

## Mean-flow data assimilation based on minimal correction of turbulence models: Application to turbulent high Reynolds number backward-facing step

Lucas Franceschini, Denis Sipp, and Olivier Marquet

*ONERA/DAAA, Université Paris Saclay, 8 rue des Vertugadins, 92190 Meudon, France*



(Received 10 March 2020; accepted 23 July 2020; published 14 September 2020)

In this article, we provide a methodology to reconstruct high Reynolds number turbulent mean-flows from few time-averaged measurements. A turbulent flow over a backward-facing step at  $Re = 28275$  is considered to illustrate the potential of the approach. The data-assimilation procedure, based on a variational approach, consists in correcting a given baseline model by tuning space-dependent source terms such that the corresponding solution matches available measurements (obtained here from direct-numerical simulations). The baseline model chosen here consists in Reynolds-averaged Navier-Stokes equations closed with the turbulence Spalart-Allmaras model. We investigate two possible tuning functions: a source term in the momentum equations, which is able to compensate for the deficiencies in the modeling of the Reynolds stresses by the Boussinesq approximation and a source term in the turbulence equation, which modifies the balance between the eddy-viscosity production and dissipation. The quality of the mean-flow reconstruction strongly depends on the baseline model and on the quantity of measurements. In the case of many measurements, very accurate reconstructions of the mean-flow are obtained with the model corrected by the source term in the momentum equations, while the reconstruction is more approximate when tuning the source term in the turbulence model. In the case of few measurements, this “rigidity” of the corrected turbulence model is favorably used and allows the best mean-flow reconstruction. The flexibility/rigidity of a model is further discussed in the light of a singular-value decomposition of the linear input/output operator between source term and measurements.

DOI: [10.1103/PhysRevFluids.5.094603](https://doi.org/10.1103/PhysRevFluids.5.094603)

### I. INTRODUCTION

Numerical simulations for turbulent aerodynamic flows are widely used in the everyday life of engineers and researchers. However, those techniques are prone to errors for several reasons: inaccurate turbulence models, erroneous numerical setup (including geometry and boundary conditions), imprecise upstream turbulence intensities, over-diffusive numerical methods, among others. For those reasons, experimental results are commonly used to validate numerical simulations. Yet, experimental techniques are also prone to uncertainties caused by different reasons, for example measurement biases, defects on the physical model or wind-tunnel (deformations, vibrations, inhomogeneous upstream conditions).

Data assimilation aims at intimately combining experiment results and numerical simulations to produce a compromise, an improved picture, by tuning uncertain parameters in numerical simulations to match measurement data within experimental uncertainty bounds. The compromise is driven by a balance between measurement accuracy (measurement covariance matrix) and model uncertainties (state covariance matrix). Several approaches have been designed to efficiently

determine such a compromise. These can be classified into two categories: gradient-based optimization techniques, also called 3D/4D Var, where uncertain parameters in the numerical model are optimized to minimize a cost-functional involving measurement mismatch and model accuracy [1]; ensemble-based techniques, where ensembles of members representing uncertainties are propagated with the model and corrected using measurements [2].

Data assimilation historically emerged in the field of meteorological forecasts (Lorenc [3], Liu *et al.* [4]), where the lack of accurate models, uncertain initial and boundary conditions yield poor predictions. In other areas of fluid mechanics, motivations for the use of data-assimilation techniques were multifold: estimation of initial or inlet boundary conditions in open flows [5,6], interpolation of velocity fields between sequences of images [7,8], identification of pollutant release location in urban areas [9], and even as a theoretical tool to investigate aspects of the decay of large scales in homogeneous isotropic turbulence [10]. In aerodynamics, predicting time-averaged quantities is the predominant interest in many industrial applications. Indeed, most of the measurements performed in industrial wind-tunnels are time-averaged pressure distributions and forces. Furthermore, the numerical simulations of turbulent flows over complex geometries are mostly achieved with the Reynolds-averaged Navier-Stokes (RANS) equations and turbulence models, thus promoting the low computational cost over the accuracy. In so far, we will therefore favor a time-averaged approach based on RANS equations and the data assimilation problem should rather be considered in the framework of inverse problems (see Foures *et al.* [11], Symon *et al.* [12] for the gradient-based optimization approach and Iglesias *et al.* [13], Kato and Obayashi [14] for the ensemble-based framework). Concerning the gradient-based optimization approach, Foures *et al.* [11] started with a low-Reynolds cylinder flow [ $\text{Re} = O(10^2)$ ] exhibiting vortex shedding. They tuned a volume-force (modeling the force associated to the Reynolds-stress) acting in the steady Navier-Stokes equations such that its corresponding solution best matches velocity measurements, mimicking a real experimental situation where such measurements are provided by a particle-image-velocimetry (PIV) setup. The same procedure was applied by Symon *et al.* [12] to reconstruct the mean-flow around an idealized airfoil at a higher Reynolds number [ $\text{Re} = O(10^4)$ ]. This time, additional difficulties related to the well-posedness of the steady Navier-Stokes equations at such high Reynolds numbers had to be faced. At even higher Reynolds numbers, the RANS equations supplemented with a turbulence model is a reasonable choice for the baseline model, since they usually provide solutions that aim at approximating the turbulent mean flow. For instance, Li *et al.* [15] optimized a set of coefficients in a  $k$ - $\omega$  RANS model to match as closely as possible some given higher-fidelity data. Yet, such an approach is strongly constrained by the structure of the turbulence model and does not allow a high enough flexibility to adjust the model, especially in flow regions where such models are known to be deficient. To overcome that limitation, Duraisamy *et al.* [16] employed gradient-based optimization techniques to tune spatially dependent production terms in turbulence models, so as to recover mean-flow data obtained by DNS or experiments. For example, Singh and Duraisamy [17] tuned space-dependent functions modulating the strength of the eddy-viscosity production term in the Spalart-Allmaras model. This procedure, called field-inversion by the authors, has been presented in more details in Parish and Duraisamy [18]. More recently, He *et al.* [19] applied this technique to reconstruct a mean-flow based on PIV measurement, which was then used to for a resolvent analysis (see also Symon *et al.* [20]). Concerning ensemble approaches, Refs. [14,21] varied arbitrarily the coefficients of a Spalart-Allmaras model to build the members of an ensemble and a Kalman filtering method then led to the tuning of optimal values of these coefficients with respect to given measurements.

The choice of the baseline RANS model is of importance in data-assimilation. Indeed, the better the model, the smaller the correction. Most of the turbulence models used in industrial design offices model Reynolds stresses with a Boussinesq approximation and an eddy-viscosity (as done in the most common turbulence models). Although this is known to be a strong weakness, the numerical procedure to find solutions with such models is far superior than for the more accurate Reynolds-stress models, which remain only seldomly used today by engineers. The robustness issue is even more critical for data-assimilation since tuning source term correction functions involves solving

stiff nonlinear state equations, that is RANS equations closed with a turbulence model and driven by an additional forcing.

In this article, we extend the time-averaged gradient-based optimization approach introduced by Refs. [11,12] to higher Reynolds number turbulent flows by using RANS equations with a turbulence model as baseline model for the data-assimilation procedure. In the present work, we choose the Spalart-Allmaras model (SA) [22] for its numerical simplicity and robustness, even though the procedure may be extended to any other turbulence model in principle. We explore two correction possibilities.

The first consists in a volume-force acting in the momentum equations (with the turbulence model still active). This force is supposed to stand for an optimal correction of the stresses induced by the eddy-viscosity term, whose spatial distribution is governed by the SA model. This allows in particular the turbulent stresses to escape the strong constraint linked to the Boussinesq assumption, which is known to be only well adapted in shear dominated regions. This approach may be considered as an extension of Foures *et al.* [11] and Symon *et al.* [12], the difference being the consideration of a background eddy-viscosity that evolves with the data-assimilation process according to solution of the full corrected model. This modeling is better suited for turbulent flows: For example, the uncorrected solution (for which the correction term is null), which is for example required to initialize the optimization algorithm, now corresponds to the RANS-SA solution. Such a solution may easily be obtained with standard numerical algorithms and is already much closer to the targeted mean-flow than an initial guess based on a laminar model. The assimilation procedure therefore only needs to compensate the *expected rather small* imperfections of the turbulence model, while using a laminar model for turbulent flows implies compensating for the whole discrepancy between the laminar steady solution and the targeted mean flow.

The second correction is a source term acting on the equation governing the strength of the eddy-viscosity. This time the correction only modifies the production and dissipation terms which drive the strength of the eddy-viscosity. Obviously, the Reynolds stresses cannot escape the constraint linked with the Boussinesq-approximation and a smaller set of reachable velocity-pressure fields is therefore expected with this second correction. Since a reference mean-flow (obtained for instance from DNS) might in fact not be compatible with a Boussinesq constraint, one expects less accurate mean-flow reconstruction using this second baseline model. Yet, we will see that such a choice may still have some advantages, both from a numerical point of view (robustness to find solutions) and from a modeling point of view. In particular, a crucial point when assessing a baseline model for data-assimilation lies in the quantity of available measurements. Complete mean velocity fields may for example be provided by advanced optical measurement methods, such as PIV measurements. However, such optical measurements are difficult to implement in industrial wind-tunnel facilities. Therefore, it is worth considering only few pointwise velocity measurements such as those given by Pitot probes or hot-wire measurements along a line to extract cross-stream velocity profiles. We will see that both the available quantity of measurements and the used correction-term in the baseline model have strong impacts on the quality of the mean-flow reconstruction. Foures *et al.* [11] and Symon *et al.* [12] have already examined pointwise measurements in the case of a laminar model at lower Reynolds numbers: In particular, Symon *et al.* [12] showed that pointwise measurements may lead to noisy reconstructions and that a measurement operator based on a spatial averaging of the measure over a finite region could improve the quality of the reconstruction by smoothing the gradients of the reconstructed velocity field in the vicinity of the measurements. We will show that we can handle noisiness associated to pointwise measurements in different ways, either by selecting a good correction term in the model or by penalizing small-scale features in the correction term. The latter can also be regarded as a covariance matrix modeling uncertainties of the correction term.

The article is organised as follows. First, in Sec. II we will describe the physical configuration of interest, by showing the reference solution (obtained by DNS) and the corresponding (uncorrected) RANS-SA solution, pointing out the differences between them and motivating the need for data-assimilation. The flow configuration is a rounded backward-facing step (BFS, Dandois *et al.* [23]), for which the SA model overestimates the recirculation length. Then, in Sec. III we will discuss

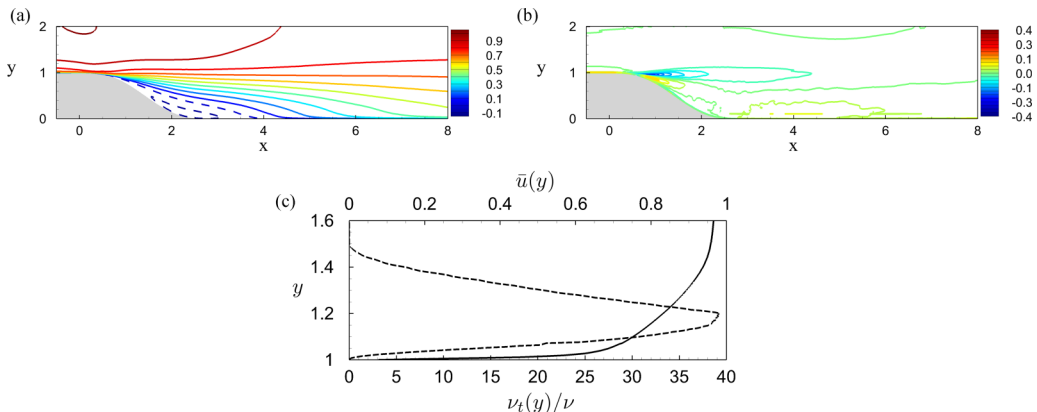


FIG. 1. Reference mean-flow solution obtained from DNS [23] at  $Re = 28275$ . (a) Streamwise velocity field  $\bar{u}$ , positive and negative isocontours being displayed with solid and dashed lines, respectively. (b) Streamwise component of the mean force  $\bar{f}_x$  induced by the Reynolds stress tensor. (c) Profiles of the streamwise velocity  $\bar{u}(y)$  (solid lines) and the eddy-viscosity ratio  $\nu_t(y)/\nu$  (dashed lines) at  $x = -2.5$ .

in more details the two above-mentioned correction functions together with the gradient-based optimization procedure underlying the data-assimilation procedure. Finally, in Sec. IV we will present the data-assimilation results on the backward-facing step configuration, by comparing to the reference mean-flow the assimilated flowfields obtained with the two baseline models, using first dense and then sparse velocity measurements. The dense velocity measurement case is an important theoretical step when addressing the performance of data-assimilation to reconstruct flow fields. Indeed, it allows to determine the best reconstructed flow-field for a given correction term, to which the flow reconstructed using sparse data can then be compared. Such ideal situation is of limited interest from an experimental point of view, but is more interesting in data-driven turbulence modeling [16]. For instance, in the field inversion and machine learning approach proposed by Refs. [17,18], the inverse modeling is first applied on a flow configuration to extract the (spatially) optimal correction terms for a given turbulence model. Machine learning is then used in a second step to transform the corrections terms computed for several flow configurations into corrective model forms.. The performance of the reconstruction as a function of the baseline model and measurement-data sparsity is finally discussed in light of an observability Gramian analysis of the linearized baseline model.

## II. FLOW CONFIGURATION AND NUMERICAL SOLUTIONS

We investigate the turbulent flow above a Backward-Facing rounded Step defined by  $y(x) = [\sin(a\pi x) - a\pi x]/(2\pi) + 1$ ,  $0 \leq x \leq 2/a$ , with  $a = 0.703$ . Direct numerical simulations (DNS) and large eddy simulations (LES) were performed by Dandois *et al.* [23] on this configuration at a Reynolds number  $Re = 28275$ , based on the height of the step and on the inflow velocity. In the following, those quantities are used to make all variables dimensionless. As detailed in Dandois *et al.* [23], the DNS was fed at the inlet boundary with fluctuations generated from a time-dependent zero-pressure-gradient turbulence simulation (see Lund [24]). This allows to describe a situation where the incoming boundary layer is fully turbulent. The streamwise component of the time- and spanwise-averaged velocity field is displayed in Fig. 1(a). The negative streamwise velocity (dashed line) indicates the existence of a mean recirculation region, extending from the separation point of the turbulent boundary layer at  $x_d = 0.53$  to the reattachment point at  $x_r = 3.93$ . The mean (time- and spanwise-averaged) velocity  $\bar{\mathbf{u}}$  and pressure  $\bar{p}$  fields satisfy the RANS equations:

$$\bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}} + \nabla \bar{p} - \nabla \cdot (\nu \nabla_s \bar{\mathbf{u}}) = -\nabla \cdot \underline{\underline{\tau}}, \quad \nabla \cdot \bar{\mathbf{u}} = 0, \quad (1)$$

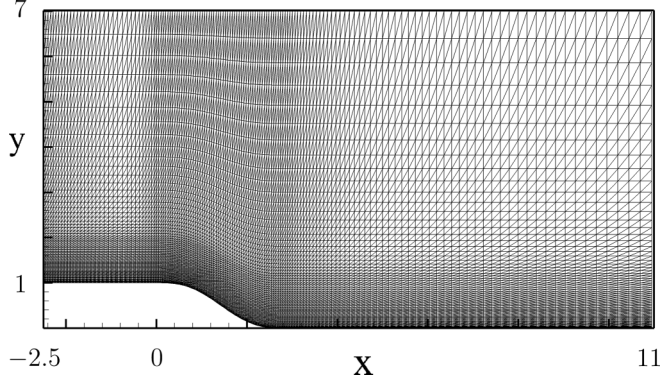


FIG. 2. Example of (coarsened) mesh used in the present study. The mesh contains  $10^5$  elements resulting in  $5 \times 10^5$  degrees of freedom using a finite-element-discretization with  $\mathcal{P}_b^1$  elements for the velocity and eddy-viscosity fields, and  $\mathcal{P}^1$  elements for the pressure.

where  $\nabla_s = (\nabla + \nabla^T)$  and  $\underline{\tau} = \overline{\mathbf{u}' \otimes \mathbf{u}'}$  is the Reynolds-stress tensor that contains all nonlinear interactions of the fluctuations  $\mathbf{u}'$ . It induces a force on the mean flow,  $\bar{\mathbf{f}} = -\nabla \cdot \underline{\tau}$ , whose streamwise component is displayed in Fig. 1(b).

Turbulence modeling consists in closing the above system of equations by approximating the Reynolds stress as a function of the mean flow. Among the several models existing in the literature, we will consider the one-equation Spalart-Allmaras model (see Ref. [22]). The mean-flow velocity and pressure  $(\bar{\mathbf{u}}, \bar{p})$ , that approximate the time- and spanwise-averaged flow  $(\bar{\mathbf{u}}, \bar{p})$ , are then solutions of the following closed system of equations:

$$\bar{\mathbf{u}} \cdot \nabla \bar{\mathbf{u}} + \nabla \bar{p} - \nabla \cdot [(v + \nu_t(\tilde{v})) \nabla_s \bar{\mathbf{u}}] = \mathbf{0}, \quad \nabla \cdot \bar{\mathbf{u}} = 0, \quad (2)$$

$$\bar{\mathbf{u}} \cdot \nabla \tilde{v} - \nabla \cdot [\eta(\tilde{v}) \nabla \tilde{v}] = s(\tilde{v}, \nabla \tilde{v}, \nabla \bar{\mathbf{u}}). \quad (3)$$

The equation on the second line is the Spalart-Allmaras model, which governs the eddy-viscosity-related variable  $\tilde{v}$ . Its diffusivity is denoted  $\eta(\tilde{v})$  and the source term is defined as  $s(\tilde{v}, \nabla \tilde{v}, \nabla \bar{\mathbf{u}}) = P(\tilde{v}, \nabla \bar{\mathbf{u}}) + D(\tilde{v}, \nabla \bar{\mathbf{u}}) + C(\nabla \tilde{v})$ , that is the sum of the production  $P(\tilde{v}, \nabla \bar{\mathbf{u}})$ , destruction  $D(\tilde{v}, \nabla \bar{\mathbf{u}})$ , and cross-diffusion  $C(\nabla \tilde{v})$  terms. More details about the exact definitions of  $\tilde{v}$  and these source terms are given in Appendix A.

A streamline-upwind Petrov-Galerkin (SUPG) finite-element method is used for the spatial discretization of Eqs. (2) and (3), instead of the classical Galerkin finite-element method used by Fournes *et al.* [11] (see Appendix B). For high Reynolds numbers flows, this allows stabilization of the convection operator. The computational domain is sketched in Fig. 2. No-slip boundary conditions  $\bar{\mathbf{u}} = 0$  are imposed at the lower wall, and symmetry boundary conditions  $(\partial_y \tilde{u}_x, \tilde{u}_y, \partial_y \tilde{v}) = (0, 0, 0)$  at the top boundary located at  $y = 7$ . At the outlet  $x = 11$ , a classical outflow condition  $(v + \nu_t) \nabla_s \bar{\mathbf{u}} \cdot \mathbf{n} + \bar{p} \mathbf{n} = 0$  is enforced for the momentum equations while a Neuman boundary condition  $\partial_x \tilde{v} = 0$  is used for the  $\tilde{v}$  equation.

At the inlet boundary ( $x = -2.5$ ), we impose the turbulent mean velocity profile of the DNS, displayed with the solid line in Fig. 1(b), and we consider two cases for the profile of the eddy-viscosity-related variable  $\tilde{v}$ . Firstly, the constant value  $\tilde{v}_v = (\tilde{v}/\nu)_\infty = 3$ , recommended in Allmaras *et al.* [25], is imposed, with a fast-decay to zero very close to the wall. Second, the actual eddy-viscosity of the DNS, displayed with the dashed line in Fig. 1(c), is imposed. It is computed from the DNS statistics as suggested by Mettrot *et al.* [26]. In that case, the maximum value of  $\tilde{v}/\nu$  is close to 40, which should result in filling the streamwise velocity profiles close to the wall and therefore increasing skin friction. The solutions of the RANS-SA model obtained with those two inflow eddy-viscosity profiles are compared in Fig. 3. We can see that, with the constant eddy-

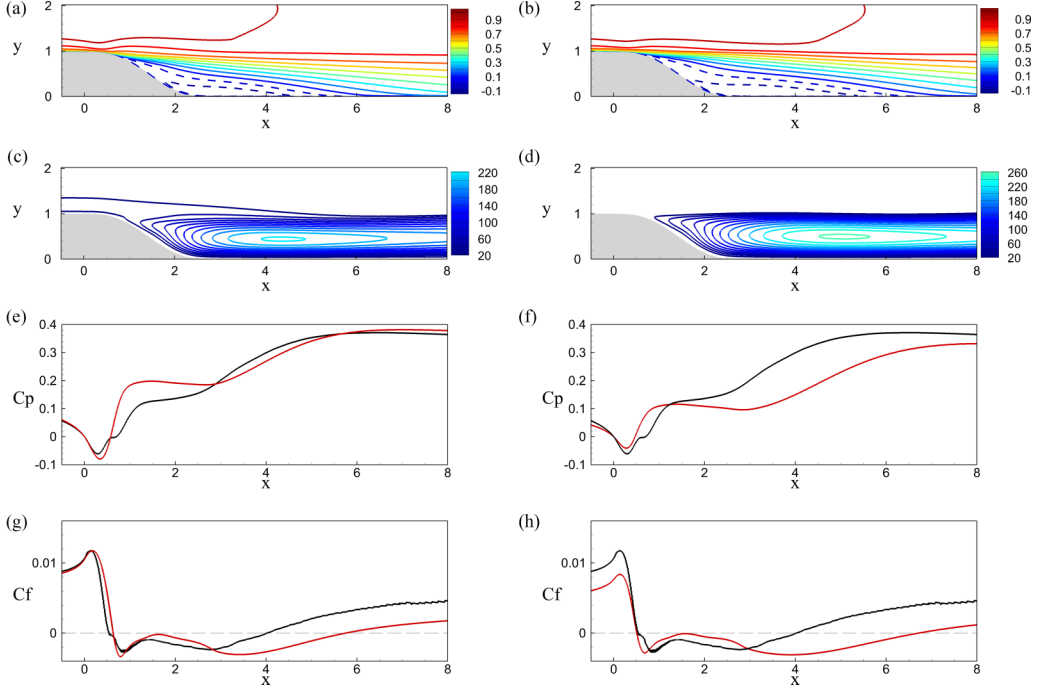


FIG. 3. Solutions of the (uncorrected) baseline RANS-SA model with two different eddy-viscosity profiles at the inlet. Left panels (a), (c), (e), (g) refer to the case of DNS-based eddy-viscosity profile [see Fig. 1(c)] and right panels (b), (d), (f), (h) to the case of constant eddy-viscosity  $\nu_t/\nu = 3$ . Isocontours in (a), (b) depict streamwise velocities  $\bar{u}$  and (c), (d) eddy viscosity fields  $\nu_t/\nu$ . Panels (e), (f) depict pressure  $C_p$  and (g), (h) friction  $C_f$  coefficients along the wall. In (e), (f), (g), (h), the black solid line refers to the reference solution of the DNS, the red solid line to the baseline RANS-SA model.

viscosity profile, the RANS solution presents a very large recirculation region ( $x_r \approx 6.6$ ), whereas, for the true eddy-viscosity profile (exhibiting much stronger values), the bubble is (accordingly much) shorter ( $x_r \approx 5.8$ ), but still much longer than the reference value  $x_r = 3.93$  of the DNS. This discrepancy is made more quantitative by examining the streamwise evolution of the wall-pressure and friction coefficients shown in Figs. 3(e) and 3(f), respectively. It is seen that the skin-friction coefficient in the attached boundary layer is well predicted in the case of the reference eddy-viscosity profile, and underestimated in the other case. With the reference eddy-viscosity profile, the incoming eddy-viscosity values at the step location are higher than with the constant eddy-viscosity profile [see Figs. 3(c) and 3(d)]. Hence, the shear-layer undergoes stronger diffusion and the bubble shortens. The discrepancies between the mean-flow results obtained with the baseline RANS-SA model and the reference solution motivate the use of data-assimilation techniques to recover the reference solution by tuning a small correction term in the baseline model. In this study, we will consider the time-averaged DNS as the reference solution from which measurements will be extracted to feed the assimilation process. The RANS-SA model with the two different inlet boundary conditions will be used as baseline models in the data-assimilation procedure, and we will solely consider turbulence modeling defects as an uncertain parameter. The inflow profiles, that could be viewed as uncertain parameters, will not be inferred in the present study.

### III. FLOW MODELS AND DATA ASSIMILATION PROCEDURE

In this section, after introducing and discussing the two source term corrections in the baseline model (Sec. III A), we present in some details the data-assimilation procedure (Sec. III B). With



respect to Foures *et al.* [11], we use an efficient low-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to find the minimum of the cost-functional from sole knowledge of the cost function and its gradient.

### A. Corrections of the RANS-SA baseline model

The baseline RANS-SA models with the two correction terms that will be considered in the following may be written as

$$\tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}} + \nabla \tilde{p} = \nabla \cdot [(\nu + \nu_t(\tilde{v}))\nabla_s \tilde{\mathbf{u}}] + \tilde{\mathbf{f}}_{\mathbf{u}}, \quad \nabla \cdot \tilde{\mathbf{u}} = 0, \quad (4)$$

$$\tilde{\mathbf{u}} \cdot \nabla \tilde{v} - \nabla \cdot (\eta(\tilde{v})\nabla \tilde{v}) = s(\tilde{v}, \nabla \tilde{v}, \nabla \tilde{\mathbf{u}}) + \tilde{f}_{\tilde{v}}, \quad (5)$$

with the boundary conditions discussed above.  $\tilde{\mathbf{f}}_{\mathbf{u}}$  and  $\tilde{f}_{\tilde{v}}$  are the two spatially dependent source terms introduced to correct the RANS-SA model. In the following we either consider  $\tilde{\mathbf{f}}_{\mathbf{u}} \neq 0$  and  $\tilde{f}_{\tilde{v}} = 0$  or vice versa. The volume-force  $\tilde{\mathbf{f}}_{\mathbf{u}}$  acts on the momentum equations without directly modifying the production of eddy-viscosity. Such a volume-force allows the Reynolds stresses to escape the Boussinesq constraint. The volume-force  $\tilde{f}_{\tilde{v}}$  acts on the equation governing the turbulent variable, without modifying the momentum equations. By modifying the balance between the production, destruction and cross-diffusion terms in the Spalart-Allmaras, this term modifies the eddy-viscosity  $\nu_t$  that appears in the momentum Eqs. (4), and has thus only an indirect effect (via  $\nu_t$ ) on the velocity  $\tilde{\mathbf{u}}$  and pressure  $\tilde{p}$  fields. The Reynolds stress tensor therefore remains constrained by the Boussinesq assumption, which is valid in shear-dominated flows. This model is sufficiently flexible to allow for example for correction of a wrong free shear-layer expansion ratio, as is the case in Fig. 3.

### B. Data assimilation procedure

In this section, we describe the data-assimilation procedure used here. Let  $\bar{\mathbf{m}}$  be a set of higher-fidelity or experimental measurements that correspond to information extracted from the flow and  $\mathcal{M}(\cdot)$  the measurement operator that allows to extract the corresponding measure from our simulation result  $(\tilde{\mathbf{u}}, \tilde{p})$ . In this work, since we are dealing only with velocity measurements, this operator will act on the velocity field,  $\tilde{\mathbf{u}}$ , yielding  $\tilde{\mathbf{m}} = \mathcal{M}(\tilde{\mathbf{u}}) \in M$ , where  $M$  is the measurement space, whose norm is given, generically, by  $\|\cdot\|_M$ . The data-assimilation problem can now be recast into an optimization one, for which the forcing terms (either  $\tilde{\mathbf{f}}_{\mathbf{u}}$  or  $\tilde{f}_{\tilde{v}}$ ) are tuned such that the cost functional:

$$J(\tilde{\mathbf{u}}) = \frac{1}{2} \|\mathcal{M}(\tilde{\mathbf{u}}) - \bar{\mathbf{m}}\|_M^2 \quad (6)$$

is minimal, with the velocity field  $\tilde{\mathbf{u}}$  satisfying the corrected RANS-SA Eqs. (4) and (5). We remark that, for now, no extra penalization of the correction term (that may be related to prior covariance matrices) is considered. Following Foures *et al.* [11], Parish and Duraisamy [18], Singh and Duraisamy [17], or Mons *et al.* [10], this optimization problem may be solved with an iterative gradient-based algorithm. It requires in particular the computation of the cost functional gradient with respect to the correction fields,  $\nabla_{\tilde{\mathbf{f}}_{\mathbf{u}}} J$  or  $\nabla_{\tilde{f}_{\tilde{v}}} J$ . To obtain an expression of the gradient, we resort to a Lagrangian formalism, that allows rewriting the constrained optimization problem into an unconstrained optimization problem. To that aim, the state is augmented with a set of Lagrange multipliers (or adjoint variables)  $(\tilde{\mathbf{u}}^\dagger, \tilde{p}^\dagger, \tilde{v}^\dagger)$  and we look for critical points of the Lagrangian functional:

$$\begin{aligned} L([\tilde{\mathbf{u}}, \tilde{p}, \tilde{v}], [\tilde{\mathbf{u}}^\dagger, \tilde{p}^\dagger, \tilde{v}^\dagger], [\tilde{\mathbf{f}}_{\mathbf{u}}, \tilde{f}_{\tilde{v}}]) \\ = J(\tilde{\mathbf{u}}) + (\tilde{\mathbf{u}}^\dagger, \tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}} + \nabla \tilde{p} - \nabla \cdot ((\nu + \nu_t(\tilde{v}))\nabla_s \tilde{\mathbf{u}}) - \tilde{\mathbf{f}}_{\mathbf{u}})_{\Omega} \\ + (\tilde{p}^\dagger, \nabla \cdot \tilde{\mathbf{u}})_{\Omega} + (\tilde{v}^\dagger, \tilde{\mathbf{u}} \cdot \nabla \tilde{v} - \nabla \cdot (\eta(\tilde{v})\nabla \tilde{v}) - s(\tilde{v}, \nabla \tilde{v}, \nabla \tilde{\mathbf{u}}) - \tilde{f}_{\tilde{v}})_{\Omega}, \end{aligned} \quad (7)$$

where  $(\mathbf{q}_1, \mathbf{q}_2)_\Omega = \int_\Omega \mathbf{q}_1 \cdot \mathbf{q}_2 \, d\Omega$  represents the inner product related to the classical  $\mathcal{L}_2$  norm. Setting to zero the variation of this Lagrangian with respect to the adjoint variables  $[\tilde{\mathbf{u}}^\dagger, \tilde{p}^\dagger, \tilde{v}^\dagger]$  yields the governing Eqs. (4) and (5). Setting to zero its variation with respect to the direct variables  $[\tilde{\mathbf{u}}, \tilde{p}, \tilde{v}]$  provides the adjoint equations of the RANS-SA model:

$$\nabla \cdot \tilde{\mathbf{u}}^\dagger = 0, \quad (8)$$

$$\begin{aligned} & \tilde{\mathbf{u}}^\dagger \cdot (\nabla \tilde{\mathbf{u}})^T - \tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}}^\dagger - \nabla \cdot [(\nu + \nu_t) \nabla_s \tilde{\mathbf{u}}^\dagger] - \nabla \tilde{p}^\dagger + \tilde{v}^\dagger \nabla \tilde{v} + \nabla \cdot (\tilde{v}^\dagger \partial_{\nabla \tilde{\mathbf{u}}} s) \\ & = - \left( \frac{\partial \mathcal{M}}{\partial \tilde{\mathbf{u}}} \right)^\dagger (\mathcal{M}(\tilde{\mathbf{u}}) - \bar{\mathbf{m}}), \end{aligned} \quad (9)$$

$$-\tilde{\mathbf{u}} \cdot \nabla \tilde{v}^\dagger - \nabla \cdot (\eta \nabla \tilde{v}^\dagger) + (\partial_{\tilde{v}} \eta) \nabla \tilde{v}^\dagger \cdot \nabla \tilde{v} + (\partial_{\tilde{v}} \nu_t) \nabla \tilde{\mathbf{u}}^\dagger : \nabla_s \tilde{\mathbf{u}} - (\partial_{\tilde{v}} s) \tilde{v}^\dagger + \nabla \cdot (\tilde{v}^\dagger \partial_{\nabla \tilde{v}} s) = 0. \quad (10)$$

We remark that, although we present the continuous formalism, in practice, we solve the discrete adjoint matrix, consisting in the transpose of the Jacobian matrix (which is necessary as well for the Newton method used to obtain the RANS-SA solutions). Indeed, one can show that this implementation of the adjoint represents a valid discretization of the above Eqs. (8), (9), and (10) in a finite-flement-method framework (see, for example, Ref. [27] for the stabilized finite element on a linear advection equation and references therein). Taking now the variation of the Lagrangian with respect to the forcing term ( $\tilde{\mathbf{f}}_{\mathbf{u}}$  or  $\tilde{f}_{\tilde{v}}$ ), we obtain the expressions of the gradients as a function of the adjoint variables:

$$\nabla_{\tilde{\mathbf{f}}_{\mathbf{u}}} J = -\tilde{\mathbf{u}}^\dagger, \quad \nabla_{\tilde{f}_{\tilde{v}}} J = -\tilde{v}^\dagger. \quad (11)$$

With this gradient information, we follow with the description of the optimization method employed.

### C. Optimization method

As mentioned above, for the optimization algorithm, we choose the (low memory) BFGS (see, for example, Ref. [28]) since it provides a second-order convergence, outperforming, in general, simple gradient descent methods. This higher-order convergence is achieved through an approximation of the Hessian  $\mathcal{H} = \nabla_{\tilde{\mathbf{F}}} \nabla_{\tilde{\mathbf{F}}} J$ , which contains the second-order derivatives of the cost functional  $J$  with respect to a generic forcing vector  $\tilde{\mathbf{F}}$ . This approximation is then used to find the descent direction by solving  $\mathcal{H}_n^{-1} \tilde{\mathbf{G}}_n$ , where  $\tilde{\mathbf{G}}_n$  is the numerical gradient at iteration  $n$ . This matrix is approximated through

$$\mathcal{H}_{n+1} = \mathcal{H}_n + \frac{\mathbf{y}_n \mathbf{y}_n^T}{\mathbf{y}_n^T \mathbf{s}_n} - \frac{\mathcal{H}_n \mathbf{s}_n \mathbf{s}_n^T \mathcal{H}_n}{\mathbf{s}_n^T \mathcal{H}_n \mathbf{s}_n}, \quad (12)$$

with  $\mathcal{H}_0 = I$ ,  $\mathbf{y}_n = \tilde{\mathbf{G}}_{n+1} - \tilde{\mathbf{G}}_n$  the difference of the gradient between two successive iterations and  $\mathbf{s}_n = \tilde{\mathbf{F}}_{n+1} - \tilde{\mathbf{F}}_n$  the difference in forcing vectors. From those relations, we can see that all inner products used in this algorithm (and as well in the line-search method, see Nocedal and Wright [29]) correspond to the Euclidean inner product  $\tilde{\mathbf{F}}^T \tilde{\mathbf{G}}$ , which may be inconsistent with the physical inner product  $(\tilde{f}, \tilde{g})_\Omega = \int_\Omega \tilde{f} \tilde{g} \, d\Omega = \tilde{\mathbf{f}}^T \mathcal{B} \tilde{\mathbf{g}}$ , where the symmetric positive matrix  $\mathcal{B}$  accounts for the metric corresponding to the spatial discretization. We propose to perform the change of variables  $\tilde{\mathbf{F}} = \mathcal{L} \tilde{\mathbf{f}}$ , where  $\mathcal{L}$  is the Cholesky decomposition of  $\mathcal{B} = \mathcal{L} \mathcal{L}^T$ ,  $\tilde{\mathbf{F}}$  is the vector used in the implemented BFGS algorithm and  $\tilde{\mathbf{f}}$  is the correction vector (the discrete counterpart of either  $\tilde{\mathbf{f}}_{\mathbf{u}}$  or  $\tilde{f}_{\tilde{v}}$  in our case). By doing so, we take into account the physical inner-product without changing the implementation of the BFGS algorithm, since  $\tilde{\mathbf{F}}^T \tilde{\mathbf{G}} = (\mathcal{L} \tilde{\mathbf{f}})^T (\mathcal{L} \tilde{\mathbf{g}}) = \tilde{\mathbf{f}}^T \mathcal{B} \tilde{\mathbf{g}}$ . A mass-lumping technique is used to perform efficiently the Cholesky decomposition, and for consistency, the gradient used in the BFGS algorithm is  $\tilde{\mathbf{G}} = \mathcal{L}^T \nabla_{\tilde{\mathbf{F}}} J$ . A sketch of the coupling of the BFGS method with our finite-element flow solver is shown in Fig. 4.





laminar) condition when the BFGS method is used without correcting the gradient direction (setting  $\mathcal{L} = \mathcal{I}$  in Fig. 4). This shows that informing BFGS algorithm the metric of the mesh is crucial for the proper convergence of BFGS.

## IV. RESULTS

We will now analyze reconstruction results of the backward-facing step flow for two different sets of measurements. The quality of the flow reconstruction will be monitored by scrutinizing the  $L_2$  norm of the reconstruction error:

$$e_\Omega = \sqrt{\left(\int_\Omega |\tilde{\mathbf{u}} - \bar{\mathbf{u}}|^2 d\Omega\right) / \text{Volume}(\Omega)}, \quad (13)$$

where  $\Omega$  denotes here the part of the domain (sketched in Fig. 2) near the wall, for  $y < 2$ .

First, in Sec. IV A, we will consider the complete mean velocity field to assess the capacity of the two corrected-models to accurately recover the velocity, pressure and Reynolds stress forcing. This is an important step since we will assess the capability of a given correction term to reproduce the various features of the reference solution. Second, in Sec. IV B, we will use only few pointwise velocity measurements and analyze the behavior of the two models when the inverse problem underlying the reconstruction procedure is strongly underdetermined. The evaluation of the performance of a given correction term if only sparse information is available is a complementary and more realistic test for the assessment of the data-assimilation procedure.

### A. Dense velocity field measurements

We consider the measure  $\bar{\mathbf{m}} = \bar{\mathbf{u}}$ , where  $\bar{\mathbf{u}}$  refers to the mean velocity field of the DNS interpolated on the mesh used for the optimization procedure (see Fig. 2). Since that mesh is coarser than the DNS-one, we consider that the error interpolation is negligible. The measure is thus defined on the same mesh as the solution in the optimization procedure. For this reason, the measurement space  $M$  ends up being the velocity space itself. The norm associated to the measurement space is chosen to be:  $\|\cdot\|_M = \|\cdot\|_\Omega$ , the  $L_2$  norm of the two velocity-components. This cost-functional is directly linked to the reconstruction error  $e_\Omega$ . In Eq. (9), we then have

$$-\left(\frac{\partial \mathcal{M}}{\partial \tilde{\mathbf{u}}}\right)^\dagger (\mathcal{M}(\tilde{\mathbf{u}}) - \bar{\mathbf{m}}) = -(\tilde{\mathbf{u}} - \bar{\mathbf{u}}). \quad (14)$$

In the following, we first (Sec. IV A 1) analyze the favorable case where, in the optimization problem, both the streamwise velocity and the eddy-viscosity profiles at the inlet correspond to those provided by the DNS [see Fig. 1(c)]. The reconstruction error related to this (uncorrected) baseline solution is  $e_\Omega = 0.06$ . Then (Sec. IV A 2), we consider the case where only the mean-velocity profile stems from the DNS and a constant eddy-viscosity profile is enforced at the inlet [ $\nu_t(y)/\nu = 3$ ]. The reconstruction error associated to the uncorrected solution is then slightly higher,  $e_\Omega = 0.094$ . In the latter case, we aim at assessing how volume source term corrections manage to compensate boundary condition defects in the model. In all cases, the optimization procedures are initialized with the baseline uncorrected RANS-SA model.

#### 1. Reference eddy-viscosity profile at the inlet

The optimization results, for both correction terms (left column figures for  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, right column for  $\tilde{f}_v$  correction), are provided in Fig. 6 in the case of the reference eddy-viscosity profile. Figures 6(a) and 6(b) show the reconstruction error  $e_\Omega$  as a function of the optimization iterations. The  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction manages to reconstruct the reference DNS solution extremely accurately due to the high flexibility of the model: the reconstruction error reaches very small values of the order of  $10^{-3}$ . This decrease is achieved in less than 200 iterations, similarly to the case of laminar flow

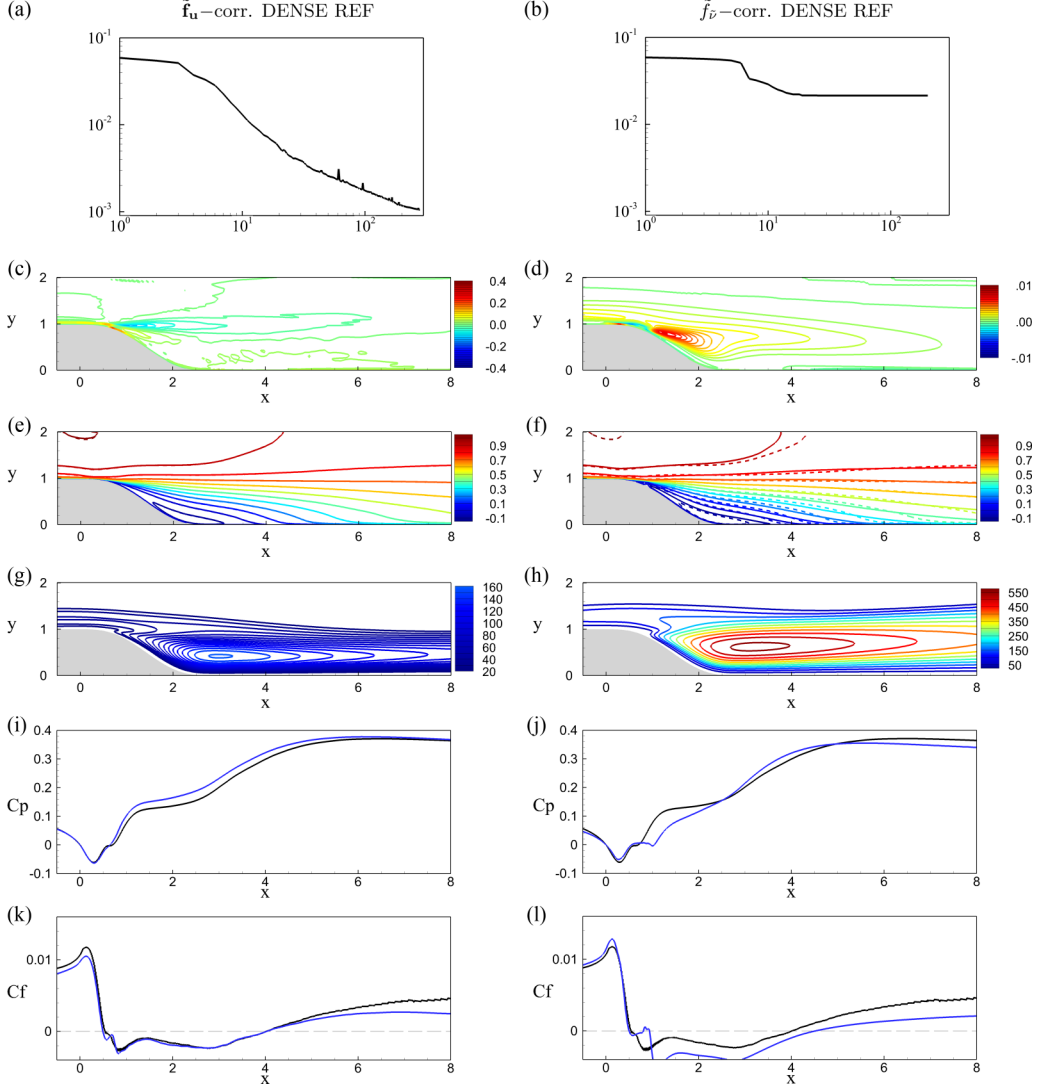


FIG. 6. Assimilation of *dense* velocity measurements with *reference eddy-viscosity profile* at the inlet. Left panels (a), (c), (e), (g), (i), (k) concern optimization with  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction and right panels (b), (d), (f), (h), (j), (l) to  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction. (a), (b) Error  $e_{\Omega}$  as a function of optimization iteration  $n$ . (c) Streamwise component of  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction. (d)  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction. (e), (f) Streamwise component of assimilated velocity field. The solid and dashed lines correspond to the assimilated and reference results, respectively [in panel (e) the contours are indistinguishable]. (g), (h) Eddy-viscosity field  $v_t/v$ . (i), (j) Wall-pressure along lower wall. (k), (l) Friction coefficient along lower wall. In panels (i), (j), (k), (l), the black solid line corresponds to the reference solution and the blue solid line to the reconstructed solution.

over a cylinder. When using the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction, the optimization procedure reaches a plateau in  $n \approx 10$  iterations, for which the reconstruction error is about  $e_{\Omega} = 0.02$  (to be compared with the baseline value that was  $e_{\Omega} = 0.06$ ). The convergence process is much faster but also less accurate than with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, due to the fact that the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction is far more constrained than the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction. We say thus that the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction is more constrained (or less controllable, or observable, in the context of data-assimilation) than the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  one. Indeed, with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, we are able to

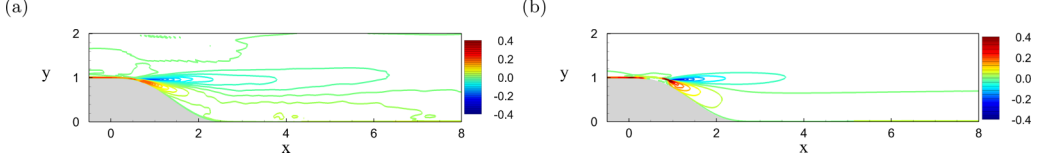


FIG. 7. Assimilation of *dense* velocity measurements with *reference eddy-viscosity profile* at the inlet. Resulting volume force  $\nabla \cdot (\nu_t \nabla_s \tilde{\mathbf{u}}) + \tilde{\mathbf{f}}_{\mathbf{u}}$ . (a) Optimization with  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction and (b) with  $\tilde{f}_{\mathbf{v}}$  correction.

encompass all corrections obtained with the  $\tilde{f}_{\mathbf{v}}$  correction, while the reverse is not true. Having observed this phenomenon, we may ask ourselves if the direct optimization of  $\nu_t$  could provide better results, since it is not restricted to some model, leading to, in theory, a more observable model. This does not happen to be the case in the configuration presented in this work. Some results on this are provided in Appendix C. As for the reconstructed velocity fields [Figs. 6(e) and 6(f)], we can accordingly see that the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction manages to accurately reconstruct the streamlines of the reference solution represented by the dashed lines [barely visible in Fig. 6(e)], whereas the  $\tilde{f}_{\mathbf{v}}$  correction only approximately achieves this goal (dashed-lines can be seen distinct from the solid-lines). In Figs. 6(k) and 6(l), the skin friction is accurately reconstructed in the separated flow region (reattachment point at  $x_r = 4.02$ ) for the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, while the agreement is more approximate for the  $\tilde{f}_{\mathbf{v}}$  correction (reattachment point around  $x_r = 4.64$ , the DNS reference and uncorrected RANS-SA values being, respectively,  $x_r = 3.93$  and  $x_r = 5.8$ ).

It is interesting to note that with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, the final eddy-viscosity values [Fig. 6(g)] have decreased with respect to the baseline values (the maximum is now around  $\nu_t/\nu \approx 160$ , compared to 220 for the uncorrected RANS-SA solution). This indicates that the baseline SA turbulence model induces errors that may be compensated most efficiently by replacing part of the Reynolds stresses modeled by a Boussinesq approximation by a general unconstrained forcing. In contrast, with the  $\tilde{f}_{\mathbf{v}}$  correction, the final eddy-viscosity values  $\nu_t$  have strongly increased for the bubble to become shorter, its maximum value being here  $\nu_t/\nu \approx 550$  [Fig. 6(h)]. Yet, the resulting reconstruction error remains much higher than with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction. Both optimized correction terms  $\tilde{\mathbf{f}}_{\mathbf{u}}$  and  $\tilde{f}_{\mathbf{v}}$  [Figs. 7(c) and 7(d)] are located in the vicinity of the baseline separation point. In Figs. 7(a) and 7(b), we can also see that the resulting overall momentum forcing terms (eddy-viscosity forcing together possibly with  $\tilde{\mathbf{f}}_{\mathbf{u}}$ ) are similar for both corrections and that they are qualitatively close to the reference Reynolds-stress force  $\tilde{\mathbf{f}} = -\nabla \cdot \underline{\tau}$  from the DNS [see Fig. 1(b)].

## 2. Constant eddy-viscosity profile at inlet

In the case of a defect in the boundary conditions of the baseline model, here an erroneous constant eddy-viscosity profile, results remain similar, as shown in Fig. 8. We just stress the main differences. The uncorrected baseline solution is slightly further away from the reference solution since it presents a larger recirculation region ( $e_{\Omega} = 0.094$  versus  $e_{\Omega} = 0.06$  before). This can be seen in Figs. 8(a) and 8(b). The  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction manages again to reach reconstruction errors of  $10^{-3}$ , while in the case of the  $\tilde{f}_{\mathbf{v}}$  correction we obtain 0.035 compared to 0.02 in the reference case. In Fig. 8(g), we observe that, contrary to before, there is no eddy-viscosity seen in the upstream boundary layer with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction: the latter forcing directly reconstructs the Reynolds-stress forcing and does not proceed through the Boussinesq term by triggering eddy-viscosity. The maximum value of  $\nu_t/\nu$  is now 160. On the contrary [Fig. 8(h)], the  $\tilde{f}_{\mathbf{v}}$  correction induces higher values of eddy-viscosity both in the upstream boundary layer (to compensate the boundary condition defect at the entrance) and in the recirculation region, where  $\nu_t/\nu = 360$ . An overshoot of the friction coefficient is even observed in the upstream boundary layer to manage to reduce as much as possible the size of the bubble (reconstruction errors over the whole bubble region dominate those in the upstream boundary layer). The resulting reattachment point in this case is  $x_r \approx 4.03$ , slightly above the value obtained in the previous section.

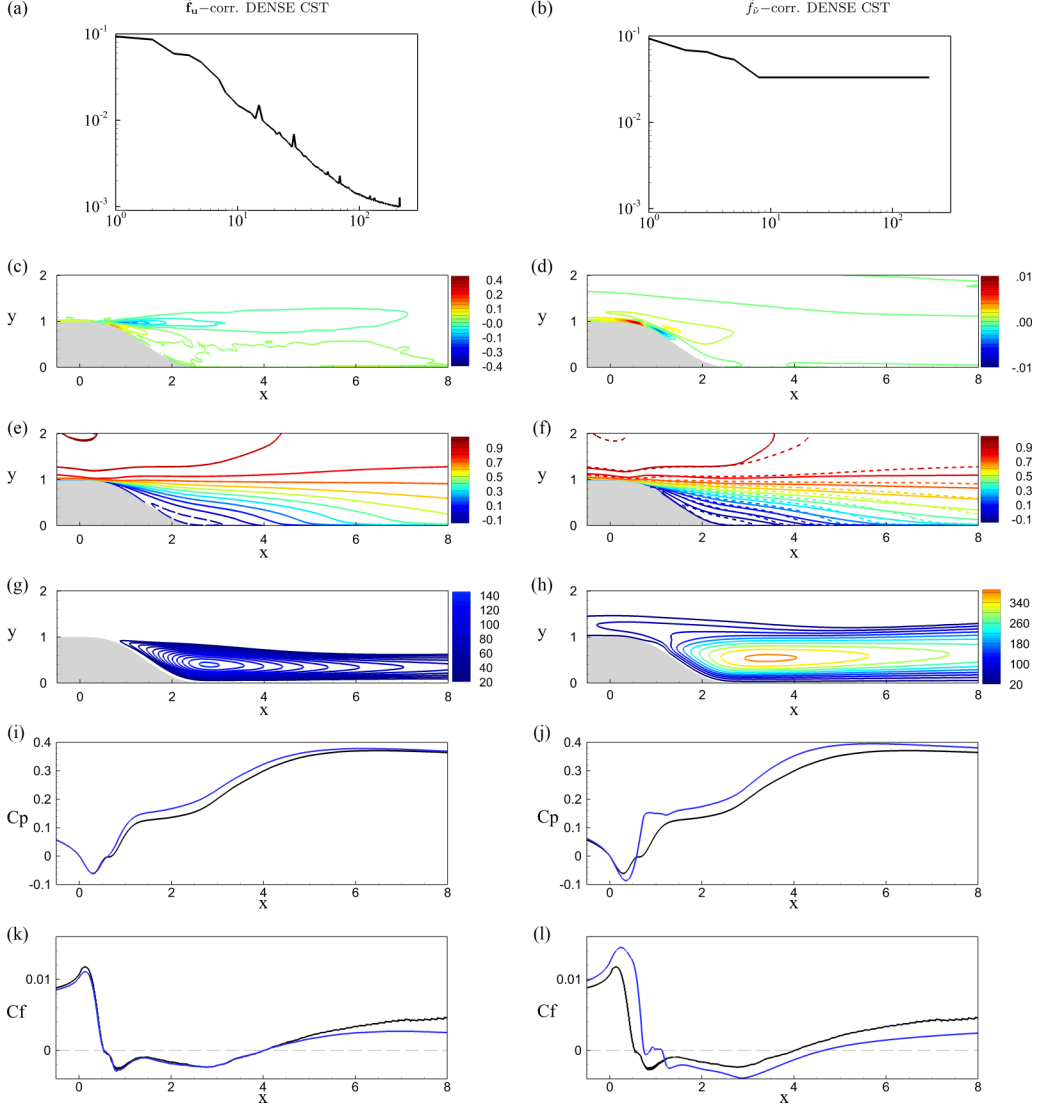


FIG. 8. Assimilation of *dense* velocity measurements with *constant eddy-viscosity profile* at the inlet. For more details see caption of Fig. 6.

### B. Sparse velocity measurements

In this section, we keep a baseline model where a constant eddy-viscosity profile is enforced at the inlet (see Sec. IV A 2). We consider the case where fewer velocity measurements are considered in the assimilation process. More precisely, the measurement operator is

$$\mathcal{M}(\tilde{\mathbf{u}}) = \{\tilde{\mathbf{u}}(\mathbf{x}_m^i)\}_{i=1,\dots,N}. \quad (15)$$

Here  $\mathbf{x}_m^i$  are  $N$  points where velocity measurements are available. Typically, these  $N$  measurements may correspond to cross-stream profiles of  $N_y$  points at  $N_x$  streamwise stations, such that

$N = N_x \times N_y$ . This way, the cost functional reads

$$J = \frac{1}{2} \sum_{i=1}^N |\tilde{\mathbf{u}}(\mathbf{x}_m^i) - \bar{\mathbf{m}}_i|^2, \quad (16)$$

where  $\bar{\mathbf{m}}_i = \bar{\mathbf{u}}(\mathbf{x}_m^i)$ . With this definition of the measurement operator, the right-hand side of Eq. (9) becomes

$$-\left(\frac{\partial \mathcal{M}}{\partial \tilde{\mathbf{u}}}\right)^\dagger (\mathcal{M}(\tilde{\mathbf{u}}) - \mathbf{m}) = -\sum_{i=1}^N (\tilde{\mathbf{u}}(\mathbf{x}_m^i) - \bar{\mathbf{m}}_i) \delta_{\mathbf{x}_m^i}, \quad (17)$$

where  $\delta_{\mathbf{x}_m}$  is the Dirac mass centered at  $\mathbf{x}_m$ .

### 1. Results for the $\tilde{f}_{\tilde{\mathbf{v}}}$ correction

Results of the optimization with  $N_x = 3$  and  $N_y = 10$  are reported in Fig. 9 with the  $\tilde{f}_{\tilde{\mathbf{v}}}$  correction. These plots should be compared to the optimization results with dense data shown in the right plots of Fig. 8. We can see that we reach  $e_\Omega = 0.037$ , which is very close to the dense value ( $e_\Omega = 0.035$ ). The  $\tilde{f}_{\tilde{\mathbf{v}}}$  correction in Fig. 9(b) is seen to be less peaked at the separation point. All other plots are close to the dense assimilation. Indeed, taking a close look at the error between the assimilated solution and the reference (see Fig. 10), we can see that the discrepancy is very close to the dense-data one, showing that, not only the global errors  $e_\Omega$  are similar but also its spatial distribution. Other optimization results for  $N_x$  and  $N_y$  are reported in Table I: it is seen that the reconstruction error  $e_\Omega$  is not very sensitive to the number of available measurements. Hence, in the case of sparse measurement data, the  $\tilde{f}_{\tilde{\mathbf{v}}}$  correction provides robust reconstruction results that are close to the dense case. We say thus that this model presented itself as “rigid,” or almost insensitive to the amount of data to be assimilated.

### 2. Results for the $\tilde{\mathbf{f}}_{\mathbf{u}}$ correction

Optimization results with  $N_x = 3$  and  $N_y = 10$  for the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction are shown in the left plots of Fig. 11. We can see from Fig. 11(e) that the isocontours of the reconstructed velocity exhibit non physical wiggles. Also, the wall-friction coefficient shown in Fig. 11(k) shows two large recirculation bubbles instead of one. This can be explained by observing that the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction is given by a linear combination of adjoint velocity fields with right-hand sides given by Eq. (17). Since those equations involve a sum of Dirac masses on the momentum equations, the adjoint velocity fields are “peaked” around the measurements, which contaminates the solution, leading to those undesired oscillations. However, this does not occur with the  $\tilde{f}_{\tilde{\mathbf{v}}}$  correction, since the  $\tilde{\mathbf{v}}$ -adjoint field is not directly forced by Dirac masses, leaving the  $\tilde{\mathbf{v}}$ -adjoint field smooth. Varying the number of measurements, Table I shows that increasing the number of measurements improves the reconstruction error  $e_\Omega$  progressively ( $e_\Omega = 0.051$  for  $N_x = 3$  and  $N_y = 5$  to  $e_\Omega = 0.027$  for  $N_x = 6$  and  $N_y = 20$ ). Hence, in the case of a larger number of measurements, the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction seems to exhibit more potential.

In the following, we show how to improve the reconstruction associated to the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction and in particular how to suppress the nonphysical wiggles that appear when sparse measurements are considered. For this, we add some knowledge about the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  field by noting that the Reynolds-stress forcing usually exhibits large-scale structures that vary on the length scales of the mean-flow. Therefore we propose to penalize unphysical small-scale features by considering the following modified cost functional  $\tilde{J}$ :

$$\tilde{J} = \frac{J}{J_0} + \frac{\gamma^2}{2} \int_{\Omega} |\nabla \tilde{\mathbf{f}}|^2 d\Omega. \quad (18)$$

Dividing the original cost functional  $J$  by its value at the first iteration, we normalize the term related to the measurement discrepancy to a unity value, making the penalization term independent of the



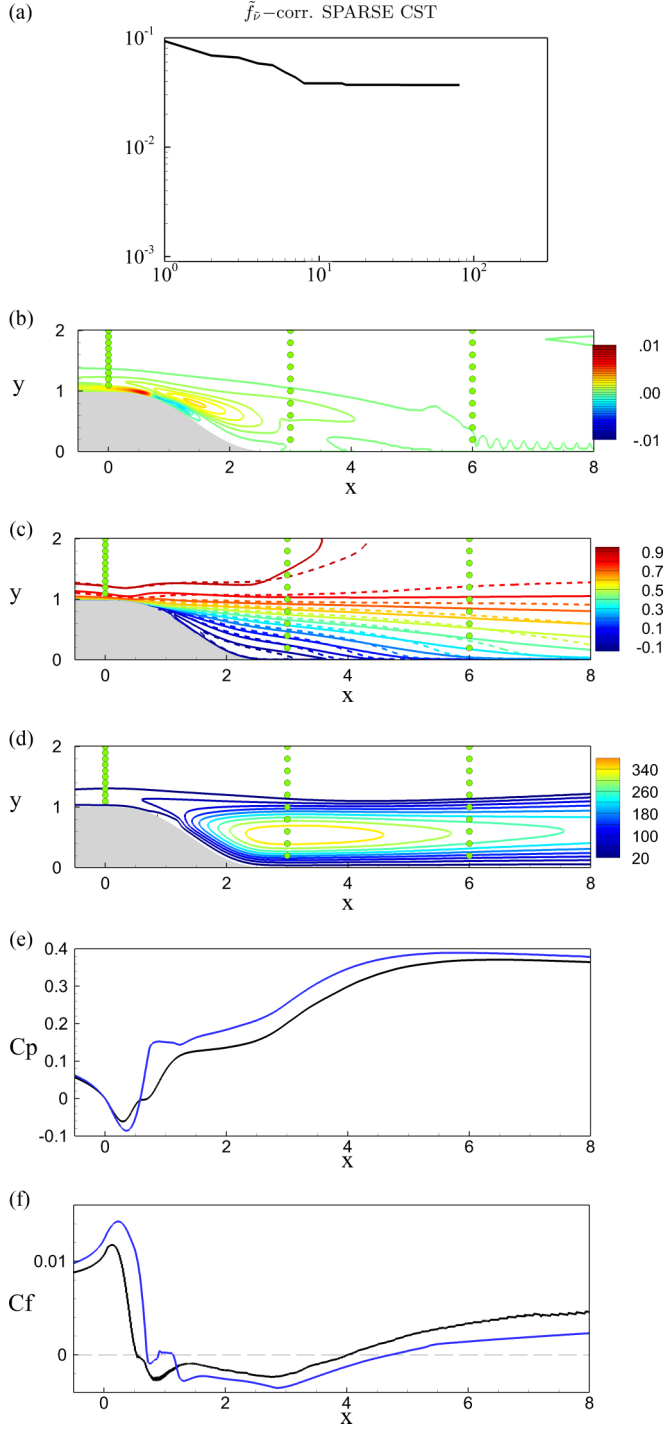


FIG. 9. Assimilation of *sparse* velocity measurements with *constant eddy-viscosity profile* at the inlet.  $N_x = 3$  and  $N_y = 10$  measurement have been used. For more details see caption of Fig. 6.

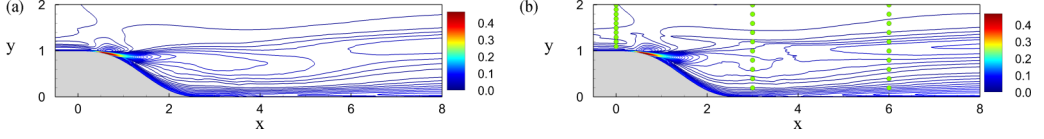


FIG. 10. Error of the assimilated solution obtained with the  $\tilde{f}_v$  model for (a) dense and (b) sparse measurements, with  $N_x = 3$  and  $N_y = 10$ .

measurement and only dependent on the parameter  $\gamma^2$ . This parameter should not be taken too small (the penalization needs to be effective) and not too large (we still want to decrease the measurement discrepancy). This cost-functional makes the algorithm favor a spatially smoother solution, since the tuning term  $\tilde{\mathbf{f}}_u$  in the momentum-equations will exhibit smaller gradient values. This is a physically sound constraint since we know that the Reynolds-stress forcing is smooth. It is interesting to notice that, in a Bayesian framework, this extra term is related to the prior covariance matrix for the forcing term  $\tilde{\mathbf{f}}_u$ ,  $\mathcal{C}_{\text{prior}}$  which, in this case, is modeled by the inverse of the stiffness matrix  $(2/\gamma^2)\mathcal{K}^{-1}$ , defined as  $\tilde{\mathbf{f}}_u^T \mathcal{K} \tilde{\mathbf{f}}_u^2 = \int_{\Omega} \nabla \tilde{\mathbf{f}}_u^1 \cdot \nabla \tilde{\mathbf{f}}_u^2 d\Omega$ . The results for this modified data-assimilation procedure are shown in the right plots of Fig. 11 for  $\gamma^2 = 1$ . We can see that the resulting solution is smoother and still matches well the measurements ( $J_n/J_0 \lesssim 1\%$ , see Table I), suggesting that the penalization (here derivatives of  $\tilde{\mathbf{f}}_u$ ) still allows enough freedom for the tuning field to match the measurements, while constraining the solution in a smooth subspace. Furthermore, in most of the situations (varying  $N_x$  and  $N_y$ , see Table I), the penalized algorithm provides an overall better reconstruction of the flow (see  $e_{\Omega}$  in Table I). Indeed, the penalized  $\tilde{\mathbf{f}}_u$  correction nearly always outperforms both the nonpenalized  $\tilde{\mathbf{f}}_u$  and the  $\tilde{f}_v$  reconstruction. Only, in the cases where very few measurements are provided (here  $N_x = 3$ ,  $N_y = 5$ ), the error with  $\tilde{f}_v$  is smaller than with the penalized  $\tilde{\mathbf{f}}_u$  term. This means that, whenever we have very few measurements, it may be preferable to use the more robust  $\tilde{f}_v$  correction than the more flexible penalized  $\tilde{\mathbf{f}}_u$  correction.

### C. Understanding the “rigidity” of the $\tilde{f}_v$ correction model through observability Gramian analysis

This section is devoted to the understanding of the “rigidity” of the  $\tilde{f}_v$  correction model, pointed out in the previous section. One possible way to understand this observation is to examine the linearized optimization problem, say, around the RANS-SA solution. In this linear framework, a variation of the control vector  $\delta \tilde{\mathbf{f}}$  (here, either  $\tilde{\mathbf{f}}_u$  or  $\tilde{f}_v$ ) induces a variation of the state variable  $\delta \tilde{\mathbf{q}} = (\delta \tilde{\mathbf{u}}, \delta \tilde{p}, \delta \tilde{v})$  that satisfies

$$\delta \tilde{\mathbf{q}} = (\partial_{\tilde{\mathbf{q}}} \mathcal{R})^{-1} P \delta \tilde{\mathbf{f}}, \quad (19)$$

TABLE I. Various optimization results (dense and sparse measurements) for the case of constant inflow eddy-viscosity profile.

Baseline		$e_{\Omega}$			$J_n/J_0$		
		$\tilde{\mathbf{f}}_u$	$\tilde{\mathbf{f}}_u, \gamma^2 = 1$	$\tilde{f}_v$	$\tilde{\mathbf{f}}_u$	$\tilde{\mathbf{f}}_u, \gamma^2 = 1$	$\tilde{f}_v$
Dense		$\sim 10^{-3}$	–	0.035	$\sim 10^{-4}$	–	12.3%
$N_x = 3$	$N_y = 5$	0.051	0.042	0.036	$\sim 10^{-6}$	0.1%	8.7%
	$N_y = 10$	0.037	0.031	0.037	$\sim 10^{-6}$	0.3%	8.7%
	$N_y = 20$	0.030	0.028	0.037	$\sim 10^{-6}$	0.4%	9.0%
$N_x = 6$	$N_y = 5$	0.042	0.035	0.037	$\sim 10^{-6}$	0.1%	11.9%
	$N_y = 10$	0.032	0.030	0.037	$\sim 10^{-6}$	0.5%	9.4%
	$N_y = 20$	0.027	0.027	0.038	$\sim 10^{-6}$	1.1%	13.4%

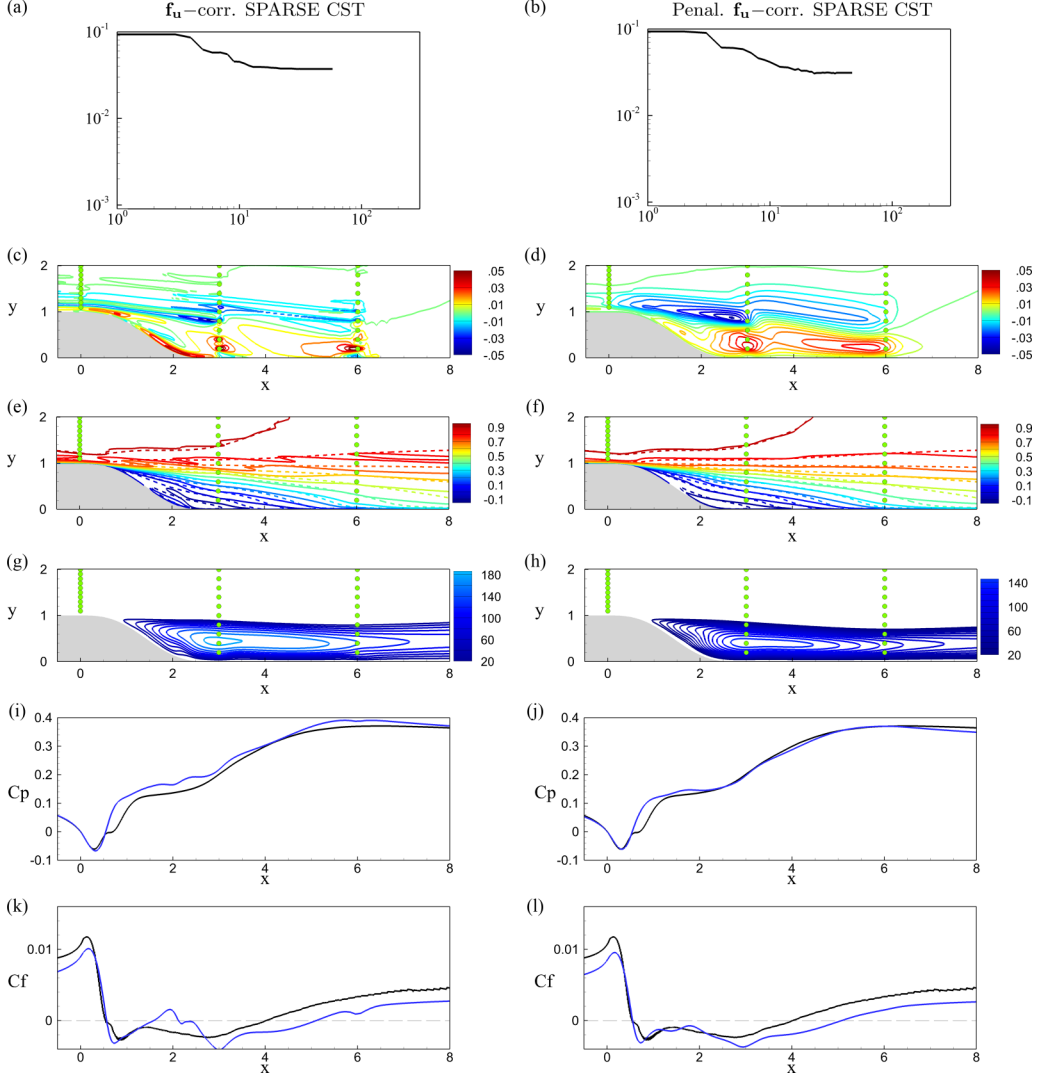


FIG. 11. Assimilation of *sparse* velocity measurements with *constant eddy-viscosity profile* at the inlet.  $N_x = 3$  and  $N_y = 10$  measurement have been used. For more details see caption of Fig. 6.

where  $\partial_{\tilde{\mathbf{q}}} \mathcal{R}$  denotes the linearization of the RANS-SA equations with the state variable and  $P$  a linear operator that maps the control vector variations to the actual forcings of the Jacobian. This variation in the state induces a variation on the measurements according to

$$\delta \tilde{\mathbf{m}} = (\partial_{\tilde{\mathbf{q}}} \mathcal{M}) \delta \tilde{\mathbf{q}} = \mathcal{A} \delta \tilde{\mathbf{f}}, \quad (20)$$

where the operator  $\mathcal{A} = (\partial_{\tilde{\mathbf{q}}} \mathcal{M})(\partial_{\tilde{\mathbf{q}}} \mathcal{R})^{-1} P$  is obtained after using Eq. (19).

To be able to identify which forcing term induces the most energetic variation on the measure, we may optimize the gain:

$$G(\delta \tilde{\mathbf{f}}) = \frac{\|\delta \tilde{\mathbf{m}}\|_M^2}{\|\delta \tilde{\mathbf{f}}\|^2} = \frac{\langle \delta \tilde{\mathbf{m}}, \delta \tilde{\mathbf{m}} \rangle_M}{\langle \delta \tilde{\mathbf{f}}, \delta \tilde{\mathbf{f}} \rangle_\Omega}. \quad (21)$$

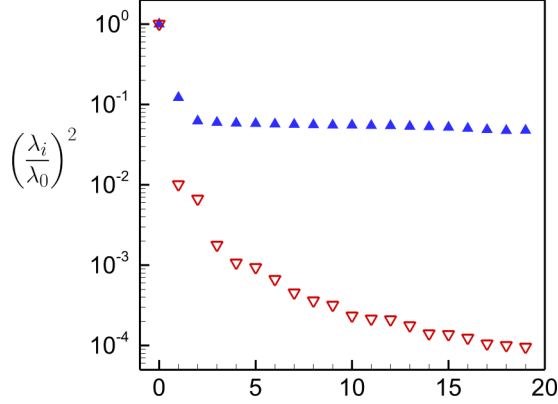


FIG. 12. Largest eigenvalues  $(\lambda_i/\lambda_0)^2$  computed for the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction model (blue upward triangles) and the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction model (red downward triangles). Linear Gramian analysis computed on the RANS-SA solution, with  $\mathcal{M}(\tilde{\mathbf{u}}) = \tilde{\mathbf{u}}$ .

This can straightforwardly be done by solving the eigenvalue problem:

$$\mathcal{A}^\dagger \mathcal{A} \delta \tilde{\mathbf{f}}_i = \lambda_i^2 \delta \tilde{\mathbf{f}}_i, \quad (22)$$

where the positive eigenvalues  $\lambda_i^2$  are ranked by decreasing order ( $\lambda_i^2 \geq \lambda_{i+1}^2$ ). By taking  $\delta \tilde{\mathbf{f}} = \delta \tilde{\mathbf{f}}_i$  in Eq. (21), we obtain  $G = \lambda_i^2$ . The values  $\lambda_i$  quantifies the measurement variations along the unit optimal measurement directions  $\delta \tilde{\mathbf{m}}_i = \lambda_i^{-1} \mathcal{A} \delta \tilde{\mathbf{f}}_i$ , induced by the unit optimal forcing directions  $\delta \tilde{\mathbf{f}}_i$ . If all the eigenvalues-values are of the same order, then we can state that any measurement  $\delta \tilde{\mathbf{m}}$  is equally reachable (or, more suitably, “observable”) and the nonlinear baseline model can be considered as more flexible. If we have a strong separation of singular values, that is  $\lambda_0^2 \gg \lambda_i^2$  for some index  $i$ , then we have measurement states that cannot be reached (“observed”) with the chosen model, that is therefore more rigid. This can be seen by considering that, for a given  $\delta \tilde{\mathbf{f}}$ , the resulting perturbation on the measure is given by

$$\delta \tilde{\mathbf{m}} = \mathcal{A} \delta \tilde{\mathbf{f}} = \mathcal{A} \sum_i \alpha_i \delta \tilde{\mathbf{f}}_i = \sum_i \alpha_i \lambda_i \delta \tilde{\mathbf{m}}_i = \lambda_0 \sum_i \alpha_i \frac{\lambda_i}{\lambda_0} \delta \tilde{\mathbf{m}}_i. \quad (23)$$

Since  $\delta \tilde{\mathbf{m}}_i$  is unit norm, the amplitude  $\alpha_i$  required for the forcing term to achieve a unit measurement variation  $\delta \tilde{\mathbf{m}}_i$  scales as  $\lambda_0/\lambda_i$ , which can be very large, and therefore not achievable.

In Fig. 12, we display the quantities  $(\lambda_i/\lambda_0)^2$  for the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  and  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction models, considering the full velocity field  $\mathcal{M}(\tilde{\mathbf{q}}) = \tilde{\mathbf{u}}$  as the measurement operator. We can see that more measurement states  $\delta \tilde{\mathbf{m}}$  can be reached with the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction compared to the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction, for which the separation between the eigenvalues is very strong and increases very rapidly for  $i > 1$ . This shows that the (linearized)  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction model is much more flexible than the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  one, in agreement with results of the previous section. In Fig. 13, we can also see that the leading forcing mode for both models is localized in the vicinity of the separation point and they both produce a similar measurement perturbation. This shows that, for both models, the separation point is the region where the tuning terms are most efficient to correct the reconstructed field and that the associated change in the reconstructed field consists in modifying the length of the separation bubble. Hence, both models are indeed able to correct easily this important feature of backward-facing step flow.

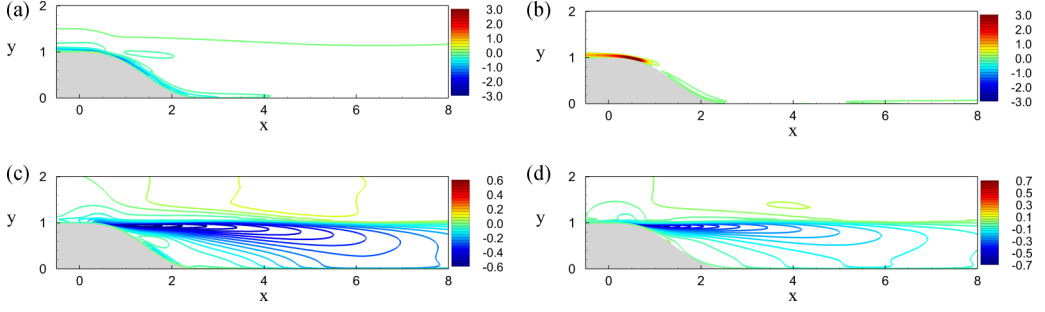


FIG. 13. Gramian modes: streamwise component of optimal forcing  $\delta\tilde{\mathbf{f}}_0$  (a), (b) and optimal response  $\delta\tilde{\mathbf{m}}_0$  (c), (d). (a), (c)  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction, (b), (d)  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction. Linear Gramian analysis computed on the RANS-SA solution, with  $\mathcal{M}(\tilde{\mathbf{u}}) = \tilde{\mathbf{u}}$ .

## V. CONCLUSION

In this paper, we have introduced a methodology for the reconstruction of mean-flow features from mean-velocity measurements for high-Reynolds number turbulent flows. The baseline model is in all cases the RANS Spalart-Allmaras model. Two correction terms to account for modeling uncertainties in the Spalart-Allmaras model have been introduced. These uncertainties have been tuned thanks to the knowledge of external measurements by minimizing the measurement discrepancy between the model and the data. The first correction-term consists in adding a volume-source term  $\tilde{\mathbf{f}}_{\mathbf{u}}$  in the momentum equations, in a similar way as in Foures *et al.* [11], the difference being that we have a background eddy-viscosity turbulence model, making the numerical procedure more robust and well-posed for high Reynolds number flows. The second one consists in the correction of the equation governing the eddy-viscosity. We have then considered a smooth turbulent backward-facing step flow to showcase the methodology. If the whole velocity field is known (like in a PIV setup), then the  $\tilde{\mathbf{f}}_{\mathbf{u}}$  correction model produces a solution that matches exactly the reference, showing the high flexibility of this model. On the contrary, the  $\tilde{\mathbf{f}}_{\mathbf{v}}$ -correction model is not capable to reproduce exactly the reference state. The model is too “rigid” and lots of measurements are not accessible with this model. If only few pointwise velocity measurements are available, then we have shown that the first model, despite its high flexibility to exactly recover the measurements, may lead to noisy/unphysical state reconstructions. For this reason, an additional penalization of the gradients of the tuning field had to be considered to obtain smooth physical solutions. As for the second model, it turned out that, independently of the quantity of available measurements, the optimal state was approximately the same. The Boussinesq hypothesis induces a strong constraint on the reconstructed state. The correction term acting only by modification of the diffusion strength  $\nu_t$ , it is possible, for example, to correct the expansion rate of the shear-layer, but not to finely adapt the flowfield in the vicinity of the attachment point. The flexibility/rigidity of the two considered models has finally been analyzed by looking at the input/output properties of the linear operator between the forcing space and the measurement space. We showed that the controllability/observability of the  $\tilde{\mathbf{f}}_{\mathbf{v}}$  correction model was actually restricted to a rather small subspace. Such a rigidity property may be helpful whenever the number of measurements is very low since the reconstructed state is almost independent of the number of measurements. Hence, the  $\tilde{\mathbf{f}}_{\mathbf{v}}$ -correction should be favored in the case of very few available measurements while the penalized  $\tilde{\mathbf{f}}_{\mathbf{u}}$ -correction is more suited in the case of many measurements.

As a possible extension to the data-assimilation performed in that study, we may consider other RANS-based turbulence modeling such as Reynolds-stress models (RSM). Those models are not based on the Boussinesq hypothesis and extra transport equations for the Reynolds-stress tensor are also solved, which could be tuned in the data-assimilation framework. This could enlarge the space

of observable modes (and thus accommodating more measurements) while retaining the physics coming from the closure of the model.

Also, even though it is difficult on physical grounds to provide prior covariance matrices for the correction terms  $\tilde{\mathbf{f}}_{\mathbf{u}}$  and  $\tilde{f}_{\tilde{v}}$ , posterior covariance matrices can, under certain conditions (related to linearity), be determined from the second derivative (Hessian) of the objective functional with respect to  $\tilde{\mathbf{f}}_{\mathbf{u}}$  and  $\tilde{f}_{\tilde{v}}$  [17,30]. Performing the optimization with successive estimations of these matrices may yield, at convergence (in terms of covariance matrices), to a reconstructed solution that better balances measurement and model uncertainties. Direct computation of the Hessian matrix being prohibitive in the case of large-scale correction terms, techniques to approximate or extract essential features of the Hessian need to be considered [17,30].

## APPENDIX A: SPALART-ALLMARAS MODEL

Based on the definition given by Eq. (3), we have the production, destruction, and cross-diffusion terms, respectively,

$$P(\tilde{v}, \nabla \bar{\mathbf{u}}) = c_{b1} \tilde{v} \tilde{S}, \quad D(\tilde{v}, \nabla \bar{\mathbf{u}}) = c_{w1} f_{w1} \left( \frac{\tilde{v}}{d} \right)^2, \quad C(\nabla \tilde{v}) = \sigma^{-1} |\nabla \tilde{v}|, \quad (\text{A1})$$

the eddy-viscosity  $\nu_t$  and the SA diffusivity  $\eta$  given by

$$\nu_t = \begin{cases} \tilde{v} f_{v1}, & \tilde{v} \geq 0 \\ 0, & \tilde{v} < 0 \end{cases}, \quad \eta(\tilde{v}) = \nu \left( 1 + \chi + \frac{\chi^2}{2} \right). \quad (\text{A2})$$

Those functions are all  $C^1$  with respect to the state, since the auxiliary function  $f_{v1}$  has a smooth and null first derivative at  $\tilde{v} = 0$ , leaving the Jacobian of the system continuous. The auxiliary functions that close the model are

$$\begin{aligned} \chi &= \frac{\tilde{v}}{\nu}, \quad f_{v1} = \frac{\chi^3}{c_{v1}^3 + \chi^3}, \quad \tilde{S} = |\nabla \times \bar{\mathbf{u}}| + \frac{\tilde{v} f_{v2}}{k^2 d^2}, \quad f_{v2} = 1 - \frac{\chi}{1 + \chi f_{v1}}, \\ f_w &= g \left[ \frac{1 + c_{w3}^6}{g^6 + c_{w3}^6} \right]^{\frac{1}{6}}, \quad g = r + c_{w2}(r^5 - r), \quad r' = \frac{\tilde{v}}{\tilde{S} k^2 d^2}, \\ r &= \begin{cases} r', & 0 \leq r' \leq 10 \\ 10, & r' < 0, r' > 10 \end{cases}, \quad g_n = 1 - 1000 \frac{\chi^2}{1 + \chi^2}. \end{aligned} \quad (\text{A3})$$

Although the function  $r = r(r')$  is not differentiable (and not even continuous at  $r' = 0$ ), the function  $r = r(\tilde{v}, \nabla \bar{\mathbf{u}})$  is (see Crivellini *et al.* [31] for more details).

## APPENDIX B: SUPG IMPLEMENTATION

The numerical implementation of the RANS-SA equations are based on the finite-element method (FEM), available in the FreeFem ++ code (see Hecht [32]). Since FEM is naturally numerically unstable at high Reynolds numbers, some stabilization scheme needs to be employed. Here, we choose the streamline-upwind Petrov-Galerkin (SUPG) formulation, as proposed by Brooks and Hughes [33]. In this formulation, the test function is advected with the local velocity field, giving an upwind effect, stabilizing the scheme. Several different formulations have been proposed in the literature (see Franca *et al.* [34], Franca and Frey [35]) for various different equations. Here, we employ a simplified version of it, common for unsteady problems (Bao *et al.* [36]), where only the advection terms are treated. In a simplified notation, we write the nonlinear residual



of the RANS-SA equations in the weak form:

$$\begin{aligned}
 R([\tilde{\mathbf{u}}, \tilde{p}, \tilde{v}], [\tilde{\mathbf{v}}, \tilde{q}, \tilde{v}]) = & \int_{\Omega} (\tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}}) \cdot \tilde{\mathbf{v}} + \int_{\Omega} [-pI + (\nu + \nu_t) \nabla_s \tilde{\mathbf{u}}] : \nabla \tilde{\mathbf{v}} - \int_{\Omega} (\nabla \cdot \tilde{\mathbf{u}}) \tilde{q} \\
 & + \int_{\Omega} (\tilde{\mathbf{u}} \cdot \nabla \tilde{v} - s) \tilde{v} + \int_{\Omega} \eta \nabla \tilde{v} \cdot \nabla \tilde{v} \\
 & + \sum_{\Omega_k} \int_{\Omega_k} \tau_{\text{SUPG}} \tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{v}} (\tilde{\mathbf{u}} \cdot \nabla \tilde{\mathbf{u}}) + \sum_{\Omega_k} \int_{\Omega_k} \tau_{\text{SUPG}} \mathbf{u} \cdot \nabla \tilde{v} (\tilde{\mathbf{u}} \cdot \nabla \tilde{v}), \quad (\text{B1})
 \end{aligned}$$

where the last two terms correspond to the SUPG formalism and the remaining terms are due to the classical (unstable) Finite-Element formulation. The function  $\tau_{\text{SUPG}}$  regulates the amount of numerical diffusivity and depends on the local Reynolds ( $\text{Re}_h$ ) number as

$$\tau_{\text{SUPG}} = \frac{\xi(\text{Re}_h) h_T}{2|\mathbf{u}|}, \quad \xi(\text{Re}_h) = \begin{cases} \text{Re}_h/3 & \text{Re}_h \leq 3 \\ 1 & \text{Re}_h > 3 \end{cases}, \quad \text{Re}_h = \frac{|\mathbf{u}| h_T}{2\nu}, \quad (\text{B2})$$

where the function  $\xi(\text{Re}_h)$  is constant for high Reynolds number, saturating this way the amount of numerical dissipation introduced. The parameter  $h_T$  indicates the local element size and is taken here as  $h_T = \sqrt{2A/h_T^{\text{max}}}$ , minimizing the numerical dissipation for highly elongated mesh elements (see Mittal [37]).

### APPENDIX C: EDDY-VISCOSITY

In this Appendix, we briefly discuss the optimization of the eddy-viscosity field  $\nu_t$  instead of  $\tilde{\mathbf{f}}_{\mathbf{u}}$  or  $\tilde{f}_{\tilde{\mathbf{v}}}$ . Arguably, this model is less constrained than the  $\tilde{f}_{\tilde{\mathbf{v}}}$  one and should produce results in between the two models described in the body of the text. However, there are additional difficulties that arise, which we discuss here.

First, the “natural” initial condition ( $\nu_t = 0$ ) for the algorithm is generally ill-posed at very large Reynolds numbers and may not even exist for many configurations. In the present case, the laminar solution would exhibit a laminar boundary layer profile and a huge recirculation bubble, which is very far from the target mean-flow solution. The optimization algorithm needs therefore to achieve a strong correction to match the actual measurement fields, which will result (at best) in many iterations for the optimization algorithm. A solution to this initial condition issue would be to start with a nonzero  $\nu_t$ , for example, the  $\nu_t$ -field of the RANS-SA model. This allows to be closer to the final solution from the start and would also ensure that the optimization algorithm has better chances to reach the global minimum (since we start the algorithm with an initial condition that is close to the optimal solution).

Second, the optimization with  $\nu_t$  may be harder than the  $\tilde{f}_{\tilde{\mathbf{v}}}$  one. This comes from the fact that the gradient involves derivatives of the velocity field and of its adjoint  $\nabla_{\nu_t} J = (\nabla \tilde{\mathbf{u}} + \nabla \tilde{\mathbf{u}}^T) : \nabla \tilde{\mathbf{u}}^\dagger$ , which are usually very strong at the separation point, meaning that optimized  $\nu_t$  field would exhibit strong values and strong gradients in the very vicinity of the wall. This poses not only the problem of the “lost” physical meaning of the eddy-viscosity (since strongest Reynolds stresses are not observed there in our configuration but in the bulk along the shear-layer) but also generates locally strong distorted  $\nu_t$  fields resulting in large gradients of the velocity field there as well, reinforcing this trend. This ends up in a local minimum, for which the cost functional is close to the initial condition. A possible remedy for this would be to “penalize” those sharp gradients (with a term  $\int_{\Omega} |\nabla \nu_t|^2$  in the objective function or with a gradient-smoothing techniques), pushing the eddy-viscosity more in the bulk. However, this is exactly what the  $\tilde{f}_{\tilde{\mathbf{v}}}$  model does, since, even if the gradient  $\nabla_{\tilde{f}_{\tilde{\mathbf{v}}}} J$  is also strong at the separation [as suggested, for example, by the  $\tilde{f}_{\tilde{\mathbf{v}}}$ -plots in Fig. 6(d)], the  $\nu_t$  field produced by that force is advected/diffused by the Spalart-Allmaras model equation, making it much more physical and therefore inducing a better decrease of the cost-functional.

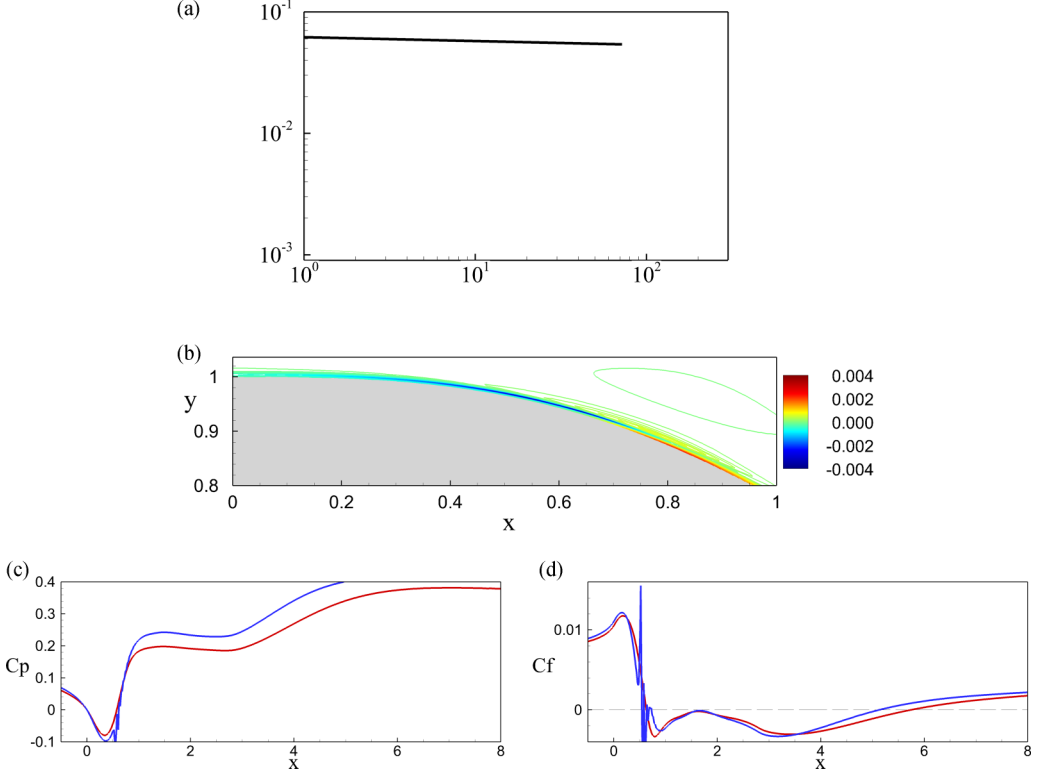


FIG. 14. Optimization of  $\nu_t$  with dense measurements for the *reference eddy-viscosity profile*: (a) Convergence history of the error  $e_\Omega$ , (b) optimized  $\nu_t^{\text{opt}}$  field, represented as a relative variation with respect to the uncorrected RANS-SA one  $\nu_t^{\text{SA}}$ ,  $(\nu_t^{\text{opt}} - \nu_t^{\text{SA}})/\|\nu_t^{\text{SA}}\|_\infty$  and pressure (c) and (d) wall-friction coefficients for optimized solution (blue) and uncorrected RANS-SA one (red) for comparison. These results should be compared to those concerning the  $\tilde{f}_v$  optimization given in the right panels of Fig. 6.

In Fig. 14, we have presented the results of the dense assimilation using  $\nu_t$ , with the reference eddy-viscosity profile at the entrance. The initial condition was provided by the  $\nu_t$  field of the uncorrected RANS-SA model (to avoid the issues with the “laminar” initial condition). The error  $e_\Omega$  barely decreases during the optimization, from  $e_\Omega = 0.060$  for the initial condition to a plateau of  $e_\Omega = 0.054$  at the end [for the  $\tilde{f}_v$ , the final error went down to  $e_\Omega = 0.02$ , see Fig. 6(b)]. Moreover, we can see in Fig. 6(b) that the variation of the eddy-viscosity with respect to the initial condition is very small (the optimal solution remains very close to the uncorrected RANS-SA solution) and localized at the separation points, generating a distorted  $[u, v, p]$  solution, shown in terms of the wall-pressure and friction coefficients in Figs. 14(c) and 14(d).

#### APPENDIX D: INCOMPRESSIBILITY AND BFGS

We notice as well that, even if the gradient has been modified through the Hessian matrix, its divergence (in the case of the volume-force correction  $\tilde{\mathbf{f}}_u$ ) seems to be null. To be able to prove this, we need to show that the application of  $\mathcal{H}_{n+1}^{-1}$  onto any divergence-free vector  $\mathbf{z}$  (here, the gradient) is also divergence-free. To do so, we use induction, where we suppose it is true at some iteration  $n$ , i.e.,  $\mathcal{D}\mathcal{H}_n^{-1}\mathbf{z} = 0$  (true for  $n = 0$  since  $\mathcal{H}_0 = \mathcal{I}$ ), where  $\mathcal{D}$  is the discrete-version of the divergence operator. We notice that the linear transformation involving the Cholesky decomposition presented before does not change this argument since it can be incorporated in the definition of  $\mathcal{D}$ . If we

suppose that  $\mathcal{D}\mathbf{s}_n = 0$ , then we have that

$$\begin{aligned}\mathcal{DH}_{n+1}^{-1}\mathbf{z} &= \mathcal{D}\left(\mathcal{I} - \frac{\mathbf{s}_n\mathbf{y}_n^T}{\mathbf{y}_n^T\mathbf{s}_n}\right)\mathcal{H}_n^{-1}\left(\mathcal{I} - \frac{\mathbf{y}_n\mathbf{s}_n^T}{\mathbf{y}_n^T\mathbf{s}_n}\right)\mathbf{z} + \mathcal{D}\frac{\mathbf{s}_n\mathbf{s}_n^T}{\mathbf{y}_n^T\mathbf{s}_n}\mathbf{z} \\ &= \mathcal{DH}_n^{-1}\left(\mathcal{I} - \frac{\mathbf{y}_n\mathbf{s}_n^T}{\mathbf{y}_n^T\mathbf{s}_n}\right)\mathbf{z} = \mathcal{DH}_n^{-1}\mathbf{z} - \mathcal{DH}_n^{-1}\frac{\mathbf{y}_n\mathbf{s}_n^T}{\mathbf{y}_n^T\mathbf{s}_n}\mathbf{z} = 0,\end{aligned}\tag{D1}$$

which is zero since both terms on the last equation are applications of  $\mathcal{DH}_n^{-1}$  onto divergence-free vectors.

- 
- [1] F. X. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspect, [Tellus A](#) **38**, 97 (1986).
  - [2] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed. (Springer-Verlag, Berlin, 2009).
  - [3] A. C. Lorenc, Analysis methods for numerical weather prediction, [Quart. J. R. Met. Soc.](#) **112**, 1177 (1986).
  - [4] C. Liu, Q. Xiao, and B. Wang, An ensemble-based four-dimensional variational data assimilation scheme. part I: Technical formulation and preliminary test, [Mon. Weather Rev.](#) **136**, 3363 (2008).
  - [5] A. Gronskis, D. Heitz, and E. Mémin, Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation, [J. Comput. Phys.](#) **242**, 480 (2013).
  - [6] V. Mons, J. C. Chassaing, T. Gomez, and P. Sagaut, Reconstruction of unsteady viscous flows using data assimilation schemes, [J. Comput. Phys.](#) **316**, 255 (1986).
  - [7] D. Heitz and E. Mémin, Variational fluid flow measurements from image sequences: Synopsis and perspectives, [Exp. Fluids](#) **48**, 369 (2010).
  - [8] Y. Yang, C. Robinson, D. Heitz, and E. Mémin, Enhanced ensemble-based 4DVar scheme for data assimilation, [Comput. Fluids](#) **115**, 201 (2015).
  - [9] V. Mons, L. Margheri, J. C. Chassaing, and P. Sagaut, Data assimilation-based reconstruction of urban pollutant release characteristics, [J. Wind Eng. Ind. Aerod.](#) **169**, 232 (2017).
  - [10] V. Mons, J. C. Chassaing, T. Gomez, and P. Sagaut, Is isotropic turbulence decay governed by asymptotic behavior of large scales? An eddy-damped quasi-normal Markovian-based data assimilation study, [Phys. Fluids](#) **26**, 115105 (2014).
  - [11] D. P. G. Foures, N. Dovetta, D. Sipp, and P. J. Schmid, A data-assimilation method for Reynolds-averaged Navier-Stokes-driven mean of reconstruction, [J. Fluid Mech.](#) **759**, 404 (2014).
  - [12] S. Symon, N. Dovetta, B. J. McKeon, D. Sipp, and P. J. Schmid, Data assimilation of mean velocity from 2D piv measurements of flow over an idealized airfoil, [Exp. Fluids](#) **61**, 1 (2017).
  - [13] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, Ensemble Kalman methods for inverse problems, [Inverse Probl.](#) **29**, 045001 (2013).
  - [14] H. Kato and S. Obayashi, Approach for uncertainty of turbulence modeling based on data assimilation technique, [Comput. Fluids](#) **85**, 2 (2013).
  - [15] Z. Li, H. Zhang, S. C. C. Bailey, J. B. Hoagg, and A. Martin, A data-driven adaptive reynolds-averaged Navier-Stokes  $\kappa$ - $\omega$  model for turbulent flow, [J. Comput. Phys.](#) **345**, 111 (2017).
  - [16] K. Duraisamy, G. Iaccarino, and H. Xiao, Turbulence modeling in the age of data, [Annu. Rev. Fluid Mech.](#) **51**, 357 (2019).
  - [17] A. P. Singh and K. Duraisamy, Using field inversion to quantify functional errors in turbulence closures, [Phys. Fluids](#) **28**, 045110 (2016).
  - [18] E. J. Parish and K. Duraisamy, A paradigm for data-driven predictive modeling using field inversion and machine learning, [J. Comput. Phys.](#) **305**, 758 (2016).
  - [19] C. He, Y. Liu, L. Gan, and L. Lesshafft, Data assimilation and resolvent analysis of turbulent flow behind a wall-proximity rib, [Phys. Fluids](#) **31**, 025118 (2019).
  - [20] S. Symon, D. Sipp, and B. J. McKeon, A tale of two airfoils: Resolvent-based modeling of an oscillator versus an amplifier from an experimental mean, [J. Fluid Mech.](#) **881**, 51 (2019).

- [21] H. Kato, A. Yoshizawa, G. Ueno, and S. Obayashi, A data assimilation methodology for reconstructing turbulent flows around aircraft, *J. Comput. Phys.* **283**, 559 (2015).
- [22] P. R. Spalart and S. R. Allmaras, A one-equation turbulence model for aerodynamic flows, *La Rech. Aéronautique* **1**, 5 (1994).
- [23] J. Dandois, E. Garnier, and P. Sagaut, Numerical simulation of active separation control by a synthetic jet, *J. Fluid Mech.* **574**, 25 (2007).
- [24] T. S. Lund, Generation of turbulent inflow data for spatially developing boundary layer simulations, *J. Comput. Phys.* **140**, 233 (1998).
- [25] S. R. Allmaras, F. T. Johnson, and P. R. Spalart, Modifications and clarifications for the implementation of the Spalart-Allmaras turbulence model, in *Proceedings of the 7th International Conference on Computational Fluid Dynamics (ICCFD'12)*, Big Island and Hawaii, 9–13 July (2012).
- [26] C. Mettot, D. Sipp, and H. Bézard, Quasi-laminar stability and sensitivity analyses for turbulent flows: Prediction of low-frequency unsteadiness and passive control, *Phys. Fluids* **26**, 045112 (2014).
- [27] P. Houston, J. A. Mackenzie, E. Süli, and G. Warnecke, *A posteriori* error analysis for numerical approximations of Friedrichs systems, *Numer. Math.* **82**, 433 (1999).
- [28] J. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects* (Springer Science & Business Media, Berlin, 2006).
- [29] J. Nocedal and S. J. Wright, in *Numerical Optimization* (Springer-Verlag, New York, 1999), pp. 490–527.
- [30] F.-X. Le Dimet, I. M. Navon, and D. N. Daescu, Second-order information in data assimilation, *Mon. Weather Rev.* **130**, 629 (2002).
- [31] A. Crivellini, V. D’Alessandro, and F. Bassi, A Spalart-Allmaras turbulence model implementation in a discontinuous Galerkin solver for incompressible flow, *J. Comput. Phys.* **241**, 388 (2013).
- [32] F. Hecht, New development in freefem++, *J. Num. Math.* **20**, 251 (2012).
- [33] A. N. Brooks and T. J. R. Hughes, Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations, *Comput. Method. Appl. M.* **32**, 199 (1982).
- [34] L. P. Franca, S. L. Frey, and T. J. R. Hughes, Stabilized finite element methods: I. Application to the advective-diffusive model, *Comput. Method. Appl. M.* **95**, 253 (1992).
- [35] L. P. Franca and S. L. Frey, Stabilized finite element methods: II. The incompressible Navier-Stokes equations, *Comput. Method. Appl. M.* **99**, 209 (1992).
- [36] Y. Bao, D. Zhou, C. Huang, Q. Wu, and Xiang-qiao Chen, Numerical prediction of aerodynamic characteristics of prismatic cylinder by finite element method with Spalart-Allmaras turbulence model, *Comput. Struct.* **89**, 325 (2011).
- [37] S. Mittal, On the performance of high aspect ratio elements for incompressible flows, *Comput. Method. Appl. M.* **188**, 269 (2000).