
Oral presentation | Turbulence simulation (DNS,LES,RANS)

Turbulence simulation(DNS,LES,RANS)-II

Wed. Jul 17, 2024 4:30 PM - 6:30 PM Room B

[9-B-03] Robust Data-driven RANS Turbulence Modeling Using Conditioned Field Inversion and Symbolic Regression

*Chenyu Wu¹, Yufei Zhang¹ (1. School of Aerospace Engineering, Tsinghua University, China)

Keywords: RANS turbulence modeling, Field inversion, Symbolic regression

Robust Data-driven RANS Turbulence Modeling Using Conditioned Field Inversion and Symbolic Regression

Chenyu Wu* and F. Yufei Zhang*

Corresponding author: zhangyufei@tsinghua.edu.cn

* School of Aerospace Engineering, Tsinghua University.

Abstract: Reynolds averaged Navier-Stokes (RANS) equations are essential for tackling engineering problems with high efficiency and reduced costs but exhibit accuracy limitations in separated flows due to inherent assumptions suitable mainly for attached flows. To address these issues, data-driven enhancements, particularly through field-inversion and machine learning (FIML), have been explored. While FIML has enhanced separated flow predictions, its machine-learning models lack robustness in simpler wall-attached flows like zero pressure-gradient flat plates. This work introduces the Conditioned Field Inversion (FI-CND) method, which innovatively combines a correction factor β with a shielding function f_d , inactive in the boundary layer and active elsewhere, to preserve the baseline Shear-stress-transport (SST) model's precision. The FI-CND method, compared to classic FI approaches (FI-CLS), maintains accuracy in wall-attached flows without compromising the performance improvements in separated flows provided by existing FIML methods.

Keywords: RANS, Turbulence Modeling, Data-driven method, Field inversion.

1 Introduction

Turbulence widely exists in engineering flows. Solving Navier-Stokes equations directly to resolve all eddies results in prohibitable computational costs. In practical engineering applications, people use turbulence modeling methods to lower the cost. Among the available turbulence modeling methods, the Reynolds Averaged Navier-Stokes (RANS) method has gained wide popularity due to its low computational cost and easy-to-use nature. Though the ever-growing computational capacity is making scale-resolved modeling methods (large eddy simulation, etc.) more and more realizable, experts still expect the RANS method will continue to be the workhorse of CFD in engineering applications for decades [1]. The industry is willing to have a turbulence model that can analyze more and more complex flows encountered in the full envelope of their product, especially the separated flows. However, after 2000, the research in RANS turbulence modeling seems to be encountering stagnation: only a few models are proposed and the shortcomings in separated flows are not thoroughly solved [2].

In recent years, the data-driven turbulence modeling method has shed new light on the development of novel and effective turbulence models that can resolve the separated flows more accurately [3]. The accumulation of high-fidelity data produced by direct numerical simulation (DNS) and experiments together with the giant leaps in machine learning has made this progress possible. Wang et al. applied evolution algorithms to recalibrate the turbulence models' coefficients [4] using the experiment data of the separated flows. Duraisamy et al. [5] proposed a method called field inversion (FI) that can extract the model correction term distribution from sparse high-fidelity data. A machine learning model can then be constructed to map the local flow features to the correction term [6]. The whole framework is often referred to as field inversion and machine learning (FIML). Many papers suggest that the FIML method can be applied to complex flows or even 3-dimensional separated flows [7–9]. On the other hand, the field inversion step in FIML makes the method applicable to experiment data or DNS data that is sparse, which is often the case for data generated earlier. Consequently, the FIML framework is considered to have good engineering potential [1]. Another type of work tried to correct the prediction

error by introducing nonlinearity using neural networks trained on DNS data to the Reynolds stress [10,11]. Besides, Yan et al. tried to use generative artificial intelligence to generate the fluctuating information of turbulence [12]. The generated information is then fed into a neural network to predict the model correction. In summary, the data-driven turbulence models have now achieved great success in tackling complex separated flows that are hard to accurately compute using traditional turbulence models.

However, there are still many problems with the current data-driven turbulence models that prevent them from being applied to real-world engineering problems. One of the fatal drawbacks is that the data-driven models struggle to resolve simple flows that are already nicely computed by the baseline model, showing very limited generalizability. Rumsey et al. [13] found that the FIML model trained on the adverse pressure-gradient flow degrades the baseline model's performance in simple wall-attached flows such as zero-gradient flat plates. Yang et al. [14] found that if the activation function in the neural network is not bounded, the data-driven model cannot produce correct asymptotic behavior near the wall as the Reynolds number approaches infinity. On the other hand, Spalart [15,16] argued that the law-of-wall must be computed with enough accuracy for all newly proposed turbulence models since the wall-attached flows still dominate in most engineering applications. To address this problem, Jackel et al. [17] used a closed-form correction term constructed by FIML, but with limited success. Joel Ho et al. [18] applied the probabilistic machine learning method in the ML step of FIML and successfully preserved the baseline model's performance in the wall-attached flows. However, the model's ability to correct separated flows is substantially compromised. Bin et al. [19,20] constructed the constrained-recalibration model that can fulfill the constraint related to the law-of-wall when the model parameters are freely adjusted. The constrained-recalibrated models show good performance in preserving the accuracy in wall-attached flows, but the parameters can only be constant throughout the field.

In this paper, we propose a modified field inversion approach called conditioned field inversion (FI-CND). Different from the classical field inversion (FI-CLS) method described in [5], the corrective term β is multiplied with a shielding function f_d in FI-CND. f_d is zero inside the boundary layer and 1 elsewhere, ensuring that any change in β won't affect the basic calibration in the boundary layer. We apply the FI-CND and FI-CLS methods in the same cases to construct corrective models for separated flows. The model with closed-form is pursued using symbolic regression (SR). The result shows that the model constructed by FI-CND (the SR-CND model) and the one constructed by FI-CLS (the SR-CLS model) have similar accuracy in separated flows tested, but only the SR-CND model can preserve the accuracy in the boundary layers. Consequently, the FI-CND method proposed by us can construct data-driven turbulence models that preserve the accuracy of simple wall-attached flows without compromising the original FIML's ability to correct separated flows.

2 The Classic Field Inversion and the Conditioned Field Inversion

We focus on the SST 2003 model described in [21]. The SST 2003 model has two transport equations for turbulent kinetic energy k and the specific dissipation frequency ω :

$$\begin{aligned} \frac{Dk}{Dt} &= P_k - \beta^* k \omega + \nabla \cdot [(v + \sigma_k v_T) \nabla k] \\ \frac{D\omega}{Dt} &= \frac{\gamma}{v_T} P_k - \theta \omega^2 + \nabla \cdot [(v + \sigma_\omega v_T) \nabla \omega] + 2(1 - F_1) \frac{\sigma_\omega^2}{\omega} \nabla k \cdot \nabla \omega \end{aligned} \quad (1)$$

More details for the terms in the SST 2003 model can be found in [21]. In the classic field inversion, the corrective term $\beta(x)$ is directly multiplied to the destruction term of the ω 's equation:

$$\frac{D\omega}{Dt} = -\beta(\mathbf{x})\theta\omega^2 + \dots \quad (2)$$

Consider the algebraic constraint required to satisfy the log law in the turbulent channel flow in the SST model [22]:

$$\theta - \beta^* \gamma = \kappa^2 \sqrt{\beta^*} \sigma_\omega, \kappa = 0.41 \quad (3)$$

For the ω equation used in the field inversion, the constraint will be:

$$\beta(\mathbf{x})\theta - \beta^* \gamma = \kappa^2 \sqrt{\beta^*} \sigma_\omega, \kappa = 0.41 \quad (4)$$

Combining Eq. (3) and Eq. (4), we find $\beta(\mathbf{x})$ must equals to 1 in the boundary layer for the log law to be held. If the classic field inversion process makes any correction of $\beta(\mathbf{x})$ in the boundary layer, the accuracy of the log law will be compromised. To safeguard the boundary layer, the following conditioned field inversion (FI-CND) approach is developed. For the FI-CND, the ω ' equation is written as:

$$\frac{D\omega}{Dt} = -[(\beta(\mathbf{x}) - 1)f_d + 1]\theta\omega^2 + \dots \quad (5)$$

$$f_d = 1 - \tanh[(8r_d)^3], r_d = \frac{\nu + \nu_T}{\kappa^2 d^2 \sqrt{u_{i,j}u_{i,j}}} \quad (6)$$

A typical distribution of f_d is shown in Figure 1. It equals 0 in the boundary layer and remains 1 elsewhere. Consequently, $\beta(\mathbf{x})$ in the boundary layer would never affect the basic calibrations because it is filtered out by the f_d . On the other hand, $\beta(\mathbf{x})$ outside the boundary layer can influence the flow just as it is in the FI-CLS. The intrinsic assumption behind the FI-CND is consistent with the hypotheses made by previous works [23,24] that the error of the RANS turbulence model mainly exists outside the boundary layer.

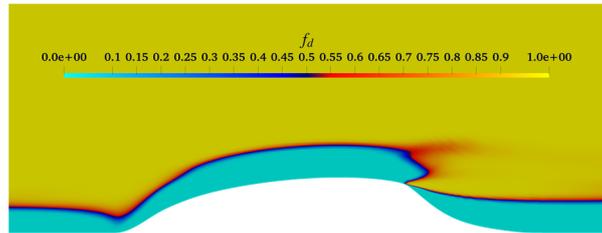


Figure 1. f_d distribution

We adjust the $\beta(\mathbf{x})$ field to minimize the discrepancy between the RANS prediction (using the corresponding $\beta(\mathbf{x})$ field) and the high-fidelity data. This task can be described as solving an optimization problem:

$$\min_{\boldsymbol{\beta}} J = \lambda_{Qol} \sum_{i=1}^K [d_i - h_i(\boldsymbol{\beta})]^2 + \lambda_{L_2} \sum_{j=1}^N (\beta_j - 1)^2 \quad (7)$$

β_j is the value of $\beta(\mathbf{x})$ in the j^{th} cell in the CFD simulation. $\boldsymbol{\beta}$ is a vector whose j^{th} element is β_j . d_i is the i^{th} high-fidelity data and $h_i(\boldsymbol{\beta})$ is the predicted value of the high-fidelity data given by RANS. λ_{Qol} and λ_2 are constants. The first term in Eq. (7) means that we adjust $\boldsymbol{\beta}$ to minimize the error between the RANS prediction and the high-fidelity data, and the second term means that we don't want the $\beta(\mathbf{x})$ field to be too bumpy. λ_{Qol} is chosen so that the first term $\sim O(1)$ when $\beta_j = 1, \forall j$ and λ_{L_2} is set to 1×10^{-5} . We use a gradient-based method [25] to solve Eq. (7) and the gradient is computed by the discrete adjoint method [26–28].

3 Data Assimilation and Model Training

In this section, we describe the data assimilation performed by the FI-CLS and the FI-CND. Machine learning models with closed-form are then constructed based on the FI data using symbolic regression.

2.1 Data Assimilation

The sparse LES data of the curved-backward facing step (CBFS) [29] and the NASA hump [30] are used for field inversion. Both flows have massive separation. The CBFS case features a low Reynolds number ($Re = 13700$) and the hump case has a high Reynolds number ($Re = 9.4 \times 10^6$). The velocity in the separated flow is used as the high-fidelity data and 30 data points are extracted and gathered for the first term in Eq. (7). The result of the field inversion is shown in Figure 2. Both the FI-CLS and the FI-CND get an optimized $\beta(\mathbf{x})$ field that can sufficiently decrease the error between the RANS prediction and the LES data. However, the $\beta(\mathbf{x})$ field given by the FI-CLS does not equal 1 in the boundary layer upstream of the step while the FI-CND gives a $\beta(\mathbf{x})$ that is uniformly 1 in the boundary layer. The $\beta(\mathbf{x})$ field obtained here is used as the label for the subsequent symbolic regression.

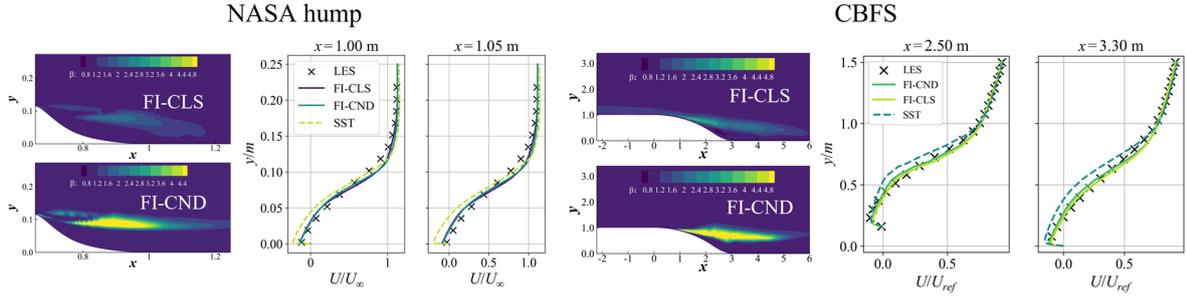


Figure 2. The field inversion result

2.2 Model Training

PySR [31] is used to distill a closed-form model from the data provided by FI-CLS and the FI-CND. The data of the CBFS case and the hump case are mixed to form the training set to capture the characteristics of both the low Reynolds number flow and the high Reynolds number flow. The input features of the model are shown in Table 1. $\hat{\mathbf{S}}$ and $\hat{\mathbf{\Omega}}$ are the non-dimensional strain tensor and rotation tensor defined as:

$$\hat{\mathbf{S}} = \frac{\mathbf{S}}{\beta^* \omega}, \hat{\mathbf{\Omega}} = \frac{\mathbf{\Omega}}{\beta^* \omega}, \beta^* = 0.09 \quad (8)$$

Table 1. The input features of the model

Name	λ_1	λ_2	λ_5	Re_Ω	P_k/ϵ	η
Definition	$\text{tr}(\hat{\mathbf{S}}^2)$	$\text{tr}(\hat{\mathbf{\Omega}}^2)$	$\text{tr}(\hat{\mathbf{\Omega}}^2 \cdot \hat{\mathbf{S}}^2)$	$ \mathbf{\Omega} d^2/\nu$	$\tau_{ij}^R u_{i,j}/(\beta^* k \omega)$	$\lambda_2 \lambda_5 / Re_\Omega$

About 10000 sample points are used for training. The trained expressions for the correction term β based on the FI-CLS dataset and the FI-CND dataset are shown in Eq. (9) and Eq. (10).

$$\text{FI-CLS: } \beta - 1 = -\frac{3}{500} \lambda_5 \tanh(-0.092 \lambda_2) \quad (9)$$

$$\text{FI-CND: } \beta - 1 = \min(0.00435 \lambda_2^2, 3.806) \quad (10)$$

The two models above are referred to as the SR-CLS and the SR-CND model in the rest of the paper. Both expressions adjust the β value based on the local strain rate or rotation rate.

4 Results

4.1 Training Set Performance

The β contour and the velocity profiles predicted by the SR-CLS and the SR-CND model are shown in Figure 3. Both the models activate the correction term β ($\beta - 1 > 0$) in the separated shear layer. The velocity profiles obtained by the SR-CLS and the SR-CND model match the high-fidelity data better than the SST model. The result indicates that the trained models are effective in the training set. On the other hand, it suggests that the SR-CND model and the SR-

CLS model have similar ability to correct the separated flow in the training set, though they are obtained by FI-CND and the FI-CLS respectively.

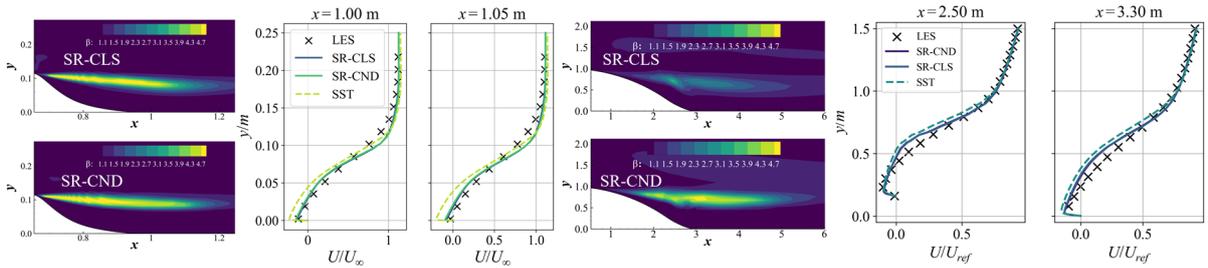


Figure 3. The performance of the SR-CND model and the SR-CLS model on the training set
 Comparing the C_f predictions of the NASA hump with the experiment data [32], we find that though both the SR-CND model and the SR-CLS model give accurate results in the separated flow, the SR-CLS model generates a large error in the attached boundary layer. Therefore, only the SR-CND model, constructed by the FI-CND method proposed, can maintain the accuracy in the wall-attached flows.

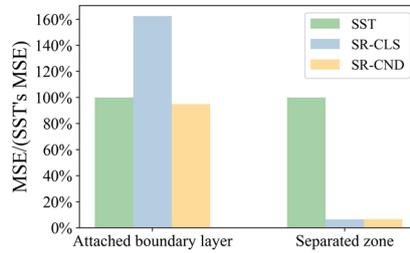


Figure 4. The MSE of C_f in different regions of the NASA hump.

4.2 Iced Airfoil (Test Set)

The GLC305 iced airfoil at $AOA = 6^\circ$ is computed using the SR-CND, SR-CLS, and the SST model. The β distribution and the pressure coefficient distribution are shown in Figure 5. Both the SR-CND and the SR-CLS models activate the correction term in the separated shear layer starting from the ice accreted at the leading edge. The C_p distribution shows that the results of the SR-CLS model and the SR-CND model are significantly closer to the experiment compared with the SST model. The SST model predicts a complete stall with a large separation bubble covering the whole upper surface of the airfoil, which is different from the experiment. The results indicate that the SR-CLS and the SR-CND models have similar abilities to correct the separated flow in the iced airfoil case.

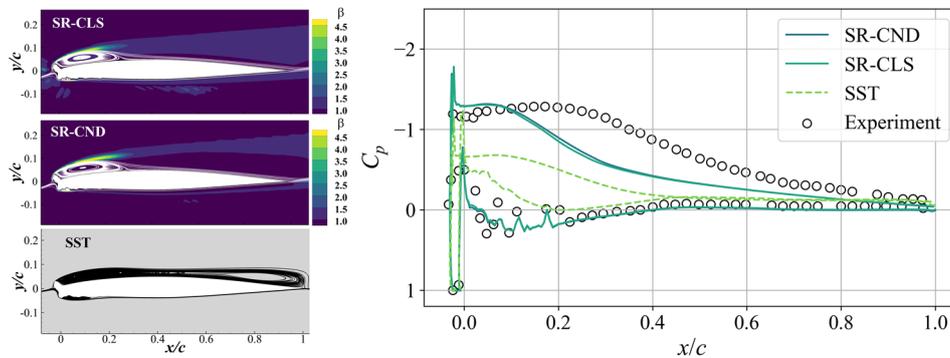


Figure 5. The result of the GLC305 iced airfoil

4.3 3D SAE body (Test Set)

The SAE body [33] is studied using the models constructed. The geometry resembles a sedan car. This case features a complex 3-D separated flow. The result shows that the separations predicted by the SR-CLS and the SR-CND model match the PIV data while the SST model overpredicts the separation size. Consequently, the model constructed by the FI-CND method proposed in this paper can achieve similar generalizability compared with the model constructed by the FI-CLS model in complex 3-D separated flows.

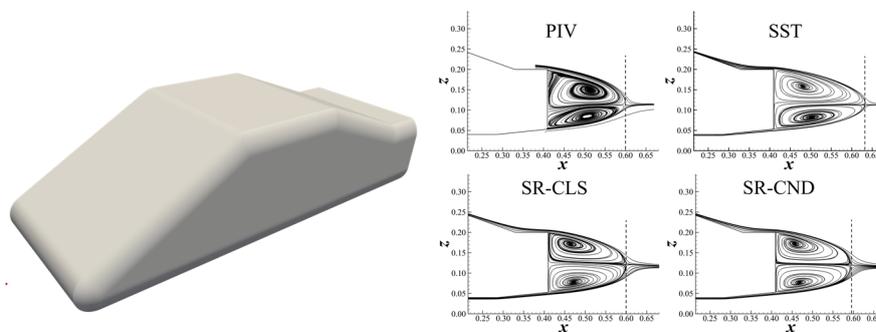


Figure 6. The result of the SAE body

4.4 Zero Pressure-Gradient Flow (Test Set)

The zero pressure-gradient (ZPG) flat plate is calculated using the models constructed in this paper. Figure 7 shows that the SR-CLS model overpredicts the C_f along the flat plate while the SR-CND model gives C_f that agrees well with the baseline SST model and the experiment [34]. The result indicates that only the model constructed by the proposed FI-CND method can preserve the accuracy of the baseline model in simple wall-attached flows.

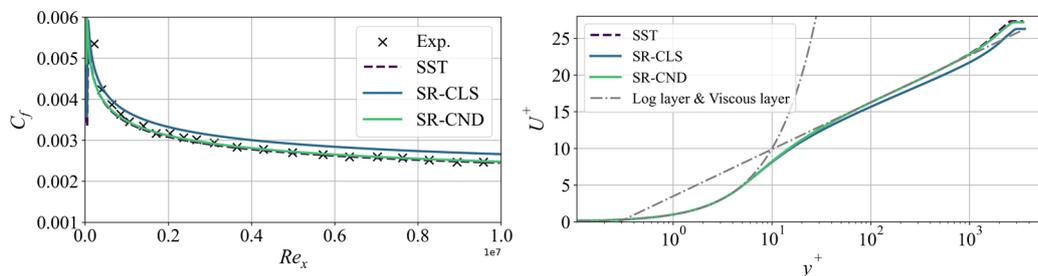


Figure 7. The C_f distribution and the velocity profile

5 Conclusion

In this research, we introduced the conditioned field inversion (FI-CND) method, which enhances the conventional field inversion approach by integrating a shielding function, f_d , with the correction factor β in the ω equation to preserve the baseline model's calibration in boundary layers. Applied to the NASA hump and the CBFS case, this method led to the development of two models: the classic SR-CLS and the novel SR-CND. Evaluations across various flow scenarios, including separated and wall-attached flows, yielded significant findings:

1. Both models performed robustly in separated flows, affirming the FI-CND method's effectiveness without compromising model generalizability across similar flow dynamics.
2. Unlike the SR-CLS, the SR-CND model maintained high accuracy in scenarios well-handled by the baseline SST model, such as the ZPG flat plate and the NACA0012 airfoil, demonstrating superior L2 generalizability.

Our results underscore that the FI-CND method not only achieves consistent performance in complex separated flows but also preserves baseline accuracy in simpler attached flows, distinguishing it from traditional field inversion approaches and advancing data-driven turbulence modeling.

Reference

- [1] K. Duraisamy, P. R. Spalart, and C. L. Rumsey, *Status, Emerging Ideas and Future Directions of Turbulence Modeling Research in Aeronautics*, (n.d.).
- [2] H. Xiao and P. Cinnella, *Quantification of Model Uncertainty in RANS Simulations: A Review*, *Progress in Aerospace Sciences* **108**, 1 (2019).
- [3] K. Duraisamy, G. Iaccarino, and H. Xiao, *Turbulence Modeling in the Age of Data*, *Annual Review of Fluid Mechanics* **51**, 357 (2019).
- [4] Z. Wang and W. Zhang, *A Unified Method of Data Assimilation and Turbulence Modeling for Separated Flows at High Reynolds Numbers*, *Physics of Fluids* **35**, 025124 (2023).
- [5] A. P. Singh and K. Duraisamy, *Using Field Inversion to Quantify Functional Errors in Turbulence Closures*, *Physics of Fluids* **28**, 045110 (2016).
- [6] E. J. Parish and K. Duraisamy, *A Paradigm for Data-Driven Predictive Modeling Using Field Inversion and Machine Learning*, *Journal of Computational Physics* **305**, 758 (2016).
- [7] K. Duraisamy, Z. J. Zhang, and A. P. Singh, *New Approaches in Turbulence and Transition Modeling Using Data-Driven Techniques*, in *53rd AIAA Aerospace Sciences Meeting* (American Institute of Aeronautics and Astronautics, Kissimmee, Florida, 2015).
- [8] C. Yan, H. Li, Y. Zhang, and H. Chen, *Data-Driven Turbulence Modeling in Separated Flows Considering Physical Mechanism Analysis*, *International Journal of Heat and Fluid Flow* **96**, 109004 (2022).
- [9] C. Yan, Y. Zhang, and H. Chen, *Data Augmented Turbulence Modeling for Three-Dimensional Separation Flows*, *Physics of Fluids* **34**, 075101 (2022).
- [10] J. Ling, A. Kurzwski, and J. Templeton, *Reynolds Averaged Turbulence Modelling Using Deep Neural Networks with Embedded Invariance*, *J. Fluid Mech.* **807**, 155 (2016).
- [11] Y. Yin, Z. Shen, Y. Zhang, H. Chen, and S. Fu, *An Iterative Data-Driven Turbulence Modeling Framework Based on Reynolds Stress Representation*, *Theoretical and Applied Mechanics Letters* **12**, 100381 (2022).
- [12] C. Yan and Y. Zhang, *Local Turbulence Generation Using Conditional Generative Adversarial Networks toward Reynolds-Averaged Navier–Stokes Modeling*, *Physics of Fluids* **35**, 105102 (2023).
- [13] C. L. Rumsey, G. N. Coleman, and L. Wang, *In Search of Data-Driven Improvements to RANS Models Applied to Separated Flows*, in *AIAA SCITECH 2022 Forum* (American Institute of Aeronautics and Astronautics, San Diego, CA & Virtual, 2022).
- [14] P. E. S. Chen, Y. Bin, X. I. A. Yang, Y. Shi, M. Abkar, and G. I. Park, *A Priori Screening of Data-Enabled Turbulence Models*, *Phys. Rev. Fluids* **8**, 124606 (2023).
- [15] P. R. Spalart, *Philosophies and Fallacies in Turbulence Modeling*, *Progress in Aerospace Sciences* **74**, 1 (2015).
- [16] P. Spalart, *An Old-Fashioned Framework for Machine Learning in Turbulence Modeling*, (n.d.).
- [17] F. Jäckel, *A Closed-Form Correction for the Spalart–Allmaras Turbulence Model for Separated Flows*, *AIAA Journal* **61**, 2319 (2023).
- [18] J. Ho, N. Pepper, and T. Dodwell, *Probabilistic Machine Learning to Improve Generalisation of Data-Driven Turbulence Modelling*, arXiv:2301.09443.
- [19] Y. Bin, G. Huang, R. Kunz, and X. I. A. Yang, *Constrained Recalibration of Reynolds-Averaged Navier–Stokes Models*, *AIAA Journal* **0**, 1 (n.d.).
- [20] Y. Bin, X. Hu, J. Li, S. J. Grauer, and X. I. A. Yang, *Constrained Re-Calibration of Two-Equation Reynolds-Averaged Navier–Stokes Models*, *Theoretical and Applied Mechanics Letters* **14**, 100503 (2024).
- [21] F. R. Menter, M. Kuntz, and R. Langtry, *Ten Years of Industrial Experience with the SST Turbulence Model*, *Heat and Mass Transfer* **4**, 625 (2003).
- [22] F. R. Menter, *Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications*, *AIAA Journal* **32**, 1598 (1994).
- [23] C. L. Rumsey, *Exploring a Method for Improving Turbulent Separated-Flow Predictions with $k-\omega$ Models*, NASA TM-2009-215952 (2009).

- [24] H. Li, Y. Zhang, and H. Chen, *Aerodynamic Prediction of Iced Airfoils Based on Modified Three-Equation Turbulence Model*, AIAA Journal **58**, 3863 (2020).
- [25] P. E. Gill, W. Murray, and M. A. Saunders, *SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization*, SIAM Rev. **47**, 99 (2005).
- [26] A. Duffy, *An Introduction to Gradient Computation by the Discrete Adjoint Method*, in (2009).
- [27] P. He, C. A. Mader, J. R. R. A. Martins, and K. J. Maki, *DAFoam: An Open-Source Adjoint Framework for Multidisciplinary Design Optimization with OpenFOAM*, AIAA Journal **58**, 1304 (2020).
- [28] O. Bidar, P. He, S. Anderson, and N. Qin, *An Open-Source Adjoint-Based Field Inversion Tool for Data-Driven RANS Modelling*, in *AIAA AVIATION 2022 Forum* (American Institute of Aeronautics and Astronautics, Chicago, IL & Virtual, 2022).
- [29] Y. Bentaleb, S. Lardeau, and M. A. Leschziner, *Large-Eddy Simulation of Turbulent Boundary Layer Separation from a Rounded Step*, Journal of Turbulence **13**, N4 (2012).
- [30] A. Uzun and M. R. Malik, *Wall-Resolved Large-Eddy Simulation of Flow Separation Over NASA Wall-Mounted Hump*, in *55th AIAA Aerospace Sciences Meeting* (American Institute of Aeronautics and Astronautics, 2017).
- [31] M. Cranmer, *Interpretable Machine Learning for Science with PySR and SymbolicRegression.Jl*, arXiv:2305.01582.
- [32] J. W. Naughton, S. Viken, and D. Greenblatt, *Skin Friction Measurements on the NASA Hump Model*, AIAA Journal **44**, 1255 (2006).
- [33] D. Wood, M. A. Passmore, and A.-K. Perry, *Experimental Data for the Validation of Numerical Methods - SAE Reference Notchback Model*, SAE Int. J. Passeng. Cars - Mech. Syst. **7**, 145 (2014).
- [34] K. Wieghardt and W. Tillmann, *On the Turbulent Friction Layer for Rising Pressure*, No. NACA-TM-1314, n.d.