Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Generalization Enhancement of Artificial Neural Network For Turbulence Closure By Feature Selection

Linyang Zhu
Northwestern Polytechnical University https://orcid.org/0000-0003-2334-8715
Weiwei Zhang
Northwestern Polytechnical University
Guohua Tu (ghtu@skla.cardc.cn)
China Aerodynamics Research and Development center

Research

Keywords: generalization, feature selection, artificial neural networks, turbulence model

Posted Date: July 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-594074/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Advances in Aerodynamics on January 13th, 2022. See the published version at https://doi.org/10.1186/s42774-021-00088-5.

Generalization enhancement of artificial neural network for turbulence closure by feature selection

Linyang Zhu[1,2], Weiwei Zhang[1,✉] Guohua Tu[2,✉]

1 Northwestern Polytechnical University, Xi 'an, 710072

2 China Aerodynamics Research and Development center, Mianyang, 621000

Abstract: Feature selection targets for selecting relevant and useful features, and is a vital challenge in turbulence modeling by machine learning methods. In this paper, a new posterior feature selection method based on validation dataset is proposed, which is an efficient and universal method for complex systems including turbulence. Different from the priori feature importance ranking of the filter method and the exhaustive search for feature subset of the wrapper method, the proposed method ranks the features according to the model performance on the validation dataset, and generates the feature subsets in the order of feature importance. Using the features from the proposed method, a black-box model is built by artificial neural network (ANN) to reproduce the behavior of Spalart-Allmaras (S-A) turbulence model for high Reynolds number (Re) airfoil flows in aeronautical engineering. The results show that compared with the model without feature selection, the generalization ability of the model after feature selection is significantly improved. To some extent, it is also demonstrated that although the feature importance can be reflected by the model parameters during training process, artificial feature selection is still very necessary.

Keywords: generalization, feature selection, artificial neural networks, turbulence model

✉ Corresponding author.
E-mail address: aeroelastic@nwpu.edu.cn
✉ Corresponding author.
E-mail address: ghtu@skla.cardc.cn

Notation

Angle of attack	α	Speed of sound	a_∞
Chord length of airfoil	c	Strain rate	$S=(u_y+v_x)/2$
Mach number	$Ma=u_\infty/a_\infty$	Rotation rate	$S=(u_y-v_x)/2$
Reynolds number	$\mathrm{Re}=u_\infty c/\upsilon$	Drag coefficient	C_d
Distance to the wall	y	Vorticity	w
Freestream speed	u_∞	Entropy (redefined)	$S'=(p/\rho^\gamma-1)/Ma,(\gamma=1.4)$
Shear stress at the wall	τ_w	Kinetic eddy viscosity	υ_t
Wall friction length unit	$\delta_\upsilon=\upsilon/u_\tau$	Kinetic viscosity	υ
Boundary layer thickness	δ	Skin friction coefficient	$C_f=2\tau_w/\rho_\infty u_\infty{}^2$
Friction velocity	$u_\tau=\sqrt{\tau_w/\rho}$	Pressure coefficient	$C_p=2(p-p_w)/\rho_\infty u_\infty{}^2$

1 Introduction

In recent years, there has been a surge in research on the combination of machine learning and turbulence. These research can be divided into two categories according to the model form: classifiers for identifying uncertainty regions and regression models for calculating turbulence-related variables. For the classifier, the researchers can use machine learning to recognize the uncertain regions calculated by the Reynolds-Averaged Navier-Stokes (RANS) turbulence model based on experimental or high-fidelity data[1,2]. The uncertainty of turbulence calculation comes from the ensemble- averaging operation of Navier-Stokes (N-S) equation, the functional form of a model, the representation of Reynolds stress and the empirical parameters of the model[3]. Then, a classifier can be constructed to predict the uncertain regions in different flow cases and a better method with higher accuracy can be used to improve the accuracy for these uncertain regions. For the regression, machine learning is mainly used to improve or substitute the conventional RANS model and the subgrid model in

large eddy simulation (LES). The linear eddy viscosity models are based on the Boussinesp assumption, which have high efficiency and good robustness, but are eclipsed by anisotropic turbulence, such as large separation and secondary flows. Thus, the researchers used machine learning to decrease the discrepancies between RANS results and high-fidelity or experimental data. There are mainly two corresponding solutions. One is to change the governing equation form of the RANS model, such as introducing the correction coefficient or adding the source term to the transport equation[4-7]. The other is to construct the deviation function by machine learning and then superimpose the results of RANS model and the output of proposed model as the Reynolds stress[8,9]. Different from the above studies, some studies directly construct the surrogate models of some turbulence variables[10-12]. Overall, the achievement is encouraging. In the future, machine learning may play a key role in turbulent modeling of complex flows[13]. With the research deeper and deeper, there are still many problems to be solved, such as feature selection and insufficient generalization ability, etc[14,15].

The *feature* means each component of the input vector and describes the sample property, which is similar to the independent variable of the function. Thus, the feature plays an important role in the model performance. It is difficult to characterize the data space thoroughly and comprehensively by inadequate features, which leads to low accuracy. On the contrary, redundant features may result in over-fitting of the model and reduce its generalization ability. Generalization generally refers to the performance of a model on unseen datasets. Feature selection is to select relevant and useful features to reduce the computation cost and improve the model performance[16]. For complex problems with unclear mechanism, it is challenging for researchers to extract features straightforwardly. A common way is selecting some related features empirically according to physical knowledge and the specific problem. The result of this way might be accidental. Therefore, efficient and reliable feature selection algorithm is needed by researchers.

In data-driven turbulence modeling, feature construction and selection are large obstacles for researchers. Many researchers select features according to existing governing equations and theorems or their own physical views. A common

recommandation is that the features should be invariant. Invariance property means that values do not change with the translation, relection and rotation of the coordinate system. Ling et al.[17] compared two methods of integrating invariant property into machine learning models and found that embedding the invariance into the features is more efficient and the corresponding model has better performance. Based on the conclusion, they took five tensor invariants as features and constructed the tensor basis neural network for scalar coefficients of the nonlinear eddy viscosity model. Besides the integrity basis derived by Pope[18], Wu et al.[19] introduced the gradient of turbulent kinetic energy and pressure to consider the effect of strong pressure changes and nonequilibrium. Yin et al.[20] further summarized feature selection criteria based on tensor analysis and flow structure identification. Although some researchers emphasize the invariance property, it is not clear whether the features without this property will definitely reduce the model performance. Wang et al.[21] studied the closure of the sub-grid stress in LES and the adopted features are the filtered velocity and first and second order derivatives. Since the geometry and coordinate system of test cases are the same as those of training cases, Cruz et al.[22] took each component of the tensors and vectors as the input features. Singh et al.[6] adopted the features mainly based on the governing equation of Spalart-Allmaras (S-A) model. The features used in most of the current work are based on physics and belong to the result of feature construction rather than feature selection to a large extent.

There are some common feature selection methods, like filter method, wrapper method and embedded method or some derivation methods based on these three methods[16]. Filter methods rank the features according to the correlation between features and output, which is efficient and priori. One of the common ways to measure the correlation is the Pearson correlation coefficient. Other correlation measurement can be found in [23,24]. The wrapper methods evaluate the feature subset according to the prediction performance. The exhaustive search of this method becomes an NP-hard problem since there are 2^N feature sunsets, where N is the candidate feature number. Embedding methods insert the feature selection into the learning process.The difference of embeded method from the previous two methods is that the learning process and the feature selection process cannot be separated, thus the feature subset

corresponding to a specific model architecture is not applicable to other model architectures. In recent years, some hybrid methods of these three methods are proposed[25,26]. More information about the feature selection method can detailed in [27-30]. In data-driven turbulence modeling, feature selection methods stated above are not widely used yet.

The principle target of this work is to enhance the model generalization ability through the proposed feature selection (FS) algorithm. For high-Re-number airfoil flows in aerospace, the constructed model in this paper is essentially a black-box model without solving the transport equation, aiming to reproduce the behavior of RANS model. Driven by the data from only one flow case, the constructed model is expected to be effective for various cases with different freestream conditions and airfoils. The rest of this paper is summarized as follows. The second section introduces the methods used in this work, including workflow, modeling strategy, artificial neural network and the feature selection algorithm. The training process is illustrated in the third section, including the datasets and the training method. The fourth section is the results and discussions, which are addressed from the aspects of accuracy and convergence. Conclusions and outlook lie in the fifth section.

2 Method

2.1 Workflow

The workflow includes two parts: neural network training and coupling of the N-S equation solver with the constructed model. The first part aims to obtain a neural network model of which the performance should be satisfying and convergent. The word “convergent” means that the loss function is hard to be lower. This part mainly consists of three aspects: data acquisition, data preprocessing and model training. Model performance is improved mainly by adjusting model parameters and hyperparameters during the training process. If the performance is not adequate, the input features and training samples can be further optimized in data preprocessing. The target of the second part is to substitute the conventional RANS model with the constructed surrogate model and obtain the results that agree well with those calculated by S-A model. The surrogate model acts as a turbulence solver, as shown in the dot box

below. The entire process is shown in Figure 1, and some details will be addressed later.

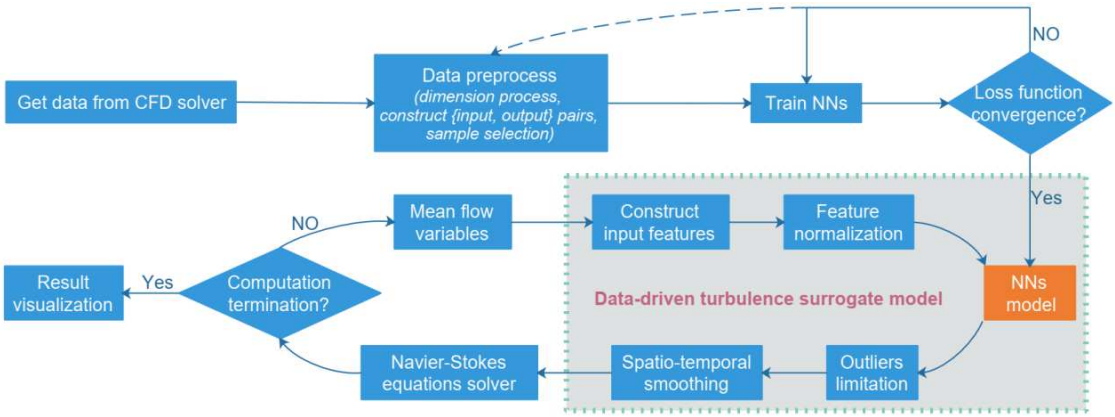


Fig. 1 Flow chart for building the turbulence surrogate model.

2.2 Modeling strategy

According to the dimensional analysis, the kinetic eddy viscosity can be expressed as the product of mixing velocity u_{mix} and mixing length l_{mix}.

$$\upsilon_t = u_{mix} l_{mix} \tag{1}$$

The "mixing length" derives from the molecular mean free path in gas dynamics, which was proposed by Prandtl[31]. The algebraic model with zero equation can express turbulent eddy viscosity explicitly in the form of eq(1). Generally, two length scales are used to characterize the turbulence from wall to the outer edge of the boundary layer, namely the wall friction scale δ_υ and the boundary layer thickness δ. The δ_υ and δ are corresponding to the inner and outer layers, respectively, where the boundary is around the outer edge of the logarithmic layer. The mixing length of the inner layer is related to the distance to the wall. There are many researches on the scale analysis of the inner layer, and the universality is good in many cases. However, the outer flow variables are only slightly influenced by the wall, and are more disorderly and irregular[32]. Therefore, the scale analysis of the outer layer is more challenging and of less universality. More information about the mixing length is detailed in [33]. The mixing length adopted in this paper is based on the Prandtl-van Driest formulation[34] and limited by the boundary layer thickness of flat plane,

$$l_{mix} = \min(ky(1 - e^{-\frac{d}{d_{y^+\approx 1} A^+}}), 0.4\delta_{flat}) \tag{2}$$

where $d_{y^+\approx 1}$ denotes the *estimated wall friction scale*, calculated by the following

address: https://www.cfd-online.com/Tools/yplus.php. Because it is found that the real wall friction scale leads to worse convergence during the iteration process. Referring to Clauser's constant eddy viscosity hypothesis[35], the mixing length of the outer layer is the same order with the boundary layer thickness. For convenience, the boundary layer thickness here is approximate to that of the flat plane, since the curvature of the airfoil is small. It should be noted that eq(2) is effective qualitatively rather than quantitatively, which is a compromise to the lack of robustness of the constructed neural network model. Once the mixing length is determined, the mixing velocity can be deduced according to eq(1). The constructed model in this paper is a nonlinear mapping between mean flow variables and the mixing velocity. Since the modeling strategy includes the distance to the wall, which can be less effective in separated flows, the flow cases in this work are all attached flows.

2.3 Artificial neural network

The term "neural network" comes from neurology and is often referred to as "artificial neural network" in machine learning. Neuron is the basic unit of neural network. The number of neurons in each layer is called *width* and the number of layers is called *depth*, which are two hyperparameters of a neural network. For a fully-connected neural network, only the neurons in adjacent layers are connected to each other, and the output of the neurons in the previous layer is the input of the neurons in the current layer. The neural network used in this work is shown in Figure 2.

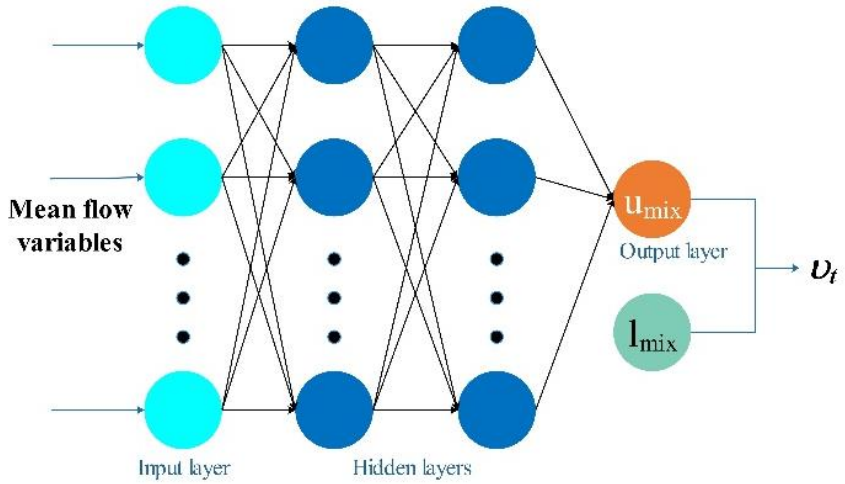


Fig. 2 Schematic of the fully-connected neural network.

The first layer is the input layer while the last layer is the output layer, and the rest

is the hidden layers. The parameters of the neuron are called *weight* and *bias*. Each neuron contains two operation. The first one is the linear addition between the product of input vector $\mathbf{X}$ and weight vector $\mathbf{W}$ and bias,

$$z = \mathbf{WX} + b \tag{3}$$

and the second one is the nonlinear activation operation.

$$y = f(z) \tag{4}$$

The common activation functions $f(\cdot)$ are Tanh, Sigmoid, Relu, LeakRelu and so on. In this work, Tanh is adopted. The structure determined by a trial-and-error approach is (20, 20, 20, 20, 20, 1). It was found that the adopted structure is enough for reaching the optimal solution. More neurons do not significantly improve model performances but increase computation time.

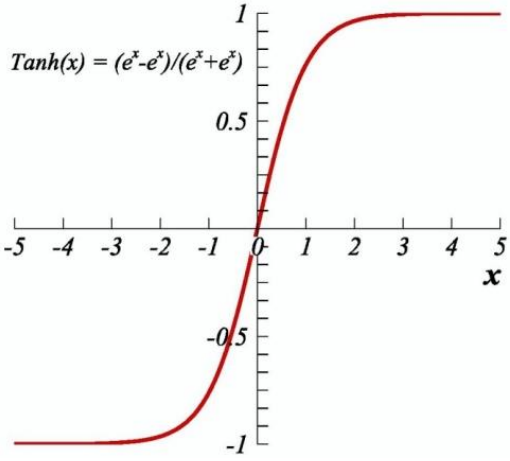


Fig. 3 The Tanh activation function.

The error back propagation (BP) algorithm based on gradient descent is used to train the neural network. The gradient is calculated by the chain rule to update the parameter values of the neurons. Because the gradient descent algorithm is prone to stop at the local minimum, the global optimum is generally approximated by several training with different initial values. More information about neural network is detailed in [36-38].

According to the universal approximation theory, neural networks with enough neurons can approximate any measurable functions[39]. The complexity of the neural network can be adjusted by changing the depth and width. The strength of mapping complex nonlinear functions makes neural network popular with researchers. It has already been used in turbulence computation[10,40].

2.4 Feature selection

We do not intend to delve into feature construction that depends on specific problems or researchers' physical view, but rather try to select some features that are helpful to model performance from existing candidate features. Although we assume that the neural network itself can play a role in feature selection more or less in the training process, the important role played by artificial feature selection is irreplaceable according to our own experience.

The proposed feature selection method includes two parts. The first one is measuring the numerical space of the features and kicking off those with obvious extrapolation according to the mean extrapolation distance (MED). The second one is ranking the feature and determine the feature subset according to the model performance. The whole process is shown in Figure 4 and to be more specifically, it includes the following steps.

Step 1: Get the dataset $\boldsymbol{D}=\{(\mathbf{Q}_{\boldsymbol{D}},\mathbf{y}_{\boldsymbol{D}}),\mathbf{Q}_{\boldsymbol{D}}\in ¡^{l_{\boldsymbol{D}}\times m},\mathbf{y}_{\mathbf{D}}\in ¡^{l_{\boldsymbol{D}}\times 1}\}$ and $\boldsymbol{T}=\{(\mathbf{Q}_{\boldsymbol{T}},\mathbf{y}_{\boldsymbol{T}}),\mathbf{Q}_{\boldsymbol{T}}\in ¡^{l_{\boldsymbol{T}}\times m},\mathbf{y}_{\boldsymbol{T}}\in ¡^{l_{\boldsymbol{T}}\times 1}\}$, where $\mathbf{Q}$ denotes the feature vectors, $\mathbf{y}$ the labels, m the number of features, $l_{\boldsymbol{D}}$ and $l_{\boldsymbol{T}}$ the sample number of $\boldsymbol{D}$ and $\boldsymbol{T}$, respectively. Normalize each feature and the output of $\boldsymbol{D}$ to $[-1,1]$, and normalize the dataset $\boldsymbol{T}$ according to the corresponding minimum and maximum values from $\boldsymbol{D}$. Mark the space domain and the domain boundary formed by each feature and the output of $\boldsymbol{D}$ as Ω_i and $\partial\Omega_i$, respectively, where $i=1,2...,m$. Put the sample of $\boldsymbol{T}$ in corresponding domain according to the feature and regard those samples outside the domain boundary as extrapolation samples. Calculate the MED of each feature by eq(5) and remove the features with large MED values. Denote the number of the remaining features as n. Extract the remaining features and the output from $\boldsymbol{D}$ and $\boldsymbol{T}$, and reshape the datasets as $\overline{\boldsymbol{D}}=\{(\mathbf{Q}_{\overline{\boldsymbol{D}}},\mathbf{y}_{\boldsymbol{D}}),\mathbf{Q}_{\overline{\boldsymbol{D}}}\in ¡^{l_{\boldsymbol{D}}\times n},\mathbf{y}_{\mathbf{D}}\in ¡^{l_{\boldsymbol{D}}\times 1}\}$ and $\overline{\boldsymbol{T}}=\{(\mathbf{Q}_{\overline{\boldsymbol{T}}},\mathbf{y}_{\boldsymbol{T}}),\mathbf{Q}_{\boldsymbol{T}}\in ¡^{l_{\boldsymbol{T}}\times n},\mathbf{y}_{\boldsymbol{T}}\in ¡^{l_{\boldsymbol{T}}\times 1}\}$, respectively.

$$MED=\frac{\sum_{i=1}^{N_e} d_i}{N_e} \tag{5}$$

where N_e is the number of the extrapolation samples.

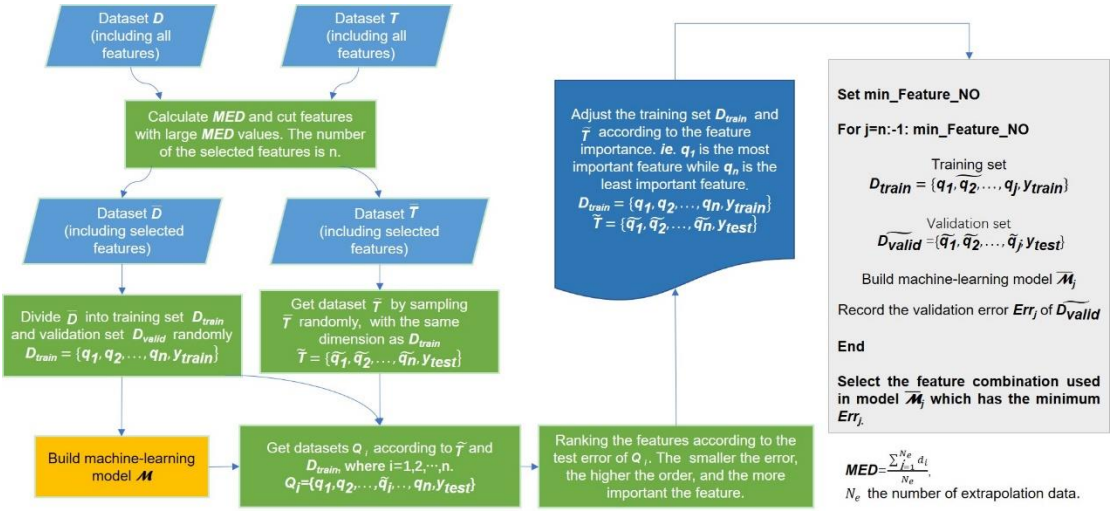


Fig. 4 The procedure of feature selection.

Step 2: Divide the dataset $\overline{\boldsymbol{D}}$ into training set $\boldsymbol{D}_{train}$ and validation set $\boldsymbol{D}_{valid}$ randomly, where $\boldsymbol{D}_{train}=\{(\mathbf{q}_1,\mathbf{q}_2,...,\mathbf{q}_n,\mathbf{y}_{train}),\mathbf{q}_i\in ¡^{l_{train}\times 1},\mathbf{y}_{train}\in ¡^{l_{train}\times 1},i=1,2,...,n\}$. Construct a neural network $\boldsymbol{\mathcal{M}}$ according to $\boldsymbol{D}_{train}$ and $\boldsymbol{D}_{valid}$. Create a new dataset $\overset{\frown}{\boldsymbol{T}}=\{(\overset{\frown}{\mathbf{q}}_1,\overset{+}{\mathbf{q}}_2,...,\overset{+}{\mathbf{q}}_n,\mathbf{y}_{test}),\overset{\frown}{\mathbf{q}}_i\in ¡^{l_{train}\times 1},\mathbf{y}_{test}\in ¡^{l_{train}\times 1},i=1,2,...,n\}$ by selecting l_{train} samples from $\overline{\boldsymbol{T}}$. Replace the feature of $\boldsymbol{D}_{train}$ by the same feature of $\overset{\frown}{\boldsymbol{T}}$ one by one and combine the output of $\overset{\frown}{\boldsymbol{T}}$, thus, we can get n datasets $\boldsymbol{Q}_i$, where $i=1,2...,n$.

$$\boldsymbol{Q}_i=\{(\mathbf{q}_1,\mathbf{q}_2,...,\overset{\frown}{\mathbf{q}}_i,...,\mathbf{q}_n,\mathbf{y}_{test}) \tag{5}$$

Test the model by datasets $\boldsymbol{Q}_i$ and get the test errors. Rank the features according to the test errors. The smaller the error, the more important the feature and the higher the rank.

Step 3: Adjust the features of $\boldsymbol{D}_{train}$ and $\overset{\frown}{\boldsymbol{T}}$ according to the feature importance and shape them as the following form:

Feature Importance : 1, 2, ..., n

$$\boldsymbol{D}_{train}=\{\mathbf{q}_1,\mathbf{q}_2,...,\mathbf{q}_n,\mathbf{y}_{train}\}$$
$$\overset{\frown}{\boldsymbol{T}}=\{\overset{\frown}{\mathbf{q}}_1,\overset{+}{\mathbf{q}}_2,...,\overset{+}{\mathbf{q}}_n,\mathbf{y}_{test}\}$$

Create the training set $\vec{\boldsymbol{D}}_{train}$ and validation set $\vec{\boldsymbol{D}}_{valid}$ by selecting the first j features from $\boldsymbol{D}_{train}$ and $\overset{\frown}{\boldsymbol{T}}$, respectively. Then, construct a neural network according to $\vec{\boldsymbol{D}}_{train}$ and $\vec{\boldsymbol{D}}_{valid}$, and record the validation error $\boldsymbol{Err}_j$ on $\vec{\boldsymbol{D}}_{valid}$. The feature number j increases from **min_Feature_NO** set by the user to n one by one. Finally, select the feature subset with the minimal validation error.

Table1 The flow cases of dataset ***D*** and ***T*** used for feature selection.

Dataset	Ma	Re ($\times 10^6$)	α (°)
D	0.27	3	2.45
T	0.17	7.24	1.72
	0.25	5.45	4.29
	0.28	3.31	2.63
	0.41	5.70	0.95
	0.47	2.22	3.65

Table 2 The total candidate features.

Feature	Description	Denotation
q_1- q_4	Velocity gradient	$u_x,\ u_y,\ v_x,\ v_y$
q_5	Projection of free stream to normal direction of streamline	$\mathrm{sgn}(y)[-v+u\tan(\alpha)]$
q_6- q_7	Pressure gradient	p_x, p_y
q_8	Redefined entropy	S'
q_9	Pressure gradient along streamline	$u_i \dfrac{\partial p}{\partial x_i}$
q_{10}	Eddy viscosity expression for the inner layer	$l_{mix}{}^2 w$
q_{11}	Velocity magnitude	$\sqrt{u^2+v^2}$
q_{12}	Velocity direction	$\arctan[\mathrm{sgn}(y)v/u]$
q_{13}- q_{14}	Velocity component	u, v
q_{15}	Density	ρ

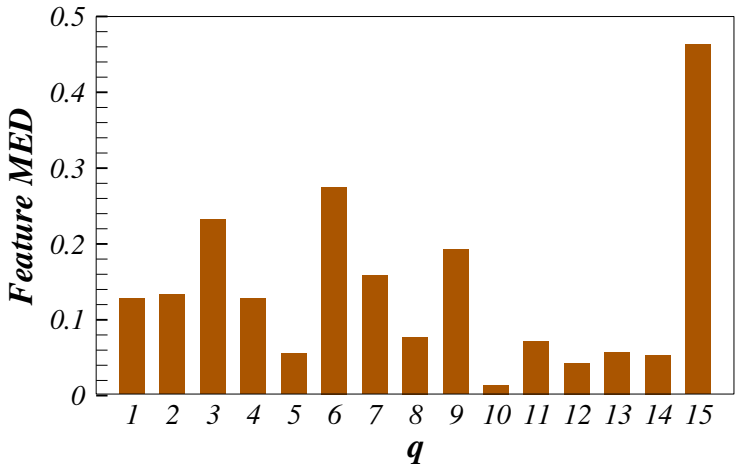
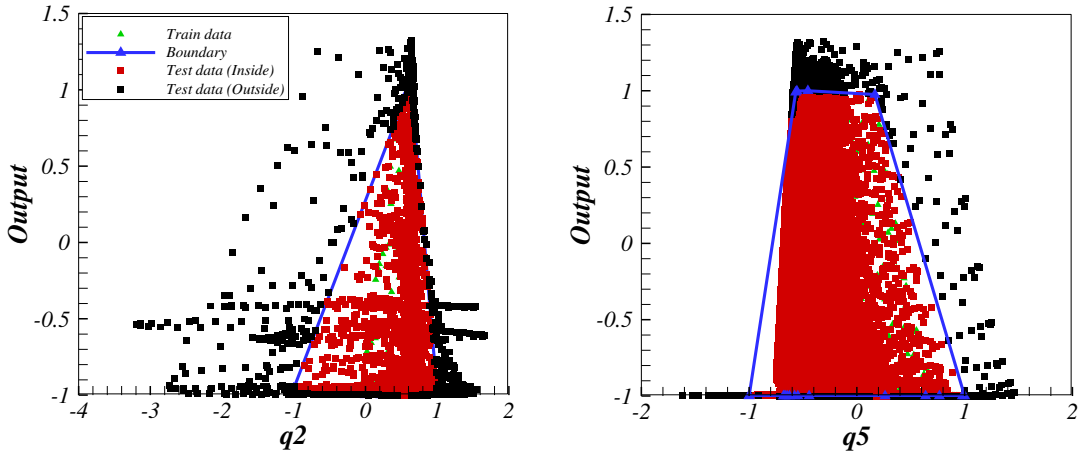


Fig.5 The MED of each feature.

In this paper, datasets $\boldsymbol{D}$ and $\boldsymbol{T}$ are composed of 2525 and 5437 samples, respectively. These samples are obtained from different flow cases (see Table 1) using the sample selection algorithm described below. The flow cases are sampled through the Latin hypercube sampling method (LHS), and the sampling space is $Ma=[0.1,0.5], \mathrm{Re}=[2\times10^6, 8\times10^6], \alpha=[0°,5°]$. The total candidate features before feature selection are shown in Table 2. According to the ***step 1***, the MSE of each feature can be calculated (see Figure 5) and the space domain formed by each feature and the output of dataset $\boldsymbol{D}$ and $\boldsymbol{T}$ can also be plotted. For example, the space domain of some features can be seen in Figure 6. After removing the features with large MED, the remaining features are $(q_5, q_8, q_{10}, q_{11}, q_{12}, q_{13}, q_{14})$. For ***step 2***, the mean square error (MSE) of 7 datasets $\boldsymbol{Q}_1 - \boldsymbol{Q}_7$ is shown in Figure 7. Thus, the feature importance is $q_8 > q_{13} > q_5 > q_{12} > q_{14} > q_{10} > q_{11}$ according to the MSE. In ***step 3***, the **min_Feature_NO** is set to be 3, which means that the feature q_8, q_{13}, q_5 are necessary.

Starting from these three features, an individual feature subset can be generated as the feature number increases by one in the order of the feature importance and a corresponding neural network model can be constructed. The best feature subset can be found by the model performance on validation set. In order to avoid the influence of occasionality, the model is trained three times for each feature subset, and the model performance is measured according to the mean value of the errors on validation set. As is shown in Figure 8 where the horizontal axis means the feature number of the feature subset, the model performance is the best when the first four features are adopted. Therefore, the features used to construct the model hereinafter are $(q_8, q_{13}, q_5, q_{12})$.



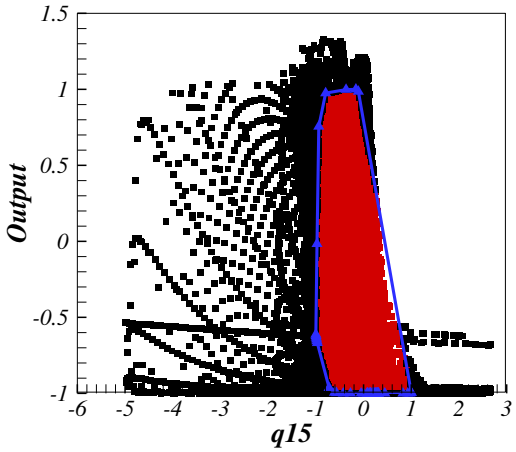


Fig.6 The numerical distribution of some features of dataset $\boldsymbol{D}$ and $\boldsymbol{T}$.

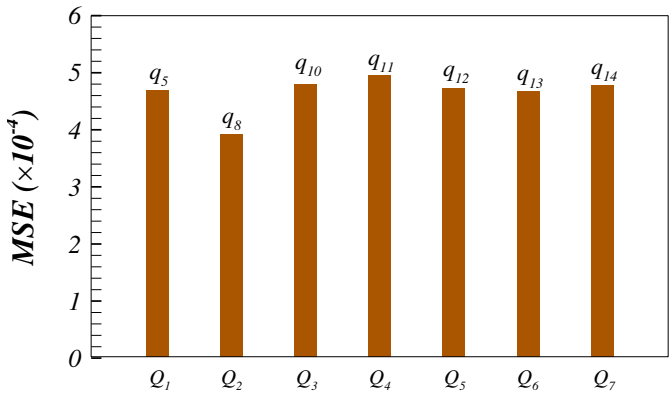


Fig.7 The model performance on dataset $\boldsymbol{Q}_1 - \boldsymbol{Q}_7$.

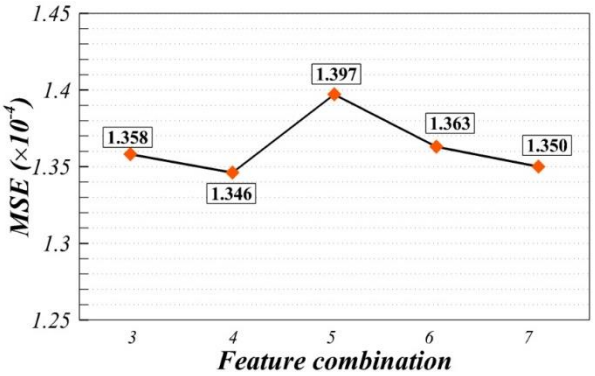


Fig.8 The model performance based on different feature subsets.

3 Model training

3.1 Dataset

The source data used in this paper is computed by the CFD solver coupled with the S-A model[41], where the transport equation of the S-A model is as follows.

$$\mu_T = \rho \hat{\upsilon} f_{v1} \tag{6}$$

$$\frac{D\hat{\upsilon}}{Dt}=C_{b1}(1-f_{t2})\hat{S}\hat{\upsilon}+\frac{1}{\sigma}\{\nabla\cdot[(\upsilon+\hat{\upsilon})\nabla\hat{\upsilon}]+C_{b2}(\nabla\hat{\upsilon})^2\}-(C_{w1}f_w-\frac{C_{b1}}{\kappa^2}f_{t2})(\frac{\hat{\upsilon}}{d})^2 \quad (7)$$

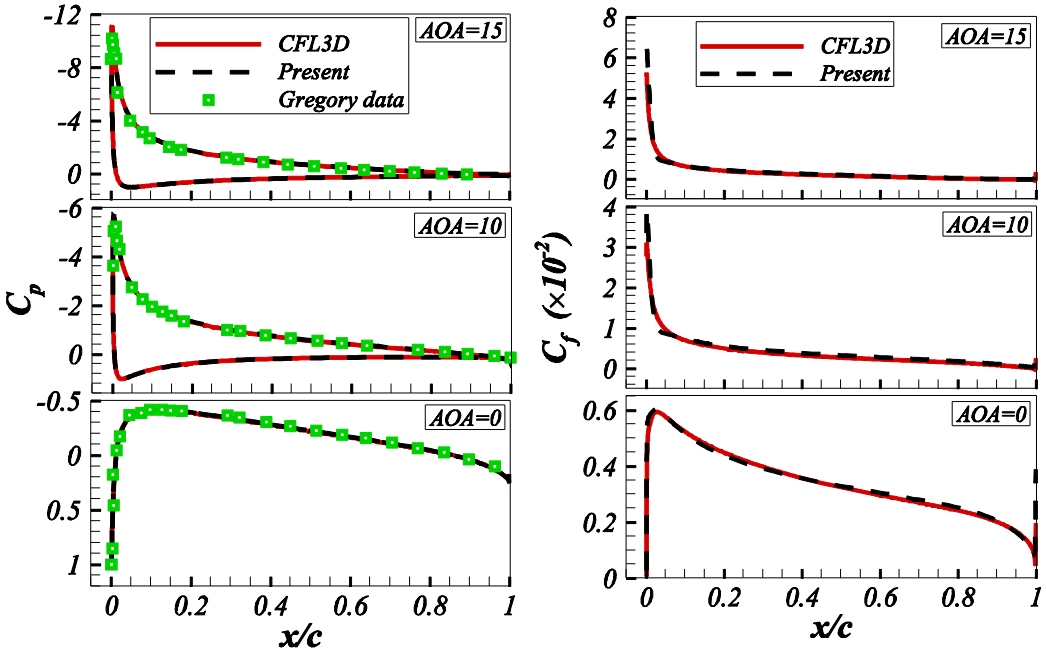


Fig.9 The comparison of surface pressure C_p and skin friction coefficient C_f of NACA0012 airfoil.

Table 3 The freestream condition of the benchmark cases.

Airfoil	Freestream condition		
	α (°)	Re ($\times10^6$)	Ma
NACA0012	0	6	0.15
	10	6	0.15
	15	6	0.15
RAE2822	2.8	6.5	0.73
	2.8	6.2	0.75

The governing equations are solved by the pseudo-time-marching method. The cell-centered finite volume method is used and the spatial discretization is the AUSM+UP scheme with second-order accuracy. The governing equations are nondimensionalized by the mean aerodynamic chord c, speed of sound a_∞ and freestream density ρ_∞. The airfoil surface is set to be *wall* with no-slip condition and the far field is *pressure-far-field* without reflection. When the variation of drag coefficient between 1000 iterations is less than 1×10^{-5}, the flow field is set to be

converged. Hybrid meshes are used. The grid near the wall is quadrilateral with the growth rate of 1.2 and the outer grid is triangular. The height of the first layer at the wall satisfies $y^+ < 1$. The benchmarks of NACA0012 and RAE2822 airfoils are adopted to validate the code and the corresponding freestream condition is shown in Table 3. The result of the CFL3D and experiment can be referred to [42-44]. Both of the agreement of C_p and C_f are good, see Figure 9 and Figure 10.

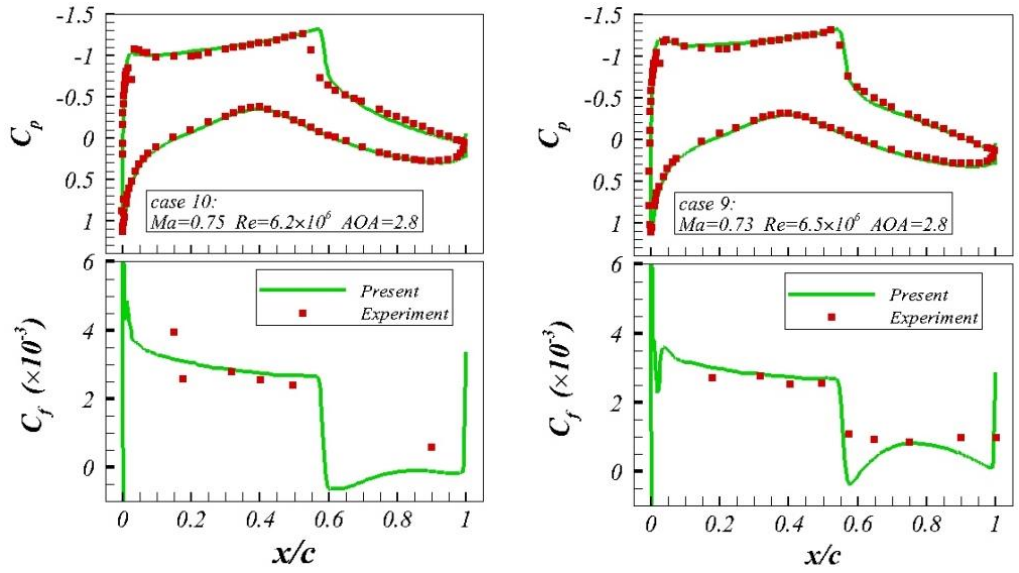
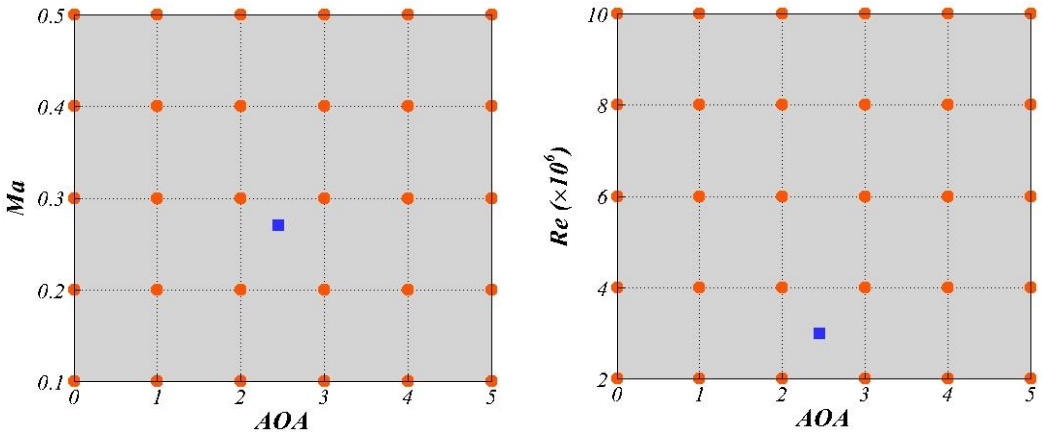


Fig. 10 The comparison of surface pressure C_p and skin friction coefficient C_f of RAE2822 airfoil.



(a) $\mathrm{Re} = 3\times10^6$ (b) $Ma = 0.27$

Fig.11 Training (blue square) and test (orange circle) cases.

In this paper, there is only one training case, and its freestream condition is $Ma = 0.27, AOA = 2.45, \mathrm{Re} = 3\times10^6$. 2525 data pairs were generated by the sample

selection algorithm, and 1/5 of them selected randomly was the validation dataset, while the rest was used as the training dataset. A total of 68 flow cases were used as the test cases, sixty of which are shown in Figure 11 and the rest cases will be illustrated later. The reason for this design of test cases is that we attempt to find which freestream condition change (Ma, Re, AOA) is more challenging for the model performance, which might be helpful for the future work.

3.2 Data process

3.2.1 Sample selection

Th goal of sample selection, in our opinion, is to select as few samples as possible to represent as much of the original sample space as possible. In other words, the diversity of physical behaviors contained in the original sample space should be retained as much as possible. The sample selection method used in this work is a combination of the law of the wall and recursive algorithm. The whole process can be achieved by two steps:

Step 1: Divide the flow field into several equal subzones according to $\ln y^+$, and the size of each subzone is a set value Δx.

Step 2: Set the standard correlation value cc_std. Take the first data in the subzone as the first sample. For each subzone, starting loop from the second data, calculate the correlation coefficient cc between each data and all samples of the current subzone, and denote the maximal cc as $cc_\max imum$. If $cc_\max imum < cc_std$, then the corresponding data is a new sample; otherwise, the data is redundant and should be removed. Loop the current step for all subzones.

The formula of correlation coefficient is,

$$cc = \frac{\mathbf{g}_1 \cdot \mathbf{g}_2}{\|\mathbf{g}_1\|\|\mathbf{g}_2\|} \tag{8}$$

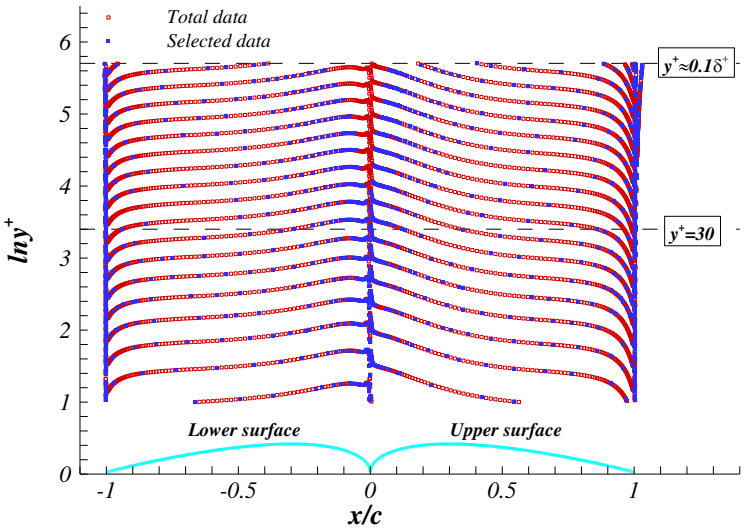
where $\mathbf{g}$ denotes the vector composed by the input features and output. The specific algorithm is shown in Table 4.

It should be noted that before the sample selection and during the coupling process later, the data of cells whose Reynolds stress is near zero in the far field is eliminated

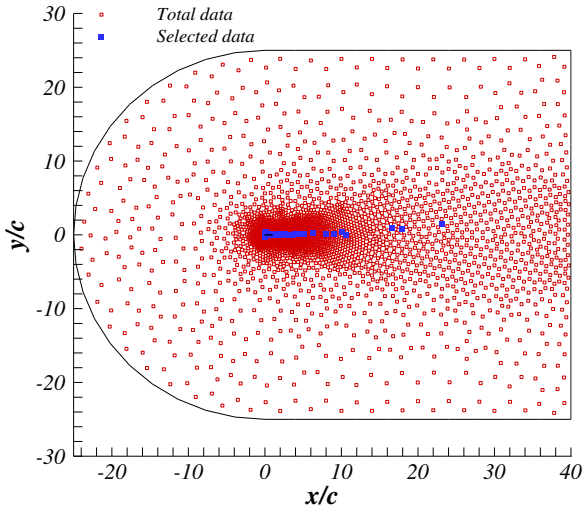
according to the redefined entropy $S' < 2\times10^{-5}$. The comparison of data distribution before and after sample selection is shown in Figure 12 when this method is applied to the training case. It can be found from the figure that, among the selected samples, many of them lie in the leading edge and trailing edge where the flow field changes rapidly. And the data at middle airfoil is obviously sparse, while there is only a few data in the far field and wake region.

Table 4 The sample selection algorithm.

Algorithm for Sample Selection
set the value of Δx *and* cc_std
SubzNO = ceil(max(lny⁺ - 1) / Δx) // calculate the number of subzone, SubzNO
For *i = 1 : SubzNO*
Sample(1) = Data(i,1) // take the first data as the first sample of the i_{th} *subzone*
SampleNO = 1
For *j = 2 : SubzData(i) // SubzData(i) denotes the number of data in* i_{th} *subzone*
For *k = 1 : Sample_NO*
compute the cc between Sample(k) and Data(i, j)
End
find the maximum of cc and the corresponding Data, denoted as Data(i, jmax)
If *cc_maximum < cc_std*
Sample_NO = Sample_NO + 1
Sample(Sample_NO) = Data(i, jmax)
End
End
Save the selected Sample of the i_{th} *subzone*
End



(a) Inner layer



(b) Outer layer

Fig. 12 Distribution of the total data and selected data for the training case $Ma = 0.27, AOA = 2.45, \mathrm{Re} = 3\times10^6$.

3.2.2 Postprocess

For the samples at the edge of model performance, the model outputs are probably outliers, resulting in negative eddy viscosity. In this case, eddy viscosity is forced to be non-negative. In addition, there may also be the phenomenon of "burr", that is the small jump of the outputs among adjacent cells. In order to obtain a smooth flow field, the spatial smoothing is introduced to smooth the eddy viscosity field calculated by the model at each iteration.

$$\mu_t = \frac{\mu_t + \beta \sum_i^N \mu_t^i}{1 + \beta N} \tag{9}$$

where i denotes the i_{th} neighboring cell, N the total number of neighboring cells, and β the weight of neighboring cells. In this work, we set $\beta \in [0.2,1]$. All the flow cases involved in this paper are steady. Therefore, the final turbulent field should be convergent. During the iteration process, we found that the turbulence field calculated by the model is convergent on the whole, but there are slight oscillation of the eddy viscosity of some cells from time to time. The slight oscillation in turn causes the residual oscillation of the N-S solver. In order to restrain such unreasonable oscillation, when the residual value of N-S solver is less than 1×10^{-5}, the mean eddy viscosity variation $\sum \Delta\mu_t / M$ between adjacent iteration steps is limited to monotonically non-increasing, where M is the number of the total cells. It is found that the influence of this limitation on the final results can be ignored.

3.3 Training process

If the loss function is simply defined as mean square error (MSE), then the relative error of the model will be large for samples with small output value. As far as this work is concerned, the above trouble happens in the near wall region, and thus the model is prone to perform poorly. Furthermore, the high shear strain rate here will lead to unreasonable results of Reynolds stress and skin friction distribution. Therefore, in the training process, the accuracy of Reynolds shear stress is considered as a constraint for the region $y^+ < 30$. In other words, the accuracy of the samples in near wall region is improved through increasing the weight in the loss function by the constraint term.

The constructed model needs to be coupled with the N-S solver like conventional turbulence models. Therefore, the robustness of the model should be enhanced during the training process to avoid the high sensitivity of output to the slight change of input. Two tricks are used. One is that we included the L1-norm and L2-norm regulation in the loss function. The other is that we introduced the stability training (ST) after the normal training process. The principle of stability training is to enhance the model robustness through adding some random noise to the original samples. The random

noise used in this work is uniformly distributed with the magnitude less than $\pm 5\%$ of the original sample.

$$\tilde{\boldsymbol{X}} = (1 \pm 5\%)\boldsymbol{X} \tag{10}$$

where $\tilde{\boldsymbol{X}}$ is called *random perturbation copy*. The model output of $\tilde{\boldsymbol{X}}$ is denoted as $\tilde{\boldsymbol{Y}}$. The stability loss term is defined as,

$$L_{stability} = \sqrt{\sum_{i}^{M} (\boldsymbol{Y}_i - \tilde{\boldsymbol{Y}}_i)^2} \tag{11}$$

The loss function of ST is:

$$Loss = MSE + \underbrace{\lambda_1 \|\mathbf{w}\|_1 + \lambda_2 \|\mathbf{w}\|_2}_{\text{Regulation loss}} + \underbrace{\lambda_3 (\|\tau_{SA}^{uv} - \tau_{ML}^{uv}\|_2)}_{\text{Constraint loss}} + \underbrace{\lambda_4 L_{stability}}_{\text{Stability loss}} \tag{12}$$

where

$$MSE = \frac{\sum_{i}^{N} (y_i - f_i)^2}{N}, \qquad y_i : \text{label}, \ f_i : \text{model output} \tag{13}$$

Table 5 The training parameters.

Parameters	Value
Batch Normalization[45]	Adopted
Activation function	tanh
Optimizer	Adamax
Learning rates	ReduceLROnPlateau
λ_1	2×10^{-5}
λ_2	1×10^{-4}
λ_3	1×10^{-6}
λ_4	1

More details on stability training can be referred to [46]. The model training is implemented on PyTorch[47], and the related parameters are shown in Table 5. There are five components in the loss function and the magnitude of each one can be different during the training process. If the magnitude of one component is much larger than that of the others, then the effect of the other components will be suppressed. Thus, the multiplier $\lambda_1 - \lambda_4$ is introduced to make the magnitude of each component on the same

order. It should be noted that the error function of the validation set does not include the stability loss term during normal training, which is defined as:

$$Error = MSE + \lambda_1(\left\|\tau_{SA} - \tau_{ML}\right\|_2) \tag{14}$$

During the stability training, the error function is redefined as:

$$Error^{ST} = MSE + \lambda_1(\left\|\tau_{SA} - \tau_{ML}\right\|_2) + \lambda_3 L_{stability} \tag{15}$$

The training process is over when the error function no longer drops within 500 steps.

4 Results and discussions

Since C_d and C_f are sensitive to the eddy viscosity field, this section mainly assesses the accuracy of the constructed model from these two aspects. The model performance is expected to be good in the test space $Ma = [0.1, 0.5], \mathrm{Re} = [2\times10^6, 10\times10^6], \alpha = [0°, 5°]$. First of all, we calculate the relative error of C_d for all the test cases in Figure 11. The error contour of the test space can be obtained by interpolation of the scattered cases, as shown in Figure 13. It can be found from Figure 13(a) that the relative error varies regularly with the angle of attack and Ma number for cases with fixed Re number. The relative error is smaller with the increase of angle of attack and larger with the increase of Ma number. However, the variation of the relative error is ambiguous with angle of attack and Re number for cases with fixed Ma number. On the whole, the relative error increases with the increase of angle of attack and Re number. The above analysis is qualitative and roughly. More detailed and quantitative information is shown in Figure 14. When $\alpha < 3°$, the influence of angle of attack on model performance is not obvious. When $\alpha > 3°$, the C_d calculated by the constructed model tends to be small with the increase of angle of attack. In addition, the C_d is smaller than that calculated by the S-A model as the Re number increases.

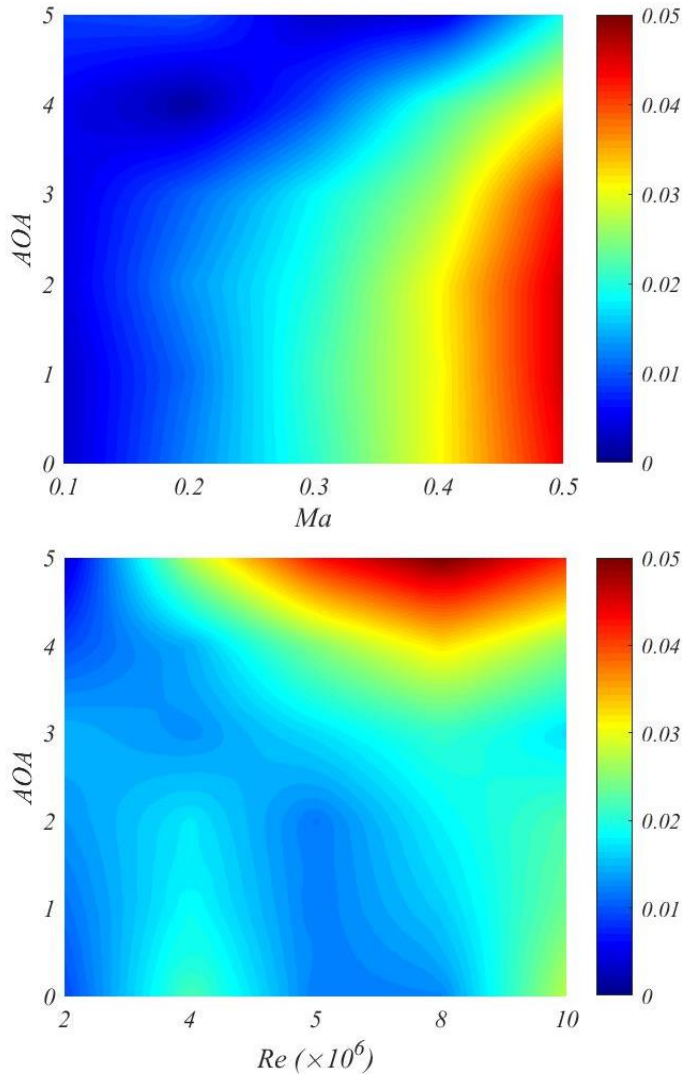
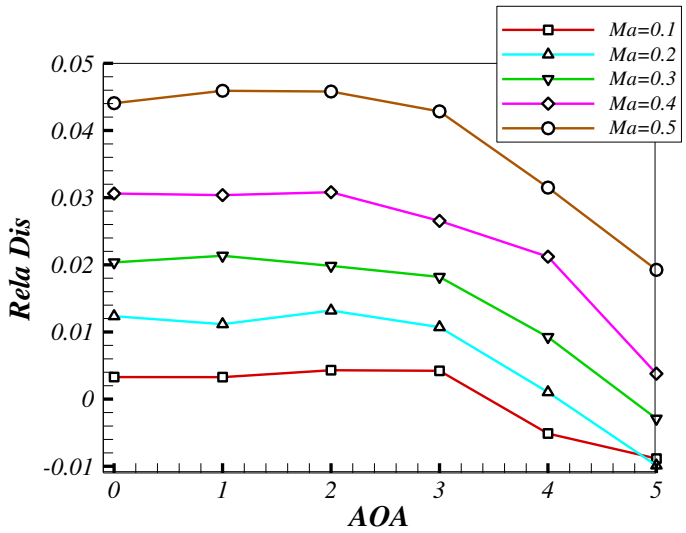


Fig.13 The relative error contour of C_d of the test cases.



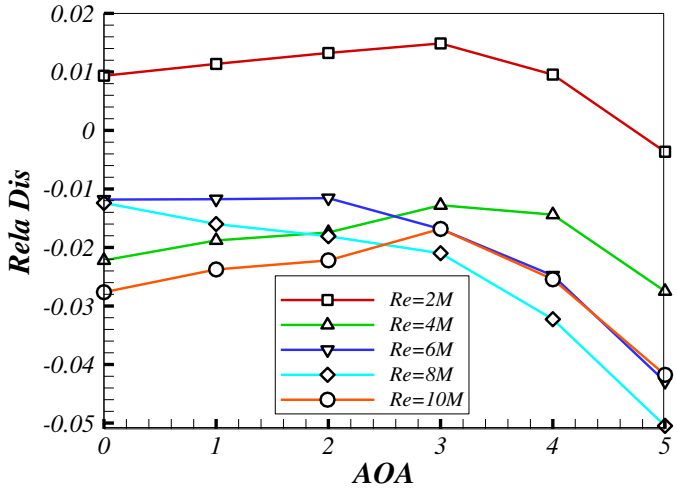
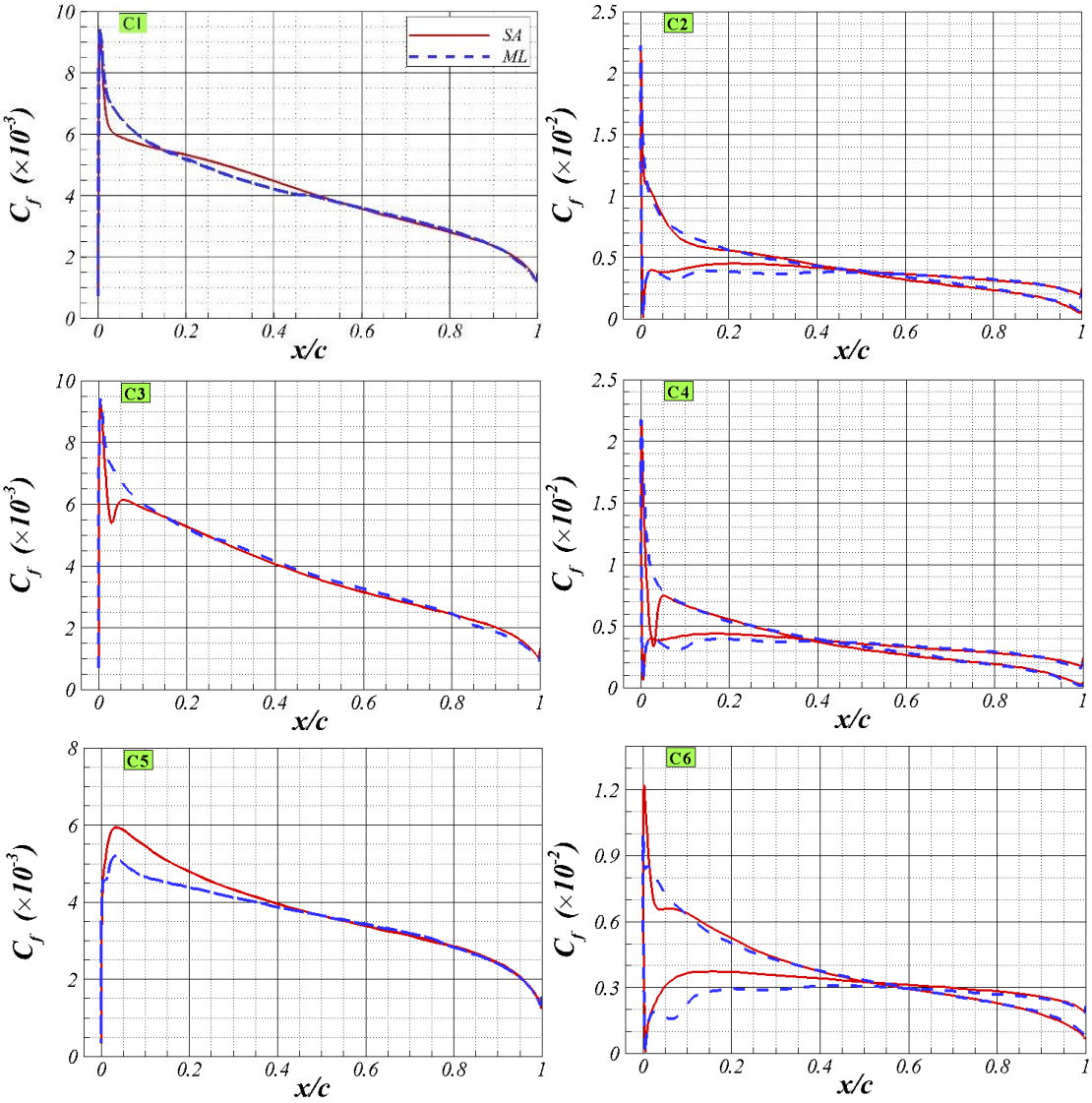


Fig. 14 The relative error of C_d of the test cases， where $\text{Rela Dis} = (C_{d,ML} - C_{d,SA}) / C_{d,SA}$



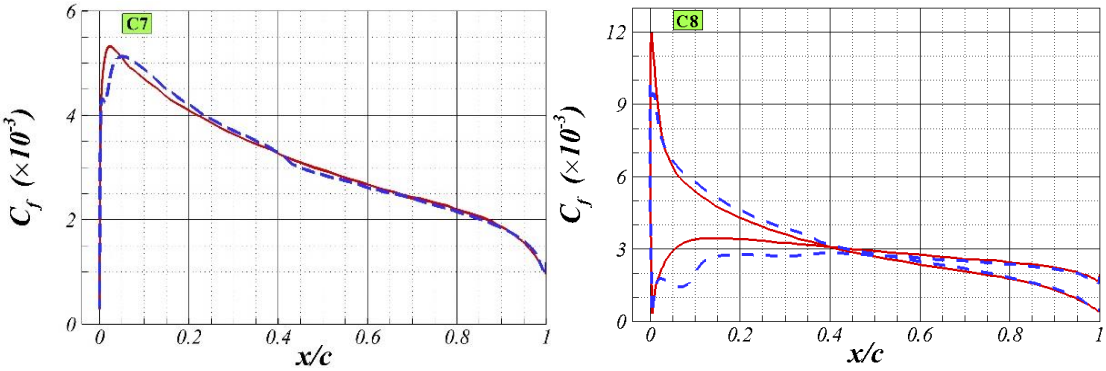


Fig. 15 The C_f comparison of the eight boundary point cases of NACA0012 airfoil.

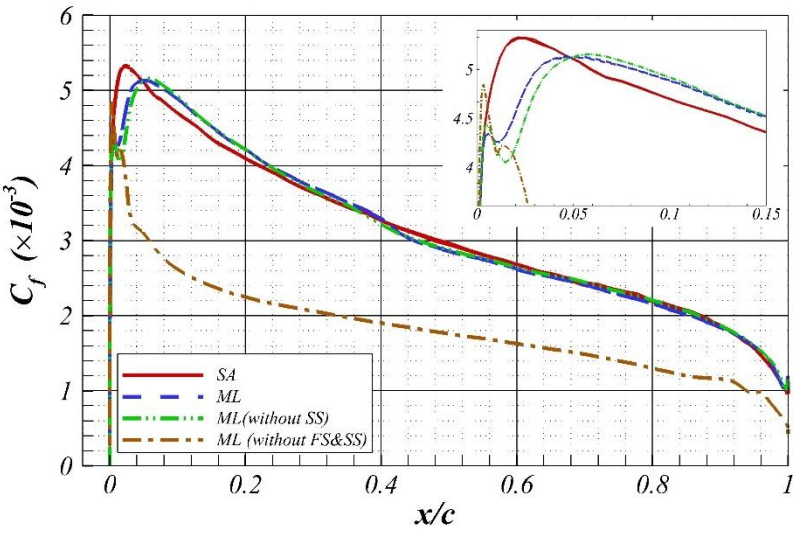
It can be found from the above results that the model performance tends to be poor at some boundaries of the test space, which is helpful to speculate the model performance in the whole test space. The freestream conditions of the eight boundary point cases and the corresponding relative errors of C_d are listed in Table 6. Only case C6 slightly exceeds 5%, while the other cases are all lower than 4%. The corresponding skin friction coefficient distribution is shown in Figure 15. The case C7 and C6 correspond to the best and worst model performance, respectively.

Table 6 The relative error of C_d of the eight boundary point cases in the test space.

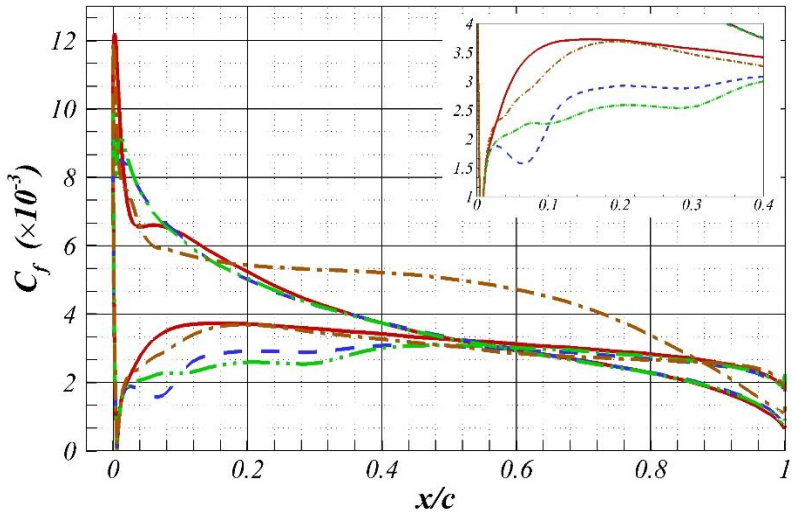
Case	Re ($\times 10^6$)	Ma	α (°)	Relative Error
C1	2	0.1	0	0.51%
C2	2	0.1	5	1.37%
C3	2	0.5	0	2.91%
C4	2	0.5	5	2.57%
C5	10	0.1	0	3.62%
C6	10	0.1	5	5.09%
C7	10	0.5	0	0.25%
C8	10	0.5	5	3.99%

In order to validate the proposed sample selection and feature selection algorithms, two models were constructed respectively under the same model architecture: one is without sample selection and feature selection and the other is only without sample selection. Take the two cases of NACA0012 airfoil as the example, and the comparison of C_f calculated by the three models is shown in Figure 16. The result is obviously wrong without feature selection, although the accuracy of the lower surface at the

leading edge is better for case C6. This indicates that more features do not mean better model performance, and the ability of neural network to select feature automatically is limited. Elaborate feature selection is necessary. Sample selection has negligible influence on the result, except for slight deviation at the leading edge of the airfoil, which indicates that the proposed sample selection algorithm is feasible and the selected samples preserve the data space and diversity of the original dataset. Thus, what should be focused on is the diversity rather than the number of samples. In terms of these two cases, the C_f agreement is very good at the upper surface, which is obviously better than that of the lower surface. The deficiency mainly manifests in two aspects. One is around the peak value of C_f at the leading edge, where the flow variables change dramatically. The other lies in the lower surface from the leading edge to the 0.4 times of chord length, where the C_f calculated by the model is smaller. When generalized to the RAE2822 airfoil, the model shows similar performance, as is shown in Figure 17.

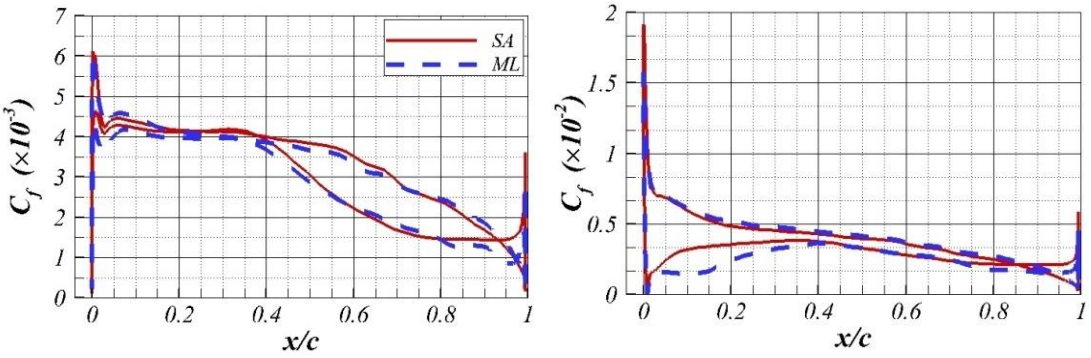


(a) C7



(b) C6

Fig. 16 The C_f comparison of NACA0012 airfoil.



(a) C7 (b) C6

Fig. 17 The C_f comparison of RAE2822 airfoil.

When the constructed models can be coupled with N-S solvers like conventional models, the research is of more practical significance. However, the error propagation between the Reynolds stress and the time-averaged flow field is a challenge for most of the research. Wu et al.[48] explained the stability of different closure models according to the local condition number. Cruz et al.[22] suggested using the Reynolds force vector instead of the Reynolds stress tensor as the model output. For the steady cases in this work, we expect that the convergent solution can be obtained when coupling the model with the N-S solver. Although we enhance the model robustness and the smoothness of the eddy viscosity field by stability training and space-time smoothing and the eddy viscosity model has its own stability advantage, these are still not enough to counteract the high sensitivity of time-averaged field to Reynolds stress, which causes the residual oscillation and the slower convergence.

Since the convergence process of all cases is similar, we take the case C6 of

NACA0012 airfoil as the example to make further analysis for the sake of convenience. In each iteration step, we tracked the Reynolds shear stress variation between the current step and the previous step of the cell with maximal residual value. In this regard, the change of S-A model decays fast and smoothly, while that of the ML model decays slower and has large oscillations frequently, as is shown in Figure 18. The oscillations of Reynolds stress variation derive from those of the model output, which disturb the convergence of the mean flow field and lead to the residual oscillations.

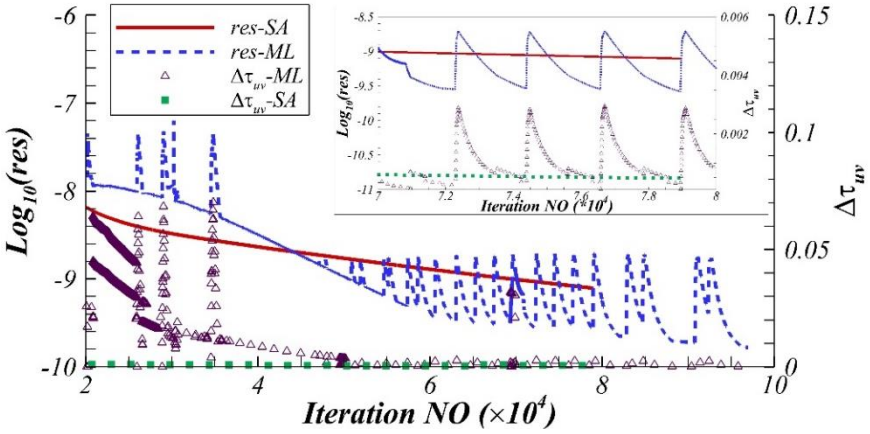


Fig. 18 The residual evolution for case C6 of NACA0012 airfoil.

5 Conclusion and outlook

In this paper, a feature selection method is developed to improve the generalization ability of machine learning models in turbulent flows. The method selects the features by three steps: firstly, measure the numerical distribution characteristics of the training and validation datasets according to the mean extrapolation distance and eliminate the features with obvious numerical extrapolation; secondly, construct a model with all features, then test and rank the effect of each feature on the model performance; finally, generate the feature subsets according to the feature importance and evaluate the model performance that each feature subset can achieve, then select the feature subset with optimal model performance. The proposed method avoids the shortcoming of priori of feature importance ranking and low efficiency of exhaustive research.

Using the features from the propposed method, a turbulence surrogate model is constructed with artificial neural network driven by the flow field data calculated by N-S solver coupled with S-A model. The training data is selected from only one flow

case, and 70 flow cases with different freestream conditions and airfoils are tested. The results indicate that the generalization ability of the constructed model is amazing. The model driven by only one flow case performs well in the whole designed test space after the feature selection. The relative error of the drag coefficient is within 5% for almost all the test cases.

This work mainly validates the proposed feature selection method. In future work, it can be compared with some existing feature selection methods for effect and efficiency.

Declarations

Acknowledgement

The authors would like to thank PhD students Xuxiang Sun and Xianglin Shan for their discussions and checking to this paper.

Authors' contributions

All authors read and approved the final manuscript.

Funding

This work is supported by the National Natural Science Foundation of China (no. 91852115), the National Numerical Wind tunnel Project (no. NNW2018-ZT1B01).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

References

[1] LING J, TEMPLETON J. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty[J]. Physics of Fluids, 2015, 27(8): 042032-042094.

[2] SINGH A P, DURAISAMY K. Using field inversion to quantify functional errors in turbulence closures[J]. Physics of Fluids, 2016, 28(4): 045110.
[3] DURAISAMY K, IACCARINO G, XIAO H. Turbulence modeling in the age of data[J]. Annual Review of Fluid Mechanics, 2019, 51: 357-377.
[4] MATAI R, DURBIN P A. Zonal Eddy Viscosity Models Based on Machine Learning[J]. Flow, Turbulence and Combustion, 2019, 103(1): 93-109.
[5] PARISH E J, DURAISAMY K. A paradigm for data-driven predictive modeling using field inversion and machine learning[J]. Journal of Computational Physics, 2016, 305: 758-774.
[6] SINGH A P, MEDIDA S, DURAISAMY K. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils[J]. AIAA Journal, 2017, : 2215-2227.
[7] ZHANG Z J, DURAISAMY K. Machine learning methods for data-driven turbulence modeling[C]. 22nd AIAA Computational Fluid Dynamics Conference, 2015 : 2460.
[8] WANG J X, WU J L, XIAO H. A Physics Informed Machine Learning Approach for Reconstructing Reynolds Stress Modeling Discrepancies Based on DNS Data[J]. Phys.rev.fluids, 2017, 2(3): 1-22.
[9] XIAO H, WU J, WANG J, et al. Physics informed machine learning for predictive turbulence modeling: Progress and perspectives[J]. AIAA Paper, 2017, (2017-1712).
[10] LING J, KURZAWSKI A, TEMPLETON J. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance[J]. Journal of Fluid Mechanics, 2016, 807: 155-166.
[11] ZHU L, ZHANG W, KOU J, et al. Machine learning methods for turbulence modeling in subsonic flows around airfoils[J]. Physics of Fluids, 2019, 31(1): 015105.
[12] ZHU L, ZHANG W, SUN X, et al. Turbulence closure for high Reynolds number airfoil flows by deep neural networks[J]. Aerospace Science and Technology, 2021, 110: 106452.
[13] NATHAN K J. Deep learning in fluid dynamics[J]. Journal of Fluid Mechanics, 2017, 814: 1-4.
[14] BRUNTON S L, NOACK B R, KOUMOUTSAKOS P. Machine learning for fluid mechanics[J]. Annual Review of Fluid Mechanics, 2019, 52.
[15] ZHANG W, ZHU L, LIU Y, et al. Progresses in the application of machine learning in turbulence modeling[J]. Acta Aerodynamica Sinica, 2019, 37(3): 444-454.

[16] GUYON I, ELISSEEFF A: An Introduction to Feature Extraction, Guyon Isabelle, Nikravesh Masoud, Gunn Steve, Zadeh Lotfi A., editor, Feature Extraction: Foundations and Applications, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 1-25.
[17] LING J, JONES R, TEMPLETON J. Machine learning strategies for systems with invariance properties[J]. Journal of Computational Physics, 2016, 318: 22-35.
[18] POPE S B. Turbulent Flows[J]. Turbulent Flows, 2000, 12(11): 806.
[19] WU J-L, XIAO H, PATERSON E. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework[J]. Physical Review Fluids, 2018, 3(7): 074602.
[20] YUHUIYIN, PUYANG, YUFEIZHANG, et al. Feature selection and processing of turbulence modeling based on an artificial neural network[J]. Physics of Fluids, 2020, 32(10): 105117.
[21] WANG Z, LUO K, LI D, et al. Investigations of data-driven closure for subgrid-scale stress in large-eddy simulation[J]. Physics of Fluids, 2018, 30(12): 125101.
[22] CRUZ M A, THOMPSON R L, SAMPAIO L E, et al. The use of the Reynolds force vector in a physics informed machine learning approach for predictive turbulence modeling[J]. Computers & Fluids, 2019, 192: 104258.
[23] BLUM A L, LANGLEY P. Selection of relevant features and examples in machine learning[J]. Artificial intelligence, 1997, 97(1-2): 245-271.
[24] JOHN G H, KOHAVI R, PFLEGER K: Irrelevant features and the subset selection problem, Machine Learning Proceedings 1994, : Elsevier, 1994: 121-129.
[25] SOLORIO-FERNáNDEZ S, CARRASCO-OCHOA J A, MARTíNEZ-TRINIDAD J F. A new hybrid filter–wrapper feature selection method for clustering based on ranking[J]. Neurocomputing, 2016, 214: 866-880.
[26] BERMEJO P, DE LA OSSA L, GáMEZ J A, et al. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking[J]. Knowledge-Based Systems, 2012, 25(1): 35-44.
[27] CHANDRASHEKAR G, SAHIN F. A survey on feature selection methods[J]. Computers & Electrical Engineering, 2014, 40(1): 16-28.
[28] GUYON I, ELISSEEFF A. An introduction to variable and feature selection[J]. Journal of machine learning research, 2003, 3(Mar): 1157-1182.
[29] CAI J, LUO J, WANG S, et al. Feature selection in machine learning: A new

perspective[J]. Neurocomputing, 2018, 300: 70-79.
[30] YU L, LIU H. Efficient feature selection via analysis of relevance and redundancy[J]. Journal of machine learning research, 2004, 5(Oct): 1205-1224.
[31] ERHARD P, ETLING D, MULLER U, et al. Prandtl-essentials of fluid mechanics[M]. 158. Springer Science & Business Media, 2010.
[32] PIROZZOLI S. Revisiting the mixing-length hypothesis in the outer part of turbulent wall layers: mean flow and wall friction[J]. Journal of Fluid Mechanics, 2014, 745: 378-397.
[33] GRANVILLE P. A modified van driest formula for the mixing length of turbulent boundary layers in pressure gradients[J]. ASME Transactions Journal of Fluids Engineering, 1989, 111: 94-97.
[34] DRIEST E V. On turbulent flow near a wall[J]. Journal of the aeronautical sciences, 1956, 23(11): 1007-1011.
[35] CLAUSER F H: The turbulent boundary layer, Advances in applied mechanics, : Elsevier, 1956: 1-51.
[36] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. nature, 2015, 521(7553): 436.
[37] BISHOP C M. Pattern recognition and machine learning[M]. springer, 2006.
[38] YU D, DENG L. AUTOMATIC SPEECH RECOGNITION[M]. Springer, 2016.
[39] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feedforward networks are universal approximators[J]. Neural networks, 1989, 2(5): 359-366.
[40] MAULIK R, SAN O. A neural network approach for the blind deconvolution of turbulent flows[J]. Journal of Fluid Mechanics, 2017, 831: 151-181.
[41] SPALART P, ALLMARAS S. A one-equation turbulence model for aerodynamic flows[C]. 30th aerospace sciences meeting and exhibit, 1992 : 439.
[42] JESPERSEN D C, PULLIAM T H, CHILDS M L. Overflow turbulence modeling resource validation results[J], 2016.
[43] BARDINA J E, HUANG P G, COAKLEY T J. Turbulence Modeling Validation, Testing, and Development[J]. Nasa Technical Memorandum, 1997.
[44] COOK P H, MCDONALD M A, FIRMIN M C P. Aerofoil RAE 2822 - Pressure Distributions, and Boundary Layer and Wake Measurements[J]. Experimental Data Base for Computer Program Assessment, 1979, AGARD AR 138.
[45] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]. International Conference on International

Conference on Machine Learning, 2015.
[46] ZHENG S, SONG Y, LEUNG T, et al. Improving the robustness of deep neural networks via stability training[C]. Proceedings of the ieee conference on computer vision and pattern recognition, 2016 : 4480-4488.
[47] PASZKE A, GROSS S, CHINTALA S, et al. Pytorch: Tensors and dynamic neural networks in python with strong gpu acceleration[J]. PyTorch: Tensors and dynamic neural networks in Python with strong GPU acceleration, 2017, 6.
[48] WU J, XIAO H, SUN R, et al. Reynolds-averaged Navier–Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned[J]. Journal of Fluid Mechanics, 2019, 869: 553-586.