit. Therefore, a column along the x direction being slightly larger than the measurement domain is considered to be illuminated by a virtual laser. Particles are initialised in the illumination region

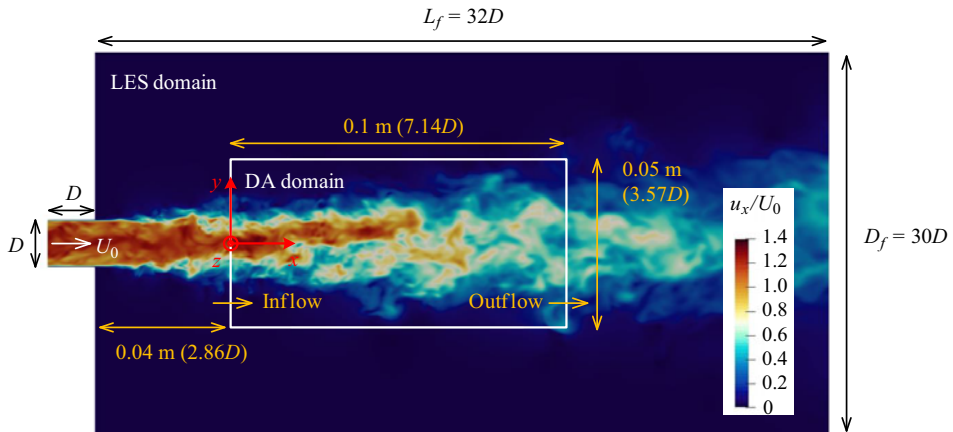


Figure 1. Schematic of the computational domains. The large region denotes the cylindrical LES domain whereas the white box indicates the cuboid DA domain. The z coordinate is directed outwards perpendicular to the paper as noted by ‘ \odot ’. Lengths L_f and D_f are plotted using different scales without showing the full range.

with a mutual distance exceeding $50 \mu\text{m}$. New particles are seeded randomly from the LES inflow boundary and are advected downstream to the illumination region with the consideration of particle entrainment from the side surface of the illumination region. After sufficient advection, the particle distribution in the illumination region reaches statistical stability and the particle concentration in the jet mainstream remains appreciably higher than the ambient concentration. The coordinates of the particles are then projected to the cameras, and synthetic images are generated following the procedure described by Tan *et al.* (2020). Images are thus produced with a particle diameter of approximately 3–5 pixels on the images and a particle concentration of 0.02 particles per pixel. Particle images are Gaussian-filtered (3×3 pixels) and noise-contaminated (standard deviation of 0.1) to approach real experimental conditions.

The cameras are calibrated using a synthetic dotted plane that can be shifted in the z direction as in the work of He *et al.* (2022). Calibration images are generated in the manner described above, and the projection matrix is then determined. During the tomography reconstruction, the virtual measurement domain is discretised to $1380 \times 690 \times 690$ voxels with a physical spatial resolution close to that of the particle images. The voxel size is approximately 0.072 mm. The particle displacement in the mainstream for each pair of images is approximately 4 pixels according to the imaging time interval. GPU-accelerated code (Zeng, He & Liu 2022) based on the MART algorithm, combined with the iterative multigrid volumetric cross-correlation approach with a final pass interrogation volume size of $32 \times 32 \times 32$ voxels and 50% overlap, is used to determine the three-dimensional velocity fields. This yields a spatial resolution of the velocity vectors of 1.2 mm. One thousand instantaneous velocity fields are obtained with a time step $\Delta t = 0.001$ s.

3.3. The DA computational set-up

The DA domain used in this study is a cuboid that has dimensions of 0.1 m ($7.17D$), 0.05 m ($3.57D$) and 0.05 m ($3.57D$) in the x , y and z directions, respectively, and is placed 0.04 m ($2.86D$) downstream of the nozzle, as shown in figure 1. This domain is typical for tomo-PIV and 4D-PTV measurements but far from sufficient for conventional numerical simulations as the inflow and outflow boundary conditions affect the flow.

No.	Case name	ΔT	Total time T (s)	λ_s^0	λ_ξ^0	α	Observations
1	DA-Clear10	$10\Delta t$	0.01	100	2×10^4	0	Clear LES
2	DA-Clear20	$20\Delta t$	0.02	100	2×10^4	0	Clear LES
3	DA-Clear30	$30\Delta t$	0.03	100	2×10^4	0	Clear LES
4	DA-Noisy5e-5	$20\Delta t$	1.0	100	2×10^4	5×10^{-5}	Noisy LES
5	DA-Noisy5e-6	$20\Delta t$	1.0	100	2×10^4	5×10^{-6}	Noisy LES
6	DA-Noisy5e-7	$20\Delta t$	1.0	100	2×10^4	5×10^{-7}	Noisy LES
7	DA-Tomo	$20\Delta t$	1.0	100	2×10^4	5×10^{-6}	Tomo-PIV
8	DA-SS	$20\Delta t$	1.0	100	2×10^4	2×10^{-5}	Tomo-PIV with SS injection
9	DA-SSN	$20\Delta t$	1.0	100	2×10^4	2×10^{-5}	Tomo-PIV with SS injection and noise

Table 1. List of DA cases and parameters.

Time Δt ($= 0.001$ s) is the time step of the DA computation; ΔT is the length of the DA window; T is the total physical time for all the DA windows.

It is noted that the DA domain is selected to be slightly smaller than the virtual measurement domain mentioned in § 3.2 for ease of grid interpolation. The grid used in DA is Cartesian with 256, 128 and 128 elements uniformly distributed along the x , y and z directions, respectively, yielding a total of 4.36 million nodes with a spatial resolution of approximately 400 μm . This grid resolution meets the requirement of LES but does not enable reliable flow predictions owing to the undersized domain and inappropriate boundary conditions. The observations are produced by linearly interpolating (as expressed by \mathbb{H} (2.3)) the LES or tomo-PIV data onto the whole DA grid with a time interval ΔT . Boundary conditions, differential schemes and step lengths are applied according to the discussion in § 2.3.

The DA cases considered in this study are listed in table 1. Clear cases (1–3) have ideal conditions with clear LES data as observations, and the regularisation in the DA computation is thus turned off. Different window lengths are tested to evaluate the sampling rate of the observations for the STR reconstruction. Noisy cases (4–6) use the LES data with white-noise contamination and box filtering ($3 \times 3 \times 3$ elements on the DA grid) as observations for evaluation of the de-noising capability. The noise standard deviation is approximately $0.3U_0$ as that used by Gillissen *et al.* (2019), and is introduced by adding random numbers in the range of $[-A_{noise}, A_{noise}]$ to the LES data, where

$$A_{noise}(\mathbf{x}) = 0.6\max(\|\mathbf{u}(\mathbf{x})\|, 0.2). \quad (3.2)$$

The coefficient 0.6 is used to tune the noise standard deviation to the desired value, and the minimum value of 0.2 is used to produce the noise in the ambient flow. The DA-Tomo case (case 7) uses the tomo-PIV data \mathbf{u}^{tomo} as observations, whereas in the DA-SS case (8), small-scale coherent structures are injected into the tomo-PIV observations according to

$$\mathbf{u}^{obs}(t) = \mathbf{u}^{tomo}(t) + \mathbf{u}^{LES}(t') - \widetilde{\mathbf{u}^{LES}}(t'). \quad (3.3)$$

Here, \mathbf{u}^{LES} represents the velocity in LES, where $\widetilde{\mathbf{u}^{LES}}$ is the box-filtered LES data on the DA grid with a filter size of $7 \times 7 \times 7$ elements, which is close to the physical size of the tomo-PIV interrogation volume. The time $t' = t + 0.5$ s is used to ensure that the injected small-scale structures are not correlated with the tomo-PIV fields. The DA-SSN

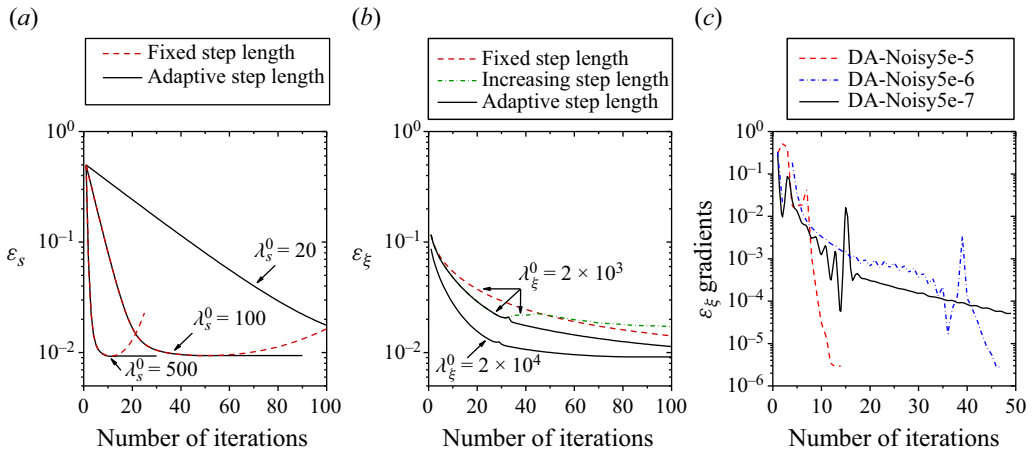


Figure 2. Residuals and their gradients in DA computations: (a) initial field DA, (b) model error DA in the case of DA-Clear30 and (c) residual gradients in the model error DA of noisy cases.

case (case 9) is similar to DA-SS except that white noise defined by (3.2) is added for further verification. The time step of the computation (i.e. the time interval of the STR reconstructed fields) is fixed at $\Delta t = 0.001$ s, while ΔT denotes the length of each DA window. Time T is the total physical time counting all the DA windows in each case.

All the computations are performed on a desktop computer with an Intel Xeon E-2144G quad-core processor and 64 G memory. The serial computation of each multi-window case (cases 4–9) takes approximately 78 hours to find the optimal solution including the initial and inflow conditions, with there being 1000 instantaneous fields within the total time of 1.0 s (50 DA windows). Multitask parallel computing is conducted by submitting four cases at the same time to accelerate the overall computation by a factor of more than 3. The data writing is the most time-consuming step and further improvements can thus be made through the appropriate compression of the result files rather than writing ASCII data directly as done presently.

4. Results and discussion

4.1. Super-temporal-resolution reconstruction with down-sampled observations

Data assimilation is first applied for STR reconstruction using the observations of down-sampled LES data containing many small-scale turbulence structures. The computational cases are listed as cases 1–6 in table 1. The length of the intermediate DA window $20\Delta t$ equals 100 LES time steps. For cases 4–6, 50 DA windows are computed within a total time T corresponding to 9 turnovers of the largest-scale mode (Hussain & Zaman 2006) according to the Strouhal number $St = 0.3$. Figure 2 presents the iteration residuals for different step-length strategies in clear and noisy cases. In the initial field DA phase as shown in figure 2(a), the instantaneous flow at a random instant is used as the first guess. The adaptive step lengths are computed using (2.19a,b). The residual has an increasing descent rate with respect to the initial step length, with only 10 iterations required using $\lambda_s^0 = 500$. Nevertheless, the residual grows rapidly after reaching a minimum when a constant step length is used; this growth is well suppressed by the adaptive formulation. In the model-error DA phase as shown in figure 2(b), there is an increasing descent rate with respect to the initial step length. When adopting the

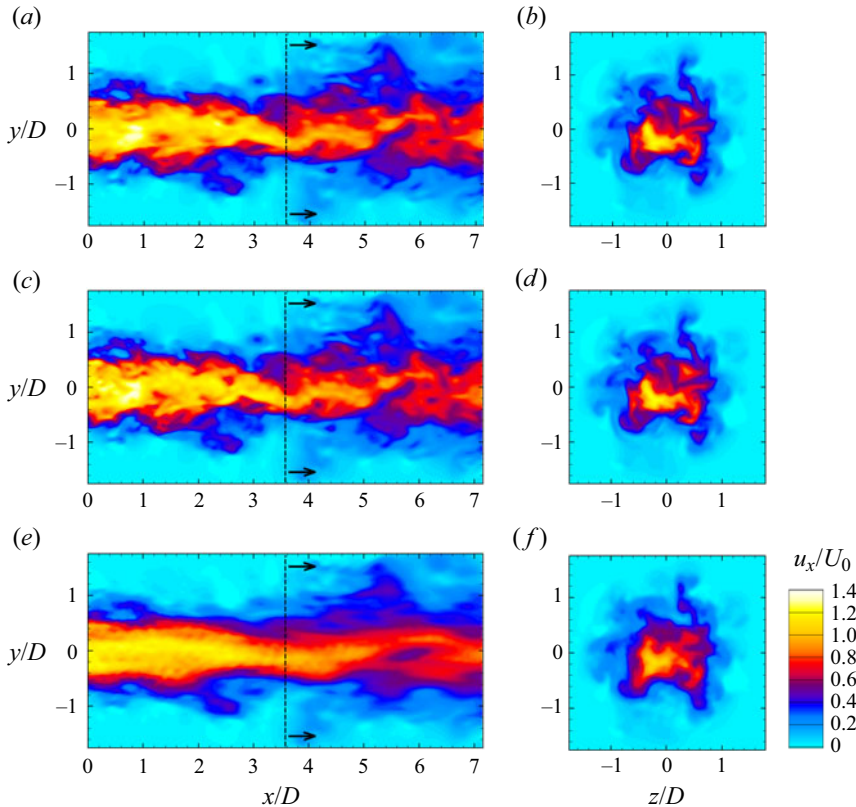


Figure 3. Instantaneous flow fields at the middle instant in the DA window: (a,b) DA reconstruction, (c,d) clear LES data and (e,f) POD optimal reconstruction. (a,c,e) The longitudinal middle sections and (b,d,f) the cross-sections at the location marked by the dashed line.

adaptive scheme, a gradual increase in the step length according to (2.20a) is effective in accelerating convergence whereas a large step length is prone to cause divergence. This problem can be solved using the adaptive step length with the combined use of (2.20a) and (2.20b). Although this adaptive scheme relies on trial and error to determine the optimal initial step length, the complementary computations can be performed only in the first time step or the first DA window, yielding an increase in the computational cost of less than 1% in total in this study. In noisy cases, the residual gradient with respect to the iteration number is better suited to convergence evaluation as the residual level is strongly associated with the regularisation coefficient α . Obviously, larger α results in better de-noising effects in the DA procedure and thus greater discrepancy between the DA results and observations. As shown in figure 2(c), the residual gradient has the highest convergence speed in the DA-*Noisy5e-5* case, reducing to below 10^{-5} after 10 rounds of iteration. The number of iterations is fixed at 50 in the model-error DA phase of all noisy cases.

In the STR reconstruction, the flow at the middle instant between the two observations is undoubtedly subject to the largest reconstruction error. This suggests the need to inspect the reconstructed flow snapshot at the middle instant in the case of DA-*Clear30* as shown in figure 3. The clear LES data and the POD optimal reconstruction (Appendix C) from the down-sampled observations with time interval $30\Delta t$ are also presented for comparison. The DA result is similar to the LES data on both longitudinal and cross-section planes

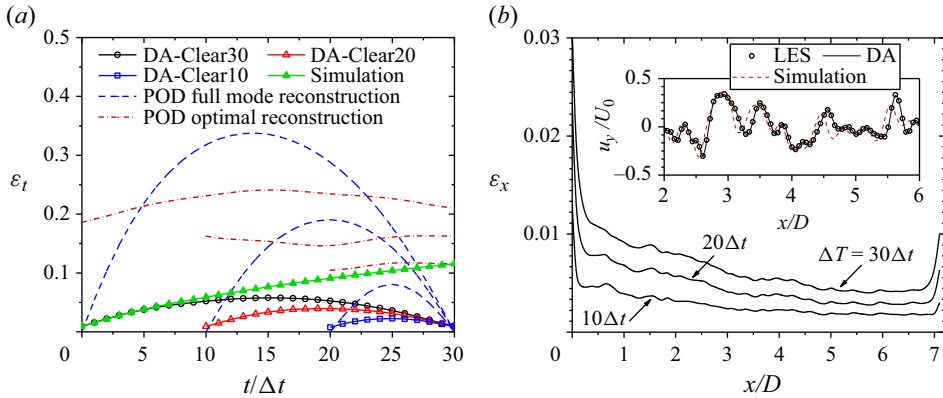


Figure 4. Pointwise error of the reconstructed flow in each clear case: (a) domain integration at different instants and (b) cross-wise and temporal integrations at different downstream locations. Instantaneous vertical velocities are inset in (b) for comparison.

as shown in figure 3(a–d). There are also slight differences in the upstream region near the inflow boundary, where some plume-like signatures are missing from the jet mainstream in the DA results. This is largely due to the error in the assimilated inflow conditions. In addition, the temporal evolution of these reconstructed fields shows the advantage of DA in STR reconstruction compared with interpolation-based approaches, which basically rely on temporally resolving the turbulence events with a sampling rate up to that meeting the Nyquist law. A POD full mode reconstruction (reconstruction using all of the modes) using the existing observations disrupts the real evolution process of the convecting small-scale vortices and is thus applicable only to slowly evolving vortices. A POD optimal reconstruction, as shown in figure 3(e, f), selects the most energetic modes that are temporally correlated for interpolation. This correctly recovers the flow convective properties but preserves less than 90 % of the total kinetic energy.

Figure 4 presents the pointwise error defined as

$$\varepsilon_t = \sum_{(x,y,z)} \|\mathbf{u}(t) - \mathbf{u}^{LES}(t)\| / \sum_{(x,y,z)} \|\mathbf{u}^{LES}(t)\|, \quad (4.1)$$

$$\varepsilon_x = \sum_{(t,y,z)} \|\mathbf{u}(x) - \mathbf{u}^{LES}(x)\| / \sum_{(t,y,z)} \|\mathbf{u}^{LES}(x)\|, \quad (4.2)$$

for each reconstruction scheme in clear cases. The summations are applied for all possible coordinates (x, y, z) and (t, y, z) , respectively. In this application, the terminal observation is fixed at $t/\Delta t = 30$ whereas the initial observations are selected at $t/\Delta t = 0, 10$ and 20 for the cases of DA-Clear30, DA-Clear20 and DA-Clear10, respectively. The direct simulation by solving the NS equation for the initial field at $t/\Delta t = 0$ yields an error that increases monotonically with respect to time, as shown in figure 4(a). This error is largely induced by the boundary conditions imposed on the compact numerical domain and is absorbed by the model error in this study. Data assimilation is performed from each selected initial field and yields the largest error at the middle time of the DA window before reaching a minimum at the end. The error is slightly higher than zero at each start and end time owing to the finite residual level at the final iterative loop. With assimilation of the model error ξ , the error of the DA reconstruction is appreciably lower than that of the direct simulation and maintains a clear descending trend with respect to

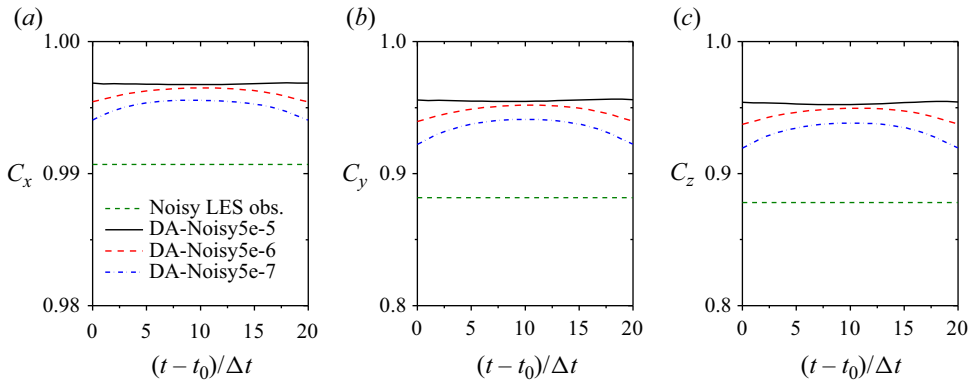


Figure 5. Cross-correlation coefficients of the reconstructed velocity fields with the referential LES data: (a) x component, (b) y component and (c) z component. The noisy LES data are averaged over all time. Different ordinate scales are used for clear illustration.

the decreasing DA window length ΔT . Additionally, the POD full modes and optimal reconstructions are compared with the DA results for each corresponding observational time interval. As the full mode reconstruction at an observational time (start or end of the DA window) produces the exact fields, the error is reduced to zero. However, the reconstruction error at the intermediate instants remains much higher than that of the DA results for each window length. The POD optimal reconstruction reduces the maximum error for large observational time intervals but performs even worse than the POD full mode reconstruction for small time intervals. The above results indicate that although POD reconstruction plays important roles in the feature recovery of large-scale evolution, the lack of temporal information on small scales in the raw observations remains a barrier for the temporal resolution enhancement of turbulence details. In the DA reconstruction, the error is mainly concentrated in the upstream region, where the jet speed is high, as shown in figure 4(b); the error is manifested as the convection lag of the flow events, as shown by the instantaneous vertical velocity distributions. In addition, the inflow and outflow boundary conditions have notable local effects. Nevertheless, most of the error is below 1% in the case of DA-Clear30, indicating the high accuracy of the proposed DA scheme for STR reconstruction even with a long time interval.

As an analogy to the measurement noise in a real experiment (e.g. PTV), the noisy observations are synthesised by adding white noise to the clear LES data according to (3.2) and box filtering using $3 \times 3 \times 3$ elements on a DA grid. In these cases, DA is performed successively in 50 windows using the terminal fields of the previous window for initialisation. Detailed parameters are given in table 1. To assess the accuracy of the DA field throughout the domain, the cross-correlation coefficient for the streamwise velocity is calculated using the clear LES data as reference for each instant in the DA window:

$$C_x(t) = \frac{\sum_x [u(x, t) \cdot u^{LES}(x, t)]}{\sqrt{\sum_x [u(x, t)]^2 \cdot \sum_x [u^{LES}(x, t)]^2}}, \quad (4.3)$$

where $t \in [t_0, t_0 + \Delta T]$, and the summations are applied for each spatial location $\mathbf{x} = (x, y, z)$ on the DA grid. The cross-correlation coefficients for other velocity components are defined similar to (4.3). The final cross-correlation coefficients are averaged for each relative time with respect to the start instants in all the DA windows. The correlation

coefficients for all the noisy cases, as well as the noisy observations, are shown in [figure 5](#). The correlation coefficients of the noisy observations remain appreciably lower than those of all the DA cases owing to the noise contamination. This clearly indicates the strong capacity of DA to recover the turbulence properties even when taking noisy measurement data as observations. Comparatively, the regularisation term in the objective function not only eliminates noise but also plays an important role in improving the facticity of the reconstructed flow fields. The correlation coefficient is the highest at the middle instant of the DA window for the cases of DA-Noise5e-6 and DA-Noise5e-7 owing to the propagation of the governing equations. This trend is obviously different from that of the clear cases, where the error is largest in the middle of the window. When a smaller value of α is used, DA applies a larger weighting to the noisy observational data and thus introduces more error into the instantaneous field near the initial and terminal instants. We thus observe straighter correlation lines along time with larger α , and the best performance with nearly constant correlations in the DA window is achieved in the case of DA-Noise5e-5. The optimal α is not readily obtained from [figure 5](#) for different DA cases. The value of α is obviously case-dependent but can be cheaply determined adopting the DA procedure with only one window. In the remaining discussion of noisy cases, only DA-Noise5e-5 is taken as the representative example.

The assimilated ξ term acts as a compensator of error between the model prediction and observation. It also plays an important role in the turbulence evolution. The physical interpretability of the compensator has increasingly drawn attention for numerous data-driven techniques. The compensator not only directly reduces the predictive error between two datasets but also is a feature vital for the flow to evolve correctly. Term ξ has equivalent effects on flows with pressure gradients in the NS equations but is always divergence-free according to (2.17). Helmholtz decomposition casts ξ in two parts as

$$\xi = \psi - \nabla\varphi, \tag{4.4}$$

where ψ is stochastic forcing due to, for example, the subgrid stress. The reconstruction error is included in ψ in the present DA framework. Term $\nabla\varphi$ is the irrotational part, which is lumped into the pressure gradient. The pressure and pressure gradient component in the x direction are first explored in [figure 6](#). There is a visible difference in the pressure and its gradient distributions between the clear LES result and the other results. Simulation by solving the original NS equations directly does not reproduce the pressure and gradient correctly, as shown in [figure 6\(a–d\)](#). The quantitative comparison presented in [figure 7](#) shows the model error effects on the confined computational domain. The results obtained by simulation and DA have high similarity, as shown in [figure 6\(c–f\)](#), with only slight discrepancy observed quantitatively in [figure 7](#). The above results demonstrate the important contribution of the irrotational part $\nabla\varphi$ in compensating for the pressure gradient discrepancies stemming from the compact computational domain. The results also indicate that further computation steps after DA are required to reconstruct the pressure fields correctly. An explicit calculation of $\nabla\varphi$ and ψ can be readily made considering the difference in the pressure gradients between the LES and DA, i.e.

$$\nabla\varphi = \nabla p^{LES} - \nabla p^{DA}, \tag{4.5}$$

$$\psi = \xi + \nabla p^{LES} - \nabla p^{DA}. \tag{4.6}$$

Probability density functions (PDFs) of ξ_x , ψ_x and $\nabla_x\varphi$ are plotted in [figure 8](#), where ∇_x is the streamwise component of the gradient. All quantities are normalised by U_0^2/D . Herein, PDFs are separately calculated for all DA windows at the second instant immediately

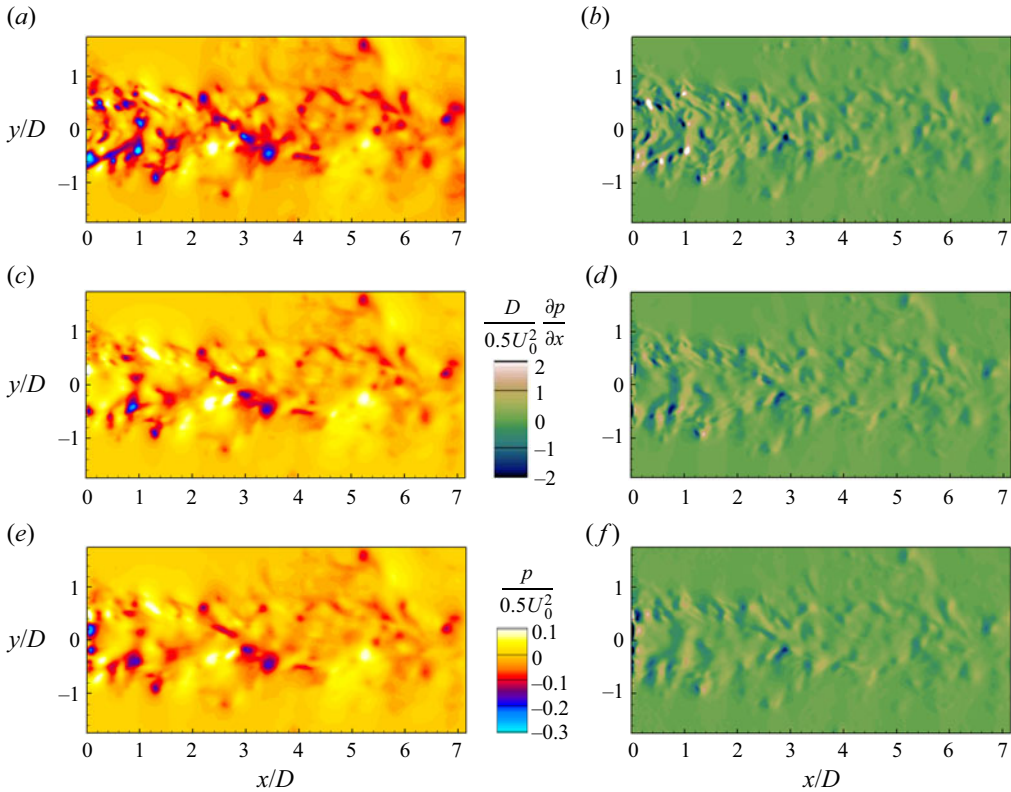


Figure 6. Instantaneous pressure (*a,c,e*) and streamwise-component pressure gradient (*b,d,f*) fields in (*a,b*) clear LES, (*c,d*) simulation and (*e,f*) the case of DA-Noise5e-5.

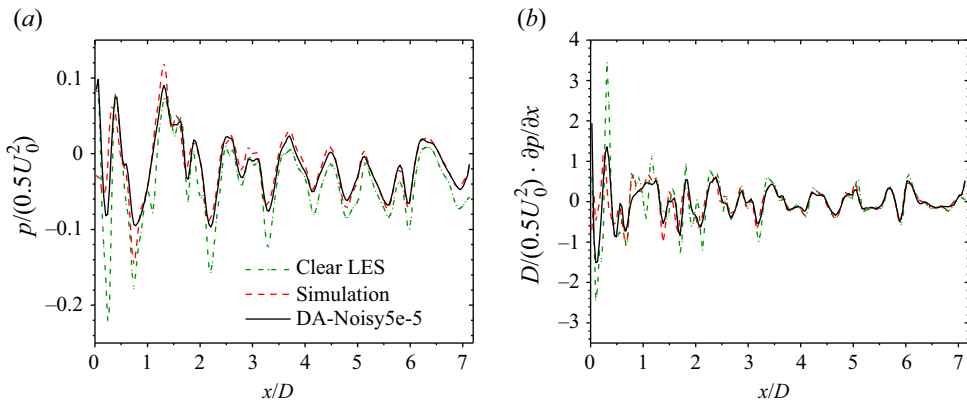


Figure 7. Quantitative comparison of (*a*) the pressure and (*b*) its streamwise-component gradient along a streamwise line at $y/D = 0.5$.

after the initial time (figure 8*a*), at the middle instant (figure 8*b*) and at the terminal instant (figure 8*c*) using data in the region $r/D < 1$ (where $r = \sqrt{y^2 + z^2}$). The results of averaging across all instants are shown in figure 8(*d*). The curves for the negative part of the abscissa are mirrored to the positive part to demonstrate the asymmetry with respect to

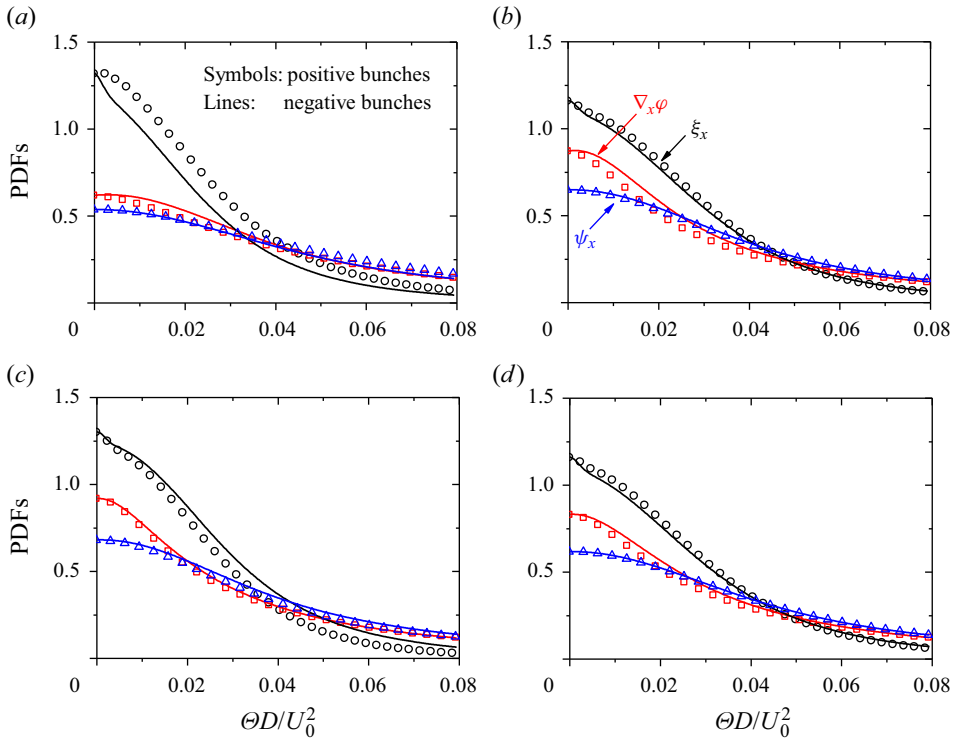


Figure 8. The PDFs of the quantities ξ_x , ψ_x and $\nabla_x\varphi$ at (a) the second instant, (b) the middle instant, (c) the terminal instant and (d) all instants. The data on the negative abscissa are mirrored to be positive. Here Θ denotes ξ_x , ψ_x or $\nabla_x\varphi$ as indicated by the arrows.

the ordinate axis. The PDFs of ξ_x have a sharp peak at the origin and rapidly decay in both positive and negative directions of the abscissa. There are clear differences between the positive and negative bunches, especially at the first instant, as shown in figure 8(a). This indicates that ξ_x has a pronounced bias towards the positive direction and complements the pressure gradient from the perspective of the whole region of interest. Recalling that direct simulation yields a lower speed of convection in the downstream direction as depicted in figure 4(b), ξ_x is physically interpreted as the driving force for the model prediction to keep up with the observation evolution. However, the importance of the driving force varies in time and largely prevails in the first half of the DA window. This depends on $\nabla_x\varphi$, which is biased against the driving force. The PDFs of ξ_x and $\nabla_x\varphi$ averaged across all the instants and shown in figure 8(d) clearly demonstrate the above mentioned complementation mechanism during the flow evolution in the DA window, although there is slightly anomalous behaviour at the terminal instant in figure 8(c), probably due to computational error. Supplemental evidence for this supposition is provided by the PDFs of ψ_x , which show good agreement for the positive and negative bunches at each instant in the figure. This result is consistent with our previous conjecture as the remaining part of ξ_x , when subtracting the irrotational part $\nabla\varphi$, has unbiased stochastic behaviours that are responsible for the production and dissipation of turbulence.

A direct view of the STR reconstruction is acquired from figure 9, where the temporal variations of the original and reconstructed velocities are plotted. The streamwise velocities at $x/D = 3.5D$ on the jet centreline are quantitatively compared in figure 9(a). The rapid variation of the clear LES results represents the small-scale coherent structures

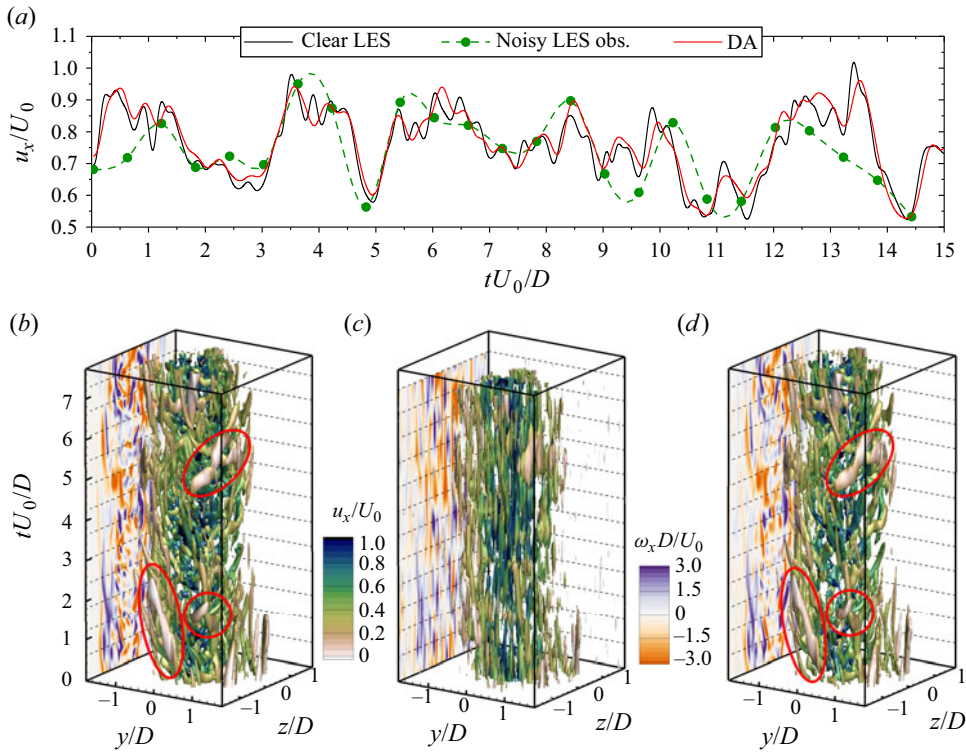


Figure 9. Temporal plots of the velocity and vortical structures ($Q=500$) from the original and DA reconstructed flow fields at $x/D = 3.5D$: (a) streamwise velocity variation on the jet centreline, (b) clear LES data with the time interval $\Delta t = 0.001$ s, (c) noisy LES observations with the time interval $\Delta T = 0.02$ s, (d) STR reconstructions with the time interval $\Delta t = 0.001$ s. The green dots and the dashed line in (a) denote the observational data and the spline fitting, respectively. The black dashed lines in (b–d) indicate the observational instants.

in the jet turbulence. The down-sampling of the observations ('Noisy LES obs.')

only captures the large-scale events even when adopting spline interpolation, and the observational noise induces appreciable deviation from the clear LES data. However, the DA results recover the small-scale fluctuations well, even though slight discrepancies are still observed. A salient feature of DA is that the STR results do not strictly follow the variation of the observations, manifesting remarkable de-noising and correcting effects when erroneous observations are used. The spatiotemporal plots of the streamwise velocities at the cross-section $x/D = 3.5D$ are present in figure 9(b–d). The clear LES data in figure 9(b) reveal appreciable temporal fluctuations along with the small-scale variation showing apparent coherent tubular structures. The detailed small vortex organisations are seen in the vorticity contours. In the down-sampled observations shown in figure 9(c), although a smooth isosurface is obtained, the loss of details between successive snapshots is an issue. This situation is equivalent to direct temporal interpolation, which makes sense only when the original sampling frequency meets the requirement of the Nyquist sampling law. The DA reconstruction shows obvious advantages in the recovery of flow details even when using the noisy down-sampled observations. The DA results with the sampling rate equalling that of the clear LES data are shown in figure 9(d). The contours of the streamwise vorticity clearly demonstrate the reconstructed flow details. The red circles

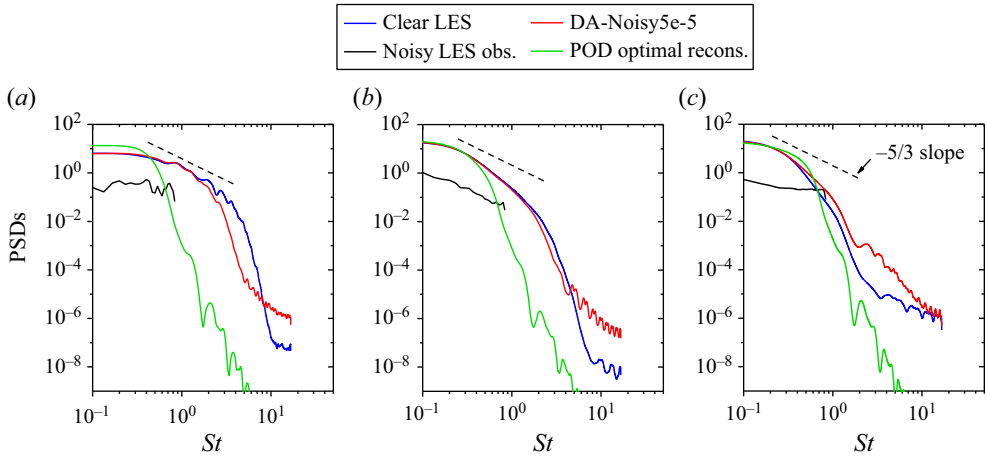


Figure 10. PSDs on the streamwise velocity at the downstream position $x/D = 3.5D$ at various radial locations: (a) $r/D = 0$, (b) $r/D = 1$ and (c) $r/D = 1.5$.

mark the moderate-scale vortices that exist in the clear LES data and are successfully reconstructed by DA but are lost in the down-sampled observations.

In-depth analysis of the small scales recovery of DA is conducted using the velocity spectra. Power spectral densities (PSDs) are computed using Welch's method (Welch 1967) with all 1000 snapshots for each case. Although the window size of 200 (corresponding to 1.8 turnovers of the largest-scale flow fluctuation) with an overlap of 100 in the PSD computation does not provide a good estimation of the spectra at low frequencies due to the data size, the results are not visibly different from those obtained using larger windows. Our present analysis focuses on high-frequency structures with Strouhal numbers larger than 0.2, which can be well captured by the present PSD algorithm. Figure 10 shows the PSDs of the streamwise velocity at various radial locations. The PSDs of the time-sparse noisy LES data are lower than the others due to the normalisation using the root mean square of the velocity signals, and it is only possible to resolve vortical structures up to $St = 0.1$ with a relatively flat distribution due to the noise. The POD optimal reconstruction recovers the large-scale vortices well but yields a rapid decay of the PSDs at $St > 0.5$ due to the lack of small-scale information. The DA reconstruction produces the best results up to $St = 5$, although with a plateau higher than that for the clear LES data due to the noise in the observations. An observable feature of the PSDs at different radial locations is the underestimation of the DA reconstruction at $r/D = 0$ but a slight overestimation at $r/D = 1.5$. This feature is reasonable in the present computations due to the high noise ratio at far radial locations where the flow speed is low.

4.2. Super-temporal-resolution reconstruction for tomo-PIV with small scales injection

Although tomo-PIV has become one of the most widely used volumetric measurement techniques in fluid mechanics, the compromise between the measurement domain size and the spatial resolution resulting from the cross-correlation (Lawson & Dawson 2014; Naka *et al.* 2016; Fiscaletti *et al.* 2022) is the major issue that limits its application in certain circumstances. The present study evaluates the field correction for the low-sampling-rate tomo-PIV results through DA reconstruction, with much effort being given to the regeneration of the turbulence structures by small scales (SS) injection from the observations. It is worth noting that the small vortices cannot be injected from the

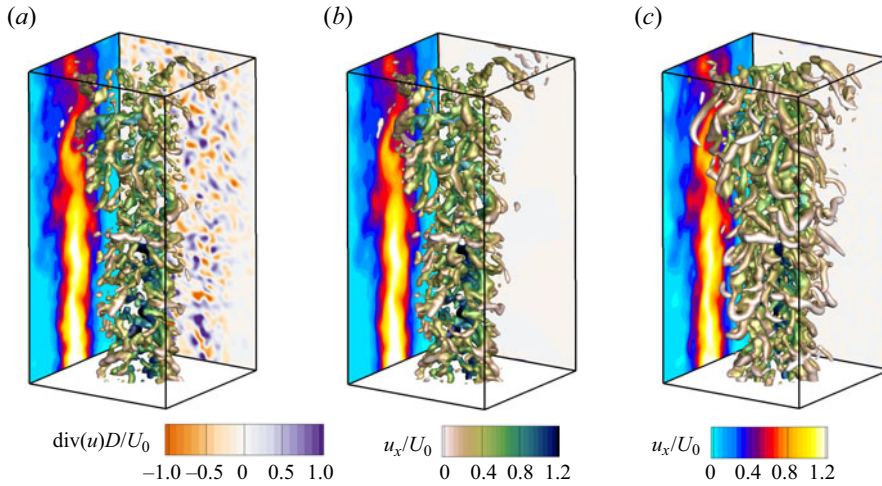


Figure 11. Instantaneous velocity fields and vortical structures ($Q = 500$) of (a) the synthetic tomo-PIV measurements, (b) DA reconstruction and (c) filtered LES data on the DA grid.

initial condition as done by Li *et al.* (2019) because of the weak-constraint nature of the present DA algorithm and the time-sparse observations.

The case of DA-Tomo (see table 1) is first discussed on the basis of the snapshots selected at the terminal instant of the DA window. The comparison of the tomo-PIV data (observation) and the DA result presented in figure 11(a,b) shows excellent similarity in terms of the velocity contours and even the three-dimensional vortical structures (Q -criterion isosurface, $Q = 500$). However, a notable feature of the DA procedure is the prominent ability to remove the velocity divergence error, which is important for post-processing, such as vorticity computation and pressure determination, using the existing three-dimensional fields. A snapshot obtained by box filtering (denoted $\widetilde{\mathbf{u}}^{LES}$, filtered using $7 \times 7 \times 7$ elements on the DA grid, which is close to the physical size of the tomo-PIV interrogation volume) from the clear LES data at a selected instant is shown in figure 11(c). The velocity contours barely show any difference between the different data, but the isosurface of the Q criterion in the box-filtered field is more coherent with many elongated vortex tubes distributed in the jet shear layer. This indicates that the cross-correlation process in tomo-PIV measurements involves far more than spatial filtering than what has been commonly understood before, and more complex effects should be considered.

The turbulence fields of the reconstructed flow can be inspected in more depth by plotting the joint PDFs of the invariants Q^* and R^* , which are defined as (Rishita & Girimaji 2022)

$$Q^* = -\frac{1}{2}b_{ij}b_{ji}, \quad (4.7)$$

$$R^* = -\frac{1}{3}b_{ij}b_{jk}b_{ki}, \quad (4.8)$$

where $b_{ij} = A_{ij}/\|A\|$. Here, the asterisk is added to the invariants to distinguish them from the vortex Q criterion. The velocity gradient tensor is $A_{ij} = \partial u_i/\partial x_j$, and $\|\cdot\|$ denotes the Frobenius norm. The zero-discriminant lines are determined according to $Q^{*3} + 27R^{*3}/4 = 0$. It has been well established that the Q^*-R^* PDFs have a characteristic teardrop shape (i.e. a Vieillefosse tail) (Vieillefosse 1984) in various turbulent flows, with

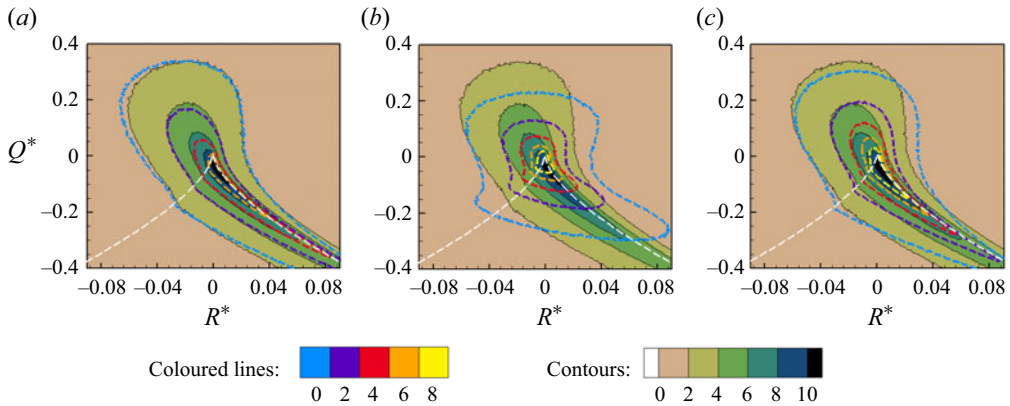


Figure 12. Joint PDFs of the velocity gradient invariables Q^* and R^* . The solid contours show the clear LES data, while the coloured dashed lines show (a) filtered LES results, (b) tomo-PIV data and (c) the DA reconstruction. White dashed lines in the third and fourth quadrants represent the zero-discriminant lines $Q^{*3} + 27R^{*3}/4 = 0$. The PDFs are normalised by the mean values on the Q^*-R^* plane.

a high probability of occurrence along the right discriminant line. This characteristic probability map represents the predominance of sheet-like structures in the fourth quadrant and that of vortex stretching in the second quadrant. In this study, Q^*-R^* maps are calculated using all the flow instantaneous fields in the region of $r/D < 0.75$. A map of the clear LES data is plotted using solid contours in each panel of figure 12 for comparison. Figure 12(a) shows that both the clear and filtered LES data have the well-known teardrop shape in the Q^*-R^* plane, whereas the filtered flow has a slightly slimmer tail indicating fewer sheet-like structures and less vortex compression effects. However, there is an appreciable difference in the tomo-PIV data, where the predominant occurrences of the high PDFs are located slightly above the discriminant lines, as shown in figure 12(b). Misalignment in tomo-PIV data has previously been observed (Khashehchi *et al.* 2010) but with high PDFs slightly below the discriminant lines. This misalignment is obviously case-dependent, whereas tomo-PIV invariably fails to capture the detailed turbulence dynamics. The DA reconstruction produces a Q^*-R^* map having a well-organised teardrop shape, which is shown in figure 12(c) and is more like the filtered LES results. This implies an improvement of the DA procedure in turbulence reconstruction even though the improved reconstruction can be hardly distinguished from the vortical structures, as shown in figure 11. Additionally, there is a slight difference, with a sharper Vieillefosse tail, relative to both the clear and filtered LES data. In contrast, Vieillefosse tails were observed to be shorter and blunter in the work of Li *et al.* (2019) owing to the use of different DA algorithms.

The filtered flow $\widetilde{\mathbf{u}}^{LES}$ is then subtracted from the clear LES fields to derive the small-scale structures, which can be added to the tomo-PIV observations for the SS injection. To remove the correlation between the injected small scales and the tomo-PIV snapshots, a time shift of 0.5 s (approximately 4.5 turnovers of the largest-scale jet oscillation) is applied for the observations in the cases of DA-SS and DA-SSN. This approach is analogous to real tomo-PIV work where small-scale information is injected from independent numerical data. Recalling that the SS injection from initial fields is subject to extremely expensive computation, taking 15 days to determine a single initial field in a long DA window as noted by Li *et al.* (2019), the observation injection in the present study requires less than 5 minutes to determine each optimal field on a desktop

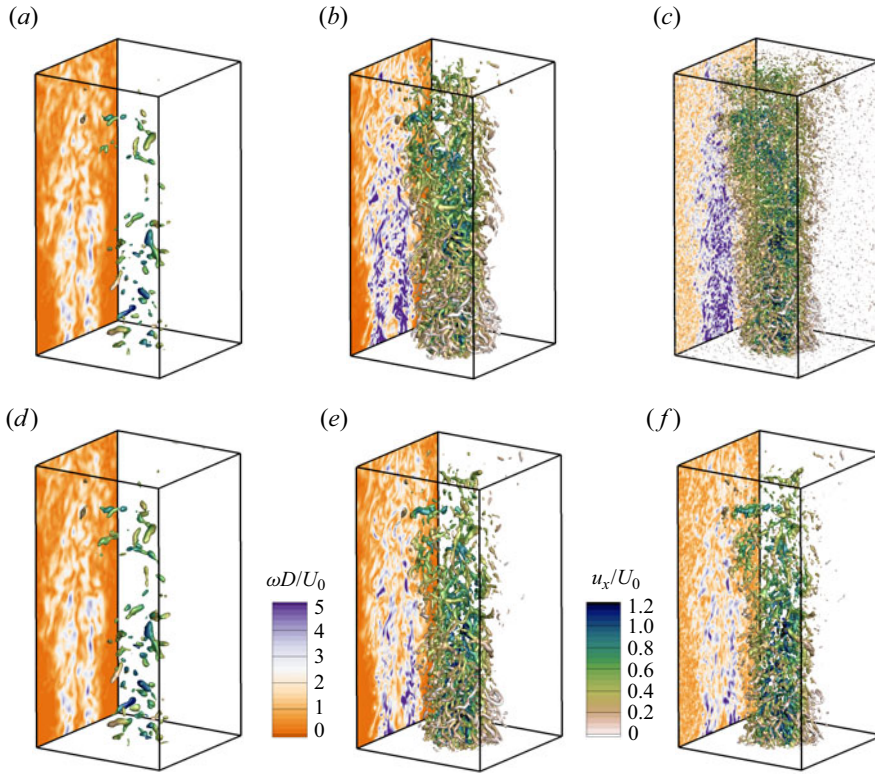


Figure 13. Instantaneous vorticity magnitude contours and vortical structures ($Q = 3000$) of the observations (*a–c*) and DA reconstructions (*d–f*): (*a*) tomo-PIV observation, (*b*) tomo-PIV observation with SS injection, (*c*) tomo-PIV observation with SS injection and noise, (*d*) DA-Tomo, (*e*) DA-SS and (*f*) DA-SSN. The instant at the terminal of a DA window is taken.

computer, with the number of computational grid elements being twice that used by Li *et al.* (2019). This SS injection is principally achieved by the adjoint source, which transfers the large-scale information from the observations to the primary flow quantities, with the system equations serving as the physical filter for the vortex correction and noise removal. A qualitative understanding of the DA reconstruction for small scales recovery is acquired from figure 13, which presents different observations and the DA results. It is noted that the Q value (3000) used to show the three-dimensional vortical structures is much larger than that used for figure 11. The tomo-PIV field and its DA result in figure 13(*a,d*) thus show few vortical structures and a low vorticity magnitude with much emphasis on the small scales in the cases of SS injection. Recalling that the regularisation coefficient α is selected to minimise the difference between the mid-time and terminal-time reconstructed fields and for de-noising in noisy cases, the DA reconstructions in figure 13(*e,f*) have a certain attenuation of the vorticity magnitude relative to the observations in figure 13(*b,c*) but are still considerably higher than those in the tomo-PIV field. These results are clear demonstrations of SS injection, with small-scale structures regenerated in the DA fields differing from those in the observations. The authenticity of these reproduced small-scale turbulence dynamics is thus evaluated subsequently.

The joint PDFs of the invariants Q^* and R^* of the observations and DA results are shown in figure 14. These invariant maps are computed according to (4.7) and (4.8). The invariant map of the clear LES data is plotted in each panel for comparison. It is observed

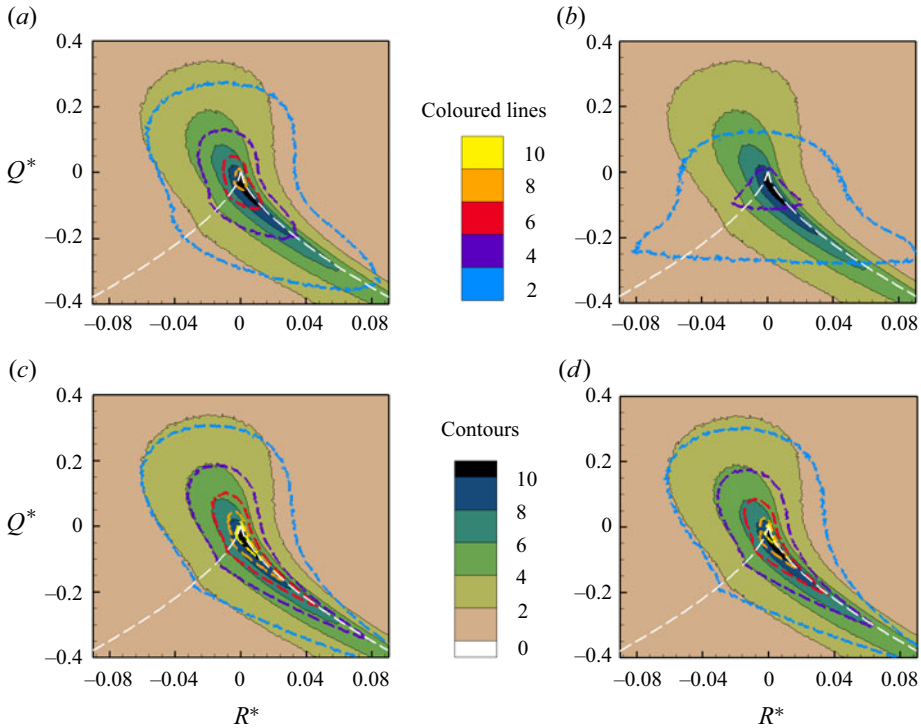


Figure 14. Joint PDFs of the velocity gradient invariables Q^* and R^* . The solid contours show the clear LES data, while the coloured dashed lines show (a) tomo-PIV observations with SS injection, (b) tomo-PIV observations with SS injection and noise, (c) the case of DA-SS and (d) the case of DA-SSN. White dashed lines in the third and fourth quadrants represent the zero-discriminant lines $Q^{*3} + 27R^{*3}/4 = 0$. The PDFs are normalised by the mean value on the Q^*-R^* plane.

that with SS injection, the tomo-PIV generates the teardrop shape of the invariant map much better than the original measurements (figure 14a). The Vieillefosse tails shift onto the right discriminant line but are still shorter than those of the clear LES data. The noise greatly deteriorates the map pattern such that the characteristic teardrop shape is not seen, as shown in figure 14(b). The induced error is prone to raise the high-PDF regions in the third and fourth quadrants above the discriminant lines, which is also seen in the original tomo-PIV data in figure 12(b). Despite the unphysical turbulence in the observations, DA is able to improve the dynamical feature substantially, resulting in the well-organised teardrop shapes of PDFs on the Q^*-R^* plane as shown in figure 14(c,d).

An important feature of velocity gradients in turbulence flows is the alignment of the vorticity along the eigenvectors of the strain-rate tensor S_{ij} (Ashurst *et al.* 1987). This alignment largely determines the vortex stretching term $S_{ij}\omega_i\omega_j$, which is the main source of the enstrophy growth. By examining the PDFs of the cosines of angles between the vorticity vector and the eigenvectors of S_{ij} (i.e. $|\hat{e}_1 \cdot \hat{\omega}|$, $|\hat{e}_2 \cdot \hat{\omega}|$ and $|\hat{e}_3 \cdot \hat{\omega}|$, where the eigenvectors \hat{e}_1 , \hat{e}_2 and \hat{e}_3 are arranged in descending order of the corresponding eigenvalues $\lambda_1 > \lambda_2 > \lambda_3$, and $\hat{\omega}$ denotes the normalised vorticity vector), the preferential alignment of $\hat{\omega}$ with \hat{e}_2 is observed. Checking the vorticity alignments in the different fields provides a useful means for assessing the effectiveness of the DA scheme in recovering the turbulence dynamic features. Figure 15 shows that the alignment of the vorticity vector with each eigenvector in the clear LES data has a tendency similar to that in a turbulent jet, as shown by Buxton & Ganapathisubramani (2010), and to that in spatially developing

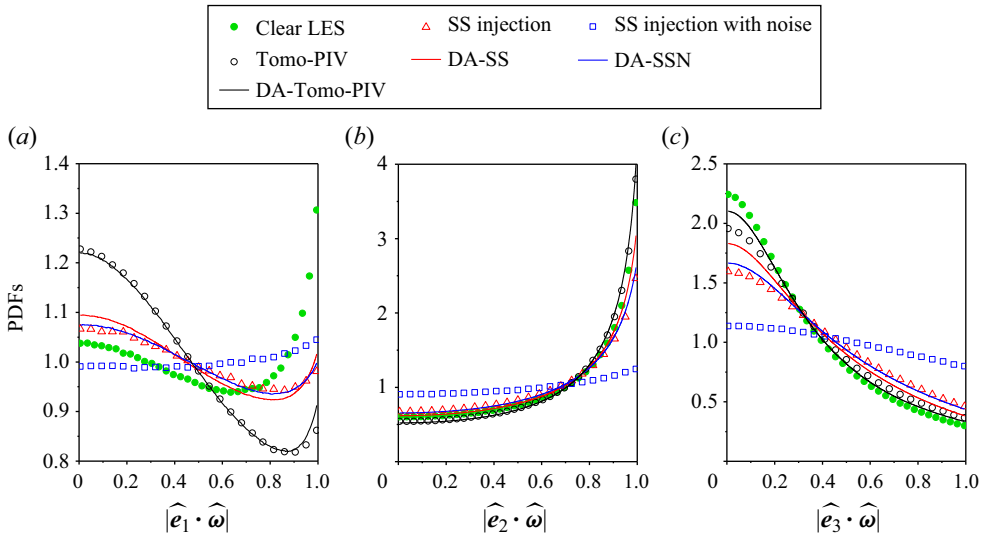


Figure 15. Alignment of the vorticity vectors with the eigenvectors of the strain-rate tensor. (a) Alignment with the first (extensive) eigenvector, (b) alignment with the second (intermediate) eigenvector and (c) alignment with the third (compressive) eigenvector. The ordinate scales differ across panels.

flow, as shown by Gomes-Fernandes, Ganapathisubramani & Vassilicos (2014), with the preferential alignment of the vorticity being along the second eigenvector. For other data in figure 15, there are certain discrepancies of the alignments compared with the clear LES. The vorticity alignments in the case of DA-tomo-PIV have a tendency almost identical to that of the original tomo-PIV results except for the third eigenvector, where there is a slight improvement, as shown in figure 15(c). This suggests that the DA procedure hardly changes the vortex orientation when using the tomo-PIV data as observations due to an absence of small-scale vortices. The SS injection improves the vorticity alignment with the first eigenvector, as shown in figure 15(a), whereas the case of DA-SS further augments the alignments with the second and third eigenvectors, as shown in figure 15(b,c). We see here evidence for the dynamic feature enhancement of small-scale turbulence structures; this is more conspicuous when there is noise in the observations, as the alignment is improved substantially in the case of DA-SSN from a rather poor tendency in the noisy observations.

Another important feature of turbulence flows is the modulation of the small-scale vortices by the large-scale gradients. Many studies have shown that large-scale structures modulate small scales both in amplitude and frequency (Mathis, Hutchins & Marusic 2009; Ganapathisubramani *et al.* 2012; Fiscaletti, Ganapathisubramani & Elsinga 2015) rather than the scales being independent of each other as had long been considered. In jet flows, the small-scale signal has a stronger amplitude for positive fluctuations of the large-scale signal independently of the radial location (Fiscaletti *et al.* 2015). There is also preferential alignment between the vorticities on the small and large scales, indicating that the vorticity direction does not vary appreciably across the scales (Fiscaletti *et al.* 2016). This body of research suggests that the anisotropy of the large scales is partially preserved at the small scale, which seems to be in contrast with the local isotropy hypothesis (Pope 2000). The small scales have been defined to be smaller than the Taylor length scale λ_T . Fiscaletti *et al.* (2016) reported that the scale modulation was weakened at larger scales but remained appreciable at scales up to $3\lambda_T$. Inspection of the scale modulation of the DA results is thus important in evaluating the DA scheme on the accurate recovery of the scale organisation in the cascade mechanism.

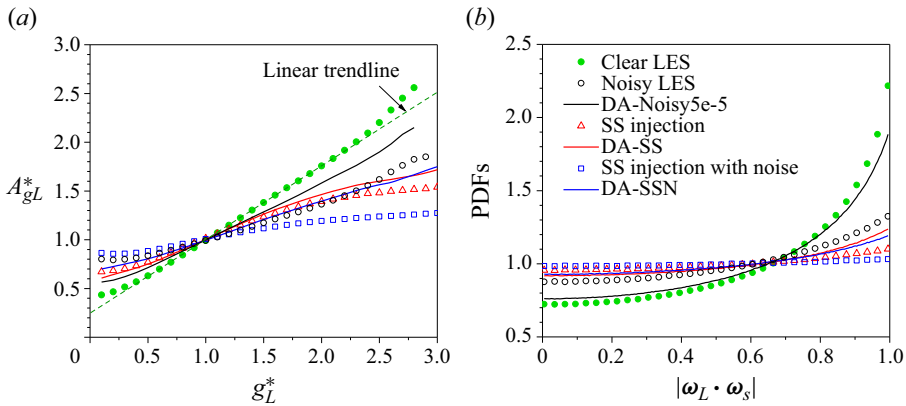


Figure 16. Modulation and alignment of vortices at different scales. (a) Local vorticity root mean square A_{gL}^* conditioned on the fluctuations of the large-scale gradients g_L^* . (b) Alignment of large-scale vorticity vectors with small-scale vorticity vectors.

Following Fiscaletti *et al.* (2016), the amplitude modulation is assessed by defining the local vorticity root mean square A_{gL}^* and the large-scale velocity gradient g_L^* as

$$A_{gL}^*(\mathbf{x}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (|\omega_i - \tilde{\omega}|)}, \quad (4.9)$$

$$g_L^*(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \sqrt{\left. \frac{\partial \tilde{u}_x}{\partial y} \right|_i^2 + \left. \frac{\partial \tilde{u}_x}{\partial z} \right|_i^2 + \left. \frac{\partial \tilde{u}_y}{\partial x} \right|_i^2 + \left. \frac{\partial \tilde{u}_y}{\partial z} \right|_i^2 + \left. \frac{\partial \tilde{u}_z}{\partial x} \right|_i^2 + \left. \frac{\partial \tilde{u}_z}{\partial y} \right|_i^2}. \quad (4.10)$$

Here, A_{gL}^* is a measure of small-scale vortical structures whereas $g_L^*(\mathbf{x})$ only includes the shear components of the large-scale velocity gradients. Both quantities are normalised by the corresponding mean values. The tilde denotes the large-scale quantities in each dataset, and N is the number of nodes in the filtering window where the large scales are computed. Herein, the definition of large scales is not straightforward, as simply imposing filters does not give a reasonable and clear separation of the different scales in the present jet. In the clear LES, noisy LES and DA-Noisy5e-5 results, the filtered LES results obtained using a filter size of $7 \times 7 \times 7$ elements on the DA grid (as shown in figure 11c) are taken as the large scales. The tomo-PIV original data are used as large scales for observations of SS injection and SS injection with noise. In the DA-SS and DA-SSN cases, the DA-Tomo results are used as large scales. With these definitions, the filter size is estimated to be approximately $3\lambda_T$ using the dissipation formulation of (3.1), and the scale modulation should still be appreciable according to the above discussion. Statistics of the activity at small scales conditioned on the large-scale gradients are shown in figure 16(a). Computations are performed in the region of $r/D \leq 0.75$. The results show that the clear LES results have a nearly linear correlation as obtained by Fiscaletti *et al.* (2016) for $0.5 < g_L^* < 2.5$. This confirms that large-scale gradients strongly modulate small-scale structures. Even with a high level of noise in observations with attenuated correlation (see ‘Noisy LES’), this modulation is substantially improved in the case of DA-Noisy5e-5. The accuracy of the modulation recovery by DA is indeed region-dependent, as much better recovery can be observed (not shown here) by checking only the jet shear layer region at $r/D \approx 0.5$. It is seen that in the case of DA-SS, there is notable improvement relative to

the SS-injected observations even with temporally sparse observational data. A loss of the scale modulation in the SS injection with and without noise is expected as the large- and small-scale structures are fully uncorrelated. Data assimilation improves the modulation properties regardless of the observational noise. This modulation recovery is an important criterion with which to assess the DA scheme in the reconstruction of real turbulence even when the small-scale structures are synthetic or contaminated with noise in measurement applications.

The modulation is also reflected in the vortex alignment between the small- and large-scale structures. The PDFs of the cosine of the angle between large and small scales (expressed as $|\omega_L \cdot \omega_S|$) are plotted in [figure 16\(b\)](#). It is seen that in the clear LES data, the small-scale vortices have a preferential alignment with the large-scale vortices, as demonstrated by the rapid increase in the PDFs as $|\omega_L \cdot \omega_S|$ approaches 1. We also see that this alignment is almost fully recovered even with noisy observations in the case of DA-Noisy5e-5. There is nearly no alignment correlation of different scales in the observations of SS injection and that with noise, with PDFs being almost horizontal in the range of $0 \leq |\omega_L \cdot \omega_S| \leq 1$. This is further evidence that the small- and large-scale structures are independent in the original SS-injected observations. However, the vortex alignment is clearly improved by DA with the PDFs increasing at large values of $|\omega_L \cdot \omega_S|$.

4.3. *Lagrangian particle tracking through STR reconstruction*

One of the motivations of the present study on STR reconstruction is to establish an auxiliary role for the successive prediction of particle positions in 4D-PTV. Four-dimensional PTV relies on the successive tracking of densely distributed particles at different instants to form long trajectories. This greatly increases the spatial resolution of the velocity vectors and improves the measurement accuracy by eliminating ghost particles that exist in tomo-PIV measurements, when using the same instrument configuration. The shake-the-box (STB) technique (Schanz, Gesemann & Schrder 2016) is a representative 4D-PTV technology that predicts the new particle positions after the trajectory initialisation, followed by position perturbation for the iterative correction. Physically constrained interpolation schemes are adopted to reconstruct velocity fields on an Euler grid (Gesemann *et al.* 2016; Schneiders & Scarano 2016). However, a high frequency of camera imaging (i.e. small displacements of the particles at two successive sampling instants) is required to identify the same particles in the image sequence (i.e. particle pairing) (Khojasteh *et al.* 2021). This requirement limits the use of the STB technique to low-speed flows for currently available measurement devices. Lagrangian particle tracking largely depends on the particle prediction for the next instant before the position correction (shaking). A Wiener filter (Wiener 1949) is used in the prediction phase of the STB technique, with the accuracy of prediction deteriorating drastically for short trajectories and large particle displacement (Schanz *et al.* 2016). Although a multi-pulse strategy using more cameras and lasers overcomes this difficulty by equivalently increasing the sampling rate and thus reducing the particle displacement (Manovski *et al.* 2021), we prefer less expensive methods based on STR reconstruction for artificially increasing the imaging frequency. This relies on the accurate recovery of the detailed turbulence between the two given snapshots, which can be tentatively determined by either tomo-PIV or conventional PTV using the double-exposure mode of the cameras. Although the practice of this measurement strategy is left as a topic of future study, Lagrangian particle tracking using the existing flow fields is evaluated here as a preliminary demonstration.

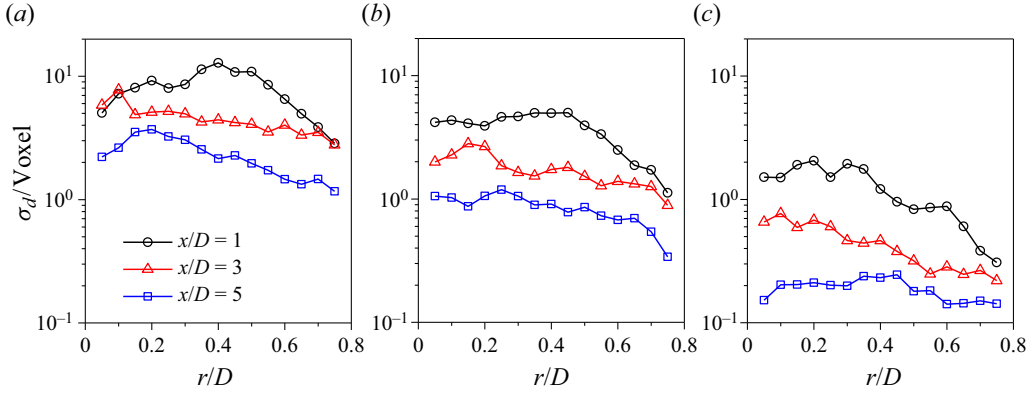


Figure 17. Particle prediction errors in the clear DA cases at the terminal instant: (a) window length $\Delta T = 30\Delta t$, (b) window length $\Delta T = 20\Delta t$ and (c) window length $\Delta T = 10\Delta t$. Results are averaged in the azimuthal direction.

Once the velocity fields are obtained, the new particle position \mathbf{x}_t is computed from the previous position \mathbf{x}_{t-1} and previous local velocity \mathbf{u}_{t-1} as

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{u}_{t-1} \cdot (\Delta t \text{ or } \Delta T). \quad (4.11)$$

The time step of the DA computation Δt or the time interval of the observations (DA window length) ΔT is used in (4.11) depending on the data used in different cases. This particle tracking formulation neglects various forces exerted on the particles by the fluid or adjacent particles; i.e. particles are distributed at a certain distance from each other and fully follow the flow as required for particle seeding in 4D-PTV measurements. We assume that the terminal field has been iteratively corrected in real experiments but includes a certain level of error. This implies that the terminal field does not need to be strictly accurate in using the present STR reconstruction. The particle prediction using the field of POD optimal reconstruction (Appendix C) and the Wiener prediction (Schanz *et al.* 2016; Tan *et al.* 2019) are evolved for comparison.

Prediction errors are evaluated at the terminal instants in different DA windows using the clear DA cases. This evaluation demonstrates the best prediction that the DA can achieve using the present STR reconstruction strategy. To compute the error, particles are initialised at different radial and streamwise locations before they are convected downstream until the terminal instant in the DA window. The errors are computed as the distance from the predicted particles to the true particles at the terminal instant, and they are normalised using the voxel size in the tomo-PIV measurement as mentioned in § 3.2. Figure 17 shows the azimuthally averaged prediction errors for different lengths of the DA window. The largest error is found to be 13 voxels in the upstream region when the window length $\Delta T = 30\Delta t$ as shown in figure 17(a). This clearly demonstrates that the error decreases as the location moves downstream and the window size decreases. For a window length $\Delta T = 20\Delta t$ as shown in figure 17(b), the error is smaller than 6 voxels at all the considered locations, and the error becomes less than 2 voxels in the downstream half-domain when the window length is reduced to $\Delta T = 10\Delta t$ as shown in figure 17(c).

Particle prediction is performed for each DA window independently in noisy cases as shown in figure 18; i.e. the particles are moved to the true positions (black dots on the true track) at each initial instant and the errors are evaluated at the terminal instant in a DA window. Data assimilation needs only two flow snapshots to compute the track whereas the Wiener prediction requires at least four previous positions for track initialisation and

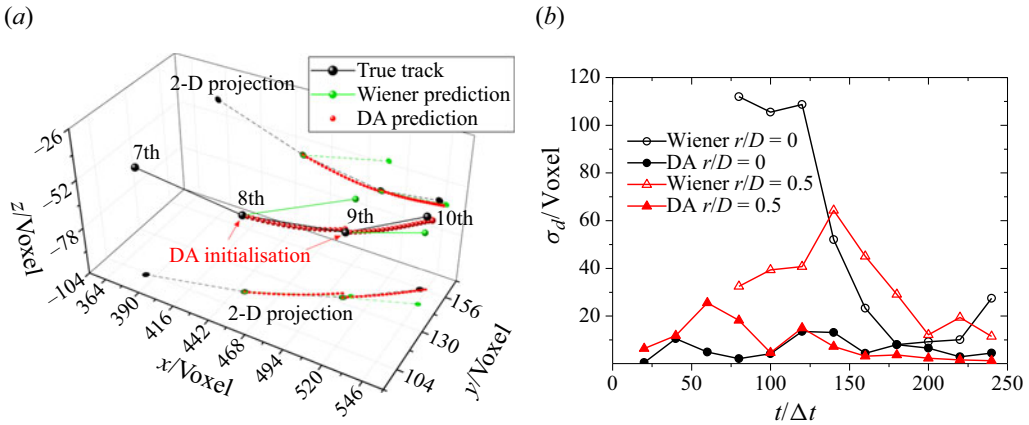


Figure 18. Particle prediction and position errors at different times: (a) three-dimensional tracking process and (b) errors at each terminal instant. The true track is computed using the clear LES data whereas the DA prediction is based on the case of DA-*Noisy5e-5*.

many more historical positions for improved accuracy. At time $t = 0$, two particles are initialised at $r/D = 0$ and $r/D = 0.5$ on the inflow surface. Figure 18(a) shows the seventh to the tenth true positions of the centreline particle. The true trajectory is computed using the clear LES data with time interval Δt in (4.11), but the time interval between two successive black dots shown in figure 18(a) is $\Delta T = 20\Delta t$. The fifth-order Wiener prediction is applied using all the previous true positions with a time interval ΔT , as we assume that only the down-sampled observations are known. It is seen that the Wiener prediction using the previous eight true positions determines the ninth position far from the truth; this also occurs for the prediction of the tenth position. The small red dots denote the DA prediction in the case of DA-*Noisy5e-5* with short-time marching realised through STR reconstruction. This prediction is thus more accurate, being much closer to the true positions. Figure 18(b) presents the error evolution with time. It is seen that the error in the DA prediction decays with time as the particle convects downstream. The largest error is approximately 30 voxels in the upstream region. The error decay in the DA prediction is mainly due to the increase in accuracy of the STR reconstruction in the downstream region. The Wiener prediction has an error greater than 100 voxels in the upstream region and decays as more historical information becomes available. However, the error in the Wiener prediction remains higher than that in the DA prediction even at the last several instants. Indeed, the Wiener prediction improves when there are more historical data, but it is not applicable as the particles exit the measurement domain on excessively long trajectories. The DA procedure is thus preferable over the Wiener filter for particle prediction with long time intervals.

Figure 19 shows the particle prediction error for conditions closer to those of a real experiment subject to various flow measurement errors. The computational procedures are similar to those used in obtaining the results in figure 17 but involve further averaging across all the DA windows. The noisy LES and tomo-PIV results are computed using the time-sparse flow data with the time interval $\Delta T = 20\Delta t$ as if the high-sampling-rate fields were not available, whereas all DA predictions use the time interval Δt . The case of DA-Clear20 is reproduced here as a lower bound for the present STR strategy, and the POD optimal reconstruction, which yields much higher particle error, is also shown for comparison. The particle prediction using the tomo-PIV fields, which has been widely

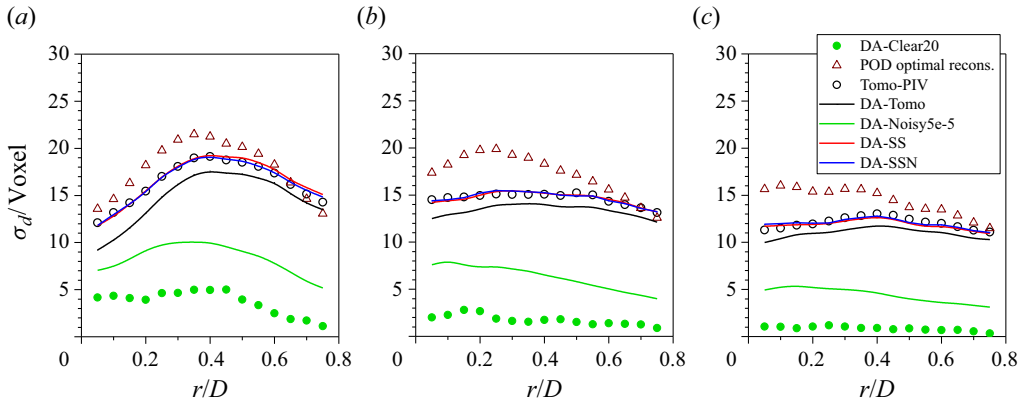


Figure 19. Particle prediction errors in different cases at the terminal instants of DA windows with $\Delta T = 20\Delta t$: (a) $x/D = 1$, (b) $x/D = 3$ and (c) $x/D = 5$. Results are averaged in the azimuthal direction and in different DA windows.

used in conventional PTV algorithms, has errors larger than 15 voxels in most of the measurement domain. The DA predictions in the cases of DA-SS and DA-SSN have errors similar to the error in the tomo-PIV results, whereas the DA prediction in the case of DA-Tomo has slightly less error owing to the STR reconstruction with a much smaller time step. Among all the discussed DA approaches, the case of DA-Noisy5e-5 is the most promising and has a particle prediction error much smaller than that of tomo-PIV used in conventional PTV algorithms. This DA case, with particle error smaller than 10 voxels, is thus much more suitable for Lagrangian particle tracking in a fluid with high particle density than the Wiener prediction. Additionally, this suggests that the uncorrelated SS injection does not benefit the particle tracking. Thus, correct small-scale structures should be included in the observations to improve the particle prediction; this can be done iteratively and is left for future work.

5. Conclusions

This study adopted 4D-Var-based data assimilation to reconstruct flow fields with high temporal resolution from time-sparse flow snapshots of a turbulent jet at Reynolds number $Re = 6000$. The highly resolved LES data were used to produce synthetic measurements that were used as observations and for validation. The DA procedure was decoupled into three phases to assimilate the initial condition, inflow boundary condition and model error separately. In the model error phase, 4D-Var was adopted by computing the primary and adjoint equations in a forward manner and backward manner, respectively, in a DA window between two observations with time interval ΔT , which was much longer than the computational time step Δt . The STR reconstruction could thus be realised after the computation. The 4D-Var was performed in each window successively with all the intermediate quantities stored in memory, and high computational efficiency was thus achieved together with a newly proposed ALSD technique.

Different types of observation were used to evaluate the STR reconstruction of the present DA method. The first type of observation was down-sampled LES data containing many small-scale turbulence structures with or without synthetic noise. These data were considered suitable for evaluating the capacity of the present DA method in the STR reconstruction of turbulence details beyond the Nyquist limit. For clear LES observations

with different window lengths, the flow instantaneous fields at the terminal instants were well recovered by DA, while the largest error was observed at the middle instant in each DA window. Even at the middle instant, the error was less than 25 % of that of the POD optimal reconstruction. Noisy cases (DA-Noisy5e-5, DA-Noisy5e-6 and DA-Noisy5e-7) demonstrated that the de-noising capacity of the present DA was strongly associated with the regularisation coefficient α . A good estimation of α was that it was approximately equal to or smaller than $1/\lambda_{\xi}^0$, where λ_{ξ}^0 is the initial steepest descent step length of the model error. It was found that $\alpha = 1/\lambda_{\xi}^0$ yielded a good balance between the noise removal and small-scale recovery in the noisy cases. The PDF of the streamwise component of the model error term ξ_x had a notable bias in the positive direction, providing good compensation for the pressure gradient difference between the truth and prediction. The temporal variation of the flow showed that small-scale structures were well recovered even with noise contaminating the observations. The spectra were resolved to a frequency approximately one order higher than that captured by the observations within the Nyquist limit.

The second type of observation was low-sampling-rate tomo-PIV data with or without SS injection. These data were obtained by synthesising the particle images from the LES results and then conducting a virtual experiment using the tomo-PIV algorithm. Additionally, small-scale structures were reconstructed by adding uncorrelated synthetic turbulence to the tomo-PIV results. The tomo-PIV results showed vortical structures different from those of the filtered LES data when using a similar filter size, although the statistical results were similar. This suggests that the PIV results were more complex than the spatially averaged results obtained from the ground truth. Data assimilation showed strong noise removal and divergence correction capacities on tomo-PIV data. Also, DA had the important feature of correcting the dynamic behaviours of the turbulence structures even when the small-scale structures were injected from uncorrelated external data. This feature was manifested by the reproduction of teardrop-shape joint PDF maps of the velocity gradient invariables and the preferential alignment of the vorticity vector with the second eigenvector of the strain-rate tensor. The modulation between the large-scale gradients captured by tomo-PIV and the small-scale structures injected externally was improved. The small-scale vortices exhibited remarkable alignment with the large-scale vortices after the DA procedure even though they were originally uncorrelated in the observations with SS injection.

Lagrangian particle tracking through STR reconstruction is a promising method for PTV conducted at a high flow speed. This study evaluated the particle prediction error in different DA cases. The results showed that in the clear DA cases, the largest particle error was approximately 13 voxels at a window length of $\Delta T = 30\Delta t$. The particle error decreased to 2 voxels at a window length of $\Delta T = 10\Delta t$. In particle prediction in noisy cases, the largest error was less than 25 % of the prediction using the Wiener filter with a window length of $\Delta T = 20\Delta t$. The particle prediction was not improved in tomo-PIV DA cases with the injection of small-scale structures. The present results suggest that the authenticity and correlation of the small-scale turbulence in observations are important to particle prediction. Such authenticity and correlation can be achieved adopting advanced measurement techniques, which is left as future work.

Funding. The authors are grateful for financial support from the National Natural Science Foundation of China (12272231, 12227803) and thank Professor H.J. Sung (KAIST) for his insightful discussion and comments on the manuscript revision.

Declaration of interests. The authors report no conflict of interest.

1. First guess for s at $t = t_0$.
2. Calculate the adjoint source $-2\gamma[s - \mathbb{H}[\mathbf{u}^{obs}(t_0)]]/U_0^2$.
3. Compute $\mathbf{v}(t_0)$ using (2.7) and (2.8) and $\mathbf{v}(t_0 + \Delta t) = 0$.
4. Update s using (2.18).
5. Calculate the residual $\varepsilon_s = ||s - \mathbb{H}[\mathbf{u}^{obs}(t_0)]||/||\mathbb{H}[\mathbf{u}^{obs}(t_0)]||$.
6. Stop in the case of convergence. Otherwise, use the new s and go to step 2.

Table 2. The DA procedure for the initial condition.

1. Set $\xi = 0$.
2. Extract $\mathbf{u}_{in}(t_0 + \Delta T)$ from $\mathbb{H}[\mathbf{u}^{obs}(t_0 + \Delta T)]$.
3. Set $n = 0$.
4. First guess for s at $t = t_0 + \Delta T - n\Delta t - \Delta t$.
5. Compute $\mathbf{u}(t_0 + \Delta T - n\Delta t)$ using (2.1) and (2.2) and s from step 4.
6. Calculate the adjoint source as $-2\gamma[\mathbf{u} - \mathbb{H}[\mathbf{u}^{obs}]]/U_0^2$ at $t = t_0 + \Delta T - n\Delta t$.
7. Compute $\mathbf{v}(t_0 + \Delta T - n\Delta t - \Delta t)$ using (2.7) and (2.8) and $\mathbf{v}(t_0 + \Delta T - n\Delta t + \Delta t) = 0$.
8. Update s at $t = t_0 + \Delta T - n\Delta t - \Delta t$ using (2.18).
9. Compute $\mathbf{u}(t_0 + \Delta T - n\Delta t)$ using (2.1) and (2.2) and s from step 8.
10. Calculate the residual $\varepsilon_s = ||\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})||/||\mathbb{H}(\mathbf{u}^{obs})||$ at $t = t_0 + \Delta T - n\Delta t$.
11. In the case of no convergence, go to step 7 and iterate.
12. In the case of convergence, at $t = t_0 + \Delta T - n\Delta t - \Delta t$, set $\mathbf{u}^{obs} = s$ and extract \mathbf{u}_{in} from s . Set $n = n + 1$ and go to step 4 for reverse recursion.
13. If $n = \Delta T/\Delta t - 1$, stop.

Table 3. The DA procedure for the inflow boundary condition.

Author ORCIDs.

-  Chuangxin He <https://orcid.org/0000-0001-7953-9380>;
-  Xin Zeng <https://orcid.org/0000-0002-9262-777X>;
-  Peng Wang <https://orcid.org/0000-0002-4589-0137>;
-  Xin Wen <https://orcid.org/0000-0002-6642-5955>.

Appendix A. Segregated DA procedures

The initial condition can be assimilated using the observation at the start time of the DA window. In each iterative loop, the adjoint velocity \mathbf{v} and adjoint pressure q are solved using (2.7) and (2.8), before the initial condition is updated using (2.18). In this strategy, the primary (2.1) and (2.2) are not solved and ξ is thus not required. The computational procedure is presented in table 2.

The above DA procedure assimilates a filtered initial field that fulfils the divergence-free condition and best fits the provided observation at the same instant. When this strategy is used for the assimilation of the flow realisation before the observation, the inflow boundary condition of the previous time step can be extracted from the assimilated initial field. This provides a computational strategy for assimilation of the inflow condition that is realisable in the present study. The computational procedure is presented in table 3. For the inflow DA procedure, the model error ξ is neglected because it does not cause an appreciable discrepancy for a short-time prediction. The historical inflow information strongly relates to the flow field immediately downstream of the inflow boundary. The adjoint equations propagate the primary flow information upstream to the inflow

1. Set $\xi = 0$.
2. Forward integrate equations (2.1) and (2.2) from $t = t_0 + \Delta t$ to $t_0 + \Delta T$.
3. Calculate the residual $\varepsilon_\xi = \|\mathbf{u} - \mathbb{H}(\mathbf{u}^{obs})\|/\|\mathbf{u}^{obs}\|$ at $t = t_0 + \Delta T$.
4. In the case of convergence, stop.
5. In the case of no convergence, calculate the adjoint source $-2\gamma[\mathbf{u} - \mathbb{H}[\mathbf{u}^{obs}]]/U_0^2$ at time $t_0 + \Delta T$.
6. Backward integrate equations (2.7) and (2.8) from $t_0 + \Delta T$ to t_0 using $\mathbf{v}(t_0 + \Delta T + \Delta t) = 0$.
7. Update ξ at each time step using (2.17).
8. Forward integrate (2.1) and (2.2) from $t_0 + \Delta t$ to $t_0 + \Delta T$.
9. Go to step 3.

Table 4. The DA procedure for the model error.

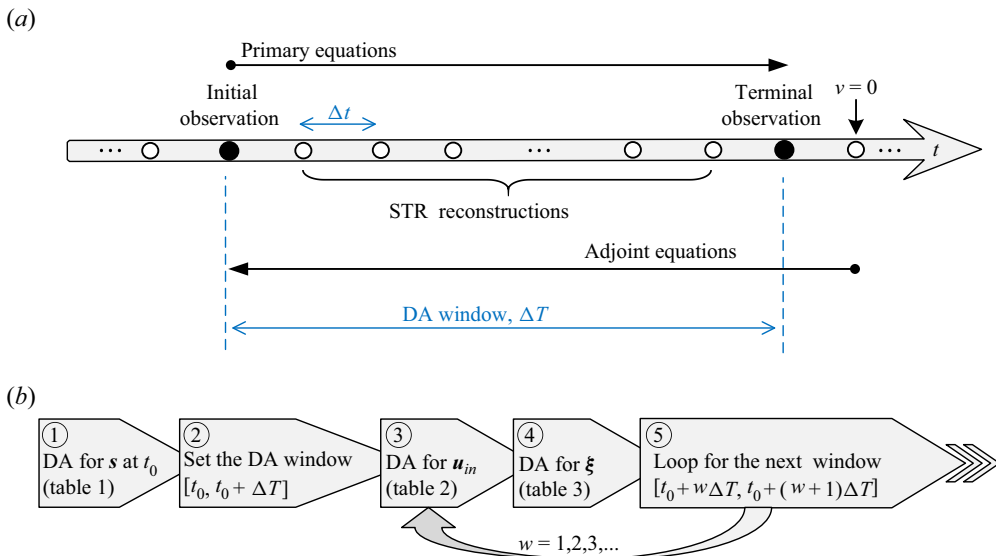


Figure 20. Schematic of the DA routine for STR reconstruction. (a) Computation process for the model error assimilation. (b) Routine for multi-window batch processing.

surface and thus determine the inflow condition in the manner of time reversal recursion. However, the error accumulation causes a certain discrepancy for early-time flow recovery; this is extensively assessed in Appendix B.

The model error ξ can be assimilated after the determination of the initial and inflow boundary conditions. This is done through the forward and backward integration of the primary and adjoint equations, respectively, in a selected DA window from time t_0 to $t_0 + \Delta T$. The computational procedure is presented in table 4. There are observations only at the initial and terminal instants of the DA window with time interval ΔT , as shown in figure 20(a). The time step of the DA computation Δt is at least one order of magnitude smaller than ΔT . The terminal time for solving the adjoint equations is shifted to $t_0 + \Delta T + \Delta t$, at which instant the terminal condition $\mathbf{v} = 0$ is applied; this treatment avoids ξ being zero when computing the primary equations at $t_0 + \Delta T$. The adjoint source is only applied at instant $t_0 + \Delta T$ and the adjoint flow develops freely with upstream advection. The STR reconstruction is thus achieved by obtaining the flow instantaneous fields with time interval Δt .

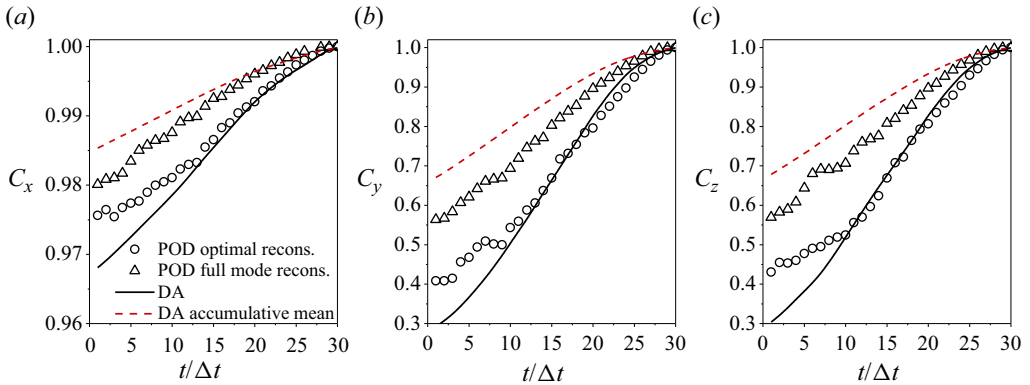


Figure 21. Cross-correlation coefficients of the reconstructed velocity fields with the clear LES data on the inflow boundary for the case DA-Clear30: (a) x component, (b) y component and (c) z component. Different ordinate scales are used in the panels for clear illustration.

Figure 20(a) shows the forward and backward integrations of the present DA algorithm for STR reconstruction. Prior to this, the DA for initial and inflow conditions should be performed. It is noted that the DA for the initial condition needs be conducted only at the start time of the first window, while the terminal field is used to initialise the computation in the next adjacent DA window. The routine for the multi-window batch processing is shown in figure 20(b), with the notation w increasing to iterate over all DA windows. It is noted that this segregated DA strategy is necessary for the present application with time-sparse observations, as simultaneous assimilation of all the quantities would result in an indeterminacy problem in which the residual evaluated at the terminal step is insensitive to the initial and boundary conditions. The present 4D-Var algorithm is indeed different from the sequential DA scheme in our previous work (He *et al.* 2020, 2022), where the backward integration was eliminated by limiting the DA window to one computational time step. Sequential DA solves the adjoint equations only at the instants that the observations exist and induces discontinuous variation of the flow quantities at the observational time. It is thus inappropriate for the purpose of STR reconstruction. Nevertheless, sequential DA is particularly suited to small-data-driven simulations with much more confidence on the predictive model.

Appendix B. Evaluation on inflow boundary condition

The inflow boundary condition is obtained through reverse recursion starting from a terminal time to the start time $t_0 = 0$ using the algorithm in table 3. The accuracy of the inflow field reconstructions is quantitatively assessed on the inflow surface using the cross-correlation coefficient defined in (4.3). The correlations of the inflow field reconstructed by DA in the case of DA-Clear30 with the clear LES data are shown in figure 21 by the black solid curves. It is seen that the correlations approach 1 close to the terminal time, i.e. $t \rightarrow 30\Delta t$. With the reverse recursion, the correlations gradually decrease as time marches from $t = 30\Delta t$ to 0. This error most probably results from neglecting the model error in the inflow DA phase. For the present window length $\Delta T = 30\Delta t$, the overall evaluation of the correlation should be the accumulative mean computed by averaging each correlation coefficient throughout the DA window. Common sense suggests that the overall correlation increases with the increasing sampling rate of the observations. If the start time is selected at t instead of time zero, the length of the DA window decreases to $\Delta T - t$. The accumulative mean should be computed by averaging

over the range $[t, \Delta T]$. By varying the start time t , the accumulative mean is obtained as shown by the red dashed curves. The accumulative mean is larger than the correlation coefficient (black solid curve) due to the strong correlation of the inflow data near the terminal time which has been averaged in the accumulation. This accumulative mean is thus more appropriate for the quality assessment of the inflow condition reconstructed by DA. It is noted that the correlation in the x direction remains appreciably higher than those in other directions owing to the jet mainstream.

In POD reconstructions, we select the clear LES data with the time interval $\Delta T - t$, and STR can be achieved by interpolating the model coefficients before the reconstruction. In the POD optimal reconstruction, the optimal number of leading modes is used for the reconstruction (see Appendix C). The POD full mode reconstruction uses all the modes and is equivalent to direct flow interpolation. The correlation of the inflow field reconstructed using the clear LES data with different time intervals can be obtained by varying time t , such that all the curves can be plotted in the same figure together with the DA results. Figure 21 shows that the correlations of the POD optimal reconstruction almost overlap with those of the DA results but become appreciably higher for $t/\Delta t < 10$. This suggests that the model error is dominant for DA-based inflow reconstruction at an early time. However, the accumulative mean of the correlation remains appreciably higher than those of all other reconstruction schemes. This suggests a better performance of the present DA scheme in the determination of the inflow boundary condition compared with solely data-driven POD reconstruction. Nevertheless, the POD full reconstruction using all the modes has appreciably stronger correlation than that of the optimal reconstruction. Although this reconstruction cannot preserve the convective properties owing to the randomness of the small-scale structures, it can still be considered as an alternative choice for inflow boundary conditions, especially in cases in which observations contain large errors and noise.

Figure 21 clearly demonstrates that when the time interval of the observations is longer than $5\Delta t$ ($t/\Delta t < 25$) in the cases of DA-Clear10, DA-Clear20 and DA-Clear30, the present DA procedure determines the inflow condition much better than POD-based reconstruction methods. This also holds in the cases of DA-Noisy5e-5, DA-Noisy5e-6 and DA-Noisy5e-7 by further calculation. Therefore, the DA reconstructions of inflow fields are used in obtaining all the results presented in § 4.1. However, the POD full model reconstruction is adopted to produce the inflow condition for the results discussed in § 4.2, as the DA approach does not produce satisfactory results in the tomo-PIV cases.

Appendix C. Proper orthogonal decomposition optimal reconstruction

Super-temporal resolution can be also achieved by interpolating the temporal coefficients a_i of the POD (Sirovich 1987; Lumley 2012) mode in time before the reconstruction. However, the fixed sampling of the observational data gives rise to chaotic variation of the temporal coefficients for the high-order modes. Direct interpolation of these coefficients incorrectly captures the variation of the flow; thus, the high-order mode coefficients are usually discarded in the reconstruction. The POD optimal reconstruction provides the best choice of the POD modes in the reconstruction procedure. The highest order of the POD modes can be determined by the requirement that the coefficient has appreciable autocorrelation in time with the specified sampling rate. For the i th-order mode, the autocorrelation C_{aa} is calculated as

$$C_{aa}(i) = \int \frac{a_i(t)a_i(t + \Delta T)}{a_i^T a_i} dt. \quad (C1)$$

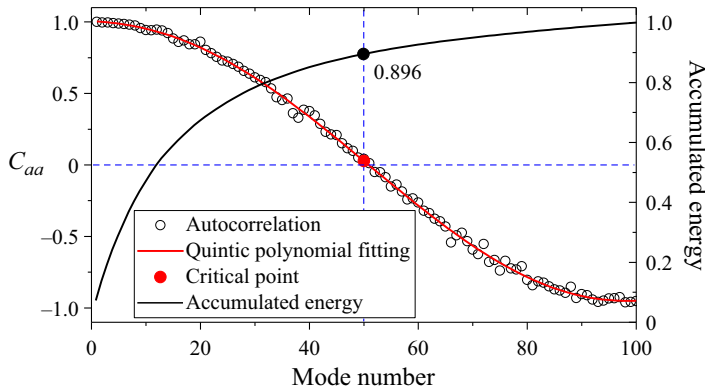


Figure 22. Autocorrelation of the mode coefficients and the accumulated energy for the clear LES data on the inflow boundary.

Figure 22 shows the autocorrelation of the model coefficients for the clear LES data on the inflow boundary. Other datasets can also be used for such computation. Model 0 denotes the mean flow, the autocorrelation C_{aa} of which equals 1. Autocorrelation C_{aa} decreases with an increasing mode number. Large C_{aa} indicates that the mode coefficients vary in time smoothly, and interpolation can thus be performed for STR reconstruction. The critical point $i = 50$ is defined, through a quintic polynomial fitting of the autocorrelation data, as the largest mode number above which C_{aa} becomes negative. We thus use the leading 50 POD modes for the optimal reconstruction. The critical point varies for different datasets and should be determined separately. In the present case, the optimal reconstruction recovers only 89.6% of the total kinetic energy. Using all of the POD modes leads to the POD full mode reconstruction, which is equivalent to direct interpolation on the flow fields.

REFERENCES

- ADRIAN, R.J. & MOIN, P. 1988 Stochastic estimation of organized turbulent structure: homogeneous shear flow. *J. Fluid Mech.* **190**, 531–559.
- ARTANA, G., CAMMILLERI, A., CARLIER, J. & MÉMIN, E. 2012 Strong and weak constraint variational assimilations for reduced order fluid flow modeling. *J. Comput. Phys.* **231**, 3264–3288.
- ASHURST, W.T., KERSTEIN, A.R., KERR, R.M. & GIBSON, C.H. 1987 Alignment of vorticity and scalar gradient with strain rate in simulated Navier–Stokes turbulence. *Phys. Fluids* **30**, 2343.
- BAO, W., LAI, W.S., ZHANG, X., GAO, Z. & YANG, M.H. 2021 MEMC-Net: motion estimation and motion compensation driven neural network for video interpolation and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 993–948.
- BENNETT, A.F. 2002 *Inverse Modeling of the Ocean and Atmosphere*. Cambridge University Press.
- BRENNER, M.P., ELDRIDGE, J.D. & FREUND, J.B. 2019 Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids* **4**, 100501.
- BRUNTON, S., NOACK, B.R. & KOUMOUTSAKOS, P. 2019 Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508.
- BUXTON, O. & GANAPATHISUBRAMANI, B. 2010 Amplification of enstrophy in the far field of an axisymmetric turbulent jet. *J. Fluid Mech.* **651**, 483–502.
- CHANDRAMOULI, P., MEMIN, E. & HEITZ, D. 2020 4D large scale variational data assimilation of a turbulent flow with a dynamics error model. *J. Comput. Phys.* **412**, 109446.
- CHI, Z., NASIRI, R.M., LIU, Z., LU, J. & PLATANOTIS, K.N. 2020 All at once: temporally adaptive multi-frame interpolation with advanced motion modeling. In *Computer Vision – ECCV 2020* (ed. A. Vedaldi, H. Bischof, T. Brox & J.M. Frahm), Lecture Notes in Computer Science, pp. 107–123. Springer.
- DENG, Z., CHEN, Y., LIU, Y. & KIM, K.C. 2019 Time-resolved turbulent velocity field reconstruction using a long short-term memory (LSTM)-based artificial intelligence framework. *Phys. Fluids* **31**, 075108.

- DURAISAMY, K., IACCARINO, G. & XIAO, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 1–23.
- DURGESH, V. & NAUGHTON, J.W. 2010 Multi-time-delay LSE-POD complementary approach applied to unsteady high-Reynolds-number near wake flow. *Exp. Fluids* **49**, 571–583.
- EVENSEN, G., VOSSEPOEL, F.C. & LEEUWEN, P.J. 2022 *Data Assimilation Fundamentals: A Unified Formulation of the State and Parameter Estimation Problem*. Springer.
- FERZIGER, J.H., PERIC, M. & STREET, R.L. 2020 *Computational Methods for Fluid Dynamics*, 4th edn. Springer.
- FISCALETTI, D., ATTILI, A., BISETTI, F. & ELSINGA, G.E. 2016 Scale interactions in a mixing layer - the role of the large-scale gradients. *J. Fluid Mech.* **791**, 154–173.
- FISCALETTI, D., GANAPATHISUBRAMANI, B. & ELSINGA, G.E. 2015 Amplitude and frequency modulation of the small scales in a jet. *J. Fluid Mech.* **772**, 756–783.
- FISCALETTI, D., RAGNI, D., OVERMARS, E.F.J., WESTERWEEL, J. & ELSINGA, G.E. 2022 Tomographic long-distance μ PIV to investigate the small scales of turbulence in a jet at high Reynolds number. *Exp. Fluids* **63**, 9.
- 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.
- FUKAMI, K., FUKAGATA, K. & TAIRA, K. 2019 Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120.
- FUKAMI, K., FUKAGATA, K. & TAIRA, K. 2021 Machine learning based spatio-temporal super resolution reconstruction of turbulent flows. *J. Fluid Mech.* **909**, A9.
- GANAPATHISUBRAMANI, B., HUTCHINS, N., MONTY, J.P., CHUNG, D. & MARUSIC, I. 2012 Amplitude and frequency modulation in wall turbulence. *J. Fluid Mech.* **712**, 61–91.
- GESEMANN, S., HUH, F., SCHANZ, D. & SCHRODER, A. 2016 From noisy particle tracks to velocity, acceleration and pressure fields using B-splines and penalties. In *18th International Symposium on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal*.
- GIANNOPOULOS, A. & AIDER, J.L. 2020 Prediction of the dynamics of a backward-facing step flow using focused time-delay neural networks and particle image velocimetry data-sets. *Intl J. Heat Fluid Flow* **82**, 108533.
- GILLISSEN, J., BOUFFANAIS, R. & YUE, D. 2019 Data assimilation method to de-noise and de-filter particle image velocimetry data. *J. Fluid Mech.* **877**, 196–213.
- GOMES-FERNANDES, R., GANAPATHISUBRAMANI, B. & VASSILICOS, J. 2014 Evolution of the velocity-gradient tensor in a spatially developing turbulent flow. *J. Fluid Mech.* **756**, 252–292.
- GRONSKIS, A., HEITZ, D. & MÉMIN, E. 2013 Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation. *J. Comput. Phys.* **242**, 480–497.
- GUEMES, A., DISCETTI, S. & IANIRO, A. 2019 Sensing the turbulent large-scale motions with their wall signature. *Phys. Fluids* **31**, 125112.
- HE, C., LIU, Y. & GAN, L. 2018a A data assimilation model for turbulent flows using continuous adjoint formulation. *Phys. Fluids* **30**, 105108.
- HE, C., LIU, Y. & GAN, L. 2020 Instantaneous pressure determination from unsteady velocity fields using adjoint-based sequential data assimilation. *Phys. Fluids* **32**, 035101.
- HE, C., LIU, Y. & YAVUZKURT, S. 2018b Large-eddy simulation of circular jet mixing: lip- and inner-ribbed nozzles. *Comput. Fluids* **168**, 245–264.
- HE, C., WANG, P. & LIU, Y. 2021 Data assimilation for turbulent mean flow and scalar fields with anisotropic formulation. *Exp. Fluids* **62**, 117.
- HE, C., WANG, P., LIU, Y. & GAN, L. 2022 Flow enhancement of tomographic particle image velocimetry measurements using sequential data assimilation. *Phys. Fluids* **34**, 035101.
- HOSSEINI, Z., MARTINUZZI, R.J. & NOACK, B.R. 2015 Sensor-based estimation of the velocity in the wake of a low-aspect-ratio pyramid. *Exp. Fluids* **56**, 13.
- HUDY, L.M., NAGUIB, A. & HUMPHREYS, W.M. 2007 Stochastic estimation of a separated-flow field using wall-pressure-array measurements. *Phys. Fluids* **19**, 024103.
- HUSSAIN, A.K.M.F. & ZAMAN, K.B.M.Q. 2006 The ‘preferred mode’ of the axisymmetric jet. *J. Fluid Mech.* **110**, 39–71.
- JIN, X., LAIMA, S., CHEN, W.L. & LI, H. 2020 Time-resolved reconstruction of flow field around a circular cylinder by recurrent neural networks based on non-time-resolved particle image velocimetry measurements. *Exp. Fluids* **61**, 114.
- KERHERVE, F., ROUX, S. & MATHIS, C.R. 2017 Combining time-resolved multi-point and spatially-resolved measurements for the recovering of very-large-scale motions in high Reynolds number turbulent boundary layer. *Exp. Therm. Fluid Sci.* **82**, 102–115.

- KHASHEHCHI, M., ELSINGA, G.E., OOI, A., SORIA, J. & MARUSIC, I. 2010 Studying invariants of the velocity gradient tensor of a round turbulent jet across the turbulent/nonturbulent interface using Tomo-PIV. In *15th Intl Symp. on Applications of Laser Techniques to Fluid Mechanics, Lisbon, Portugal*.
- KHOJASTEH, A.R., YANG, Y., HEITZ, D. & LAIZET, S. 2021 Lagrangian coherent track initialisation (LCTI). *Phys. Fluids* **33**, 095113.
- KIM, H., KIM, J., WON, S. & LEE, C. 2021 Unsupervised deep learning for super-resolution reconstruction of turbulence. *J. Fluid Mech.* **910**, A29.
- KIM, J. & LEE, C. 2020 Prediction of turbulent heat transfer using convolutional neural networks. *J. Fluid Mech.* **882**, A18.
- LAWSON, J.M. & DAWSON, J.R. 2014 A scanning PIV method for fine-scale turbulence measurements. *Exp. Fluids* **55**, 1857.
- LEE, S. & YOU, D. 2019 Data-driven prediction of unsteady flow over a circular cylinder using deep learning. *J. Fluid Mech.* **879**, 217–254.
- LEMKE, M. 2015 *Adjoint Based Data Assimilation in Compressible Flows with Application to Pressure Determination from PIV Data*. Technischen Universität Berlin.
- LI, Y., ZHANG, J., DONG, G. & ABDULLAH, N.S. 2019 Small scale reconstruction in three-dimensional Kolmogorov flows using four-dimensional variational data assimilation. *J. Fluid Mech.* **885**, A9.
- LIU, B., TANG, J., HUANG, H. & LU, X.Y. 2020 Deep learning methods for super-resolution reconstruction of turbulent flows. *Phys. Fluids* **32**, 025105.
- LUMLEY, J.L. 2012 *Stochastic Tools in Turbulence*. Academic Press.
- MANOVSKI, P., NOVARA, M., MOHAN, N., GEISLER, R. & SCHRDER, A. 2021 3D Lagrangian particle tracking of a subsonic jet using multi-pulse shake-the-box. *Exp. Therm. Fluid Sci.* **123**, 110346.
- MATHIS, R., HUTCHINS, N. & MARUSIC, I. 2009 Large-scale amplitude modulation of the small-scale structures in turbulent boundary layers. *J. Fluid Mech.* **628**, 311–337.
- MONS, V. & MARQUET, O. 2021 Linear and nonlinear sensor placement strategies for mean-flow reconstruction via data assimilation. *J. Fluid Mech.* **923**, A1.
- NAKA, Y., TOMITA, K., SHIMURA, M., FUKUSHIMA, N., TANAHASHI, M. & MIYAUCHI, T. 2016 Quad-plane stereoscopic PIV for fine-scale structure measurements in turbulence. *Exp. Fluids* **57**, 63.
- NIKLAUS, S., MAI, L. & LIU, F. 2017 Video frame interpolation via adaptive convolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA*.
- NÓVOA, A. & MAGRI, L. 2022 Real-time thermoacoustic data assimilation. *J. Fluid Mech.* **948**, A35.
- OTHMER, C. 2008 A continuous adjoint formulation for the computation of topological and surface sensitivities of ducted flows. *Intl J. Numer. Meth. Fluids* **58**, 861–877.
- PANCHAPAKESAN, N.R. & LUMLEY, J.L. 1993 Turbulence measurements in axisymmetric jets of air and helium. Part 1. air jet. *J. Fluid Mech.* **246**, 197–223.
- PAPOUTSIS-KIACHAGIAS, E.M. & GIANNAKOGLU, K.C. 2016 Continuous adjoint methods for turbulent flows, applied to shape and topology optimization: industrial applications. *Arch. Comput. Meth. Engng* **23**, 255–299.
- PARK, J. & CHOI, H. 2020 Machine-learning-based feedback control for drag reduction in a turbulent channel flow. *J. Fluid Mech.* **904**, A24.
- PINO, F., SCHENA, L., RABAULT, J. & MENDEZ, M.A. 2023 Comparative analysis of machine learning methods for active flow control. *J. Fluid Mech.* **958**, A39.
- PODVIN, B., NGUIMATSIA, S., FOUCAUT, J.-M., CUVIER, C. & FRAIGNEAU, Y. 2018 On combining linear stochastic estimation and proper orthogonal decomposition for flow reconstruction. *Exp. Fluids* **59**, 58.
- POPE, S.B. 2000 *Turbulent Flows*. Cambridge University Press.
- RISHITA, D. & GIRIMAJI, S.S. 2022 The effect of large-scale forcing on small-scale dynamics of incompressible turbulence. *J. Fluid Mech.* **941**, A34.
- SCHANZ, D., GESEMANN, S. & SCHRDER, A. 2016 shake-the-box: Lagrangian particle tracking at high particle image densities. *Exp. Fluids* **57**, 70.
- SCHNEIDERS, J.F.G. & SCARANO, F. 2016 Dense velocity reconstruction from tomographic PTV with material derivatives. *Exp. Fluids* **57**, 139.
- SIRIGNANO, J. & MACART, J.F. 2023 Deep learning closure models for large-eddy simulation of flows around bluff bodies. *J. Fluid Mech.* **966**, A26.
- SIROVICH, L. 1987 Turbulence and the dynamics of coherent structures. Part I: coherent structures. *Q. Appl. Maths* **45**, 561–571.
- STEFANESCU, R., SANDU, A. & NAVON, I.M. 2015 POD/DEIM reduced-order strategies for efficient four dimensional variational data assimilation. *J. Comput. Phys.* **295**, 569–595.
- TAN, S., SALIBINDLA, A., MASUK, A.U.M. & NI, R. 2019 An open-source shake-the-box method and its performance evaluation. In *13th International Symposium on Particle Image Velocimetry, Munich, Germany*.

Four-dimensional variational data assimilation

- TAN, S., SALIBINDLA, A., MASUK, A.U.M. & NI, R. 2020 Introducing OpenLPT: new method of removing ghost particles and high-concentration particle shadow tracking. *Exp. Fluids* **61**, 47.
- TINNEY, C.E., UKEILEY, L.S. & GLAUSER, M.N. 2008 Low-dimensional characteristics of a transonic jet. Part 2. Estimate and far-field prediction. *J. Fluid Mech.* **615**, 53–92.
- TREMOLET, Y. 2007 Model-error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.* **133**, 1267–1280.
- TU, J.H., GRIFFIN, J., HART, A., ROWLEY, C.W., CATTAFESTA, L.N. & UKEILEY, L.S. 2013 Integration of non-time-resolved PIV and time-resolved velocity point sensors for dynamic estimation of velocity fields. *Exp. Fluids* **54**, 1429.
- VIDARD, P.A., BLAYO, E., DIMET, F. & PIACENTINI, A. 2000 4D variational data analysis with imperfect model. *Flow Turbul. Combust.* **65**, 489–504.
- VIEILLEFOSSE, P. 1984 Internal motion of a small element of fluid in an inviscid flow. *Physica A: Stat. Mech. Applics.* **125**, 150–162.
- WANG, G. & PAN, Y. 2021 Phase-resolved ocean wave forecast with ensemble-based data assimilation. *J. Fluid Mech.* **918**, A19.
- WANG, Q., WANG, M. & ZAKI, T.A. 2022 What is observable from wall data in turbulent channel flow? *J. Fluid Mech.* **941**, A48.
- WATERSON, N.P. & DECONINCK, H. 2007 Design principles for bounded higher-order convection schemes – a unified approach. *J. Comput. Phys.* **224**, 182–207.
- WELCH, P.D. 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **AU-15**, 70–73.
- WIENER, N. 1949 *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press.
- ZAUNER, M., MONS, V., MARQUET, O. & LECLAIRE, B. 2022 Nudging-based data assimilation of the turbulent flow around a square cylinder. *J. Fluid Mech.* **937**, A38.
- ZENG, X., HE, C. & LIU, Y. 2022 GPU-accelerated MART and concurrent cross-correlation for tomographic PIV. *Exp. Fluids* **63**, 91.