32nd Congress of the International Council of the Aeronautical Sciences
September 6-10, 2021
Pudong Shangri-La, Shanghai, China

RESEARCH OF TURBULENCE MODELING BASED ON MACHINE LEARNING

Yin Yuhui[1], Zhang Yufei[1]

[1]Tsinghua University，Beijing 100084, China

Abstract

Improving the prediction accuracy of separation flow based on Reynolds-averaged N-S equation is a widely concerned research problem. In this paper, the random forest and neural network are adopted to train the mapping model from the average characteristic of the flow field to the Reynolds stress. The linear eddy viscosity model is used as the baseline model and the corresponding results are selected as input variables. The results of a modified transition model and the DNS results, which have better agreement with the true values, are selected as true values. Firstly, certain mean flow characteristic are selected as the input of the machine learning algorithms according to the mechanism of turbulence evolution. The discrepancy of eddy viscosity and Reynolds stress eigenvalues/eignvectors are respectively selected as targets. The results show that the feature selection and processing, and the local regional modeling strategy have benefits on the convergence and prediction accuracy of the training model. The prediction of the eddy viscosity and Reynolds stress and the resulting mean flow fields can both have better performance, and the smoothness problem needs further concern.

Keywords: turbulence modeling; Navier-Stokes equations; machine learning; feature selection; flow separation

1. Introduction

It is of great significance to understand the formation and evolution mechanism of turbulence structure and to predict turbulence accurately in modern engineering problems. At present, due to the limitation of the computation technology, high-precision numerical methods such as direct numerical simulation (DNS) or large eddy simulation (LES) are still difficult to solve practical engineering problems. Reynolds-averaged N-S equation (RANS) is still the most commonly used numerical simulation method in aerodynamic design. However, many complex flow phenomena such as swirl, strong pressure gradient flow, large curvature of streamline, etc., can not be precisely simulated using RANS method to give the results that meet the requirements of engineering designs[1]. Taking the flow separation problem in flight as an example, stall at large angle of attack, separation caused by shock / boundary layer interference and deterioration of lift & drag performance after icing on the surface area of wing have a decisive influence on the flight safety. However, for the RANS method, the researchers compared several common turbulence models and found that they can not accurately predict the flow separation caused by the strong adverse pressure gradient on the upper surface of the airfoil [2]. For the calculation error of RANS method, the researchers agreed that the difference of the predicted Reynolds stress are mainly to blame. To improve the prediction effect, we need to develop a more accurate model of Reynolds stress prediction, that is, to find the mapping relationship between the Reynolds stress and the mean flow [3][4].

The earliest and still commonly used Reynolds stress closure method is the linear eddy viscosity model (LEVM), which assumes that the Reynolds stress is proportional to the mean flow strain rate, and determines the proportional coefficient, referred to as eddy viscosity, based on the empirical correlation or solving the governing equations. The turbulence model based on LEVM has many advantages, such as good computational robustness, low computational time cost, and good simulation effect for most flows with attached boundary layer. However, due to ignoring the influence of streamline curvature and the actual anisotropy of Reynolds stress, LEVM can not predict complex flows, for an example, the calculation of secondary flow in square tube flow [5] and separation zone in periodic hill flow [6]. In order to overcome the limitation of LEVM, the nonlinear eddy viscosity model and the Reynolds stress model are proposed. These models can improve the prediction accuracy in some complex flows. However, the additional terms and coefficients added in the nonlinear eddy viscosity model may vary with the flow problem and be very sensitive, while the Reynolds stress model increases the calculation cost and has poor robustness [7]. To sum up, there is no turbulence calculation model that can not only give accurate results of different types of complex flows, but also maintain robustness and save cost. However, this problem is difficult to be completely solved by the traditional turbulence modeling framework.

A large number of high-fidelity turbulence data have been produced in decades of turbulence research, including the numerical results of DNS and LES, as well as the experimental observation results. At the same time, with the development of statistical inference algorithm and machine learning technology, it is possible to use these data to assist turbulence modeling. This framework is referred to as data-driven modeling. The current development ideas of data-driven methods can be summarized into the following three categories [8][9]:

1) The uncertainty quantification of turbulence model in RANS method. The purpose of this category is to quantify the uncertainty in the Reynolds stress tensor [10][11], and identify the areas with high uncertainty in the flow field [12] by statistical inference method, so as to help constrain the uncertainty and identify the specific sources of RANS prediction discrepancy.

2) The inference of the coefficients and the terms of generation and dissipation based on the turbulence data. In this category, the quantities of interest (QoIs), which have a key influence on turbulence calculation, are selected as the inference variables, which can be the model coefficient [13] or the terms of turbulence model governing equation [14] [15] (its initial value is usually called a priori result). Then, these selected variables are regarded as random fields of the flow distribution and are varied in order to acquire certain flow field that matches the given high-fidelity data (such as surface pressure distribution [15] or velocity value at observation position [16]). If the selected attention is reasonable, the flow results calculated by using the modified attention (called posterior results) will be more consistent with the target.

3) Modification of the turbulence model using machine learning strategy. The purpose of this category is to establish the mapping between the mean flow characteristics and the Reynolds stress characteristics. Combined with big data, machine learning method can extract multi-level features and give regression or classification results without a lot of prior knowledge. This powerful feature extraction and regression ability makes it suitable for building turbulence model mapping.

Among the three categories mentioned above, the first two can't produce prediction models, that is, they can't be used to improve the cases that haven't appeared in the process of training and inference, while the machine learning model based on big data training can be extrapolated to the problems with the same type of flow phenomena and flow structures. The generalization ability of machine learning model makes it more valuable for engineering application, and helps to reduce the cost of experiment or DNS calculation. Therefore, this study adopts the third research idea, that is, using machine learning method to improve the prediction results of RANS model.

In recent years, turbulence modeling based on machine learning has received extensive attention. The selection of training target reflects researchers' different understanding of the main sources of turbulence prediction errors and how to deal with them, which can be used to classify the related research.

The first idea is to modify the relative magnitude of some terms in the original turbulence model equations. Tracey et al. [17] used machine learning to reconstruct the source term in S-A model, and verified the feasibility of this method. Zhu et al. [18] established an artificial neural network directly aiming at the eddy viscosity calculated by S-A model. Using network instead of the original model improves the computational efficiency and reduces the dependence on grid density. The above work is to reproduce the results of the baseline model. If the goal is to improve the effect of the original turbulence model in complex problems, Bayesian inversion method should be used to infer the modified source term first, and then machine learning method is used for learning a predictive model [15][19][20]. In these studies, researchers modify the model equation and the mean flow equation by multiplying a coefficient varying in the full field to the generation term or dissipation term. Because this method affects the Reynolds stress through the transport equation, it is easier to achieve the smoothness of the modified Reynolds stress field. However, the eddy viscosity assumption used in the baseline turbulence model limits the anisotropy of the Reynolds stress. No matter how the amplitude changes, the principal direction of the Reynolds stress is always consistent with the strain rate tensor. Therefore, there are still some differences between the current theoretical optimal results and the real Reynolds stress.

The second view is that the whole Reynolds stress should be chosen as the training target. In Ling's work [21], Reynolds stress is expressed as a linear combination of ten integral bases consisting of strain rate tensor and rotation rate tensor of mean flow. The output variables of the neural network are the combined coefficients, and then combined with the corresponding tensor bases to form the Reynolds stress anisotropy tensor. This idea similar to the nonlinear eddy viscosity model inspired some later studies [22]. In addition to linear combination, another representation of Reynolds stress is eigenvalue decomposition. Since Reynolds stress is a symmetric second-order tensor, eigenvalue decomposition will give a set of real eigenvalues and orthogonal eigenvectors, which can be further described by Euler angle [23] or quaternion [24]. The real Reynolds stress and the Reynolds stress predicted by RANS are decomposed to obtain the difference field of each feature, which is used as the target of machine learning.

Compared with the two methods, the method based on a baseline turbulence model has less variables to be modified and is less difficult to predict, but it can not predict the complete Reynolds stress characteristics; The direct representation and learning of Reynolds stress can reproduce the real results to the greatest extent in theory, but the quantity to be predicted becomes more and more. At the same time, because the predicted quantity directly affects the mean flow equation, it also brings the problem of smoothness.

This paper will study the possibility of applying this framework to the typical and key separated flow in practical engineering problems such as the periodic hill and complex flow around icing airfoil. The research mainly focuses on the problem of input / output feature selection in the process of building machine learning prediction framework, and the difficulty of model prediction caused by uneven data distribution in the flow around the outflow. In this paper, the selection criteria of input and output characteristics are proposed, and the local regional modeling of airfoil flow field is carried out to verify the feasibility of local freezing. The method is applied to the flow calculation of periodic hills and the icing airfoil, and the prediction accuracy of the original turbulence model is improved.

2. Methods

2.1 Machine learning assisted turbulence modeling

Firstly, the framework of machine learning aided turbulence modeling used in this paper is described. The RANS equations of three-dimensional compressible flow are shown in equation (1).

$$\begin{aligned}
&\frac{\partial \rho}{\partial t}+\nabla\cdot(\rho \boldsymbol{u})=0\\
&\frac{\partial}{\partial t}(\rho \boldsymbol{u})+\nabla\cdot(\rho \boldsymbol{u}\boldsymbol{u})=\rho \boldsymbol{f}+\nabla\cdot \boldsymbol{T}\\
&\frac{\partial}{\partial t}(\rho e)+\nabla\cdot(\rho e\boldsymbol{u})=\nabla\cdot(\boldsymbol{T}\cdot\boldsymbol{u})+\rho \boldsymbol{f}\cdot\boldsymbol{u}\\
&\boldsymbol{T}=-p\boldsymbol{I}+\left[\lambda(\nabla\cdot\boldsymbol{u})\boldsymbol{I}+\mu(\nabla\boldsymbol{u}+\boldsymbol{u}\nabla)\right]+\boldsymbol{\tau}
\end{aligned} \tag{1}$$

Among them, ρ、 u、f、e、p、 λ and μ are separately density, velocity, volume force, specific energy, pressure, volume viscosity and molecular viscosity. **T** represents the total stress tensor, including pressure, molecular viscous stress and Reynolds stress due to Reynolds averaging **τ**. In traditional turbulence modeling, Reynolds stress is determined by turbulence model equation. When both RANS equation and turbulence model equation converge, the solution process ends. The role of machine learning in turbulence modeling is to provide direct or indirect prediction results of Reynolds stress given input characteristics, that is, to provide a regression function with predictive ability. To establish this regression function, a large number of high-fidelity data are needed. This process is called machine learning training process. After the training, the function can be used to predict the new unseen flow, which is called test process. It reflects the generalization ability of machine learning model on unknown problems. The selection of prediction target, input and output characteristics and machine learning algorithm have a decisive impact on the prediction performance.

The choice of prediction target varies with the type of learning truth value. In this paper, we select two typical flows: (1) periodic hill, (2) the iced airfoil. The first case has been simulated using DNS and the data can be accessed online. As for the second case, there is a lack of high confidence DNS and Les simulation results in this aspect. The truth value selected in this study is the SPF modified three equation k-v^2- ω developed by the research group (referred to as SPF k-v^2- ω below). The model is applied to the flow prediction of typical icing airfoil at high Reynolds number, and the force coefficient curve, flow field fluctuation and average measurement results are in good agreement [26]. The three equation model is still based on the framework of the linear eddy viscosity model, so the eddy viscosity field of the iterative convergence result of the model is taken as the true value and recorded as. After obtaining the true value, the prediction target is taken as the difference field of the Reynolds stress in the periodic hill case and the eddy viscosity in the iced airfoil case.

No matter the eddy viscosity or the Reynolds stress are selected as the target, the input features all come from the calculation results of the baseline model. When it is substituted into the average equation, it needs to be frozen in the iterative process until the residual of the average equation converges again, which means that the machine learning model is only called once before the second iteration.

The turbulence modeling framework assisted by machine learning is arranged as follows, and is shown in Figure 1.

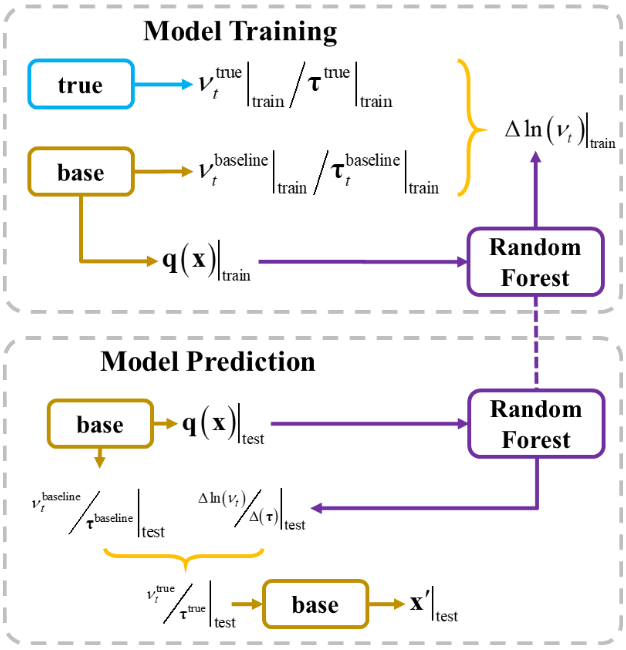


Figure 1 – Schematic of the machine learning assisted turbulence modeling framework.

2.2 Input feature selection based on tensor analysis and physical feature analysis

According to the description in Section 2.1, the input features of the machine learning framework used in this paper should be constructed from the mean flow features calculated by the baseline turbulence model. These input characteristics should also meet the following three basic principles:

1) Completeness, that is, the input feature set should include all possible information related to Reynolds stress distribution;

2) Compactness, that is, invalid or redundant information in the input feature set should be identified and deleted;

3) Realizability, that is, as long as the flow structure remains unchanged, the variables of the input feature set should be consistent in different situations (such as different reference coordinate systems or flow directions).

Satisfying the first two principles can improve the performance of machine learning model and make it more extrapolative and accurate; The third principle ensures that the data-driven model has the same physical rationality as the traditional turbulence model. There have been many studies on the selection of input characteristics [27] - [29]. In this paper, two selection criteria based on tensor analysis and physical feature recognition are proposed. The inputs in previous studies can be classified into these two perspectives. At the same time, some new effective features can be constructed and added to the input set under the guidance of the criteria.

2.2.1 Feature selection based on Reynolds stress representation

Firstly, the tensors or vectors on which the Reynolds stress depends are found, and complete tensor bases are constructed to represent the Reynolds stress. The nonlinear eddy viscosity assumption is a typical turbulence model based on the representation of Reynolds stress, which assumes the Reynolds stress as the combination of polynomials formed by the mean flow strain rate **S** and the rotation rate **Ω**. Pope [30] used tensor function representation theory to deduce the integral bases containing 10 components, thus obtaining the compact form of polynomial. Inspired by this, Wu et al. [24] further added the pressure gradient Δp and turbulent kinetic energy gradient Δk in order to supplement the influence of strong pressure gradient and non-uniform distribution of turbulence intensity. For convenience of representation, these two gradients are transformed into corresponding antisymmetric tensors $\mathbf{A}_p$ and $\mathbf{A}_k$. According to the tensor representation theorem, the integral bases of 47 components are obtained. Because the components of integral basis are tensors, they can not be used as input features directly, therefore their traces are adopted. For the two dimensional problems, the repetitive features with zero in two dimensions and the same absolute value are removed from the above features (at this time, the remaining 17 features). In addition, for the iced airfoil case, due to the complexity of the airfoil mesh and the characteristics of external flow, it is difficult to achieve high smoothness, so some high-order features which are very sensitive to the smoothness of the mesh are further removed, and finally the remaining 6 features are left, As shown in Table 1（$|\cdot|$ represents the vector norm and $\|\cdot\|$ the Frobenius matrix norm. variable ε Is the turbulent dissipation rate.)

Table 1 Input features selected based on Reynolds stress representation

$(n_{\mathrm{sym}}, n_{\mathrm{anti}})$	Feature No.	Invariants
(1, 0)	q_1	$\hat{\boldsymbol{S}}^2$
(0, 1)	q_2, q_3	$\hat{\boldsymbol{A}}_p^2$, $\hat{\boldsymbol{A}}_k^2$
(1, 1)	q_4	$\hat{\boldsymbol{A}}_k^2\hat{\boldsymbol{S}}$
(0, 2)	q_5	$\hat{\boldsymbol{A}}_p\hat{\boldsymbol{A}}_k$
(1, 2)	q_6	$\hat{\boldsymbol{A}}_p^2\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{S}}$

where $\hat{\boldsymbol{S}} = \frac{\boldsymbol{S}}{\|\boldsymbol{S}\| + \varepsilon / k}$, $\hat{\boldsymbol{\Omega}} = \frac{\boldsymbol{\Omega}}{2\|\boldsymbol{\Omega}\|}$, $\hat{\boldsymbol{A}}_k = \frac{\boldsymbol{A}_k}{\|\boldsymbol{A}_k\| + \varepsilon / \sqrt{k}}$, $\hat{\boldsymbol{A}}_p = \frac{\boldsymbol{A}_p}{\|\boldsymbol{A}_p\| + \rho|(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}|}$, n_{sym} and n_{anti} correspond to the number of symmetric and anti-symmetric tensors used in the current row.

2.2.2 Feature selection based on flow identification

Since the input features are constructed entirely based on the baseline mean field, it is difficult to give a reasonable expression when the difference between the true value and the baseline value is large. In order to further supplement the input features, a second perspective is introduced, that is, to locate and quantitatively describe the large difference between the baseline and the true value by finding or constructing marking variables. Compared with the above representation method, it is more direct to find and identify the difference markers as input, and has higher correlation with the prediction target. In addition to the function of marking, the selected features should also have a clear physical meaning to ensure similar performance in flows with similar flow structures. On the one hand, it can be realized by constructing markers identifying the key flow characteristics, because the difference between the reference value and the true value often appears in the complex flow structure. Taking the airfoil with horn-shaped ice as an example, the flow separation bubble is formed after ice and attached to the airfoil. The prediction error of the baseline turbulence model mainly comes from the prediction of shear layer, strong reverse pressure gradient and swirl in the separation bubble. The region to be corrected can be found by identifying the above flow structure, which is also used in the traditional modeling. On the other hand, the relative ratio of turbulence related physical quantities can be constructed to reflect

the identification degree of the baseline model for key flow phenomena. Adding it into the input quantity can better locate the error of the baseline turbulence model and give the correction direction. After considering the feature smoothness, six scalar features are constructed as follows.
1) Identification function of flow shear layer and vortex flow [14]: $f_1 = d^2\Omega / \left(\nu_t + \nu\right)$, where D is the distance to the wall, ν_t is the eddy viscosity predicted by the baseline turbulence model, $\Omega = \sqrt{2\Omega_{ij}\Omega_{ij}}$;
2) Boundary layer identification function (protection function used by DDEs) [31]: $f_2 = 1 - \tanh\left(\left[8r_d\right]^3\right)$, where $r_d = \dfrac{\nu_t + \nu}{\kappa^2 d^2 \sqrt{u_{i,j}u_{i,j}}}$;
3) The ratio of turbulent kinetic energy to mean flow kinetic energy: $f_3 = 2k / \left(u^2 + v^2 + w^2\right)$;
4) Turbulent Reynolds number based on wall distance [32]: $f_4 = \dfrac{\sqrt{k}d}{50\nu}$;
5) The ratio of turbulence time scale to mean flow time scale: $f_5 = Sk / \varepsilon$;
6) The ratio of eddy viscosity to molecular viscosity: $f_6 = \nu_t / \nu$.
To sum up, 12 normalized invariants are constructed from the mean flow variables as the input features of the machine learning framework. Due to the characteristics of random forest method, the importance of the constructed feature set can be ranked, especially the importance of the last six features added based on flow structure recognition. The results will be presented in Section 4.

2.3 Local regional modeling for airfoil flow.

Most of the previous researches on the machine learning assisted turbulence modeling framework have focused on the internal flow. The internal flow and the external flow may be relatively close in key flow structures (such as the separation of periodic hill flow and the separation of flow behind an iced airfoil). However, in order to ensure that the far-field boundary conditions have little influence on the flow field near the flow object, the external flow is generally large. The characteristic length of the far field needs to be 50-100 times of that of the flow object. Even if the grid becomes sparse, the flow data far away from the wall still occupy a large part of the flow field data. Because this part of the flow data is not disturbed too much, most of the input characteristics are consistent and there is no clear difference, and the difference between the predicted Reynolds stress and the true value is small, so it has a negative impact on the training of machine learning model. In order to solve this problem, Zhu et al. [18] used the partition modeling method when reconstructing the S-A model in the whole field. According to the distance to the wall, the whole flow field was divided, and different regions were trained separately. In this problem, this paper further considers that only the region near the wing is used for training, while other regions are still calculated using the baseline turbulence model, which is different from the original regional modeling method. The idea of this paper can be called local regional modeling. Compared with partition modeling, local region modeling firstly reduces the training cost of machine learning model, because far-field training is not needed; In addition, local area modeling can reasonably limit the influence area of machine learning algorithm, and the influence range of machine learning model is basically constrained near the wall. For the complex combination of flow problems, this idea can ensure that the error of machine learning will not affect the original small difference area, causing additional error.
Next, the idea of local region modeling is explained in detail from the perspective of model training and substitution calculation. Local region modeling is easy to deal with from the perspective of machine learning model training. In this paper, according to the significant region size of the predicted target quantity (difference field) in space, the split wall distance range is determined, and according to the ijk index range of the generated airfoil structure mesh, only the flow data in this region is used in the training and prediction of machine learning model.
For local regional modeling, it is an important problem to couple the prediction model with the CFD mean flow calculation process. The basic idea adopted in this paper is to still calculate the turbulence model equation in the whole field. When the eddy viscosity needs to be substituted into the mean flow equation in each iteration step, the eddy viscosity field in the designated area is replaced by the frozen value and frozen, while the rest still uses the results calculated by turbulence model. This method can freely specify the region to be replaced without modifying the solution region and the corresponding boundary conditions of the turbulence model equation, and is easy to be implemented in the program; And through the calculation verification, the smooth average velocity field can be obtained in the case of convergence. However, this local freezing method challenges the convergence of CFD. Due to the

great difference between the frozen eddy viscosity and that calculated by the baseline turbulence model, there are discontinuities on the boundary of the zone. The verification calculation shows that the direct freezing will cause the turbulence model equation to diverge quickly in the iteration process, and it can not be improved by adjusting the CFL number or increasing the iteration steps.

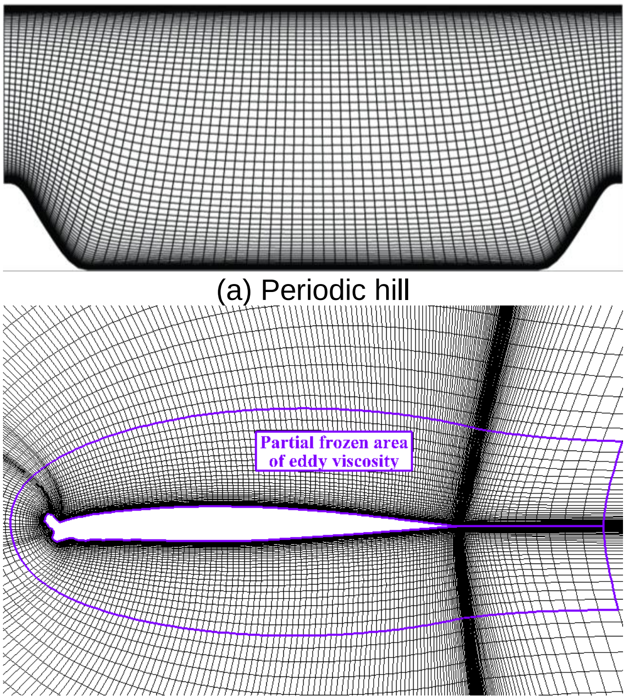
In order to solve the problem of divergence in CFD calculation of local freezing region, the idea of mixing is adopted: in the local freezing region, the eddy viscosity predicted by the machine learning model and the eddy viscosity calculated by the baseline turbulence model are mixed according to a certain proportion, and the mixing coefficient is recorded as λ Hybrid, as shown in equation (2), when λ When hybrid is 0, it means that it is completely frozen, and when hybrid is 1, it means that it is not substituted into all the baseline calculation results.

$$
\begin{aligned}
\nu_t &= \lambda_{\text{hybrid}} \nu_t^{\text{baseline}} + \left(1-\lambda_{\text{hybrid}}\right) \nu_t^{\text{ML}} \\
\boldsymbol{\tau} &= \lambda_{\text{hybrid}} \boldsymbol{\tau}^{\text{baseline}} + \left(1-\lambda_{\text{hybrid}}\right) \boldsymbol{\tau}^{\text{ML}}
\end{aligned}
\tag{2}
$$

In the actual calculation, the convergence results of the configuration to be modified are calculated by using the benchmark turbulence model, and then several steps are calculated by using the above hybrid freezing method, and then the frozen eddy viscosity is completely replaced. The calculation results show that this treatment can effectively solve the problem of turbulence model equation divergence caused by the large initial gap between the inside and outside of the frozen region.

3. Numerical Verification

In this part, we will verify the fully frozen and partial frozen methods of the eddy viscosity. In this process, we will not introduce the machine learning model, but directly use the truth value to calculate, that is, to obtain the theoretical upper limit that the current method can get, and investigate the rationality of the method. The canonical flow separation case periodic hill with five different hill slopes[33] and the flow around GLC305 airfoil with 944 ice shape [34] (denoted as GLC305 / 944) are used in the selected example. The schematic of the grids is shown in Figure 2. In this paper, CFL3D v6.7 [35], an open source program for compressible flow, is used to solve the flow, and the functions of freezing the full / partial eddy viscosity or Reynolds stress and mixing with the original values are added to the program.

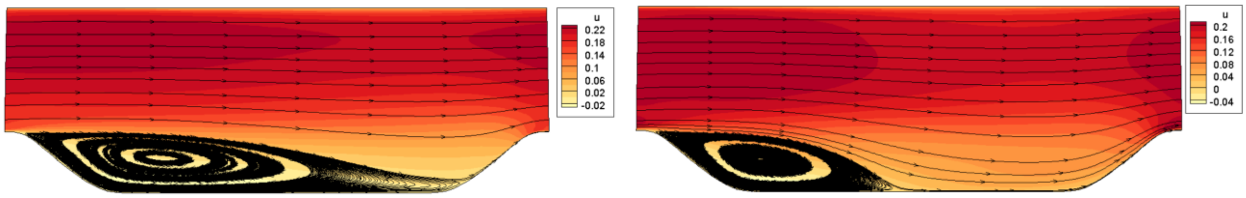


(a) Periodic hill

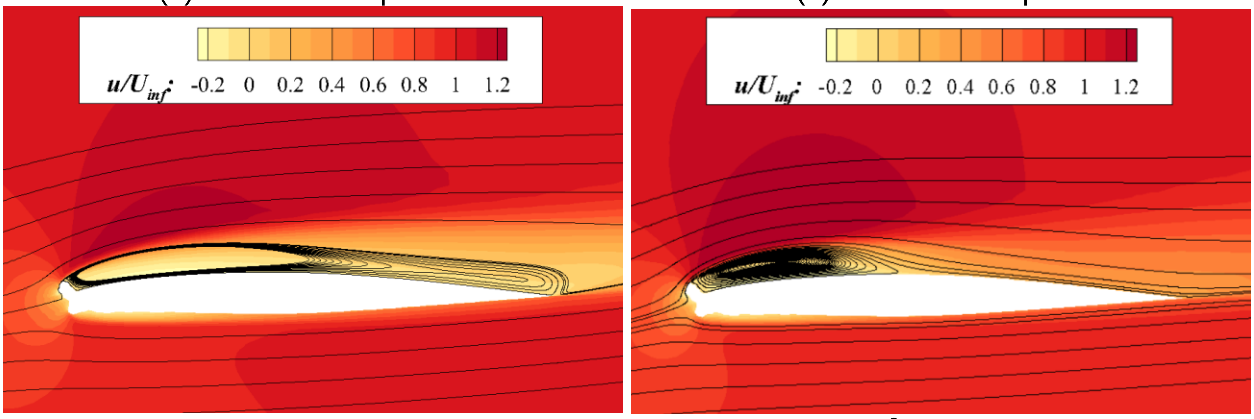
(b) GLC305/944 airfoil

Figure 2 – Schematic of the computation grid of periodic hill and GLC305/944 airfoil with the partial frozen area.

First, use SST model to simulate the cases. The results will be used as the baseline results in the follow-up machine learning training and prediction. For the periodic hill case, the DNS results provided in [33] are regarded as the true value. For the GLC305/944 airfoil case, the SPF k-v^2- ω is used to compute the true value. The comparison of baseline results and true values are shown in Figure 3.



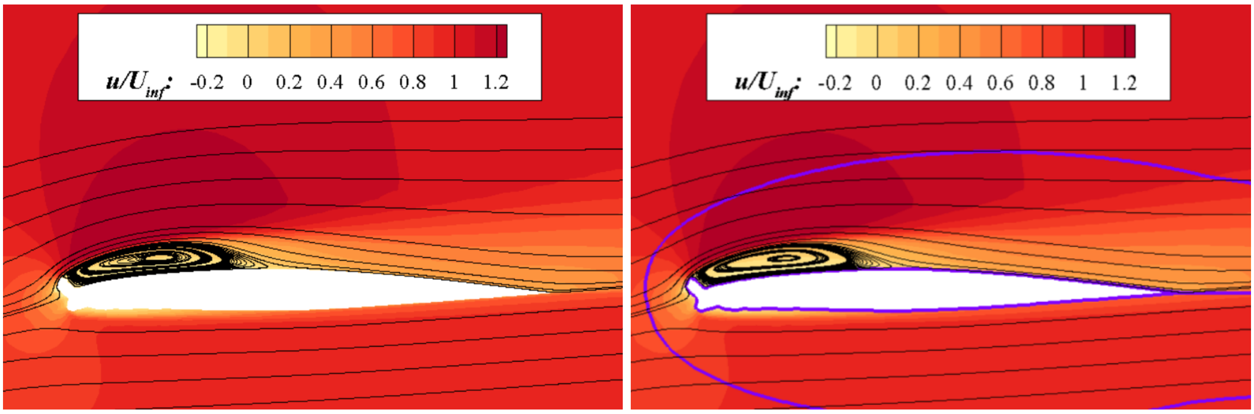
(a) SST result of periodic hill (b) DNS result of periodic hill



(c) SST result of GLC305/944 (d) SPF k-v^2-ω result of GLC305/944

Figure 3 – Comparison of baseline results and true values of two cases.

For the external flow of airfoil, two frozen calculations are carried out as described in Section 2.3. First, thee eddy viscosity of the whole computational domain using SPF k-v^2-ω is extracted and substituted into the convergent calculation results of SST model to restart the calculation until convergence is resumed; The second frozen calculation is to freeze the eddy viscosity in the purple region shown in Figure 2 and mix in the first 55000 iteration (take λ_{hybrid} = 0.5), then freeze all the data as the true eddy viscosity and recalculate until the flow convergence. The mean velocity contours obtained by the two iterations (at angle of attack 6°) are shown in Figure 4. From the comparison of Figure 3 and Figure 4, it can be seen that the mean flow fields of SPF k-v^2-ω, full frozen, and partial frozen are close enough, which proves that the validity of the current method and the theoretical limit that the machine learning model can reach are enough to correct the mean flow.



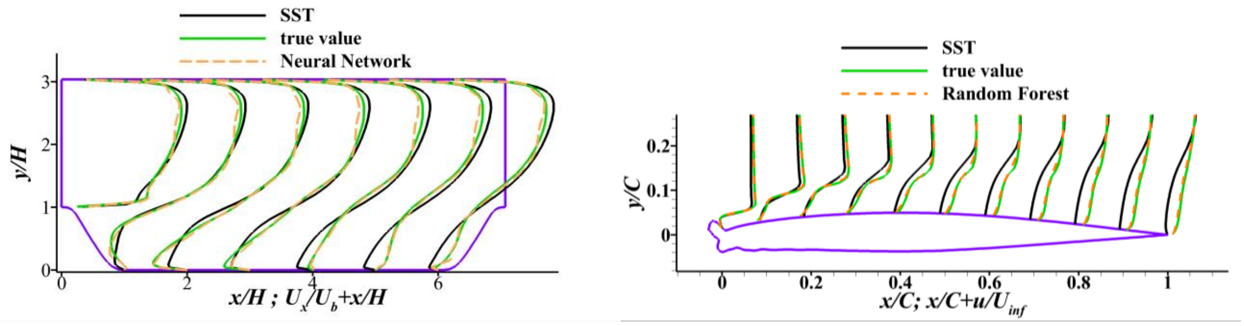
(a) Full field frozen (b) Partial field frozen

Figure 4 – Comparison of two frozen methods of GLC305/944 airfoil.

4. Model Training and Prediction Performance

For the periodic hill, the training set and test set are constructed by changing the slope of the hill. The slope 0.8, 1.2 are selected as training set and 0.5, 1.0, 1.5 are selected as test set. For the GLC305/944 airfoil, the angles of attack are changed. 4° and 6° are selected as training set and 5° and 7° are selected as test set. The machine learning algorithm used in this paper is neural network for the periodic

hill case and random forest for the airfoil case.
Taking the test set (slope 0.5 of periodic hill and angle of attack 7° of airfoil) to evaluate the prediction performance, the velocity profiles of original SST model, true value, and the machine learning results are shown in Figure 5. It can be seen that the mean flow for both cases are significantly improved compared with original results.



(a) Periodic hill (b) GLC305/944 airfoil

Figure 5 – Velocity profile comparison of two cases.

5. Conclusion

This paper discusses the establishment of turbulence modeling framework assisted by machine learning, including the construction of input set, feature selection and machine learning method selection. The machine learning methods show significant improvements both in the internal flow and the external flow cases. Especially when the machine learning model is applied to the outflow flow, the method of local regional modeling and the Reynolds stress hybrid calculation method are helpful to improve the convergence of the calculation. The results show that the local regional modeling method can reduce the time cost of the machine learning framework applied to the complex outflow flow problem, improve the training efficiency and accuracy, and achieve better correction effect for the mean flow.

6. Copyright Statement

References

[1] CRAFT T J, LAUNDER B E, SUGA K. Development and application of a cubic eddy-viscosity model of turbulence. Int.j.heat Fluid Flow, 1996, 17(2):108-115.

[2] CELIC A, HIRSCHEL E H. Comparison of Eddy-Viscosity Turbulence Models in Flows with Adverse Pressure Gradient. AIAA Journal, 2006, 44(10):2156-2169.

[3] SPALART P R . Philosophies and fallacies in turbulence modeling. Progress in Aerospace ences, 2015, 74(apr.):1-15.

[4] SHE Z, TANG F, XIAO M. Structural ensemble dynamics theory for engineering turbulence models. Acta Aerodynamica Sinica, 2019, 37(1): 1-18. (in Chinese)

[5] SPEZIALE, CHARLES G . On nonlinear K-l and K-ε models of turbulence. Journal of Fluid Mechanics, 1987, 178(-1):459-475.

[6] BREUER M, PELLER N, RAPP C, et al. Flow over periodic hills – Numerical and experimental study in a wide range of Reynolds numbers. Computers & Fluids, 2009, 38(2):433-457.

[7] SPALART P R. Strategies for turbulence modelling and simulations . International Journal of Heat & Fluid Flow, 2000, 21(3):252-263.

[8] DURAISAMY K , IACCARINO G , XIAO H . Turbulence Modeling in the Age of Data. Annual Review of Fluid Mechanics, 2019, 51:357-377.

[9] ZHANG W, ZHU L, LIU Y, KOU J. Progresses in the application of machine learning in turbulence modeling. Acta Aerodynamica Sinica, 2019, 3:444-454.(in Chinese)

[10]EMORY M , LARSSON J , IACCARINO G . Modeling of structural uncertainties in Reynolds-averaged Navier-Stokes closures. Physics of Fluids, 2013, 25(11):1155.

[11]XIAO H , WANG J X , GHANEM R G . A Random Matrix Approach for Quantifying Model-Form Uncertainties in Turbulence Modeling. Computer Methods in Applied Mechanics & Engineering, 2017, 313(JAN.1):941-965.

[12]LING J , TEMPLETON J . Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. Physics of Fluids, 2015, 27(8):042032-94.

[13]EDELING W N , CINNELLA P , DWIGHT R P . Predictive RANS simulations via Bayesian Model-Scenario Averaging. Journal of Computational Physics, 2014, 275:65-91.

[14]DURAISAMY K , ZHANG Z J , SINGH A P . New Approaches in Turbulence and Transition Modeling Using Data-driven Techniques// AIAA Aerospace Sciences Meeting. 2015.

[15]SINGH A P, MEDIDA S, DURAISAMY K. Machine Learning Augmented Predictive Modeling of Turbulent Separated Flows over Airfoils. AIAA Journal, 2017, 55(7): 2215-2227.

[16]XIAO H, WU J-L, et al. Quantifying and reducing model-form uncertainties in Reynolds-averaged Navier–Stokes simulations: A data-driven, physics-informed Bayesian approach. Journal of Computational Physics, 2016, 324: 115-136.

[17]TRACEY B D , DURAISAMY K , ALONSO J J . A Machine Learning Strategy to Assist Turbulence Model Development// 53rd AIAA Aerospace Sciences Meeting. 2015.

[18]ZHU L, ZHANG W, KOU J, LIU Y. Machine learning methods for turbulence modeling in subsonic flows around airfoils. Physics of Fluids, 2019, 31:015105.

[19]DURAISAMY K, SINGH A P. Using field inversion to quantify functional errors in turbulence closures. Physics of Fluids, 2016, 28(4): 045110.

[20]DURAISAMY K, PARISH E J. A paradigm for data-driven predictive modeling using field inversion and machine learning. Journal of Computational Physics, 2016, 305:758-774.

[21]LING J , KURZAWSKI A , TEMPLETON J . Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. Journal of Fluid Mechanics, 2016, 807:155-166.

[22]SCHMELZER M , DWIGHT R P , CINNELLA P . Discovery of Algebraic Reynolds-Stress Models Using

Sparse Symbolic Regression. Flow Turbulence and Combustion, 2019, 104(2-3): 579-603.

[23]WANG J X , WU J L , XIAO H . Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. Phys.rev.fluids, 2017, 2(3):1-22.

[24]WU J L , XIAO H , PATERSON E . Physics-Informed Machine Learning Approach for Augmenting Turbulence Models: A Comprehensive Framework. Physical Review Fluids, 2018, 3(7): 074602.

[25]YIN Y H, YANG P, ZHANG Y F, CHEN H X, FU S. Feature selection and processing of turbulence modeling based on an artificial neural network. Physics of Fluids, 2020, 32(10): 105117.

[26]LI H, ZHANG Y, CHEN H. Aerodynamic Prediction of Iced Airfoils Based on Modified Three-Equation Turbulence Model, AIAA Journal, 2020, 58(9): 3863-3876.

[27]TRACEY B, DURAISAMY K, ALONSO J J. Application of Supervised Learning to Quantify Uncertainties in Turbulence and Combustion Modeling, AIAA paper, 2013, 0259.

[28]DURAISAMY K , ZHANG Z J , SINGH A P . New Approaches in Turbulence and Transition Modeling Using Data-driven Techniques// AIAA Aerospace Sciences Meeting. 2015.

[29]LING J , JONES R , TEMPLETON J . Machine learning strategies for systems with invariance properties. Journal of Computational Physics, 2016, 318:22-35.

[30]POPE S B. A more general effective-viscosity hypothesis. Journal of Fluid Mechanics, 1975, 72(2):331-340.

[31]SPALART P R , DECK S , SHUR M L , et al. A New Version of Detached-eddy Simulation, Resistant to Ambiguous Grid Densities. Theoretical & Computational Fluid Dynamics, 2006, 20(3):181.

[32]RISTORCELLI J R. Diagnostic Statistics for the Assessment and Characterization of Complex Turbulent Flows, NASA-CR, 1995, 198221.

[33]XIAO H, WU J.-L., Laizet S, and DUAN L. Flows over periodic hills of parameterized geometries: A dataset for data-driven turbulence modeling from direct simulations Comput. Fluids, 2020, 200:104431 .

[34]ADDY H, BROEREN A, ZOECKLER J, LEE S. A Wind Tunnel Study of Icing Effects on a Business Jet Airfoil// AIAA Aerospace Sciences Meeting. 2003.

[35]NASA. CFL3D Version 6 [EB/OL]. https://cfl3d.larc.nasa.gov