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

Multi-Resolution Data Fusion for Turbulent Flows in Large Eddy Simulation Using a Kalman Filter

Yujia Liu[1], Haiting Yu[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 turbulent flows remains challenging due to the high computational cost of large eddy simulation (LES) at fine resolutions. In this study, a Kalman-filter-based multi-resolution data fusion framework is proposed within an LES context, where sparse high-resolution information is fused with downsampled low-resolution flow fields derived from the same high-fidelity data. Gaussian process regression (GPR) is employed to propagate the influence of sparse high-resolution data and model spatial correlations during the fusion process. The methodology is first validated using a prescribed analytical field, in which noisy downsampled data represent low-resolution fields and sparsely sampled coefficients emulate high-resolution information. The framework is then applied to turbulent channel flow, demonstrating that the proposed approach effectively reconstructs flow features and improves accuracy compared to low-resolution data alone, highlighting its potential for efficient multi-resolution LES data fusion.

KEY WORDS: multi-resolution, data fusion, Kalman filter, large eddy simulation (LES), Gaussian process regression (GPR), channel flow

INTRODUCTION

Turbulence remains one of the central challenges in computational fluid dynamics (CFD) due to its intrinsic multi-scale, nonlinear and non-equilibrium nature. Despite decades of theoretical, experimental and numerical research, the accurate and efficient prediction of turbulent flows in realistic engineering configurations is still an open problem. From a practical standpoint, the choice of turbulence modelling strategy is largely governed by the trade-off between physical fidelity and computational affordability, which has led to the coexistence of multiple modelling paradigms in current CFD practice, most notably Reynolds-averaged Navier-Stokes (RANS) methods and large-eddy simulation (LES).

With the continuous growth of available computational resources and the steady progress of numerical algorithms, high-fidelity approaches such as direct numerical simulation (DNS) and wall-resolved LES have become increasingly feasible for canonical flows and simplified geometries. However, their computational cost remains prohibitive for most industrial applications involving high Reynolds numbers, complex geometries and parametric studies. Consequently, industrial CFD simulations still rely predominantly on RANS solvers combined with turbulence models, a situation that is expected to persist in the foreseeable future, particularly for outer-loop applications such as design optimization and uncertainty quantification (Slotnick et al., 2014).

Most turbulence models employed in RANS frameworks are linear eddy-viscosity models, such as the k-ε (Launder and Sharma, 1974) and Spalart-Allmaras (Spalart and Allmaras, 1992) models, which are built upon the weak-equilibrium assumption and the Boussinesq hypothesis (Pope, 2001). While these assumptions enable computational efficiency and numerical robustness, they also limit the ability of RANS models to accurately represent complex turbulent phenomena, including strong anisotropy, separation, rotation and secondary flows. More advanced closures, such as Reynolds stress transport models, nonlinear eddy-viscosity models and explicit algebraic stress models, have been proposed to address these deficiencies (Launder et al., 1975; Spalart, 2000). Although such models can offer improved physical realism in specific flow regimes, their wider adoption has been hindered by increased computational cost, implementation complexity and reduced robustness across a broad range of applications. As a result, RANS modelling in practice often involves a compromise between predictive performance and engineering usability (Xiao and Cinnella, 2018).

In parallel, LES has emerged as an attractive alternative to Reynolds-averaged approaches by explicitly resolving the large, energy-containing turbulent structures while modelling only the smaller, more universal scales. LES has demonstrated clear advantages in capturing unsteady flow features, flow separation, and complex turbulence dynamics that are often inadequately represented by RANS models. However, for wall-bounded flows at high Reynolds numbers, wall-resolved LES requires extremely fine grid resolution in the near-wall region, leading to prohibitively high computational costs. In practical configurations, wall-resolved LES often demands that the vast majority of grid points be concentrated near solid boundaries, which severely limits its applicability in industrial-scale simulations (Piomelli, 2008). To reduce this computational burden, wall-modelled LES has been introduced, in which the near-wall region is treated using simplified models while LES is applied in the outer flow. Although this strategy significantly reduces the resolution requirements, its accuracy and robustness strongly depend on the choice of wall model and its interaction with the resolved outer-layer turbulence. In particular, issues related to model sensitivity, grid dependence, and predictive reliability remain active research topics.

In view of these limitations, there has been growing interest in augmenting conventional CFD frameworks with data-driven modelling strategies. Rather than relying solely on empirical assumptions and fixed functional forms, data-driven turbulence

modelling aims to leverage high-fidelity simulation data or experimental measurements to infer improved turbulence closures or model corrections. Various approaches have been proposed for data-driven turbulence modelling, including direct a priori training methods (Ling et al., 2016), adjoint-based differentiable learning frameworks (Holland et al., 2019), ensemble-based gradient approximation methods (Ströfer et al., 2021). By exploiting modern machine-learning techniques, these approaches can explore a much richer functional space and capture complex, nonlinear relationships that are difficult to represent using traditional models. As a result, data-driven methods have shown promise in enhancing the representational capability of turbulence models and extending their applicability beyond classical calibration regimes.

However, purely data-driven turbulence modelling also faces fundamental challenges, particularly in terms of data availability, generalization and consistency with governing equations. Despite the continuous development of turbulence models and numerical methods, significant discrepancies between CFD predictions and experimental observations are still frequently encountered in practical turbulent flow simulations. These discrepancies stem not only from modelling assumptions inherent in RANS and LES frameworks, but also from epistemic uncertainties associated with turbulence closures, boundary conditions, and model parameters. As a result, improving the reliability and predictive confidence of CFD simulations, especially for complex, high-Reynolds-number flows, remains a critical challenge. Moreover, models trained in an a priori manner may perform well in isolation but exhibit degraded performance when embedded in CFD solvers due to the intrinsic ill-conditioning of the RANS and LES equations. These limitations highlight the need for a more integrated framework in which data-driven models are developed and applied in close conjunction with the underlying flow solvers.

To address the persistent deficiencies in turbulence modelling and to enhance the reliability of CFD predictions, an increasingly important strategy is to incorporate experimental or observational data directly into the computational framework. While improvements in turbulence closures can reduce model-form errors to some extent, they alone are often insufficient to guarantee reliable predictions in complex, high-Reynolds-number flows, where modelling assumptions may be violated and uncertainties can accumulate. In contrast, experimental measurements provide direct information about the true flow state and can serve as an external constraint to correct model biases and reduce prediction uncertainty. The systematic integration of model predictions and observational data is commonly referred to as data fusion or data assimilation. In this framework, neither the numerical model nor the data is assumed to be perfect; instead, both are treated as uncertain and complementary sources of information. Data assimilation seeks to combine these two sources in an optimal statistical sense, yielding an improved estimate of the flow state or model parameters that is more accurate and reliable than either source alone. Several open-source software packages have been developed for data assimilation, which can be broadly classified into ensemble-based methods, such as the ensemble Kalman filter, and variational, gradient-based approaches. Representative platforms include the NCAR Data Assimilation Research Testbed (DART), a mature and widely used ensemble data assimilation framework with strong parallel capabilities but a relatively high learning cost (Anderson et al., 2009), and OpenDA, a general-purpose toolkit supporting both ensemble and variational methods that has been successfully coupled with engineering solvers such as OpenFOAM, albeit with similar accessibility challenges (Verlaan et al., 2010). DATeS provides a more lightweight, Python-based environment for testing data assimilation algorithms, supporting both ensemble and variational formulations (Attia and Sandu, 2019).

Among the various data assimilation techniques, Kalman filter-based methods provide a rigorous and well-established statistical framework for data fusion. The classical Kalman filter formulates the evolution of the system state and its uncertainty through a prediction–correction cycle, in which the model forecast is first advanced in time and then updated using available observations. In practical turbulent flow applications, the direct use of classical Kalman filtering is often infeasible due to the prohibitive cost of explicitly forming and storing large covariance matrices. To address this issue, reduced and approximate representations of the covariance structure are commonly adopted, enabling the Kalman filtering concept to be applied efficiently to large-scale flow problems. In this context, statistical learning techniques such as Gaussian process regression can be naturally integrated into the Kalman filtering framework to model spatial correlations and propagate the influence of sparse or high-resolution information across the flow field. Owing to its non-intrusive nature and its ability to accommodate sparse, indirect, and multi-resolution data, Kalman-filter-based data fusion has become an effective tool for improving turbulent flow predictions within large eddy simulation frameworks. By combining physics-based LES solutions with additional information in a probabilistic manner, this approach provides a robust pathway to enhance predictive accuracy and reduce uncertainty in complex turbulent flow simulations.

THE MATHEMATICAL FORMULATIONS

Wall-resolved large-eddy simulation (WRLES)

The high-fidelity WRLES methodology is based on the finite volume method (FVM). The governing equations for LES are derived by applying a spatial filter to the incompressible N-S equations.

$$\frac{\partial \tilde{u}_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial \tilde{u}_i \tilde{u}_j}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \tilde{p}}{\partial x_i} + \nu \frac{\partial^2 \tilde{u}_i}{\partial x_j \partial x_j} + \frac{\partial \tau_{ij}^{sgs}}{\partial x_j} \tag{2}$$

The filtered equations include the momentum equation and continuity equation, where $\tilde{u}_i$ (for i = 1, 2, 3) represents the filtered velocity component in the x_i direction, $\tilde{p}$ denotes the filtered pressure field, ν is the kinematic viscosity of the fluid, and τ_{ij}^{sgs} is the SGS stress tensor.

$$\tau_{ij}^{sgs} = 2\nu_{sgs}\widetilde{S}_{ij} + \frac{1}{3}\tau_{kk}^{sgs}\delta_{ij} \tag{3}$$

The SGS stress term is modeled as the deviatoric part of the stress tensor, expressed in terms of the SGS eddy viscosity ν_{sgs} and the resolved strain-rate tensor $\widetilde{S}_{ij}$. Here, $\widetilde{S}_{ij}$ is defined as one-half the sum of the velocity gradient and its transpose, while δ_{ij} represents the Kronecker delta.

In this study, the wall-adapting local eddy-viscosity (WALE) model proposed by Nicoud and Ducros (1999) is employed to compute the SGS eddy viscosity.

$$\nu_{sgs} = (C_w\Delta)^2 \frac{(S_{ij}^d S_{ij}^d)^{3/2}}{(\widetilde{S}_{ij}\widetilde{S}_{ij})^{5/2} + (S_{ij}^d S_{ij}^d)^{5/4}} \tag{4}$$

The WALE model formulation involves the model constant C_w (equal to 0.325). Δ is the filter width (taken as the cube root of the local cell volume). And S_{ij}^d is the traceless symmetric part of the square of the velocity gradient tensor. The WALE model effectively accounts for both the strain and rotation rates of the smallest resolved turbulent fluctuations.

Kalman Filter

In many engineering and physical systems, the quantities of interest cannot be measured directly or continuously, and the available observations are often sparse, noisy, and incomplete. At the same time, the evolution of such systems is typically governed by known physical laws, which can be expressed in the form of mathematical models. However, these models are inevitably imperfect due to modelling assumptions, parameter uncertainties, and unresolved physical

processes. As a result, neither numerical models nor experimental measurements alone are sufficient to provide reliable and accurate descriptions of the system behaviour.

To address this fundamental challenge, it is natural to seek a framework that combines model predictions with observational data in a systematic and statistically consistent manner. Kalman filtering provides such a framework by formulating the evolution of a physical system as a state transition process and embedding it within a probabilistic state-space representation. In this context, the system state is assumed to be representable by a vector $X \in R^n$ in an n-dimensional space, which encapsulates the essential variables describing the system at a given time. For convenience of description, the following assumptions are made: (i) the state transition of the physical system can be described as a discrete-time stochastic process; (ii) the system state is influenced by control inputs; (iii) both the system state and the observation process are inevitably affected by noise; and (iv) the system state is not directly observable.

Under these assumptions, the system state variable is defined as $X_k \in R^n$, the control input as U_k, and the process noise as W_k. The stochastic state transition equation of the system can then be written as:

$$X_k = FX_{k-1} + BU_{k-1} + W_{k-1} \tag{5}$$

where the observation variable is defined as $Z_k \in R^m$ with observation noise V_k, leading to the measurement equation:

$$Z_k = HX_k + V_k \tag{6}$$

The process noise W_k and the observation noise V_k are assumed to be mutually independent, zero-mean Gaussian white noises, with covariance matrices Q and R, respectively, i.e.

$$W_k \sim N(0, Q) \tag{7}$$

$$V_k \sim N(0, R) \tag{8}$$

Here, F, B, and H are collectively referred to as the state transition matrices, which serve as adjustment coefficients in the state evolution process and are derived from the mathematical model of the system. Starting from the mathematical model of a physical system, the computational framework of the Kalman filter can be derived, which consists of two main components: the time-update (prediction) equations and the measurement-update (correction) equations. For clarity, the following definitions are introduced. Let $\hat{X}_k \in R^n$ denote the a priori (predicted) state estimate at time step k, given the system information up to time k-1, and $\hat{X}_k$ denote the a posteriori (updated) state estimate at time step k, after incorporating the measurement Z_k. Based on these definitions, the a priori and a posteriori estimation errors are defined as:

$$E_k^- = X_k - \hat{X}_k^- \tag{9}$$

$$E_k = X_k - \hat{X}_k \tag{10}$$

with their corresponding error covariance matrices given by:

$$P_k^- = E(E_k^- E_k^{-T}) \tag{11}$$

$$P_k = E(E_k E_k^T) \tag{12}$$

The Kalman filter constructs the posterior state estimate as a linear combination of the prior estimate and the residual between the observation and its prediction, expressed as:

$$\hat{X}_k = \hat{X}_k^- + K(Z_k - \hat{X}_k^-) \tag{13}$$

Here, the term $(Z_k - \hat{X}_k^-)$ represents the measurement residual, which quantifies the discrepancy between the predicted observation and the actual measurement. The matrix $K \in R^{n \times m}$, known as the Kalman gain, determines the relative weighting between the model prediction and the measurement. Its role is to minimize the posterior estimation error covariance. By minimizing P_k with respect to K_k, one form of the matrix K:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \tag{14}$$

The determination of the Kalman gain is one of the key steps in filter design, as it has a significant impact on estimation accuracy and convergence efficiency.

The Kalman filter operates through two sequential processes: prediction and correction. In the prediction step, the time-update equations are used to propagate the state estimate and its uncertainty forward in time, yielding a prior estimate for the current state. In the correction step, the measurement-update equations incorporate the current observation to refine the prior estimate and obtain an improved posterior estimate. This recursive procedure is commonly referred to as the prediction-correction framework.

For a discrete-time linear system, the Kalman filter equations are given as follows. Time-update (prediction) equations:

$$\hat{X}_k^- = F\hat{X}_{k-1} + B\hat{U}_{k-1} \tag{15}$$

$$P_k^- = FP_{k-1}F^T + Q \tag{16}$$

Measurement-update (correction) equations:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-2} \tag{17}$$

$$\hat{X}_k = \hat{X}_k^- + K_k(Z_k - H\hat{X}_k^-) \tag{18}$$

$$P_k = (I - K_k H)P_k^- \tag{19}$$

In the above equations, $F \in R^{n \times n}$ is the state transition matrix, B is the control input matrix, $H \in R^{m \times m}$ is the observation matrix mapping the state space to the observation space, Q and R are the process noise and observation noise covariance matrices, respectively, I is the identity matrix, and K_k is the Kalman gain.

THE RESULTS AND ANALYSIS

Simple function experiment

This section presents a simple function experiment designed to verify the effectiveness of Kalman-filter-based data fusion in a fully controlled environment. Owing to the availability of an analytical solution, both qualitative visualization and quantitative error assessment can be performed.

The reference field is defined by an analytical function with an explicit closed-form expression:

$$\begin{aligned} P_{true}(x,y) = {} & 1 - 0.2\tanh[(x-2\pi)^2 + (y-2\pi)^2] \\ & + 0.15\sin(2x)\cos(1.5y) + 0.1e^{-\frac{(y-0.5)^2}{2}} \\ & + 0.05(\frac{x}{4\pi}) + 0.03\sin(3x)\cos(2y) \end{aligned} \tag{20}$$

Prior to visualization and subsequent analysis, the analytical reference field is normalized in order to ensure numerical stability and to facilitate a consistent comparison with the noisy numerical and observational data. The normalization is performed according to:

$$P_{normalized}(x,y) = 0.8 + 0.4 \cdot \frac{P_{true} - P_{\min}}{P_{\max} - P_{\min}} \tag{21}$$

where P_{min} and P_{max} denote the minimum and maximum values of the analytical field over the computational domain. All subsequent operations, including noise injection, sampling, and data fusion, are conducted on the normalized field.

To provide an intuitive understanding of the reference solution, the analytical field is visualized over the entire computational domain.

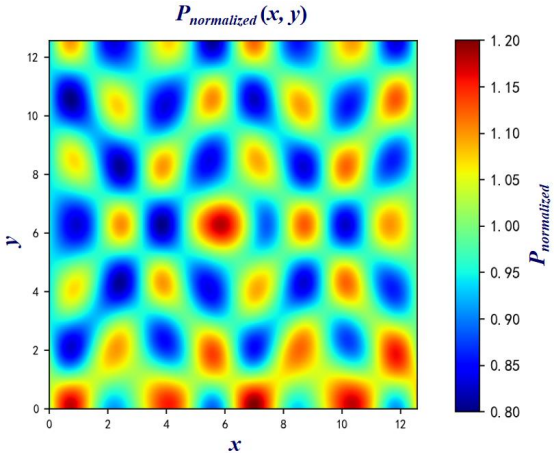

Fig.1 Visualization of the analytical reference field $P_{normalized}$

To mimic the uncertainty introduced by numerical discretization errors and turbulence modelling deficiencies in practical CFD simulations, a synthetic numerical field is constructed by perturbing the analytical reference field with additive process noise. Specifically, the numerical field is defined as:

$$P_{cfd}(x,y) = P_{normalized}(x,y) + w(x,y) \tag{22}$$

where $w(x,y)$ denotes the process noise, which is assumed to follow a prescribed stochastic distribution with user-defined statistical properties. The process noise is spatially distributed over the entire computational domain, representing unresolved physics and modelling errors. The noise amplitude and correlation characteristics are selected to emulate realistic levels of numerical uncertainty encountered in CFD simulations. In the present study, the process noise is further decomposed into two distinct components, namely a systematic noise and a random noise, in order to better represent different sources of uncertainty commonly encountered in CFD simulations. Accordingly, the process noise field $w(x,y)$ is expressed as:

$$w(x,y) = \sigma_{systematic}(x,y) + \sigma_{random}(x,y) \tag{23}$$

where $\sigma_{systematic}(x,y)$ denotes the systematic noise component, which accounts for structured modelling errors such as turbulence model bias and numerical discretization effects, and $\sigma_{random}(x,y)$ represents the random noise component associated with stochastic fluctuations and unresolved small-scale dynamics. The specific systematic noise is prescribed as:

$$\sigma_{systematic}(x,y) = 0.02\sin(0.5x) + 0.01\cos(0.3y) \tag{24}$$

which introduces spatially correlated perturbations with a prescribed amplitude and characteristic length scale. The random noise is defined as:

$$\sigma_{random}(x,y) \sim N(0, {\sigma_{model}}^2) \tag{25}$$

where σ_{model} denotes the variance of the random process noise. The two noise components are assumed to be statistically independent. This decomposition allows the process noise model to simultaneously capture both structured and unstructured uncertainties in the numerical solution, providing a more realistic representation of modelling errors in CFD and enabling a clearer assessment of the robustness of the data fusion procedure.

The resulting noisy numerical field is visualized to illustrate the deviation from the analytical reference.

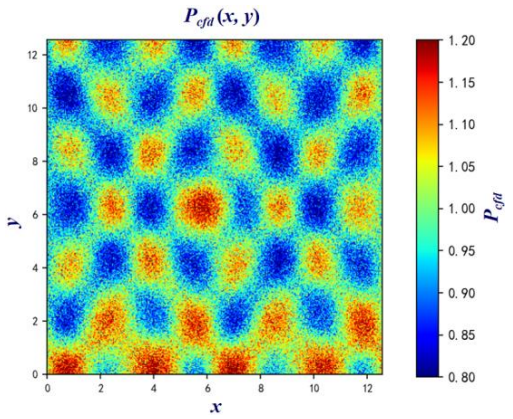

Fig.2 Visualization of the numerical field with process noise P_{cfd}

To emulate experimental measurements in a more realistic manner, the observational data are generated through a non-uniform, importance-based sampling strategy. Instead of sampling uniformly over the computational domain, observation locations are preferentially distributed in regions where the analytical reference field exhibits large spatial gradients. This reflects the practical consideration that experimental measurements are often concentrated in flow regions with strong variations or complex dynamics.

Specifically, a spatial sampling probability density function $P_{meas}(x,y)$ is defined based on the magnitude of the gradient of the analytical reference field as:

$$P_{meas}(x,y) = \frac{\| \nabla P_{normalized}(x,y) \|}{\iint \| \nabla P_{normalized}(x,y) \| \, dxdy} \tag{26}$$

where $\nabla P_{normalized}(x,y)$ denotes the spatial gradient of the analytical pressure field. This normalization ensures that $P_{meas}(x,y)$ forms a valid probability density function over the domain.

Observation locations $\{(x_i, y_i)\}_{i=1}^{N}$ are then drawn according to the distribution $P_{meas}(x,y)$, such that regions with larger pressure gradients are sampled with higher probability. This gradient-weighted sampling strategy enhances the information content of the observations by focusing measurements in dynamically important regions.

Given the sampled locations, the corresponding observational data are generated according to the measurement model:

$$z_i = P_{normalized}(x_i, y_i) + \varepsilon_i \tag{27}$$

where ε_i represents the measurement noise, which is assumed to be independently and identically distributed Gaussian noise,

$$\varepsilon_i \sim N(0, {\sigma_{meas}}^2) \tag{28}$$

with σ_{meas} denoting the variance of the measurement noise.

This observational data generation process yields sparse, noisy measurements that are unevenly distributed over the domain and biased toward regions of high gradient magnitude. Such characteristics are representative of many experimental datasets and pose a meaningful challenge for data fusion and state estimation methods.

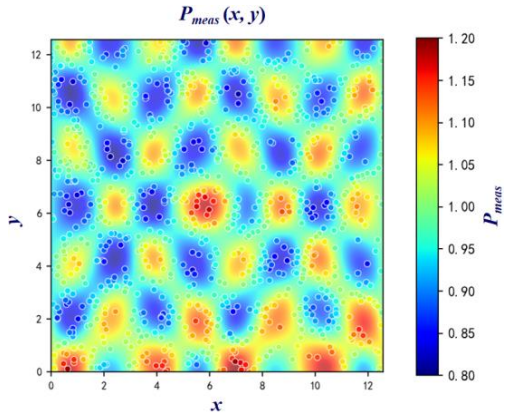

Fig.3 Locations of observation points generated by gradient-based importance sampling P_{meas}

Before performing data fusion via Kalman filtering, a Gaussian Process Regression (GPR) method is employed to reconstruct a continuous spatial field from the sparse and noisy observational data. Gaussian process regression is a non-parametric Bayesian approach that infers a probability distribution over functions by explicitly modelling spatial correlations through a covariance kernel.

In the GPR framework, the unknown field is modelled as a realization of a Gaussian process,

$$f(\boldsymbol{x}) \sim GPR(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')) \tag{29}$$

where $\boldsymbol{x} = (x, y)$ denotes the spatial coordinate, $m(x)$ is the mean function, and $k(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance kernel that encodes the spatial correlation structure of the field. In this study, the mean function is taken as $m(x) = 0$, without loss of generality, as all data have been normalized beforehand. To characterize spatial smoothness and locality, a radial basis function (RBF) kernel is adopted:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 e^{-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2l^2}} \tag{30}$$

where σ^2 denotes the signal variance controlling the overall magnitude of fluctuations, and l is the characteristic length scale governing the spatial correlation range. These hyperparameters determine how information propagates from observed locations to unobserved regions and are selected according to the spatial characteristics of the reference field.

Gaussian process regression yields a posterior predictive distribution for the field at arbitrary spatial locations. This posterior is Gaussian, fully characterized by its mean and variance, providing both a reconstructed field and an associated uncertainty estimate. In the present study, the posterior mean of the GPR prediction is used as a smooth, spatially continuous estimate of the field inferred solely from experimental observations. This reconstructed field serves two purposes: (i) it provides an intermediate representation that mitigates the effects of sparse sampling and measurement noise, and (ii) it facilitates a consistent comparison and subsequent fusion with the noisy numerical field within the Kalman filtering framework.

Under the assumed statistical models for the process and observation noise, the Kalman filter produces a fused field P_{fused} that minimizes the posterior estimation error covariance in a least-squares sense. The fusion procedure is applied to the entire field, yielding a spatially continuous posterior estimate.

The resulting fused field is visualized and compared with both the numerical and analytical fields.

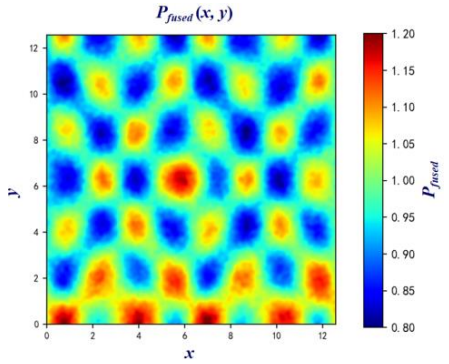

Fig.4 Fused field P_{fused} obtained by Kalman filtering from sparse noisy observations and numerical field

To quantitatively evaluate the effectiveness of the data fusion procedure, several error metrics are computed by comparing the numerical results with the analytical reference field, including the root-mean-square error (RMSE), the standard deviation of pointwise errors, and the RMSE evaluated in high-gradient regions. Both the prior numerical field and the posterior fused field are assessed against the analytical reference. As summarized in Table 1, the data fusion procedure leads to a substantial reduction in all error metrics, with improvement ratios of approximately 67%, clearly demonstrating the effectiveness of the proposed data fusion approach in enhancing solution accuracy.

Table 1. Quantitative comparison of error metrics for numerical and fused fields

	Priori	Posteriori	Improvement-ratio
RMSE	0.052407	0.017425	66.75%
std	0.052392	0.017382	66.82%
High-grad-RMSE	0.052514	0.017261	67.13%

In addition to the quantitative metrics, the spatial distributions of the prior and posterior error fields are illustrated in the following figure, where the error contour maps provide a visual comparison of the error reduction achieved after data fusion.

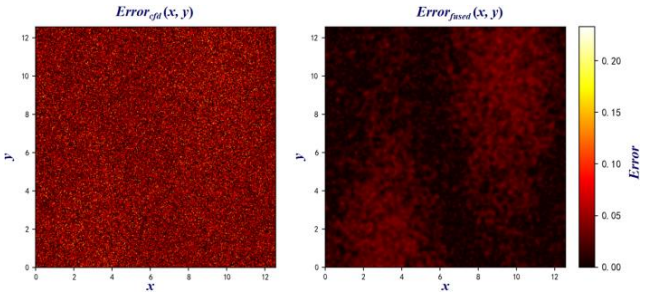

Fig.5 Comparison of error distributions before and after data fusion

Super-Resolution Reconstruction

Following the validation on the analytical benchmark problem, the proposed data fusion framework is further applied to a more realistic and challenging scenario: the super-resolution reconstruction of turbulent pressure fields in a channel flow. Compared with the simple function experiment, this case involves complex multi-scale flow structures, strong spatial intermittency, and realistic numerical and observational uncertainties, thereby providing a test of the robustness and effectiveness of the proposed methodology.

The turbulent channel flow considered in this study is simulated using wall-resolved large-eddy simulation (WRLES), which provides the high-fidelity reference data for the subsequent super-resolution

reconstruction experiments. The computational domain is illustrated in Fig. 11, with dimensions $L_x \times L_y \times L_z = 2\pi\delta \times 2\delta \times \pi\delta$, where $\delta = 1m$ denotes the channel half-height. The coordinates x, y, and z correspond to the streamwise, wall-normal, and spanwise directions, respectively.

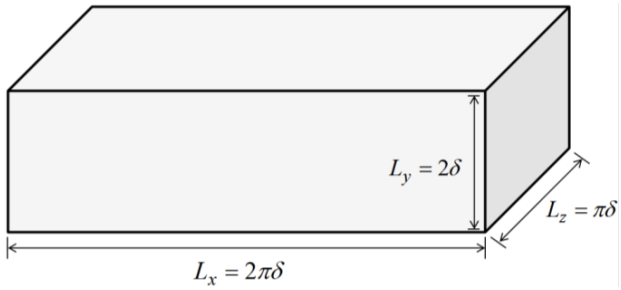

Fig. 6 The illustration of the computational domain

Periodic boundary conditions are imposed in the streamwise and spanwise directions. At the top and bottom walls, a no-slip boundary condition is applied for the velocity field, while a zero-gradient boundary condition is prescribed for pressure. To maintain a statistically stationary flow, a source term is introduced into the momentum equations to enforce a constant bulk mean velocity U_b.

The Reynolds numbers based on the bulk velocity and the friction velocity are $Re_b = U_b\delta/\nu = 20000$ and $Re_\tau = u_\tau\delta/\nu = 1000$, where $u_\tau = \sqrt{\tau_\omega/\rho}$ is the friction velocity at the wall, τ_ω denotes the wall shear stress, and ν is the kinematic viscosity.

The computational mesh is designed to fully resolve near-wall turbulence structures. The grid parameters are summarized in Table 1, where $N_x \times N_y \times N_z$ denotes the number of grid points in the three spatial directions, N_{total} is the total number of grid cells, and $\Delta x^+, \Delta y^+, \Delta z^+$ represent the grid spacings in wall units. The nondimensional time step is denoted by Δt^+. In this study, quantities with superscript "+" are nondimensionalized using the characteristic length scale ν/u_τ and velocity scale u_τ.

In the wall-normal direction, the grid spacing is gradually stretched from $\Delta y^+ = 0.79$ at the wall to $\Delta y^+ = 15$ using a linear stretching ratio of 1.05, after which a uniform spacing is employed. The total number of grid cells is approximately 22.1 million, ensuring adequate resolution for wall-resolved LES at $Re_\tau = 1000$.

Regarding the numerical methods, the pressure-velocity coupling is solved using the PISO algorithm with three pressure correction steps per time step. Temporal discretization is performed using a second-order implicit backward scheme. The gradient and Laplacian terms are discretized with second-order linear schemes, while the convective term in the Navier-Stokes equations is treated using second-order central differencing.

Table 2. The summary of the key mesh parameters used in the WRLES simulations

Case	$N_x \times N_y \times N_z$	Δx^+	Δy^+	Δz^+	Δt^+
WRLES	321×217×321	19.6	0.79~15.0	9.8	0.1

Fig. 7 illustrates the instantaneous vortical structures in the near-wall region identified using the Q-criterion. Owing to the approximate symmetry of the channel flow, only the vortical structures between the lower wall and the channel centerline are shown. It can be observed that vortex structures become progressively larger with increasing wall-normal distance. This spatial variation clearly reflects the multi-scale characteristics inherent to wall-bounded turbulent flows.

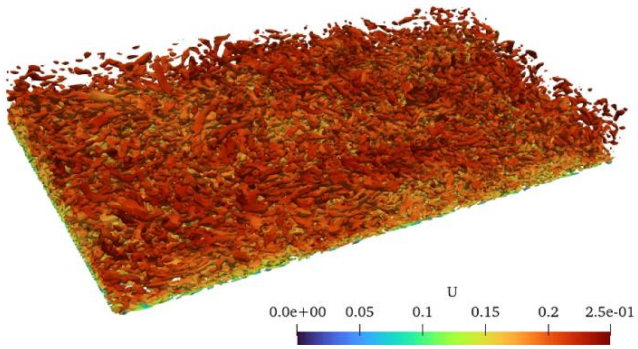

Fig. 7 Instantaneous near-wall vortical structures visualized using the Q-criterion ($Q = 0.05$).

The following figures present two-dimensional snapshots of the pressure field obtained from the WRLES simulation of turbulent channel flow. The pressure fields are extracted on a representative cross-sectional plane and visualized on a uniform grid with a resolution of 320×320. This resolution corresponds to the high-fidelity reference field used in the super-resolution reconstruction experiments, providing a detailed representation of the spatial distribution and multiscale characteristics of turbulent pressure fluctuations in the channel flow.

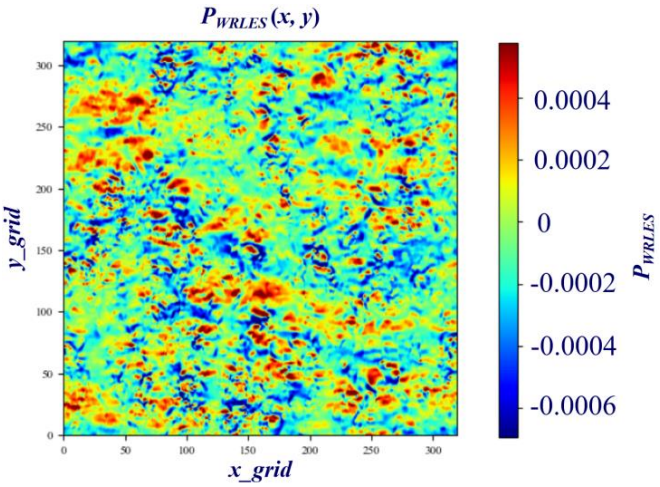

Fig. 8 The illustration of the pressure field from the WRLES simulation of turbulent channel flow visualized on a uniform grid with a resolution of 320×320

In analogy with the observation generation procedure in the simple function experiment, synthetic experimental measurements are generated by sparsely sampling the high-resolution WRLES pressure field. The sampling locations are distributed over the computational domain according to a prescribed sampling strategy, and only a limited number of observation points are retained to reflect practical experimental constraints.

Measurement noise is added to the sampled pressure values to account for sensor uncertainty and experimental errors. The observation noise is assumed to follow a prescribed stochastic distribution with user-defined statistical properties, consistent with the observation noise model adopted in the analytical benchmark.

The resulting dataset represents sparse and noisy experimental observations that provide localized but high-fidelity information about the true pressure field.

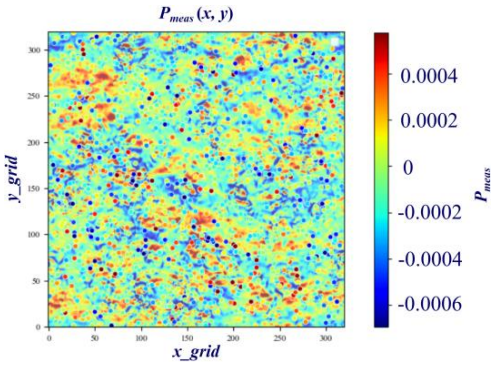

Fig. 9 Locations of the spatial distribution of the sampled observation points by gradient-based importance sampling P_{meas}

To emulate coarse-grid numerical simulations, the reference WRLES pressure field is spatially downsampled by resolution reduction factors of 4, 8, and 16 in both spatial directions, yielding low-resolution (LR) pressure fields with grid sizes of 80×80, 40×40, and 20×20, respectively. These LR fields represent numerical predictions obtained under limited computational resources and play a role analogous to the noisy numerical fields in the simple function experiment. The resulting low-resolution fields retain the large-scale flow features but fail to resolve fine-scale turbulent structures, leading to significant loss of spatial detail and increased modeling error. Each downsampled field is treated as a numerical prior in the subsequent data fusion process.

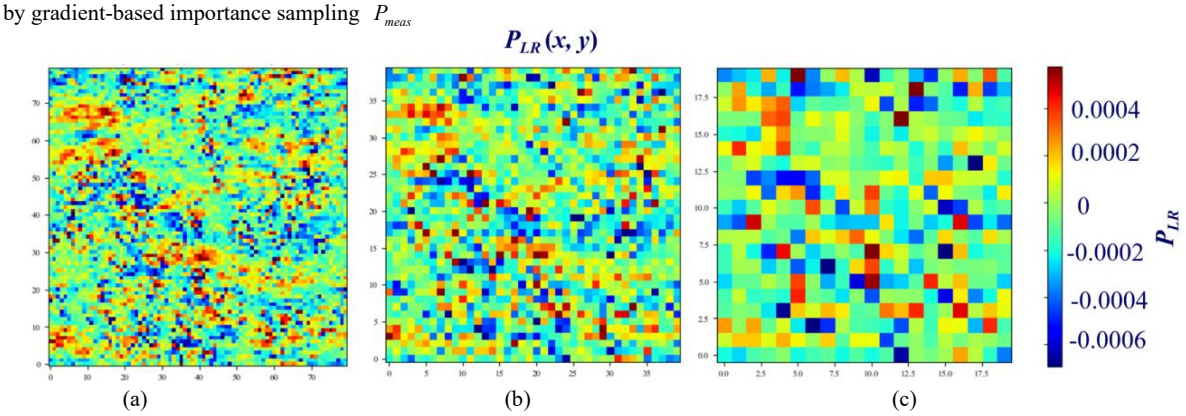

Fig. 10 The illustrations of the low-resolution pressure fields corresponding to different resolution reduction ratios: (a) resolution reduction factor is 4; (b) resolution reduction factor is 8; (c) resolution reduction factor is 16

Using the same data fusion strategy as in the simple function experiment, Kalman filtering is employed to combine the low-resolution numerical fields with the sparse noisy observations. The low-resolution numerical field serves as the background estimate, while the observational data provide corrective information to compensate for unresolved spatial scales and modeling errors.

Process noise is introduced to represent numerical uncertainty associated with coarse-grid simulations, while observation noise accounts for measurement uncertainty. The Kalman filter systematically balances these two sources of uncertainty to produce a posterior estimate of the pressure field. The resulting posterior field corresponds to a super-resolved pressure reconstruction defined on the high-resolution grid, with enhanced spatial fidelity and improved representation of turbulent structures.

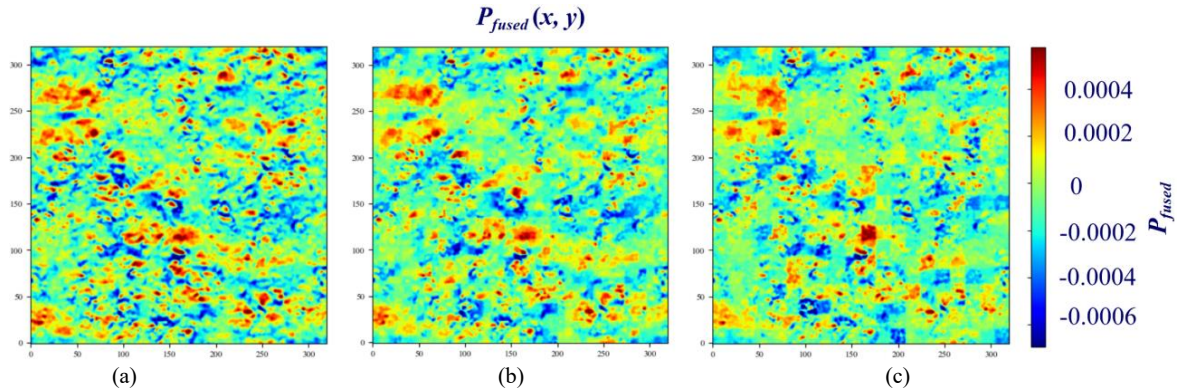

Fig. 11 The illustrations of the confused pressure fields corresponding to different resolution reduction ratios: (a) resolution reduction factor is 4; (b) resolution reduction factor is 8; (c) resolution reduction factor is 16

To quantitatively evaluate the effectiveness of the super-resolution data fusion procedure, the reconstructed pressure fields are assessed through both qualitative visualization and quantitative error analysis by comparison with the high-resolution WRLES reference solution. In addition, the prior coarse-resolution numerical fields are included in the evaluation to explicitly quantify the improvement achieved through Kalman-filter-based fusion. The evaluation strategy is consistent with that adopted in the simple function experiment, enabling a direct and coherent comparison between the analytical benchmark and the turbulent flow case.

Several error metrics are computed to characterize reconstruction accuracy, including the root-mean-square error (RMSE), the standard deviation of pointwise errors, and spectral error measures that reflect the recovery of fine-scale turbulent structures. These metrics are

evaluated for both the prior numerical fields and the posterior fused fields at different resolution reduction ratios, allowing a systematic assessment of the dependence of reconstruction performance on the level of spatial degradation.In this study, the root-mean-square error (RMSE) is employed as a global quantitative metric to evaluate the discrepancy between the reconstructed and reference flow fields.

Table 3. Quantitative comparison of error metrics for numerical and fused fields corresponding to different resolution reduction ratios

Reduction ratio	Error metric	Priori ($\times 10^{-4}$)	Posteriori ($\times 10^{-4}$)	Improvement-ratio
4	RMSE	5.82	1.96	66.3%
	std	5.64	1.88	66.7%
	High-grad-RMSE	8.25	2.71	67.2%
8	RMSE	6.35	2.45	61.4%
	std	6.12	2.28	62.7%
	High-grad-RMSE	9.18	3.52	61.7%
16	RMSE	7.86	3.92	50.1%
	std	7.45	3.68	50.6%
	High-grad-RMSE	11.24	5.87	47.8%

This systematic error analysis reveals well-defined scaling relationships between the reduction ratio and the error metrics for the turbulence modeling approach. The RMSE, standard deviations, and high-gradient-region errors all exhibit proportional scaling with respect to the reference value range of $\pm 6\times 10^{-4}$, while maintaining consistent hierarchical ordering where high-gradient errors exceed global RMSE values, which in turn surpass the standard deviation measures. The modeling framework demonstrates robust error reduction capabilities across all tested configurations, achieving 47-67% reductions in posteriori errors compared to their priori counterparts, though the improvement ratio predictably diminishes at higher reduction ratios due to increasingly significant subfilter-scale effects. These observed trends quantitatively confirm the expected resolution-dependent behavior of the turbulence model while validating the consistency of the error metric formulation, particularly in capturing the enhanced sensitivity of high-gradient regions to resolution degradation. The systematic variation of errors within one order of magnitude of the reference values further substantiates the physical meaningfulness of the quantitative results.

To further illustrate the effectiveness of the data fusion process, Fig. 12 shows the spatial distributions of the pressure reconstruction errors before and after data fusion. The error contour maps clearly demonstrate a substantial reduction in both the magnitude and spatial extent of reconstruction errors in the posterior fields.

Overall, the results demonstrate that the Kalman-filter-based data fusion framework, previously validated on an analytical benchmark, remains effective for complex turbulent flows. By systematically integrating coarse-grid numerical predictions with sparse high-fidelity observations, the method achieves substantial improvements in spatial resolution and reconstruction accuracy, highlighting its potential for practical applications in turbulent flow diagnostics and data-driven CFD enhancement.

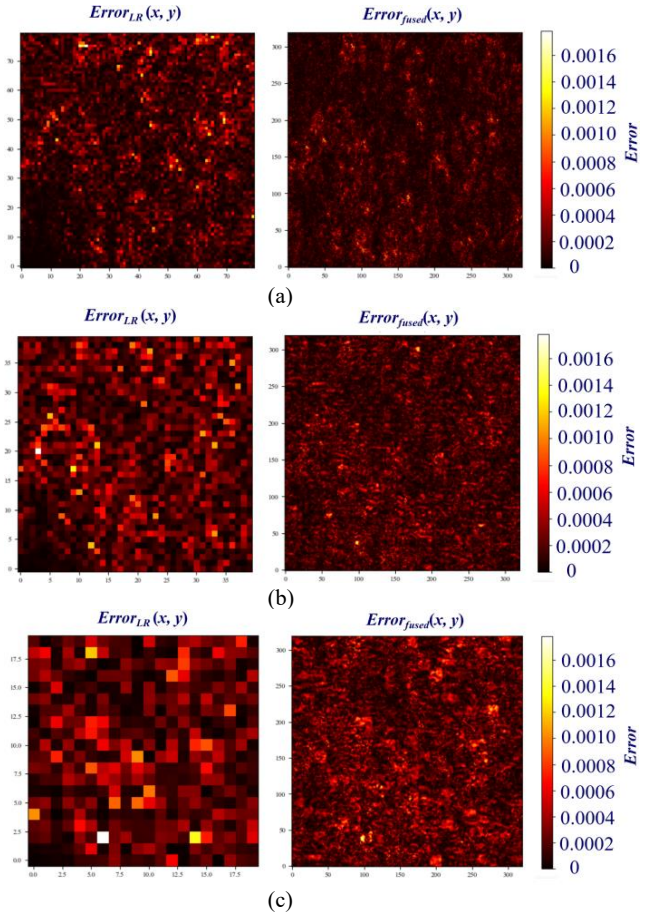

Fig. 12 The spatial distributions of the pressure reconstruction errors before and after data fusion corresponding to different resolution reduction ratios: (a) resolution reduction factor is 4; (b) resolution reduction factor is 8; (c) resolution reduction factor is 16

CONCLUSIONS

This study proposes a multi-resolution data fusion framework based on the Kalman filter, integrated with Gaussian process regression to characterize spatial correlations and propagate the constraints of high-resolution data, for large eddy simulation (LES) of turbulent flows.

Initially, the proposed framework is validated using a prescribed analytical field. This validation experiment confirms the framework's efficacy in fusing data with different levels of noise. It integrates sparse low-noise high-fidelity information with noisy low-fidelity data. Comprehensive assessments, including quantitative error analysis and visual comparison of error distributions, collectively confirm that the framework yields a significant improvement in the accuracy of flow field reconstruction relative to standalone noisy data.

Building on this validated performance, the framework is further extended to the canonical case of turbulent channel flow. Specifically, three distinct resolution reduction ratios are adopted in the simulation to investigate the influence of spatial degradation on fusion performance. The results from this practical application indicate that the proposed method can effectively capture and reconstruct key flow features of turbulent channel flow, such as coherent structures and velocity profiles, achieving a notable enhancement in prediction accuracy compared to low-resolution simulations without data fusion. Moreover, it is found that the fusion effectiveness exhibits a negative correlation with the resolution reduction ratio: the greater the degree

of resolution reduction, the poorer the fusion performance. In fact, The accuracy of the proposed framework is inherently influenced by the resolution of the input LES data, which is determined by the grid size and time step. As demonstrated in the turbulent channel flow case, the fusion performance shows a clear dependence on the resolution reduction ratio. When the resolution is excessively reduced, fine-scale turbulent structures are no longer adequately resolved, leading to weakened spatial correlations and reduced effectiveness of both the Gaussian process regression and Kalman filter. Nevertheless, within a reasonable range of resolution degradation, the proposed method consistently improves prediction accuracy compared to standalone low-resolution simulations. This phenomenon can be attributed to the fact that excessive resolution reduction leads to the loss of more fine-scale turbulent information and weakens the spatial correlation characteristics of the flow field, which in turn limits the ability of Gaussian process regression to propagate high-fidelity data constraints and the Kalman filter to optimize fusion results. This extension verifies the framework's adaptability from idealized analytical fields to complex real-world turbulent flow systems, while also clarifying the influence of resolution degradation on its performance, underscoring its robustness and practical applicability beyond controlled validation scenarios. Moreover, it should be noted that the primary advantage of the proposed framework lies in its ability to significantly reduce computational cost compared to fully resolved high-resolution LES. In the present approach, the computational burden is mainly associated with the low-resolution LES, while the additional cost introduced by the Kalman filter and Gaussian process regression is relatively minor. Unlike conventional approaches that require substantial grid refinement and reduced time steps, the proposed method achieves improved accuracy through data fusion without increasing the resolution of the underlying simulation. As a result, the framework provides an efficient trade-off between computational cost and prediction accuracy.

In summary, this work develops an efficient and accurate data fusion strategy that enables an effective trade-off between computational cost and simulation precision in turbulent flow simulations. By synergizing the merits of Kalman filtering and Gaussian process regression, the framework provides a viable solution to mitigate the high computational burden associated with high-resolution LES, exhibiting considerable application potential in turbulent flow prediction and related engineering contexts. For future research, several directions can be explored to further advance the framework's performance and applicability. First, more efficient nonlinear Kalman filters, such as the Ensemble Kalman Filter (EnKF), can be incorporated to replace the traditional Kalman filter, which is expected to enhance the framework's adaptability to the strong nonlinearity inherent in complex turbulent flows and improve fusion robustness. Second, the integration of time-scale dynamics into the framework is a key focus, aiming to achieve dynamic flow field prediction rather than static reconstruction. This extension would enable the framework to forecast the temporal evolution of turbulent flows, expanding its utility in time-dependent engineering scenarios and real-time flow control applications. Additionally, future efforts may further expand the framework's applicability to more complex flow scenarios and explore its integration with real-time experimental data to augment its practical utility.

ACKNOWLEDGEMENTS

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

REFERENCES

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Avellano, A., 2009. The data assimilation research testbed: A community facility. Bull. Amer. Meteor. Soc. 90, 1283–1296.

Attia, A., Sandu, A., 2019. DATeS: A highly extensible data assimilation testing suite v1.0. Geoscientific Model Development 12, 629–649.

Holland, J.R., Baeder, J.D., Duraisamy, K., 2019. Field inversion and machine learning with embedded neural networks: Physics-consistent neural network training, in: AIAA Aviation 2019 Forum. Presented at the AIAA Aviation 2019 Forum, American Institute of Aeronautics and Astronautics, Dallas, Texas.

Launder, B.E., Reece, G.J., Rodi, W., 1975. Progress in the development of a reynolds-stress turbulence closure. J. Fluid Mech. 68, 537–566.

Launder, B.E., Sharma, B.I., 1974. Application of the energy-dissipation model of turbulence to the calculation of flow near a spinning disc.

Ling, J., Templeton, J., Kurzawski, A., 2016. Reynolds averaged turbulence modeling using deep neural networks with embedded invariance.

Piomelli, U., 2008. Wall-layer models for large-eddy simulations. Progress in Aerospace Sciences 44, 437–446.

Pope, S.B., 2001. Turbulent flows. Meas. Sci. Technol. 12, 2020.

Slotnick, J., Khodadoust, A., Alonso, J., Darmofal, D., Gropp, W., Lurie, E., Mavriplis, D., 2014. CFD vision 2030 study: A path to revolutionary computational aerosciences.

Spalart, P., Allmaras, S., 1992. A one-equation turbulence model for aerodynamic flows, in: 30th Aerospace Sciences Meeting and Exhibit. American Institute of Aeronautics and Astronautics.

Spalart, P.R., 2000. Strategies for turbulence modelling and simulations.

Ströfer, C.A.M., Zhang, X.-L., Xiao, H., 2021. Ensemble gradient for learning turbulence models from indirect observations.

Verlaan, M., van Velzen, N., Hummel, S., Gerritsen, H., 2010. OpenDA, a generic toolbox for data assimilation in numerical modelling.

Xiao, H., Cinnella, P., 2018. Quantification of model uncertainty in RANS simulations: A review.

Yang, X.I.A., Griffin, K.P., 2021. Grid-point and time-step requirements for direct numerical simulation and large-eddy simulation. Physics of Fluids 33, 015108.