Nudging-based data assimilation of the turbulent flow around a square cylinder

Markus Zauner, Vincent Mons, Olivier Marquet, Benjamin Leclaire

Nudging-based data assimilation of the turbulent flow around a square cylinder

M. Zauner[1], V. Mons[1]†, O. Marquet[1], and B. Leclaire[1]

[1]DAAA, ONERA, Université Paris Saclay, F-92190 Meudon, France

(Received xx; revised xx; accepted xx)

We investigate the estimation of the turbulent flow around a canonical square cylinder at $Re = 22000$ based on temporally-resolved but spatially sparse velocity data and solving the unsteady Reynolds-averaged Navier-Stokes (URANS) equations. Flow reconstruction from sparse data is achieved through the application of a nudging data assimilation technique. It involves the introduction of a feedback control term in the momentum equations which allows to drive URANS predictions towards reference data, which are here extracted from a direct numerical simulation (DNS). Such a data assimilation approach induces negligible supplementary computational cost compared to that of a standard URANS simulation. The influence of the spatial resolution of the reference data on the reconstruction performances is systematically investigated. Using a spacing of the order of one cylinder length between data points, we already observe synchronisation of the low-frequency vortex shedding between the full reference flow and the one that is estimated by URANS. The present data assimilation procedure allows to compensate for deficiencies in standard URANS calculations and leads to a significant decrease in temporal and spectral errors as computed by spectral proper orthogonal decomposition. Furthermore, high accuracy in terms of mean flow prediction by URANS is achieved. When considering spacings between measurements that are of the order of the wavelength of the Kelvin-Helmholtz vortices, such phenomena in the shear layers at the top and bottom of the cylinder are correctly estimated, while they are not self-sustained in standard URANS. The influence of the structure of the feedback control term in the data assimilation procedure is also investigated.

Key words:

1. Introduction

Fine characterisation of turbulent flows is usually performed via experiments, scale-resolving numerical approaches, or simulation methods relying on turbulence models. Experiments reproduce the entire complexity of flow physics, but the spatial and temporal resolutions of experimental data are generally limited, which prevents from a full characterisation of the flow of interest. Direct numerical simulations (DNS) may provide such detailed results with high spatio-temporal resolution for the full flow field. But those scale-resolving simulations require massive computational power, that may even be excessive when investigating high Reynolds number flows of interest for the aeronautics, civil engineering or automotive industry. Lower-fidelity steady or unsteady methods based on Reynolds-averaged Navier-Stokes (RANS) equations are popular low-cost alternative

† Email address for correspondence: vincent.mons@onera.fr

approaches. However, the accuracy of such computations may largely be affected by the fidelity of the closure model for the Reynolds stress tensor.

In order to overcome the respective limitations of the above mentioned approaches, data assimilation (Hayase 2015*b*) appears as a valuable tool. Data assimilation consists in merging experimental, or reference data with computationally efficient numerical models. On the one hand, it allows to infer or compensate for uncertainties in input and/or model parameters in simulations. On the other hand, it provides an augmented and full flow estimation from possibly sparse reference data. Popular state-of-the-art data assimilation techniques include adjoint-based (Le Dimet & Talagrand 1986) and (ensemble) Kalman filter-based methodologies (Evensen 2009). Adjoint-based or variational data assimilation is formulated as an optimal control problem where model parameters, which may vary from initial conditions (Chandramouli *et al.* 2020; Li *et al.* 2020; Wang & Zaki 2021) to space-dependent model corrections (Franceschini *et al.* 2020), are optimised in order to match measurements. Kalman filter techniques are derived in a stochastic formulation of data assimilation and correspond to sequential methodologies, where the flow state is directly updated in time based on data (Colburn *et al.* 2011; Da Silva & Colonius 2020; Moldodovan *et al.* 2021; Suzuki 2012). These techniques are thus more straightforwardly designed for state estimation, while adjoint-based data assimilation is also suited to solve inverse problems. Despite their significant differences in terms of formulation and implementation, these two classes of approaches may require similar computation costs, which are typically of the order of 100 baseline simulations (Mons *et al.* 2016). Such costs are induced by the number of direct and adjoint calculations that are required to reach convergence in variational techniques, while in Kalman filtering they originate from the need of propagating the covariance matrix that is associated to the estimated flow.

Exploring a cost-efficient alternative data-assimilation methodology focusing on state estimation, we here consider the so-called nudging technique (Hoke & Anthes 1976; Lakshmivarahan & Lewis 2013). It may also be referred to as state observer technique in the framework of control theory, or more specifically measurement-integrated simulation (Hayase 2015*a*) in the context of computational fluid dynamics (CFD). Nudging consists of adding a feedback term to the flow governing equations that acts at specified measurement locations (nudging points) and is directly proportional to the difference between reference data and numerical prediction. In contrast to previously mentioned data assimilation approaches, the supplementary computational cost that is associated to nudging is essentially negligible compared to that of a standard simulation.

Incompressible flows have generally been considered in previous studies discussed below, and, depending on the considered type of observation (namely velocity or pressure measurements), the feedback term may act on the momentum equations (Imagawa & Hayase 2010) and/or on the pressure Poisson equation (Neeteson & Rival 2020). The intensity of this feedback term is tuned through a single scalar, which thus allows to adjust to which extent the simulation is driven towards the reference data. As such, nudging may be considered as a simplification of the Kalman filter. Indeed, Kalman filtering and derived sequential data-assimilation techniques also amount to introduce a feedback term in the governing equations which involves the discrepancies between observations and prediction (Lewis *et al.* 2008). In this case, the feedback term is however weighted through a matrix gain instead of a single scalar. Nudging may also be categorised as a special case of proportional–integral–derivative (PID) controller that only includes the proportional contribution. The benefits of considering temporal integral and derivative of the discrepancies between observations and prediction have been recently investigated by Neeteson & Rival (2020); Saredi *et al.* (2021).

In the context of CFD, the application of nudging was first performed in a purely

numerical context by Hayase & Hayashi (1997); Imagawa & Hayase (2010) considering the turbulent flow in a square pipe. Based on punctual observations of the velocity field of a reference DNS, nudging was employed to drive a second simulation towards the reference one. Nudging was able to compensate for uncertainties in the initial condition (Imagawa & Hayase 2010) or numerical errors due to the use of a coarser grid in the second simulation (Hayase & Hayashi 1997), and thus successfully reproduced the reference solution. Neeteson & Rival (2020) also considered nudging and more general PID controllers to drive a coarse-grid simulation based on synthetic pressure data of the flow past a square cylinder at $Re = 100$. Suzuki & Hasegawa (2017) combined a nudging/state observer-like approach with linear stochastic estimation to reconstruct turbulent channel flows at friction Reynolds number $Re_\tau = 100$ from wall measurements, the same DNS setup was used to both generate data and perform estimation. In Di Leoni *et al.* (2020), nudging was applied to homogeneous turbulence, still relying on DNS. In particular, nudging was used to recover a reference forced isotropic turbulent flow, assuming that the forcing of the reference flow was unknown. Various types of observations of the reference flow were considered among Eulerian observations, as in Hayase & Hayashi (1997); Imagawa & Hayase (2010); Neeteson & Rival (2020), Lagrangian data, or in spectral space. The effect of the observation density on the ability of nudging in recovering unmeasured scales was also investigated. While the reference data were generated with the same numerical methods as those employed for data assimilation in the above studies, nudging in conjunction with a turbulence model was investigated by Buzzicotti & Di Leoni (2020) in the context of large eddy simulations of isotropic turbulence. Filtered reference data were generated by DNS, and nudging was performed assimilating the full reference data at all resolved scales, and was thus only useful to correct model errors.

Aside from these purely numerical studies, Nisugi *et al.* (2004); Yamagata *et al.* (2008) demonstrated the ability of nudging to deal with various types of experimental data of the flow past a square cylinder at $Re = 1200$, such as wall pressure and particle image velocimetry (PIV) measurements, based on the laminar Navier-Stokes equations. Suzuki *et al.* (2009) performed the assimilation of rearranged particle tracking velocimetry (PTV) measurements of flows past airfoils at low Reynolds numbers. To efficiently handle experimental data of higher Reynolds number flows, Nakao *et al.* (2009) considered relying on unsteady RANS (URANS) modelling to perform nudging. Recently, steady nudging in conjunction with RANS was performed by Saredi *et al.* (2021) based on three-dimensional PIV measurements of the mean flow around a wall-mounted bluff obstacle. Nudging proved to be effective in correcting deficiencies in the RANS prediction, such as the overestimation of the extent of the recirculation region past the obstacle. However, while these few studies confirmed the feasibility of applying nudging to experimental data and ambitious flow configurations, the use of only experimental data precluded a quantitative assessment of the improvement in the estimation of the considered high Reynolds number flows outside of the measurement locations.

In the present contribution, we aim to assess in detail the full potential of the nudging approach for unsteady flow estimation in an aerodynamic context, based on URANS modelling in view of complex applications and synthetic velocity data. The unsteady turbulent flow around a square cylinder at $Re = 22000$ is chosen as flow configuration (Lyn *et al.* 1995; Trias *et al.* 2015) because it exhibits several flow phenomena occurring at various time and spatial scales that are of interest for aeronautical application. They include low-frequency quasi-periodic vortex shedding (VS) past the square cylinder, along with broad-band high-frequency Kelvin-Helmholtz (KH) fluctuations in the shear layers at the top and bottom sides of the cylinder. For the present flow configuration, Iaccarino *et al.* (2003) showed that URANS simulations allow to capture the periodic

vortex shedding but slightly mispredict the associated Strouhal number and spatial structures. On the other hand, URANS does not necessarily capture the small-scale high-frequency fluctuations that are associated to Kelvin-Helmholtz vortices. This applies in particular when relying on eddy viscosity-based models such as the Spalart-Allmaras model (Spalart & Allmaras 1994), which is employed here. According to Menter *et al.* (2010), this inability in capturing the amplification of flow phenomena such as shear-layer instabilities is not directly due to the Reynolds (or phase) averaging operation but rather to the quality of the turbulence model (Menter *et al.* 2010). For instance, Palkin *et al.* (2016) compared results of URANS calculations of the flow past a circular cylinder at $Re = 1.4 \times 10^5$ which were obtained with Reynolds-stress and eddy viscosity-based turbulence models. Thanks to a lower effective eddy-viscosity, Reynolds-stress model-based calculations satisfactorily predicted the high-frequency content ($1 \leqslant St \leqslant 10$) of the flow, while the energy of these fluctuations was largely underestimated by eddy-viscosity models. The point of view that the characteristics of the turbulence model determine the portion of turbulent scales that are resolved is shared by various hybrid-simulation approaches such as scale-adaptive simulations (Menter & Egorov 2010) and partially-averaged Navier-Stokes models (Girimaji 2006), among others, and will be adopted here.

Interestingly, to compensate the above-mentioned deficiencies of too-dissipative hybrid/RANS models, Menter *et al.* (2010) also proposed to introduce a forcing in the momentum equations that sustain and enable the emergence of unsteady flow phenomena such as shear-layer instabilities by transferring turbulence energy from the model to the resolved flow. Following this path, we will first demonstrate the ability of the URANS equations (complemented with the Spalart-Allmaras model) in capturing high-frequency fluctuations by performing white-noise pointwise-forced simulations. After characterising the frequencies and spatial structures that are most amplified and emerge through the model, the accurate estimation of these high-frequency fluctuations will be achieved using the nudging approach, for which the feedback term also forces the URANS simulations, but at several locations and with amplitude and frequency content that correspond to the measurement data.

The objective of the present contribution is thus to assess the potential of the nudging method in unsteady flow reconstruction based on the relatively affordable URANS approach, which could not alone provide an accurate prediction of both low- and higher-frequency features of the flow. It may be emphasized that the role of nudging will however not consist in correcting the employed RANS model, but rather in directly updating the URANS prediction based on data, focusing on state estimation only. Synthetic observations for the considered two-dimensional URANS simulations will be generated through a spanwise average of a three-dimensional DNS of the present flow configuration. Nudging will then be employed to reconstruct the full reference flow from sparse velocity observations, whose spatial density will be systematically varied. A crucial aspect of the present study will be to correctly reproduce flow structures evolving coherently in space and time. This will be assessed through the spectral proper orthogonal decomposition (Towne *et al.* 2018).

The paper is organised as follows. Standard numerical simulations of the turbulent flow around a square cylinder are described in §2. After introducing the numerical methods used for DNS and URANS simulations in §2.1 and §2.2, respectively, we describe methods to analyse dynamic as well as mean properties of flow fields in §2.3. DNS and standard URANS results are discussed in §2.4. The potential of nudged URANS simulations is then discussed in §3. After a description of the nudging method and used data-sets of measurements in §3.1, the present approach is validated using a dense set of measurements in §3.2. Then, the performance of nudging in the estimation of low- and high-frequency

flow phenomena based on sparse data is discussed in sections §3.3 and §3.4, respectively. Variations in the structure of the feedback term in the nudging methodology and its impact on flow reconstruction are investigated in §3.5. Implications for experimental studies are discussed in §3.6. Finally, concluding remarks are drawn in §4.

2. Standard URANS simulations compared to DNS

We investigate the turbulent flow developing around a square cylinder (or square prism) of length D facing an upstream uniform flow of velocity U_∞. The Reynolds number based on these reference length and velocity scales is $Re = U_\infty D/\nu = 2.2 \cdot 10^4$. Three-dimensional direct numerical simulation (DNS) presented in §2.1 captures the full range of flow scales, unlike standard two-dimensional unsteady Reynolds-Averaged Navier-Stokes (URANS) equations introduced in §2.2 that may capture only a portion of the unsteady flow structures. A turbulence model then accounts for the effect of incoherent and unresolved structures so that the overall computational costs are significantly lower for URANS computations compared to DNS. In order to assess the quality of URANS predictions, instantaneous DNS snapshots are averaged in the spanwise direction and provide reference data. After describing respectively temporal and modal diagnosis tools in §2.3, detailed comparison between DNS and URANS simulations is provided in §2.4.

2.1. *Direct numerical simulations*

Three-dimensional DNS of the turbulent flow around the square cylinder are performed with the ONERA's compressible solver FastS (Dandois *et al.* 2018; Mary & Sagaut 2002). The three-dimensional computational domain is formed by a two-dimensional circle of radius $50D$ in the $\boldsymbol{x} = (x, y)^{\mathrm{T}}$ plane, centered around the square cylinder, which is extruded in the spanwise homogeneous direction z with length $L_z = 4D$. The two-dimensional mesh consists of approximately $2.6 \cdot 10^5$ grids points and $N_z = 960$ planes uniformly discretise the spanwise direction. Periodic boundary conditions are applied in the spanwise direction. Homogeneous freestream conditions are enforced at the outer boundaries, where only the horizontal velocity component (parallel to upper and lower cylinder surfaces) is non-zero and equal to U_∞. The simulation is run close to incompressible flow regime at $M = 0.1$ enforcing no-slip conditions at the cylinder surface. In the following, all quantities are non-dimensionalised based on the reference length D and free stream parameters U_∞, ρ_∞. The DNS time step is $\Delta t_{DNS} = 3.3188 \cdot 10^{-4}$, which corresponds to a maximal CFL number of approximately 0.7. More details about this DNS solver are provided in Dandois *et al.* (2018); Franceschini *et al.* (2020); Mary & Sagaut (2002).

Obtained unsteady three-dimensional velocity fields $\boldsymbol{u}$ can be decomposed as

$$\boldsymbol{u}(\boldsymbol{x}, z, t) = \bar{\boldsymbol{u}}_r(\boldsymbol{x}, t) + \boldsymbol{u}'(\boldsymbol{x}, z, t) \tag{2.1}$$

where $\bar{\boldsymbol{u}}_r$ is the statistically-averaged two-dimensional velocity field of interest while $\boldsymbol{u}'$ refers to the associated three-dimensional deviation, not used in the following. Then, it is assumed that $\bar{\boldsymbol{u}}_r$ can be extracted from DNS velocity fields $\boldsymbol{u}$ by averaging instantaneous snapshots in the spanwise direction according to

$$\bar{\boldsymbol{u}}_r(\boldsymbol{x}, t) = \frac{1}{L_z} \int_0^{L_z} \boldsymbol{u}(\boldsymbol{x}, z, t)\, dz \sim \frac{1}{N_z} \sum_{k=1}^{N_z} \boldsymbol{u}(\boldsymbol{x}, z = k\Delta z, t) \tag{2.2}$$

where $\Delta z \sim 0.004$ is the uniform discretisation step in the spanwise direction. This spatial average well approximates the statistical average if the spanwise extent L_z of

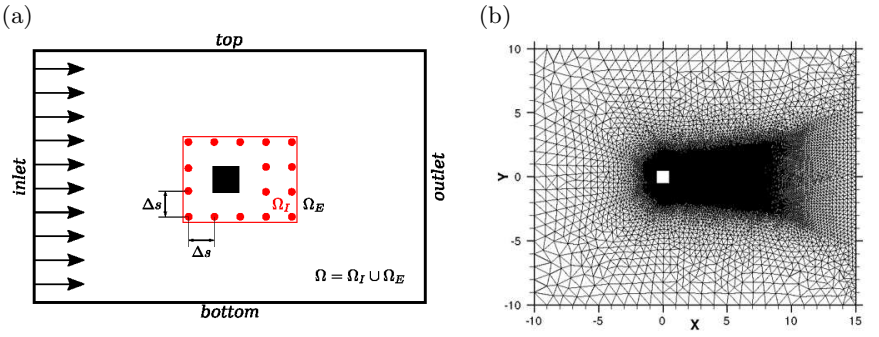

FIGURE 1. (a) Sketch of the computational domain Ω for URANS simulations, where red dots denote positions $(\boldsymbol{x}_k)_{k=1\cdots M}$ of velocity measurements in the internal domain Ω_I (bounded by the red frame). These so-called nudging points are uniformly separated in both directions by a spacing-distance Δs. The domain that is external to the measurement area is denoted as Ω_E. (b) Typical triangular mesh used for the standard and nudged URANS simulations around a square cylinder.

the computational domain and the number of planes N_z are large enough. This two-dimensional field extracted from DNS will be used as reference solution (hence the subscript r), first to assess the accuracy of the standard URANS simulation methods in §2.4, and will then form the targeted flow that we aim to reconstruct based on sparse observations thanks to the nudging method in §3. To complete the description of this reference data set, it is obtained by down-sampling the DNS snapshots every 63 time steps. The time interval Δt_r between snapshots in this reference data set is thus $\Delta t_r = 63 \times \Delta t_{DNS} = 0.0209$ convective time units.

2.2. *Unsteady Reynolds-averaged Navier-Stokes simulations*

Rather than extracting the temporal variation of the spatially-averaged flow from DNS simulations, one may rather estimate the latter through the two-dimensional URANS equations which are supplemented with a turbulence model to take into account the effect of the non-resolved component. The two-dimensional resolved velocity ($\bar{\boldsymbol{u}}$) and pressure ($\bar{p}$) fields then satisfy the following (incompressible) flow equations

$$\frac{\partial \bar{\boldsymbol{u}}}{\partial t} + (\bar{\boldsymbol{u}} \cdot \boldsymbol{\nabla})\bar{\boldsymbol{u}} + \boldsymbol{\nabla}\bar{p} - \boldsymbol{\nabla} \cdot \left[2\left(Re^{-1} + \nu_t(\tilde{\nu})\right)\boldsymbol{\nabla}_{\mathrm{s}}\bar{\boldsymbol{u}}\right] = \boldsymbol{0}\,, \qquad \boldsymbol{\nabla} \cdot \bar{\boldsymbol{u}} = 0, \tag{2.3}$$

where $\boldsymbol{\nabla}_{\mathrm{s}}\bar{\boldsymbol{u}}$ refers to the symmetric part of the velocity gradient and ν_t denotes the turbulent eddy-viscosity. Here, the latter is determined using the Spalart-Allmaras model (Spalart & Allmaras 1994), where the following governing equation for the intermediate eddy-viscosity-like variable $\tilde{\nu}$ is solved

$$\frac{\partial \tilde{\nu}}{\partial t} + (\bar{\boldsymbol{u}} \cdot \boldsymbol{\nabla})\tilde{\nu} = s(\tilde{\nu}, \bar{\boldsymbol{u}}). \tag{2.4}$$

The source term s accounts for the production, dissipation and diffusion terms for $\tilde{\nu}$. Details about the present implementation of the Spalart-Allmaras model are provided in Franceschini *et al.* (2020).

The URANS equations (2.3)-(2.4) are discretised in space using the finite-element method implemented in the software FreeFEM++ (Hecht 2012). Second-order polynomial elements (P_2) are employed for the velocity field, while first-order elements (P_1) are considered for the pressure and eddy-viscosity variables. For the relatively high

Reynolds number flow of the present contribution ($Re = 22000$), streamline-upwind Petrov-Galerkin (SUPG) (Brooks & Hughes 1982) and grad-div (Olshanskii *et al.* 2009) stabilisations are implemented. The rectangular computational domain is sketched in figure 1(a) with an inlet located at $x = -10$, an outlet at $x = 15$ and top and bottom planes at $y = 10$ and $y = -10$, respectively. Remaining details of figure 1(a) will be discussed later when introducing the nudging approach. The unstructured mesh, shown in figure 1(b), consists of triangles with a total number of 46228 nodes. At the inlet, the boundary condition $(\bar{u}, \bar{v}, \tilde{\nu}) = (1, 0, 0)$ is imposed. At the cylinder surface, the no-slip boundary condition $\bar{\boldsymbol{u}} = \boldsymbol{0}$ is enforced in conjunction with $\tilde{\nu} = 0$. The conditions $\frac{\partial \tilde{\nu}}{\partial x} = 0$ and $2(Re^{-1} + \nu_t)\boldsymbol{\nabla}_{\mathrm{s}}\bar{\boldsymbol{u}} \cdot \boldsymbol{n} - \bar{p}\boldsymbol{n} = \boldsymbol{0}$ are used at the outlet, where $\boldsymbol{n}$ denotes the normal vector. At the top and bottom boundaries, symmetry conditions are imposed according to $(\frac{\partial \bar{u}}{\partial y}, \bar{v}, \frac{\partial \tilde{\nu}}{\partial y}) = \boldsymbol{0}$.

Time integration is performed in a fully implicit way based on a second-order accurate finite-difference approximation of the time derivative. A quasi-Newton method is used to solve the resulting nonlinear problem at each iteration. The time step of URANS simulation is denoted by Δt. Results of the standard simulations presented in §2.4 are obtained with $\Delta t = \Delta t_r$, i.e. matching the time interval of the reference data set. For nudged URANS simulations in §3, the value of Δt will be equal to the time interval of data sets (not necessarily the reference data set) inserted into the simulations through the feedback term. By varying the sampling of these data sets, we will thus modify the time step Δt.

2.3. *Temporal and spectral errors*

We introduce in this paragraph various mathematical definitions for assessing errors between the reference data set and the output of the standard or nudged URANS simulations. Note that the latter have not yet been introduced, but we will use the same notation for denoting the corresponding variables in both simulations. In the following, both standard and nudged URANS solutions may be referred to as the estimated flow, while the field $\bar{\boldsymbol{u}}_r$ which is extracted from DNS in (2.2) is referred to as the reference flow, as mentioned above. Definitions of the instantaneous and time-averaged errors are introduced in this section before considering the spectral error based on a SPOD analysis of the data.

The instantaneous error field $e(\boldsymbol{x}, t)$ between the estimated velocity field $\bar{\boldsymbol{u}}$ and the reference one $\bar{\boldsymbol{u}}_r$ is defined as

$$e(\boldsymbol{x}, t) = \left((\bar{u}_r - \bar{u})^2 + (\bar{v}_r - \bar{v})^2\right)^{\frac{1}{2}}. \tag{2.5}$$

The instantaneous and time-averaged global errors are then respectively defined as

$$E(t) = \int_{\Omega} e(\boldsymbol{x}, t)^2 \, d\boldsymbol{x}, \quad \langle E \rangle = \frac{1}{T_e} \int_{t_i}^{t_i + T_e} E(t) \, dt, \tag{2.6}$$

where Ω denotes the entire computational domain, $t_i > 0$ is an arbitrary time that is chosen to avoid transient effects (occurring in $[0, t_i]$) and T_e refers to the considered time window for averaging the instantaneous global error. Note that the time average is denoted $\langle \cdot \rangle$, and thus $\langle \bar{\boldsymbol{u}} \rangle$ is the time-averaged velocity field of the estimated flow. Its error compared to the time-averaged reference velocity $\langle \bar{\boldsymbol{u}}_r \rangle$ is then defined as

$$e_{\langle \bar{\boldsymbol{u}} \rangle}(\boldsymbol{x}) = \left(\left(\langle \bar{u}_r \rangle - \langle \bar{u} \rangle\right)^2 + \left(\langle \bar{v}_r \rangle - \langle \bar{v} \rangle\right)^2\right)^{\frac{1}{2}}. \tag{2.7}$$

To better understand the unsteady effect, a spectral analysis of the data is performed

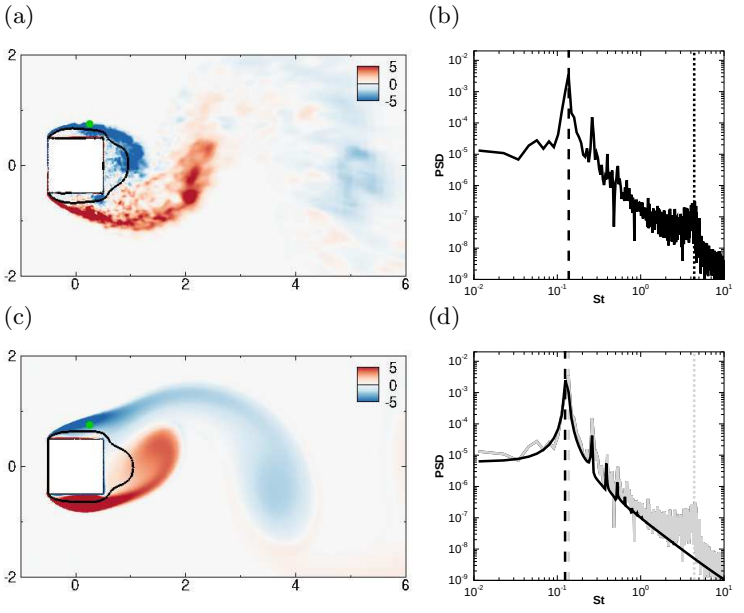

FIGURE 2. Results for (a-b) reference (DNS) and (c-d) estimated (standard URANS) flows. (a,c): Instantaneous spanwise vorticity field $\bar{\omega}_z$ at $t = 50$, where black iso-curves denote $\langle\overline{u}\rangle = 0$. (b,d): Fourier spectrum of the streamwise velocity at the green monitor point in (a,c) (full black line). For the sake of comparison, the DNS spectrum is duplicated in (d) with the grey curve. Vertical dashed and dotted lines indicate the low-frequency peak associated with large-scale vortex-shedding ($St_{VS} = 0.137$ for DNS while $St_{VS} = 0.126$ for URANS) and the high-frequency bump linked to Kelvin-Helmholtz instabilities ($St_{KH} = 4.384$ for DNS), respectively.

which relies on the spectral proper orthogonal decomposition (SPOD) described in Towne *et al.* (2018). Taking a temporal series of velocity snapshots as inputs, the SPOD analysis provides a set of SPOD modes $\boldsymbol{\Phi} = (\Phi_u, \Phi_v)^T$ that each oscillate at a single non-dimensional frequency (Strouhal number) $St = fD/U_\infty$. To each frequency corresponds a set of normalized SPOD modes ($\int_\Omega \boldsymbol{\Phi}^* \boldsymbol{\Phi} d\Omega = 1$, where the asterisk denotes the Hermitian transpose) ranked by their kinetic energy λ, that are the eigenvalues of the cross-spectral density tensor which is estimated from snapshots. In the following, we will focus on the dominant SPOD modes, namely which are associated to the largest eigenvalue, for frequencies of interest. Denoting hereinafter the SPOD modes of the reference and URANS data set as $[\lambda_r, \boldsymbol{\Phi}_r]$ and $[\lambda, \boldsymbol{\Phi}]$ respectively, a spectral error field may be defined as

$$\Phi_e = \left(\left(\sqrt{\Phi^*_{u,r}\Phi_{u,r}} - \sqrt{\Phi^*_u \Phi_u} \right)^2 + \left(\sqrt{\Phi^*_{v,r}\Phi_{v,r}} - \sqrt{\Phi^*_v \Phi_v} \right)^2 \right)^{\frac{1}{2}}. \tag{2.8}$$

In the following, the dynamical content of standard and nudged URANS simulations will thus be assessed and compared to the reference solution through their frequency content, the kinetic energy of the dominant SPOD modes, and the shape of the latter through (2.8).

2.4. *Comparisons between standard URANS and DNS*

A preliminary URANS simulation is performed to obtain a periodic evolution of the aerodynamic coefficients. It allows to determine an instant which minimises the discrepancy between the lift coefficients computed from URANS and DNS simulations.

The flow field at that instant is then used as initial condition $t = 0$ for the URANS simulation that is discussed below. By doing so, the DNS and URANS solutions are initially phased in time and we can thus properly compare the temporal evolution of the error between the two.

Figure 2 provides a first comparison of the flow extracted from DNS [figures 2(a-b)] and estimated by the URANS equations [figures 2(c-d)]. The instantaneous vorticity contours are shown in figures 2(a) and (c) for $t = 50$. In both cases, we observe that large-scale clockwise (blue) and counter-clockwise (red) vortices are shed in the wake of the square cylinder. Small-scale structures are also visible in the reference flow field shown in figure 2(a). In the upper and lower shear layers, emerging from the leading edge corners of the cylinder, they correspond to two-dimensional roll-up structures associated with Kelvin-Helmholtz instabilities. In the wake, they interact with the large-scale vortices of the von-Kármán vortex street. Such small-scale structures are clearly absent in the estimated vorticity field that is shown in figure 2(c). Large-scale and small-scales structures are associated with different parts of the Fourier spectra displayed in figures 2(b,d), which are obtained from the streamwise velocity u monitored at a grid point located in the upper shear layer, at $(x, y) = (0.25, 0.75)$ [green dot in figures 2(a,c)]. The large-scale periodic vortex shedding is associated with several peaks, one at the fundamental frequency $St_{VS} = U_\infty f_{VS}/D$ (indicated in both figures by the vertical dashed lines) and the others at multiples of the fundamental frequency, i.e. kSt_{VS} with $k \geqslant 2$. The fundamental frequency is $St_{VS} = 0.137$ for the DNS simulation and $St_{VS} = 0.126$ for the standard URANS simulation. The small-scale structures that are observed in the shear layers in the reference snapshot (figure 2(a)) correspond to a high-frequency broadband bump which is centered around $St_{KH} \approx 4.384$, as indicated by the vertical dotted line in figure 2(b). This high-frequency bump is absent in the spectrum of the URANS signal, displayed with the black curve in figure 2(d). This confirms that the present standard URANS simulation does not capture the emission of Kelvin-Helmholtz vortices in the shear layers, at least in a self-sustained way, which will be further discussed at the end of this section.

Although the URANS simulation well approximates the periodic vortex-shedding phenomenon, the difference between Strouhal numbers noticed above is responsible for desynchronisation of the wake flow as illustrated in figure 3. The instantaneous vorticity fields displayed in figures 3(a-c) and (d-f) for DNS and URANS simulations, respectively, seem in phase after 2 vortex-shedding periods (a,d), but, after 4 shedding cycles (b,e), large vortex structures observed in URANS snapshots clearly lag behind corresponding structures in DNS snapshots. After 19.5 periods (c,f), the two solutions are out-of-phase by a half-wavelength. This desynchronisation of the large-scale structures results in large values of the instantaneous local error in the wake of the cylinder, as illustrated by figures 3(g-i). The temporal evolution of the instantaneous global error $E(t)$ in (2.6) is reported in figure 3(j) from $t = 0$ to $t = 200$, which corresponds to a time window of 26 vortex-shedding periods. The error $E(t)$ increases during the first 20 periods ($0 \leqslant t \leqslant 150$), in agreement with the previous observations, before slightly decreasing. All the above findings are thus related to the lack of synchronisation between reference and estimated flows, which originates from the misprediction of the fundamental frequency St_{VS} by URANS. This may already justify the need of sequential data assimilation for full flow reconstruction, as only phasing an initial condition from data is not sufficient in the present case to correctly track the reference flow.

The statistical and dynamic discrepancies between estimated and reference flows are further examined in the following, where we first consider the time average of these flows. The streamwise velocity of the time-averaged reference (DNS) and estimated (URANS)

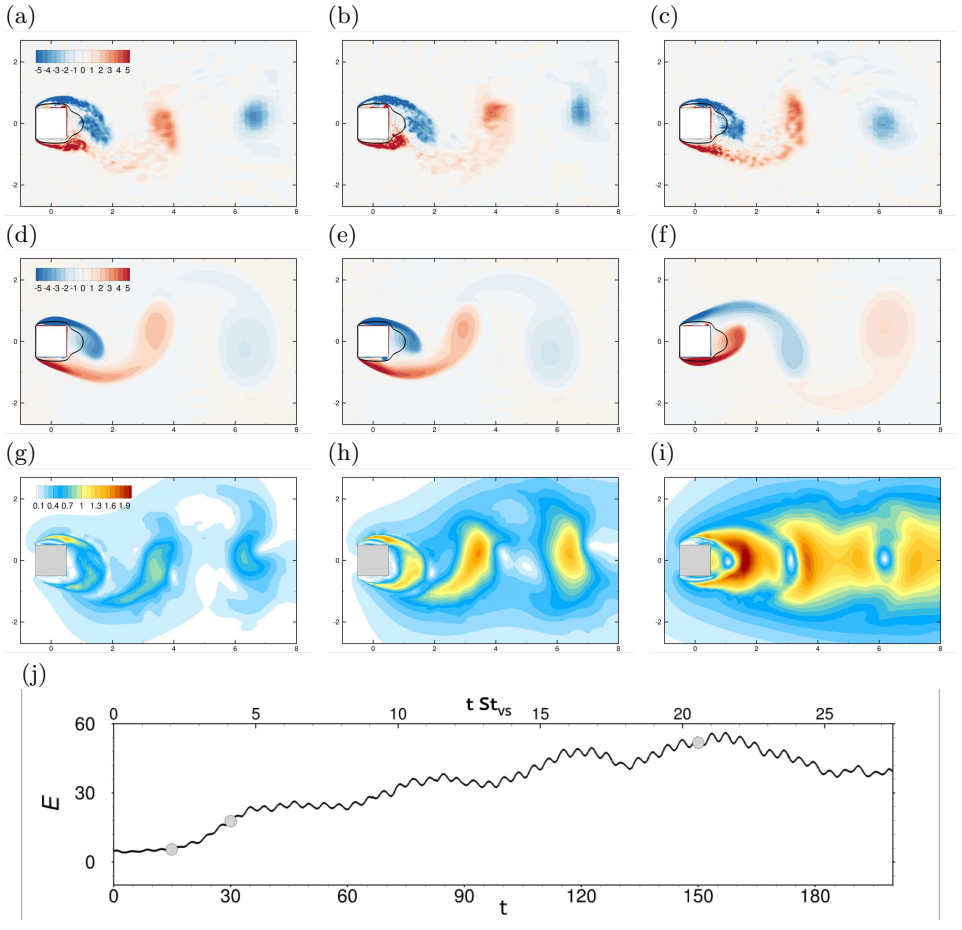

FIGURE 3. Instantaneous vorticity fields of (a-c) the reference flow $\bar{\boldsymbol{u}}_r$ (DNS) and (d-f) estimated flow $\bar{\boldsymbol{u}}$ (standard URANS) at times (a,d) $t = 15$, (b,e) $t = 30$ and (c,f) $t = 150$. (g-i) Instantaneous error fields $e(\boldsymbol{x}, t)$ between the reference and estimated flow (definition in eq. 2.5) at the corresponding instants. (j) Temporal evolution of the global error $E(t)$ (see eq. 2.6) where the bottom axis reports the (nondimensional) time while the top axis reports the number of low-frequency cycles determined as $t/\tau_{VS} = t\ St_{VS}$ with $St_{VS} = 0.137$. The grey circles indicates time $t = 15$, 30 and 150.

flows are depicted in figures 4(a) and 4(b), respectively, while the associated discrepancy field $e_{\langle\bar{\boldsymbol{u}}\rangle}(\boldsymbol{x})$ in (2.7) is displayed in figure 4(c). Interestingly, the mean flows agree qualitatively well, and the error field $e_{\langle\bar{\boldsymbol{u}}\rangle}(\boldsymbol{x})$ reaches overall significantly lower values than the discrepancy field $e(\boldsymbol{x}, t)$ for the instantaneous flow in figures 3(g-i), at least in the near wake. Largest errors are observed in the shear layers that emerge from the cylinder upstream corners. The recirculation region, which is delineated by black curves, is slightly larger for the estimated flow ($L = 0.55$) compared to that of the reference flow ($L = 0.45$). Substantial errors are also visible in the far wake where streamwise velocity is overestimated by URANS.

The SPOD decomposition of the reference and estimated flows are both performed by considering as input 8420 snapshots that are sampled at $\Delta t_r = 0.021$ in a time interval of around 176 convective times. Using the Matlab implementation provided by Towne *et al.* (2018), this series is typically divided in 3 overlapping bins, each bin containing about

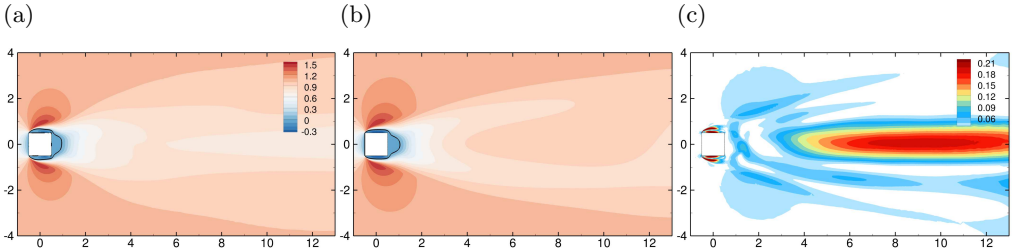

FIGURE 4. (a-b) Streamwise component of the time-averaged (a) reference flow $\langle \bar{\boldsymbol{u}}_r \rangle$ (DNS) and (b) estimated flow $\langle \bar{\boldsymbol{u}} \rangle$ (standard URANS). (c) Associated discrepancy field $e_{\langle \bar{\boldsymbol{u}} \rangle}$ as defined in (2.7). In (a-b), black curves correspond to contours of zero streamwise time-averaged velocity, which delineate the recirculation region.

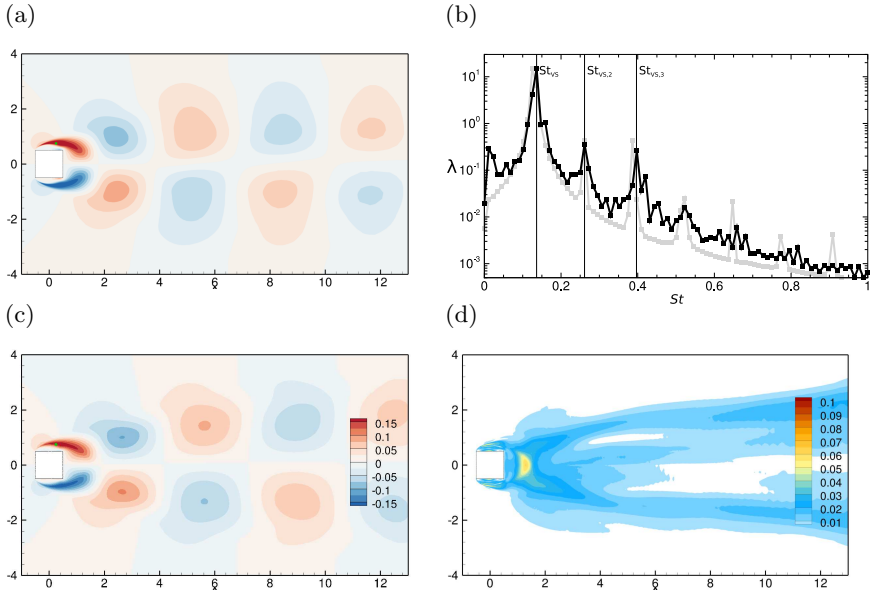

FIGURE 5. SPOD analysis of the vortex-shedding phenomenon in the reference (DNS) and estimated (standard URANS) flows: streamwise velocity contours of the imaginary part of the dominant (a) DNS and (c) URANS SPOD modes for the fundamental frequency St_{VS}. Modes are phased so that each of them reaches its maximum value at the green monitor point. (b) Spectra showing the largest eigenvalue λ as a function of the frequency St in the low-frequency range $0 \leqslant St \leqslant 1$ for DNS (black curve) and URANS (grey curve). The vertical lines indicate low-frequency peaks at $St_{VS} = 0.137$ and its harmonics for DNS results. The modal discrepancy field Φ_e in (2.8) between the modes in (a) and (c) is reported in (d).

4200 instantaneous snapshots, and overlapping by 50% of their size. The time interval of a bin is thus around 88 convective time units.

The SPOD analysis of the vortex-shedding phenomenon as predicted by DNS and standard URANS is first illustrated in figure 5. Spectra of the most-energetic SPOD modes for a low-frequency range ($0 \leqslant St \leqslant 1$) as functions of Strouhal number St are reported in figure 5(b). DNS and URANS results are denoted by black and light grey dots, respectively. As in the Fourier analysis of figure 2, several peaks are observed at the fundamental frequency $St_{VS} = 0.137$ of the vortex-shedding phenomenon and at frequencies corresponding to second and third harmonics at $St_{VS,2} = 0.261$ and $St_{VS,2} = 0.398$, respectively. Frequencies of URANS modes (grey peaks) are slightly lower, peaking

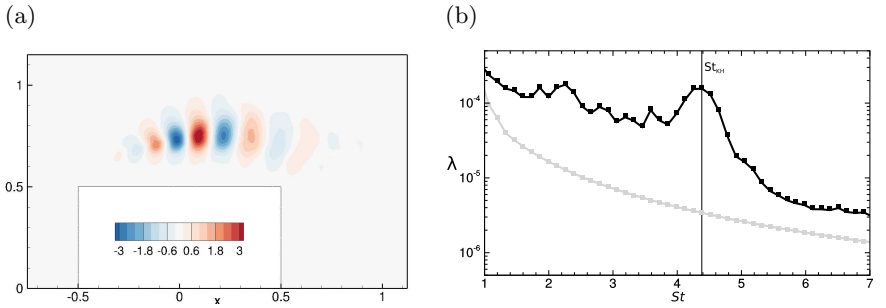

FIGURE 6. (a) Contours of transverse velocity fluctuations are shown for the real part of the dominant SPOD mode of DNS that is associated to Kelvin-Helmholtz instabilities ($St_{KH} = 4.384$, only captured with DNS). (b) Spectra showing the largest eigenvalue λ as a function of the frequency St in the high-frequency range $1 \leqslant St \leqslant 7$. Black and light grey curves correspond to DNS and URANS results, respectively.

at $St_{VS} = 0.126$, $St_{VS,2} = 0.261$, and $St_{VS,2} = 0.386$. Streamwise velocity contours of the imaginary part of the dominant DNS and URANS SPOD modes for the fundamental frequency St_{VS} are displayed in figures 5(a) and 5(c), respectively. Both modes are phased so that they reach their maximum amplitude at the monitor point indicated by the green dot, which is the same as in figure 2. Their spatial structure is typical of modes corresponding to the alternate shedding of vortices in the wake, with a streamwise wavelength equal to ~ 4 cylinder lengths. Furthermore, we show in figure 5(d) the so-called modal error field $\Phi_e(\boldsymbol{x})$, which is defined in equation (2.8). Dominant errors are observed in boundary and shear layers, as well as in the near-wake region, which may be related to the slightly longer recirculation region predicted by URANS.

Figure 6(b) shows the spectrum of the most dominant SPOD modes corresponding to high-frequency fluctuations. A broadband peak centered around $St_{KH} = 4.384$ is clearly visible in the DNS spectrum (black) but not in the URANS one (grey). In order to prevent modes from being corrupted by numerical oscillations in regions of decreased grid resolution, we limit the domain that is considered for SPOD targeting Kelvin-Helmholtz phenomena to $-1.5 < x < 1.5$ and $-1.5 < y < 1.5$. Furthermore, we consider 1440 snapshots at a sampling rate of $\Delta t_r = 0.021$ in a time interval of around 30 convective times using 6 overlapping bins, each containing about 33 Kelvin-Helmholtz cycles. In order to better visualise convective small-scale instabilities originating from the front edge of the cylinder, a magnification of the upper-side shear layer in figure 6(a) shows the real part of transverse velocity contours for the dominant SPOD mode from DNS results at $St_{KH} = 4.384$. Typical wavelengths of these structures are around 0.25, i.e. a quarter of the cylinder's length. Similar spatial structures are also observed on the lower side of the cylinder, but of much weaker magnitude. As expected when considering the spatial symmetry of the time-averaged flow, a second SPOD mode with spatial structures predominant on the lower-side shear layer is also obtained but associated to a slightly lower kinetic energy.

As the present standard URANS results do not include such structures, one may suspect the Spalart-Allmaras model to be here too dissipative for self-sustained Kelvin-Helmholtz instabilities, as discussed in the introduction. Following proposals in Menter *et al.* (2010), one may wonder if the Spalart-Allmaras model may at least allow the emergence of Kelvin-Helmholtz phenomena from sustained disturbances in the present case. This is investigated through figure 7 showing SPOD results for a URANS calculation, where uniform white noise is injected at $(x, y) = (-0.4, 0.6)$ (black dot in the

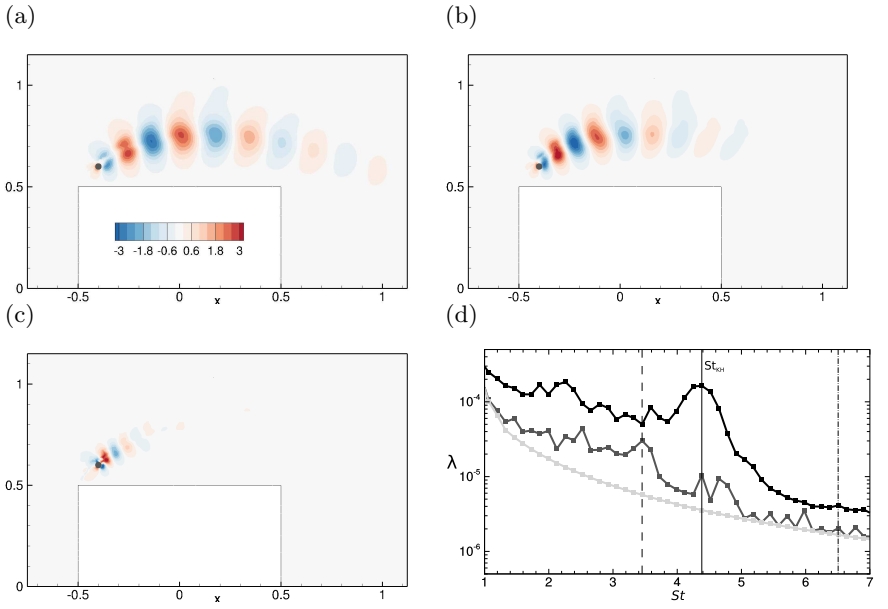

FIGURE 7. (a)-(c) Contours of transverse velocity fluctuations are shown for the real part of the dominant SPOD mode at (a) $St = 3.454$ (vertical dashed line in (d)), (b) $St = St_{KH} = 4.384$ (vertical full line in (d)) and (c) $St = 6.510$ (vertical dash-dotted line in (d)) for a URANS calculation in which white noise is injected at the black dot. (d) Spectra showing the largest eigenvalue λ as a function of the frequency St in the high-frequency range $1 \leqslant St \leqslant 7$. Black and light grey curves correspond to DNS and unperturbed URANS results, respectively, while the dark grey curve refers to the URANS calculation with white noise.

figure). More specifically, the spectrum range of the white noise is limited to $St \geqslant 0.5$ in order to avoid interference with low-frequency vortex shedding. Independent realisations of this white noise affect both the streamwise and transverse velocity components at $(x, y) = (-0.4, 0.6)$, and the associated standard deviation is $5 \cdot 10^{-3}$. Qualitatively, the results in figure 7 exhibit minor sensitivity with respect to small changes in the amplitude and location of the noise source (not shown here for the sake of brevity). From this punctual disturbance, figures (a-b) confirm the development of Kelvin-Helmholtz-like structures. These figures, which may be compared to the reference mode in figure 6(a), correspond to dominant SPOD modes at $St = 3.454$ and $St = St_{KH} = 4.384$. We can observe significantly increased energy content at these frequencies, especially for $St = 3.454$, comparing spectra of perturbed (dark grey curve) and unperturbed (light grey curve) URANS simulations in figure 7(d). On the other hand, higher frequencies ($St \geqslant 5$) are not sustained by the model and the corresponding fluctuations remain localised around the source of white noise, as illustrated by figure 7(c), which reports the dominant SPOD mode at $St = 6.510$. While these results do not necessarily agree with the DNS ones on a quantitative basis, they confirm the ability of the URANS equations in frequency selection and in developing shear-layer instabilities. These findings thus support the relevance of employing the present RANS model in the following data assimilation procedure even for completing reference measurements for the estimation of Kelvin-Helmholtz phenomena.

Previously discussed SPOD results for DNS and (unperturbed) standard URANS are summarised in table 1.

	ΔC_L	$\overline{C_D}$	St_{VS}	λ_{VS}	$St_{VS,2}$	$\lambda_{VS,2}$	$St_{VS,3}$	$\lambda_{VS,3}$	St_{KH}	λ_{KH}
DNS	4.683	2.091	0.137	14.847	0.261	0.359	0.398	0.261	4.384	0.000164
standard URANS	4.198	2.017	0.126	15.082	0.261	0.442	0.386	0.440	–	–

TABLE 1. Time-averaged and unsteady characteristics of the flow obtained with DNS and standard URANS simulations. The first two columns report the maximal lift variation and time-averaged drag coefficients, respectively. The remaining columns report the dominant eigenvalue λ for various frequencies St from the SPOD analysis, considering the fundamental harmnonic for the von-Kármán vortex shedding (VS), its second ($VS, 2$) and third ($VS, 3$) harmonics, and the dominant frequency for the broadband emission of Kelvin-Helmholtz vortices (KH).

3. Nudging-based data-assimilation

In §2.4, the limitations in estimating the considered reference flow based on URANS modelling alone have been established. In the present section, we will introduce a data assimilation technique with the objective of enabling URANS in accurately estimating and reconstructing the reference flow from limited observations. The proposed nudging approach, which in its simplest form here amounts to incorporating pointwise measurements extracted from the reference DNS into the URANS simulations, is introduced in §3.1 and is then applied in §3.2-§3.4 considering various scenarios in terms of temporal and spatial resolutions of measurement data. First, results for a baseline configuration using dense measurement data for nudging are presented in §3.2. As a further step, the influence of reduced resolution of measurement data is assessed in §3.3 and §3.4 with respect to low- and high-frequency phenomena, respectively. Variations in the nudging approach are investigated in §3.5, while the potential and relevance of nudging from modelling and experimental perspectives is discussed in §3.6.

3.1. *Data assimilation methodology*

3.1.1. *Nudged equations and measurements*

Nudging allows a dynamic adjustment of the estimated flow through the addition of a feedback term in the URANS momentum equations in (2.3), which is directly proportional to the discrepancies between flow measurements denoted $\boldsymbol{m}$ and the estimated velocity field at the measurement locations. The intensity of this feedback term is adjusted through a parameter called α. The estimated velocity field, which is now denoted $\bar{\boldsymbol{u}}_\alpha$, satisfies the so-called nudged equations

$$\frac{\partial \bar{\boldsymbol{u}}_\alpha}{\partial t} + (\bar{\boldsymbol{u}}_\alpha \cdot \boldsymbol{\nabla})\bar{\boldsymbol{u}}_\alpha + \boldsymbol{\nabla}\bar{p}_\alpha - \boldsymbol{\nabla}\cdot\left[2\left(Re^{-1} + \nu_t(\tilde{\nu}_\alpha)\right)\boldsymbol{\nabla}_{\mathrm{s}}\bar{\boldsymbol{u}}_\alpha\right] = \alpha\,\mathcal{H}^\dagger[\boldsymbol{m} - \mathcal{H}(\bar{\boldsymbol{u}}_\alpha)], \quad (3.1)$$

which are supplemented by the divergence-free condition in (2.3) and the Spalart-Allmaras model (2.4) to determine the eddy-viscosity field $\nu_t(\tilde{\nu}_\alpha)$ in the above equation. The feedback term on the right-hand-side involves the measurement operator $\mathcal{H}$, which acts on the estimated velocity field $\bar{\boldsymbol{u}}_\alpha$ and returns a vector $\mathcal{H}(\bar{\boldsymbol{u}}_\alpha)$ in the measurement space. Further details about the measurement operator and its adjoint $\mathcal{H}^\dagger$ are provided below. Setting $\alpha = 0$, one recovers the standard URANS equations, while for $\alpha > 0$, the feedback term in the momentum equations drives the estimated flow $\bar{\boldsymbol{u}}_\alpha$ towards measurements $\boldsymbol{m}$. The rationale behind the choice of an appropriate value for α is detailed in appendix A, where minor sensitivity of results is shown as soon as $\alpha > 1$. The value $\alpha = 100$ is thus chosen in the following nudged URANS simulations. Incidentally, this value appears consistent with the considerations of Di Leoni *et al.* (2020), who

established that α should scale as the inverse of the time step of the simulations, which corresponds to $1/\Delta t = 50$ for the smallest value of Δt investigated here (see §3.1.2).

In the present study, the measurement operator $\mathcal{H}$ is defined so as to extract the two components of a velocity field at selected spatial locations, thus mimicking the particle image velocimetry (PIV) approach but without taking into account any measurement error that may occur in real experiments. These M 'nudging points', which are denoted by $\boldsymbol{x}_k$ with $k = 1 \cdots M$, are sketched with red dots in figure 1(a). They are chosen to be equidistantly distributed (in both directions) by a spatial distance Δs inside a nudging region Ω_I which is delimited by the red frame in this figure. The nudging points do not have to coincide with grid nodes. Note that the complementary region is denoted as Ω_E. In the following, the nudging region is fixed to $\Omega_I = [-1.5, 3.5] \times [-1.5, 1.5]$ and we will investigate the effect of the spatial sampling Δs.

The measurements $\boldsymbol{m}$ are here extracted from the reference velocity field $\bar{\boldsymbol{u}}_r$ from DNS results (see §2.1). More precisely, $\boldsymbol{m}$ is defined as

$$\boldsymbol{m}(t) = \mathcal{H}(\bar{\boldsymbol{u}}_r(t)) = [\bar{\boldsymbol{u}}_r(\boldsymbol{x}_1, t), \cdots, \bar{\boldsymbol{u}}_r(\boldsymbol{x}_M, t)]^{\mathrm{T}} = [\boldsymbol{m}_1(t), \cdots, \boldsymbol{m}_M(t)]^{\mathrm{T}}, \tag{3.2}$$

where $\boldsymbol{m}_k$ is a vector that contains the two velocity components at measurement point $\boldsymbol{x}_k$. The size of the data vector $\boldsymbol{m}$ is therefore $2M$. At an instant t, the right-hand side term in (3.1) is obtained from the discrepancy between the estimated flow and this data vector, i.e. $[\boldsymbol{m} - \mathcal{H}(\bar{\boldsymbol{u}}_\alpha)] = [\boldsymbol{m}_1 - \bar{\boldsymbol{u}}_\alpha(\boldsymbol{x}_1), \cdots, \boldsymbol{m}_M - \bar{\boldsymbol{u}}_\alpha(\boldsymbol{x}_M)]^{\mathrm{T}}$, and the adjoint measurement operator $\mathcal{H}^\dagger$ then allows to transform this discrepancy vector of size $2M$ into a field in the model space of pointwise forcings at the same locations $\boldsymbol{x}_k$ according to

$$\mathcal{H}^\dagger[\boldsymbol{m} - \mathcal{H}(\bar{\boldsymbol{u}}_\alpha)] = \sum_{k=1}^{M} \delta(\boldsymbol{x} - \boldsymbol{x}_k) \left(\boldsymbol{m}_k - \bar{\boldsymbol{u}}_\alpha(\boldsymbol{x}_k)\right). \tag{3.3}$$

The implementation of the observation operator $\mathcal{H}$ and of its adjoint in the present finite-element solver is detailed in appendix B.

The temporal and spatial methods that are used to discretise the nudging equation (3.1) are similar to those described in §2.2 for the standard URANS simulations. Note that the feedback term is discretised as an implicit term so as to improve the numerical stability of the algorithm, thus allowing to choose the time step Δt arbitrarily large. For all simulations considered hereinafter, it is assumed that the measurement vector $\boldsymbol{m}(t_i)$ is available at every instant $t_i = i\Delta t$ of the simulation. In other words, we do not make any difference between the time step Δt of the simulation and the time interval between measurements. In the following, Δt is therefore used to characterise the temporal sampling of a data set that can be written as

$$[\boldsymbol{m}(t_1), \cdots, \boldsymbol{m}(T_m)], \tag{3.4}$$

where N_m is the number of measurement vectors and $T_m = t_{N_m} = N_m \Delta t$ is the measurement time (or equivalently here the time window of the simulations). In the next paragraph, it will be detailed how the spatial sampling Δs, the temporal sampling Δt and the measurement time T_m are chosen to define several data sets that are used to assess the nudging method in its ability to improve the fidelity of the estimated flow at different spatial and temporal flow scales.

As final comments on the feedback term in the nudged URANS equations (3.1) in this section, it should first be emphasised that its structure, which involves the discrepancies

Data-set group	St	τ	λ_x	Δt	N_m	$T_m = N_m \Delta t$	$\tau/\Delta t$	T_m/τ
VS (§3.3)	0.137	7.335	4.00	0.209	908	189.85	35	26
KH (§3.4)	4.384	0.228	0.25	0.021	1441	30.13	11	132
VS+KH (§3.2)	0.137	7.335	4.00	0.021	8421	176.1	349	24
	4.384	0.228	0.25				11	772

TABLE 2. Description of three groups of measurement data-sets which are named from the targeted flow phenomenon by the nudging approach: the von-Kármán Vortex Street (VS), the Kelvin-Helmholtz vortices (KH) and both phenomena (VS+KH). Their frequency St and period $\tau = 1/St$ as captured by DNS are recalled in the second and third columns, respectively. These data-set groups are characterized by the time interval between consecutive measurements Δt (fifth column), the number of measurement snapshots N_m (sixth column) and the total measurement time T_m (seventh column). The last two columns report the number of snapshots per cycle (of the target flow phenomenon) and the total number of cycles.

between measurements and the estimated flow, is typical of that of sequential data assimilation and state estimation techniques. For example, the much more expensive Kalman filter is recovered by replacing the contribution $\alpha \mathcal{H}^\dagger$ with the Kalman gain $\boldsymbol{K}$ after discretisation (Lewis *et al.* 2008). Secondly, the role of this feedback term includes phasing and synchronisation with the reference measurements, independently of the presence or not of turbulence model, as would be the case with other sequential approaches. In the present case, it may in addition be interpreted as compensating for deficiencies in the present RANS model. Accordingly, one could wonder if this feedback term could be instead written in a divergence form, which would be more in accordance with the original structure of the URANS equations. This could be achieved in particular if measurements of the Reynolds stress tensor were available. For the present study, however, we assume to only have access to observations of the (resolved) velocity field, which may often be the case in practical applications. Furthermore, we here do not intend to correct the RANS model itself, but rather compensate adverse effects on physical flow structures (e.g. excessive dissipation), focusing on state estimation of the resolved scales only. The present shape of the feedback term is thus fully justified by data assimilation concepts and, again, would be similar in many sequential schemes. While the results in §3.3-§3.4 will confirm the relevance and efficacy of the present approach in the vast majority of the following data assimilation experiments, the punctual aspect of the feedback term in (3.1) and (3.3) can however be a matter of debate, in particular in the case of very sparse measurements. In §3.5, interpolant-based nudging will be investigated as an alternative to the present punctual approach to further enhance reconstruction results in some of the sparsest test cases.

3.1.2. *Measurement data-sets*

Three groups of data-sets are defined in table 2 to specifically investigate the ability of nudging to accurately estimate low-frequency vortex shedding [group VS], high-frequency Kelvin-Helmholtz vortices [group KH], and both phenomena [group VS+KH]. The frequency St, period $\tau = 1/St$, and typical wavelength λ_x of the structures that are associated to each flow phenomenon are first recalled in the first four columns of this table. The time interval Δt between consecutive measurements and the total time T_m of

Δs	0.03125	0.0625	0.125	0.25	0.5	0.75	1	1.25	1.5	1.66	2
Data-set group	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
VS (§3.3)	128	64	32	16	8	5.33	4	3.2	2.66	2.41	2
KH (§3.4)	8	4	2	1	×	×	×	×	×	×	×
VS+KH (§3.2)	(128; 8)	×	×	×	×	×	×	×	×	×	×

TABLE 3. Test matrix indicating, for the three data-set groups VS, KH and VS+KH in table 2, the number of measurement points $N_s = \lambda_x/\Delta s$ that sample the wavelength λ_x of the corresponding spatial structures (indicated in the fourth colmun of table 2) when varying the spatial sampling Δs. For instance, the data-set VS(a) corresponds to ($\lambda_x = 4.00$, $\Delta s = 0.03125$), thus $N_s = 128$.

the simulations are then chosen to properly capture one of the two phenomena (VS or KH) or both of them (VS+KH). The ratios $\tau/\Delta t$ and T_m/τ, which are indicated in the last two columns, correspond to the number of measurement snapshots per cycle and to the total number of cycles, respectively. They confirm that the temporal resolution and time window of the data assimilation experiments are sufficiently large to capture the physics of interest.

It may be noted that data-sets of group VS+KH combine the temporal resolution of group KH ($\Delta t = 0.021$) and the time window of group VS ($T_m \simeq 180$), which makes them sufficiently well resolved for describing the high-frequency shedding of Kelvin-Helmholtz vortices (11 snapshots per cycle), while also capturing a significant number (24) of the low-frequency von-Kármán vortex shedding cycles. Data-sets in this group (VS+KH) are therefore characterised by a large number of measurement snapshots ($N_m \sim 10^4$), which makes them less convenient for parametric studies. They will mainly be used in §3.2 to demonstrate the ability of the nudging method to reconstruct both phenomena based on measurements that are dense in both time and space.

The influence of the spatial resolution Δs on the performances of the data assimilation procedure is mainly investigated using the data-set groups VS and KH in §3.3 and §3.4, respectively. Table 3 summarises the values of spatial sampling Δs that are tested for each data-set group. More precisely, it indicates the number of measurement points $N_s = \lambda_x/\Delta_s$ that sample the characteristic wavelength λ_x of the corresponding flow structure. The data-sets of group VS cover a wide range of spatial resolutions, from the highest resolution with 128 points per wavelength when $\Delta s = 0.03125$, to the lowest spatial resolution with 4 points per wavelength for $\Delta s = 1$. The spatial resolution covered by the data-sets of group KH is more restricted, from 8 points to only 1 point per wavelength of Kelvin-Helmholtz vortices.

3.2. *Results with the data-set of highest spatio-temporal resolution (data-set VS+KH(a))*

We first investigate results of the nudging method using the data set VS+KH(a), which corresponds to the highest temporal and spatial resolutions investigated here, as detailed in tables 2 and 3. Results from the nudged URANS simulation are first examined and compared with standard URANS through the instantaneous vorticity field at $t = 50$, similarly as in figure 2, which reported DNS and standard URANS results. Small-scale structures missing in the URANS prediction without nudging (figure 8(a)) are clearly recovered by the nudging procedure (figure 8(c)). The estimated flow obtained with nudging looks very similar to the reference solution (figure 2(a)), especially in the nudging

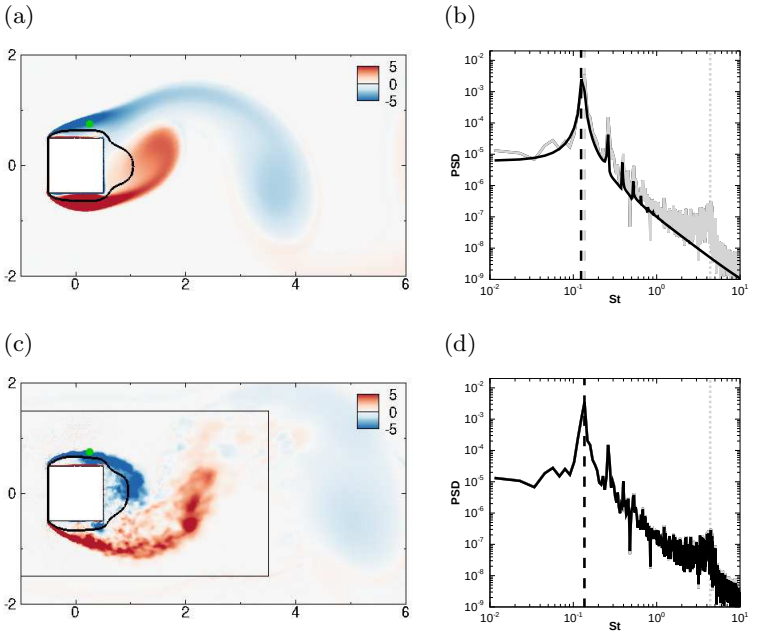

FIGURE 8. Results for (a-b) standard and (c-d) nudged URANS with the data-set VS+KH(a). (a,c): Instantaneous spanwise vorticity field $\bar{\omega}_z$ at $t = 50$, where black iso-curves denote $\langle\overline{u}\rangle = 0$. The extent of the nudging region is delineated by black lines in (c). (b,d): Fourier spectrum of the streamwise velocity at the green monitor point in (a,c) (full black line). For the sake of comparison, the DNS spectrum is also reported (full grey line, well-hidden in (d)). Vertical dashed and dotted lines indicate the low-frequency peak associated with large-scale vortex-shedding and the high-frequency bump linked to Kelvin-Helmholtz instabilities, respectively.

region. More downstream ($x > 3.5$), we may already note that these small scale structures are missing, as further discussed in the following.

Figures 8(b) and (d) display the Fourier spectrum of the streamwise velocity at $(x, y) = (0.25, 0.75)$ for the estimated flow without and with nudging, respectively. As mentioned in §2.4, the high-frequency content ($1 < St < 10$) corresponding to Kelvin-Helmholtz fluctuations in the reference flow (grey curve in figure 8(b)) is not captured by the estimated flow without nudging (black curve in figure 8(b)). On the other hand, the Fourier spectrum corresponding to the URANS prediction with nudging (black curve in figure 8(d)) is almost perfectly superimposed to the reference spectrum, for all frequencies. In particular, the low-frequency peak corresponding to vortex-shedding (dashed vertical line) is reached at $St_{VS} = 0.137$ with nudging, instead of $St_{VS} = 0.126$ without nudging, which is in prefect agreement with the reference value.

This improvement in the prediction of the fundamental frequency St_{VS} is associated with a good synchronisation of the estimated flow with the reference one. This effect, which could have already been identified by comparing figure 8(c) with 2(a), is clearly visible in figure 9, which displays three instantaneous snapshots of the reference and nudged flows in figures 9(a-c) and 9(d-f), respectively. The large scale vortices are now perfectly synchronised, in contrast with standard URANS and figure 3. The small-scale structures seem also satisfactorily captured in the nudging region, but not further downstream. This defect, which can be also observed in the spatial error distribution in figures 9(g-i), may be due to imperfections of the turbulence model,

leading to excessive dissipation of the von-Kármán vortices. Nevertheless, the induced discrepancies remain significantly smaller compared to standard URANS in figure 3(g-i). The large improvement in the estimated flow through the nudging approach is further confirmed by considering the temporal evolution of the global error $E(t)$ (defined in eq. 2.6), as illustrated in figure 9(j). This error is strongly reduced for the nudged simulation (black curve) compared with the standard one (grey curve). In the former case, $E(t)$ decreases during an initial transient and eventually maintains a fairly constant low level. This is in contrast with the standard URANS results, for which $E(t)$ significantly increases and then oscillates due to insufficient synchronisation with respect to the reference solution, as discussed in §2.4.

In order to get a better insight in the extrapolation capabilities of nudging, the evaluation of $E(t)$ is decomposed according to

$$E(t) = E_I(t) + E_E(t) = \int_{\Omega_I} e(\boldsymbol{x},t)^2 \, d\boldsymbol{x} + \int_{\Omega_E} e(\boldsymbol{x},t)^2 \, d\boldsymbol{x}. \tag{3.5}$$

This decomposition allows to distinguish between an internal error $E_I(t)$ defined over the nudging domain Ω_I (see figure 1), which is directly influenced by the measurements, and an external error $E_E(t)$ defined over the complementary domain $\Omega_E = \Omega \setminus \Omega_I$. Both error components E_E and E_I are shown as functions of time in figures 10 (a) and (b), respectively.

Initially, the internal and external errors in the estimated flow are of the same order of magnitude. The internal error E_I for nudged URANS (black curve) achieves a large reduction by three orders of magnitude in about 5 convective times, which indicates an excellent agreement with the reference flow in Ω_I. More time is necessary to decrease the external error E_E as the corrected flow in the nudging region has to be convected in the complementary region Ω_E. The external error remains, however, significantly larger than the internal one and thus dominates the total error E. The reduction in the external error compared to standard URANS results (grey curve) is still noticeable (by one order of magnitude), which confirms the significant benefits of nudging even outside of the nudging region Ω_I.

So far, the potential of nudging the URANS equations for unsteady data assimilation using a measurement data-set with high spatio-temporal resolution has been demonstrated. In the following, we will study the ability of the present data assimilation procedure in flow reconstruction based on measurements with low spatial resolution, focusing on either the estimation of large-scale vortex shedding (§3.3) or that of Kelvin-Helmholtz instabilities (§3.4).

3.3. *Nudging the low-frequency large-scale structures (data-set group VS)*

We now specifically analyse the performance of the nudging approach in the accurate estimation of low-frequency phenomena ($St < 1$) using data-sets of lower resolution. More specifically, we now consider data-set group VS in table 2, which is characterised by the temporal sampling $\Delta t = 0.209$, this value being 10 times larger than for the data-set group VS+KH used in §3.2. Moreover, for most of the data-sets in group VS, the spatial sampling Δs is also decreased compared to previous results. As summarised in table 3, the investigated values of Δs amount to $2 \leqslant N_s \leqslant 128$, where $N_s = \lambda_x/\Delta s$ is the number of nudging points per characteristic wavelength $\lambda_x \sim 4$ of the large-scale vortices.

Figure 11 shows the vorticity fields at $t = 50$ and the Fourier spectra of the streamwise velocity at point $(x, y) = (0.25, 0.75)$, similarly as in figures 2 and 8, which are obtained

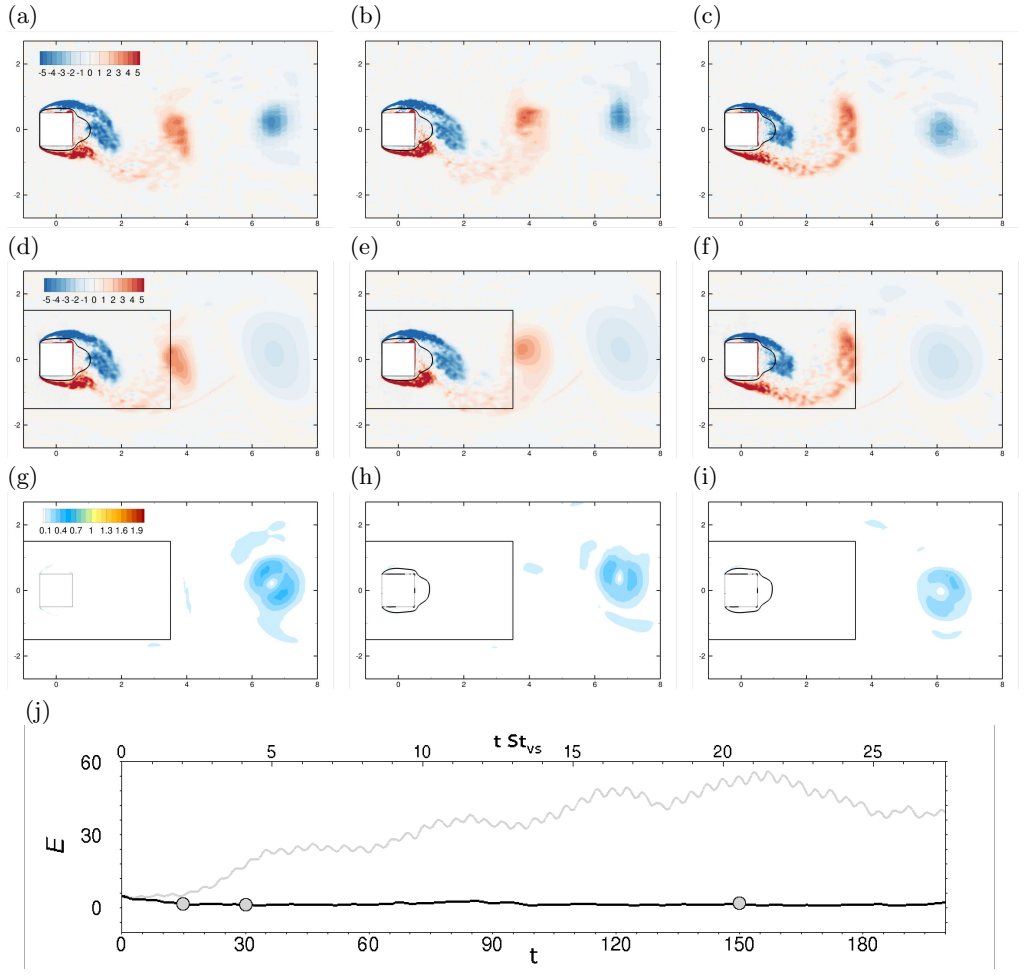

FIGURE 9. Instantaneous vorticity fields of (a-c) the reference flow $\bar{\boldsymbol{u}}_r$ and (d-f) the estimated flow $\bar{\boldsymbol{u}}$ with nudging relying on the data-set VS+KH(a) at times (a,d) $t = 15$, (b,e) $t = 30$ and (c,f) $t = 150$. (g-i) Instantaneous error fields $e(\boldsymbol{x}, t)$ between the reference and nudged flows (see eq. 2.5). (i) Temporal evolution of the global error $E(t)$ (see eq. 2.6) between the nudged and reference flow (black curve) and between the URANS and reference flows (grey curve). The bottom axis reports the (nondimensional) time while the top axis reports the number of low-frequency cycles determined as $t/\tau_{VS} = t\, St_{VS}$ with $St_{VS} = 0.137$. The grey circles indicates time $t = 15$, 30 and 150

for nudged URANS simulations with decreasing values of the spatial sampling Δs. In the Fourier spectra, results corresponding to nudged URANS are displayed with black curves, while the grey curves correspond to results of the reference flow when sampled at $\Delta t = 0.209$, i.e. the temporal sampling of the present data-set. This spectrum is therefore different from the reference spectrum in figures 2 and 8 that corresponds to a temporal sampling ten times smaller, ($\Delta t = 0.021$). Their comparison clearly shows the filtering effect when down-sampling the reference flow.

Let us first examine results obtained with the largest spatial samplings displayed in figure 11(a-b) for $\Delta s = 1.66$ and in figure 11(c-d) for $\Delta s = 1.5$. For $\Delta s = 1.66$, the nudged simulation is not synchronised with the measurement data. The largest amplitude peak

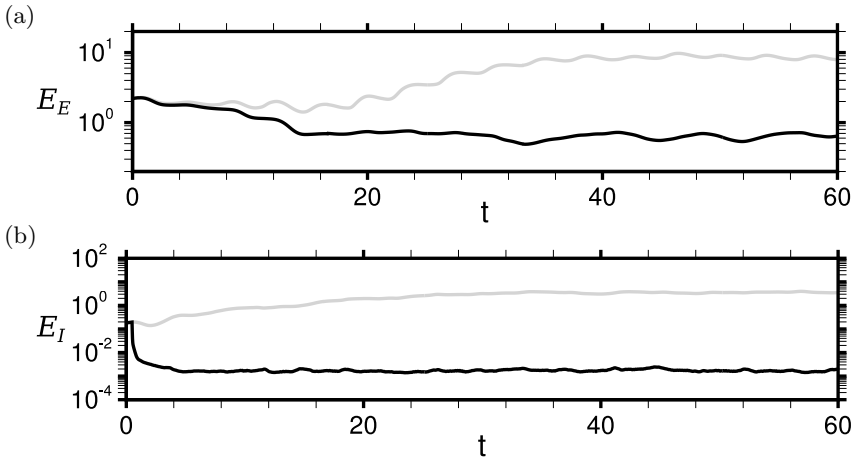

FIGURE 10. Temporal evolution of (a) external error E_E and (b) internal error E_I (see (3.5)) for standard URANS (grey lines) and nudged URANS with the data-set VS+KH(a) (black lines).

is obtained for a frequency that is slightly lower than in the reference simulation (see figure 11(b)) and is the same as in the standard URANS results ($St_{VS} = 0.126$ versus $St_{VS} = 0.137$ in the reference simulation). By increasing the spatial sampling of the observations to $\Delta s = 1.5$, the vortex-shedding frequency of the nudged simulation already locks to the reference frequency (figure 11(d)). Furthermore, this appears to be true not only for the fundamental frequency St_{VS}, but also for its second and third harmonics. For the present set-up, the smallest number of measurement points that is necessary to lock the low-frequencies associated with large-scale vortex shedding is thus around only $N_s \sim 2.5$ points per characteristic wavelength. However, despite this frequency locking, it may be noticed that the estimated flow is still not perfectly in phase with the reference one for $\Delta s = 1.5$, as further discussed in the following. Moreover, the largest amplitude peak is under-estimated, as well as the power spectral densities at higher frequencies. Furthermore, as for $\Delta s = 1.66$, the vorticity field in figure 11(c) (as in figure 11(a)) exhibits undesired small-scale wake-like structures around some nudging points. This appears as one limitation of nudging based on pointwise forcings at the measurement locations in the case of very sparse data. This issue will be revisited and addressed in §3.5 by modifying the structure of the feedback term in the nudged URANS equations (3.1).

A better flow estimation is obtained by further decreasing the spatial sampling to $\Delta s = 0.5$ (figure 11(e-f)). The large-scale vortices are now well synchronised with the reference flow (figure 2(a)) and the dominant peak in the spectrum is almost perfectly recovered. Moreover, the spurious small-scale structures around nudging points that were observed for larger spatial samplings are not visible anymore. Decreasing the spacing between nudging points to $\Delta s = 0.125$ adds more small-scale features to the nudged URANS simulation, as illustrated by figure 11(g), which gets even closer to the reference snapshot of figure 2(a). In addition, the corresponding spectrum in figure 11(h) agrees perfectly well with the down-sampled reference one. It may be emphasised that $\Delta s = 0.125$ still corresponds to a much lower resolution compared to that of the mesh that is used in the URANS calculations.

To better assess the frequency locking and phasing effects when varying the number of nudging points, we display in figures 12(a) and (b) the frequency of the largest amplitude peak St_{VS} and the phase shift ϕ, respectively, as a function of Δs (and N_s). As noticed

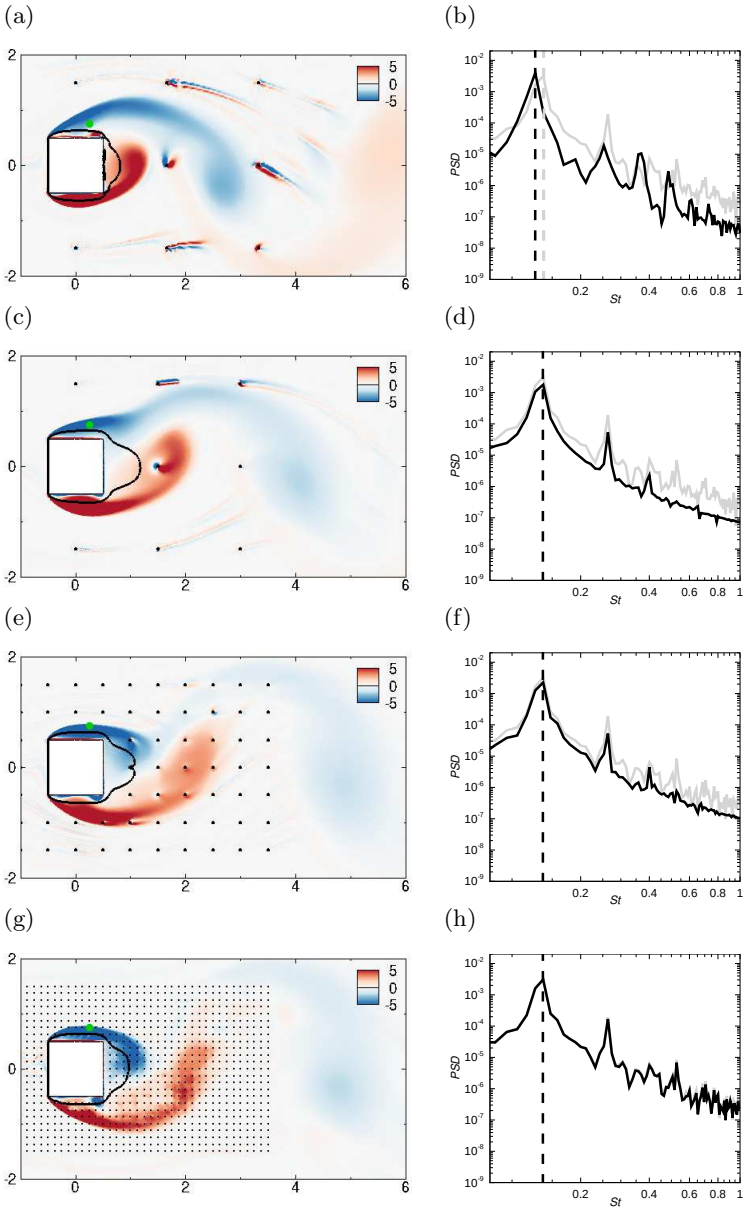

FIGURE 11. Results for nudged URANS simulations in data-set group VS with spatial spacings (a-b) $\Delta s = 1.66$, (c-d) $\Delta s = 1.5$, (e-f) $\Delta s = 0.5$, and (g-h) $\Delta s = 0.125$. (a,c,e,g): Instantaneous spanwise vorticity field $\bar{\omega}_z$ at $t = 50$, where black iso-curves denote $\langle \overline{u} \rangle = 0$, while dots refer to nudging points. (b,d,f,h): Fourier spectrum of the streamwise velocity at the green monitor point in (a,c,e,g) (full black line), the reference spectrum to recover is also reported (full grey curve). Vertical dashed lines indicate the low-frequency peak associated with large-scale vortex-shedding.

above, the fundamental frequency of the nudged simulations is equal to that in standard URANS results ($St_{VS} = 0.126$) for low spatial resolution ($N_s < 2.5$) and locks to the frequency of the reference flow ($St_{VS} = 0.137$) for higher spatial spatial resolution ($N_s > 2.5$). This frequency locking does not guarantee a perfect flow synchronisation. Indeed,

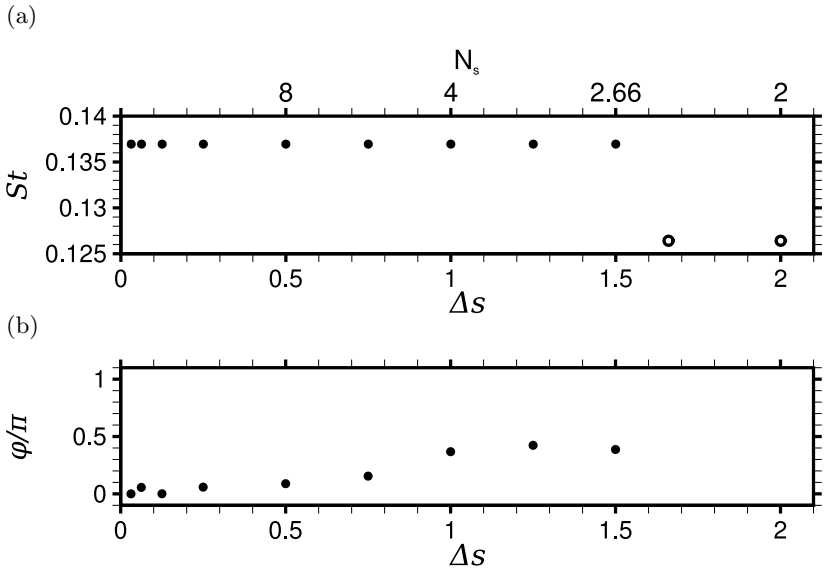

FIGURE 12. (a) Dominant Strouhal number and (b) phase angle with respect to reference results as a function of the spacing between nudging points Δs, or equivalently the number N_s of nudging points per characteristic wavelength, for nudged URANS simulations of data-set group VS. Results are obtained from the Fourier analysis of the signal at the same monitor point as in figure 11. Filled symbols correspond to the case where the estimated shedding frequency is $St_{VS} = 0.137$, as in DNS results, while open symbols correspond to $St_{VS} = 0.126$, as in standard URANS.

for $2.5 \leqslant Ns \leqslant 5.75$, we observe a significant phase shift between temporal evolutions of the streamwise velocity in nudged and reference solutions at the monitor point in figure 11, which vanishes only for higher spatial resolution ($N_s \geqslant 8$).

The effect of the spatial resolution of measurement points on the time-averaged global error $\langle E \rangle$ in (2.7) is now examined through figure 13(a). In this figure, $\langle E \rangle$ is normalised by its value for standard URANS results. For the smallest number of nudging points $N_s = 2\,(\Delta s = 2)$, the normalised error is slightly smaller than one, indicating a very weak improvement compared to the standard simulation. When increasing to $N_s = 2.66$, i.e the smallest number of nudging points leading to a frequency locking, the normalised error decreases to 0.3, which corresponds to an improvement of 70% compared to the standard URANS results. Interestingly, for $2.66 \leqslant N_s \leqslant 5.75$, the evolution of the error follows that of the phase shift shown in figure 12(b). In this range of spatial resolution, the decrease of the error is thus strongly related to the decrease of the phase shift noticed before. Finally, for $N_s \geqslant 8$, the time-averaged error reaches a constant value that is close to 0.05 times the standard URANS error.

To assess the extrapolation capability of the nudged simulations, $\langle E \rangle$ is decomposed into internal and external components, $\langle E_I \rangle$ and $\langle E_E \rangle$, corresponding to the error inside and outside the nudging region, respectively, similarly as in §3.2. The time-averaged internal and external errors, normalised by the total error of the standard simulation, are displayed as a function of N_s in figure 13(b) using black and grey dots, respectively. We clearly note that the external error is the dominant component, whatever the number of nudging points, as for the data-set VS+KH(a) discussed in §3.2. For a small number of nudging points ($N_s < 5.75$), these two errors are of similar order of magnitude, and they decrease in a similar way when increasing the number of nudging points. This is related

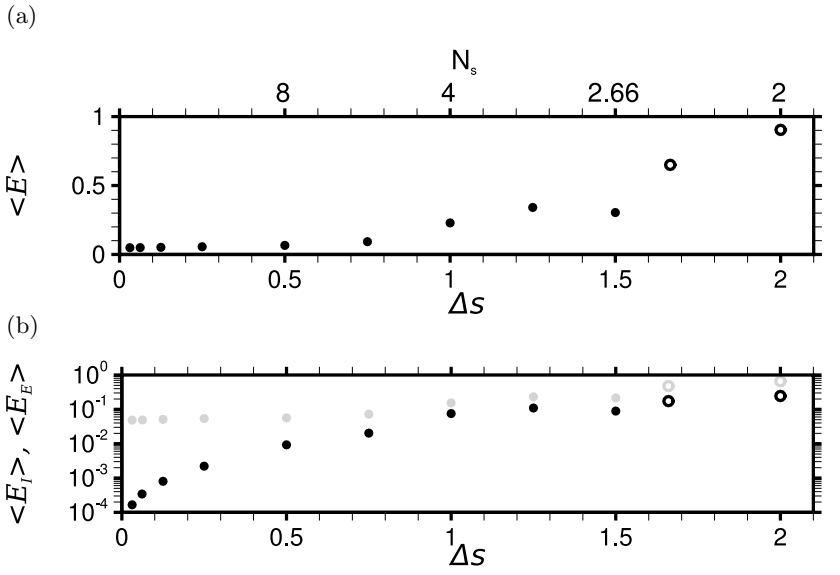

FIGURE 13. Time-averaged errors as a function of the spacing between nudging points Δs, or equivalently the number N_s of nudging points per characteristic wavelength, for nudged URANS simulations of data-set group VS. (a) Total error $\langle E \rangle$ in (2.6), which may be decomposed as the sum of (b) internal $\langle E_I \rangle$ (black symbols) and external errors $\langle E_E \rangle$ (grey symbols). The reported values are normalised by the total time-averaged error for the standard URANS solution. Filled symbols correspond to the case where the estimated shedding frequency is $St_{VS} = 0.137$, as in DNS results, while open symbols correspond to $St_{VS} = 0.126$, as in standard URANS.

to the improved flow synchronisation (frequency locking and phase-shift decrease) which has been previously noticed for that range of spatial resolution and occurs everywhere in the flow. Once the number of nudging points is large enough to perfectly synchronise the flow ($N_s > 5.75$), the internal error strongly decreases unlike the external error that reaches a constant value around 5% of the total error obtained by standard URANS simulation. This constant value agrees with the total error in figure 13(a) for $\Delta s \leqslant 0.75$. This irreducible error may be attributed to (turbulence) model errors which limit the extrapolation capability of the nudging approach. Its interpolation capability, i.e. the ability to estimate the flow in between nudging points, is nevertheless very good as the internal error decreases to 10^{-4}.

Having carefully characterised in time the flow estimations for the data-set group VS, we now analyse these results in the frequency space through SPOD, as performed in previous sections. For the present data-set, SPOD is computed based on overlapping bins that cover about 63 time units corresponding to approximately 9 vortex-shedding cycles. The frequency St_{VS} and eigenvalue λ_{VS} (kinetic energy) corresponding to the largest amplitude peak (see figure 5(b)) are shown in figure 14(a) and (b), respectively, as a function of the spacing between measurements Δs. For $\Delta s > 1.5$, the frequency of the nudged mode is equal to that of the URANS mode ($St_{VS} = 0.126$), while for $\Delta s \leqslant 1.5$, it locks to the frequency of the reference mode ($St_{VS} = 0.137$), in agreement with previous observations deduced from a pointwise temporal signal (figure 12). Figure 14(a) thus confirms the global aspect of frequency locking for $\Delta s \leqslant 1.5$. Examining now the estimated kinetic energy λ_{VS} of a dominant SPOD mode at St_{VS} in figure 14(b), it is almost equal to that of the reference mode (dash-dotted line) once the frequency

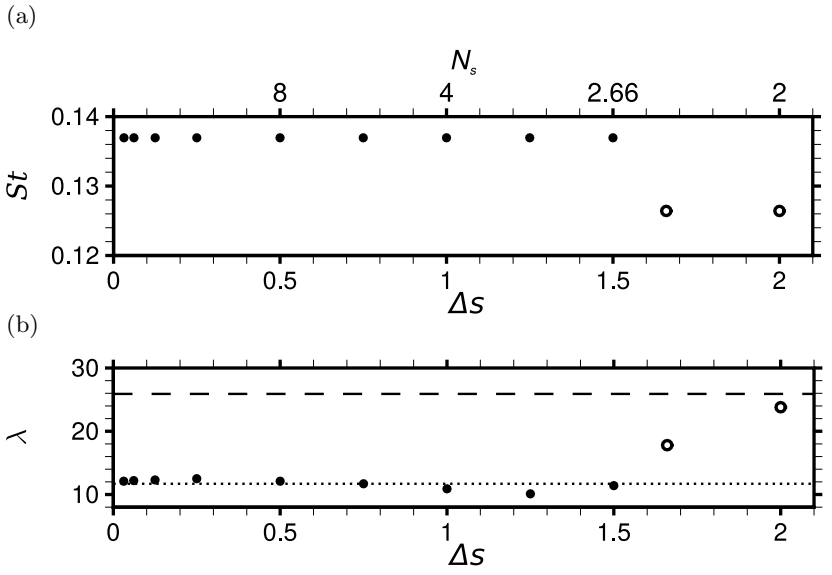

FIGURE 14. (a) Frequency and (b) kinetic energy of the dominant SPOD modes corresponding to vortex-shedding as a function of the spatial sampling Δs for nudged URANS simulations of data-set group VS. The dashed and dash-dotted horizontal lines in (b) correspond to energy of the SPOD mode for the URANS and reference data set, respectively. The spatial structure of few modes are displayed in figure 15. Filled symbols correspond to the case where the estimated shedding frequency is $St_{VS} = 0.137$, as in DNS results, while open symbols correspond to $St_{VS} = 0.126$, as in standard URANS.

is locked to the reference frequency. Using $N_s = 2$ points per wavelength ($\Delta s = 2$ in the figure), the kinetic energy is still close to that of the URANS mode (dashed line). But a slight increase to $N_s = 2.66$ points ($\Delta s = 1.5$) yields an almost perfect agreement with the reference value. Further increasing the number of nudging points (decreasing their spacing Δs) does not significantly improve the estimation of λ_{VS}. The spatial structure of the corresponding mode is shown in figures 15(a,c,e,g) for several spatial spacings Δs. In addition, the corresponding spectral error fields Φ_e in (2.8) are reported in figures 15(b,d,f,h). For the SPOD mode associated to the unlocked frequency (figure 15(a-b)), the largest errors are obtained close to some of the nudging points downstream of the cylinder. They are associated to errors in the cross-stream velocity (not displayed here) since the streamwise velocity is zero on the symmetry axis $y = 0$. These specific defects vanish once the frequency is locked to the reference frequency (figures 15(c-d)). Largest errors are now obtained in between nudging points, in a (red) pocket immediately downstream of the cylinder ($x < 1.5$) and in two symmetric (yellow) pockets around $x = 2$. Both the shape and the intensity of the errors are actually similar to those for standard URANS (figure 5(d)). Reducing the spacing between nudging points to $\Delta s = 0.5$ (figures 15(e-f)), these errors decrease in the near wake. One may directly identify for the mode in figure 15(e) a change in the structures around $2 \leqslant x \leqslant 4$ compared to previous cases, these structures being now closer to the ones in the reference mode (figure 5(a)). However, modal errors are significantly reduced only for spacings $\Delta s \leqslant 0.25$, as well as for the case $\Delta s = 0.125$, which is illustrated in figures 15(g-h). The residual errors in the far wake ($x > 8$) are barely modified when further increasing the number of nudging points. Accurately recovering the detailed shape of the dominant SPOD mode associated to vortex shedding thus seems to require more spatially dense measurements

($N_s \geqslant 16$ nudging points per characteristic wavelength) compared to the estimation of its frequency and kinetic energy/magnitude ($N_s \sim 2.5$). However, it may be emphasised that the estimation of the shape of such a mode, which is already acceptable with standard URANS (see figure 5), may be considered as only secondary to the correct prediction of associated frequency and kinetic energy. These results from the SPOD analysis agree with the low values for the time-averaged global errors $\langle E \rangle$ that are already reached for $N_s \geqslant 2.66$, i.e. after frequency locking and phasing are achieved, and also with a correct estimation of λ_{VS}, as discussed previously.

Finally in this section, the quality of the mean flow estimated by the nudged URANS simulations for data-set group VS is discussed. The streamwise mean velocity field is depicted in figures 16(a,c,e,g) for several values of Δs and can be compared to the reference and URANS mean flows displayed in figures 4(a) and (b), respectively. The corresponding error fields $e_{\langle \bar{\boldsymbol{u}} \rangle}$ in (2.7) are reported in figures 16(b,d,f,h). For the lowest spatial resolution ($\Delta s = 1.66$), significant improvement is already observed in the far wake outside of the nudging region, which is remarkable as the fundamental frequency St_{VS} is not locked to the reference one in this case. Once it is locked for $\Delta s = 1.5$ (figures 16(c-d)), the error increases in the near wake, similarly as the spectral error of the SPOD mode in figure 15(d) for the same spacing. This mean-flow error can be decreased by increasing the number of nudging points. Significant improvements may already be identified for $\Delta s = 0.5$ (figures 16(e-f)). For the largest spatial resolution ($\Delta s = 0.125$) displayed in figure 16(g-h), the mean-flow error drops below 3% in the near and far wake. Only the error in the shear layers cannot be further decreased when increasing the number of nudging points.

A more quantitative description of the mean-flow error is provided in figure 17 which displays the spatially-integrated mean-flow error $\int_{\Omega_{\langle \bar{\boldsymbol{u}} \rangle}} e_{\langle \bar{\boldsymbol{u}} \rangle}(\boldsymbol{x})^2 d\boldsymbol{x}$, where $\Omega_{\langle \bar{\boldsymbol{u}} \rangle}$ is the domain of integration. Figure 17(a) illustrated this global error, divided by the value for standard URANS results, as a function of Δs for two domains having the same streamwise extent ($-1.5 \leqslant x \leqslant 12$) but different cross-stream ones. Squares correspond to the mean-flow error when integrated in a domain of size $-4 \leqslant y \leqslant 4$, which is identical to the y-extension used for the figures. In this case, the global mean-flow error is only moderately decreased (by $\sim 50\%$ or less) even for the smallest value of Δs. However, when restricting the domain $\Omega_{\langle \bar{\boldsymbol{u}} \rangle}$ to the extension of the nudging region in the cross-stream direction ($-1.5 \leqslant y \leqslant 1.5$), the global mean-flow error (circles) exhibits a larger decrease for $\Delta s \leqslant 1$, while it remains fairly constant for $\Delta s > 1$. This evolution is better understood by examining the internal (black circles) and external (grey circles) components of this error, which are displayed in figure 17(b). For $\Delta s > 1$, the decrease of the external error is actually balanced by the increase of the internal error. These two opposite effects were already visible in figure 16(a-d). For $\Delta s \leqslant 1$, both internal and external mean-flow errors decrease almost monotonically. The external contribution reaches a constant value for $\Delta s \leqslant 0.25$ while the internal one still decreases, which is reminiscent of the evolution of the time-averaged spatially-integrated errors in figure 13. To conclude, the nudging approach using the low-frequency data-set also allows for an improvement in the mean-flow prediction, not only inside but also outside of the nudging region. This extrapolation capability is mainly significant for a domain of limited extension in the cross-stream direction, which seems related to the extension of the nudging region and appears to be directly downstream of the latter, according to the present results.

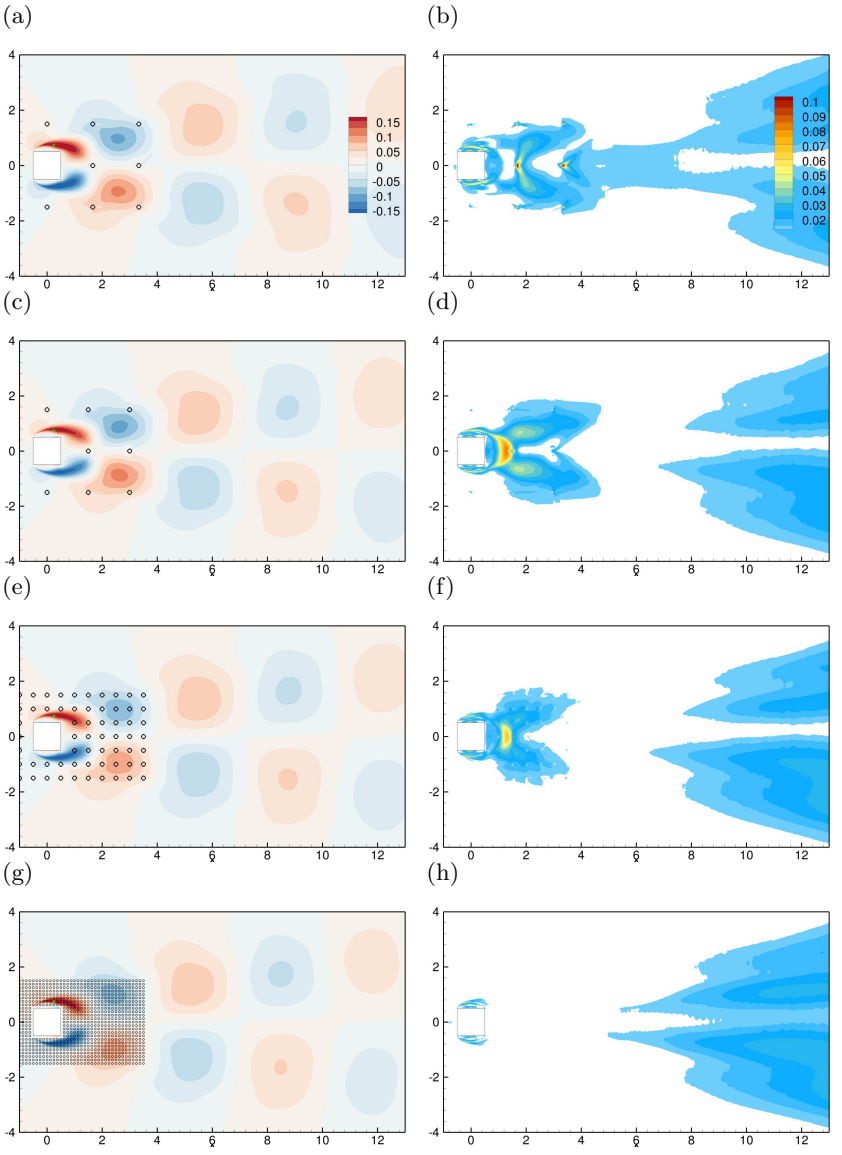

FIGURE 15. SPOD analysis of the vortex-shedding phenomenon for nudged URANS simulations in data-set group VS with spatial spacings (a-b) $\Delta s = 1.66$, (c-d) $\Delta s = 1.5$, (e-f) $\Delta s = 0.5$, and (g-h) $\Delta s = 0.125$. (a,c,e,g): Streamwise velocity contours of the imaginary part of the dominant SPOD mode for the fundamental frequency St_{VS}. While the frequency of the mode in (a) is $St_{VS} = 0.126$, remaining modes oscillate at $St = 0.137$. Modes are phased so that each of them reaches its maximum value at the green monitor point. (b,d,f,h): Corresponding modal discrepancy field Φ_e in (2.8).

3.4. *Nudging the high-frequency small-scale structures (data-set group KH)*

As detailed in §2.4, DNS results capture the shedding of Kelvin-Helmholtz (KH) vortices in the shear layers around the cylinder. As illustrated by figure 6, a broadband peak is observed around a frequency of $St = 4.384$ in the spectrum obtained by the SPOD analysis, and the dominant SPOD mode is characterised by spatially oscillating structures of wavelength $\lambda_x \sim 0.25$ located in the upper shear layer. While such structures are

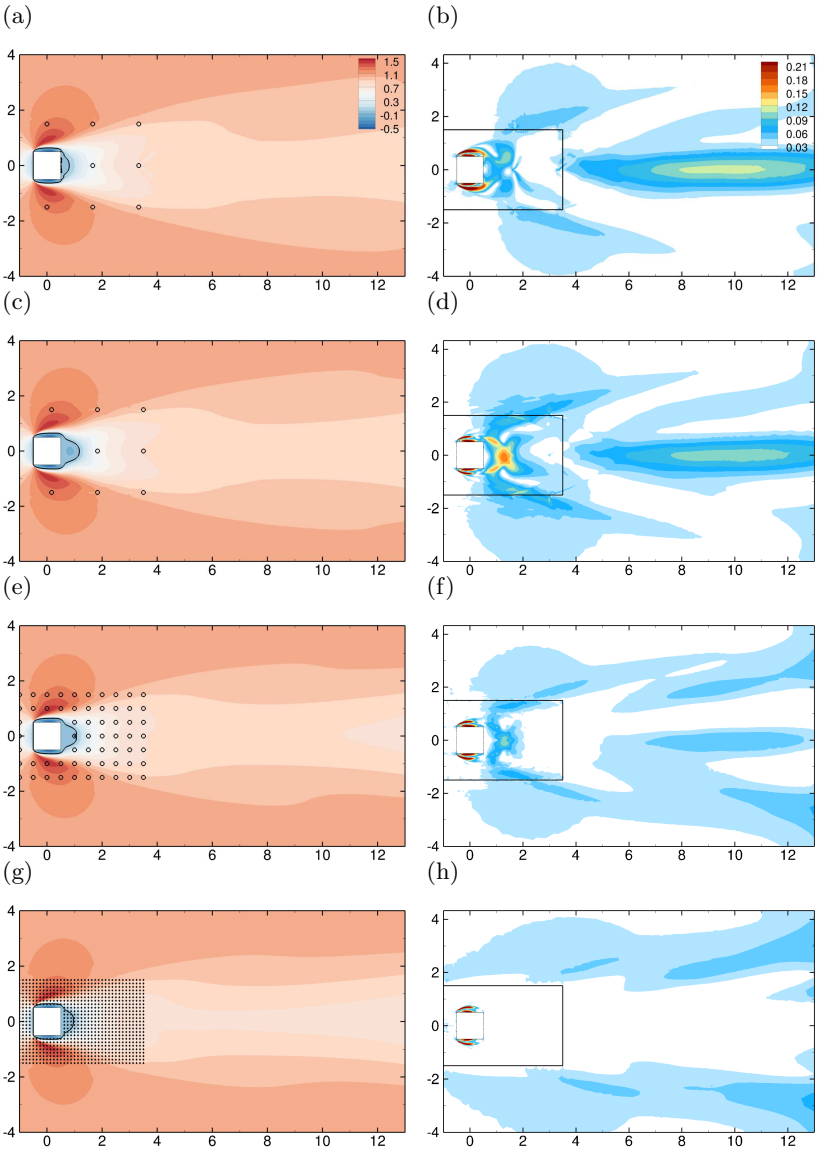

FIGURE 16. Mean-flow results for nudged URANS simulations in data-set group VS with spatial spacing (a,b) $\Delta s = 1.66$, (c,d) $\Delta s = 1.5$, (e,f) $\Delta s = 0.5$ and (g,h) $\Delta s = 0.125$. (a,c,e,g): Streamwise component of the mean flow. Black curves correspond to the isocontours of zero streamwise velocity. (b,d,f,h): Mean-flow error field $e_{\langle \bar{\boldsymbol{u}} \rangle}$ in (2.7). The extent of the nudging region is delineated by black lines.

absent in the present standard URANS simulations, their formation could be triggered by sustained perturbations. In this section, we now investigate the potential of the nudging approach in the accurate estimation of these phenomena by using the data-sets of group KH (see tables 2 and 3). While data-sets in this group generally correspond to higher temporal and spatial samplings compared to previously discussed data-sets (group VS in §3.3), it may be noticed that these resolutions are actually moderate compared to the time and spatial scales that are associated to the KH vortices. The time sampling Δt indeed corresponds to approximately 11 snapshots per KH cycle, while the investigated

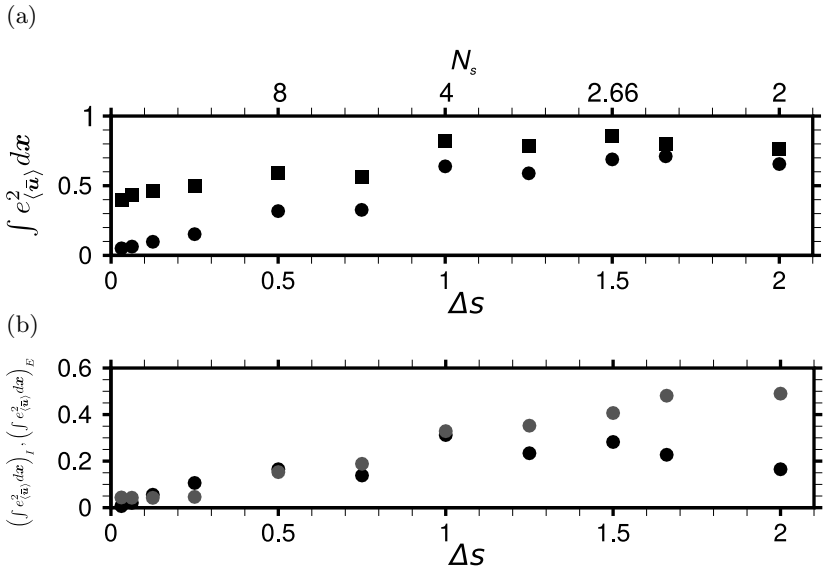

FIGURE 17. (a) Spatially-integrated mean-flow error $\int_{\Omega_{\langle \bar{u} \rangle}} e_{\langle \bar{u} \rangle}(\boldsymbol{x})^2 d\boldsymbol{x}$ of the nudged URANS simulations of data-set group VS as a function of the spatial sampling Δs. This integrated error is evaluated over a spatial domain Ω that extends in the streamwise direction between $-1.5 \leqslant x \leqslant 12.5$ and is restricted in the cross-stream direction to (squares) $-4 \leqslant y \leqslant 4$ or (circles) $-1.5 \leqslant y \leqslant 1.5$. (b) The integrated error in the restricted domain $-1.5 \leqslant y \leqslant 1.5$ is decomposed as the sum of internal (black) and external (grey) errors with respect to the nudging region. Reported values are normalised by the corresponding error of the standard URANS solution.

values of Δs amount to a number of nudging points per wavelength of the KH vortices verifying $1 \leqslant N_s \leqslant 8$. Results are here directly analysed based on a SPOD analysis that is adapted to KH phenomena as performed in §2.4.

The real part of the transverse component of the dominant SPOD mode at $St_{KH} = 4.384$ obtained with nudging is shown in the left-hand-side plots of figure 18 for the different values of the spatial sampling Δs in data-set group KH. SPOD spectra are also displayed in the right-hand-side plots. The SPOD mode in figure 18(a), already captures fluctuations that are reminiscent of Kelvin-Helmholtz structures visible in the reference SPOD mode in figure 6(a). The shape of the estimated mode is already quite close to the reference one for $x \geqslant 0$ (comparisons may also be performed with the mode in figure 18(g) which is very close to the reference one). This is remarkable, as the associated spatial sampling $\Delta s = 0.25$ corresponds to $N_s = 1$, i.e. one nudging point per wavelength of coherent vortex structures. This notable result is likely favoured by the ability of the present URANS model in allowing the emergence of KH structures from sustained excitation as demonstrated in §2.4. The present findings may also be related to the observation in Di Leoni *et al.* (2018) that nudging was able to compensate for modelling limitations in the form of badly-known physical parameters in the case of rotating homogeneous turbulence.

However, the spectrum in figure 18(b) indicates significantly under-estimated amplitudes of all dominant modes around $St_{KH} = 4.384$. In particular, the amplitude λ_{KH} of the mode in figure 18(a) corresponds to only $\sim 4\%$ of the reference value. Similarly as in the results of figure 7, the nudging approach excites modes that are associated with KH instabilities. Their energy amplification, however, is not large enough to reach

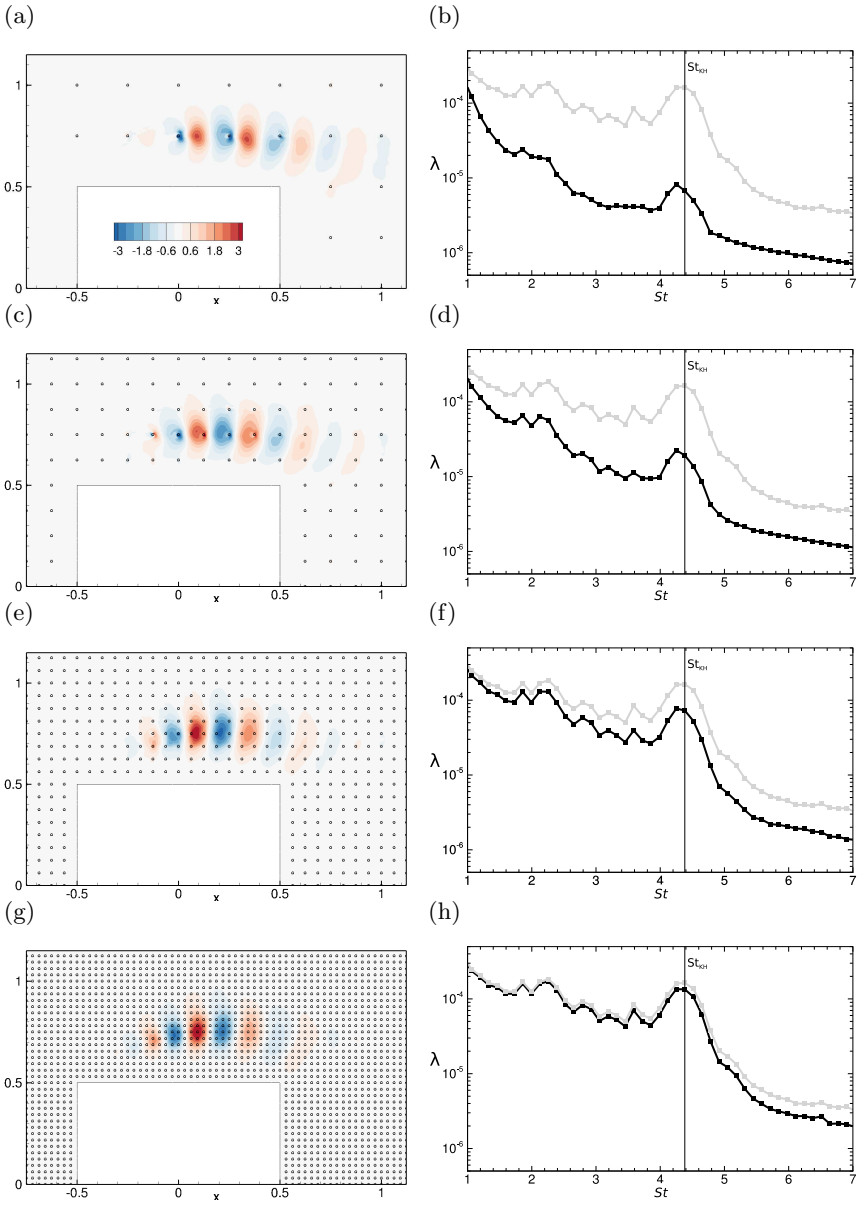

FIGURE 18. SPOD analysis of Kelvin-Helmholtz phenomena as estimated by nudged URANS simulations for the data-set group KH with (a,b) $N_s = 1$ ($\Delta s = 0.25$), (c,d) $N_s = 2$ ($\Delta s = 0.125$), (e,f) $N_s = 4$ ($\Delta s = 0.0625$) and (g,h) $N_s = 8$ ($\Delta s = 0.03125$) nudging points per wavelength of the KH instabilities. (a,c,e,g) Real part of the transverse component of the SPOD mode for $St_{KH} = 4.384$ and (b,d,f,h) SPOD spectrum for nudged URANS (black), the DNS results (grey) are also reported for the sake of comparison.

sufficient amplitudes, probably because of the eddy-viscosity provided by the Spalart-Allmaras model, which dissipates too quickly the energy at that frequency. Increasing the spatial resolution of nudging points leads to increasingly pronounced spectral peaks at $St_{KH} = 4.384$, as more energy is introduced to promote KH structures. This can be confirmed in figures 18(c,d) for $\Delta s = 0.125\,(N_s = 2)$. Eventually, the spatial pattern

obtained for smaller spatial samplings (see figures 18(e) and (g)) becomes even closer to the reference SPOD mode (figure 6(a)). The good performance of nudging is also evident in the SPOD spectra shown in figures 18(f) and (h). $N_s = 8$ nudging points per KH wavelength (figures 18(g-h)) thus appear sufficient to satisfactorily predict the kinetic energy of the modes around $St_{KH} = 4.384$. In particular, the estimated amplitude λ_{KH} corresponds to $\sim 80\%$ of the reference value.

We have shown that applying nudging with sufficiently well-resolved measurement data allows to correctly estimate KH vortices, which are not captured by the standard URANS equations in a self-sustained way. Similarly as for the estimation of low-frequency phenomena (see §3.3), it may be emphasised that nudging improves URANS simulations in a 'global' sense and is here not simply 'over-ruling' local values in shear layers. This is particularly evident for the case with large spacing between nudging points (see figure 18(a)), where local maxima and minima in SPOD modes associated with KH instabilities do not coincide with nudging positions.

3.5. *Interpolant-based nudging*

As previously discussed, the present data assimilation procedure appears efficient in correctly estimating various flow phenomena from limited observations. However, in the case of very sparse data ($\Delta s \geqslant 1.5$), it was noticed that nudging may introduce spurious structures which emerge from the measurement locations, as illustrated by figures 11(a,c). This originates from the localised nature of the feedback term in the nudged URANS equations in (3.1) and (3.3). As an alternative to punctual nudging, we consider in this section the application of interpolant-based nudging as proposed by Azouani *et al.* (2014). In this approach, instead of directly considering the punctual measurements $\boldsymbol{m}(t)$ in (3.2), they are first spatially interpolated to get a more regular representation of the data. The thus obtained field $\tilde{\boldsymbol{m}}(\boldsymbol{x}, t)$ may belong to the same function space as the estimated velocity field $\bar{\boldsymbol{u}}_\alpha(\boldsymbol{x}, t)$. This interpolation process is denoted as

$$\tilde{\boldsymbol{m}}(\boldsymbol{x}, t) = \tilde{\mathcal{I}}(\boldsymbol{m}(t)), \tag{3.6}$$

where $\tilde{\mathcal{I}}$ refers to the interpolation operator. As detailed in appendix B, the interpolation procedure is performed based on a mesh whose nodes coincide with the measurement locations and is thus fully independent from the mesh in figure 1(b), which is used to solve the URANS equations. To enable a consistent comparison with $\tilde{\boldsymbol{m}}(\boldsymbol{x}, t)$ without any approximation, one has to pass the estimated velocity field through the same interpolation process. In other words, $\tilde{\boldsymbol{m}}(\boldsymbol{x}, t)$ should be compared with $\tilde{\boldsymbol{u}}_\alpha(\boldsymbol{x}, t) = \tilde{\mathcal{I}}(\mathcal{H}(\bar{\boldsymbol{u}}_\alpha(\boldsymbol{x}, t)))$, where $\mathcal{H}$ extracts the values of the estimated velocity field at the measurement locations, as detailed in §3.1.1 and appendix B. In this framework, the interpolant-based nudged URANS equations may be written as

$$\frac{\partial \bar{\boldsymbol{u}}_\alpha}{\partial t} + (\bar{\boldsymbol{u}}_\alpha \cdot \boldsymbol{\nabla})\bar{\boldsymbol{u}}_\alpha + \boldsymbol{\nabla}\bar{p}_\alpha - \boldsymbol{\nabla} \cdot \left[2\left(Re^{-1} + \nu_t(\tilde{\nu}_\alpha)\right)\boldsymbol{\nabla}_{\mathrm{s}}\bar{\boldsymbol{u}}_\alpha\right] = \alpha[\tilde{\boldsymbol{m}} - \tilde{\mathcal{I}}(\mathcal{H}(\bar{\boldsymbol{u}}_\alpha))]. \tag{3.7}$$

Contrary to the previous nudged URANS equations (3.1), the feedback term in the right-hand-side of (3.7) is not punctual and may affect the URANS equations in the whole measurement domain Ω_I in figure 1(a).

Interpolant-based nudging (3.7) is applied in figure 19 for a spacing between measurements $\Delta s = 1.5$ and data set VS. For the sake of comparison, the results in figures 11(c-d), which have been obtained with punctual nudging for the same case are reproduced in figures 19(a-b). In accordance with theoretical considerations (Azouani & Titi 2014), the estimated field obtained through interpolant-based nudging (figure 19(c)) appears much more regular than the one with punctual nudging (figure 19(a)). This is particularly

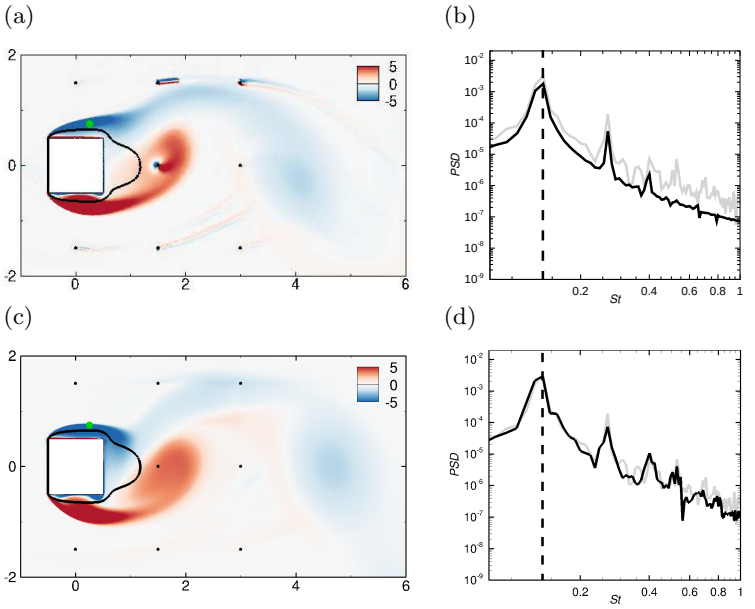

FIGURE 19. Results for nudged URANS simulations in data-set group VS with spatial spacing $\Delta s = 1.5$ where nudging is applied either (a-b) in its standard punctual form or (c-d) through interpolant. (a,c): Instantaneous spanwise vorticity field $\bar{\omega}_z$ at $t = 50$, where black iso-curves denote $\langle \overline{u} \rangle = 0$, while dots refer to nudging points. (b,d): Fourier spectrum of the streamwise velocity at the green monitor point in (a,c) (full black line), the reference spectrum to recover is also reported (full grey curve). Vertical dashed lines indicate the low-frequency peak associated with large-scale vortex-shedding.

striking in the measurement domain where the spurious structures that are observed in the punctual case are absent in the flow that is retrieved through interpolant-based nudging. While for $\Delta s = 1.5$ punctual nudging already enables frequency locking, as discussed in §3.3, it appears from the comparison between figures 19(b) and 19(d) that interpolant-based nudging further improves the estimation of the Fourier spectrum of the velocity fluctuations at the monitor point in figures 19(a,c), which is in this case quite far from all measurement points. In addition, the flow that is estimated by interpolant-based nudging is better phased with the reference one, which results in a decrease of the time-averaged global error $\langle E \rangle$ in (2.6) by a factor of ~ 2.4 compared to punctual nudging. On the other hand, no significant variation in the quality of the estimation of the mean flow or that of the dominant SPOD mode at $St = St_{VS}$ was noticed between punctual and interpolant-based nudging. This may be consistent with the fact that the spurious structures that are introduced by punctual nudging are partly smoothed out through time average and have less impact on statistical quantities than on instantaneous ones.

Finally, interpolant-based nudging is applied to data set KH for $\Delta s = 0.125$. This spacing still corresponds to a low resolution with respect to Kelvin-Helmholtz phenomena ($N_s = 2$ points per characteristic wavelength). The dominant SPOD mode at $St = St_{KH}$ that is estimated by punctual nudging in figure 18(c) is reproduced in figure 20(a) and may be compared with the one that is obtained through interpolant-based nudging in figure 20(b). While both modes are similar in terms of global spectral errors $\int_\Omega \Phi_e(\boldsymbol{x})^2 \, d\boldsymbol{x}$ (see equation (2.8)), it appears that the interpolant-based nudging approach has removed any trace of the measurement locations in the SPOD mode, which seems significantly smoother than with punctual nudging. Figure 20(b) even better illustrates the ability of

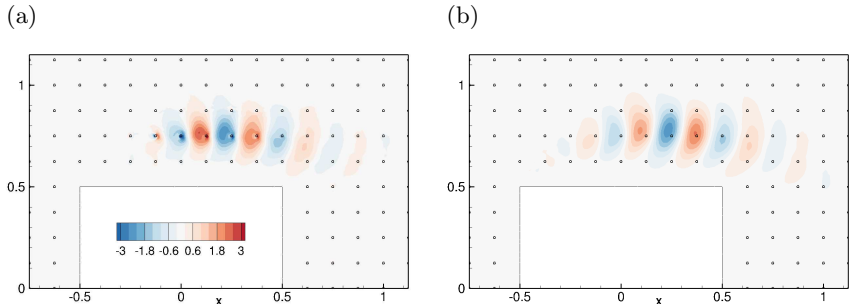

FIGURE 20. Real part of the transverse component of the SPOD mode for $St_{KH} = 4.384$ for nudged URANS simulations in data-set group KH with $\Delta s = 0.125$ where nudging is applied either (a) in its standard punctual form or (b) through interpolant.

the (nudged) URANS equations in providing an estimation of the spatial structure of the developing Kelvin-Helmholtz instabilities in between measurement points.

While not substantially affecting the statistics of the flow and despite of requiring more implementation work, interpolant-based nudging thus appears as an interesting alternative to punctual nudging to ensure the smoothness of the estimated velocity field in the case of too sparse measurements.

3.6. *Modelling and experimental considerations*

As discussed in §3.3-§3.5, the use of nudging in conjunction with URANS here appears to be an efficient flow estimation approach even when considering moderately dense data with respect to the phenomenon of interest. For both large-scale vortex shedding and Kelvin-Helmholtz instabilities, with a time interval between measurements that is below one tenth of the considered characteristic period, less than $N_s = 3$ nudging points per characteristic wavelength is already enough to correct the frequency content of URANS. The kinetic energy that is associated to the (SPOD) modes that oscillate at the frequencies of interest is also satisfactorily recovered with $N_s = 3$ and $N_s = 8$ for large-scale vortex shedding and Kelvin-Helmholtz phenomena, respectively. The detailed shape of these modes may be further adjusted when increasing the measurement spatial density. If one is only interested in mean quantities, a spatial sampling of $\Delta s = 0.25$, or equivalently $N_s = 4$ if the cylinder size is here considered as the characteristic length, is sufficient to significantly improve the estimation of the mean flow compared to standard URANS for the present flow configuration.

It may be worth emphasising that the present results have been obtained in conjunction with a specific RANS turbulence model, namely the Spalart-Allmaras one. Through the consideration of other models, one could possibly further decrease the requirements in terms of temporal and spatial samplings of the measurements, in particular concerning the estimation of the Kelvin-Helmholtz instabilities. As discussed in the introduction and reported in the study of Palkin *et al.* (2016), the use of Reynolds stress models should enable the prediction of Kelvin-Helmholtz phenomena in a self-sustained way. When applying data assimilation to further improve their estimation, it could thus be imagined in this case that less nudging points could be required in order to accurately predict the associated kinetic energy. Aside from the choice of the turbulence model in the URANS framework, more generally, it seems worth investigating the use of nudging to enhance the estimation of turbulent flows in conjunction with related modelling approaches, namely hybdrid approaches (Chaouat 2017; Sagaut *et al.* 2013) such as scale-adaptive simulations (SAS) (Menter & Egorov 2010), partially averaged Navier–Stokes (PANS) (Girimaji

2006), or partially integrated transport modelling (PITM) (Chaouat & Schiestel 2005), among others. Such approaches generally provide clearer and explicit definitions of the resolved and modeled scales, and thus may disambiguate which aspects, i.e. scales, in reference/experimental data the estimated flow is supposed to recover.

In the perspective of using the current nudging approach in an actual experimental context, it is interesting to note that the above mentioned temporal and spatial resolutions may be considered as quite realistic in an experimental point of view, i.e. if one were to consider performing a PIV experiment of the present flow. Similar to the current nudging procedure, PIV measurements are also typically characterised by regular spatial sampling of measurements within a rectangular-shaped measurement domain. Current state-of-the-art in high frame rate PIV includes cameras with a typical sensor size of roughly 2000 pixels. Using standard seeding conditions and current processing algorithms, an interrogation window would consist of 16 pixels with 75% overlap (Raffel *et al.* 2018). The final vector spacing would be $\Delta s_{PIV} = 0.03$, which roughly corresponds to the lowest spatial sampling investigated here. Considering a water-flow experiment and a cylinder size of e.g. 20 mm (leading to a flow velocity of roughly 1.1 m/s), the frequency of shear-layer instabilities would be roughly equal to 230 Hz. The present requirement of acquisition to be performed at a roughly ten times higher frequency (thus 2.3 kHz) is thus also reachable, as a high frame rate PIV system can nowadays operate with full laser energy and camera resolution at frequencies of $2 - 3$ kHz minimum.

Note that this remains of course a rapid and preliminary verification. Turning to an actual experiment would raise a number of additional challenges (local spatial averaging due to the PIV interrogation windows, measurement noise, etc.), which will be addressed in future steps. However, it is already quite promising that the amount of data necessary for the nudging method to yield excellent results is at first order compatible with current experimental capabilities.

4. Conclusion

Data assimilation for the accurate prediction of the turbulent flow past a square cylinder has been performed through nudging and in conjunction with a relatively low-fidelity and affordable simulation method, namely URANS modelling. Through a careful examination of mean and dynamical flow features in both physical and spectral spaces, this study has confirmed the potential of nudging in the state estimation of a complex flow that exhibits a wide range of spatio-temporal scales. Nudging was indeed able to enhance the URANS estimation based on sparse velocity observations from DNS. With very few measurements per characteristic wavelength, nudging allowed synchronisation and the accurate characterisation of large-scale vortex shedding. Despite the limitations of the considered RANS model in self-sustaining Kelvin-Helmholtz instabilities, the associated structures were also successfully estimated through nudging, based on a still moderate number of measurements per characteristic wavelength. SPOD and mean-flow analyses confirmed that nudging allowed a global improvement in the flow prediction, and in particular outside of the region that is directly driven by the measurements.

Variational and Kalman filter-based data assimilation techniques are derived on firmer theoretical grounds than nudging. Therefore, they could possibly provide even better reconstruction results than those reported in the present study. However, such approaches would be of the order of 100 times more expensive than nudging, as mentioned in the introduction. Furthermore, as non-sequential data assimilation techniques, variational approaches may be limited in the size of the assimilation window by the Lyapunov time scale of the flow (Chandramouli *et al.* 2020; Li *et al.* 2020; Wang & Zaki 2021).

α	0	10^{-5}	10^{-4}	0.001	0.01	0.1	1	1000
$\langle E \rangle$	30	29.6	26	4.8	1.86	1.58	1.52	1.52
$\langle E_I \rangle$	7.8	7.8	7.0	1.24	0.090	0.032	0.026	0.024
$\langle E_E \rangle$	22	21.6	19.0	3.4	1.78	1.54	1.5	1.48
$\left\langle \|\|f_x\|\|^2 \right\rangle$	1	1.00	0.94	0.20	$9.1 \cdot 10^{-3}$	$5.1 \cdot 10^{-4}$	$9.9 \cdot 10^{-6}$	$1.1 \cdot 10^{-11}$
$\left\langle \|\|f_y\|\|^2 \right\rangle$	1	1.00	0.94	0.15	$5.9 \cdot 10^{-3}$	$4.1 \cdot 10^{-4}$	$8.7 \cdot 10^{-6}$	$1.0 \cdot 10^{-11}$
$\alpha^2 \left\langle \|\|f_x\|\|^2 \right\rangle$	0	$1.7 \cdot 10^{-8}$	$1.6 \cdot 10^{-6}$	$3.5 \cdot 10^{-5}$	$1.5 \cdot 10^{-4}$	$8.6 \cdot 10^{-4}$	$1.7 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$
$\alpha^2 \left\langle \|\|f_y\|\|^2 \right\rangle$	0	$2.9 \cdot 10^{-8}$	$2.7 \cdot 10^{-6}$	$4.3 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$

TABLE 4. Sensitivity of time-averaged total, internal, external errors (respectively $\langle E \rangle$, $\langle E_I \rangle$, and $\langle E_E \rangle$), and forcing magnitudes ($\langle ||f_x||^2 \rangle$ and $\langle ||f_y||^2 \rangle$) to the nudging coefficient α for the data-set of group VS with $\Delta s = 0.125$ (see tables 2-3). In the fourth and fifth rows, the values of $\langle ||f_x||^2 \rangle$ and $\langle ||f_y||^2 \rangle$ are normalized by their values for standard URANS.

Accordingly, in complement to these approaches, nudging appears as a valuable and cost-efficient data assimilation technique for the state estimation of complex turbulent flows. This is reinforced by the fact that accurate reconstruction results may be obtained even when nudging is applied in conjunction with relatively low-fidelity simulations, as demonstrated by the present study, while nudging itself induces negligible supplementary computational cost.

This effectiveness of the nudging approach should foster its application in future work to even more complex flows than the one considered here, whatever in its standard or interpolant-based form. As detailed in §3.6, another perspective is the consideration of more advanced modelling approaches than URANS in conjunction with nudging that could provide a clearer definition of resolved/reconstructable scales and modelled ones. Perhaps more importantly, future work could deal with the consideration of actual experimental data. As discussed in §3.6, this may be encouraged by the fact that the identified requirements of nudging in terms of spatio-temporal resolution of the measurements seem compatible with the possibilities that are offered by optical techniques.

Acknowledgements

This work was partly supported by the HOMER project from the European Union's Horizon 2020 research and innovation program under grant agreement No. 769237.

Declaration of interests

The authors report no conflict of interests.

Appendix A. Influence of nudging parameter α

In this appendix, the sensitivity of the reconstruction results with respect to the parameter α in the nudged URANS equations (3.1) is investigated. Time-averaged global errors $\langle E \rangle$, $\langle E_I \rangle$ and $\langle E_E \rangle$ (see (2.6) and (3.5)) are summarized in rows 2 to 4 of table 4 for values of the nudging coefficient that range from $\alpha = 0$ (corresponding to standard URANS) to $\alpha = 1000$. The considered measurement data-set belongs to group VS with $\Delta s = 0.125$. Fifth and sixth rows in table 4 report the time average of the norm of the

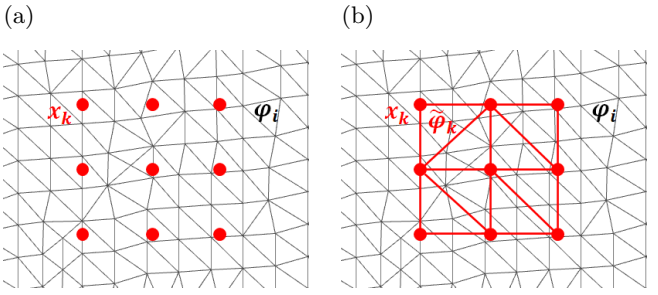

FIGURE 21. Sketch of (a) standard punctual and (b) interpolant-based nudging. The black triangular mesh corresponds to the one that is used to solve the nudged URANS equations (with associated basis functions $(\varphi_i)_{i=1\cdots N_u}$). Red dots refer to measurement locations $(\boldsymbol{x}_k)_{k=1\cdots M}$. In (b), these locations are used as vertices of the so-called measurement mesh which is reported in red (with associated basis functions $(\tilde{\varphi}_k)_{k=1\cdots M}$).

forcing vector

$$\boldsymbol{f} = \boldsymbol{m} - \mathcal{H}(\bar{\boldsymbol{u}}_\alpha), \tag{A 1}$$

which thus allows to quantify the discrepancies between estimated and reference solution at the measurement locations, distinguishing between the streamwise and crossflow components.

Table 4 confirms that the magnitude of the components of $\boldsymbol{f}$ (fifth and sixth rows) is reduced by more than three orders of magnitude for $\alpha \geqslant 0.1$ compared to standard URANS and, as expected, converges towards zero as α is further increased. On the other hand, the effective amplitude of the forcing that is introduced in the nudged URANS equations (3.1), which may be evaluated as $\alpha^2\langle||\boldsymbol{f}||^2\rangle$, converges towards constant values, as the errors with respect to the reference data seem to decrease proportionally to increasing values of the nudging coefficient (seventh and eighth rows in table 4). As this effective amplitude but also the global errors in the second to fourth rows of table 4 show minor sensitivity with respect to α for $\alpha \geqslant 1$, it was decided to maintain $\alpha = 100$ for all runs in the present contribution.

Appendix B. Nudging in a finite-element framework

We here provide some details about the implementation of nudging in both punctual and interpolant-based forms in conjunction with the finite-element solver FreeFEM++ (Hecht 2012). The finite-element representation of the streamwise component of the estimated velocity field (and similarly for its transverse component in the following) is given by

$$\bar{u}_\alpha(\boldsymbol{x}) = \sum_{i=1}^{N_u} q_{u,i}\,\varphi_i(\boldsymbol{x}), \tag{B 1}$$

where N_u is the number of degrees of freedom of the finite-element space which is defined on the computational mesh (see figures 1(b) and 21), $(\varphi_i)_{i=1\cdots N_u}$ are shape functions, here piece-wise second-order continuous polynomials (P2), and $(q_{u,i})_{i=1\cdots N_u}$ are unknown real coefficients which have to be determined and form the vector $\boldsymbol{q}_u$. The equations on $\boldsymbol{q}_u$ are obtained by integrating over the computational domain Ω the streamwise component of the nudged URANS equations multiplied by the shape functions. The left-hand-side

of the streamwise component of the nudged URANS equations thus becomes

$$n_{u,i} = \int_\Omega \left[\frac{\partial \bar{\boldsymbol{u}}_\alpha}{\partial t} + (\bar{\boldsymbol{u}}_\alpha \cdot \boldsymbol{\nabla})\bar{\boldsymbol{u}}_\alpha + \boldsymbol{\nabla}\bar{p}_\alpha - \boldsymbol{\nabla} \cdot \left[2 \left(Re^{-1} + \nu_t(\tilde{\nu}_\alpha) \right) \boldsymbol{\nabla}_{\mathrm{s}} \bar{\boldsymbol{u}}_\alpha \right] \right] \cdot \boldsymbol{e}_x \varphi_i \, d\boldsymbol{x}, \quad \text{(B 2)}$$

for the i-th degree of freedom. For the sake of simplicity, the discretisation of the time derivative is not addressed in the following. In vectorial form, the streamwise component of the nudged URANS equations (3.1) with punctual feedback becomes after spatial discretisation

$$\boldsymbol{n}_u(\boldsymbol{q}) = \alpha \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{m}_u - \boldsymbol{H}\boldsymbol{q}_u), \quad \text{(B 3)}$$

where the vector $\boldsymbol{q}$ concatenates the coefficients of all state variables in their finite-element representation (it thus includes $\boldsymbol{q}_u$), $\boldsymbol{m}_u$ refers to the streamwise component of the measurements $\boldsymbol{m}$ of the reference flow in (3.2), namely

$$m_{u,k} = \bar{u}_r(\boldsymbol{x}_k), \quad \text{(B 4)}$$

while $\boldsymbol{H}$ corresponds to the discretised version of the observation operator $\mathcal{H}$. The components of this matrix are obtained by evaluating the shape functions at the measurement locations $(\boldsymbol{x}_k)_{k=1\cdots M}$ (see figure 21(a)), which do not have to coincide with the degrees of freedom of the computational mesh, according to

$$H_{ki} = \varphi_i(\boldsymbol{x}_k). \quad \text{(B 5)}$$

As the shape functions have compact support on the elements of the computational mesh, most of the elements of $\boldsymbol{H}$ will be zero, resulting in a very localized right-hand-side in (B 3) through the application of $\boldsymbol{H}^{\mathrm{T}}$.

In the interpolant-based approach, as illustrated by figure 21(b), measurements are first interpolated on a mesh that is built from the measurement locations. We here consider the case where the interpolation is performed through piecewise linear continuous polynomials (P1). In this case, all degrees of freedom of the so-called measurement mesh identify with its nodes, i.e. the measurement locations. The streamwise component of the interpolated measurement field $\tilde{\boldsymbol{m}}$ in (3.6) is given by

$$\tilde{m}_u(\boldsymbol{x}) = \tilde{\mathcal{I}}(\boldsymbol{m}_u) = \sum_{k=1}^{M} m_{u,k} \, \tilde{\varphi}_k(\boldsymbol{x}), \quad \text{(B 6)}$$

where $(\tilde{\varphi}_k)_{k=1\cdots M}$ are the P1 shape functions that are defined on the measurement mesh. The coefficients in (B 6) are thus directly given by the measurement values. The discretised form of the streamwise component of the interpolant-based nudged URANS equations (3.7) is

$$\boldsymbol{n}_u(\boldsymbol{q}) = \alpha \tilde{\boldsymbol{I}} (\boldsymbol{m}_u - \boldsymbol{H}\boldsymbol{q}_u), \quad \text{(B 7)}$$

with

$$\tilde{I}_{ik} = \int_\Omega \varphi_i \tilde{\varphi}_k \, d\boldsymbol{x}. \quad \text{(B 8)}$$

The only difference with the punctual-nudging equations (B 3) thus lies in the above matrix $\tilde{\boldsymbol{I}}$ which originates from the interpolation operator $\tilde{\mathcal{I}}$. Contrary to $\boldsymbol{H}^{\mathrm{T}}$, the application of this matrix will enable a non-zero feedback term in (B 7) for any computational degree of freedom i that lies in the measurement domain, as in this case there will be at least one degree of freedom k of the measurement mesh for which the supports of φ_i and $\tilde{\varphi}_k$ have a non-empty intersection.

REFERENCES

Azouani, A., Olson, E. & Titi, E. S. 2014 Continuous Data Assimilation Using General Interpolant Observables. *Journal of Nonlinear Science* **24**, 277–304.

Azouani, A. & Titi, E. S. 2014 Feedback control of nonlinear dissipative systems by finite determining parameters - A reaction-diffusion paradigm. *Evolution Equations & Control Theory* **3** (4), 579–594.

Brooks, A. N. & Hughes, T. J. R. 1982 Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering* **32**, 199–259.

Buzzicotti, M. & Di Leoni, P. C. 2020 Synchronizing subgrid scale models of turbulence to data. *Physics of Fluids* **32**, 125116.

Chandramouli, P., Mémin, E. & Heitz, D. 2020 4D large scale variational data assimilation of a turbulent flow with a dynamics error model. *Journal of Computational Physics* **412**, 109446.

Chaouat, B. 2017 The State of the Art of Hybrid RANS/LES Modeling for the Simulation of Turbulent Flows. *Flow, Turbulence and Combustion* **99**, 279–327.

Chaouat, B. & Schiestel, R. 2005 A new partially integrated transport model for subgrid-scale stresses and dissipation rate for turbulent developing flows. *Physics of Fluids* **17**, 065106.

Colburn, C. H., Cessna, J. B. & Bewley, T. R. 2011 State estimation in wall-bounded flow systems. Part 3. The ensemble Kalman filter. *Journal of Fluid Mechanics* **682**, 289–303.

Da Silva, A. F. C. & Colonius, T. 2020 Flow state estimation in the presence of discretization errors. *Journal of Fluid Mechanics* **890**, A10.

Dandois, J., Mary, I. & Brion, V. 2018 Large-eddy simulation of laminar transonic buffet. *Journal of Fluid Mechanics* **850**, 156–178.

Di Leoni, P. C., Mazzino, A. & Biferale, L. 2018 Inferring flow parameters and turbulent configuration with physics-informed data assimilation and spectral nudging. *Physical Review Fluids* **3**, 104604.

Di Leoni, P. C., Mazzino, A. & Biferale, L. 2020 Synchronization to Big Data: Nudging the Navier-Stokes Equations for Data Assimilation of Turbulent Flows. *Physical Review X* **10** (1), 011023.

Evensen, G. 2009 *Data Assimilation: The Ensemble Kalman Filter (2nd edition)*. Springer-Verlag.

Franceschini, L., Sipp, D. & Marquet, O. 2020 Mean-flow Data Assimilation based on minimal correction of turbulence models: Application to turbulent high-Reynolds number backward-facing step. *Physical Review Fluids* **5**, 094603.

Girimaji, S. S. 2006 Partially-Averaged Navier-Stokes Model for Turbulence: A Reynolds-Averaged Navier-Stokes to Direct Numerical Simulation Bridging Method. *Journal of Applied Mechanics* **73**, 413–421.

Hayase, T. 2015*a* A Review of Measurement-Integrated Simulation of Complex Real Flows. *Journal of Flow Control, Measurement & Visualization* **03** (02), 51–66.

Hayase, T. 2015*b* Numerical simulation of real-world flows. *Fluid Dynamics Research* **47**, 051201.

Hayase, T. & Hayashi, S. 1997 State estimator of flow as an integrated computational method with the feedback of online experimental measurement. *Journal of Fluids Engineering, Transactions of the ASME* **119** (4), 814–822.

Hecht, F. 2012 New development in FreeFem++. *Journal of Numerical Mathematics* **20**, 251–265.

Hoke, J. E. & Anthes, R. A. 1976 The Initialization of Numerical Models by a Dynamic-Initialization Technique. *Monthly Weather Review* **104**, 1551–1556.

Iaccarino, G., Ooi, A., Durbin, P. A. & Behnia, M. 2003 Reynolds averaged simulation of unsteady separated flow. *International Journal of Heat and Fluid Flow* **24**, 147–156.

Imagawa, K. & Hayase, T. 2010 Numerical experiment of measurement-integrated simulation to reproduce turbulent flows with feedback loop to dynamically compensate the solution using real flow information. *Computers and Fluids* **39** (9), 1439–1450.

Lakshmivarahan, S. & Lewis, J. M. 2013 Nudging Methods: A Critical Overview. In *Data*

Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II), , vol. II, pp. 27–57.

Le Dimet, F.-X. & Talagrand, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **38A**, 97–110.

Lewis, F. L., Xie, L. & Popa, D. 2008 *Optimal and Robust Estimation: With an Introduction to Stochastic Control Theory (2nd edition)*. CRC Press.

Li, Y., Zhang, J., Dong, G. & Abdullah, N. S. 2020 Small-scale reconstruction in three-dimensional Kolmogorov flows using four-dimensional variational data assimilation. *Journal of Fluid Mechanics* **885**, A9.

Lyn, D. A., Einav, S., Rodi, W. & Park, J.-H. 1995 A laser-Doppler velocimetry study of ensemble-averaged characteristics of the turbulent near wake of a square cylinder. *Journal of Fluid Mechanics* **304**, 285–319.

Mary, I. & Sagaut, P. 2002 Large Eddy Simulation of Flow Around an Airfoil Near Stall. *AIAA Journal* **40**, 1139–1145.

Menter, F. R. & Egorov, Y. 2010 The Scale-Adaptive Simulation Method for Unsteady Turbulent Flow Predictions. Part 1: Theory and Model Description. *Flow, Turbulence and Combustion* **85**, 113–138.

Menter, F. R., Garbaruk, A., Smirnov, P., Cokljat, D. & Mathey, F. 2010 Scale-adaptive simulation with artificial forcing. In *Progress in Hybrid RANS-LES Modelling* (ed. Shia-Hui Peng, Piotr Doerffer & Werner Haase), pp. 235–246. Springer Berlin Heidelberg.

Moldodovan, G., Lehnasch, G., Cordier, L. & Meldi, M. 2021 A multigrid/ensemble Kalman filter strategy for assimilation of unsteady flows. *Journal of Computational Physics* **443**, 110481.

Mons, V., Chassaing, J.-C., Gomez, T. & Sagaut, P. 2016 Reconstruction of unsteady viscous flows using data assimilation schemes. *Journal of Computational Physics* **316**, 255–280.

Nakao, M., Kawashima, K. & Kagawa, T. 2009 Application of MI Simulation Using a Turbulent Model for Unsteady Orifice Flow. *Journal of Fluids Engineering* **131** (11), 1114011–1114016.

Neeteson, N. J. & Rival, D. E. 2020 State observer-based data assimilation: a PID control-inspired observer in the pressure equation. *Measurement Science and Technology* **31**, 014003.

Nisugi, K., Hayase, T. & Shirai, A. 2004 Fundamental study of hybrid wind tunnel integrating numerical simulation and experiment in analysis of flow field. *JSME International Journal, Series B: Fluids and Thermal Engineering* **47** (3), 593–604.

Olshanskii, M., Lube, G., Heister, T. & Löwe, J. 2009 Grad–div stabilization and subgrid pressure models for the incompressible Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering* **198**, 3975–3988.

Palkin, E., Mullyadzhanov, R., Hadziabdić, M. & Hanjalić, K. 2016 Scrutinizing URANS in Shedding Flows: The Case of Cylinder in Cross-Flow in the Subcritical Regime. *Flow, Turbulence and Combustion* **97**, 1017–1046.

Raffel, M., Willert, C.E., Scarano, F., Kähler, C., Wereley, S.T. & Kompenhans, J. 2018 *Particle Image Velocimetry: A Practical Guide*. Springer International Publishing.

Sagaut, P., Deck, S. & Terracol, M. 2013 *Multiscale and multiresolution approaches in turbulence*. Imperial College Press.

Saredi, E., Ramesh, N. T., Sciacchitano, A. & Scarano, F. 2021 State observer data assimilation for RANS with time-averaged 3D-PIV data. *Computers and Fluids* **218**, 104827.

Spalart, P. R. & Allmaras, S. R. 1994 A one-equation turbulence model for aerodynamic flows. *La Recherche Aérospatiale* **1**, 5–21.

Suzuki, T. 2012 Reduced-order Kalman-filtered hybrid simulation combining particle tracking velocimetry and direct numerical simulation. *Journal of Fluid Mechanics* **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. *Journal of Fluid Mechanics* **830**, 760–796.

Suzuki, T., Ji, H. & Yamamoto, F. 2009 Unsteady PTV velocity field past an airfoil solved

with DNS: Part 1. Algorithm of hybrid simulation and hybrid velocity field at $Re \simeq 10^3$. *Experiments in Fluids* **47**, 957–976.

Towne, A., Schmidt, O. T. & Colonius, T. 2018 Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *Journal of Fluid Mechanics* **847**, 821–867.

Trias, F.X., Gorobets, A. & Oliva, A. 2015 Turbulent flow around a square cylinder at reynolds number 22,000: A dns study. *Computers & Fluids* **123**, 87–98.

Wang, M. & Zaki, T. A. 2021 State estimation in turbulent channel flow from limited observations. *Journal of Fluid Mechanics* **917**, A9.

Yamagata, T., Hayase, T. & Higuchi, H. 2008 Effect of Feedback Data Rate in PIV Measurement-Integrated Simulation. *Journal of Fluid Science and Technology* **3** (4), 477–487.