Proceedings of the Thirty-sixth (2026) International Ocean and Polar Engineering Conference **www.isope.org**
Orlando, Florida, USA, May 31-June 5, 2026

ISBN 978-1-880653-72-2; ISSN 1098-6189

Ensemble-Based Data Assimilation for Improving Near-Wall Velocity Prediction in Backward-Facing Step Flows

Wenbin Zhang[1], Jie Zhu[2], Maokun Ye[1], Decheng Wan[1]*

[1] Computational Marine Hydrodynamics Lab (CMHL), School of Ocean and Civil Engineering,
Shanghai Jiao Tong University, Shanghai, China
[2] Wuhan Second Ship Design and Research Institute, Wuhan, China
[*]Corresponding Author

ABSTRACT

Accurate prediction of near-wall flow characteristics in backward-facing step (BFS) flows remains challenging for RANS turbulence models, due to inherent limitations in turbulence closure assumptions. This study presents a data assimilation framework using the Ensemble Kalman Filter (EnKF) within the DAFI toolkit to improve near-wall velocity prediction in simulations of BFS flows. The baseline simulation is conducted on the open-source software OpenFOAM using the k−ω SST turbulence model. The turbulence viscosity field (ν_t) after dimensionality reduction via Karhunen-Loève (KL) decomposition is taken as the field to be assimilated. The velocity at selected points obtained from the experiment data is treated as the observations of the real physics process. Then the EnKF method is applied in an iterative manner until an optimal estimation of the turbulence viscosity field is achieved. Validation against high-fidelity experimental data shows that the posterior results significantly reduce the root-mean-square errors (RMSE) in near-wall velocity profiles compared to baseline k−ω SST simulations. Improvements are evident not only in the vicinity of observation points but also in unobserved regions, confirming the framework's ability to extrapolate corrections effectively over the computational domain.

KEY WORDS: data assimilation; ensemble Kalman filter; CFD; backward facing step

INTRODUCTION

Reynolds-averaged Navier-Stokes (RANS) methods are widely adopted for turbulent flow analysis, primarily owing to their high computational efficiency that enables rapid acquisition of engineering-oriented results. RANS equations are derived via time-averaging of the primitive Navier-Stokes equations and serve to predict the steady-state behaviors of turbulent flows. In the RANS framework, instantaneous flow variables are decomposed by Reynolds decomposition into time-averaged and fluctuating components. The arising Reynolds stresses, stemming from unresolved fluctuation correlations, render the equations unclosed, thus requiring dedicated closure models. A classic closure model is the k−ω SST model, which combines the k−ω model and the k−ε model, thus exhibiting good adaptability to flows in different regions.

Although significant progress has been made in the improvement of Reynolds-Averaged turbulence closure models over the past decades, it remains challenging for the turbulence closure models to simulate the strongly separated wall-bounded turbulent flow. Therefore, a trend that utilizes data-driven approaches to enhance closure models is emerging in recent years. A recently popular method involves substituting the closure model with a trained machine learning (ML) model. Ling et al. (2016) developed a tensor basis neural network (TBNN) for RANS turbulence modelling, integrating Galilean invariance through an invariant tensor basis to predict the Reynolds stress anisotropy tensor. Trained on high-fidelity data, the TBNN outperforms conventional eddy viscosity models and generic neural networks in capturing anisotropy and key flow phenomena across test cases. Wu et al. (2021) proposed a machine learning surrogate model for the k−ω SST turbulence model to address the closure problem in RANS simulations. Validated on backward-facing step flows across different Reynolds numbers, the model achieves high accuracy when test Reynolds numbers are close to the training range, shows improved performance at higher Reynolds numbers, and significantly reduces computational time compared to the traditional k−ω SST model. Similar work has been conducted by numerous researchers (Grabe et al., 2023; Wang et al., 2017; Zhu et al., 2019).

However, the training of a ML model usually requires massive high-fidelity numerical or experiment data, which is very expensive to acquire. Recently, a new branch of data-driven method, namely, ensemble-based data assimilation has emerged. This method employs the ensemble Kalman Filter (EnKF) (Lakshmivarahan and Stensrud, 2009) and is able to utilize noisy and sparse observation data to improve RANS closures. Aulakh et al. (2024) utilized the EnKF method for improving the performance of the Spalart-Allmaras (SA) turbulence model in separated flows while preserving generalization for unseen cases. The method calibrates key coefficients of the SA model using sparse and noisy experimental data from a single backward-facing step (BFS) flow. The recalibrated model shows enhanced accuracy in predicting skin friction coefficient (C_f) and pressure coefficient (C_p) for strong separated flows and maintains the original SA model's proficiency in attached and unbounded flows. Zhang et al. (2022) proposed an EnKF-based acoustic inversion method to reduce uncertainties in RANS-based jet noise prediction. The method infers turbulent kinetic energy (k) and dissipation rate (ε) by assimilating far-field noise data and can incorporate acoustic model parameters as uncertain quantities. It improves the reconstruction of turbulent flow fields and far-field noise at unobserved locations, and its practicality is verified through a realistic case with experimental data. Ge et al. (2024) applied the EnKF to optimize the k–ω turbulence model for ship wake field simulations, aiming to address parametric uncertainty in model coefficients. The method calibrates key coefficients using pseudo-experimental velocity data, significantly improving the accuracy of wake field predictions compared to the baseline model.

Most data assimilation studies aim to obtain a more accurate and generalizable turbulence closure model. Therefore, the state vector usually consists of the parameters of the existing closure model. However, there remains a case-by-case demand for adjusting the ν_t field to reach a better estimation of the velocity field using sparse experimental data. Simply adjusting closure model parameters may not be feasible for the inherent limitations of the closure model. This study aims to test the ability of EnKF with Karhunen-Loève (KL) decomposition on improving the near-wall velocity prediction in BFS flows. Although BFS flow is a classic and simple test case that has been studied via numerous approaches, traditional RANS closure models have difficulty capturing the near-wall velocity. Through performing data assimilation on the ν_t field of the BFS flow, advantages, limitations and potential applications of EnKF in the CFD, especially on enhancing the performance of under-resolved CFD simulations, are discussed.

METHODOLOGY

Governing Equations

The incompressible, steady and Newtonian fluid can be described by the Reynolds Averaged Navier Stokes (RANS) equation:

$$\nabla \cdot \boldsymbol{u} = 0 \tag{1}$$

$$\boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\nabla p + \nu \nabla^2 \boldsymbol{u} - \nabla \cdot \boldsymbol{\tau} \tag{2}$$

where $\boldsymbol{u} = (u, v)$ is the two-dimensional mean velocity, p denotes the mean pressure normalized by the constant fluid density, ν is the kinematic viscosity and the Reynolds stress $\boldsymbol{\tau}$ indicates the effects of instantaneous turbulence fluctuations on the mean flow quantities.

Modeling the Reynolds Stress

In this paper, the linear eddy viscosity model based on Boussinesq hypothesis is used to model the Reynolds stress:

$$\boldsymbol{\tau} = 2\nu_t \boldsymbol{S} - \frac{2}{3} k \boldsymbol{I} \tag{3}$$

where ν_t is the isotropic turbulent eddy viscosity, $\boldsymbol{S} = (\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T)/2$ represents the mean strain-rate tensor, $\boldsymbol{k}$ is the turbulent kinetic energy, and $\boldsymbol{I}$ is the identity tensor.

As the baseline turbulence model, the Menter 2003 version of the k–ω SST model (Menter et al., 2003) is adopted for modeling the turbulent eddy viscosity ν_t. The k–ω SST model is a hybrid model combining the k–ω model and the k–ε model. It activates k–ω model near the boundary to better capture the characteristics of the boundary layer and k–ε model in the far field to reduce the model's sensitivity to inlet conditions and enhance simulation stability.

The key equation of the k–ω SST model is given as follows:

$$\begin{aligned} &\frac{\partial \alpha\rho k}{\partial t} + \nabla \cdot (\alpha\rho \mathbf{U} k) - \nabla \cdot (\alpha\rho D_k \nabla k) \\ &= \alpha\rho P_k - \frac{2}{3}\alpha\rho k(\nabla \cdot \mathbf{U}) - \alpha\rho\beta^* \omega k + S_k \end{aligned} \tag{4}$$

$$\begin{aligned} &\frac{\partial \alpha\rho\omega}{\partial t} + \nabla \cdot (\alpha\rho \mathbf{U}\omega) - \nabla \cdot (\alpha\rho D_\omega \nabla \omega) \\ &= \alpha\rho\gamma \min\left(\frac{G}{\nu_t}, \frac{c_1}{a_1}\beta^* \omega \max\left(a_1\omega, b_1 F_2 \sqrt{S_2}\right)\right) \\ &- \frac{2}{3}\alpha\rho\gamma(\nabla \cdot \mathbf{U})\omega - \alpha\rho\beta\omega^2 - \alpha\rho(F_1 - 1)\frac{CD_{k\omega}}{\omega}\omega + S_\omega \end{aligned} \tag{5}$$

where the details of the SST model are omitted here as they are not the focus of this study.

In the present ensemble-based data assimilation framework, the Karhunen-Loève decomposition with the squared exponential (SE) kernel and boundary condition considered (Michelén Ströfer et al., 2020) is adopted to capture the nonlinear spatial correlations of the turbulent eddy viscosity ν_t field and reduce the dimension. The expression of the SE kernel is as follows:

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sigma_f^2 \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{2l^2}\right) \tag{6}$$

where σ_f^2 describes the amplitude of the correlations, l is the correlation length, and $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ are the coordinates of two points in the domain.

The KL decomposition of the ν_t field consists of the following steps:
(i) Construct the covariance matrix of the ν_t field using the SE kernel and the spatial coords of the grid cells.
(ii) Perform eigenvalue decomposition on the covariance matrix to get the eigenvalues and the eigenvectors, which quantify the energy contribution of each latent mode.
(iii) Map back the eigenvectors to the original physical space to obtain KL modes, which serve as a set of nonlinear orthogonal bases of the ν_t field.
(iv) Extract the corresponding order of KL modes based on the pre-set energy coverage as the basis vectors for the nut field reconstruction.

Ensemble Kalman Filter

The classic Kalman filter was proposed to optimally estimate the states of linear dynamic systems from noisy measurements in 1960. The algorithm consists of two core phases: prediction and update. In the prediction step, the current state estimate and its error covariance are propagated forward via the system's dynamic model to generate a prior state estimation. In the update step, this prior estimate is corrected by incorporating new measurement data, weighted by the relative uncertainty of the model predictions and measurements, to produce an optimal posterior state estimate.

The ensemble Kalman filter (EnKF) is a Monte Carlo implementation of the classic Kalman filter for nonlinear and complex problems like field inversion. In this study, the state vector is composed of the weights of the KL modes for the ν_t field. The observation space is the velocity field. Observations of the truth Y are obtained from the experimental data. To begin with, an ensemble of the state vector X is generated by Lognormal random sampling: $\ln(X) \sim N(\mu, \sigma^2)$, where μ is the ensemble mean and σ is the standard deviation. Let X_i be the i-th sample in the ensemble, each sample is used to reconstruct the corresponding ν_t field. Since no dynamic model exists for the KL modes, the predicted state vector X_i^f is equal to the original X_i. Then the velocity field is obtained based on each ν_t field. After that, predictions of each sample $Z_i^f = HX_i$ are obtained by projecting the velocity field onto the observation space, where H is the transformation operator from state vector space to observation space.

The core of the ensemble Kalman filter is the analysis step. The analysis step consists of the calculation of the Kalman gain K and the update of the state vector. At first, the covariance matrices of and between the state vector and the prediction are calculated as follows:

$$C_X^f = \frac{1}{N}\sum_{i=1}^{n}\left(X_i^f - \bar{X}^f\right)^2 \tag{7}$$

$$C_Z^f = \frac{1}{N}\sum_{i=1}^{n}\left(Z_i^f - \bar{Z}^f\right)^2 \tag{8}$$

$$C_{XZ}^f = \frac{1}{N}\sum_{i=1}^{n}\left(X_i^f - \bar{X}^f\right)\left(Z_i^f - \bar{Z}^f\right) \tag{9}$$

Given the predefined observation error covariance matrix C_Y, the Kalman gain is calculated as follows:

$$K = C_{XZ}^f (C_Z^f + C_Y)^{-1} \tag{10}$$

Then the state vector of each sample X_i^a and its covariance C_X^a are updated by:

$$X_i^a = X_i^f + K(Y_i - Z_i^f) \tag{11}$$

$$C_X^a = (I - KH)C_X^f \tag{12}$$

To iterate the prediction-analysis cycle and obtain a converged state vector, the X_i^a becomes the new X_i^f is the next step. The iteration terminates when the error covariance of the state vector drops to a predefined threshold or the maximum step is reached. The flow chart of utilizing the EnKF method to improving the near-wall velocity prediction in BFS flow is shown in Fig. 1.

The DAFI framework is adopted in this study to realize the iterative EnKF and the operation of the FOAM fields. Further details of the DAFI framework and the EnKF can be found in the original publication by the DAFI developers(Carlos A. Michelén Ströfer et al., 2021).

TEST CASE DESCRIPTION

Physical Model

The test case is the standard BFS flow with physics settings corresponding to the experiments by Driver and Seegmiller (Driver and Seegmiller, 1985). A sketch of the physics domain is shown in Fig. 2. The height of the step is $H = 0.0127m$. The distance between the bottom wall and the top wall is 9H. The x length before the step is $x_{in} = 130H$, with a startup length of $x_{start} = 20H$, and after the step is $x_{out} = 50H$. The uniform inlet velocity is $U = 44.2m/s$, and the kinematic viscosity is $\nu = 1.56 \times 10^{-5} s/m^2$, thus the Reynolds number based on the step height is $Re_H = 3.6 \times 10^4$.

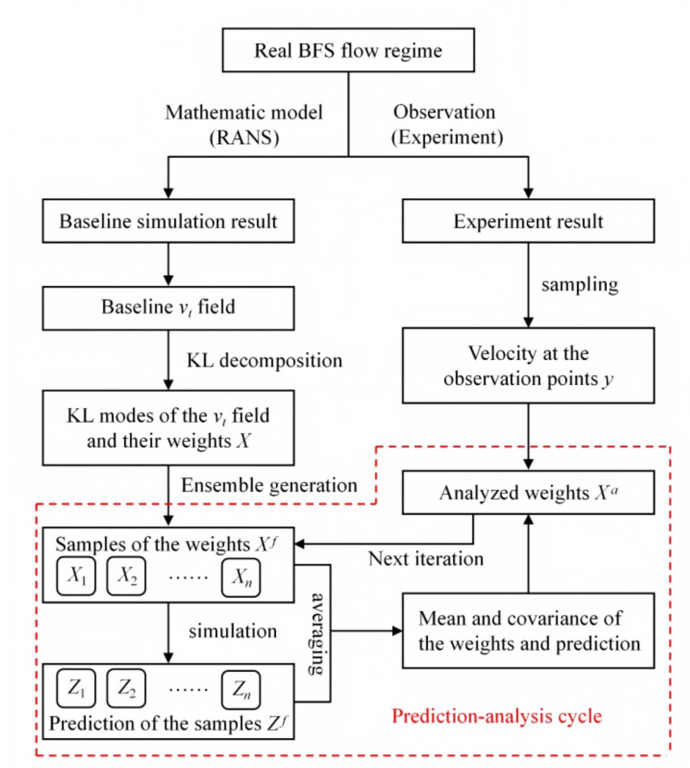


Fig. 1 Flow chart of the EnKF data assimilation on BFS flow

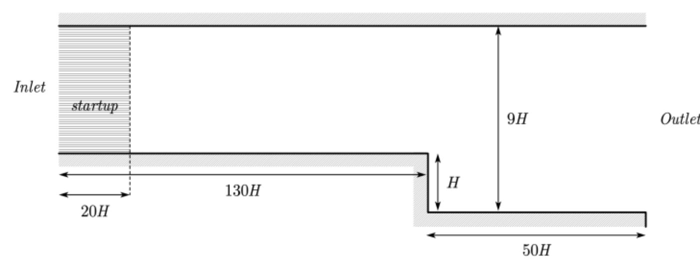
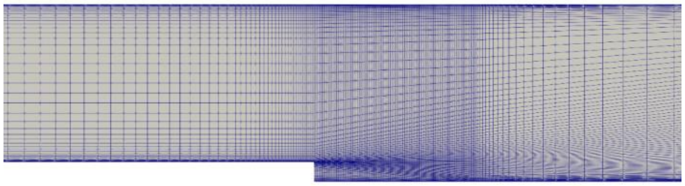


Fig. 2 A sketch of the physics domain

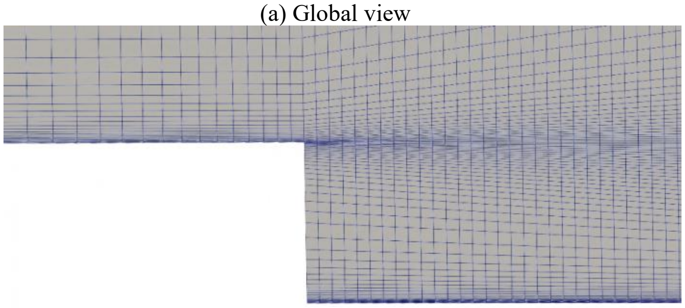
CFD Settings

This subsection discusses the computational settings of numerical simulation. For the boundary conditions, a no-slip boundary condition is imposed on the lower and upper walls, except for the startup region. Symmetry boundary conditions are imposed on the lower and upper walls of the startup region to to improve the numerical stability of the simulation. A velocity inlet boundary condition is imposed at the inlet, and a pressure outlet boundary condition is used at the outlet.

The computational mesh is established utilizing the blockMesh tool in OpenFOAM. The computational mesh consists of 20540 cells, with local mesh refinement near the wall boundaries and at the step corner. The current mesh is consistent with the OpenFOAM test case. Therefore, the baseline CFD result is reliable.



(a) Global view



(b) Mesh near the backward facing step

Fig. 3 Computational mesh of the backward facing step flow

Data Assimilation Settings

The data assimilation process requires the definition of the state vector extracted from the numerical model and the observations of the real physics process. To construct the state vector, KL decomposition is performed on the computational mesh of the BFS flow described above. 17 KL modes are retained to capture 99.5% of the total energy of the covariance matrix. The observations consist of velocity measurements at eight discrete points from the experimental dataset. The detail locations of the observation points are shown in Fig. 4. Two different standard deviation of the observations are considered: $\sigma_Y = 1\times10^{-3}$, 1×10^{-6}.

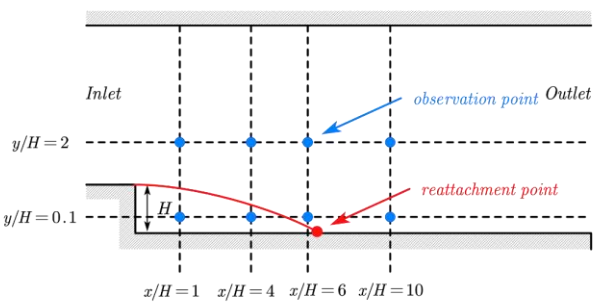


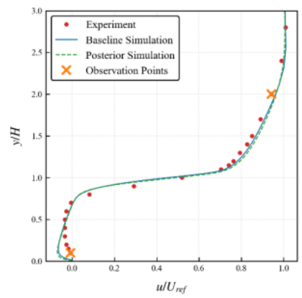
Fig. 4 Locations of the observation points in the BFS domain

To initialize the data assimilation process, 32 samples are generated by the Monte Carlo sampling of the weights of the KL modes. Each sample is propagated with the prediction and analysis iteration cycle. The iterative data assimilation process has a maximum number of five iteration steps.

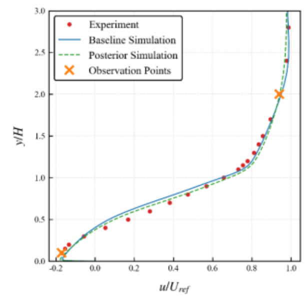
RESULTS AND DISCUSSIONS

Velocity and ν_t profile comparison

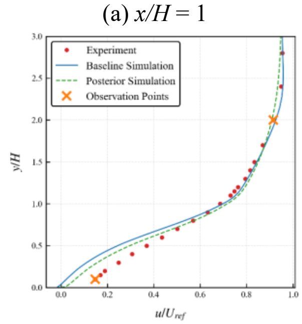
At first, the standard deviation of the observations from the experiments is set to $\sigma_Y = 1\times10^{-3}$. Fig. 5 shows the u component of the near-wall velocity with $\sigma_Y = 1\times10^{-3}$. The red dots represent the experimental data, and the orange crosses denote the observation points. The solid blue line represents the baseline prediction of the near-wall velocity using the k–ω SST model, and the dashed green line represents the posterior prediction of the near-wall velocity after the data assimilation process. The results show that the baseline simulation accurately predicts the streamwise velocity in the far-wall region from the lower wall. However, the baseline simulation struggles to predict the near-wall velocity, particularly in the vicinity of the reattachment point. At x/H = 1, the predictions of the baseline simulation and posterior simulation are nearly the same. At x/H = 4, it is observed that although the observation points are at y/H = 0.1 and 2, the velocity prediction within the range of y/H = 0.3~1.1 improves. It indicates that the KL decomposition with the SE kernel and the EnKF based data assimilation can capture the correlations between different locations in the ν_t field, thus leading to improved CFD predictions of the entire flow field via sparse observations. Results at x/H = 6 and x/H = 10 further strengthen the conclusion, where consistent improvements of the near-wall velocity are observed. The v component of near-wall velocity results with $\sigma_Y = 1\times10^{-3}$ are shown in Fig. 6. Prominent improvement of the normal velocity at x/H = 1,4 and y/H > 1 is observed. However, the improvement in other regions is limited. In general, the assimilation performance of the EnKF with $\sigma_Y = 1\times10^{-3}$ is moderately satisfactory, and there is still room for improvement.



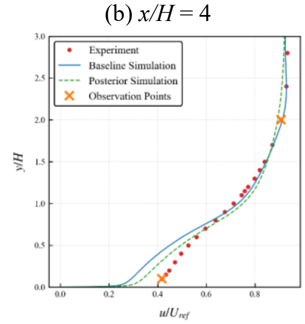
(a) x/H = 1



(b) x/H = 4

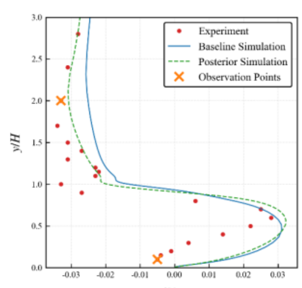


(c) x/H = 6

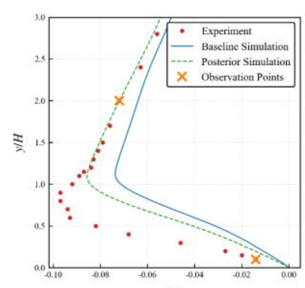


(d) x/H = 10

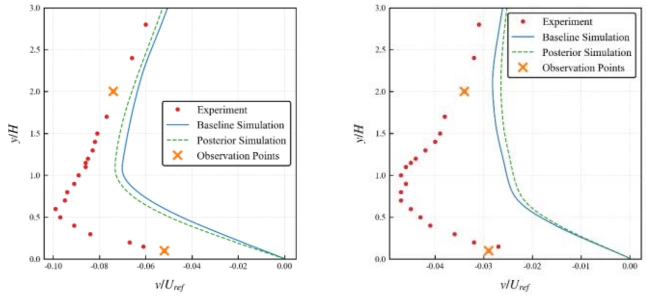
Fig. 5 u component of near-wall velocity results with $\sigma_Y = 1\times10^{-3}$



(a) x/H = 1



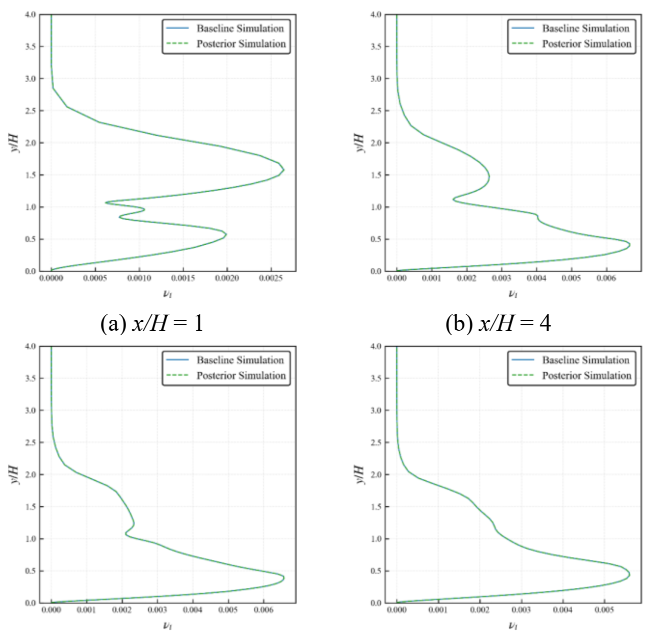
(b) x/H = 4



(c) $x/H = 6$ (d) $x/H = 10$

Fig. 6 v component of near-wall velocity results with $\sigma_Y = 1\times10^{-3}$

Fig. 7 compares the ν_t profiles of the baseline simulation and the posterior simulation with $\sigma_Y = 1\times10^{-3}$. It is found that the ν_t field is largely unchanged after the data assimilation process. The underlying reason for the phenomenon is that the standard deviation of the prediction is much smaller than the set deviation of the observations. Fig. 8 shows the averaged σ_Z during each round of the assimilation process. The result shows that the values of σ_Z are on the order of 10^{-6}. It leads to a very small value of the Kalman gain, and then results in a limited correction range of the state vector. This also explains the reason why the assimilation performance of the EnKF with the present setting $\sigma_Y = 1\times10^{-3}$ has room for improvement.



(a) $x/H = 1$ (b) $x/H = 4$

(c) $x/H = 6$ (d) $x/H = 10$

Fig. 7 ν_t profile result with $\sigma_Y = 1\times10^{-3}$

Therefore, a smaller standard deviation of the observation $\sigma_Y = 1\times10^{-6}$, which is smaller than the deviation of the predictions, is further tested. The results of $\sigma_Y = 1\times10^{-6}$ are shown in Figs. 9-11. It is observed from Fig. 9 that the u component of near-wall velocity result of the posterior simulation is in much better agreement with the experimental data compared to the baseline simulation at x/H = 4,6,10. However, an "overfitting" issue is observed in the far-wall region. The predictions from the posterior simulation are less accurate than those from the baseline simulation. The underlying mechanism for this behavior is rooted in the fundamental principles of KL decomposition. When the weights of the KL modes are influenced by the error of the numerical prediction near the wall, the square-exponent kernel of the KL decomposition will spread the influence to the neighboring locations. Therefore, when the ν_t field changes in the region where the standard k–ω SST model provides a good prediction, the result may worsen by the data assimilation process. This also explains the result of the v component near-wall velocity in Fig. 10. While at x/H = 1, the error of numerical simulation is further corrected, at x/H = 10, the error is further amplified. Fig. 11 shows the ν_t field comparison between the baseline simulation and the posterior simulation. Comparing to Fig. 7, the ν_t field with $\sigma_Y = 1\times10^{-6}$ has a significant difference from the baseline simulation. Furthermore, the corrected ν_t field retains the physical structure of the original field, indicating that the assimilation framework combining the EnKF and KL decomposition does not deviate significantly from fundamental physical principles.

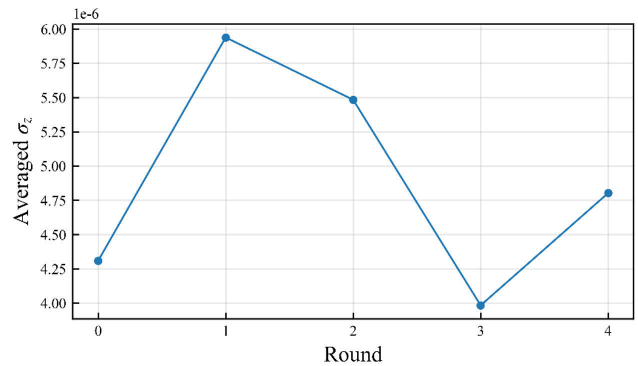
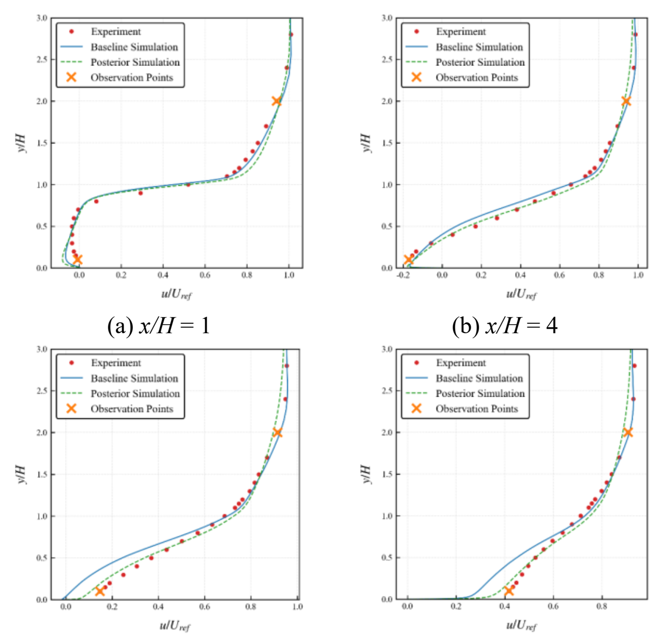


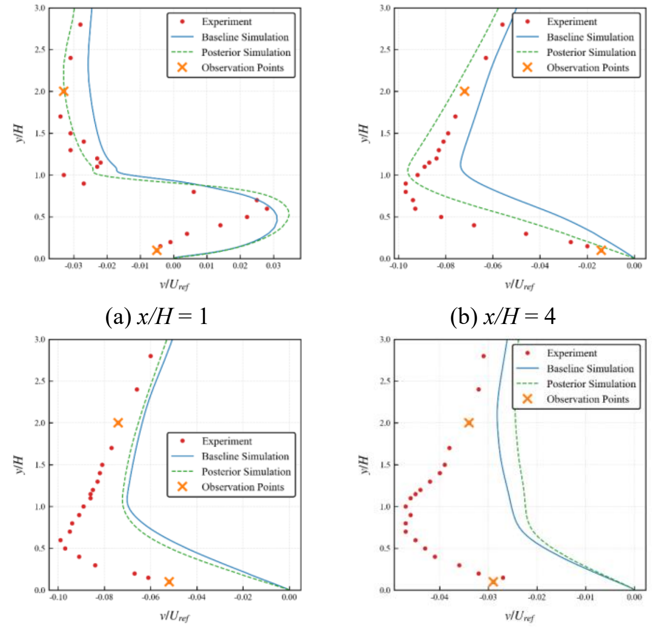
Fig. 8 Averaged σ_z during the assimilation process



(a) $x/H = 1$ (b) $x/H = 4$

(c) $x/H = 6$ (d) $x/H = 10$

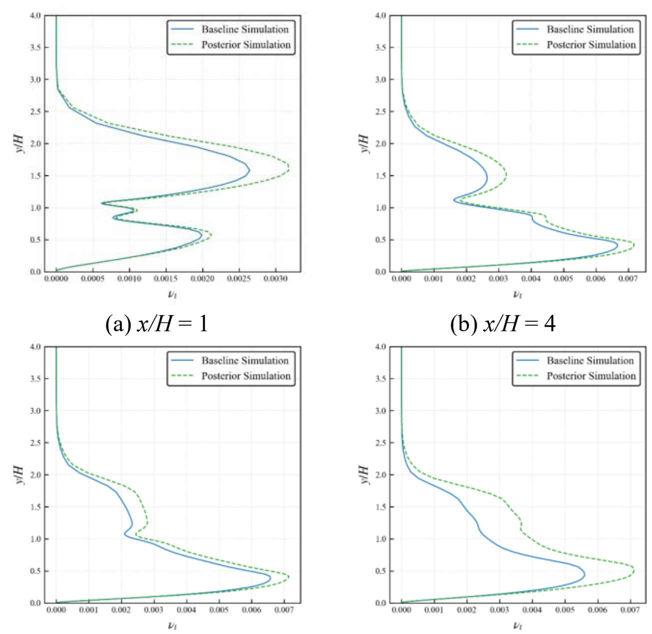
Fig. 9 u component of near-wall velocity results with $\sigma_Y = 1\times10^{-6}$



(a) $x/H = 1$ (b) $x/H = 4$

(c) $x/H = 6$ (d) $x/H = 10$

Fig. 10 v component of near-wall velocity results with $\sigma_Y = 1\times10^{-6}$



(a) $x/H = 1$ (b) $x/H = 4$

(c) $x/H = 6$ (d) $x/H = 10$

Fig. 11 ν_t profile result with $\sigma_Y = 1\times10^{-6}$

Quantitative RMSE Comparison

To quantitatively compare the assimilation performance for the two different observation error standard deviations, the root-mean-square error (RMSE) between the predictions from the baseline/posterior simulations and the experimental data is calculated at the four vertical profiles. The definition of RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

where n represents the number of points on a certain profile, y_i denotes the velocity value from the experiment data, and $\hat{y}_i$ denotes the velocity value from the simulation results. The detailed results of RMSE are shown in Table. 1~2. Each column lists the RMSE of baseline simulation, posterior simulation with $\sigma_Y = 1\times10^{-3}$, and posterior simulation with $\sigma_Y = 1\times10^{-6}$ on a certain profile. The last column is the average RMSE improvement of two posterior simulations. The average RMSE improvements of the streamwise and normal velocity component with $\sigma_Y = 1\times10^{-6}$ are 33.78% and 13.01%, which are 41.1% and 22.9% higher than the result with $\sigma_Y = 1\times10^{-3}$, respectively.

The result indicates that lower standard deviation of the observation leads to lower RMSE based on the experiment, which conforms to the principle and the target of EnKF-based data assimilation. Moreover, the value is an average of four distinct profiles. This means the value is strongly brought down by some locations where present EnKF assimilation performs poorly. On the other hand, the present EnKF assimilation is indeed not effective for some locations, and even have negative effects. To be exact, the assimilation works well at regions where the absolute errors of the baseline simulation were large. The main reason of this phenomenon is that the analysis step of the assimilation uses the absolute error, therefore, the assimilation tends to preferentially correct the large absolute error locations to get an optimum estimation of the entire field. Nevertheless, the results show that ensemble-based data assimilation has an overall positive effect on the correction of the numerical simulation.

Table. 1 RMSE comparison of the u component between the baseline simulation and posterior simulation with $\sigma_Y = 1\times10^{-3}$, 1×10^{-6}

	x/H=1	x/H=4	x/H=6	x/H=10	Average RMSE improvement
baseline	0.041	0.047	0.075	0.061	-
$\sigma_Y = 1\times10^{-3}$	0.041	0.034	0.050	0.039	23.94%
$\sigma_Y = 1\times10^{-6}$	0.046	0.033	0.029	0.026	33.78%

Table. 2 RMSE comparison of the v component between the baseline simulation and posterior simulation with $\sigma_Y = 1\times10^{-3}$, 1×10^{-6}

	x/H=1	x/H=4	x/H=6	x/H=10	Average RMSE improvement
baseline	0.015	0.026	0.033	0.020	-
$\sigma_Y = 1\times10^{-3}$	0.013	0.019	0.030	0.020	10.58%
$\sigma_Y = 1\times10^{-6}$	0.014	0.014	0.029	0.022	13.01%

CONCLUSIONS

In this study, data assimilation based on the ensemble Kalman filter is performed on the ν_t field of the backward facing step (BFS) flow. The observations are extracted from the experimental dataset of Driver and Seegmiller (Driver and Seegmiller, 1985). Two different standard deviations of the observation data $\sigma_Y = 1\times10^{-3}$, 1×10^{-6} are chosen and compared to study the influence of the standard deviation on the assimilation results. The main conclusions of this study are as follows:

(1) The data assimilation framework based on the ensemble Kalman filter and the KL decomposition method works as an effective way to improve the prediction of the RANS simulation using sparse experimental data. The maximum improvement of the u and v component of the velocity profile is 33.78% and 13.01%.

(2) The standard deviation of the observation data has a significant influence on the effect of the data assimilation. Smaller standard deviation means having more faith in experiment data, resulting in further reduced RMSE of the posterior velocity profile prediction. The average RMSE improvements of the streamwise and normal velocity component with $\sigma_Y = 1\times10^{-6}$ are 41.1% and 22.9% higher than the result with $\sigma_Y = 1\times10^{-3}$, respectively.

(3) The present data assimilation framework has two key limitations. Firstly, it performs poorly and even has negative effects in regions where absolute error of the baseline simulation is small. Secondly, the present simulation is only based on the spatial correlation of the ν_t field, thus uncapable to fully capture the complete underlying physical process.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Program of China (2025YFF0519800), to which the authors are most grateful.

REFERENCES

Aulakh, DJS, Yang, X, and Maulik, R (2024). "Robust experimental data assimilation for the Spalart-Allmaras turbulence model," *Phys Rev Fluids*, 9(8), 084608.

Carlos A. Michelén Ströfer, CAMS, Xin-Lei Zhang, X-LZ, and Heng Xiao, HX (2021). "DAFI: An Open-Source Framework for Ensemble-Based Data Assimilation and Field Inversion," *CICP*, 29(5), 1583–1622.

Driver, DM, and Seegmiller, HL (1985). "Features of a reattaching turbulent shear layer in divergent channelflow," *AIAA Journal*, 23(2), 163–171.

Ge, G, Zhang, W, Xie, B, and Li, J (2024). "Turbulence model optimization of ship wake field based on data assimilation," *Ocean Engineering*, 295, 116929.

Grabe, C, Jäckel, F, Khurana, P, and Dwight, RP (2023). "Data-driven augmentation of a RANS turbulence model for transonic flow prediction," *International Journal of Numerical Methods for Heat & Fluid Flow*, 33(4), 1544–1561.

Lakshmivarahan, S, and Stensrud, D (2009). "Ensemble Kalman filter," *Control Systems, IEEE*, 29, 34–46.

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

Menter, FR, Kuntz, M, and Langtry, R (2003). "Ten Years of Industrial Experience with the SST Turbulence Model," *Turbulence, Heat and Mass Transfer*,.

Michelén Ströfer, CA, Zhang, X-L, Xiao, H, and Coutier-Delgosha, O (2020). "Enforcing boundary conditions on physical fields in Bayesian inversion," *Computer Methods in Applied Mechanics and Engineering*, 367, 113097.

Wang, J-X, Wu, J-L, and Xiao, H (2017). "Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data," *Phys Rev Fluids*, 2(3), 034603.

Wu, J, Li, J, Qiu, X, Xie, X, and Liu, Y (2021). "Machine learning based Reynolds averaged simulation of backward-facing step flows at different Reynolds numbers," *Mod Phys Lett B*, 35(25), 2150430.

Zhang, X-L, Xiao, H, Wu, T, and He, G (2022). "Acoustic Inversion for Uncertainty Reduction in Reynolds-Averaged Navier–Stokes-Based Jet Noise Prediction," *AIAA Journal*, 60(4), 2407–2422.

Zhu, L, Zhang, W, Kou, J, and Liu, Y (2019). "Machine learning methods for turbulence modeling in subsonic flows around airfoils," *Physics of Fluids*, 31(1), 015105.