DOI: 10.13009/EUCASS2023-401
Aerospace Europe Conference 2023 – 10TH EUCASS – 9TH CEAS

Data-driven Turbulence Modeling Using Field Inversion and Deep Symbolic Regression

Chenyu Wu and Yufei Zhang*†*
** School of Aerospace Engineering, Tsinghua University, Beijing, China*
Qing Hua Road, Hi Dian District
wcy22@mails.tsinghua.edu.cn – zhangyufei@tsinghua.edu.cn
† Corresponding Author

Abstract
Neural network-based, data-driven turbulence models substantially enhance the predictive capabilities for separated flows, but their generalizability to new, distinct scenarios is limited and interpretation challenging. To overcome these issues, we employ deep symbolic regression to create an analytical correction term for the Shear-Stress Transport (SST) model. The training data is generated via the field inversion method from Large Eddy Simulation (LES) data of a curved-backward-facing step. This augmented model demonstrates promising performance in separated flows that differ significantly from the training set, inclusive of complex 3D cases.

1. Introduction

Effective turbulence modeling in CFD plays an important role in a vast number of engineering applications, such as the aerodynamic design of wings, turbine blades, and automobiles. High-fidelity methods such as direct numerical simulation (DNS) and large eddy simulation (LES) can resolve complex turbulence structures with good accuracy, but the computational cost is too high to be applied to the daily engineering design process. Reynolds averaged Navier-Stokes equation (RANS), on the other hand, has a relatively low computational cost, which is suitable for evaluating multiple configurations in a limited time in the trial-and-error design phase. However, RANS turbulence models struggle to give accurate predictions in complex separated flows, since most of its components are constructed and calibrated in simple attached flows. Widely used Spalart-Allmaras (SA) model and the Shear-stress-transport (SST) model predicted erroneously large separation zone in the NASA hump case and the periodic hill [1]. Li et al. found that the SST model also fails to predict the pressure distribution on an iced airfoil [2]. Similar issues are also observed in the wake prediction of the simplified car model (Ahmed body) [3].

Many attempts have been made to enhance the RANS model's ability in predicting separated flows. Rumsey proposed a correction factor to the destruction term of ω's transport equation for the SST model based on his observation that the RANS model often underpredicts eddy viscosity in the separated shear layer. The corrected model gave improved results in the periodic hill [1]. Li et al. proposed a similar correction factor to the $k-\overline{v^2}-\omega$ model and improves its ability in predicting the flow separation on an iced airfoil [2,4,5]. Though the interpretability of these models is good, deriving these correction terms from physical knowledge and experience can be very difficult. And the modeler's own bias may hinder the discovery of more novel expressions of the correction term.

In recent years, the data-driven method has been introduced to the turbulence modeling community. The uncertainty of classic turbulence models was quantified using mathematical methods. Xiao et al. [6] used the Ensemble-Karman filter (EnKF) method to quantify the discrepancy between the Reynolds stress given by RANS models and the high-fidelity data. Duraisamy [7] added a multiplicative correction factor to the production term of the SA model and derived its distribution by minimizing the error between the quantity of interest (QoI) given by RANS prediction and high-fidelity data. This framework is called field inversion (FI). On the other hand, many data-driven turbulence models based on the black-box machine learning model that learns from high-fidelity turbulence data emerge. These models interreact with the RANS solver dynamically in the CFD solution process to improve the accuracy of separation prediction. Yin et al. [8] proposed a novel set of input features and fed them into an ANN to predict the Reynolds stress, which predicted the separation in the periodic hill case successfully. Yan et al. [9] utilized field inversion to derive the distribution of the correction term and then uses the ANN model to predict it dynamically. This approach succeeded in both 2D and 3D separation prediction [10]. Though these black-box models excel in flows similar to their training set, they often fail in completely different flows. Some of them may even deteriorate the original turbulence model's

DOI: 10.13009/EUCASS2023-401

ability to predict simple attached flows [11]. On the other hand, the black-box model lacks interpretability, making its behavior unpredictable and prone to give surprising results.

Recently, the symbolic regression (SR) method, which is a classic ML algorithm, is introduced to turbulence modeling. Like other ML methods, SR also learns from data, but it produces a (set of) compact analytical expressions of the output relative to the input features rather than a black-box model which is often over-parameterized. Since the analytical expression is short, interpretable, and has only a few parameters, it is less prone to overfit the data and has a larger potential to be generalized to new cases. Some SR frameworks [12,13] even allow the modeler to define custom operators to reflect their physical understanding of the problem, which also increases the output expression's interpretability. SR has been used for discovering algebraic Reynolds-Stress models [14] and nonlinear eddy viscosity relations in multi-phase flows [15]. However, in these works the generalization ability of the expression is not thoroughly studied, and the final expression is simple polynomials.

In this study, we propose a new data-driven turbulence modeling framework to obtain an analytical correction factor for the SST turbulence model. In this framework, the correction factor $\beta(\mathbf{x})$ is multiplied by the destruction term of $\omega's$ equation. Field inversion is then performed on the curved-backward-facing-step (CBFS) case using LES data to derive $\beta(\mathbf{x})$'s distribution. After that, we leverage the power of a newly developed deep symbolic regression (DSR) tool [13] to find out the analytical expression (denoted as $\beta(\boldsymbol{\chi})$) of β with respect to selected flow features (denoted as $\boldsymbol{\chi}$). The expression $\beta(\boldsymbol{\chi})$ is then modified based on physical a priori and integrated into the original SST model to obtain the SST-DSR model. The SST-DSR model is then applied to its training set (CBFS) and distinct testing cases with separated flows including 2D bumps with various heights and 3D Ahmed body (simplified car body). The SST-DSR model outperforms the original SST model in all these testing cases, showing its high generalizability. The model is also applied to the turbulent boundary layer on a flat plate, giving nearly identical results compared with the original SST model. This implies that the correction term obtained by DSR does not affect the original SST model's performance in simple attached flows.

2. The framework of uncertainty quantification and deep symbolic regression

In this paper, the field inversion method proposed by [7] is used to quantify the uncertainty of the SST turbulence model. After field inversion, the distribution of the correction term β is obtained. Then, DSR [13] developed by Petersen et al. is applied to learn the analytical relationship between β and the selected flow features $\boldsymbol{\chi}$. The overall framework of this study including the testing part is sketched in Figure 1. In the following section, the outline of field inversion will be discussed.

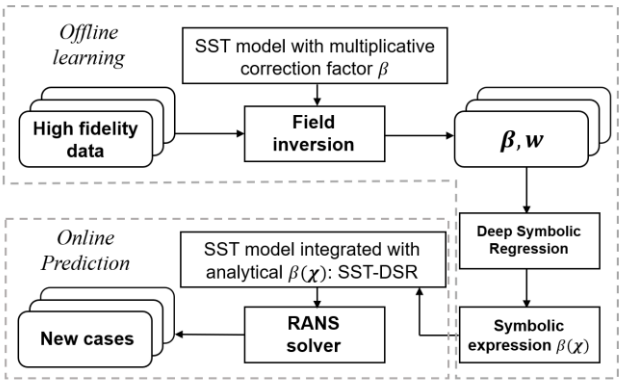


Figure 1. The framework of field inversion and Deep symbolic regression as well as the testing process.

2.1 Field inversion

The SST model [16] takes the following form. It has two transport equations to describe the distribution of turbulent kinetic energy k and specific dissipation rate ω :

DOI: 10.13009/EUCASS2023-401

$$\frac{\partial(\rho k)}{\partial t}+\frac{\partial(\rho u_j k)}{\partial x_j}=P-\beta^*\rho\omega k+\frac{\partial}{\partial x_j}\left[(\mu+\sigma_k\mu_t)\frac{\partial k}{\partial x_j}\right]$$
$$\frac{\partial(\rho\omega)}{\partial t}+\frac{\partial(\rho u_j \omega)}{\partial x_j}=\frac{\gamma}{\nu_t}P\underline{-\theta\rho\omega^2}+\frac{\partial}{\partial x_j}\left[(\mu+\sigma_\omega\mu_t)\frac{\partial \omega}{\partial x_j}\right]+2(1-F_1)\frac{\rho\sigma_{\omega 2}}{\omega}\frac{\partial k}{\partial x_j}\frac{\partial \omega}{\partial x_j} \tag{1}$$

The production term of k is related to the total stress and the velocity gradient:

$$P=\tau_{ij}\frac{\partial u_i}{\partial x_j},\tau_{ij}=\mu_t\left(2S_{ij}-\frac{2}{3}\frac{\partial u_k}{\partial x_k}\delta_{ij}\right)-\frac{2}{3}\rho k\delta_{ij}$$
$$S_{ij}=\frac{1}{2}(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}) \tag{2}$$

Here we use $\boldsymbol{S}=S_{ij}\boldsymbol{e_i}\boldsymbol{e_j}$ denotes the strain rate tensor and $\boldsymbol{\tau}=\tau_{ij}\boldsymbol{e_i}\boldsymbol{e_j}$ denotes the total stress tensor. The eddy viscosity μ_T is determined by:

$$\mu_T=\frac{\rho a_1 k}{max(a_1\omega,|\boldsymbol{\Omega}|F_2)},\Omega_{ij}=\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j}-\frac{\partial u_j}{\partial x_i}\right) \tag{3}$$

$\boldsymbol{\Omega}=\Omega_{ij}\boldsymbol{e_i}\boldsymbol{e_j}$ denotes the rotation rate tensor and $|\boldsymbol{\Omega}|=\sqrt{\Omega_{ij}\Omega_{ij}}$ is its magnitude. The definition of other functions and constants can be found in [16]. To derive the transport equations of the SST model, many empirical assumptions have been made, which cause error in real applications.
The SST model is widely used in engineering applications, showing great robustness. Several modifications were made to broaden the SST model's application range, including the SST-sf model and the SST-RC-Hellsten model. In these works, the modification term is always multiplied by the destruction term (which is underlined in Eq (1)) of ω's transport equation:

$$-\theta\rho\omega^2\xrightarrow{modification}-f(\boldsymbol{\chi})\theta\rho\omega^2 \tag{4}$$

where $\boldsymbol{\chi}$ represents some flow features. In our study, we follow this method and introduce the multiplicative factor $\beta(\mathbf{x})$ to the destruction term of ω to reflect the uncertainty of the SST model:

$$\frac{\partial(\rho\omega)}{\partial t}+\frac{\partial(\rho u_j \omega)}{\partial x_j}=\frac{\gamma}{\nu_t}P\underline{-\beta(\mathbf{x})\theta\rho\omega^2}+\frac{\partial}{\partial x_j}\left[(\mu+\sigma_\omega\mu_t)\frac{\partial \omega}{\partial x_j}\right]+2(1-F_1)\frac{\rho\sigma_{\omega 2}}{\omega}\frac{\partial k}{\partial x_j}\frac{\partial \omega}{\partial x_j} \tag{5}$$

Every time the distribution of $\beta(\mathbf{x})$ is adjusted, the flow field predicted by the RANS simulation changes. If the $\beta(\mathbf{x})$ field is adjusted so that the quantity of interest (QoI, i.e., velocity) predicted by the RANS method is close to the QoI given by high-fidelity method (i.e., experiment, DNS, and LES), we can say that this 'optimum' $\beta(x)$ distribution can quantify the uncertainty (error) caused by the empirical assumptions made in the form of the SST model. We derive the distribution of $\beta(\mathbf{x})$ by minimizing the following function:

$$\min_{\beta} J=\lambda_{obs}\sum_i\left(\hat{d}_i-d_i(\boldsymbol{\beta})\right)^2+\lambda_{prior}\sum_j(\beta_j-1)^2 \tag{6}$$

$\boldsymbol{\beta}$ is the vector whose j th component is the value of $\beta(\mathbf{x})$ on the j th cell. β_j is $\boldsymbol{\beta}$'s j th component. $\hat{d}_i$ denotes the i th high-fidelity QoI data and $d_i(\boldsymbol{\beta})$ is the i th QoI predicted by the RANS method given the vector $\boldsymbol{\beta}$. The first term in Eq. (6) represents the error of QoI. The second term is added to prevent β from deviating from its original value (i.e., 1) too far away, punishing β distribution that is not smooth enough. $\lambda_{obs}\approx\left[\sum_i\left(\hat{d}_i-d_i(1)\right)^2\right]^{-1}$ to make $J\approx 1$ when $\beta_j=1,\forall j$. λ_{prior} is often at the level of $1\times10^{-6}\sim1\times10^{-4}$.

The optimization problem (which is also called the field inversion problem) in Eq. (6) is solved by SNOPT [17] using gradient-based algorithms. The discrete adjoint method, which is designed for optimization problems involving PDE, is used to compute the gradient. The details of the discrete adjoint algorithm can be found in [18]. In this paper, the RANS solver and its adjoint solver are developed upon the open-source discrete adjoint toolkit DAFoam [19]. DAFoam is remarkably flexible and is friendly to secondary development.

DOI: 10.13009/EUCASS2023-401

2.2 Deep symbolic regression

The overall framework of symbolic regression (SR) is quite similar to ordinary ANN models. A set of input features is selected and the dataset of input features and output labels is generated. Then the dataset is fed into the algorithm, which then outputs a model mapping the input features to the labels. In this study, five flow features are chosen as the input :

$$x_0 = tr(\hat{\boldsymbol{S}}^2) = \lambda_1, x_1 = tr(\hat{\boldsymbol{\Omega}}^2) = \lambda_2, x_2 = tr(\hat{\boldsymbol{\Omega}}^2 \cdot \hat{\boldsymbol{S}}^2) = \lambda_5, x_3 = \frac{|\boldsymbol{\Omega}|d^2}{\nu}, x_4 = P/\epsilon \tag{7}$$

$\hat{\boldsymbol{S}}, \hat{\boldsymbol{\Omega}}$ are nondimensional strain tensor and rotation tensor respectively:

$$\hat{\boldsymbol{S}} = \frac{k}{\epsilon}\boldsymbol{S} = \frac{k}{\epsilon}\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)\boldsymbol{e}_i\boldsymbol{e}_j, \hat{\boldsymbol{\Omega}} = \frac{k}{\epsilon}\boldsymbol{\Omega} = \frac{k}{\epsilon}\frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right)\boldsymbol{e}_i\boldsymbol{e}_j \tag{8}$$

d is the distance to the nearest wall. P is the production term of the turbulent kinetic energy. The definition can be found in Eq. (2). ϵ in Eq. (11), (12) is defined as:

$$\epsilon = \beta^* \rho \omega k, \beta^* = 0.09 \tag{9}$$

$x_0 \sim x_2$ mainly reflect the strain and rotation of the flow, and is proposed by Pope [20]. x_3 is first used in [2] to detect regions with strong shear above the wall. x_4 is used by Rumsey in [1] to measure the non-equilibrium characteristic of turbulence. The label of the dataset is set to be the difference between the β derived by field inversion and 1 :

$$y_i = \beta_i - 1 \tag{10}$$

One additional requirement of SR is that the user should also specify the operators used to construct the expression. In this paper, the following operators are used :

Table 1 The operators used to construct the final expression

Unary operators	$\exp(x_i), \tanh(x_i), \frac{1}{x_i}$
Binary operators	$x_i + x_j, x - x_j, x_i * x_j, \frac{x_i}{x_j}$

The deep symbolic regression (DSR) tool developed by Petersen [13] is used in this paper. RNN is used to predict the next symbol in a sequence of symbols. In particular, RNN predicts a probability distribution over all available symbols specified by the user of the next token based on the context. The next token is then obtained by sampling this probability distribution. The prediction terminates when the number of symbols balances the sum of arity provided by all symbols (2 for binary operators, 1 for unary operators, and 0 for constants and variables) :

$$N - 1 - \sum_{i=1}^{N} A_i = 0 \tag{11}$$

where N is the total number of symbols and A_i is the arity of the i th symbol. The resulting sequence of symbols is interpreted as a pre-order transversal of an expression tree and the corresponding expression tree is reconstructed by placing the symbols in the sequence from root to node and from left to right, as shown in Figure 2. The process of calling RNN to generate the symbol sequence is shown in Figure 3. Note that the context of a token is its sibling and parent node in the expression tree.

DOI: 10.13009/EUCASS2023-401

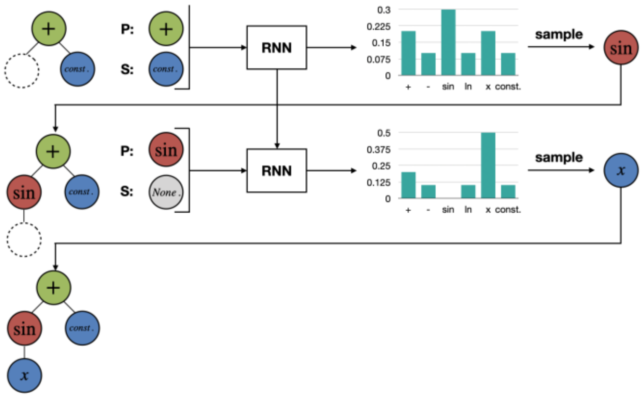


Figure 2. Calling RNN recursively to generate a sequence of symbols

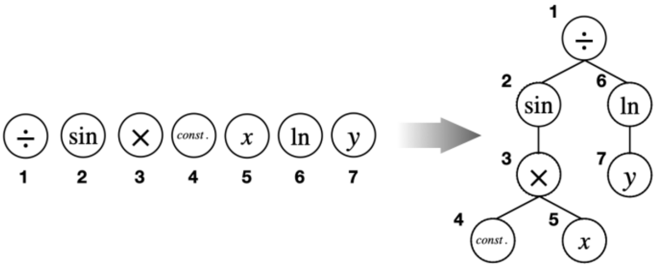


Figure 3. Reconstruct the expression tree by treating the sequence of symbols as a pre-order traversal

During the training process, the dataset is fed in, and a reinforcement learning method is used to adjust the parameters of the RNN to generate more accurate expressions. The objective of reinforcement learning is to maximize the following reward function:

$$R = \frac{1}{1 + NRMSE} \tag{12}$$

$$NRMSE = \frac{1}{\sigma_y}\left[\frac{1}{n}\sum\nolimits_{j=1}^{n}(y_j - \hat{y}_j)^2\right]^{\frac{1}{2}}, \sigma_y = \left[\frac{1}{n}\sum\nolimits_{j=1}^{n}(\hat{y}_j - \bar{\hat{y}})^2\right]^2 \tag{13}$$

where $\hat{y}_j$ is the j th label and y_j is the j th predicted value. σ_y is the standard deviation of the label in the training set, acting as a normalizer. For a perfect recovery of the symbolic expression (i.e., $NRMSE = 0$), the reward function R should be 1. The program will keep track of the expression with the highest R in the training process. When the training ends (i.e., the program hits the upper limit of the number of expressions evaluated), this best expression is chosen as the final expression.

3. Performing field inversion on the CBFS case and learning the expression using DSR

In this section, field inversion is conducted using the LES data of the CBFS case. After the optimized β field is derived, the features in Eq. (7) are calculated from the flow field corresponding to the optimum β and then fed into the DSR to train the analytical expression of $\beta - 1$. The expression obtained is then integrated into the original SST model, forming the SST-DSR model. The SST-DSR model is then applied to the training case (CBFS) to predict β dynamically. The result shows that the SST-DSR model outperforms the original SST model.

3.1 Field inversion on the CBFS case

Figure 4 shows the configuration of the CBFS case. The velocity profile is given at the inlet based on the LES simulation data [21]. The maximum velocity of the profile is $U_{ref} = 1\ m/s$. The height of the curved backward-facing

DOI: 10.13009/EUCASS2023-401

step is $h = 1\ m$. The corresponding Reynolds number based on U_{ref} and h is 13700. A total of 37093 cells are used and the height of the first layer grid satisfies that $\Delta y^+ \leq 1$.

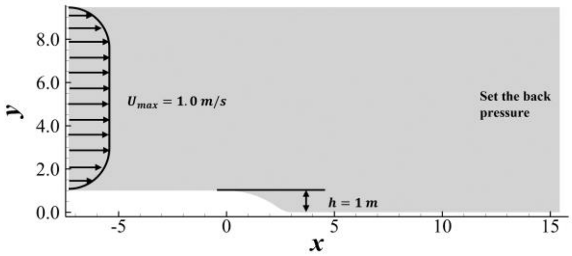
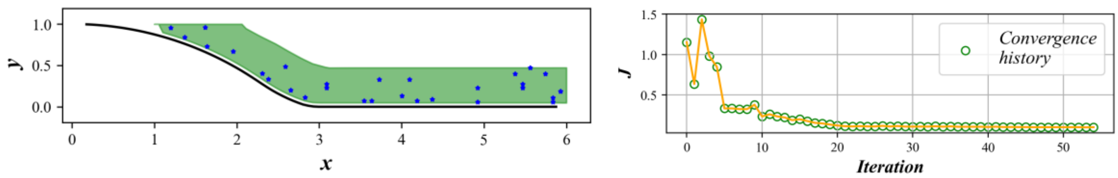


Figure 4. The configuration of the computational domain of the CBFS case

The field inversion problem of the CBFS case based on the LES data in [Bentaleb] is defined as:

$$\min_{\boldsymbol{\beta}} J = 2.0 \sum_{i=1}^{30} \left(\hat{u}_i - u_i(\boldsymbol{\beta})\right)^2 + 1.0 \times 10^{-4} \sum_{j=1}^{37093} \left(\beta_j - 1\right)^2 \tag{14}$$

The x component of the velocity (u) is chosen as the QoI. A total of 30 LES data points are used. These points are randomly sampled near the separation zone, as denoted by the shaded area in Figure 5(a).



(a) (b)

Figure 5. (a) Sample points of the LES data for field inversion. The stars represent the sample points.(b) The convergence history of J

The field inversion problem defined in Eq. (14) is solved after 55 iterations of the SNOPT optimizer. The convergence history of the objective function J is shown in Figure 5(b) It can be seen that the value of J drops about 90% after optimization.

The optimized β field identified by the optimizer is depicted in Figure 6, showing an increase in the separated shear layer. Streamlines surrounding the curved backward-facing step, corresponding to both the optimized β field and the original SST model ($\beta_j = 1, \forall j$), are presented in Figure 7. The reattachment point exhibits a significant upstream shift with the optimized β field. Following field inversion, the velocity profile aligns closely with the Large Eddy Simulation (LES) data. In contrast, the original SST model's result deviates considerably from the LES data, as demonstrated in Figure 8.

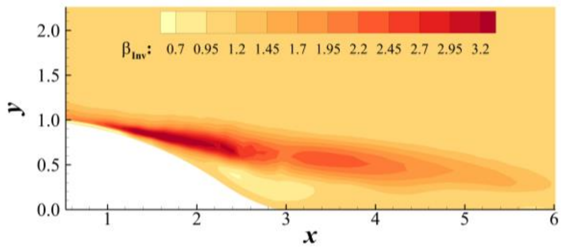


Figure 6. The optimized β distribution

DOI: 10.13009/EUCASS2023-401

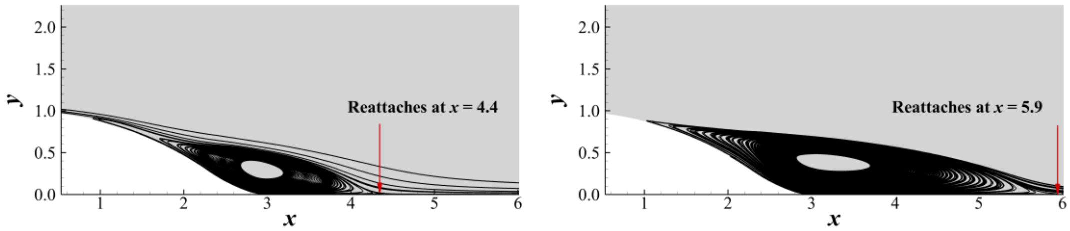


Figure 7. The separation zone predicted by optimized β and the original SST model ($\beta_j = 1, \forall j$)

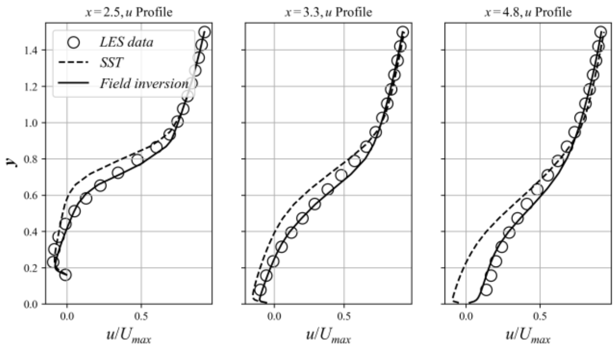


Figure 8. Velocity profiles at different $x = const.$ locations

An increased β in the separated shear layer suppresses the dissipation of k, which subsequently amplifies turbulence activity. Consequently, the optimized β field aligns physically with Rumsey's observation that conventional turbulence models frequently underestimate μ_T (which positively correlates with turbulence activity) within the separated shear layer.

3.2 Deep symbolic regression on the field inversion data

The optimized β distribution and the corresponding flow features in Eq. (7) are used to train the analytical expression of β, denoted as $\beta(\chi)$. Sample points are extracted from both the mainstream region and the separation region (without the near wall region where $d < 0.05$), as shown in . However, this extraction leads to an unbalanced β distribution: 82% of β falls in the interval $(0.9, 1.1)$, meaning that most of the samples correspond to minor correction. If we use the dataset directly to train the model, the algorithm would pay much attention to the minor correction samples, affecting the prediction accuracy of the significantly corrected samples ($\beta > 1.1$). Consequently, we use random under-sampling to limit the amount of β in the interval $(0.9, 1.1)$. After the under-sampling procedure, a total of 949 samples (note that symbolic expression does not require a large amount of data to train, since there are just few parameters to determine) are selected and the ratio of minor correction β is approximately 50%.

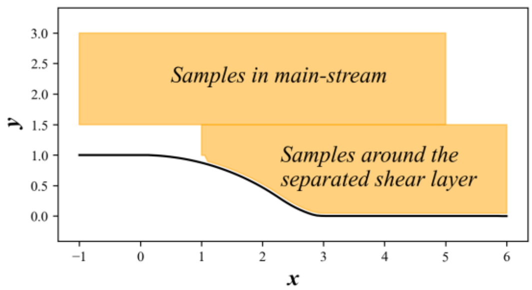


Figure 9. Samples for training the analytical expression

Deep symbolic regression (DSR) is then performed on the under-sampled dataset, using the operators in Table 1. The length of every expression is restricted to be larger than 4 and smaller than 12. A total of 2×10^6 expressions are evaluated and the batch size is 1000. The best equation, with $R = 0.718$, is:

DOI: 10.13009/EUCASS2023-401

$$\beta - 1 = \frac{P}{\epsilon}\frac{\eta}{\eta + 0.58}, \eta = \frac{\lambda_2\lambda_5}{Re_\Omega} \tag{15}$$

Eq. (15) discovered by DSR has a clear physical meaning. To interpret the equation, we start by analyzing the behavior of η. For 2D incompressible flow, λ_2 and λ_5 can rewritten using the incompressible relation:

$$\lambda_2 = tr(\widehat{\mathbf{\Omega}}^2) = -|\widehat{\mathbf{\Omega}}|^2, \lambda_5 = tr(\widehat{\mathbf{\Omega}}^2 \cdot \widehat{\boldsymbol{S}}^2) = -\frac{1}{2}|\widehat{\mathbf{\Omega}}|^2|\widehat{\mathbf{S}}|^2 \tag{16}$$

Using Eq. (8), Eq. (9), Eq. (16), and boundary layer approximation, the following expression can be derived in the boundary layer:

$$\eta \propto \frac{\nu}{d^2\omega^6}\left|\frac{\partial u}{\partial y}\right|^5 \tag{17}$$

On the other hand, [22] suggests that ω can be written as

$$\omega = \alpha\nu/d^2 \tag{18}$$

in the viscous sublayer, where α is a constant. Combining Eq. (17) and Eq. (18):

$$\eta \propto \left|\frac{\partial u}{\partial y}\right|^5 \frac{\nu}{d^2\omega^6} \propto \left(\left|\frac{\partial u}{\partial y}\right|\frac{d^2}{\nu}\right)^5 \tag{19}$$

Eq. (19) suggests that η tends to zero near the wall as $d \to 0$. On the other hand, P/ϵ tends to a finite value near the wall. This makes $\beta - 1 \to 0$, meaning that the correction term is turned off by η in the viscous sublayer, which is a desirable feature since the original SST model can already resolve the near-wall region with good accuracy. η can also be expanded in a different way, leading to other characteristics of this feature. We define:

$$T_t = k/\epsilon \tag{20}$$

T_t has a unit of time, and is called the turbulence time scale. Meanwhile, for the shear flow, we have approximately:

$$T_m \coloneqq \frac{1}{|\boldsymbol{S}|} \approx \frac{1}{|\mathbf{\Omega}|} \tag{21}$$

T_m can be viewed as the time scale of the mean flow. With Eq. (20) and Eq. (21), η can be written as:

$$\eta \approx \left(\frac{T_t}{T_m}\right)^5 \frac{T_t\nu}{d^2} = |\widehat{\boldsymbol{S}}|^5 \frac{T_t\nu}{d^2} \tag{22}$$

Within the separated shear layer, the mean flow shear exhibits high intensity, resulting in a significantly reduced mean flow time scale, potentially even smaller than T_t, as per Eq. (21). As a result, the term $(T_t/T_m)^5 = |\widehat{\boldsymbol{S}}|^5$ in Eq. (22) can become quite large, thereby augmenting the value of η. In the curved backward-facing step (CBFS) scenario, this inflated $(T_t/T_m)^5$ may lead to $\eta \sim O(10)$ in the separated shear layer. A high η brings the second term $\eta/(\eta + 0.58)$ in Eq. (15) close to 1, activating the correction term.

The above arguments demonstrate that η serves as a feature to identify the separated shear layer; it activates the correction (making $\eta/(\eta + 0.58)$ approach 1) within the separated shear layer and deactivates it near the wall. We now examine the first term in Eq. (15). When the correction term is turned on by η, its magnitude is approximately:

$$\beta - 1 \approx \frac{P}{\epsilon} \tag{23}$$

In the separated shear layer, the non-equilibrium turbulence is usually strong [1], leading to a large P/ϵ. Eq. (23) suggests that a larger P/ϵ leads to larger β, rendering a stronger ω destruction according to Eq. (5), thereby

DOI: 10.13009/EUCASS2023-401

intensifying turbulence activity (higher k). The increase in k will elevate the value of μ_T, enhancing the momentum exchange. The high momentum in the mainstream is then transferred into the separated shear layer, promoting reattachment. The physical process discussed above is shown graphically in Figure 10.

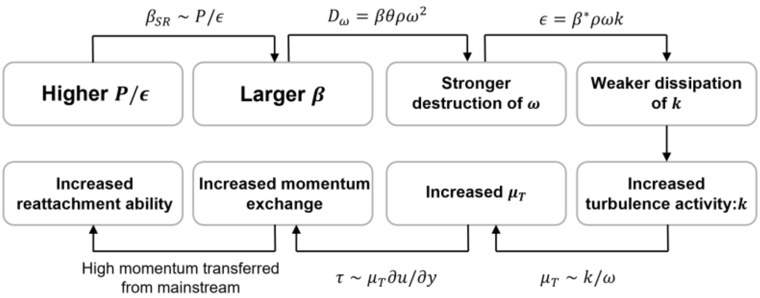


Figure 10 The physical process in which the higher P/ϵ leads to increased reattachment ability through the correction term in Eq. (15)

The discussion above analyzes Eq. (15) term by term, showing the good physical interpretability of the expression generated by DSR. However, Eq. (15) still has some flaws. η is quite large at the edge of the attached boundary layer, where we don't want the correction term to be activated. Moreover, in the region of expansion or compression, $|\widehat{\boldsymbol{S}}|$ might be high, activating the correction term. To get rid of these scenarios, two switches are added to the final expression:

$$\beta_{DSR} = [\beta(\chi) - 1]s_{\lambda_5}s_I + 1.0 \tag{24}$$

$$s_{\lambda_5} = (1/2)\tanh\left[C_{\lambda_5,1}\left(\lambda_5 - C_{\lambda_5,2}\right)\right] + 1/2, s_I = (1/2)\tanh\left[C_{I,1}\left(I - C_{I,2}\right)\right] + 1/2 \tag{25}$$

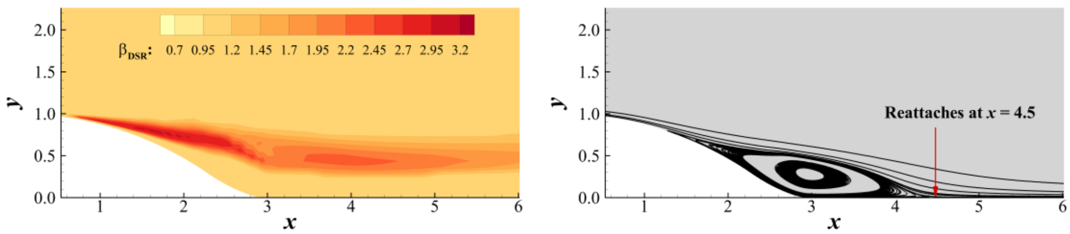
$$I = k/|\mathbf{u}|^2 \tag{26}$$

$$C_{\lambda_5,1} = -5.0, C_{\lambda_5,2} = -27.0, C_{I,1} = 800, C_{I,2} = 0.007 \tag{27}$$

The physical meaning of these two switches is that the correction term won't be turned on unless $|\lambda_5| > |C_{\lambda_5,2}|$ and $I > C_{I,2}$. The constants' value in Eq. (27) is calibrated on the CBFS case.

3.3 Online prediction of the SST-DSR model on the CBFS case

Eq. (24) is integrated into the original SST model's code to obtain the SST-DSR model. The SST-DSR model is then applied to the CBFS case. Though the geometry is the same as the one used in field inversion, the value of input features of Eq. (24) is different from its training data in the DSR process, since in the CFD iteration process the value of the input features changes. The contour of β_{DSR} is shown in Figure 11 (a). β_{DSR} is increased in the separated shear layer, which is similar to the result of field inversion. The position of the reattachment point is much closer to the result of field inversion compared with the result given by the SST model in Figure 7. The velocity profiles at different standpoints in x direction shown in Figure 12. The u_x given by the SST-DSR model aligns well with the LES data. In summary, the result shows that the SST-DSR model outperforms the original SST model on the CBFS case, showing that the expression constructed by the DSR method is effective, though its length is quite short.



(a) The contour of β_{DSR} (b) The streamlines near the corner

Figure 11. The contour of β_{DSR} and the streamline plot of the separation zone

DOI: 10.13009/EUCASS2023-401

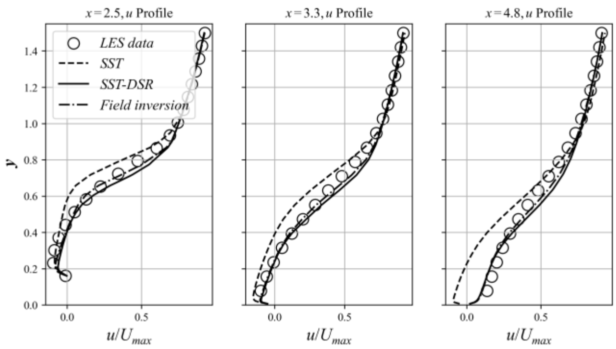


Figure 12. Velocity profile at different x positions

4. Testing the SST-DSR on distinct cases

The SST-DSR model is applied to various test cases that have completely different geometry compared with the CBFS case to test the model's generalizability. The result shows that the SST-DSR model performs well in these cases.

4.1 2D-bumps with various height

In this section, two bumps with different heights ($h = 42\ mm, h = 31\ mm$) are studied. The geometry of the bump is shown in Figure 13. There are three main differences between the 2D-bump case and the CBFS case:

1. The geometry is completely different. In the CBFS case, the flow only undergoes expansion caused by the step. However, in the 2D-bump case, the flow undergoes both compression and expansion.
2. The shape of the separation zone is different. In the CBFS case, the size of the separation zone is large, but in the 2D-bump case with $h = 31\ mm$, the separation zone is relatively small and resides in the corner.
3. The numerical magnitudes of flow variables are different. In the CBFS case, the inlet velocity is normalized with $|\boldsymbol{U}|_{max} = 1\ m/s$. However, in the 2D-bump case, $|\boldsymbol{U}|_{max} \approx 18\ m/s$ at the inlet. Henceforth, the $|\boldsymbol{S}|$ in the CBFS case is around 3~6 in the separated shear layer, whereas in the 2D-bump case, $|\boldsymbol{S}|$ is around 1000 in similar regions.

All of the differences listed above challenge the SST-DSR model's generalization ability.

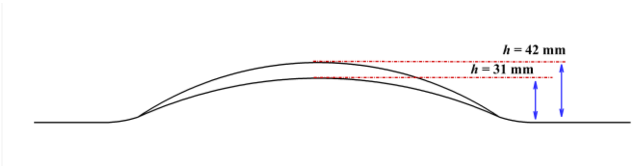
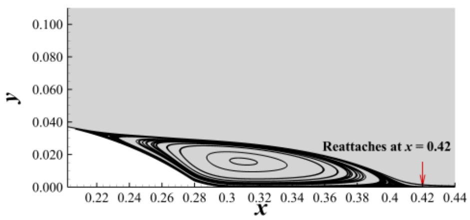


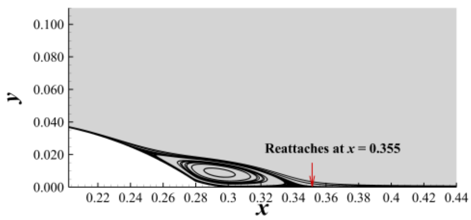
Figure 13. The geometry of bumps with different heights.

The mesh used in this test case is shown in A total of 72100 grid cells are used for CFD computation and the Δy^+ for the first layer of the grid is smaller than 1. At the inlet, the velocity profile is given. The static pressure is specified at the outlet.

The separation zones of $h = 42\ mm$ case predicted by the SST-DSR model, the SST model, and the LES [23] are shown in Figure 14. The SST model predicts an erroneously large separation zone, while the result given by the SST-DSR model is much closer to the LES data. The velocity profiles in the separation zone are shown in Figure 15. The result of the SST-DSR model aligned better with the LES data.

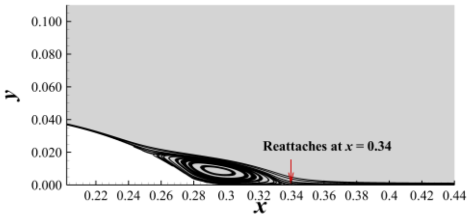


(a) SST



(b) SST-DSR

DOI: 10.13009/EUCASS2023-401



(c) LES [23]

Figure 14. Separation zone given by different method, $h = 42\ mm$

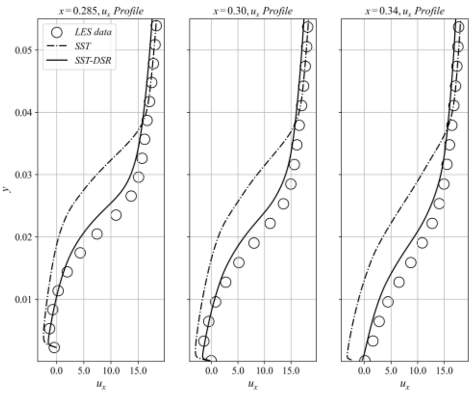


Figure 15. The velocity profiles at different x locations in the separation zone, $h = 42\ mm$

Figure 16 shows that β_{DSR} is increased over the separated shear layer, which is consistent with [23]'s observation that the non-equilibrium turbulence prevails above the separated shear layer in this case.

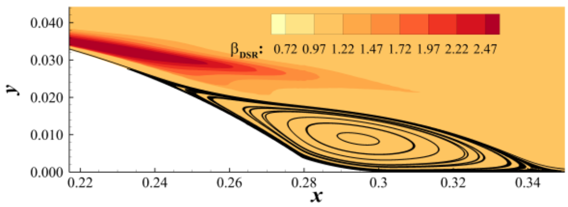
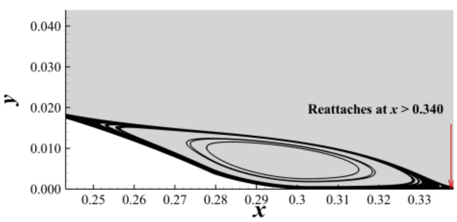
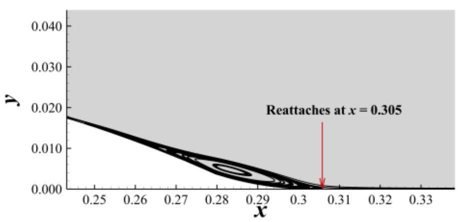


Figure 16. β_{DSR} distribution, $h = 42\ mm$

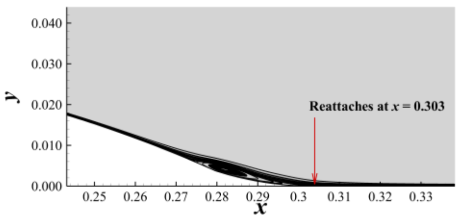
In the $h = 31\ mm$ case, a similar analysis can be carried out. The plot of the separation zone in Figure 17 shows that the SST-DSR model still predicts a smaller separation bubble compared with the SST model and agrees better with the LES data. Note that the size of the separation zone is much smaller than the $h = 42$ mm case and the CBFS case. The velocity profiles are shown in Figure 18, with the SST-DSR model outperforming the SST model substantially. The region where β_{DSR} is increased is similar to the $h = 42\ mm$ case, as shown in Figure 19.



(a) SST



(b) SST-DSR



DOI: 10.13009/EUCASS2023-401

(c) LES [23]

Figure 17. Separation zone given by different method, $h = 31\ mm$

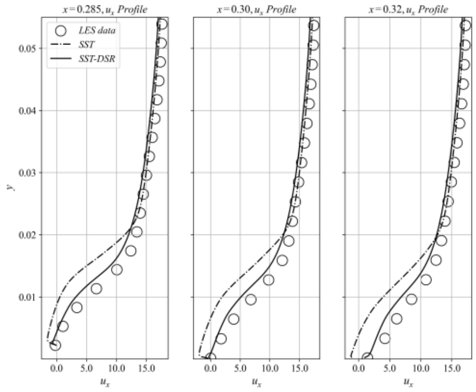


Figure 18. The velocity profiles at different x locations in the separation zone, $h = 31\ mm$

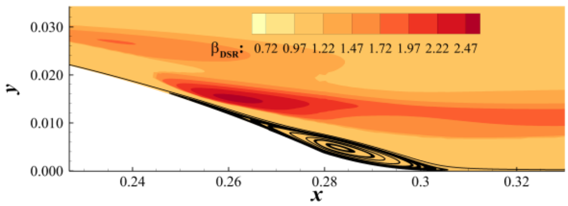


Figure 19. β_{DSR} distribution, $h = 31\ mm$

In summary, the 2D-bump test case illustrates that the SST-DSR model can generalize to completely different 2D geometry with different sizes of separation zone and different numerical values.

4.2 3D Ahmed's body

In this case, the performance of the SST-DSR model on 3D complex separated flow is tested by applying it to the 3D Ahmed body [24] (the simplified car body), which is a benchmark test case in the automobile industry. The shape of Ahmed's body is shown in Figure 20 (a). The slant angle is $\phi = 25°$, and the Reynolds number based on the body length is $Re_L = 2.78 \times 10^6$. We use a half model for CFD computation, the grid used is shown in Figure 20 (b). A total of 3.6×10^6 hexahedra cells are used.

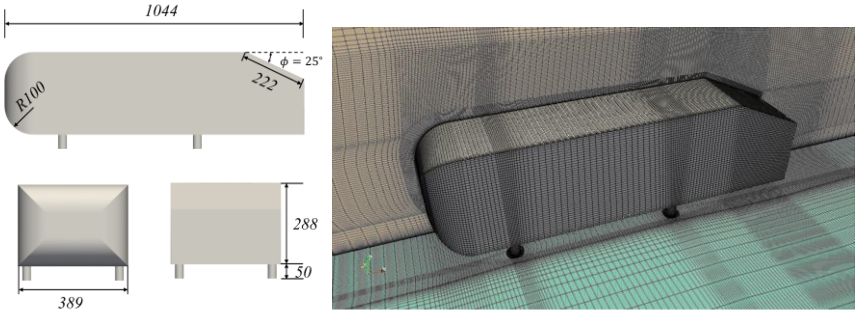
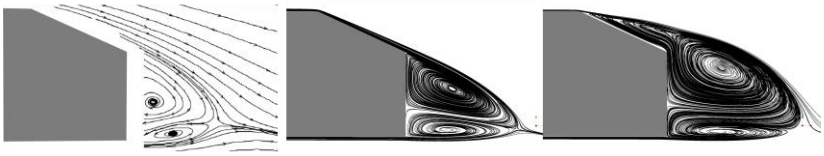


Figure 20. The shape of the Ahmed body and the grid are used for computation.

[3,25] shows that the SST model and the DES method predict a separation bubble that extends from the bottom of the body to the top of the body in the wake and the separation zone covers the entire slant. However, the experiment in [26] suggests that the flow remains largely attached throughout the slant and only separates mildly at the beginning of it. Moreover, the separation in the wake does not extend from the bottom to the top and is much smaller. Our computation using the SST model also shows the same large separation zone similar to [3,25], as shown in Figure 21. However, the separation zone given by the SST-DSR model is much smaller and agrees better with the PIV data [26]. Figure 22 shows the velocity profiles on the slant and in the wake. The velocity given by the SST-DSR model is closer to the experimental data than the result of the SST model.

DOI: 10.13009/EUCASS2023-401



(a) PIV [26] (b) SST-DSR (c) SST

Figure 21. Separation zone at the symmetry plane predicted by different methods

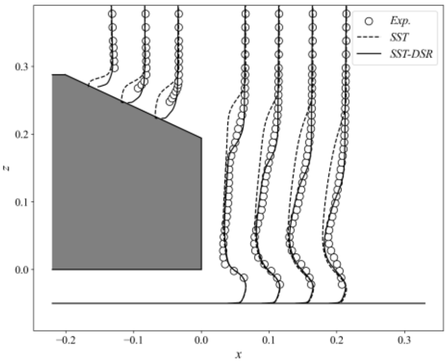
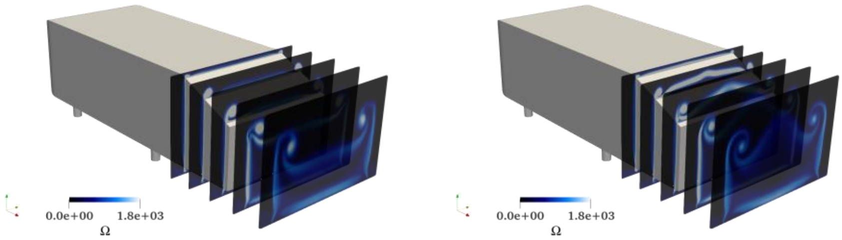


Figure 22. The velocity profiles at the symmetry plane

The 3-D structure of the separation zone near the wake is visualized by plotting the magnitude of $|\mathbf{\Omega}|$ at several planes perpendicular to the main flow direction, as shown in Figure 23. The vortex structure computed by the SST-DSR model resides lower than the one given by the SST model, and the C-pillar vortex (generated by the edge of the slant) in the SST-DSR model's result is stronger.



(a) SST-DSR (b) SST

Figure 23. The contour of $|\mathbf{\Omega}|$ at multiple stations in the wake.

The correction term β_{DSR} is activated on the slant where the adverse pressure gradient dominates, as shown in Figure 24. The increased β_{DSR} enhances momentum exchange, making the flow remains attached to the slant.

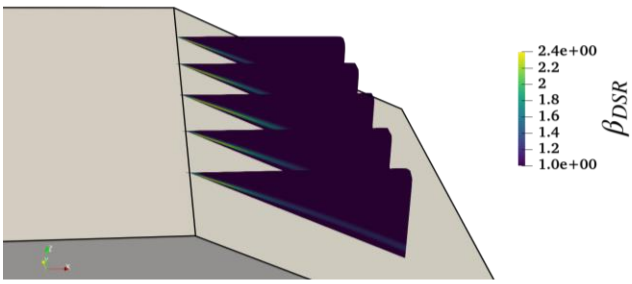


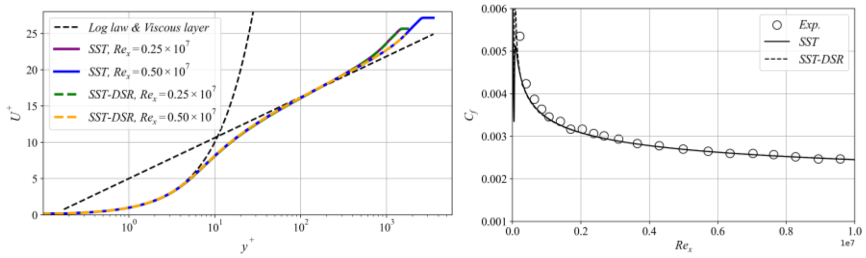
Figure 24. The β_{DSR} distribution on multiple $y = const.$ planes

In summary, the SST-DSR model gives better results compared with the SST model, demonstrating its ability to generalize to 3-D complex separated flows.

DOI: 10.13009/EUCASS2023-401

4.3 Turbulent boundary layer on a flat plate, $Re_L = 1 \times 10^7$

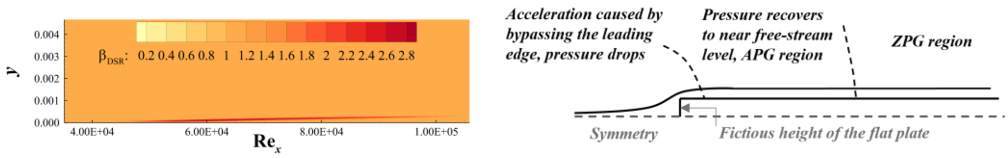
In this test case, we apply the SST-DSR model to calculate the zero-pressure-gradient turbulent boundary layer on a flat plate, in which the SST model can already give satisfactory results. The Reynolds number based on the plate's length is $Re_L = 1 \times 10^7$ and the height of the first grid layer satisfies $\Delta y^+ \approx 0.05$. The velocity profiles at $Re_x = 0.25 \times 10^7$ and $Re_x = 0.5 \times 10^7$ are shown in Figure 25(a). The result of the SST-DSR model almost coincides with the result of the SST model, both agree well with the logarithmic law and the linear law in the viscous sublayer. The friction coefficient is shown in Figure 25(b), with the SST-DSR model giving results similar to the SST model. The experimental data is from [27]. This suggests that the SST-DSR model does not degrade the baseline SST model's performance on simple elementary flows.



(a) Velocity profiles (b) C_f along the flat plate

Figure 25. The velocity profile and the friction coefficient

The friction coefficient given by the SST-DSR model is a bit higher at the leading edge. This is because the correction term β_{DSR} is activated near the leading edge where $Re_x < 1 \times 10^5$, leading to an increased k that renders higher friction. The increased β_{DSR} at the leading edge is believed to be caused by the slight adverse pressure gradient near the leading edge, which is illustrated in Figure 26.



(a) (b)

Figure 26. (a)The distribution of β_{DSR} (b) The adverse pressure gradient near the leading edge

5. Conclusion

In this study, we introduce a unique data-driven methodology employing field inversion and deep symbolic regression for creating a generalizable turbulence model for separated flows. The generated dataset via field inversion is trained with the deep symbolic regression algorithm, yielding an expression with distinct physical interpretation. This expression is adjusted to align with our physical a priori, subsequently integrated into the SST model, resulting in the SST-DSR model.

The SST-DSR model excels in performance over the original SST model for both the CBFS case (training case) and the 2D-bump case, signifying its effective predictions on separated flows across various geometries. Furthermore, it predicts separation in the wake of the Ahmed body more accurately than the SST model, emphasizing its capacity to adapt to 3D complex flows. In the flat-plate scenario, the SST-DSR model aligns closely with the SST model in predicting velocity profiles and the friction coefficient of the zero-pressure-gradient turbulent boundary layer, further corroborating its robustness.

Overall, the SST-DSR model, formulated through deep symbolic regression, demonstrates remarkable generalizability - a characteristic not yet reported in current black-box data-driven turbulence models, as known by the authors. This model's strong performance suggests the potential of merging field inversion and symbolic regression for developing physically interpretable and widely applicable turbulence models.

References

[1] Rumsey, C. L. "Exploring a Method for Improving Turbulent Separated-Flow Predictions with k-ω Models." *NASA TM-2009-215952*, 2009.

[2] Li, H., Zhang, Y., and Chen, H. "Aerodynamic Prediction of Iced Airfoils Based on Modified Three-Equation Turbulence Model." *AIAA Journal*, Vol. 58, No. 9, 2020, pp. 3863–3876. https://doi.org/10.2514/1.J059206.

DOI: 10.13009/EUCASS2023-401

[3] Guilmineau, E., Deng, G. B., Leroyer, A., Queutey, P., Visonneau, M., and Wackers, J. Aessessment of RANS and DES Methods for The Ahmed Body. Presented at the VII European Congress on Computational Methods in Applied Sciences and Engineering, Crete Island, Greece, 2016.

[4] Li, H., Zhang, Y., and Chen, H. "Numerical Simulation of Iced Wing Using Separating Shear Layer Fixed Turbulence Models." *AIAA Journal*, Vol. 59, No. 9, 2021, pp. 3667–3681. https://doi.org/10.2514/1.J060143.

[5] Li, H., Zhang, Y., and Chen, H. "Optimization of Supercritical Airfoil Considering the Ice-Accretion Effects." *AIAA Journal*, Vol. 57, No. 11, 2019, pp. 4650–4669. https://doi.org/10.2514/1.J057958.

[6] Xiao, H., Wu, J.-L., Wang, J.-X., Sun, R., and Roy, C. J. "Quantifying and Reducing Model-Form Uncertainties in Reynolds-Averaged Navier-Stokes Simulations: A Data-Driven, Physics-Based Bayesian Approach." *Journal of Computational Physics*, Vol. 324, 2016, pp. 115–136. https://doi.org/10.1016/j.jcp.2016.07.038.

[7] Singh, A. P., and Duraisamy, K. "Using Field Inversion to Quantify Functional Errors in Turbulence Closures." *Physics of Fluids*, Vol. 28, No. 4, 2016, p. 045110. https://doi.org/10.1063/1.4947045.

[8] Yin, Y., Yang, P., Zhang, Y., Chen, H., and Fu, S. "Feature Selection and Processing of Turbulence Modeling Based on an Artificial Neural Network." *Physics of Fluids*, Vol. 32, No. 10, 2020, p. 105117. https://doi.org/10.1063/5.0022561.

[9] Yan, C., Li, H., Zhang, Y., and Chen, H. "Data-Driven Turbulence Modeling in Separated Flows Considering Physical Mechanism Analysis." *International Journal of Heat and Fluid Flow*, Vol. 96, 2022, p. 109004. https://doi.org/10.1016/j.ijheatfluidflow.2022.109004.

[10] Yan, C., Zhang, Y., and Chen, H. "Data Augmented Turbulence Modeling for Three-Dimensional Separation Flows." *Physics of Fluids*, Vol. 34, No. 7, 2022, p. 075101. https://doi.org/10.1063/5.0097438.

[11] Rumsey, C. L., Coleman, G. N., and Wang, L. In Search of Data-Driven Improvements to RANS Models Applied to Separated Flows. Presented at the AIAA SCITECH 2022 Forum, San Diego, CA & Virtual, 2022.

[12] Cranmer, M. Interpretable Machine Learning for Science with PySR and SymbolicRegression.Jl. http://arxiv.org/abs/2305.01582. Accessed May 31, 2023.

[13] Petersen, B. K., Landajuela, M., Mundhenk, T. N., Santiago, C. P., Kim, S. K., and Kim, J. T. Deep Symbolic Regression: Recovering Mathematical Expressions from Data via Risk-Seeking Policy Gradients. http://arxiv.org/abs/1912.04871. Accessed May 31, 2023.

[14] Schmelzer, M., Dwight, R. P., and Cinnella, P. "Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression." *Flow, Turbulence and Combustion*, Vol. 104, Nos. 2–3, 2020, pp. 579–603. https://doi.org/10.1007/s10494-019-00089-x.

[15] Xie, H., Zhao, Y., and Zhang, Y. "Data-Driven Nonlinear K-L Turbulent Mixing Model via Gene Expression Programming Method." *Acta Mechanica Sinica*, Vol. 39, No. 2, 2023, p. 322315. https://doi.org/10.1007/s10409-022-22315-x.

[16] Menter, F. R., Kuntz, M., and Langtry, R. "Ten Years of Industrial Experience with the SST Turbulence Model." *Heat and Mass Transfer*, Vol. 4, No. 1, 2003, pp. 625–632.

[17] Gill, P. E., Murray, W., and Saunders, M. A. "SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization." *SIAM Review*, Vol. 47, No. 1, 2005, pp. 99–131. https://doi.org/10.1137/S0036144504446096.

[18] Duffy, A. An Introduction to Gradient Computation by the Discrete Adjoint Method. 2009.

[19] He, P., Mader, C. A., Martins, J. R. R. A., and Maki, K. J. "DAFoam: An Open-Source Adjoint Framework for Multidisciplinary Design Optimization with OpenFOAM." *AIAA Journal*, Vol. 58, No. 3, 2020, pp. 1304–1319. https://doi.org/10.2514/1.J058853.

[20] Pope, S. B. "A More General Effective-Viscosity Hypothesis." *Journal of Fluid Mechanics*, Vol. 72, No. 02, 1975, p. 331. https://doi.org/10.1017/S0022112075003382.

[21] Bentaleb, Y., Lardeau, S., and Leschziner, M. A. "Large-Eddy Simulation of Turbulent Boundary Layer Separation from a Rounded Step." *Journal of Turbulence*, Vol. 13, 2012, p. N4. https://doi.org/10.1080/14685248.2011.637923.

[22] OpenFOAM: User Guide: OmegaWallFunction. https://www.openfoam.com/documentation/guides/v2112/doc/guide-bcs-wall-turbulence-omegaWallFunction.html. Accessed Apr. 18, 2023.

[23] Matai, R., and Durbin, P. "Large-Eddy Simulation of Turbulent Flow over a Parametric Set of Bumps." *Journal of Fluid Mechanics*, Vol. 866, 2019, pp. 503–525. https://doi.org/10.1017/jfm.2019.80.

[24] Ahmed, S. R., Ramm, G., and Faltin, G. Some Salient Features Of The Time-Averaged Ground Vehicle Wake. Presented at the SAE International Congress and Exposition, 1984.

[25] Haase, W., Ed. *FLOMANIA: A European Initiative on Flow Physics Modelling: Results of the European-Union Funded Project, 2002-2004*. Springer, Berlin ; New York, 2006.

[26] Lienhart, H., Stoots, C., and Becker, S. "Flow and Turbulence Structures in the Wake of a Simplified Car Model (Ahmed Model)."

[27] Wieghardt, K., and Tillmann, W. *On the Turbulent Friction Layer for Rising Pressure*. Publication NACA-TM-1314.