eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

# Sensor placement for data assimilation of turbulence models using eigenspace perturbations

O. Bidar ✉ 🅾 ; S. R. Anderson 🅾 ; N. Qin 🅾

Check for updates

View Online     Export Citation     CrossMark

AIP Publishing

# Sensor placement for data assimilation of turbulence models using eigenspace perturbations

View Online    Export Citation    CrossMark

O. Bidar,[1,a] (iD) S. R. Anderson,[2] (iD) and N. Qin[3] (iD)

AFFILIATIONS

[1]Department of Automatic Control and Systems Engineering and Department of Mechanical Engineering, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom
[2]Department of Automatic Control and Systems Engineering, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom
[3]Department of Mechanical Engineering, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

[a]Author to whom correspondence should be addressed: obidar1@sheffield.ac.uk

## ABSTRACT

We present an approach to sensor placement for turbulent mean flow data assimilation in the context of Reynolds-averaged Navier–Stokes (RANS) simulations. It entails generating a spatial uncertainty map through the eigenspace perturbations (ESPs) of the baseline turbulence model (e.g., the $k - \omega$ shear stress transport model) to quantify the epistemic structural errors in the model. A novel greedy search algorithm is proposed to place sensors targeting regions of highest uncertainty in the spatial uncertainty map generated from ESP. The algorithm is computationally efficient (e.g., computational cost negligible compared to a RANS solution) and is both easy to implement and tune. It involves two hyper-parameters (a constraint to avoid sensor clustering and the number of sensors) which we investigate in-depth. A variational (adjoint-based) data assimilation approach is used for flow reconstruction. The proposed strategy was tested on three two-dimensional wall-bounded flows (Reynolds number ranging $5.6 \times 10^3$–$9.36 \times 10^5$) involving flow separation and reattachment. For the wall-mounted hump case, we found that data assimilation using 33 sensors with our proposed sensor placement algorithm reduced the average velocity prediction error by 60% vs 38% with a simple uniform placement of sensors. Furthermore, we found that we could achieve 61% error reduction using our algorithm with only three sensors. Notably, in all tested cases, the error reduction using our method for sensor placement was close in accuracy to the instances where the entire flow field data were used for flow reconstruction, which involved two to three orders of magnitude more data points than the placed sensors.

## I. INTRODUCTION

Turbulent flows are broadly investigated using physical experiments and/or computational fluid dynamics (CFD). In both approaches, the accuracy, level of fidelity, and associated costs play competing roles. For practical CFD analyses, there is a heavy reliance on the Reynolds-averaged Navier–Stokes (RANS) simulations due to their relative simplicity and low computational resources requirement. It comes at the cost of limited accuracy for complex flows primarily due to the need for modeling all turbulent scales.[1] In recent years, a large number of data assimilation techniques have emerged that rely on the synergy of RANS models, such as high-fidelity simulations or experimental measurements for uncertainty quantification

(e.g., Ref. 2), turbulent flow reconstruction (e.g., Ref. 3), and data-driven modeling and predictions (e.g., Refs. 4–8). In the context of data-driven RANS-based models, a number of frameworks have been shown to achieve considerable improvements using limited, experimentally measurable, data (e.g., Refs. 9–12). However, most applications, thus far, have been primarily dictated by the availability of existing data, usually generated for benchmarking and validation. The important task of sensor placement for generating the experimental data *a priori*—and desirably tailored for RANS-based data assimilation—remains less well-explored.

Sensor placement in fluid flows have been studied in the context of model reduction, for example, using proper orthogonal decomposition (POD),[13–15] resolvent analysis,[16] and deep learning.[17] It has also been

20 February 2024 15:08:21

explored for flow control, for example, Refs. 18 and 19. Manohar *et al.*[20] introduced a sparse sensor placement strategy using singular value decomposition and QR pivoting. Callaham *et al.*[21] proposed using sparse representation techniques to reconstruct flow fields given sparse measurements. Lu and Papadakis[22] reconstructed flow around a surface-mounted prism using a combination of POD and a data-driven estimator using sparse velocity and scalar measurements. Recently, Karnik *et al.*[23] introduced a constrained sensor placement strategy using a greedy algorithm for flow reconstruction in nuclear digital twins. High-dimensional fields were reconstructed based on sparse sensor data using reduced order models.

In the context of RANS-based turbulent flow reconstruction, broadly, two main data assimilation classes exist: the stochastically formulated ensemble-based and the variational or adjoint-based, derived from an optimal control approach.

The ensemble Kalman filter (EnKF)[24] is a popular ensemble-based data assimilation technique. Developed from the Kalman filter (KF), the EnKF technique uses ensemble realizations of the system state to estimate the error covariance. This solves the computationally expensive standard KF which needs to propagate and store a high-rank error covariance matrix, not suitable for high dimensional problems, such as turbulent flow reconstruction. Examples of EnKF applications include data assimilation of turbulent wall-bounded flows with wall pressure and skin friction data;[25] enhanced flow reconstruction with disparate data (combinations of sparse data for multiple engineering quantities of interest, such as surface pressure and streamwise velocities) with a regularized EnKF;[10] and a comparison of EnKF with other ensemble-based techniques[26] using different quantities. Ensemble-based techniques have the advantages of allowing for incorporating and propagating the statistics of the assimilated flow state and relatively easy implementation due to their non-intrusive nature.

The adjoint-based or variational data assimilation involves the solution of a gradient-based optimization problem that minimizes the error between the high-fidelity data and RANS output. The high dimensions of such an inverse problem require the use of the adjoint method for efficient derivative computations of the cost function. Examples of applications include reconstructing flow by adjusting the term corresponding to the divergence of the Reynolds stress tensor in the RANS momentum equation (e.g., Refs. 27, 28) and the so-called field inversion (FI) approach where the RANS turbulence model transport equation is modified by a scalar field that is optimized using the adjoint method (e.g., Refs. 12, 29, and 30). The latter has also been extended for data-driven predictive RANS models using machine learning (e.g., Refs. 9 and 31–33). Adjoint-based methods can, in principle, assimilate finer scales of turbulent flow with limited sample sizes compared to ensemble-based methods.[10,34] The adjoint method allows scalability and robustness in variational data assimilation, however, due to its intrusive nature involving the governing equations (i.e., RANS equations with turbulence model equations), it requires a time-consuming and laborious software development phase. In the context of turbulent mean flow data assimilation, we have addressed this issue through an open-source implementation.[35]

Sensor placement has been investigated in the context of both ensemble-based and variational RANS-based data assimilation. Recently, Deng *et al.*[36] proposed a deep neural network (DNN)-based strategy to obtain the spatial sensitivity of the velocity field with respect to perturbations to RANS model constants. The *a priori* sensitivity analysis requires running ensembles of flow simulations with modified RANS model constants (100 CFD calculations in Ref. 36) and using a feature importance layer in a deep neural network to map the velocity fields to the respective modified constants. The feature importance layer is effectively weights in the entire CFD mesh, and after DNN training, the sensors are placed at the locations with the highest weights. The EnKF-based data assimilation by calibrating model coefficients may not be sufficient to reconstruct turbulent flows, as RANS-based discrepancy is thought to be mainly due to the structural forms of the turbulence model equations (i.e., uncertainties due to the Boussinesq hypothesis), rather than model parameters.[2] In addition, it is not clear how the bounds for model constants should be set for generating the ensemble data for DNN training.

Mons *et al.*[37] investigated sensor placement in the context of variational data assimilation for an unsteady *laminar* flow past a rotationally oscillating cylinder. The proposed framework involves the maximization of the sensitivity of observations with respect to changes in initial and boundary conditions using a first-order adjoint approach. It is unclear how this method can be extended to turbulent flows, and if the initial and boundary conditions are effective metrics to account for discrepancies due to parametric, functional, and structural uncertainties in turbulence models. Recently, Mons *et al.*[38] also proposed a more comprehensive linear and non-linear sensor placement strategies for laminar mean-flow variational data assimilation. The data assimilation is based on inferring a forcing term that corresponds to the divergence of the Reynolds stress tensor in the RANS momentum equation (as in Ref. 27). The optimization problem was formulated in two contexts: linear and non-linear. In the linear approach, the sensor placement strategy involves identifying the forcings that results in the most important variations in the mean velocity field and placing sensors at the dominant locations to allow for accurate reconstruction. In order for this approach to be most effective, the analysis needs to be performed close to the true states of the flow—which is a limitation for *a priori* analyses. To overcome this issue, a non-linear approach based on the second-order adjoint method was proposed that involves the minimization of the condition number of the Hessian of the assimilated flow. While this approach was found to be considerably more effective compared to the linear approach, the Hessian evaluation is computationally expensive—especially when applied to high Reynolds number three-dimensional turbulent flows—and difficult to implement.

In this work, we propose an optimization-based approach to variational sensor placement that involves the following:

1. Initially, we generate a spatial sensitivity map of the flow by employing the eigenspace perturbation approach for epistemic structural uncertainty in a baseline turbulence model.[39] Unlike the examples presented from literature, here the aim is to directly address the main source of discrepancy—the epistemic structural errors in turbulence models[2]—and use it to tailor sensor placement for mean turbulent flow assimilation.
2. After the uncertainty map is generated, an optimization problem is solved where the regions of highest uncertainties are targeted. This process is decoupled from the RANS-based solution, unlike the variational sensor placement strategies discussed above. The advantages are a very low computational cost, an easy implementation, and algorithm tuning, compared to the adjoint-based analyses in Refs. 37 and 38.

20 February 2024 15:08:21

3. Once sensor locations have been identified, we then perform the variational data assimilation using the field inversion framework. This approach has the advantage of model-consistency,[4] and the aforementioned capacity to perform reasonably with limited data. In addition, unlike the method in Ref. 36, this approach is not limited to parametric adjustment of the baseline turbulence model and addresses the functional discrepancy.

The rest of the paper is structured as follows. In Sec. II, we formulate the sensor placement problem, introduce the $k - \omega$ SST baseline RANS model, the eigenspace perturbation approach, the sensor placement strategy, and the field inversion method. Results are presented and analyzed in Sec. III, with the entire framework and analysis consolidated in Sec. IV, and final conclusions are drawn in Sec. V.

## II. METHODS

We will begin by motivating the approach to sensor placement, with a general overview of our proposed strategy, followed by the details of components involved in the framework, described sequentially.

### A. Problem formulation and framework overview

Very broadly, the task of sensor placement can be posed as an optimization problem as follows:

$$\max_{\mathbf{x}} f[\mathbb{S}(\mathbf{x})], \quad \text{or} \quad \min_{\mathbf{x}} \quad f[\mathbb{S}(\mathbf{x})]^{-1},$$
$$\text{subject to} \quad \mathscr{C}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ represents the desired sensor locations, which we will also denote $\mathscr{S}$ henceforth, $\mathbb{S}$ is an operator for the desired sensor locations, and $\mathscr{C}$ represents any/all the constraints (e.g., number of sensors, regions of interest, minimum distance between sensors, other physics-based constraints including the governing equations of the system, etc.).

Short of a brute-force exhaustive search, the so-called "generate and test" approach, which will have huge time and cost implications, the goal is to have an *a priori* sensor placement approach, i.e., placing sensors by approximating regions of uncertainty based on physical and/or expert knowledge before undertaking the experiment(s). The function $f[\mathbb{S}(\mathbf{x})]$ can be based on the determinant of the Fisher information matrix (e.g., Ref. 40) some measure from a deep neural network (e.g., Refs. 17 and 36) or any other appropriate cost function (e.g., Refs. 23, 37, and 38). We stress that the goal is for $f[\mathbb{S}(\mathbf{x})]$ in Eq. (1) to effectively capture the uncertainty in the underlying system and target spatial sensors in the regions of high uncertainty, some of which have been highlighted already in Sec. I.

Figure 1 illustrates the proposed framework. Starting with a linear-eddy viscosity RANS closure, we perturb the Reynolds stress tensor to their extremal states to generate an uncertainty map, a surrogate for the operator $\mathbb{S}(\mathbf{x})$ in Eq. (1). Specifically, five CFD simulations are run sequentially: two simulations perturbing the eigenvectors and three simulations perturbing the eigenvalues. All five simulations result in as many realizations of the flow prediction, allowing us to generate uncertainty maps—based on the variances—for various quantities of interest, e.g., a variance map of the streamwise velocity, with a value for each mesh cell, etc. Then, for a prescribed number of sensors and for a given quantity of interest, an optimization problem is solved by a
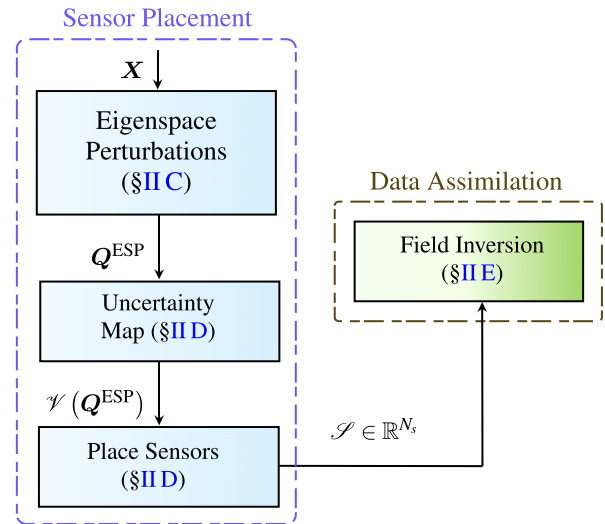


**FIG. 1.** Overview of the proposed framework for sensor placement and data assimilation. The sensor placement and data assimilation problems are decoupled. $\boldsymbol{X}$: spatial coordinates of flow domain; $\boldsymbol{Q}$: flow variables such as velocity and pressure; $\mathscr{V}$ variance of any given quantity $(q)$ following the five eigenspace perturbations; and $\mathscr{S}$: coordinates of optimized sensor locations. The relevant sections (§) of the paper are highlighted in each block.

greedy search to ensure sensors are placed in regions of flow field with highest uncertainty. After identifying these locations, the turbulent flow is reconstructed through an adjoint-based data assimilation approach.

### B. Underlying RANS model

In RANS-based analyses, derived from Reynolds (and time-averaging) of the Navier–Stokes equations, all turbulence scales—represented by the Reynolds stress tensor, $\overline{u_i u_j}$—are modeled.[41] The most commonly used model are based on the Boussinesq approximation, and the Reynolds stress tensor can be approximated as a linear function of the mean rate of strain, $S_{ij}$,

$$\rho \overline{u_i u_j} = \frac{2}{3} \rho k \delta_{ij} - \mu_t \left( S_{ij} - \frac{2}{3} \frac{\partial U_k}{\partial x_k \delta_{ij}} \right), \tag{2}$$

where $\rho$, $k$, $\delta_{ij}$, and $\mu_t$ are the density, turbulent kinetic energy, the Kronecker delta function, and the turbulent viscosity respectively.

A commonly employed RANS model is the two-equation $k - \omega$ shear stress transport (SST) model.[42] It blends the $k - \omega$ model in regions close to walls, and $k - \varepsilon$ in the farfield. The transport equations of the turbulent kinetic energy $k$ and dissipation rate $\omega$ are

$$\frac{Dpk}{Dt} = \rho P - \beta^\star \rho k \omega + \frac{\partial}{\partial x_j} \left[ (\mu + \sigma_k \mu_t) \frac{\partial k}{\partial x_j} \right], \tag{3}$$

$$\frac{D\rho\omega}{Dt} = \frac{\gamma}{\nu_t} \rho P - \beta \rho \omega^2 + \frac{\partial}{\partial x_j} \left[ (\mu + \sigma_\omega \mu_t) \frac{\partial \omega}{\partial x_j} \right]$$
$$+ 2\rho (1 - F_1) \sigma_{\omega 2} \frac{1}{\omega} \frac{\partial k}{\partial x_j} \frac{\partial \omega}{\partial x_j}, \tag{4}$$

where $P$ is the production of turbulent kinetic energy, $\mu$ is the dynamic viscosity, $\beta^{\star}$, $\sigma_k$, $\gamma$, $\beta$, $\sigma_\omega$, $\sigma_{\omega 2}$ are model constants, and $F_1$ is a blending function.

The turbulent viscosity is calculated as

$$\mu_t = \rho \frac{a_1 k}{\max(a_1 \omega, SF_2)}, \tag{5}$$

where $a_1$ is a model constant, $S$ is the magnitude of the shear stress tensor, and $F_2$ is another blending function.

### C. Eigenspace perturbations

The eigen-decomposition of the Reynolds stress tensor in Eq. (2) results in

$$\overline{u_i u_j} = 2k \left( \frac{\delta_{ij}}{3} + \nu_{in} \Lambda_{nl} \nu_{lj} \right), \tag{6}$$

where $\nu_{in}$ is the orthornormal eigenvectors matrix and $\Lambda_{nl}$ is a diagonal matrix composed of eigenvalues, $\lambda_l$. The eigen-decomposition is useful because now the shape and orientation of the Reynolds stress tensor are directly represented by the eigenvalues and eigenvectors, respectively.

The eigenvalues can be represented in the Barycentric map,[43] shown in Fig. 2. All realizable turbulent states are encompassed by the equilateral triangle, with the vertices representing the limiting states. Any given turbulence state $x$ can be represented as linear functions of the eigenvalues, $\lambda_l$ and the vertices of the Barycentric map, as follows:[43]

$$x = x_{1c}(\lambda_1 - \lambda_2) + x_{2c}(2\lambda_2 - 2\lambda_3) + x_{3c}(3\lambda_3 + 1), \tag{7}$$

where the three vertices $x_{1c}$, $x_{2c}$, and $x_{3c}$ represent one-component, two-component (axisymmetric), and three-component (isotropic) turbulence.

Uncertainty can now be introduced in the Reynolds stresses by perturbing the eigenvalues to their limiting states, i.e., the three vertices in the Barycentric map. The perturbed states $x^*$ can be defined as[39,44,45]

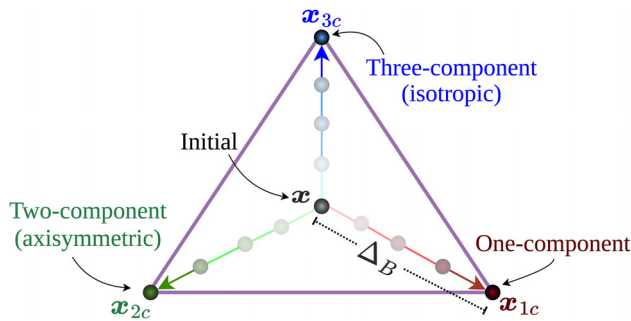$$x^* = x + \Delta_B(x^{(t)} - x), \tag{8}$$



**FIG. 2.** Barycentric triangle used to perturb eigenvalues to their three limits. Mapping turbulence states on the Barycentric map is useful as Banerjee et al.[43] showed that all realizable turbulence states are bounded and presented by the edges and vertices of the equilateral triangle. In eigenspace perturbations an arbitrary Reynolds stress state $x$ is *sequentially* perturbed from the initial state to the three vertices $x_{1c}$, $x_{2c}$, and $x_{3c}$.

where $x$ is the arbitrary initial state, $\Delta_B \in [0, 1]$ is the relative distance between $x$, and $x^{(t)}$ is the target vertex. Thus, the perturbed eigenvalues, $\lambda_l^*$ can be calculated as

$$\lambda_l^* = B^{-1} x^*, \tag{9a}$$

$$= (1 - \Delta_B) B^{-1} x + \Delta_B B^{-1} x^{(t)}, \tag{9b}$$

$$= (1 - \Delta_B) \lambda_l + \Delta_B B^{-1} x^{(t)}, \tag{9c}$$

where $B$ is a linear mapping with $B^{-1} x_{1c} = (2/3, -1/3, -1/3)^T$, $B^{-1} x_{2c} = (1/6, 1/6, -1/3)^T$, and $B^{-1} x_{3c} = (0, 0, 0)^T$.

The eigenvector perturbations modulate the turbulent kinetic energy production, $\mathscr{P}_k = -\overline{u_i u_j}(\partial U_i / \partial x_j)$, which represents the transfer of kinetic energy from the mean flow to the fluctuating velocity field.[39] This corresponds to varying the alignment of the Reynolds stress ellipsoid. Mathematically, this modulation is achieved by varying the Frobenius inner product $\langle A, R \rangle_F = \text{tr}(AR)$, with $A$ being representing the mean velocity gradient, and $R$ being the Reynolds stress tensor. As with the eigenvalue perturbations, the extremal states are sought by considering the bounding values of the inner product to consider all permissible dynamics. Iaccarino et al.[39] show that the bounds on the inner product are

$$\langle A, R \rangle_F \in [\lambda_1 \gamma_3 + \lambda_2 \gamma_2 + \lambda_3 \gamma_1, \lambda_1 \gamma_1 + \lambda_2 \gamma_2 + \lambda_3 \gamma_3], \tag{10}$$

with $\lambda_1 \geq \lambda_2 \geq \lambda_3$ representing the eigenvalues of the symmetric components of the mean velocity gradient $A$, which is also the strain rate tensor. Thus, in the coordinate system defined by the eigenvectors of the strain rate tensor, the bounds for the Reynolds stress alignments of the Reynolds stress eigenvectors are

$$\nu_{\min} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \nu_{\max} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The eigenspace perturbation (ESP) implementation of Mishra et al.[45] in the SU2 CFD suite is used in this work. The implementation in Ref. 45 has been tested on both two- and three-dimensional flows.

### D. Sensor placement

The uncertainty map generated as described in Sec. II C can now be used as a surrogate for the $f[\mathbb{S}(\mathbf{x})]$ operator in Eq. (1). One approach of formulating the optimization problem for sensor placement, with a prescribed number of sensors, $N_s$, can be

$$\max_{\mathbf{x}} \quad \mathscr{J} = \frac{1}{\|\mathscr{J}_0\|} \sum_{i=1}^{N_s} \mathscr{V}^q(\mathbf{x}_i), \quad \text{subject to} \quad d(\mathbf{x}) \geq d_{\min}, \tag{11}$$

where $\mathscr{V}^q$ represents the variance-based uncertainty map from eigenspace perturbation for a quantity $q$, $\mathbf{x} \in \chi$ represents the sensor coordinates, where $\chi \subseteq X$ represents the space of potential sensor locations, and for $N_s > 1$, $d(\mathbf{x})$ represents a minimum distance constraints between any two sensors, prescribed by the user (based, for example, on operational parameters/constraints, etc.). Additionally, the distance constraint is essential for the problem to be well-posed in order to avoid sensor clustering.

One approach to solving the optimization problem in Eq. (11) can be using global search methods, such as genetic algorithms, with many off-the-shelf implementations to choose from. However, this

can be a computationally challenging optimization problem due to the very large design space involved in turbulent flows, i.e., the computational budget required in high Reynolds number, three-dimensional flows of industrial interest is not anticipated to scale well. Given this, we propose a more computationally efficient solution based on a greedy search, which uses the heuristic of making the locally optimal decision at each choice of sensor location, i.e., the sensor is placed at the next permissible location of maximal uncertainty in each iteration, summarized in Fig. 3 and Algorithm 1.

The sensor placement procedure is as follows:

1. Given a prescribed number of sensors, $N_s$, the potential sensor sites $\chi \subseteq X$, uncertainty map for a given quantity, $\mathcal{V}^q$, place the first sensor in the cell with the highest variance, $\mathcal{V}^q$, and store the spatial coordinates, $\mathcal{S}_n$. An initial radius, $r_{\text{initial}}$, must also be prescribed for defining the circular (in two-dimensional cases)— or spherical in three-dimensional flows—exclusion zone, $\Omega_n$ with $n = 1$, to avoid subsequent sensors clustering around the sensor placed previously. This is an important hyper-parameter which we will return to shortly.
2. Exclude the cells in the domain $\Omega_n$, for $n \geq 1$, and update $\chi$ and $\mathcal{V}^q$ accordingly.
3. Place the next sensor at the cell with highest variance (based on the updated $\mathcal{V}^q$ and $\chi$), and store the coordinates $\mathcal{S}_n$.
4. Select the radius, $r_n$ for $n > 1$, for the exclusion domain $\Omega_n$ using as follows:

$$r_n = \left[1 + \left(\mathcal{V}^q_{n-1} - \mathcal{V}^q_n\right)\right] r_{n-1}. \tag{12}$$

Using the relation in Eq. (12), we linearly increase the size of the exclusion domain based on the change in variance between subsequent sensors.

5. Repeat steps 2–4 until all sensors are placed.

There are two hyper-parameters in the proposed greedy algorithm. The first one is the initial radius for the circular exclusion zone. This determines how far the sensors are placed. For a given number of sensors, we propose an iterative approach to guide this selection, using the following metric $\mathcal{M}_1$,
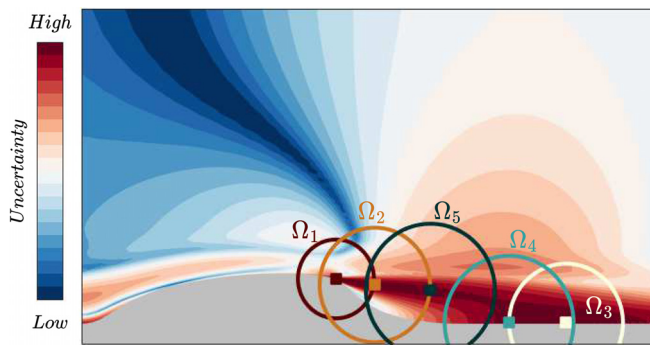


**FIG. 3.** Sensor placement visualization using the proposed greedy algorithm, in Algorithm 1. The contour represents the uncertainty in the streamwise velocity predictions from eigenspace perturbations, $\mathcal{V}(U_x)$. The number of sensors $N_s = 5$ in this example, placed sequentially. The sensors are represented by the square markers, and the circles labeled $\Omega_n$ represent the exclusion domain defined by radius $r_n$, where the radius is calculated using the expression in Eq. (12) for $n > 1$, and the initial radius $r_1$ is selected iteratively using $\mathcal{M}_1$ in Eq. (13).

$$\mathcal{M}_1 = \frac{\sum\limits_{i}^{N_\chi} \mathcal{V}^q_i}{\sum\limits_{n}^{N_s} \sum\limits_{m}^{\Omega_n} \mathcal{V}^q(\mathbf{x}_m)}, \tag{13}$$

where we iteratively vary the initial radius, illustrated in Fig. 4. The numerator is the sum of variances, $\mathcal{V}^q$, over the entire potential sensor sites, $\chi$, and the denominator is the sum of variances over the excluded circular domains with the sensors locations at the centers (refer to illustrative exclusion domains, $\Omega_n$, in Fig. 3). The initial radius can be defined as a proportion of some characteristic length of the case under consideration. As it increases, $\mathcal{M}_1$ in Eq. (13) approaches unity. Thus, we can select the initial radius iteratively, using some relative threshold (i.e., with respect to unity). This will be practically demonstrated and assessed numerically in Sec. III A. In addition, this approach of selecting the initial radius parameter is a practical one, given the low computational cost of the entire sensor placement algorithm.

**Algorithm 1.** Greedy algorithm for sensor placement.

---

**Inputs**: $N_s$ = number of sensors; $\chi \subseteq X = x, y, z$ Cartesian coordinates of mesh cell centers; $\mathcal{V}^q$ = eigenspace perturbation-based variance of quantity to be measured; $r_{\text{initial}}$ = initial radius for exclusion domain.

**Output**: $\mathcal{S}$ = sensor locations ($x, y, z$ coordinates)

1: **procedure** PLACESENSORS($N_s, \chi, \mathcal{V}^q, r_{\text{initial}}$)
2: $\quad r \leftarrow r_{\text{initial}}$　　　　　　　▷ radius for exclusion domain
3: $\quad$ **for all** $n \in 1, \ldots, N_s$ **do**
4: $\quad\quad i = \max(\mathcal{V}^q)$　　▷ $i$ = index of cell with max. variance
5: $\quad\quad \mathcal{S}[n, :] = \chi[i, :]$　　▷ save $x, y, z$ coordinates of sensor $n$
6: $\quad\quad \hat{\chi}, \hat{\mathcal{V}} = []$　　　　　▷ storage for non-excluded regions
7: $\quad\quad M \leftarrow \text{size}(\chi)$
8: $\quad\quad$ **for all** $m \in 1, \ldots, M$ **do**
9: $\quad\quad\quad x, y, z \leftarrow \chi[m, :]$
10: $\quad\quad$ **if** $x, y, z \notin \Omega_n$ **then**　　　▷ only keep cells outside $\Omega_n$
11: $\quad\quad\quad \hat{\chi} \leftarrow \text{concatenate}(x, y, z)$
12: $\quad\quad\quad \hat{\mathcal{V}} \leftarrow \text{concatenate}(\mathcal{V}^q[m])$
13: $\quad\quad j \leftarrow \max(\hat{\mathcal{V}})$　　▷ $j$ = index of cell with max. variance
14: $\quad\quad r \leftarrow [1 + (\mathcal{V}^q[i] - \hat{\mathcal{V}}[j])] * r$　　▷ update radius for $\Omega_{n+1}$
15: $\quad\quad \chi \leftarrow \hat{\chi}$
16: $\quad\quad \mathcal{V}^q \leftarrow \hat{\mathcal{V}}$
17: **return** $\mathcal{S}$

---

The number of sensors can be considered as the second hyper-parameter. We note that based on the experimental approach, this can be dictated by other constraints (e.g., in the instance where only a discrete given set of sensors are available). Nonetheless, as an approximate guide to how many sensors may be considered, we propose the following metric, $\mathcal{M}_2$,

$$\mathcal{M}_2 = \left(\sum\limits_{n}^{N_s} \mathcal{V}^q_n\right)^{-1}, \tag{14}$$
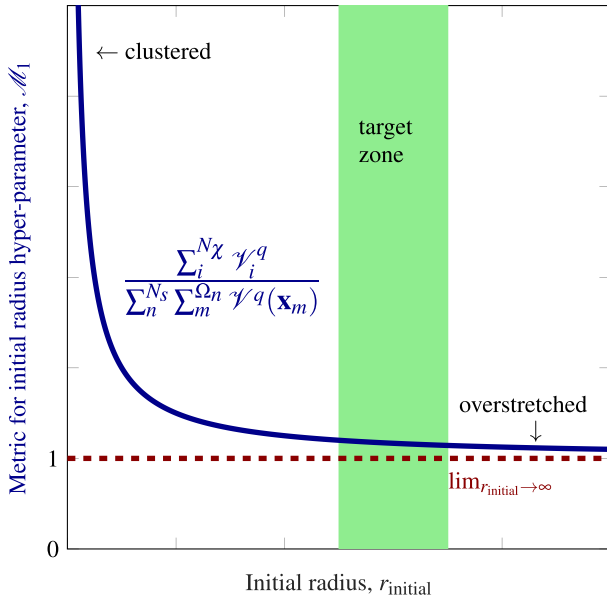
**FIG. 4.** Illustrative plot of the metric in Eq. (13) for guiding the selection of the initial radius hyper-parameter. The initial radius for the exclusion domain, required by the greedy sensor placement algorithm, can be iteratively selected, i.e., any point in the "target zone" when $\mathcal{M}_1$ levels off.

which is the sum of variances over the sensor locations, $\mathcal{S}$. As the number of sensors increase $N_s$ will level off, as illustrated in Fig. 5, with detailed numerical analysis in the results section.

### E. Data assimilation using field inversion

We employ the adjoint-based field inversion (FI) approach for data assimilation. This is essentially an inverse problem, where the transport equation for an existing turbulence model is modified by introducing a multiplicative spatially varying scalar field ($\beta^{\text{FI}}$) to the production term.[9,29–31,33] The scalar field is iteratively tuned (by solving an optimization problem) such that the error between RANS-based flow predictions and high-fidelity data are minimized. For the $k - \omega$ SST, the equation for the turbulence dissipation rate [Eq. (4)], shown in a generalized form below, is thus modified as[46]

$$\frac{D\omega}{Dt} = \beta^{\text{FI}}(\boldsymbol{X})\mathcal{P}_\omega(\boldsymbol{Q}) + \mathcal{T}_\omega(\boldsymbol{Q}) - \mathcal{D}_\omega(\boldsymbol{Q}), \qquad (15)$$

where $\beta^{\text{FI}}(\boldsymbol{X}) \in \mathbb{R}^{N_\beta}$ is the spatial scalar field with $N_\beta$ equivalent to the number of mesh cells, $\boldsymbol{Q}$ represents the Reynolds-average conserved flow variables, $\mathcal{P}_\omega$, $\mathcal{T}_\omega$, and $\mathcal{D}_\omega$, represent the production, transport, and destruction terms in turbulence dissipation transport equation. The original/baseline model is recovered with $\beta^{\text{FI}} = 1$.

The inverse problem is the formulated as finding the optimum discrepancy field, $\beta^{\text{FI}}(\boldsymbol{X})$, by minimizing the following objective function:

$$\min_{\beta^{\text{FI}}} \quad \mathcal{L} = ||q_i^{\text{RANS}}(\beta^{\text{FI}}) - q_i^{\text{data}}||_2^2 + \lambda||\beta^{\text{FI}} - 1||_2^2, \qquad (16)$$

which reduces the functional errors in the baseline turbulence model, where $q_i^{\text{data}} \in \mathscr{D}$ represents high-fidelity physical quantities of size $N_s$,
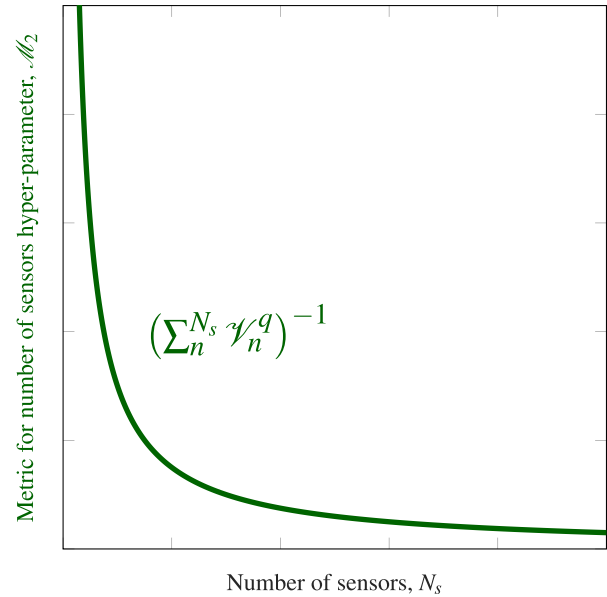


**FIG. 5.** Illustrative plot of the metric in Eq. (14) for estimating the number of sensors. An appropriate number of sensors can be approximated *a priori*, iteratively, as the points on the curve when the metric $\mathcal{M}_2$ levels off.

$q_i^{\text{RANS}}$ represents the equivalent quantity predicted by the turbulence model, and $||.||_2$ is the $L_2$ norm. The second term in Eq. (16) tuned by the relaxation parameter $\lambda$ is to regularize the problem by avoiding huge deviations from the baseline turbulence model and avoiding an ill-posed optimization problem.
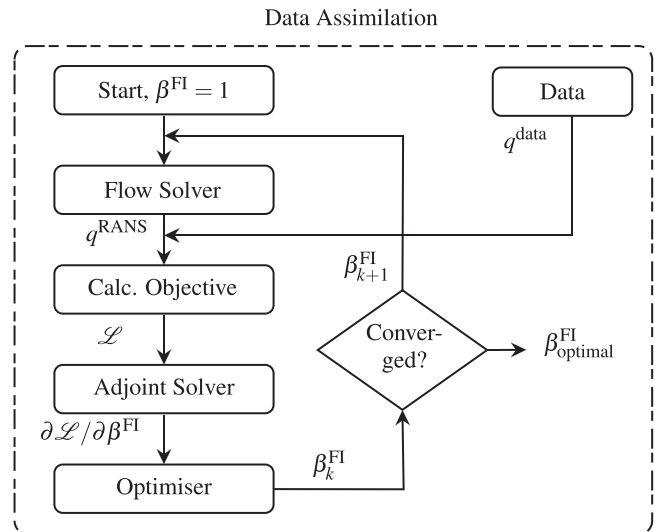


**FIG. 6.** Flow diagram for the iterative adjoint-based method for field inversion. The process is labeled as data assimilation in Fig. 1. The corrective scalar-field $\beta^{\text{FI}}$ perturbs the turbulence transport equation [the production term in Eq. (4)] and is optimized such that the error between high-fidelity data and the baseline RANS predictions are minimized.
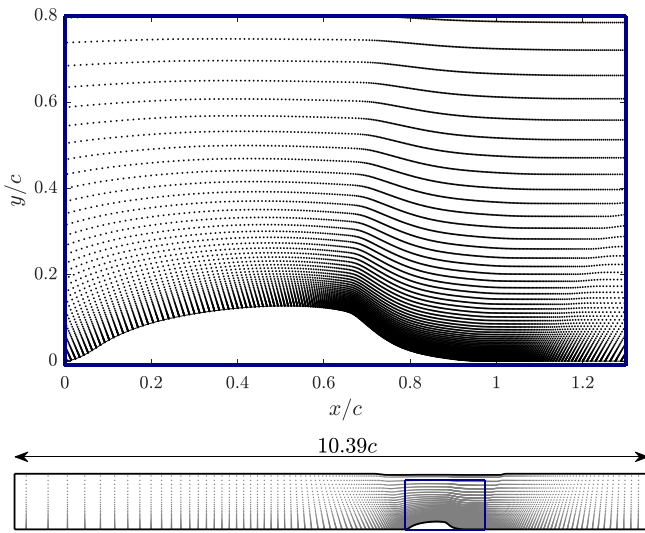
**FIG. 9.** The uncertainty map in streamwise velocity for the NASA hump case.

**FIG. 7.** 2D NASA smooth hump, with the dots marking cell-centers, **X**, in the mesh used.[47] The blue bounding box represent the region of potential sensor sites, $\chi \subseteq X$.

Due to the high-dimensional nature of the optimization problem, it is essential to use the adjoint-method for efficient gradient calculations, details provided in Ref. 35, with the general process summarized in the flow diagram in Fig. 6.

## III. RESULTS AND ANALYSIS

In this section, we apply the proposed framework to three fully-turbulent wall-bounded benchmark flows, all involving challenging flow physics (e.g., separation and reattachment) for RANS-based simulations. In all the cases, we use the streamwise velocity as the quantity from high-fidelity data, since it is one of the most commonly measured (in terms of volume data, i.e., measured quantities in the flow domain). All flows are simulated as two-dimensional, steady, and incompressible.

### A. 2D NASA wall-mounted hump

The flow over the well-established 2D wall-mounted hump, Fig. 7, part of the NASA Turbulence Modeling Resource database,[47] is selected
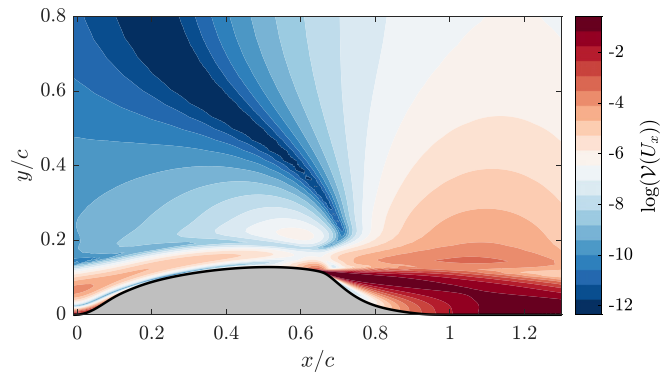
as the first test case. It involves flow separation as a result of adverse pressure gradients over the smooth hump surface. The flow is commonly used as a benchmark case for verifying and validating turbulence models since most linear eddy viscosity models perform poorly in predicting the separation, reattachment, and boundary recovery by over-predicting the size of the separation bubble due to under-predicted turbulent shear stress in the separation region. The chord-based Reynolds number is $9.36 \times 10^5$ with a Mach number of 0.1.

We use wall-resolved large eddy simulation (LES) data by Uzun et al.[48] as a surrogate for experimental data. These results have been validated against experimental data and are preferred over limited publicly available experimental results as it allows for benchmarking data assimilation with sparse experimental data against the scenario when extensive data are available, such as those generated in particle image velocimetry (PIV) experiments.

The baseline $k - \omega$ SST, LES, and eigenspace perturbation streamwise velocity profiles are shown in Fig. 8. The SST predictions are most erroneous near the wall, in the separated shear layer, due to a very high flow reversal aft of the hump apex, resulting in an over-predicted circulation zone. The eigenspace perturbation results show a high degree of variability compared to the SST predictions in the near-wall region. Increasing the turbulent production mechanism (e.g., $\boldsymbol{x}^{(t)} = \boldsymbol{x}_{1c}, \nu = \nu_{\min}$) reduce the flow separation, thus shifting the results closer to the reference data, while damping the production
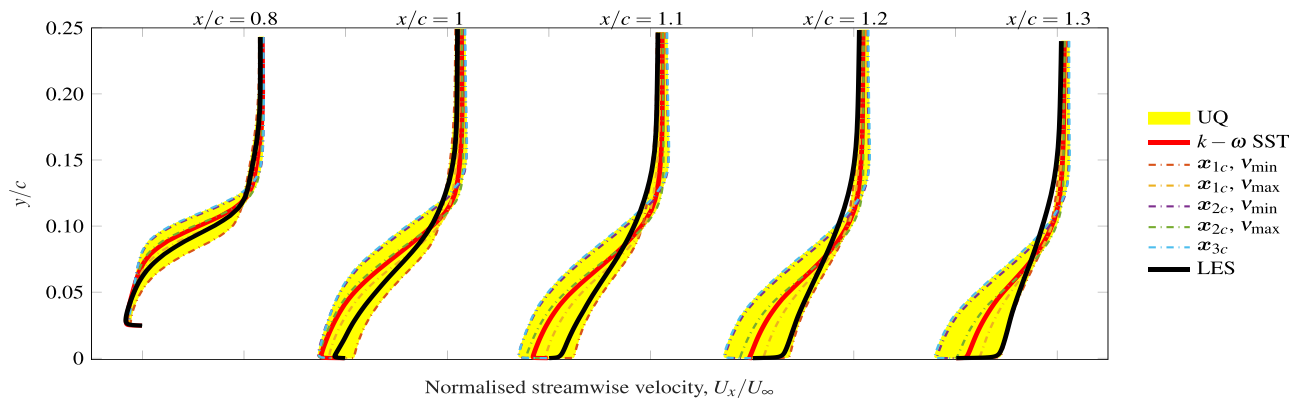
**FIG. 8.** Streamwise velocity profiles with uncertainty bounds based on eigenspace perturbations. The dashed-dotted lines represent the five ESP scenarios.
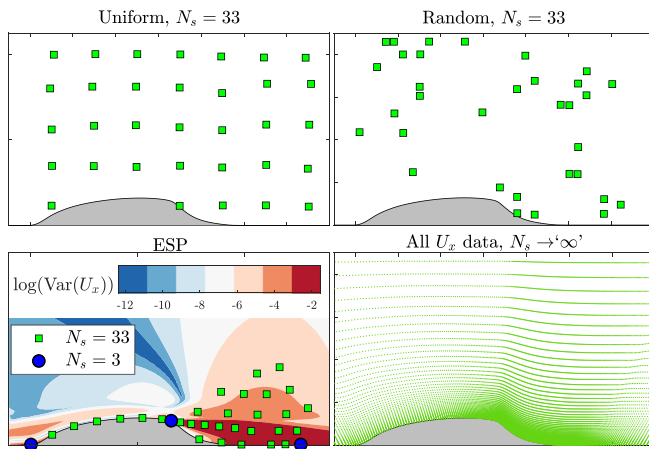
**FIG. 10.** Various sensor placement scenarios used for data assimilation. ESP refers to the proposed placement algorithm, and the contour lines represent the uncertainty map in Fig. 9.

(e.g., $x^{(t)} = x_{3c}$) lead to higher separation. The ESP results essentially subsume the LES references data, especially near the wall, in the separated shear layer, and becomes negligible as we approach the free-stream (as expected).

We can now map the uncertainty for any given physical quantity, $q \subseteq Q$ (in the case of vector quantities, we can look at individual scalar components, as at the present work, or some normalized sum of different components) as a function of the variance based on eigenspace perturbation results. As the reference data will be the streamwise velocity, $U_x$, Fig. 9 shows the uncertainty map, normalized by the maximum variance. We use a logarithmic scale to better illustrate the degree of uncertainty in the different regions of the flow domain. The high regions of uncertainty are in the near the hump, aft of hump apex, increasing to the highest uncertainty in the separated shear layer—overall, qualitatively, bearing engineering scrutiny.

Sensors are now placed using the uncertainty map, and our proposed heuristic greedy algorithm. In order to benchmark the framework, we will present results for the following scenarios, shown in Fig. 10: (a) uniformly spaced, manually placed sensors, ending up with
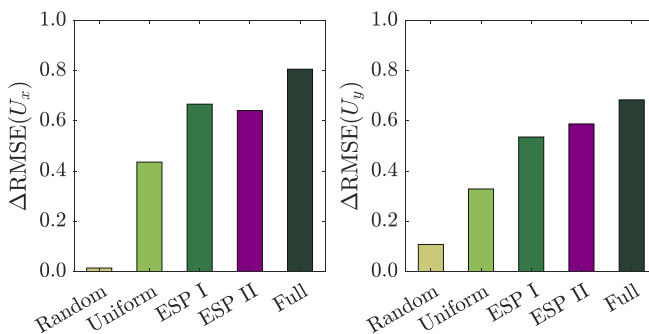
$N_s = 33$ over the entire potential sensor sites, $\chi \subseteq X$, fixed for the subsequent two scenarios, (b) randomly placed sensors, (c) sensor placed using our algorithm with $N_s = 33$, (d) using our algorithm but reducing the number of sensors by over one order of magnitude, i.e., $N_s = 3$, and (e) using the streamwise velocity over the entire potential sensor sites, $\chi \subseteq X$.

We compare the turbulent mean flow reconstruction results for the stated scenarios in Fig. 11, in terms of the change in root-mean-square error in the baseline SST predictions of the velocity components. As anticipated, the random sensor placement case is the least effective (negligible improvement in the streamwise velocity despite the use of $U_x$ data in field inversion), highlighting the need for informed sensor placement. The uniform sensor placement case reduces the average error in the velocity components by 38%. For the same number of sensors as the previous two discussed scenarios (i.e., $N_s = 33$), with our proposed sensor placement algorithm, the average errors reduce by 60% (labeled "ESP I" in Fig. 11), compared to a 74% reduction when using all the streamwise velocity data (labeled "Full" in Fig. 11). The case with all the data shows that using very large datasets does not necessarily results in proportionally similar data assimilation error reduction—also observed in the other two cases in Secs. III B and III C. Notably, the results for this flow show that with our approach the same error reduction is achievable even if the number of sensors is reduced by over an order of magnitude, i.e., $N_s = 3$ (labeled "ESP II" in Fig. 11), with an average error reduction of 61%.

Next, we investigate the two hyper-parameters for the proposed sensor placement algorithm as discussed in Sec. II D: initial radius for the circular exclusion domain, and the number of sensors.

Figure 12 compares the initial radius against the metric $\mathcal{M}_1$ in Eq. (13) (proposed to guide selecting an appropriate $r_{\text{initial}}$ *a priori*)



**FIG. 11.** Comparing the root-mean-square error change in velocity predictions (left: streamwise component, $U_x$; right: wall-normal component, $U_y$) for various sensor placement scenarios shown in Fig. 10. ESP I and II refer to the cases with $N_s = 33$, and $N_s = 3$, respectively, where sensors are placed using our proposed algorithm.
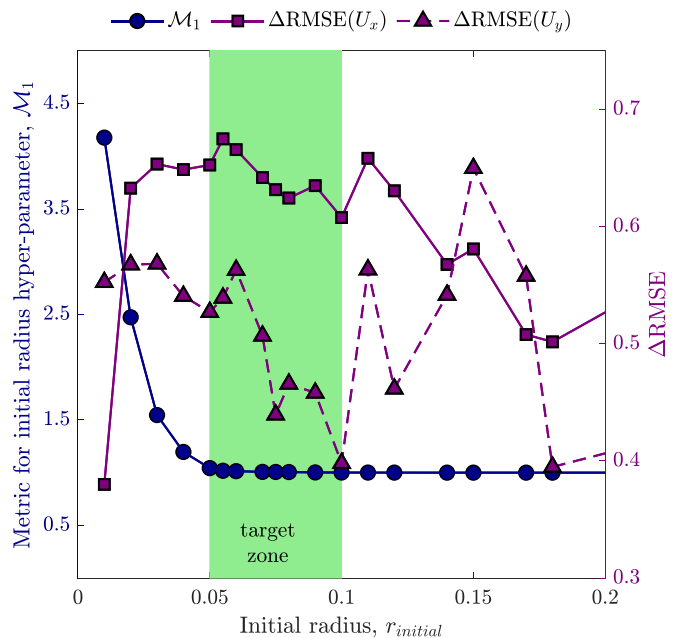


**FIG. 12.** The tuning metric $\mathcal{M}_1$ [Eq. (13)] on the left axis, and the effect of initial radius on flow reconstruction error reduction on the right axis. Number of sensors is set to ten.
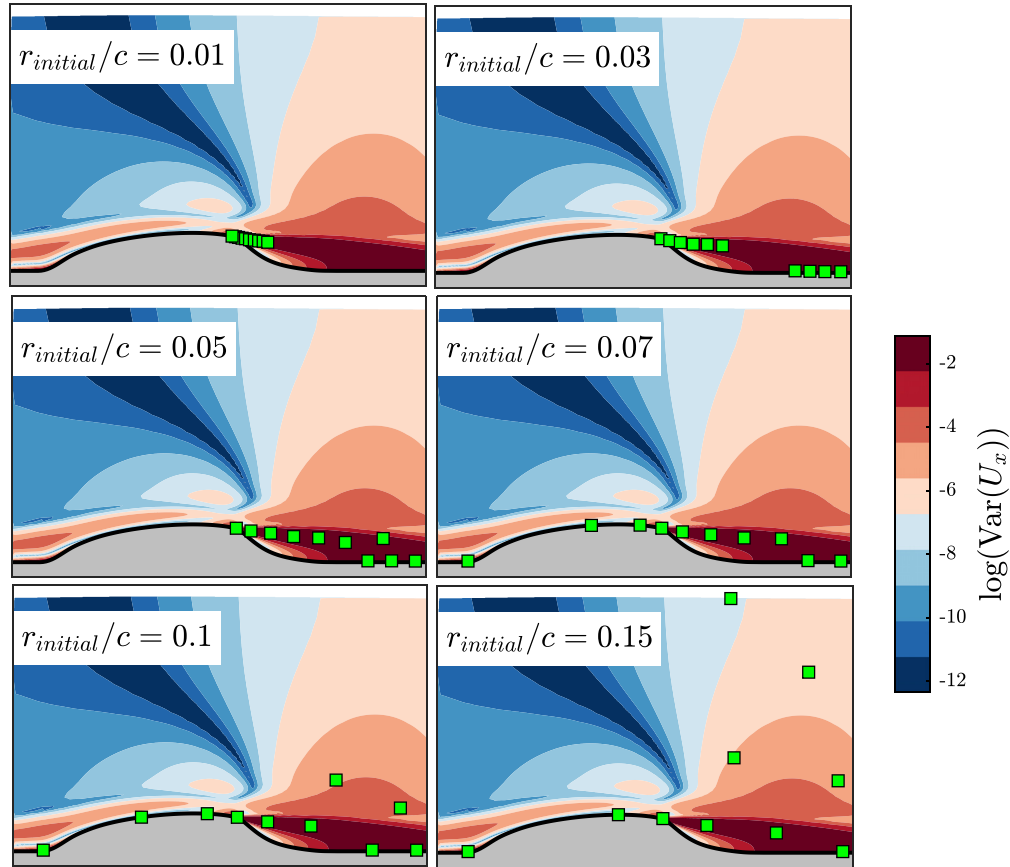
**FIG. 13.** Some illustrative sensor configurations, for investigating the effect of the initial radius on data assimilation. The number of sensors is fixed to $N_s = 10$.

and the subsequent error reduction through data assimilation, with $0.01 \leq r_{initial}/c \leq 0.20$, where $c$ is chord length for the hump (Fig. 7). The number of sensors is fixed to $N_s = 10$. The metric, $\mathcal{M}_1$ sharply decreases in the range $0.01 \leq r_{initial}/c \leq 0.05$, reaching $\leq 5\%$ of the asymptotic value of $\mathcal{M}_1$, i.e., $\lim_{r_{initial} \to \infty} \mathcal{M}_1(r_{initial}) \leq 1.05$. The average percentage error reduction in the velocity is: $44\% \leq \Delta\text{RMSE}(\boldsymbol{U}) \leq 61\%$. (For reference, with only a third of the sensors, the lower bound is still an improvement on the uniformly placed sensors case, Fig. 11.) Illustrative placement scenarios are shown in Fig. 13. Near the lower initial radius bound the sensors are clearly too clustered, while near the upper bound, the sensors are overstretched (some placed in regions of low uncertainty, while many regions of higher uncertainty are not considered during the placement). The configuration in and close to the "target zone" in Fig. 12, $0.05 \leq r_{initial}/c \leq 0.1$, seem to be good potential candidates. This is broadly confirmed by the similar error reduction in the streamwise component of velocity. Admittedly, the prescribed metric $\mathcal{M}_1$ does not guarantee the optimal solution and requires the exercises of an element of engineering judgment.

Investigations of the second hyper-parameter, the number of sensors, are summarized in Figs. 14 and 15. The analysis is with reference to the second metric, $\mathcal{M}_2$ introduced in Sec. II D. For all the number of sensors considered, the initial radius is chosen iteratively using the previously described approach: the final $r_{initial}$ is chosen before $\mathcal{M}_1$

reaches an asymptotic value of 1.02 (see illustrative placement configurations in Fig. 15). The value of the metric $\mathcal{M}_2$ initially reduces sharply up to 10–15 sensors, then gradually leveling from $N \geq 35$, in Fig. 14. The average percentage error reduction in the velocity is: $53\% \leq \Delta\text{RMSE}(\boldsymbol{U}) \leq 71\%$. The error reduction in the streamwise velocity (the quantity used as reference data for field inversion, and the dominant component in terms of contribution to velocity magnitude) levels after around 20 sensors and is reasonably close to the scenario when all the data are used. On the other hand, the reduction in the wall-normal velocity component is less clear as a function of number of sensors. Figure 14 demonstrates that the metric introduced, $\mathcal{M}_2$, can be used as a reasonable guide to approximating the number of sensors a priori (i.e., by setting the number of sensors around the values when $\mathcal{M}_2$ levels off). However, there is no guarantee of optimality. In addition, as previously mentioned, the number of sensor may also be dictated by other constraints (e.g., operational budget, etc.).

Finally, we present some more detailed results in terms of normalized streamwise velocity profiles (Fig. 16) and surface pressure distribution on the hump wall (Fig. 17). The velocity profiles show that all field inversion scenarios are able to considerably reduce the errors in the separated shear layer. Broadly, increasing the number of sensors leads to an improved reconstructed quantity; however, as previously mentioned, using a very large dataset does not mean a proportionally
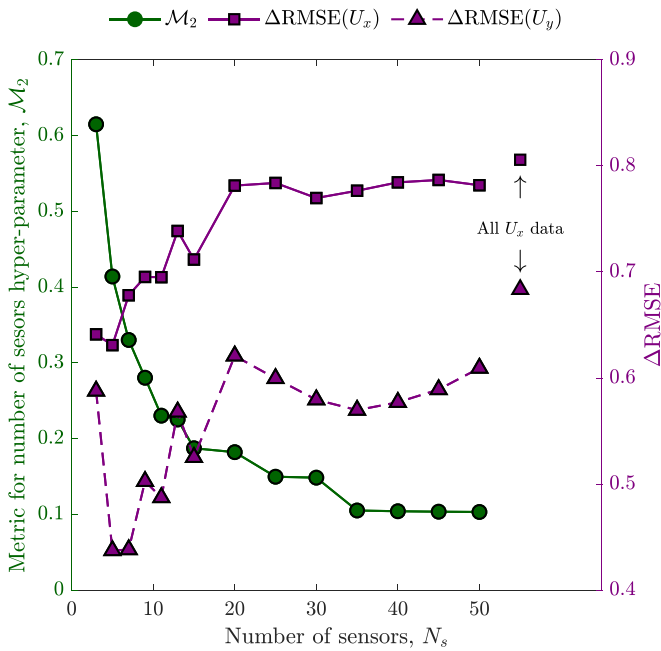
**FIG. 14.** The tuning metric $\mathcal{M}_2$ [Eq. (14)] on the left axis, and the effect of number of sensors on flow reconstruction error reduction on the right axis.

similar error reduction. Encouragingly, Fig. 17 shows that using relatively sparse velocity data also leads in better reconstruction of the surface pressure distribution, notably in and around the separation zone, $0.5 \lesssim x/c \lesssim 1.5$.

### B. 2D converging-diverging channel flow

The second case is flow in a channel with a smooth converging-diverging lower wall, Fig. 18, at a moderate Reynolds number of $Re_H = 12\,600$ based on the channel half-height, $H$, and the maximum inlet velocity. Similar to the previous case, we employ high-fidelity simulation data, from direct numerical simulations in this case,[50,51] as surrogate measurement data. The flow is again characterized by an adverse pressure gradient, resulting in a separation bubble on the lower wall, but at a much milder Reynolds number. In addition, it is another well-established flow for testing data-driven RANS models, e.g., Refs. 8, 52, and 53 The aim is to highlight the limitations of the baseline RANS model and to reinforce the applicability of the proposed framework across multiple scenarios. For brevity, only the key results will be discussed.

The eigenspace perturbation-based uncertainty map for the streamwise velocity field is shown in Fig. 19. The regions of highest uncertainty are aft of the channel bump apex, particularly concentrated in the separated shear layer. We show comparison of the streamwise velocity profiles in Fig. 20. The baseline $k - \omega$ SST model significantly over-predicts the flow separation and the recirculation zone. For data
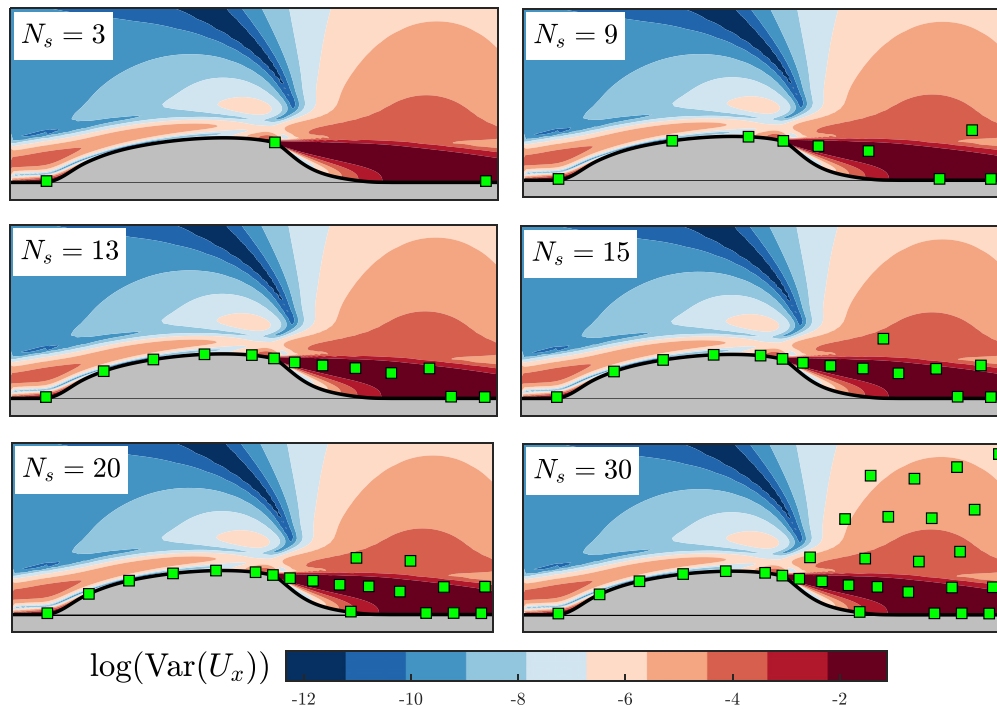


**FIG. 15.** Some illustrative sensor configurations, for investigating the effect of number of sensors on data assimilation.
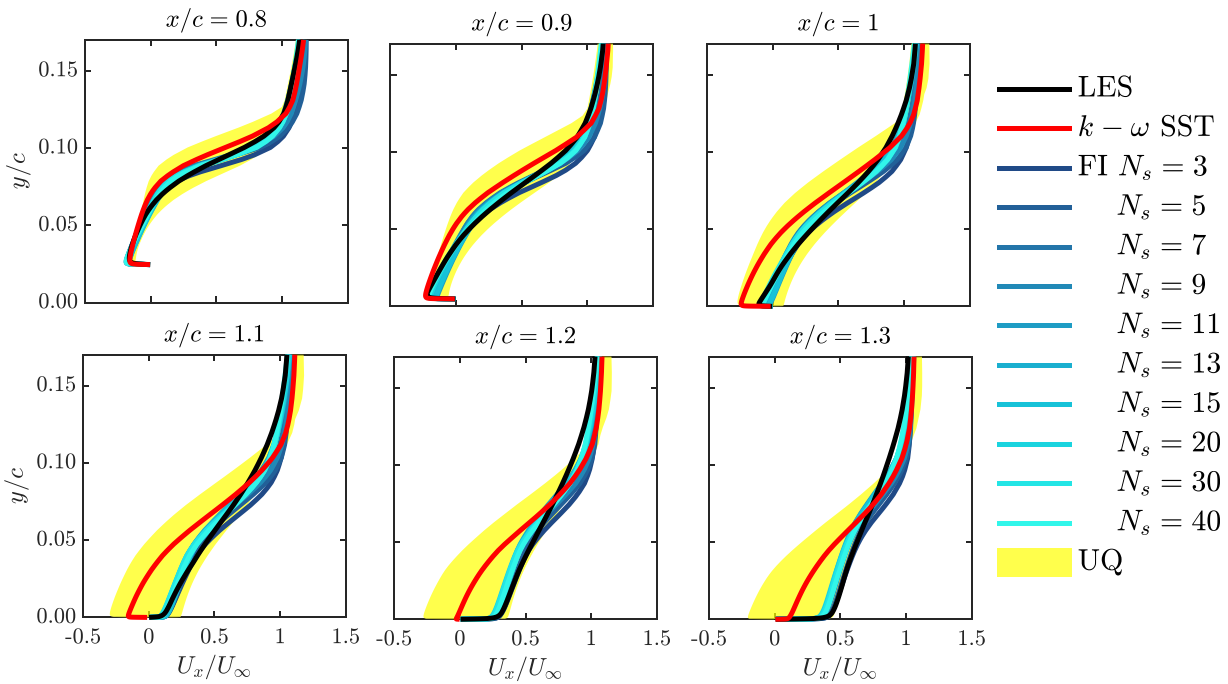
**FIG. 16.** Normalized streamwise velocity profiles comparison for the 2D NASA hump case before and after data assimilation for different number of sensors.

assimilation, we prescribe $N_s = 15$ sensors and place these using the proposed algorithm. The entire domain is assumed to be potential sites for sensors, i.e., $\chi = X \in R^{N_{cells}}$, with $N_{cells} = 98\,700$. The initial radius is selected using the metric, $\mathcal{M}_1$ [Eq. (13)] leading to



**FIG. 17.** Surface pressure predictions on the 2D NASA hump wall. For the legend refer to Fig. 16.

$r_{initial} = 0.53H$. The resulting sensor locations are shown in Fig. 19. Reconstructed flow result with this configuration demonstrates significant improvements to the baseline predictions, especially in the separation and flow recovery region near the lower wall. As the velocity profiles demonstrate, the flow reconstruction with the streamwise velocity data in the entire flow field leads to a marginally better field inversion result, also summarized in terms of the root-mean-square error reduction in Table I.
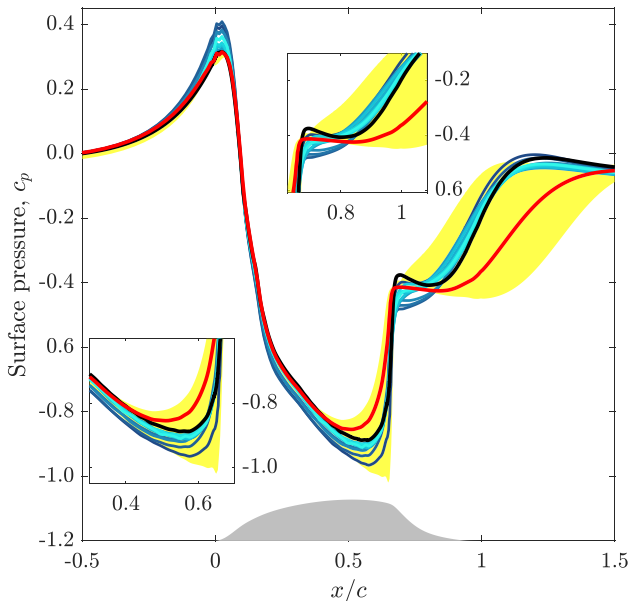
## C. Separated periodic hill flow

The final test case is the separated periodic hill flow. The geometry is shown in Fig. 21. It involves two curved surfaces (hills) connected by a flat plate. The case involves flow separation on the initial hill and reattachment on the flat plate. The flow separation is poorly predicted by the commonly used models (e.g., Refs. 12 and 35) making it a popular benchmarking case with extensive experimental (e.g., Refs. 54 and 55) and high-fidelity simulations (LES, e.g., Refs. 56 and 57 and DNS, e.g., Refs. 58 and 59) data available in the literature.

We use the DNS data from Xiao *et al.*[59] The flow has a Reynolds number of $Re_H = 5\,600$, based on the hill height $H$, and the bulk velocity. $U_b$ Cyclic boundary conditions are applied at the inlet and outlet, and the flow is driven by a mean velocity which is maintained by adding a source term to the momentum equations, with the bulk velocity and Reynolds number defined as follows:

$$Re_H = \frac{U_b H}{\nu}, \, U_b = \frac{1}{2.035H} \int_H^{3.035H} U_x(y) dy, \quad (17)$$

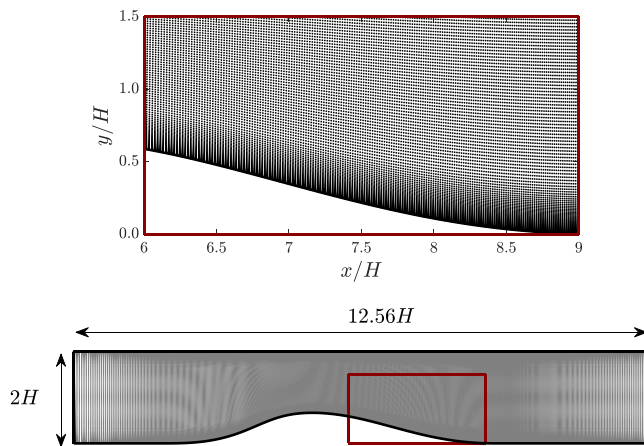where $\nu$ is the kinematic viscosity.

**FIG. 18.** Converging-diverging channel geometry, and mesh. The red box shows a close-up of the mesh from Ref. 49, with the dots representing cell-centers. DNS data are available over the entire flow field.

The uncertainty map from eigenspace perturbations are shown in Fig. 22. The uncertainty map is based on the combined variances of the streamwise and wall-normal velocity components, and the pressure. The maximum regions of uncertainty are concentrated near the upper and lower walls, in general, and around the hills, in particular. Again, we place 15 sensors, with the initial exclusion domain radius, $r_{initial} = 0.68H$, which is selected iteratively using the metric $\mathcal{M}_1$ [Eq. (13)].

We present flow reconstruction results using $N_s = 15$ sensors placed using the proposed framework and using the entire streamwise velocity field data (i.e., $N_s \equiv N_{cells} = 14\,751$). The root-mean-square errors for the velocity components in the two field inversion scenarios are reported in Table II. We note that the flow reconstruction with just 15 sensors is marginally inferior to the case when all the streamwise velocity data are used.

Streamwise velocity profiles are shown in Fig. 23. The baseline SST model is inaccurate near the walls, especially in the separated shear layer near the lower wall, as captured in the uncertainty ranges from the eigenspace perturbations. The baseline model over-predicts the size of the separation bubble. Both field inversion scenarios are able to improve the predictions; however, some discrepancies remain in the region close to the second hill. Once again, the improvements using only 15 data points vs the entire streamwise velocity field is noteworthy.
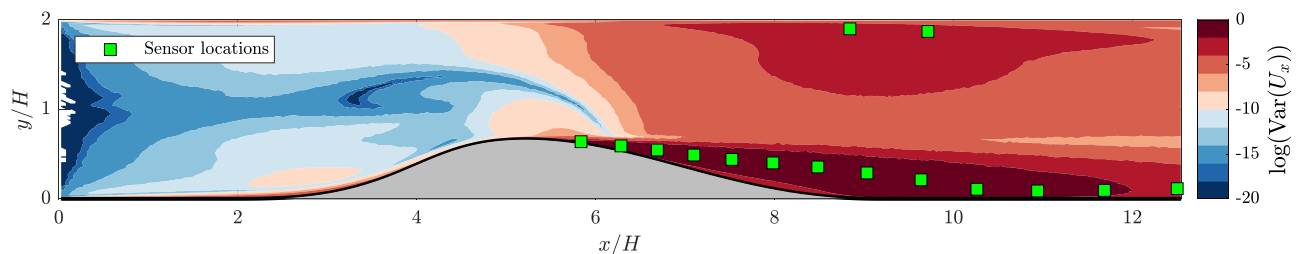
## IV. DISCUSSION

In this section, we will consolidate the features of the proposed framework, and the results presented above.

We presented a sensor placement strategy for the data assimilation of RANS-based turbulence models. The key component is the systematic eigenspace perturbations of the modeled Reynolds-stress tensor to generate a spatial uncertainty map—in practice, five RANS-based simulations, with the uncertainty map generated as a function of the variance of a particular quantity based on the five perturbed flow realizations. We then propose a computationally efficient greedy algorithm to sensor placement. The greedy algorithm targets sensors, sequentially, at the regions of highest uncertainty, while avoiding the clustering of sensors in any one particular region by using a circular (spherical in three-dimensional space) exclusion domain after a sensor is placed. The first hyper-parameter required is the radius for the initial exclusion zone, with the subsequent zones extending as a function of the change in the uncertainty [i.e., variance $\mathcal{V}(q)$]. We proposed a metric to guide the tuning of this hyper-parameter iteratively [$\mathcal{M}_1$ in Eq. (13)]. The second hyper-parameter is the number of sensors, which can be estimated using $\mathcal{M}_2$ in Eq. (14).

The proposed strategy has the following advantages:

1. Through eigenspace perturbations we directly address epistemic structural uncertainties in turbulence models (not fully captured in the perturbations to model constants in Ref. 36, for instance).
2. The strategy is computationally cheap. The computational overhead involves five RANS-based flow simulations for eigenspace perturbations, plus the sensor placement using the proposed greedy algorithm. The computational time for sensor placement is on the order of seconds for all three test cases presented. For comparison, in Ref. 36, parametric perturbations involving around 100 CFD flow simulations are required, followed by deep neural network training for sensor placement—which is easily one or more order of magnitude more expensive. In Refs. 37 and 38, a first or second-order adjoint-based optimization problem is solved to place sensors, which can be orders of magnitude more expensive compared to the proposed method. We note that the computational overhead discussed here is purely for *a priori* sensor placement and does not include the costs for data assimilation once high-fidelity data is generated.
3. The proposed algorithm is relatively easy to implement (for example, an open-source implementation of the eigenspace perturbations, and the data assimilation were highlighted) and the
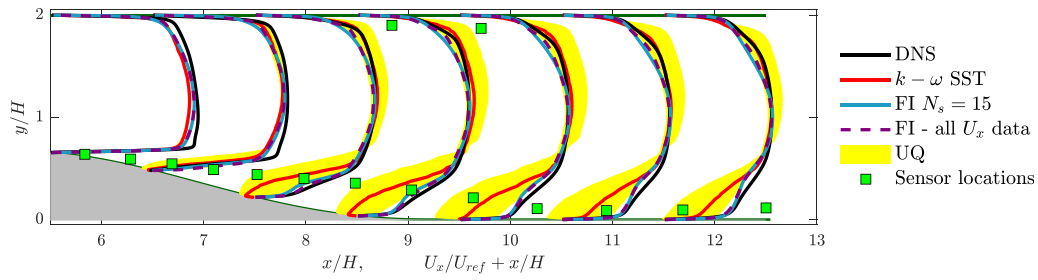


**FIG. 19.** The streamwise velocity uncertainty map for the converging-diverging channel.

**FIG. 20.** Velocity profiles comparison for the converging-diverging channel.

**TABLE I.** Data assimilation root-mean-square error reduction for the converging-diverging channel flow.

| Case | $N_s$ | $\Delta$RMSE($U_x$) (%) | $\Delta$RMSE($U_y$) (%) | Average (%) |
|------|-------|-------------------------|-------------------------|-------------|
| ESP | 15 | 51.4 | 58.4 | 54.9 |
| All $U_x$ | 98 700 | 57.3 | 59.5 | 58.4 |

greedy sensor placement algorithm (Algorithm 1) can be implemented straight-forwardly in a few dozen lines of code.

Once sensors are placed, and the data collected (synthetic high-fidelity simulation data from direct numerical or large-eddy simulations in the case studies presented), we performed data assimilation using the variational (adjoint-based) field inversion. This involved perturbing the transport equation of the baseline RANS model by a spatial scalar-field, $\beta^{FI}$, and solving an inverse problem that optimizes the field values such that the error between the RANS predictions and the reference data are minimized. While we used this particular data assimilation approach due to its advantages (model consistency, directly addressing functional errors in RANS-based models, and ability to work with relatively sparse datasets) and our previous experience with this method,[12,35,60] the proposed sensor placement framework should, in principle, be similarly applicable to other approaches (e.-g., ensemble-based field inversion) since the sensor placement problem is, inherently, decoupled from the data assimilation process. The decoupled nature of the strategy may be a limitation as we do not consider the sensitivity of the RANS model states with respect to the regions of uncertainty *during* the sensor placement. The relative simplicity and low computational cost of the strategy and the results presented demonstrate the effectiveness in spite of this.
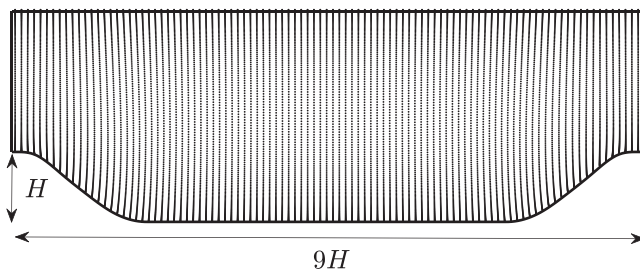
We tested the framework on three two-dimensional wall-bounded turbulent flows (I) NASA wall-mounted hump: $Re = 9.36 \times 10^5$, (II) converging-diverging channel, $Re = 12\,600$, and (III) periodic hill, $Re = 5600$, all three involving flow separation and reattachment. We showed that the baseline $k - \omega$ SST model struggles to accurately capture the flow in the separated shear layer in all three cases and over-predicted the extent of flow separations. We placed sensors based on the uncertainty map of the streamwise velocity for cases I and II and a combination of velocity components and pressure for case III. Streamwise velocity data were used for data assimilation since it is a routinely measured quantity in experimental studies. Regions of highest uncertainty identified through eigenspace perturbations were consistently concentrated in the flow areas involving complex turbulent structures such as boundary layer close to the walls, points of flow separation and re-attachment, and areas of reversed flow (such as those caused by adverse pressure gradients). All regions of interest from a fluid dynamics standpoint and, thus, appropriate for sensor placement.

Key findings of the results for the hump case are: an average mean velocity error reduction of 61% was achieved with just three sensors, compared to 74% when using all the available streamwise velocity field data; data assimilation with just the streamwise velocity also resulted in significant error reductions in the wall-normal velocity, and a better match of the surface pressure predictions on the hump wall; and detailed numerical investigations of proposed metrics for hyperparameter tuning demonstrated that these were effective measures to
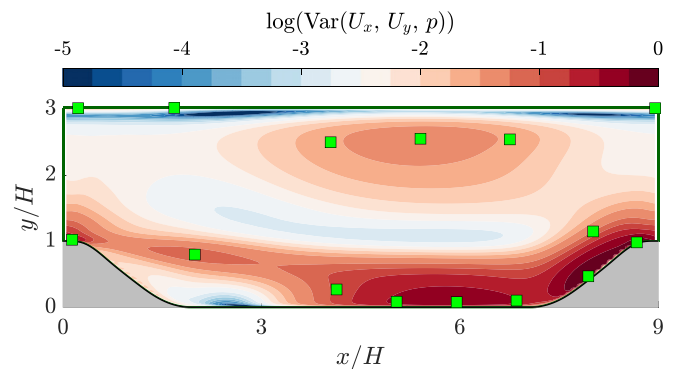


**FIG. 21.** The periodic hill geometry and mesh cell centers.



**FIG. 22.** The uncertainty map for the periodic hill case. The variance is based on a combination of the streamwise and wall normal velocity components, and the pressure.

TABLE II. Data assimilation root-mean-square error reduction for the periodic hill flow.

| Case | $N_s$ | $\Delta$RMSE$(U_x)$ (%) | $\Delta$RMSE$(U_y)$ (%) | Average (%) |
|------|-------|------------------------|------------------------|-------------|
| ESP | 15 | 42.5 | 50.1 | 46.7 |
| All $U_x$ | 14 751 | 55.6 | 43.3 | 49.4 |
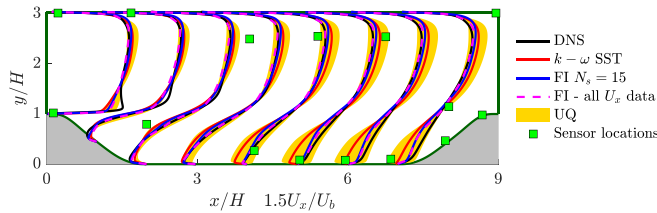


FIG. 23. The streamwise velocity profiles comparison for the periodic hill case.



FIG. 25. The corrective field $\beta$ for the converging-diverging channel flow, using $N_s = 15$.

guide practitioners. In the case of converging-diverging channel flow, we demonstrated an error reduction in the mean velocity of 55% with just 15 sensors, compared to 58% reduction when using the streamwise velocity in the entire flow domain. Finally, for the periodic hill case, an average velocity error reduction of 46% was achieved with 15 sensors, compared to 49% when using the entire streamwise velocity field.

The corrective fields $\beta^{\text{FI}}$ modifying the transport equation for the turbulence dissipation ($\omega$) were presented for all three cases in Fig. 24 (NASA hump), Fig. 25 (converging-diverging channel), and Fig. 26 (periodic hill). These demonstrate that very complex non-linear modifications of the baseline model are required for effective flow reconstruction.

All three test cases presented were two-dimensional flows, while the methodology presented is directly applicable to three-dimensional cases. One of the key challenges for the absence of a three-dimensional test case in the present work is the lack of such a benchmark flow with rich high-fidelity, well-established, publicly available data that would allow similar benchmarking to the two-dimensional cases presented. Future work will generate the high-fidelity data to allow testing the framework for three-dimensional flows.

## V. CONCLUSION

We presented a novel strategy to sensor placement for data assimilation in the context of RANS-based turbulent flow reconstruction. The framework relied on generating a spatial uncertainty map through systematic perturbations o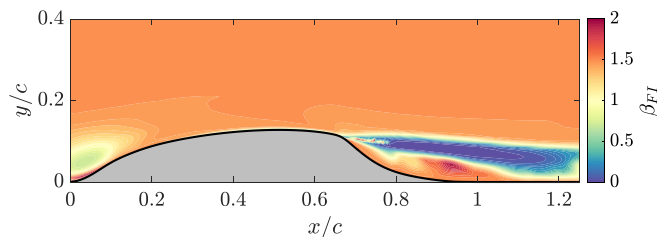f the Reynolds stress tensor. An optimization-based greedy search was proposed to place sensors. Data assimilation was performed using a variational (adjoint-based) approach.

The proposed framework was tested on three two-dimensional wall-bounded fully turbulent flows with Reynolds numbers ranging $5.6 \times 10^3$–$9.36 \times 10^5$. All flows involved mild to massive separation which were over-predicted by the baseline $k - \omega$ SST model. Data assimilation was performed using streamwise velocity data. Average root-mean-square error reduction in the velocity predictions (compared to the baseline SST model) after data assimilation are summarized as follows:

1. 2D NASA wall-mounted hump: the error was reduced by 61% just using 3 sensor data using the proposed placement method. For comparison, using 33 sensors with uniform sensor placement led to an error reduction of 38%, while using the entire available streamwise velocity data ($30.7 \times 10^3$ data points) achieved 74% error reduction.
2. Converging-diverging channel: average error reduction of 55% with 15 sensors using the proposed method, compared to a 58% reduction when using all the streamwise velocity field ($98.7 \times 10^3$ data points).
3. Periodic hill flow: average velocity error reduction of 47% with 15 sensors, compared to 49% when using all the data ($14.7 \times 10^3$ data points).

The framework was tested on relatively well-established two-dimensional flows. In principle, it can be extended to three-dimensional cases using the same methodology outlined in Sec. II. Future work will investigate the framework for complex three-dimensional flows (e.g., transonic and supersonic aerospace flows, flows involving heat transfer, etc.). In addition we will consider sensor placement for experimentally measurable surface data (e.g., skin friction and surface pressure), investigate the eigenspace perturbations and data assimilation for additional turbulence models and test the framework with other data assimilation techniques, e.g., EnKF-based field inversion.



FIG. 24. The corrective field $\beta$ for the 2D NASA hump case, modifying the SST transport equation after data assimilation. This is for the case with $N_s = 33$.
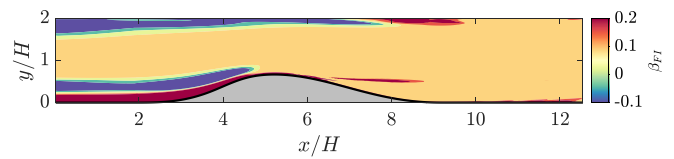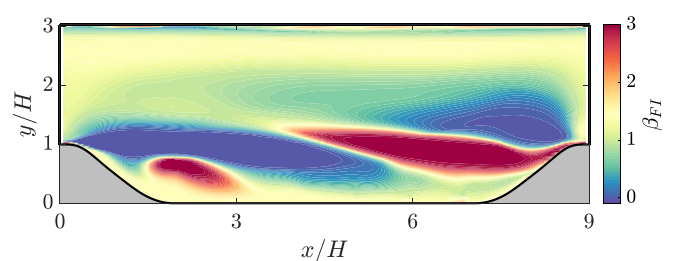


FIG. 26. The corrective field $\beta$ for the separated periodic hill flow.

## AUTHOR DECLARATIONS
### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Omid Bidar:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Sean Anderson:** Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Ning Qin:** Conceptualization (equal); Formal analysis (equal); Methodology (equal); Supervision (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

[1] P. A. Durbin, "Some recent developments in turbulence closure modeling," Annu. Rev. Fluid Mech. **50**, 77–103 (2018).

[2] H. Xiao and P. Cinnella, "Quantification of model uncertainty in RANS simulations: A review," Prog. Aerosp. Sci. **108**, 1–31 (2019).

[3] K. Fukami, K. Fukagata, and K. Taira, "Super-resolution reconstruction of turbulent flows with machine learning," J. Fluid Mech. **870**, 106–120 (2019).

[4] K. Duraisamy, G. Iaccarino, and H. Xiao, "Turbulence modeling in the age of data," Annu. Rev. Fluid Mech. **51**, 357–377 (2019).

[5] J. Ling, A. Kurzawski, and J. Templeton, "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance," J. Fluid Mech. **807**, 155–166 (2016).

[6] J. Weatheritt and R. Sandberg, "The development of algebraic stress models using a novel evolutionary algorithm," Int. J. Heat Fluid Flow **68**, 298–318 (2017).

[7] J.-L. Wu, H. Xiao, and E. Paterson, "Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework," Phys. Rev. Fluids **3**, 074602 (2018).

[8] M. Schmelzer, R. P. Dwight, and P. Cinnella, "Discovery of algebraic Reynolds-stress models using sparse symbolic regression," Flow Turbul. Combust. **104**, 579–603 (2019).

[9] A. P. Singh, S. Medida, and K. Duraisamy, "Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils," AIAA J. **55**, 2215–2227 (2017).

[10] X.-L. Zhang, H. Xiao, G.-W. He, and S.-Z. Wang, "Assimilation of disparate data for enhanced reconstruction of turbulent mean flows," Comput. Fluids **224**, 104962 (2021).

[11] C. He, P. Wang, and Y. Liu, "Sequential data assimilation of turbulent flow and pressure fields over aerofoil," AIAA J. **60**, 1091–1103 (2022).

[12] O. Bidar, P. He, S. Anderson, and N. Qin, "Turbulent mean flow reconstruction based on sparse multi-sensor data and adjoint-based field inversion," AIAA Paper No. 2022-3900.

[13] P. Mokhasi and D. Rempfer, "Optimized sensor placement for urban flow measurement," Phys. Fluids **16**, 1758–1764 (2004).

[14] K. Cohen, S. Siegel, and T. McLaughlin, "A heuristic approach to effective sensor placement for modeling of a cylinder wake," Comput. Fluids **35**, 103–120 (2006).

[15] B. Yildirim, C. Chryssostomidis, and G. Karniadakis, "Efficient sensor placement for ocean measurements using low-dimensional concepts," Ocean Modell. **27**, 160–173 (2009).

[16] B. Jin, S. J. Illingworth, and R. D. Sandberg, "Resolvent-based approach for $h_2$-optimal estimation and control: an application to the cylinder flow," Theor. Comput. Fluid Dyn. **36**, 491–515 (2022).

[17] M. L. Provost, W. Hou, and J. Eldredge, "Deep learning and data assimilation approaches to sensor reduction in estimation of disturbed separated flows," AIAA Paper No. 2020-0799.

[18] P. Kumar, Y. M. E. Sayed, and R. Semaan, "Optimized sensor placement using stochastic estimation for a flow over a 2D airfoil with coanda blowing," AIAA Paper No. 2014-2101.

[19] N. Ali, M. Calaf, and R. B. Cal, "Clustering sparse sensor placement identification and deep learning based forecasting for wind turbine wakes," J. Renewable Sustainable Energy **13**, 023307 (2021).

[20] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns," IEEE Control Syst. Mag. **38**, 63–86 (2018).

[21] J. L. Callaham, K. Maeda, and S. L. Brunton, "Robust flow reconstruction from limited measurements via sparse representation," Phys. Rev. Fluids **4**, 103907 (2019).

[22] S. Lu and G. Papadakis, "Flow reconstruction around a surface-mounted prism from sparse velocity and/or scalar measurements using a combination of POD and a data-driven estimator," Flow Turbul. Combust. **110**, 1059–1090 (2023).

[23] N. Karnik, M. G. Abdo, C. E. Estrada Perez, J. S. Yoo, J. J. Cogliati, R. S. Skifton, P. Calderoni, S. L. Brunton, and K. Manohar, "Optimal sensor placement with adaptive constraints for nuclear digital twins," arXiv:2306.13637 (2023).

[24] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," Ocean Dyn. **53**, 343–367 (2003).

[25] C. H. Colburn, J. B. Cessna, and T. R. Bewley, "State estimation in wall-bounded flow systems. Part 3. The ensemble Kalman filter," J. Fluid Mech. **682**, 289–303 (2011).

[26] V. Mons, J.-C. Chassaing, T. Gomez, and P. Sagaut, "Reconstruction of unsteady viscous flows using data assimilation schemes," J. Comput. Phys. **316**, 255–280 (2016).

[27] D. P. G. Foures, N. Dovetta, D. Sipp, and P. J. Schmid, "A data-assimilation method for Reynolds-averaged Navier–Stokes-driven mean flow reconstruction," J. Fluid Mech. **759**, 404–431 (2014).

[28] 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).

[29] A. P. Singh and K. Duraisamy, "Using field inversion to quantify functional errors in turbulence closures," Phys. Fluids **28**, 045110 (2016).

[30] C. He, Y. Liu, and L. Gan, "A data assimilation model for turbulent flows using continuous adjoint formulation," Phys. Fluids **30**, 105108 (2018).

[31] E. J. Parish and K. Duraisamy, "A paradigm for data-driven predictive modeling using field inversion and machine learning," J. Comput. Phys. **305**, 758–774 (2016).

[32] J. R. Holland, J. D. Baeder, and K. Duraisamy, "Field inversion and machine learning with embedded neural networks: Physics-consistent neural network training," AIAA Paper No. 2019-3200.

[33] C. Wu and Y. Zhang, "Enhancing the shear-stress-transport turbulence model with symbolic regression: A generalizable and interpretable data-driven approach," Phys. Rev. Fluids **8**, 084604 (2023).

[34] Y. Yang, C. Robinson, D. Heitz, and E. Mémin, "Enhanced ensemble-based 4dvar scheme for data assimilation," Comput. Fluids **115**, 201–210 (2015).

[35] O. Bidar, P. He, S. Anderson, and N. Qin, "An open-source adjoint-based field inversion tool for data-driven RANS modelling," AIAA Paper No. 2022-4125.

[36] Z. Deng, C. He, and Y. Liu, "Deep neural network-based strategy for optimal sensor placement in data assimilation of turbulent flow," Phys. Fluids **33**, 025119 (2021).

[37] V. Mons, J.-C. Chassaing, and P. Sagaut, "Optimal sensor placement for variational data assimilation of unsteady flows past a rotationally oscillating cylinder," J. Fluid Mech. **823**, 230–277 (2017).

20 February 2024 15:08:21

[38]V. Mons and O. Marquet, "Linear and nonlinear sensor placement strategies for mean-flow reconstruction via data assimilation," J. Fluid Mech. **923**, A1 (2021).

[39]G. Iaccarino, A. A. Mishra, and S. Ghili, "Eigenspace perturbations for uncertainty estimation of single-point turbulence closures," Phys. Rev. Fluids **2**, 024605 (2017).

[40]K. Nakai, K. Yamada, T. Nagata, Y. Saito, and T. Nonomura, "Effect of objective function on data-driven greedy sparse sensor optimization," IEEE Access **9**, 46731–46743 (2021).

[41]B. E. Launder and N. D. Sandham, eds., *Closure Strategies for Turbulent and Transitional Flows* (Cambridge University Press, 2001).

[42]F. Menter, "Zonal two equation k-w turbulence models for aerodynamic flows," in *23rd Fluid Dynamics, Plasmadynamics, and Lasers Conference* (American Institute of Aeronautics and Astronautics, 1993).

[43]S. Banerjee, R. Krahl, F. Durst, and C. Zenger, "Presentation of anisotropy properties of turbulence, invariants versus eigenvalue approaches," J. Turbul. **8**, N32 (2007).

[44]M. Emory, J. Larsson, and G. Iaccarino, "Modeling of structural uncertainties in Reynolds-averaged Navier–Stokes closures," Phys. Fluids **25**, 110822 (2013).

[45]A. A. Mishra, J. Mukhopadhaya, G. Iaccarino, and J. Alonso, "Uncertainty estimation module for turbulence model predictions in SU2," AIAA J. **57**, 1066–1077 (2019).

[46]Z. Belligoli, R. P. Dwight, and G. Eitelberg, "Reconstruction of turbulent flows at high Reynolds numbers using data assimilation techniques," AIAA J. **59**, 855–867 (2021).

[47]C. L. Rumsey, "Recent developments on the turbulence modeling resource website (invited)," AIAA Paper No. 2015-2927.

[48]A. Uzun and M. R. Malik, "Large-eddy simulation of flow over a wall-mounted hump with separation and reattachment," AIAA J. **56**, 715–730 (2018).

[49]R. McConkey, E. Yee, and F.-S. Lien, "A curated dataset for data-driven turbulence modelling," Sci. Data **8**, 255 (2021).

[50]J.-P. Laval and M. Marquillie, "Direct numerical simulations of converging–diverging channel flow," in *ERCOFTAC Series* (Springer, Netherlands, 2011), pp. 203–209.

[51]M. Marquillie, U. Ehrenstein, and J.-P. Laval, "Instability of streaks in wall turbulence with adverse pressure gradient," J. Fluid Mech. **681**, 205–240 (2011).

[52]X. Zhang, T. Gomez, and O. Coutier-Delgosha, "Bayesian optimisation of RANS simulation with ensemble-based variational method in convergent-divergent channel," J. Turbul. **20**, 214–239 (2019).

[53]S. Li, C. He, and Y. Liu, "A data assimilation model for wall pressure-driven mean flow reconstruction," Phys. Fluids **34**, 015101 (2022).

[54]C. Rapp and M. Manhart, "Flow over periodic hills: An experimental study," Exp. Fluids **51**, 247–269 (2011).

[55]M. Breuer, N. Peller, C. Rapp, and M. Manhart, "Flow over periodic hills - numerical and experimental study in a wide range of Reynolds numbers," Comput. Fluids **38**, 433–457 (2009).

[56]X. Gloerfelt and P. Cinnella, "Large eddy simulation requirements for the flow over periodic hills," Flow Turbul. Combust. **103**, 55–91 (2019).

[57]J. Frohlich, C. P. Mellen, W. Rodi, L. Temmerman, and M. A. Leschziner, "Highly resolved large-eddy simulation of separated flow in a channel with streamwise periodic constrictions," J. Fluid Mech. **526**, 19–66 (2005).

[58]B. Krank, M. Kronbichler, and W. A. Wall, "Direct numerical simulation of flow over periodic hills up to $Re_H = 10\,595$," Flow, Turbul. Combust. **101**, 521–551 (2018).

[59]H. Xiao, J.-L. Wu, S. Laizet, and L. Duan, "Flows over periodic hills of parameterized geometries: A dataset for data-driven turbulence modeling from direct simulations," Comput. Fluids **200**, 104431 (2020).

[60]O. Bidar, "Relative importance of physical quantities for data-driven RANS-based turbulence modelling," paper presented at the UK Fluids Conference, Sheffield, UK, 2022.

20 February 2024 15:08:21