JM PAPERS

State estimation in turbulent channel flow from limited observations

Mengze Wang[1] **and Tamer A. Zaki**[1],†

[1]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

Estimation of the initial state of turbulent channel flow from limited data is investigated using an adjoint-variational approach. The data are generated from a reference direct numerical simulation that is subsampled at different spatiotemporal resolutions. When the velocity data are at 1/4096 the spatiotemporal resolution of the direct numerical simulation, the correlation coefficient between the true and adjoint-variational estimated state exceeds 99 %. The robustness of the algorithm to observation noise is demonstrated. In addition, the impact of the spatiotemporal density of the data on estimation quality is evaluated, and a resolution threshold is established for a successful reconstruction. The critical spanwise data resolution is proportional to the Taylor microscale, which characterizes the domain of dependence of an observation location. Owing to mean advection, either the streamwise or temporal data resolution must satisfy a criterion based on the streamwise Taylor microscale. A second configuration is considered where the subsampled data comprise velocities in the outer layer and wall shear stresses only. The near-wall flow statistics and coherent structures, although not sampled, are accurately reconstructed, which is possible because of the coupling between the outer flow and near-wall motions. Finally, the most challenging configuration is addressed where only the spatiotemporally resolved wall stresses are observed. The estimation remains accurate within the viscous sublayer and deteriorates significantly with distance from the wall. In wall units, this trend is nearly independent of the Reynolds number considered, and is indicative of the fundamental difficulty of reconstructing wall-detached motions from wall data.

Key words: turbulence simulation, turbulent boundary layers

1. Introduction

Estimation of instantaneous turbulent flows from the assimilation of limited observations is a challenging problem due to the chaotic nature of turbulence (Kim & Bewley 2007).

† Email address for correspondence: t.zaki@jhu.edu

Given flow-field information with limited resolution, such as particle image velocimetry (PIV) data or pressure measurements, there are potentially multiple solutions that satisfy the Navier–Stokes solutions and match the observations. In addition, a small error in the initial state or boundary conditions will amplify exponentially in time, and thus the estimated state will diverge from the true one (Deissler 1986). Adjoint-variational methods address the state estimation problem by constructing an optimal initial condition that generates a trajectory in state space as close to the observations as possible. In this work, we evaluate the accuracy of the adjoint-variational approach for estimating turbulent channel flow and the dependence of estimation quality on the locations and resolution of the observations.

Three classes of state estimation techniques have been applied to flow problems: linear stochastic estimation (LSE) (e.g. Adrian & Moin 1988; Naguib, Morrison & Zaki 2010; Encinar & Jiménez 2019), filtering and smoothing (Law, Stuart & Zygalakis 2015). LSE utilizes prior knowledge of two-point correlation to estimate the flow state from observations. The correlation, which is not always available in practice, sets an upper bound on the estimation accuracy. In addition, LSE does not satisfy the Navier–Stokes equations and, as such, does not provide a trajectory in state space.

Filtering, or sequential, techniques consist of a prediction step, which involves marching the governing equations until the observation time, and an update step, where the prediction is augmented with observations (Evensen 1994). When the governing equations are linear, the optimal weight for the observations can be analytically derived, and the corresponding method is the so-called Kalman filter, which has been adopted for estimating the disturbance of laminar flows (Hœpffner *et al.* 2005). For nonlinear problems, the weight can be calculated by either linearizing the governing equations (extended Kalman filter) or marching an ensemble of different states in time (ensemble Kalman filter). Both of these methods have been evaluated for estimating turbulent channel flow from wall observations (Chevalier *et al.* 2006; Colburn, Cessna & Bewley 2011; Suzuki & Hasegawa 2017). Ultimately, the accuracy of the filtering techniques is limited because they only focus on fitting the data at one moment rather than a time interval (Kim & Bewley 2007). Also, the filtered state may not satisfy the Navier–Stokes equations due to observation noise and the difference between estimation and observations.

Smoothing techniques utilize a time series of data to search for the optimal initial condition, boundary conditions and model parameters which ensure that the evolution of the predicted state reproduces available data. Therefore, an accurate forecast of the flow evolution beyond the observation window is possible. This class of techniques is also capable of optimizing sensor placement and weighting in order to achieve the best prediction accuracy (see e.g. Mons, Chassaing & Sagaut 2017; Mons, Wang & Zaki 2019; Buchta & Zaki 2020). Mons *et al.* (2016) compared three of the most popular smoothing techniques: the adjoint-variational method (referred to as 4DVar in numerical weather prediction (Le Dimet & Talagrand 1986)), the ensemble Kalman smoother, and the ensemble variational method. The objective was to estimate the unsteady free-stream condition for laminar flow around a cylinder, and 4DVar achieved the lowest estimation error, for a specified computational cost. Adjoint techniques were also demonstrated to be viable in transitional (Mao, Blackburn & Sherwin 2013; Mao *et al.* 2017) and turbulent flows (Bewley & Protas 2004; Vishnampet, Bodony & Freund 2015), including, for example, for estimating scalar sources from remote observations (Cerizza *et al.* 2016; Wang, Hasegawa & Zaki 2019*b*). Wang, Wang & Zaki (2019*a*) derived the discrete adjoint of the incompressible Navier–Stokes equations in general curvilinear coordinates and applied it to estimating the turbulent state of circular Couette flow; they demonstrated

accuracy of the forward-adjoint relation to within eight significant figures. We herein adopt the adjoint-variational approach to examine the influence of available observations on the accuracy of the estimated turbulent fields in channel flow.

Previous efforts in the context of channel flow have all attempted to estimate the entire state from wall observations only, namely the wall stresses and pressure (Bewley & Protas 2004; Hœpffner *et al.* 2005; Chevalier *et al.* 2006; Colburn *et al.* 2011; Suzuki & Hasegawa 2017; Liu & Hasegawa 2020). No matter which method was adopted, the estimated state was only correlated with the true state up to the buffer layer. The literature on wall-bounded turbulence has not, however, examined how the accuracy of turbulence reconstruction changes with spatiotemporal resolution and placement of the observations, e.g. if more information about the flow state is available from PIV data. Recent state estimation tests in homogeneous isotropic turbulence (Yoshida, Yamaguchi & Kaneda 2005; Lalescu, Meneveau & Eyink 2013; Di Leoni, Mazzino & Biferale 2019; Li *et al.* 2020) demonstrated that the reconstruction of turbulence is successful only when the highest wavenumber k_m of velocity data satisfies $k_m \eta > 0.2$, where η is the Kolmogorov scale of the flow. In wall turbulence, however, flow inhomogeneity in the wall-normal direction, the wall-normal dependence of mean advection and the turbulence production all preclude adopting the same criterion from the homogeneous case. For the same reasons, it is also anticipated that the critical data resolution for reconstructing the turbulence is anisotropic – a matter that we will explore herein. Our focus is on reconstruction of turbulence at all scales using the nonlinear Navier–Stokes equations, and thus the critical data resolution is more restrictive than that for designing a reduced-order model for flow control (Jones *et al.* 2011, 2015).

In § 2, we introduce the adjoint-variational state estimation algorithm, and provide the details of the flow configuration and problem set-up. The state estimation results are presented in § 3. A benchmark case with subsampled volume data of velocity is analysed, followed by the effect of observation noise. Then a range of streamwise, spanwise and temporal data resolutions are explored. We propose criteria for minimal data required to successfully reconstruct the turbulent state. The possibility of estimating near-wall structures, which are difficult to measure experimentally, from data in the outer region and at the wall is subsequently investigated. At the end of § 3, the Reynolds-number effect on state estimation is discussed in the context of wall observations. The main conclusions that are drawn from these tests are summarized in § 4.

2. Adjoint-variational state estimation

A schematic of the channel-flow configuration is shown in figure 1. The domain is periodic in the streamwise and spanwise directions, and bounded by two no-slip surfaces in the vertical direction. The relevant Reynolds numbers are $Re = U_b h/\nu$ and $Re_\tau = u_\tau h/\nu$, where U_b is the bulk velocity, u_τ is the friction velocity, h is the half channel height and ν is the kinematic viscosity.

The adjoint-variational state estimation is formulated as a constrained optimization problem. The constraint is the numerical model $u^{n+1} = \mathcal{N}(u^n)$, which governs the evolution of the velocity field u from one time instant n to the next $n + 1$. The control vector, or the subject of the optimization, is the initial condition u^0. Given the observation data $\{m\}_{n=0}^N$, we define a cost function,

$$J(u^0) = \sum_{n=0}^N \tfrac{1}{2}\|m^n - \mathcal{M}(u^n)\|_O^2, \tag{2.1}$$
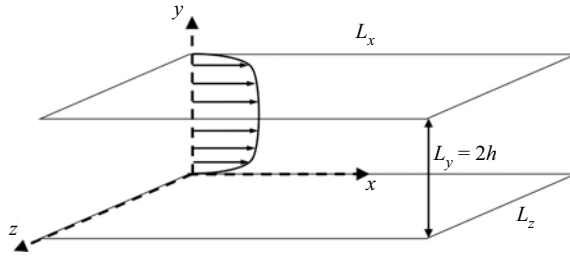
Figure 1. Schematic of channel flow and the coordinate system.

which is the L_2 norm of the difference between the observation data and their estimation from an initial condition $\boldsymbol{u}^0$. The subscript O represents the observation space, and $\mathcal{M}$ is an observation operator, which generates the measured quantity from any velocity field. The adjoint model is invoked to calculate the gradient of the cost function, which is necessary for its minimization procedure. The minimizer is the estimated initial condition, and the velocity field marched from this initial condition is the estimated flow. A detailed derivation and validation of the adjoint-variational method is provided by Wang *et al.* (2019*a*). In the following, we briefly summarize the forward model, adjoint equations and the optimization procedures.

2.1. *Forward equations and data acquisition*

The flow evolution is governed by the incompressible Navier–Stokes equations,

$$\nabla \cdot \boldsymbol{u} = 0, \tag{2.2}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u}\boldsymbol{u}) = -\nabla p + \frac{1}{Re}\nabla^2 \boldsymbol{u}, \tag{2.3}$$

where t is time and p is pressure. These equations are also referred to as the forward model because they are adopted to advance the flow state in time.

The Navier–Stokes equations are solved using a fractional-step method with a local volume flux formulation on a staggered grid (Rosenfeld, Kwak & Vinokur 1991). The advection terms are discretized by the Adams–Bashforth scheme, and the Crank–Nicolson scheme is adopted for the diffusion terms. The pressure Poisson equation is solved using a Fourier transform in the periodic directions and tridiagonal inversion in the wall-normal direction. The algorithm has been applied in a number of direct numerical simulations (DNS) of transitional and turbulent flows (Zaki & Durbin 2005; Zaki *et al.* 2010; Zaki 2013; Lee & Zaki 2017). For simplicity, the discretized Navier–Stokes equations will be denoted as

$$\boldsymbol{q}^{n+1} = \boldsymbol{G}^n \boldsymbol{q}^n, \tag{2.4}$$

where $\boldsymbol{q}^n$ is the state vector, including the velocity and pressure at every grid point, and $\boldsymbol{G}^n$ is a matrix that represents the discretized Navier–Stokes operator. Note that $\boldsymbol{G}^n$ is also a function of $\boldsymbol{q}^n$ because the equations are nonlinear.

The true state is statistically stationary turbulence, and is sustained by a known constant pressure gradient in the streamwise direction. Except in § 3.5, where we explore the effect of Reynolds numbers, we set $Re_\tau = 180$. While the forward model at these conditions has been extensively studied (Kim, Moin & Moser 1987; Jelly, Jung & Zaki 2014), this Reynolds number is higher than previously attempted in the context of adjoint-variational

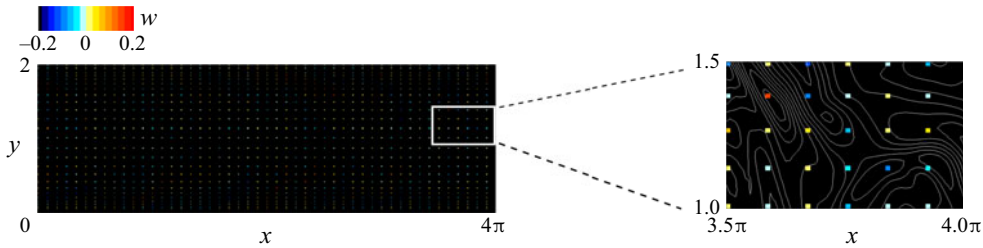| Domain size			Grid points			Grid resolution			
L_x/h	L_y/h	L_z/h	N_x	N_y	N_z	Δx^+	Δy_{min}^+	Δy_{max}^+	Δz^+
4π	2	2π	384	256	320	5.89	0.20	2.95	3.53

Table 1. Domain size and grid resolution.

Figure 2. Visualization of observation data of spanwise velocity in the x–y plane, subsampled at one-eighth the DNS resolution. Coloured regions represent the observation locations, and line contours are the full DNS field.

state estimation in channel flow. The domain size and the grid resolution are summarized in table 1. The computational domain is the same as that adopted by Kim *et al.* (1987), who used a pseudo-spectral algorithm. For our finite-volume scheme, we have doubled the resolution in each direction and performed extensive validation (see e.g. Jelly *et al.* 2014). The grid resolution is also reported in viscous units, denoted by superscript plus $(\cdot)^+$, $\Delta x^+ \equiv (\Delta x/h)Re_\tau$. The time-step size is $\Delta t^+ \equiv (\Delta t U_b/h)(Re_\tau^2/Re) = 0.058$ such that the Courant–Friedrichs–Lewy (CFL) number is lower than one-half.

We consider two types of observations: subsampled velocity data (e.g. figure 2) and stresses on both channel walls. The observations set-up is summarized in table 2. In all cases, the estimation window is $T = 4.5$ ($T^+ = 50$). This choice is motivated by the following considerations. The duration T should be sufficiently long that each point in the fluid is within the domain of dependence of observations. It should also be longer than the time to 'fill' the turbulence energy spectra starting solely form observations. Finally, T should not appreciably exceed the Lyapunov time scale ($\tau_\sigma^+ = 48$ at $Re_\tau = 180$ according to Nikitin (2018)). If $T \gg \tau_\sigma$, any infinitesimal perturbation in the initial condition will exponentially amplify and the accuracy of the state estimation deteriorates (Li *et al.* 2020; Chandramouli, Mémin & Heitz 2020). We start with a benchmark case (§ 3.1, case B1), where the velocity field is observed every eighth point in each dimension, including space and time. The velocity at the observation locations is assumed to be known precisely, without measurement noise. Subsequently (§ 3.2, cases N1–N2), the data will be contaminated with Gaussian random noise with standard deviation proportional to the local velocity, for example, for the streamwise velocity,

$$u_m = u_{true} + \eta, \quad \eta \sim N(0, \sigma |u_{true}|). \qquad (2.5a,b)$$

The effect of spatial and temporal data resolution is explored in § 3.3 (cases RZ1–RZ4, RX1–RX4 and RT1–RT4), and in § 3.4 the velocity observations in the viscous sublayer and buffer layer are removed. Finally, in § 3.5, we evaluate the influence of Reynolds number when the only observations are the wall stresses.

§	Case	ΔM_x	ΔM_y	ΔM_z	ΔM_t	Δx_m^+	Δy_m^+	Δz_m^+	Δt_m^+	T^+	σ (%)
3.1	B1	8	8	8	8	47	[1.8, 24]	28	0.46	50	0
3.2	N1, N2	8	8	8	8	47	[1.8, 24]	28	0.46	50	5, 10
3.3	RZ1–4	1	1	4–32	8	5.9	[0.2, 2.95]	14–112	0.46	50	0
3.3	RX1–4	4–32	1	1	8	24–192	[0.2, 2.95]	3.53	0.46	50	0
3.3	RT1–4	32	1	1	16–128	192	[0.2, 2.95]	3.53	0.92–7.4	50	0

Table 2. Parameters of observations relative to the reference DNS. The sampling rate is ΔM, with a subscript that denotes the spatial or temporal dimension; subscript m denotes the resolution of the observation grid, superscript '+' denotes viscous scaling, T is the observation time horizon, and σ is the standard deviation of the Gaussian noise.

2.2. Adjoint equations and the state estimation algorithm

In order to minimize the cost function (2.1) while satisfying the Navier–Stokes constraint, we introduce the Lagrangian,

$$L = J - \sum_{n=0}^{N-1} (q^{\dagger(n+1)})^{\mathrm{T}}(q^{n+1} - G^n q^n). \tag{2.6}$$

Note that the Lagrangian is a function of $\{q\}_{n=0}^N$ and $\{q^\dagger\}_{n=1}^N$. Taking the derivative of the Lagrangian with respect to $q^{\dagger(n)}$ and setting it to zero, we obtain the forward Navier–Stokes equations (2.4). By setting the derivatives of the Lagrangian with respect to q^n to zero ($1 \leqslant n \leqslant N-1$), we obtain the discrete adjoint equations,

$$q^{\dagger(n)} = (G^n)^{\mathrm{T}} q^{\dagger(n+1)} + \frac{\partial J}{\partial q^n}, \quad 1 \leqslant n \leqslant N-1, \tag{2.7}$$

which are marched backwards in time, and are forced by the gradient of the cost function with respect to the state. The forward operator G^n contains the forward variables q^n, which means that the full spatiotemporal evolution of the forward fields are required and must be stored for the solution of the adjoint equations. The second term on the right-hand side of (2.7) can be analytically derived from the expression of the cost function.

When the adjoint equations are marched back to $n = 0$, the following relation is obtained:

$$\frac{\partial L}{\partial q^0} = (G^0)^{\mathrm{T}} q^{\dagger 1} + \frac{\partial J}{\partial q^0} \equiv q^{\dagger 0}. \tag{2.8}$$

The initial adjoint field is therefore the gradient of the Lagrangian, and also the gradient of the cost function when both the forward and adjoint equations are satisfied,

$$\nabla_{q^0} J = \frac{\partial L}{\partial q^0} = q^{\dagger 0}. \tag{2.9}$$

Note that q^0 contains the velocity field u^0 only, because the initial pressure field is not required to solve the incompressible Navier–Stokes equations. Similarly, the initial adjoint field $q^{\dagger 0}$ comprises $u^{\dagger 0}$ only. Since the above derivation starts from the discrete Navier–Stokes equations, the gradient obtained using the discrete adjoint in (2.8) is accurate to machine precision. Detailed expressions of the discrete adjoint and verification of the forward-adjoint relation are provided in Wang *et al.* (2019*a*).

Algorithm 1: Adjoint-variational state estimation.

Step 1: Forward model;
- Start with an estimate of the initial condition u^{0^-} and project it onto a divergence-free space, $u^0 = Pu^{0^-}$;
- March u^0 using the forward equations (2.4) from $n = 0$ to $n = N - 1$ and store the forward velocity fields at every time step;
- Evaluate the cost function (2.1);

Step 2: Adjoint model;
- Solve the discrete adjoint equations (2.7) from $n = N - 1$ to $n = 1$;
- Obtain the gradient of the cost function, $u^{\dagger 0}$, defined in (2.8);
- Project the gradient onto solenoidal space $Pu^{\dagger 0}$;

Step 3: Update the estimated initial state;
- Compute the search direction using L-BFGS algorithm;
- Find an appropriate step size along the search direction;
- Update the estimate of the initial state and repeat Steps 1–3 until a prescribed maximum number of iterations is reached.

With the gradient of the cost function, we adopt the limited-memory Broyden–Fletcher–Goldfarb–Shanno(L-BFGS) optimization algorithm to minimize the cost function (Nocedal 1980). In order to guarantee that the estimated initial condition is divergence-free, we slightly modify the L-BFGS algorithm by introducing a symmetric projector. The basic idea is to update the new estimate of the initial condition using

$$u_{k+1}^0 = P(u_k^0 + \alpha_k d_k), \tag{2.10}$$

where subscript k denotes the kth iteration of the optimization procedure, and $d_k = -B_k P u^{\dagger 0}$ is the search direction; the matrix B_k is a rank-two approximation of the inversion of the Hessian matrix of the cost function; and the matrix P is a symmetric projector, which projects any velocity field u^0 or gradient $u^{\dagger 0}$ onto the divergence-free space. The symmetry of P ensures that $Pu^{\dagger 0}$ is the gradient of the cost function with respect to u^0 when u^0 is projected and before being advanced by the forward equations. The step size α_k is computed by the line search routine CVRSCH (Moré & Thuente 1994), which enforces the strong Wolfe condition and adopts cubic interpolation to update α_k.

Combining the forward solver (§ 2.1) with the adjoint solver and optimization algorithm (§ 2.2), we obtain the adjoint-variational state estimation algorithm. A summary is provided in algorithm algorithm 1. In all the examined configurations, the algorithm is always performed for 100 L-BFGS iterations, and, as such, comparisons are made using the same computational cost. It is important to emphasize that, in addition to the storage requirements associated with saving the forward fields at full spatiotemporal resolution, the computational cost is also substantial because each of the 100 L-BFGS iterations comprises at least one forward and one adjoint computation.

3. State estimation results

3.1. *Performance of the algorithm: the benchmark case*

In order to provide a qualitative perspective on the performance of the algorithm, figure 3 shows realizations of the flow from case B1 (see table 2), at both (i) the initial and (ii)

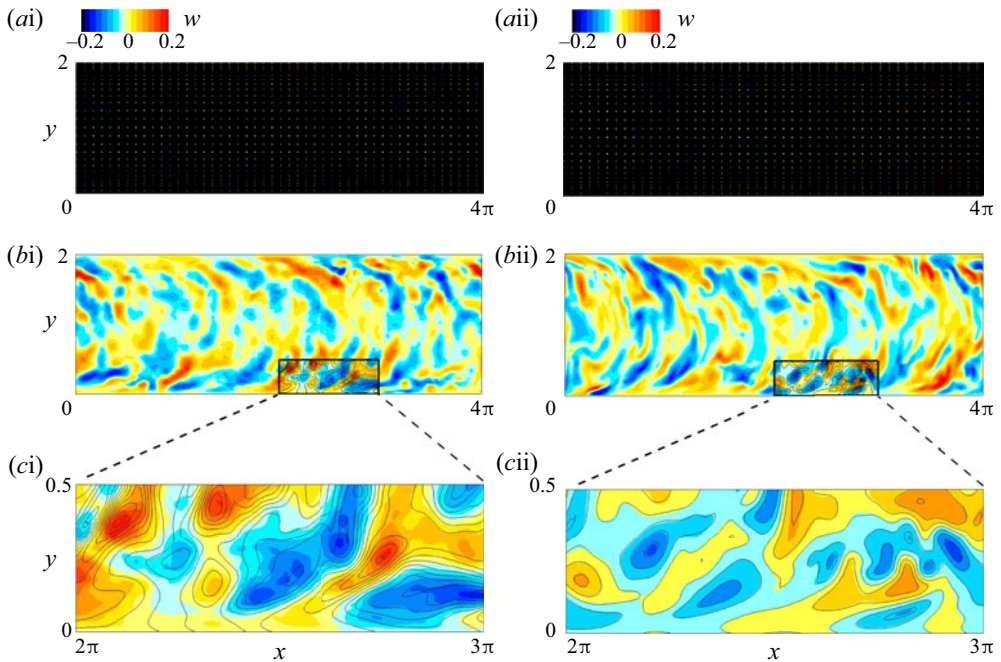Figure 3. Instantaneous spanwise velocity w at $z = L_z/2$ visualized using (*a*) observational data and (*b*) true fields (lines) and adjoint-variational prediction (colours) at (i) $t = 0$ and (ii) $t = T$ ($T^+ = 50$). (*c*) Zoom-in views of (*b*).

the final ($t^+ = T^+ = 50$) times within the assimilation window. At each instant, the field is visualized using (*a*) the observational data only, (*b*) the adjoint-variational prediction, and (*c*) a detailed comparison of the predicted (colour) and true (lines) states. Recall that the observations from benchmark case B1 are at 1/4096 the resolution of the simulation, since the velocity is observed at every eighth point in each spatial dimension and in time. The quality of the reconstruction is evident in figure 3, with the predictions at the initial time capturing the large scales of the flow, but appearing to contain some small-scale, or high-wavenumber, deviations (figure 3*c*i). At the final time, in contrast, these small-scale deviations have mostly vanished and the reconstruction quality is, qualitatively, improved (figure 3*c*ii).

The convergence history of the cost function that generated the assimilated initial condition is shown in figure 4(*a*). The normalization is performed using the cost function associated with advancing the initial guess which was obtained by spline interpolation of the observations and their projection onto the divergence-free space. The monotonic decrease of the cost function is ensured by the accurate evaluation of its gradient using the discrete adjoint. After 100 L-BFGS iterations, the cost function is reduced to 2.8 % of its initial value.

A quantitative assessment of prediction accuracy starts with an evaluation of the root-mean-square (r.m.s.) errors between the predicted and true states,

$$\mathcal{E}_V(\boldsymbol{u}) = \langle \|\boldsymbol{u} - \boldsymbol{u}_{true}\|^2 \rangle_V^{1/2}, \tag{3.1}$$

where $\langle \cdot \rangle$ denotes averaging, and the subscript indicates the averaging domain, here over the volume V; the same convention is adopted throughout the work for errors $\mathcal{E}$ and
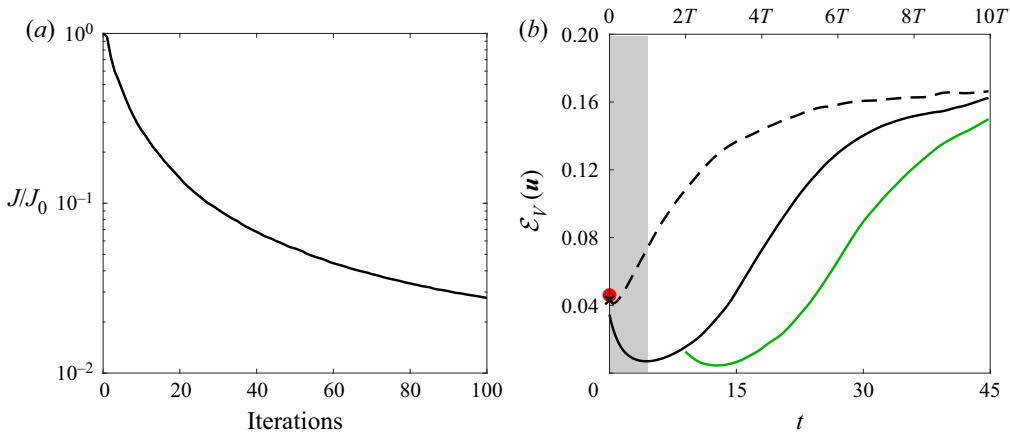
917 A9-8

Figure 4. (*a*) Convergence history of the cost function, normalized by its initial value. (*b*) Volume-averaged error (3.1) of the instantaneous velocity field estimated using different strategies: black dashed line, interpolated initial condition (red circle) is projected onto divergence-free space (black cross) and advanced using the Navier–Stokes equations; black solid line, estimated initial condition using data within $t \in [0, T]$ (grey region) is evolved in time; green solid line, estimated state at $t = 2T$ using data within $t \in [2T, 3T]$ (green region) is evolved in time.

correlation coefficients $\mathcal{C}$. When errors and correlation coefficients between the predicted and true fields are reported, they are evaluated throughout the domain, and not only at observation locations.

The time dependence of the estimation error is plotted in figure 4(*b*). At the initial time, the spline interpolation of the observations (red circle), projection onto the solenoidal space (cross) and the adjoint-variational state all have seemingly similar errors – the last being obtained after 100 L-BFGS iterations. A mild reduction in the initial errors is achieved by the projection, and a further modest improvement is achieved by the variational approach, but the initial errors remain approximately 4 % of the bulk velocity, which is of the same order of magnitude as the r.m.s. fluctuations. An important difference arises, however, when the initial conditions are advanced in time. When the interpolated observations are projected onto the divergence-free field and marched using the Navier–Stokes equations (dashed line), without any data assimilation, the errors amplify as expected due to the chaotic nature of the flow. At long times, after a transient divergence, nonlinear effects become dominant and thus the estimation error saturates. Simply performing spatiotemporal interpolation of the observations would maintain lower errors, similar to the red circle, although that would not be a solution to the governing equations. Now consider the errors when the initial condition is obtained from the adjoint-variational state estimation (solid line). The initial errors in the assimilated field decay with time, and the flow more closely tracks the trajectory of the true field in state space during the observation window (shaded region, $0 \leqslant t \leqslant 4.5$). At $t = T$, the errors are an order of magnitude smaller than those from interpolating and advancing the initial condition or performing spatiotemporal interpolation of observations.

Beyond the observation window, the estimated state again diverges from the truth but remains more accurate than evolving the interpolated initial field during the interval $[T, 3T]$. These results demonstrate the potential of the algorithm to provide a better prediction of the future state $t > T$, when observations are not available. If new data do become available at later times, the estimated state can be adopted as the initial guess, and the same variational procedure can be applied to drive the estimated state towards
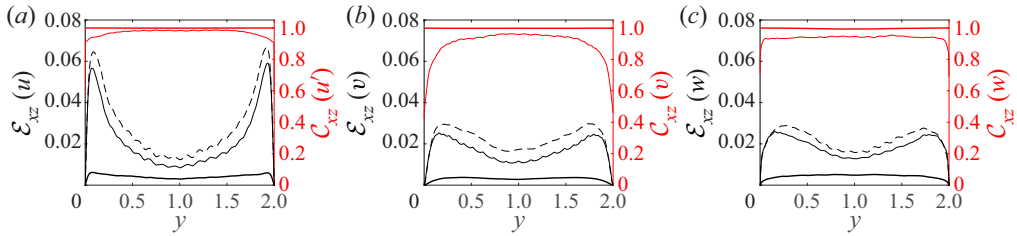
Figure 5. Wall-normal profiles of (black curves) horizontally averaged errors $\mathcal{E}_{xz}(q)$ and (red curves) correlation coefficient $\mathcal{C}_{xz}(q)$ (3.2) between the estimated and true flow states: (*a*) streamwise, (*b*) wall-normal and (*c*) spanwise components. Lines: black dashed line, spline interpolation of observed velocity field at $t = 0$; black and red solid thin lines, adjoint-variational estimation at $t = 0$; and black and red solid thick lines, estimation at $t = T$ ($T^+ = 50$).

the true state, again. This idea is demonstrated in figure 4(*b*): new observation data were provided in the interval $t \in [2T, 3T]$, which is marked by the light grey region. The adjoint-variational algorithm was applied in that new interval, and the predicted state at $t = 2T$ now yields a new trajectory that more closely follows the true flow (green line). While this point was noteworthy, the focus of the present effort is on characteristics of the state estimation in the first window $t \in [0, T]$, which are generally applicable to any subsequent interval of new observations.

The r.m.s. estimation error evaluated in the horizontal plane, $\mathcal{E}_{xz}(q)$, are plotted as a function of wall-normal direction in figure 5. The errors in the interpolated initial guess (dashed lines) are proportional to the level of physical fluctuations in the velocity field and, as a result, the errors in the streamwise component are most dominant especially in the near-wall region where u-perturbations are most energetic. The estimated initial condition (thin black line) is slightly improved relative to the interpolated state. The key observation is, however, at $t = T$, where all three components of errors in the estimated state are an order of magnitude more accurate than advancing the interpolated state using the forward model. For the streamwise component, the error is less than 0.5 % of the bulk velocity, or, equivalently, 8 % of the peak value of r.m.s. streamwise fluctuation. Figure 5(*b*) also shows the correlation coefficient,

$$\mathcal{C}_{xz}(q) = \frac{\langle q\, q_{true}\rangle_{xz}}{\langle q^2\rangle_{xz}^{1/2} \langle q_{true}^2\rangle_{xz}^{1/2}}. \tag{3.2}$$

At $t = T$, the estimated and the true fields are nearly perfectly correlated, which highlights the capacity of the assimilated field to reproduce the true trajectory of the flow in state space.

In order to explain the improvement in accuracy during the observation time horizon, we evaluate the spectra of the estimation error,

$$\mathcal{E}_y(\hat{\boldsymbol{u}}) = \langle \|\hat{\boldsymbol{u}} - \hat{\boldsymbol{u}}_{true}\|^2\rangle_y^{1/2}, \tag{3.3}$$

where $\hat{\boldsymbol{u}}$ is the Fourier transform of $\boldsymbol{u}$ in the streamwise and spanwise directions. The spectra of the errors are reported in figure 6(*a*) as a function of the horizontal (k_x, k_z) wavenumbers; also shown in figure 6(*b*) is the spectrum of the true velocity field. Errors appear largest in the low-wavenumber modes, but they should be viewed relative to the high energy content in these modes in figure 6(*b*). Normalized by the spectral density in the true field (see figure 6*c*), low wavenumbers are better reconstructed since they are encoded in the sparse observations. A more important observation is the behaviour of the
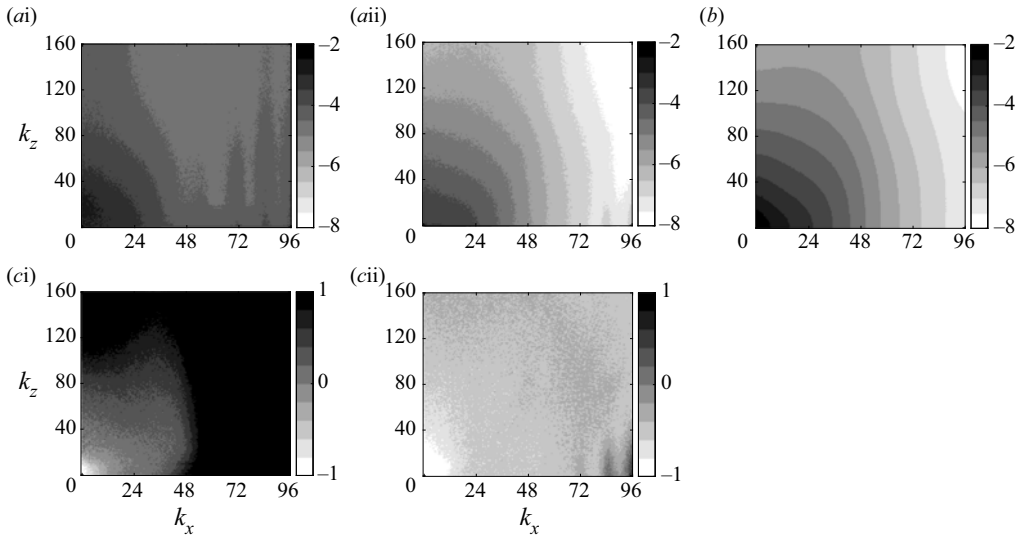
917 A9-10

Figure 6. (*a*) Fourier spectrum of errors from adjoint-variational estimation, averaged in y and normalized by the bulk velocity, $\log_{10} \mathcal{E}_y(\hat{\boldsymbol{u}})$. (*b*) Spectrum of the true velocity field averaged in time and wall-normal direction, $\log_{10}\langle\|\hat{\boldsymbol{u}}_{true}\|^2\rangle_{yt}^{1/2}$. (*c*) Same as (*a*) with the errors normalized by $\langle\|\hat{\boldsymbol{u}}_{true}\|^2\rangle_{yt}^{1/2}$. Panels: (i) $t = 0$ and (ii) $t = T$ ($T^+ = 50$). Wavenumbers (k_x, k_z) are normalized by the half channel height.

high-wavenumber components of the errors. The estimated initial condition (figure 6*a*i) has appreciable errors in those wavenumbers (higher than the interpolated state – not shown). However, since most of these modes decay with time (figure 6*a*ii), they have little impact on the estimation quality later within the assimilation time horizon. In addition, due to their rapid decay, these high-wavenumber initial errors do not appreciably affect the value of the cost function, which is integrated over the entire observation window. As a result, they persist in the initial condition during the optimization procedure. An effective strategy to reduce these initial high-wavenumber errors is to incorporate time-dependent weights in the cost function, which amplify the contribution of early observations near $t = 0$ (Wang *et al.* 2019*a*). One should caution, however, that not all the initial high-wavenumber errors are stable and decaying. Small components of that error are unstable and amplify at the Lyapunov rate, and, although they are not perceptible within the present time horizon, they will dominate at longer times.

Consistent with the above integral and spectral measures of the estimation quality, the reconstructed velocity field at $t = T$ is almost identical to the true one (side views in figure 7). An accurate estimation of vortical structures is, however, more challenging since the computation involves velocity gradients. These structures are visualized using the λ_2 vortex identification criterion (Jeong & Hussain 1995) and compared in figure 7 (grey isosurfaces). The prediction of vortical structures is very compelling, in both the near-wall and outer regions. In the former, the vortical tubes are attached to the wall and elongated in the streamwise direction. Farther from the wall, the lifted vortical tubes break down and generate small-scale structures that are mostly captured by the estimated state. In practice, reconstructing the vorticity field from coarse-resolution experimental data is a challenge (Suzuki 2012). The satisfactory estimation quality in figure 7(*b*) demonstrates the potential of our algorithm to augment under-resolved turbulent data from experiments.
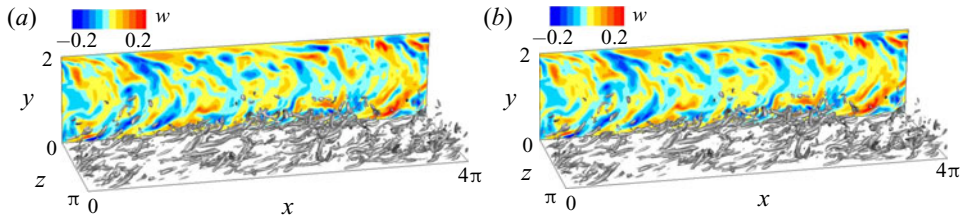
Figure 7. Comparison of (*a*) true and (*b*) adjoint-variational estimated state at $t = T$ ($T^+ = 50$). Grey isosurfaces: vortical structures visualized using the λ_2 vortex identification criterion with threshold $\lambda_2 = -2$. Side view: contours of spanwise velocity.

3.2. *The influence of noise in the observation data*

In the previous section, the observation data were free of any noise. In practice, however, experimental measurements invariably contain errors and, as such, may violate the governing equations, lead to statistical errors and severely preclude accurate evaluation of derivatives, especially in turbulent flows where strong velocity gradients are expected (see e.g. Bardet, Peterson & Savaş 2010). For example, Abrahamson & Lonnes (1995) assessed the ability of conventional circulation and least-squares methods to reproduce the vorticity field from a randomly perturbed DNS velocity field. When 5 % noise was superposed onto the fully resolved velocity data, the uncertainty in the computed vorticity field reached 40 %, and most of the small-scale structures were absent in the reconstruction.

In order to evaluate the performance of the adjoint-variational algorithm with noisy data, we contaminate the observed velocities by Gaussian random noise with standard deviation that is proportional to the local velocity component (2.5*a,b*). The spatiotemporal resolution of the data and the estimation window $T^+ = 50$ remain the same as in the benchmark case. We consider three levels of noise: $\sigma = \{0, 5, 10\}$ %, which correspond to cases B1, N1 and N2 in table 2. The first guess of the initial condition is interpolation of the noisy data at $t = 0$.

After 100 L-BFGS iterations, the cost function is decreased to $\{2.8, 27, 45\}$ % for the three levels of contamination. Since the cost function is defined as the difference between the estimated and contaminated data, and the latter do not satisfy the Navier–Stokes equations, the cost function cannot decrease to zero. The r.m.s. error of the estimated state relative to the true, uncontaminated flow field was evaluated and shows a similar decay from the initial to the final time as in the benchmark case without noise. Therefore, only the results at $t = T$ are examined here (figure 8). The estimation error increases with the noise level (from light grey to black lines), but remains within 2 % of the bulk velocity. Note that, due to the mean flow, the observation noise in the streamwise direction can exceed σ times the bulk velocity, which means that the estimation accuracy of the streamwise velocity actually exceeds the precision of observation data. Comparatively, the estimation errors of the wall-normal and spanwise velocity components are bounded by the observation noise, which is approximately σ times the r.m.s. turbulence fluctuations. Overall, even with the highest noise level ($\sigma = 10$ %), the correlation coefficient (3.2) between the estimated and true state is close to unity at all the y locations, as shown by the red lines in figure 8(*b*).

The reconstructed vortical structures are visualized in figure 9. Although some of the small-scale structures are not captured, most of the reconstructed wall-attached and detached vortical structures remain nearly identical to the true flow, and the estimation quality is almost independent of the noise level. A quantitative assessment of the quality of the vorticity field is provided in figure 10. When noisy observations are interpolated and
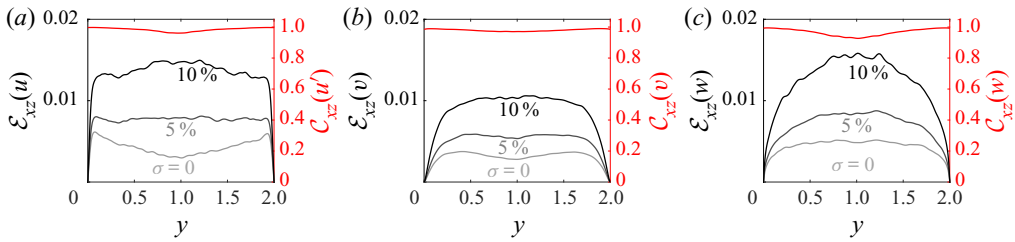
Figure 8. Effect of observation noise level on (grey to black, $\sigma = \{0, 5, 10\}\%$) estimation error $\mathcal{E}_{xz}(q)$ and (red, $\sigma = 10\%$) correlation coefficient $\mathcal{C}_{xz}(q)$ at $t = T$ ($T^+ = 50$): (a) streamwise, (b) wall-normal and (c) spanwise components.
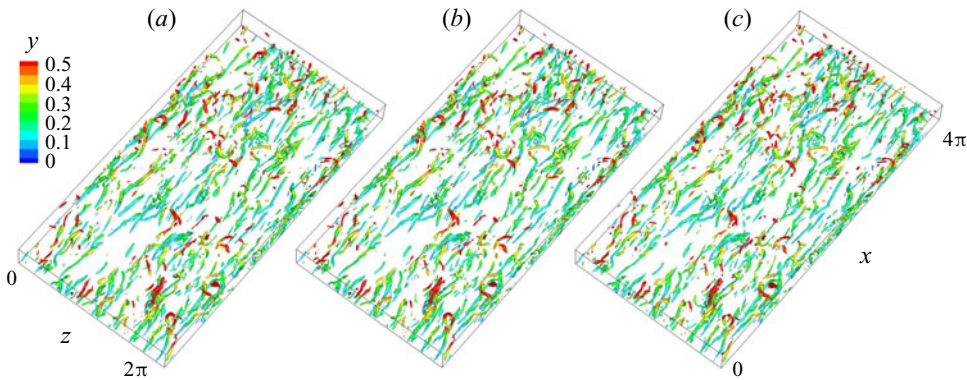

Figure 9. Comparison of (a) true vortical structures and (b,c) the adjoint-variational reconstruction when $\sigma = \{5, 10\}\%$ at $t = T$ ($T^+ = 50$) within the bottom half channel, visualized using the λ_2 vortex identification criterion with threshold $\lambda_2 = -2$.
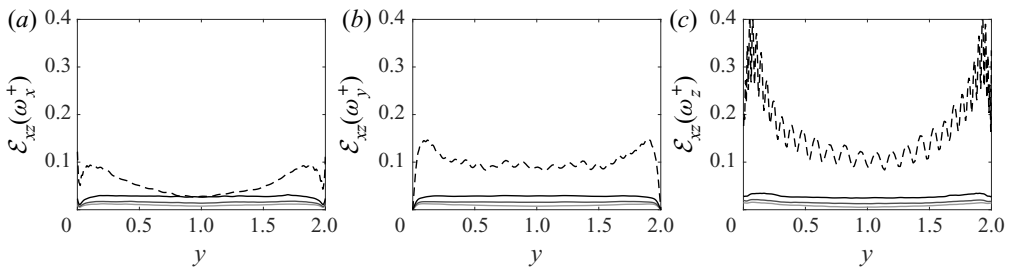

Figure 10. Horizontally averaged error $\mathcal{E}_{xz}(q)$ of vorticity field at $t = T$ ($T^+ = 50$), estimated by (black dashed line) interpolating noisy velocity data and (black solid line) adjoint-variational approach (grey to black, noise level $\sigma = \{0, 5, 10\}\%$): (a) streamwise, (b) wall-normal and (c) spanwise components. The error is normalized by mean vorticity $\langle du/dy \rangle$ at the wall.

vorticity is evaluated, the error (dashed lines) in the near-wall region reaches 10–40 % of the mean vorticity at the wall. The error of the vorticity field from adjoint-variational estimation (black solid lines) is within 4 % of the wall vorticity, and the estimation accuracy is robust against observation noise. These results demonstrate the superiority of the adjoint-variational approach for evaluating velocity gradients and its robustness to observation noise.
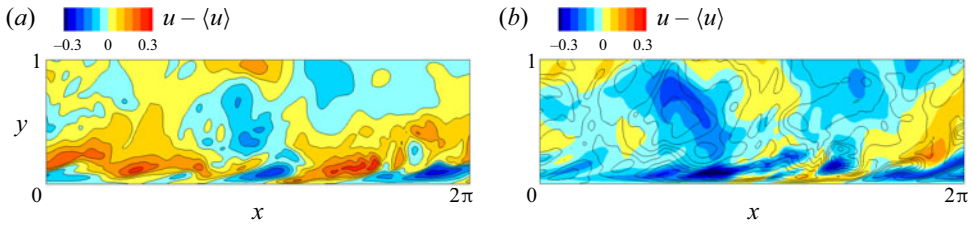
Figure 11. The (lines) true streamwise fluctuation $u - \langle u \rangle$ at $t = T$ ($T^+ = 50$) and (colours) estimation with data resolution $\Delta z_m^+ = 112$ (case RZ4 in table 2). The fields are visualized at (*a*) an observation plane, $z = z_m$, and (*b*) the midpoint between two observation planes, $z = z_m + 0.5\Delta z_m$.

3.3. *The effect of spatial and temporal data resolution*

Although the results thus far have demonstrated the accuracy of the flow reconstructions, it is expected that the estimation quality depends on the spatiotemporal resolution of the observations. And it is also of interest to query the lowest resolution requirement for accurate estimation. In homogeneous isotropic turbulence, it has been reported that reconstruction of the full field is successful only if the resolution of spatial observations satisfies $\Delta_m < 15\eta$ (Li *et al.* 2020), where η is the Kolmogorov length scale. An equivalent criterion has not, however, been proposed for anisotropic, wall-bounded turbulence, where the effects of mean shear, advection and the no-slip boundary may alter the resolution requirements of the observations. Hereafter, we revert to adopting noise-free data and focus on the influence of spatiotemporal resolution of observations on the accuracy of state estimation within the time horizon $T^+ = 50$.

We first consider the impact of spanwise spacing of observations that are fully resolved in the *x*–*y* plane (cases RZ1–RZ4 in table 2). With the most coarse observations ($\Delta z_m^+ = 112$, case RZ4), the estimated state is visualized in figure 11 and compared to the true one. At the observation locations ($z = z_m$), velocity data are reproduced by the algorithm. At the midpoint between observation planes ($z = z_m + 0.5\Delta z_m$), however, the estimation accuracy is notably compromised.

Since our interest is in the distribution of errors between observation planes, the estimation error is phase-averaged in the span in addition to averaging in the streamwise direction, and is denoted $\mathcal{E}_{xz_m}(q)$. The results for cases RZ1–RZ4 are shown in figure 12. Only the error for the *u* component is plotted, and the results for the *v* and *w* components are similar. With the most poorly resolved data (figure 12*a*), the estimation error increases by an order of magnitude from the observation planes to the midpoint between them. The error in the near-wall region is of the same order of magnitude as the local turbulence fluctuations, which is significantly higher than the error at the channel centre. As better-resolved observations are adopted (figure 12*b*–*d*), the inhomogeneity of the error distribution in the spanwise direction becomes weaker.

The effect of spanwise data resolution can be explained by examining the spanwise two-point correlations $\mathcal{R}_u(\Delta z)$ in figure 13(*a*i). Since the dominant structures in the wall layer are streaks and streamwise vortices that are narrow in the span, the two-point correlation decays faster at locations closer to the wall. As a result, the domain of dependence of the observation planes shrinks from channel centre to the wall, which explains the high estimation error in the wall layer. Figure 13(*a*ii) shows the reconstruction quality at $t = T$ in terms of the correlation coefficient between the true and estimated states from the least-resolved data ($\Delta z_m^+ = 112$). The profiles are qualitatively similar to the two-point correlations within $\Delta z^+ < 20$. However, while the turbulent structures
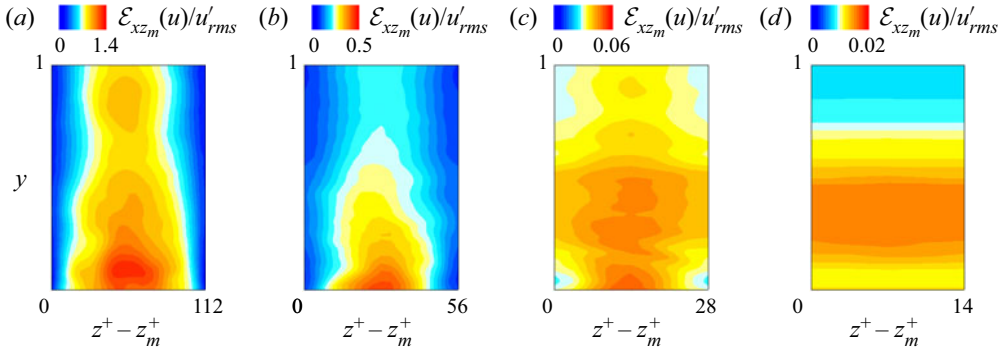
917 A9-14

Figure 12. Streamwise- and phase-averaged estimation error $\mathcal{E}_{xz_m}(u)$ between two x–y observation planes at $t = T$ ($T^+ = 50$), normalized by the true local r.m.s. fluctuations u'_{rms}: (a–d) $\Delta z_m^+ = \{112, 56, 28, 14\}$, respectively.
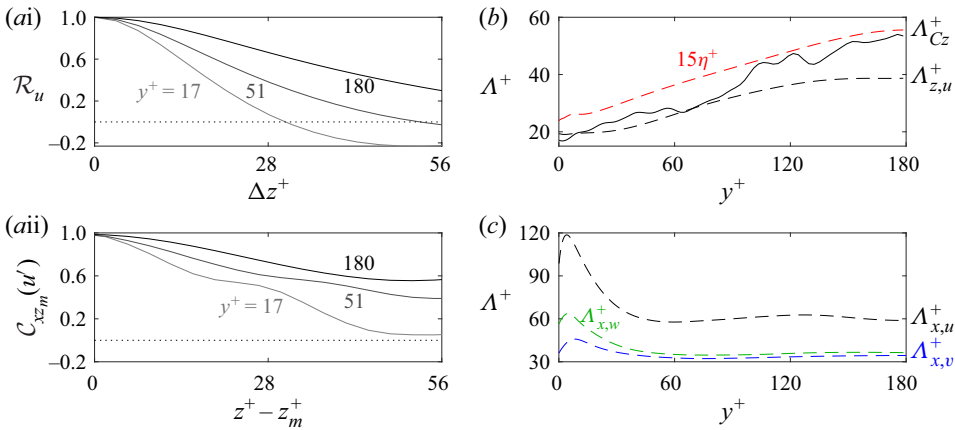

Figure 13. (ai) Two-point correlations $\mathcal{R}_u$ of the true streamwise fluctuation as a function of spanwise separation; and ($a$ii) correlation coefficient $\mathcal{C}_{xz_m}(u')$ between estimated (case RZ4, $\Delta z_m^+ = 112$) and true streamwise fluctuations at $t = T$ ($T^+ = 50$) as a function of distance from observation planes. Grey to black correspond to $y^+ = \{17, 51, 180\}$. The horizontal dotted line marks zero correlation. (b) Wall-normal profiles of (red dashed line) Kolmogorov length scale and Taylor microscales based on (black solid thin line) correlation coefficient $\mathcal{C}_{xz_m}(u')$ and (black dashed line) spanwise two-point correlations. (c) Streamwise Taylor microscales based on two-point correlations of (black dashed line) u, (blue dashed line) v and (green dashed line) w components.

decorrelate at larger distances, the accuracy of the reconstruction remains relatively higher, and $\mathcal{C}_{xz_m}(u')$ returns to approximately unity as we approach the next observation plane at $z^+ - z_m^+ = 112$.

Similar to the notion of the Taylor microscale,

$$\Lambda_{z,u} = \left(-\frac{1}{2} \frac{\mathrm{d}^2 \mathcal{R}_u}{\mathrm{d}(\Delta z)^2} \bigg|_{\Delta z = 0} \right)^{-1/2}, \tag{3.4}$$

we introduce a length scale for the domain of dependence of one observation plane by replacing $\mathcal{R}_u$ in (3.4) by the correlation coefficient $\mathcal{C}_{xz_m}(u')$. The resulting length scale, which we denote Λ_{Cz}, is representative of the domain of accurate estimation. Both $\Lambda_{z,u}$

and Λ_{Cz} are plotted in figure 13(*b*), and have similar values across the height of the channel: the spanwise size of the domain of dependence of observations is similar in the Taylor microscale. Therefore, the criterion

$$\Delta z_m \lesssim 2\Lambda_{z,u} \tag{3.5}$$

must be satisfied to guarantee an accurate estimation of the local u field. Similar criteria can be adopted for accurate estimation of the v and w components using the Taylor microscales $\Lambda_{z,v}$ and $\Lambda_{z,w}$, respectively. Since those length scales are commensurate with $\Lambda_{z,u}$, the condition (3.5) suffices. Using (3.5), we can also interpret the estimation results with different spanwise data resolutions (figure 12). When $\Delta z_m^+ = 56$ (figure 12*b*), the criterion (3.5) is satisfied for the bulk of the channel ($2\Lambda_{z,u}^+ \approx 80$, cf. figure 13*b*) and starts to be violated for $y^+ < 70$. As such, while figure 12(*b*) reports high prediction accuracy in the bulk ($\mathcal{C}_{xz_m}(u') = 0.97$), errors increase in that near-wall region ($\mathcal{C}_{xz_m}(u') = 0.83$) and become inhomogeneous in the span. When $\Delta z_m^+ = 28 < \min_y 2\Lambda_{z,u}^+$ (figure 12*c*), every point in the flow is covered by the domain of dependence of observations, so the estimation quality becomes more accurate and uniform at all the y locations ($\mathcal{C}_{xz_m}(u') = 0.99$).

We compare the criterion (3.5) to that for homogeneous isotropic turbulence (Yoshida *et al.* 2005; Lalescu *et al.* 2013; Li *et al.* 2020) that $\Delta_m \lesssim 15\eta$. The Kolmogorov scale in our case is $\eta \equiv (Re^3\mathcal{D})^{-1/4}$, where $\mathcal{D} = (2/Re)\langle s'_{ij}s'_{ij}\rangle$ is the viscous rate of dissipation and $s'_{ij} = (\partial_i u'_j + \partial_j u'_i)/2$ is the fluctuating strain-rate tensor. The criterion 15η is plotted as the red dashed line in figure 13(*b*). Since the average dissipation is affected by the streamwise-elongated structures in the channel, η is larger than the spanwise size of the smallest eddies. Nonetheless, the trend is similar to the Taylor microscale condition provided above. Physically, both criteria demonstrate that the critical data resolution for accurate estimation of the entire flow is within the transition zone between the inertial and viscous dissipation ranges.

Next, we consider fully resolved observations in the cross-flow y–z planes, and under-resolved in the streamwise direction (cases RX1–RX4 in table 2). Before examining the state estimation results, we report the Taylor microscales $\Lambda_{x,i}$ ($i = u, v, w$) in figure 13(*c*), which is computed from the streamwise two-point correlations. Owing to the near-wall elongated streaky structures, $\Lambda_{x,u}$ is largest among all three components, peaks near the wall and decays towards the channel centre. We therefore expect that the estimation, particularly of the u component, should be most accurate in the inner layer relative to the accuracy in the outer flow.

The r.m.s. estimation error is plotted in figure 14, where $\mathcal{E}_{x_m z}(u)$ is averaged in the span and phase-averaged in the streamwise direction, and also normalized by the r.m.s. fluctuations. Overall, the estimation error decreases as better-resolved data are included (figure 14*a–d*). For each observation resolution, the estimation quality deteriorates from the wall to the channel centre, as expected. Two notable differences from the effect of spanwise resolution are observed: (i) the estimation error is not symmetric with respect to the midpoint between observation planes, especially in figure 14(*a,b*); and (ii) although the separation of observation planes in figure 14(*a*) is larger than $2\Lambda_{x,u}^+$ (≈ 118) in the bulk (cf. figure 13*c*), the estimation error does not increase appreciably between streamwise observation locations and remains within approximately 10 % of the local r.m.s. fluctuations ($\mathcal{C}_{x_m z}(u') \geqslant 0.95$ while $\mathcal{R}_u(\Delta x^+ = 96) \approx 0.6$). Both points are caused by the mean advection in the streamwise direction. Conceptually, every instant when data are recorded in the cross-flow y–z plane corresponds to an accurately estimated 'layer' in the spatiotemporal evolution of the flow, propagating by the advection velocity U_a. The thickness of such layers is approximately the Taylor microscale $\Lambda_{x,u}$, and the
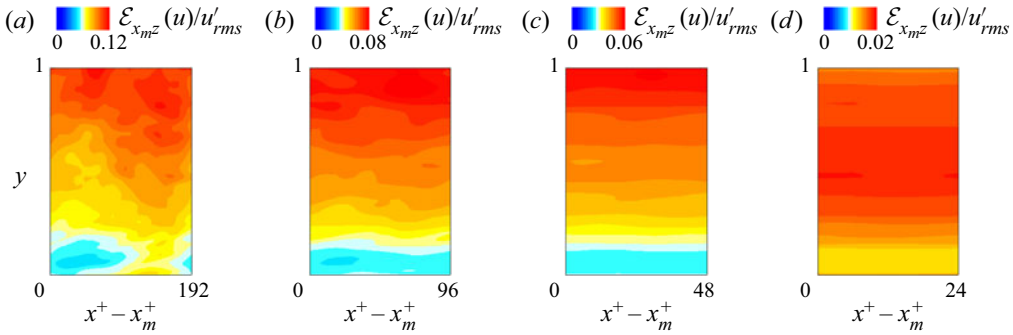
Figure 14. Spanwise- and phase-averaged estimation error $\mathcal{E}_{x_{mz}}(u)$ between two y–z observation planes at $t = T$ ($T^+ = 50$), normalized by the true local r.m.s. fluctuations u'_{rms}: (a–d) $\Delta x_m^+ = \{192, 96, 48, 24\}$, respectively.

distance between two adjacent ones is approximately $U_a \Delta t_m$. Since the observation data are temporally well resolved, $U_a \Delta t_m \ll 2\Lambda_{x,u}$, the accurately estimated layers overlap with one another and lead to commensurate accuracy between observation locations. An example that further highlights this conceptual interpretation is considered next, where the temporal sampling rate is sufficiently low ($U_a \Delta t_m > 2\Lambda_{x,u}$) in order to distinguish the accurately predicted layers associated with different (x_m, t_m).

Consider the same coarse spatial resolution from case RX4 ($\Delta x_m^+ = 192$), and we now adopt a long time between observations, $\Delta t_m^+ = 7.4$. This case is denoted RT4, and a temporal evolution of the estimation error from $t = 0$ to $t = \Delta t_m$ is shown in figure 15. Three layers of small errors are observed: A originates at the observation station at $t = 0$ and advects downstream; B emerges between observation locations and times, such that the field accurately reproduces measurements at a downstream observation point at a subsequent measurement time $t = \Delta t_m$; and B' is similar to B but reaches the observation station at $t = 2\Delta t_m$. Once A leaves the observation plane, the error increases, at least initially, due to the chaotic nature of turbulence and the absence of a nearby observation station to correct the field – behaviour later in time is discussed below. By comparison, the error in B decays with time as it approaches the observation location at time $t = \Delta t_m$.

A detailed space–time representation of the estimation error at the channel centre is shown in figure 16. All the regions of small errors (blue) belong to one of the three classes described above, and can be associated with at least one observation station (black plus signs). The pattern of the errors is repeated, anchored at $(\Delta x_m, \Delta t_m)$ and inclined according to the advection speed. Since $3U_a \Delta t_m \approx 2\Delta x_m$, regions A_i that originate from $(x, t) = (x_m, 0)$ will reach another observation station at $(x, t) = (x_m + 2\Delta x_m, 3\Delta t_m)$; the errors in that region thus initially increase moving away from $(x_m, 0)$ and reduce as that later observation point is approached. For a general spatiotemporal resolution of observations, regions A_i may not coincide with another observation position and time, and hence the errors would not undergo the later decay. From figure 16, it is evident that an accurate estimation of the entire flow state can be achieved by refining either the spatial or temporal data resolution such that the low-error regions overlap. This view is demonstrated in figure 17(a–d) where the panels correspond to increasing temporal resolution (cases RT4–RT1, respectively, in table 2). When the true state is observed sufficiently frequently, the estimation error in the bulk region approaches homogeneity in the streamwise direction (figure 17c,d). Specifically, accurate reconstruction is achieved when the distance between the thin layers that sample observation points is within the domain of dependence of
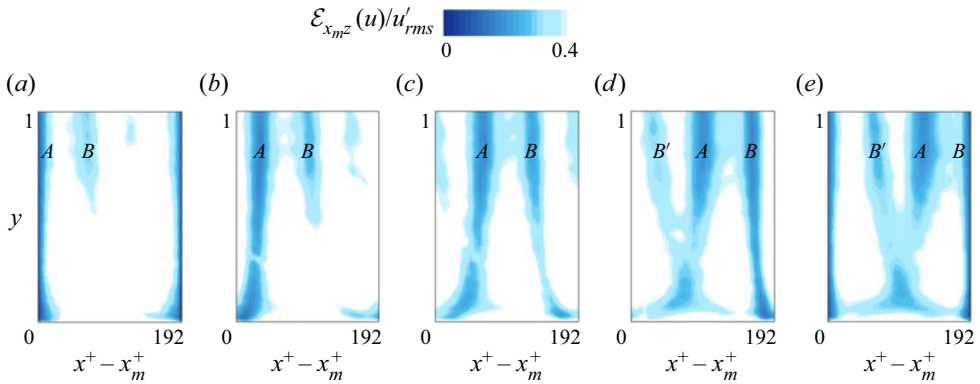
917 A9-17

Figure 15. Spanwise- and phase-averaged estimation error $\mathcal{E}_{x_m z}(u)$ within the first observation time interval ($\Delta t_m^+ = 7.4$, $T^+ = 50$, case RT4): (*a–e*) $t = \{0.00, 0.25, 0.50, 0.75, 1.00\}\Delta t_m$, respectively. Blue regions: *A*, regions where error is low at $(x, t) = (x_m, 0)$ and advected downstream; *B*, emergent region of small error that later coincides with observation plane $(x, t) = (x_m, \Delta t_m)$; and similarly *B'* coincides with $(x, t) = (x_m, 2\Delta t_m)$.
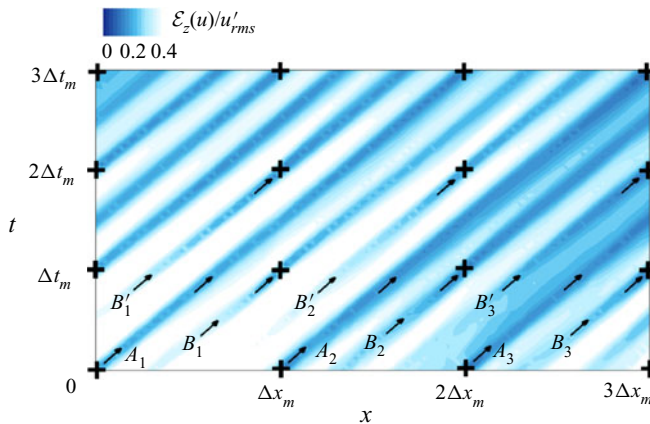

Figure 16. Space–time evolution of the spanwise-averaged estimation error $\mathcal{E}_z(u)$ at the channel centre. Black plus sign, observation location and time; A_i, B_i, B_i', same as *A*, *B*, *B'* in figure 15; and the subscript denotes the *i*th interval between streamwise observations.

observations,

$$U_a \Delta t_m \lesssim 2\Lambda_{x,u}. \qquad (3.6)$$

Recalling that $\Lambda_{x,v} \approx \Lambda_{x,w} < \Lambda_{x,u}$ (cf. figure 13*c*), the equivalent criterion for accurate estimation of v and w is more restrictive. When the observations only satisfy the bound for u, the reconstruction is less accurate for the other two velocity components (e.g. $\mathcal{C}_{x_m z}(u', v, w) = \{0.95, 0.86, 0.79\}$ when $\Delta x_m^+ = 192$ and $\Delta t_m^+ = 7.4$).

In summary, to guarantee an accurate estimation of local velocity component *i*, the streamwise and temporal data resolution must satisfy either one of two conditions:

(i) $\Delta x_m \lesssim 2\Lambda_{x,i}$ and $\Delta t_m \lesssim \tau_\sigma$ or   (ii) $\Delta t_m \lesssim 2\Lambda_{x,i}/U_a$ and $\Delta x_m \lesssim U_b \tau_\sigma$. (3.7)

The first condition is similar to that for the spanwise data resolution, namely that the Taylor microscale must be resolved by observations. Note that it is supplemented by a statement regarding the observation time, which must be smaller than the Lyapunov time
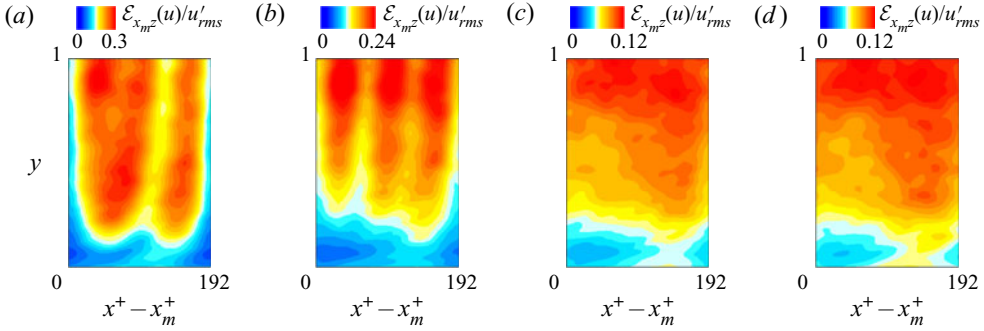
917 A9-18

Figure 17. Spanwise- and phase-averaged estimation error $\mathcal{E}_{x_m z}(u)$ at $t = T$ ($T^+ = 50$) with different observation time interval: (*a*–*d*) $\Delta t_m^+ = \{7.4, 3.7, 1.8, 0.9\}$, respectively.
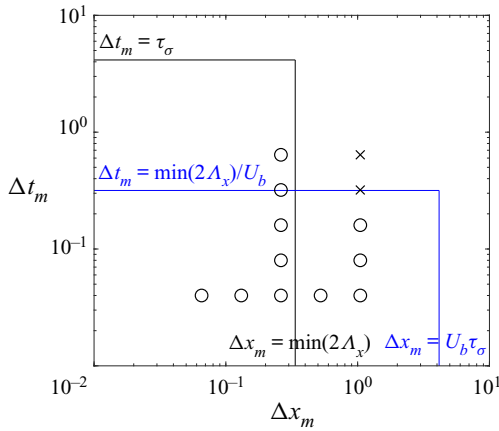

Figure 18. Criteria for streamwise and temporal data resolutions: (black line) $\Delta t_m = \tau_\sigma$ and $\Delta x_m = \min_{y,i}(2\Lambda_{x,i})$; (blue line) $\Delta t_m = \min_{y,i}(2\Lambda_{x,i})/U_b$ and $\Delta x_m = U_b\tau_\sigma$; (empty circles) successful estimation with correlation coefficient higher than 0.9; (crosses) unsuccessful estimation.

scale τ_σ because even a near-perfect estimate of the flow state will diverge from the true solution over that period unless additional data are available for assimilation. The second condition is a reinterpretation that accounts for mean advection: should the temporal resolution resolve the advected Taylor scale, the only requirement for spatial resolution becomes a condition based on the travel distance within the Lyapunov time. These two conditions are plotted in figure 18. We performed additional tests with $\Delta x_m^+ = 48$ and $\Delta t_m^+ = \{0.45, 0.90, 1.80, 3.70, 7.40\}$. When the correlation coefficient between the true and estimated states was higher than 90 %, the outcome was deemed successful; these cases are marked by circles. Inaccurate, or unsuccessful, reconstructions are marked by crosses. The outcomes of all the cases agree with the above two conditions.

3.4. *Estimation without wall-layer data*

In all the above cases, observations were distributed throughout the wall-normal extent of the channel, spanning both the inner and outer layers. In reality, however, experimental observations become progressively more difficult to obtain near the wall (Hutchins, Hambleton & Marusic 2005; Smits, McKeon & Marusic 2011). And since most of

the turbulence kinetic energy is produced near the wall (peak production at $y^+ \approx 12$) and transported into the bulk region, lack of observation in this region may severely compromise the estimation of the full state. We therefore explore the accuracy of reconstructing the flow within the inner region when observations are only available in the outer layer and directly on the wall. Previous efforts for this configuration are limited and have generally focused on use of linear models (Baars, Hutchins & Marusic 2016; Illingworth, Monty & Marusic 2018). The present test case is the first attempt to perform a reconstruction based on the full nonlinear Navier–Stokes equations.

Specifically, we consider two types of observations: (i) subsampled velocity data available in the outer layer, $y^+ > 30$, with the same spatial and temporal resolution as the benchmark case; (ii) shear stresses $\boldsymbol{\tau} = (\tau_{xy}, \tau_{zy})$ on both channel walls with the same resolution as velocity data. The observation time horizon is the same as the benchmark case, $T^+ = 50$. These observations are weighted in the cost function,

$$J(\boldsymbol{u}^0) = \tfrac{1}{2} \sum_{n=0}^{N} [\|\boldsymbol{u}_m^n - \mathcal{M}_u(\boldsymbol{u}^n)\|^2 + (\beta Re/Re_\tau)^2 \, \|\boldsymbol{\tau}_m^n - \mathcal{M}_\tau(\boldsymbol{u}^n)\|_w^2], \qquad (3.8)$$

where $\|\cdot\|_w$ represents integration over top and bottom walls and the weight $\beta = 1/5$. The choice of weighting parameter β may be motivated by different objectives, such as balancing the two contributions in the cost function or their gradients or minimizing the condition number of the Hessian of $\mathcal{J}$. Here we adopt $\beta = 1/5$ such that the stress term is commensurate with the velocity at the first fluid grid point that was observed in our benchmark case. The estimated state is compared with the benchmark case to highlight the impact of missing observations in the wall layer.

The predicted statistics without wall-layer observations, evaluated at $t = T$, are plotted in figure 19 and are compared with the true statistics. The quantities shown are horizontally averaged; the corresponding turbulence kinetic energy equation is

$$\frac{\partial E}{\partial t} = \underbrace{-\langle u_i' u_j' \rangle \frac{\partial \langle u_i \rangle}{\partial x_j}}_{\mathcal{P}} - \frac{\partial}{\partial x_j} \left(\frac{1}{2} \langle u_i' u_i' u_j' \rangle + \langle u_j' p \rangle - \frac{2}{Re} \langle u_i' s_{ij}' \rangle \right) - \underbrace{\frac{2}{Re} \langle s_{ij}' s_{ij}' \rangle}_{\mathcal{D}}. \qquad (3.9)$$

The estimated profiles are accurate, even within the viscous sublayer and buffer layer, which are void of observations and where production and dissipation predominate. Although the interpolated initial condition also converges to the true statistics after a long time ($t \approx 4T$), our estimated state reaches the statistically stationary state faster ($t \approx 1.5T$), especially for dissipation. The errors in the estimated instantaneous velocity fields are shown in figure 20. Compared with the benchmark case (grey lines), the estimation quality in the outer layer is almost unaffected when the wall-layer data are removed. In the near-wall region, however, the estimation error reaches a maximum and exceeds twice the error in the benchmark case. The streamwise component is the most poorly reconstructed. Without data in the wall layer, the data assimilation algorithm starts with an interpolated velocity field as a first guess, which underestimates the mean flow near the wall; the final prediction of the algorithm retains a similar deficit in the mean, and hence the estimation error is largest in the streamwise component (black lines in figure 20). At $t = T$, the highest estimation error is approximately 4 % of the bulk velocity, which is smaller than the 16 % local r.m.s. streamwise turbulence fluctuations. In addition, the correlation coefficient between the true and estimated states at $t = T$ is above 0.8 for all three components of the velocity field.
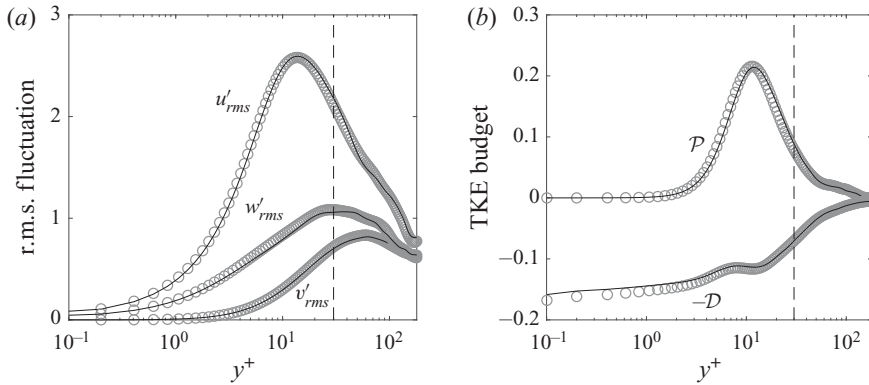
917 A9-20

Figure 19. Comparison of the instantaneous ($t = T$, $T^+ = 50$), horizontally averaged flow statistics for the (grey circles) true and (black lines) estimated fields, without wall-layer observation data: (*a*) r.m.s. fluctuations; and (*b*) turbulence kinetic energy (TKE) production and dissipation. Dashed lines, $y^+ = 30$.
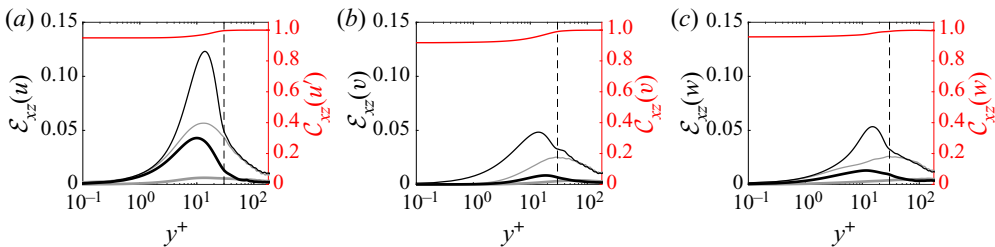

Figure 20. Comparison of the estimation errors $\mathcal{E}_{xz}(q)$ (grey) with and (black) without wall-layer observation data: (*a*) streamwise, (*b*) wall-normal and (*c*) spanwise components at (thin) $t = 0$ and (thick) $t = T$ ($T^+ = 50$). Red: correlation coefficient $\mathcal{C}_{xz}(q)$ between the true and estimated states without wall-layer observation data at $t = T$. Dashed line, $y^+ = 30$.

The wall layer is host to coherent structures, such as streaks and streamwise vortices, that play an important role in the dynamics of the near-wall turbulence cycle (Hamilton, Kim & Waleffe 1995) and in flow control. For example, the streak spacing must be resolved by a controller in order to effectively relaminarize turbulent channel flow (Sharma *et al.* 2011). Accurate prediction of these structures is therefore important. The error of our estimation at different length scales is reported in figure 21. The most accurately reconstructed structures are long in the streamwise direction ($\lambda_x \approx O(L_x)$) and large waves in the span ($\lambda_z^+ \in (10^2, 10^3)$), which are typical features of near-wall streaks and streamwise vortices (Jiménez 2018). The estimation error of these structures is mildly affected by removing the wall-layer data, due to their coherence across wall-normal locations, which facilitates the estimation even without near-wall observations. In contrast, reconstruction of small-scale near-wall structures is more sensitive to the lack of observations in that region, which is symptomatic of the weaker sensitivity of the outer flow to these structures.

Despite the overall increase of estimation errors when wall-layer data are not available, the instantaneous visualization of the coherent structures can still be compelling. Figure 22 shows a visualization of the predicted instantaneous streamwise fluctuation velocities for two predictions: without and with observations in the wall layer. The estimated streaky structures without wall-layer observations (figure 22*a*i) are irregularly spaced in the spanwise direction and meander downstream, as expected from a realistic channel flow.
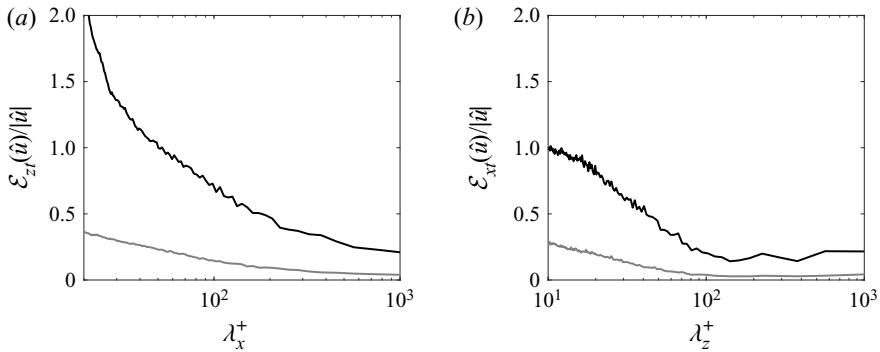
Figure 21. One-dimensional Fourier spectra of estimation error at $y^+ = 15$, averaged within $t \in [0.5, 1.5]T$ and normalized by the amplitude of the corresponding Fourier modes: (*a*) streamwise spectrum averaged in the span; and (*b*) spanwise spectrum averaged in the streamwise direction. Estimation (grey line) with and (black line) without wall-layer observation data.
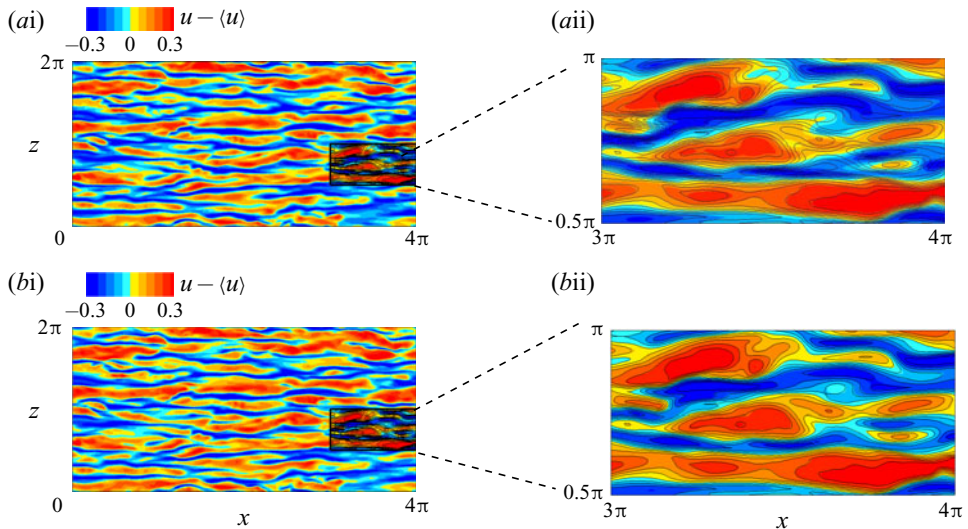

Figure 22. Streamwise fluctuations at $y^+ = 15$ and $t = T$ ($T^+ = 50$), calculated by subtracting the true mean velocity: estimations (*a*) without and (*b*) with wall-layer data. Lines: true fluctuation fields.

The zoomed-in view (figure 22*a*ii) also includes the true state (line contours), which confirms that the state estimation is predictive: the true streaks are reproduced. The estimation accuracy is in fact comparable to the case when observations are available in the wall layer (figure 22*b*). The successful estimation of the missing wall layer is directly tied to the sensitivity of the observations in the outer flow and at the walls to the state in this region at the initial time. In other words, accurate reconstruction of the inner layer at the initial time is indispensable to match future observations far from the wall, and is manifest in figure 22 by the accurate prediction of the flow state at the final time.

3.5. *Estimation from wall observations*

In practice, control of wall-bounded turbulence may rely on sensing and actuation at the wall, and effective strategies are often predicated on decoding the wall signature

Re_τ	Re	L_x/h	L_z/h	N_x	N_y	N_z	Δx^+	Δy^+_{min}	Δy^+_{max}	Δz^+
100	1429	4π	2π	128	128	128	9.8	0.74	2.2	4.9
180	2800	4π	2π	384	256	320	5.9	0.20	3.0	3.5
392	6875	2π	π	256	320	192	9.6	0.34	5.1	6.4
590	10935	2π	π	384	384	384	9.6	0.44	6.5	4.8

Table 3. Computational set-up of cases with wall observations (§ 3.5).

of outer turbulence structures. It has been demonstrated both mathematically (Lighthill 1963; Constantin & Iyer 2011) and numerically (Eyink, Gupta & Zaki 2020*a,b*) that all the interior vorticity is generated at the wall. The converse problem, specifically to what extent the initial flow state can be predicted from wall observations, has not been addressed as comprehensively. Previous efforts at $Re_\tau = 100$, and using a variety of approaches, yielded an estimated state that is nearly uncorrelated with the true flow beyond the buffer layer (Bewley & Protas 2004; Suzuki & Hasegawa 2017; Liu & Hasegawa 2020). The modest Reynolds number in those studies does not support the presence of large-scale outer structures that appreciably influence the wall stress fluctuations (Abe, Kawamura & Choi 2004). LSE has been applied at higher Re_τ to reconstruct large-scale structures in the log layer from wall observations (Encinar & Jiménez 2019). However, the predicted state from LSE does not satisfy the Navier–Stokes equations. Here we focus on adjoint-variational state estimation, specifically on the accuracy of the predicted state and the domain of dependence of wall observations. The dependence on Reynolds number will be examined for $Re_\tau = \{100, 180, 392, 590\}$. The computational domain and grid resolution are summarized in table 3. Note that a smaller domain size is adopted for the $Re_\tau = \{392, 590\}$ cases due to the limited computational resources.

The observations are fully resolved shear stresses $\boldsymbol{\tau} = (\tau_{xy}, \tau_{yz})$ and pressure p at both walls, similar to those adopted by Bewley & Protas (2004) at the lower Reynolds number. The corresponding cost function is defined as

$$J(\boldsymbol{u}^0) = \tfrac{1}{2} \sum_{n=0}^{N} [\|\boldsymbol{\tau}_m^n - \mathcal{M}_\tau(\boldsymbol{u}^n)\|_w^2 + \|p_m^n - p^n\|_w^2]. \tag{3.10}$$

For all the different Re cases, the estimation window is the same in viscous units, $T^+ = 50$, which is long enough for perturbation in the bulk region to affect wall signals. The first guess of the initial condition is an LSE of the flow using observations at $t = 0$. The stochastic estimator was constructed with a completely independent time series. All the estimated flows are obtained after 100 L-BFGS iterations of the state estimation 4DVar algorithm.

The correlation coefficients of the estimated and true fluctuating velocities at $t = T$ are shown in figure 23. The $Re_\tau = 100$ results are comparable to those in figure 6(*a*) of Bewley & Protas (2004) at the same Reynolds number. As Re_τ is increased, the correlation near the wall slightly deteriorates but remains sufficiently high to provide confidence in predictions. Precisely, for all Reynolds numbers, $C_{xz} > 0.8$ when $y^+ < 15$. The correlations start to decay beyond the buffer layer, and those for v and w (figure 23*b,c*) do so nearly monotonically and with a similar slope for all Re_τ. This trend highlights the challenge of interpreting turbulent flows from wall observations: the accurately predicted near-wall layer is a diminishingly smaller physical region as the Reynolds number is increased. A noteworthy exception is recorded in the correlation coefficients of the streamwise velocity
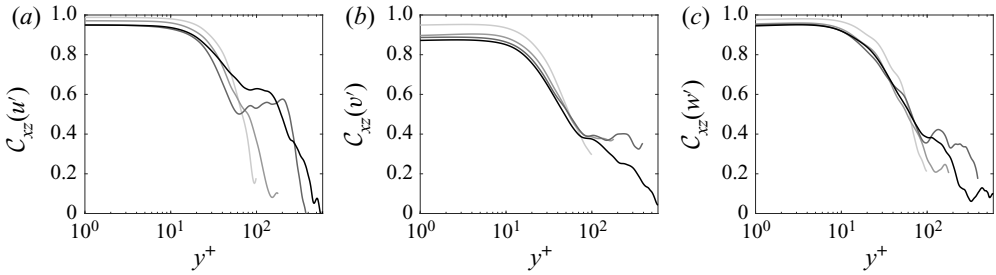
Figure 23. Correlation coefficient $\mathcal{C}_{xz}(q)$ (3.2) between the true and estimated fluctuation fields at $t = T$ ($T^+ = 50$), evaluated relative to the true mean flow: (*a*) streamwise, (*b*) wall-normal and (*c*) spanwise components. Grey to black: $Re_\tau = \{100, 180, 392, 590\}$, respectively.
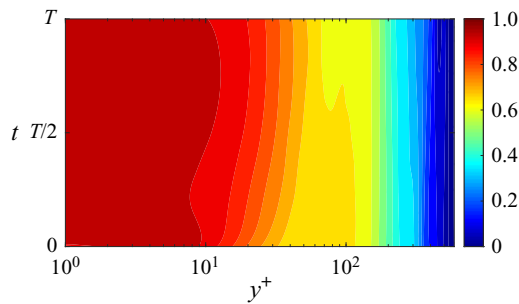

Figure 24. Time dependence of the horizontally averaged correlation coefficient $\mathcal{C}_{xz}(u')$ between true state and adjoint-variational estimation at $Re_\tau = 590$.

(figure 23*a*) at $Re_\tau = \{392, 590\}$: the initial decay outside the buffer layer is followed by a plateau within ($30 < y^+$, $y < 0.3$) where the reconstruction remains marginally accurate because the large-scale structures in that region superimpose a footprint on wall signals (Abe *et al.* 2004; Mathis, Hutchins & Marusic 2009).

The history of the correlation coefficient $C_{xz}(u')$ at the highest Reynolds number $Re_\tau = 590$ is shown in figure 24. The adjoint-variational approach yields an evolution of the near-wall flow that is strongly correlated with the true state throughout the observation horizon. It also sustains the plateau in $C_{xz}(u')$ in the outer region that was observed at the final time in figure 23, and which is related to the estimation of the outer large-scale motions. This relation is evident in visualizations of the perturbation fields, e.g. the instantaneous views in figure 25 at $t = T$. In the outer layer, the estimated field matches the mean flow and large-scale motions, while the small-scale fluctuations in the bulk and vortices that are detached from the wall are not captured. These results indicate that the wall signature at these Reynolds numbers is not sensitive to the wall-detached motions in the initial condition – a demonstration of the inherent difficulty of turbulence reconstruction from wall observations.

In light of the importance of the near-wall turbulence regeneration cycle, we turn to the detail of the true and estimated states in that region. The near-wall streaks (bottom x–z planes in figure 25) are accurately reconstructed, and many of the vortical structures (grey isosurfaces in figure 26) are reproduced in the estimated field. Sheng, Malkiel & Katz (2009) reported that violent near-wall ejections and sweeps contribute appreciably to the local turbulence energy. Examples of both events are shown in panels (i) and (ii) in figure 26: the true vortex lines above the stress local extrema are closely followed by the
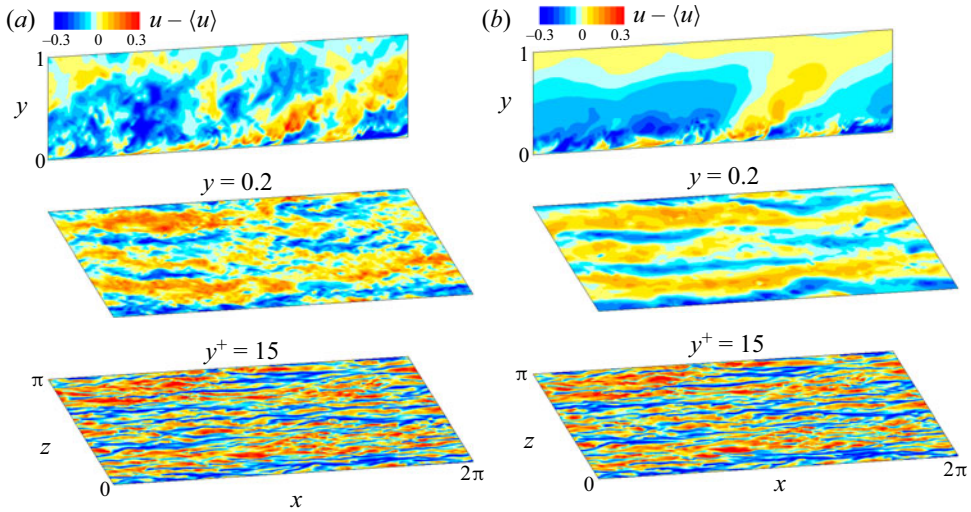
Figure 25. Comparison of (*a*) true streamwise fluctuation and (*b*) estimation using wall observations at $t = T$ ($T^+ = 50$), $Re_\tau = 590$.
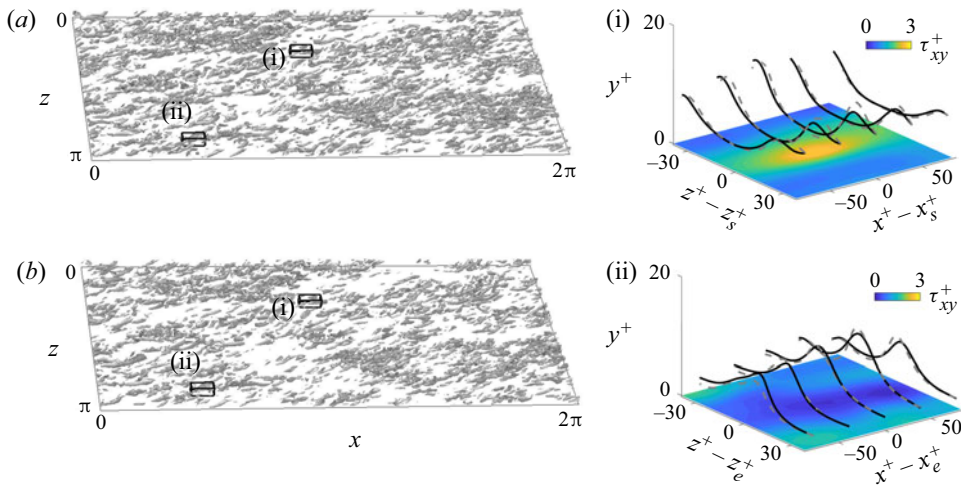

Figure 26. (*a*) True and (*b*) estimated vortical structures at $t = T$ ($T^+ = 50$) within $y^+ \in [0, 30]$ ($Re_\tau = 590$), visualized using the λ_2 vortex identification criterion with threshold $\lambda_2 = -15$. Panels (i) and (ii) show (dashed grey) true and (black solid line) estimated vortex lines associated with extreme events: (i) sweep, $(x_s, z_s) = (3.0, 0.9)$, vortex lines initiated at $y^+ = 7.6$; (ii) ejection, $(x_e, z_e) = (1.3, 2.8)$, vortex lines initiated at $y^+ = 2.3$.

estimation, which demonstrates that the extreme events in the buffer layer are encoded in the wall signatures.

The accuracy of the estimated state can be viewed against the backdrop of the near-wall 'autonomous cycle' (Jiménez & Pinelli 1999; Jiménez & Moser 2007). By artificially removing the outer flow in DNS, Jiménez & Pinelli (1999) demonstrated that the near-wall dynamics, especially the regeneration cycle of streaks and vortices, is self-sustained. And, as observed in experiments, 'ejection' and 'sweep' events in that cycle have a wall signature in the form of minima and maxima of the wall shear stress (Sheng *et al.* 2009).

Here we demonstrate that the wall stress is in fact encoded with the entire dynamics of the regeneration cycle, and our state estimation algorithm decodes these observations to discover the entire flow within the wall layer. The algorithm can also reconstruct the outer large-scale motions when their footprint is encoded in the wall measurements.

4. Conclusions

Starting from sparse observations, we attempted to reconstruct the initial state of turbulence in channel flow. The problem was formulated as an adjoint-variational minimization of a cost function that is defined in terms of the difference between the available observations and their estimates from fully resolved Navier–Stokes simulations. The gradient of the cost function was computed by solving the forward equations and their discrete adjoints, and an L-BFGS algorithm was adopted to update the estimate of the initial state during successive iteration; this step was supplemented by a symmetric projector that constrains both the search direction and estimated initial state to be solenoidal.

The performance of the algorithm was evaluated in a benchmark case, where the observations were low-resolution data, at 1/4096 of the required sampling to spatially and temporally resolve the flow in DNS. The variational state estimation algorithm achieved more than 80 % error reduction compared to interpolating the coarse-resolution velocity data, and ensured that the predicted flow not only satisfies the Navier–Stokes equations but also tracks the evolution of the true field in state space over the observation time horizon. The estimation errors are initially high wavenumber, and decay within the assimilation time window which was designed to be of the order of the Lyapunov time scale. The error characteristics were explained in terms of the sensitivity of observations to the initial state. Specifically, observations are insensitive to the high-wavenumber content of the initial condition. In addition, the estimation error decays because the optimization problem is dominated by late observations since any mismatch with available data amplifies exponentially in the adjoint reverse time. We cautioned, however, that longer time horizons than the Lyapunov time scale would lead to diverging trajectories because small unstable errors in the initial condition would amplify sufficiently and compromise accuracy.

Using the same benchmark configuration, the observations were contaminated with Gaussian noise and yet the variational state estimation algorithm successfully reconstructed a noise-free Navier–Stokes solution. The correlation between the estimated and true states exceeds 95 %, even when observation noise reaches 10 % of local velocity. The vortical structures, which are generally difficult to reproduce from noisy data, were also accurately reconstructed. Quantitatively, the error of the estimated vorticity field is within 4 % of the wall vorticity. That the noise in the observations was not amplified in the estimated state demonstrates the robustness of the method to reconstruct velocity gradients.

Criteria for the density of observations in the horizontal plane and in time were identified, and are related to the Taylor microscale of the turbulence. Physically, in order to ensure accurate reconstruction, the separation of observation stations cannot exceed their domain of dependence. The criteria are consistent with that obtained in homogeneous isotropic turbulence (Yoshida *et al.* 2005; Lalescu *et al.* 2013; Li *et al.* 2020), and can accommodate anisotropy of wall turbulence. The presence of mean advection can be exploited to relax the critical streamwise data resolution, when the frequency of temporal sampling can resolve the advected Taylor scale. These criteria were also supplemented with a condition that accounts for divergence of trajectories during the Lyapunov time scale due to the stochasticity of the flow.

Another important configuration was considered where no observations were available in the wall layer. Instead, observations only comprised subsampled velocities in the outer flow and wall shear stresses. This test case is, to date, the first attempt to reconstruct the instantaneous flow field in the wall layer $y^+ < 30$ from such observations and using the full nonlinear Navier–Stokes equations. In spite of the lack of observations in the region of peak turbulence kinetic energy production, the estimated profiles of flow statistics and the streaky structures in the wall layer were almost indistinguishable from the true state. These results demonstrate the sensitivity of the outer flow and wall shear stress to the turbulence field in the wall layer at the initial state – a result consistent with our notion that the turbulence produced in that region imprints onto the wall stresses (Sheng *et al.* 2009) and extends and is transported into the outer flow (Jiménez & Moser 2007).

One final configuration that is of both theoretical and practical interest is state estimation from wall observations. From a theoretical perspective, it is known that all the vorticity in the field can be traced back in time to its origin at the wall (Lighthill 1963; Constantin & Iyer 2011; Eyink *et al.* 2020*a*,*b*); here the converse problem is examined, where the wall vorticity is traced back to the initial state of the flow. From a practical perspective, the capacity to control wall turbulence is often reliant on our ability to predict its state from wall measurements. State estimation was performed using fully resolved wall observations of shear stresses and pressure, at four Reynolds numbers. The first guess of the initial condition was constructed from an LSE, and was updated using the iterative adjoint-variational approach. The adjoint estimation accurately reproduces the evolution of the true state in the near-wall region and also captures the evolution of outer large-scale motions when their footprint reaches the wall. Despite the accurate estimation of u' large-scale structures in the log layer, overall the estimation quality deteriorates appreciably beyond the buffer layer. This deterioration is due to a lack of sensitivity of wall observations to the wall-detached outer structures, which is consistent with the autonomy of the near-wall regeneration cycle that was demonstrated by Jiménez & Pinelli (1999). Finally, it is important to recall that the accurately predicted region below the buffer layer corresponds to a progressively smaller physical height at higher Reynolds numbers, which highlights the challenge of estimating wall-bounded turbulence in that regime.

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

Author ORCIDs.
Ⓓ Tamer A. Zaki https://orcid.org/0000-0002-1979-7748.

REFERENCES

ABE, H., KAWAMURA, H. & CHOI, H. 2004 Very large-scale structures and their effects on the wall shear-stress fluctuations in a turbulent channel flow up to $Re_\tau = 640$. *Trans. ASME J. Fluids Engng* **126** (5), 835–843.

ABRAHAMSON, S. & LONNES, S. 1995 Uncertainty in calculating vorticity from 2D velocity fields using circulation and least-squares approaches. *Exp. Fluids* **20** (1), 10–20.

ADRIAN, R.J. & MOIN, P. 1988 Stochastic estimation of organized turbulent structure: homogeneous shear flow. *J. Fluid Mech.* **190**, 531–559.

BAARS, W.J., HUTCHINS, N. & MARUSIC, I. 2016 Spectral stochastic estimation of high-Reynolds-number wall-bounded turbulence for a refined inner-outer interaction model. *Phys. Rev. Fluids* **1**, 054406.

BARDET, P.M., PETERSON, P.F. & SAVAȘ, Ö. 2010 Split-screen single-camera stereoscopic PIV application to a turbulent confined swirling layer with free surface. *Exp. Fluids* **49** (2), 513–524.

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

BUCHTA, D.A. & ZAKI, T.A. 2020 Observation-infused simulations of high-speed boundary layer transition. *J. Fluid Mech.* (accepted) arXiv:2011.08265.

CERIZZA, D., SEKIGUCHI, W., TSUKAHARA, T., ZAKI, T.A. & HASEGAWA, Y. 2016 Reconstruction of scalar source intensity based on sensor signal in turbulent channel flow. *Flow Turbul. Combust.* **97** (4), 1211–1233.

CHANDRAMOULI, P., MÉMIN, E. & HEITZ, D. 2020 4D large scale variational data assimilation of a turbulent flow with a dynamics error model. *J. Comput. Phys.* **412**, 109446.

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

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

CONSTANTIN, P. & IYER, G. 2011 A stochastic-Lagrangian approach to the Navier–Stokes equations in domains with boundary. *Ann. Probab.* **21** (4), 1466–1492.

DEISSLER, R.G. 1986 Is Navier–Stokes turbulence chaotic? *Phys. Fluids* **29** (5), 1453–1457.

DI LEONI, P.C., MAZZINO, A. & BIFERALE, L. 2019 Synchronization to big-data: nudging the Navier–Stokes equations for data assimilation of turbulent flows. arXiv:1905.05860.

ENCINAR, M.P. & JIMÉNEZ, J. 2019 Logarithmic-layer turbulence: a view from the wall. *Phys. Rev. Fluids* **4** (11), 114603.

EVENSEN, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* **99** (C5), 10143–10162.

EYINK, G.L., GUPTA, A. & ZAKI, T.A. 2020*a* Stochastic Lagrangian dynamics of vorticity. Part 1. General theory for viscous, incompressible fluids. *J. Fluid Mech.* **901**, A2.

EYINK, G.L., GUPTA, A. & ZAKI, T.A. 2020*b* Stochastic Lagrangian dynamics of vorticity. Part 2. Application to near-wall channel-flow turbulence. *J. Fluid Mech.* **901**, A3.

HAMILTON, J.M., KIM, J. & WALEFFE, F. 1995 Regeneration mechanisms of near-wall turbulence structures. *J. Fluid Mech.* **287**, 317–348.

HŒPFFNER, J., CHEVALIER, M., BEWLEY, T.R. & HENNINGSON, D.S. 2005 State estimation in wall-bounded flow systems. Part 1. Perturbed laminar flows. *J. Fluid Mech.* **534**, 263–294.

HUTCHINS, N., HAMBLETON, W.T. & MARUSIC, I. 2005 Inclined cross-stream stereo particle image velocimetry measurements in turbulent boundary layers. *J. Fluid Mech.* **541**, 21–54.

ILLINGWORTH, S.J., MONTY, J.P. & MARUSIC, I. 2018 Estimating large-scale structures in wall turbulence using linear models. *J. Fluid Mech.* **842**, 146–162.

JELLY, T.O., JUNG, S.Y. & ZAKI, T.A. 2014 Turbulence and skin friction modification in channel flow with streamwise-aligned superhydrophobic surface texture. *Phys. Fluids* **26** (9), 095102.

JEONG, J. & HUSSAIN, F. 1995 On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94.

JIMÉNEZ, J. 2018 Coherent structures in wall-bounded turbulence. *J. Fluid Mech.* **842**, P1.

JIMÉNEZ, J. & MOSER, R.D. 2007 What are we learning from simulating wall turbulence? *Phil. Trans. R. Soc. Lond.* A **365** (1852), 715–732.

JIMÉNEZ, J. & PINELLI, A. 1999 The autonomous cycle of near-wall turbulence. *J. Fluid Mech.* **389**, 335–359.

JONES, B.L., HEINS, P.H., KERRIGAN, E.C., MORRISON, J.F. & SHARMA, A.S. 2015 Modelling for robust feedback control of fluid flows. *J. Fluid Mech.* **769**, 687–722.

JONES, B.L., KERRIGAN, E.C., MORRISON, J.F. & ZAKI, T.A 2011 Flow estimation of boundary layers using DNS-based wall shear information. *Intl J. Control* **84** (8), 1310–1325.

KIM, J. & BEWLEY, T.R. 2007 A linear systems approach to flow control. *Annu. Rev. Fluid Mech.* **39** (1), 383–417.

KIM, J., MOIN, P. & MOSER, R. 1987 Turbulence statistics in fully developed channel flow at low Reynolds number. *J. Fluid Mech.* **177**, 133–166.

LALESCU, C.C., MENEVEAU, C. & EYINK, G.L. 2013 Synchronization of chaos in fully developed turbulence. *Phys. Rev. Lett.* **110** (8), 084102.

LAW, K., STUART, A. & ZYGALAKIS, K. 2015 *Data Assimilation: A mathematical introduction*. Springer.

LE DIMET, F.-X. & TALAGRAND, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* A **38** (2), 97–110.

LEE, S.J. & ZAKI, T.A. 2017 Simulations of natural transition in viscoelastic channel flow. *J. Fluid Mech.* **820**, 232–262.

LI, Y., ZHANG, J., DONG, G. & ABDULLAH, N.S. 2020 Small-scale reconstruction in three-dimensional Kolmogorov flows using four-dimensional variational data assimilation. *J. Fluid Mech.* **885**, A9.

LIGHTHILL, M.J. 1963 Boundary layer theory. In *Laminar Boundary Layers* (ed. L. Rosenhead), pp. 46–113. Oxford University Press.

LIU, Z. & HASEGAWA, Y. 2020 Estimation of turbulent channel flow based on time-series wall measurements. *Seisan Kenkyu* **72** (1), 5–8.

MAO, X., BLACKBURN, H.M. & SHERWIN, S.J. 2013 Calculation of global optimal initial and boundary perturbations for the linearised incompressible Navier–Stokes equations. *J. Comput. Phys.* **235**, 258–273.

MAO, X., ZAKI, T.A., SHERWIN, S.J. & BLACKBURN, H.M. 2017 Transition induced by linear and nonlinear perturbation growth in flow past a compressor blade. *J. Fluid Mech.* **820**, 604–632.

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

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

MONS, V., WANG, Q. & ZAKI, T.A. 2019 Kriging-enhanced ensemble variational data assimilation for scalar-source identification in turbulent environments. *J. Comput. Phys.* **398**, 108856.

MORÉ, J.J. & THUENTE, D.J. 1994 Line search algorithms with guaranteed sufficient decrease. *ACM Trans. Math. Softw.* **20** (3), 286–307.

NAGUIB, A.M., MORRISON, J.F. & ZAKI, T.A. 2010 On the relationship between the wall-shear-stress and transient-growth disturbances in a laminar boundary layer. *Phys. Fluids* **22** (5), 054103.

NIKITIN, N. 2018 Characteristics of the leading Lyapunov vector in a turbulent channel flow. *J. Fluid Mech.* **849**, 942–967.

NOCEDAL, J. 1980 Updating quasi-Newton matrices with limited storage. *Math. Comput.* **35** (151), 773–782.

ROSENFELD, M., KWAK, D. & VINOKUR, M. 1991 A fractional step solution method for the unsteady incompressible Navier–Stokes equations in generalized coordinate systems. *J. Comput. Phys.* **94**, 102–137.

SHARMA, A.S., MORRISON, J.F., MCKEON, B.J., LIMEBEER, D.J.N, KOBERG, W.H. & SHERWIN, S.J. 2011 Relaminarisation of $Re_\tau = 100$ channel flow with globally stabilising linear feedback control. *Phys. Fluids* **23** (12), 125105.

SHENG, J., MALKIEL, E. & KATZ, J. 2009 Buffer layer structures associated with extreme wall stress events in a smooth wall turbulent boundary layer. *J. Fluid Mech.* **633**, 17–60.

SMITS, A.J., MCKEON, B.J. & MARUSIC, I. 2011 High Reynolds number wall turbulence. *Annu. Rev. Fluid Mech.* **43** (1), 353–375.

SUZUKI, T. 2012 Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation. *J. Fluid Mech.* **709**, 249–288.

SUZUKI, T. & HASEGAWA, Y. 2017 Estimation of turbulent channel flow at $Re_\tau = 100$ based on the wall measurement using a simple sequential approach. *J. Fluid Mech.* **830**, 760–796.

VISHNAMPET, R., BODONY, D.J. & FREUND, J.B. 2015 A practical discrete-adjoint method for high-fidelity compressible turbulence simulations. *J. Comput. Phys.* **285**, 173–192.

WANG, M., WANG, Q. & ZAKI, T.A. 2019*a* Discrete adjoint of fractional-step incompressible Navier–Stokes solver in curvilinear coordinates and application to data assimilation. *J. Comput. Phys.* **396**, 427–450.

WANG, Q., HASEGAWA, Y. & ZAKI, T.A. 2019*b* Spatial reconstruction of steady scalar sources from remote measurements in turbulent flow. *J. Fluid Mech.* **870**, 316–352.

YOSHIDA, K., YAMAGUCHI, J. & KANEDA, Y. 2005 Regeneration of small eddies by data assimilation in turbulence. *Phys. Rev. Lett.* **94** (1), 014501.

ZAKI, T.A. 2013 From streaks to spots and on to turbulence: exploring the dynamics of boundary layer transition. *Flow Turbul. Combust.* **91** (3), 451–473.

ZAKI, T.A. & DURBIN, P.A. 2005 Mode interaction and the bypass route to transition. *J. Fluid Mech.* **531**, 85–111.

ZAKI, T.A., WISSINK, J.G., RODI, W. & DURBIN, P.A. 2010 Direct numerical simulations of transition in a compressor cascade: the influence of free-stream turbulence. *J. Fluid Mech.* **665**, 57–98.