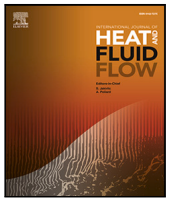
International Journal of Heat and Fluid Flow 97 (2022) 109038



Contents lists available at ScienceDirect

International Journal of Heat and Fluid Flow

journal homepage: www.elsevier.com/locate/ijhff





On the explainability of machine-learning-assisted turbulence modeling for transonic flows

Xiao He [a,*], Jianheng Tan [b], Georgios Rigas [b], Mehdi Vahdati [a]

[a] Department of Mechanical Engineering, Imperial College London, London SW7 2AZ, UK
[b] Department of Aeronautics, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Keywords:
Turbulence modeling
Machine learning
Data-driven
Transonic flow

ABSTRACT

Machine learning (ML) is a rising and promising tool for Reynolds-Averaged Navier–Stokes (RANS) turbulence model developments, but its application to industrial flows is hindered by the lack of explainability of the ML model. In this paper, two types of methods to improve the explainability are presented, namely the intrinsic methods that reduce the model complexity and the post-hoc methods that explain the correlation between the model inputs and outputs. The investigated ML-assisted turbulence model framework aims to improve the prediction accuracy of the Spalart–Allmaras (SA) turbulence model in transonic bump flows. A random forest model is trained to construct a mapping between the input flow features and the output eddy viscosity difference. Results show that the intrinsic methods, including the hyperparameter study and the input feature selection, can reduce the model complexity at a limited cost of accuracy. The post-hoc Shapley additive explanations (SHAP) method not only provides a ranked list of input flow features based on their global significance, but also unveils the local causal link between the input flow features and the output eddy viscosity difference. Based on the SHAP analysis, the ML model is found to discover: (1) the well-known scaling between eddy viscosity and its source term, which was originally found from dimensional analysis; (2) the well-known rotation and shear effects on the eddy viscosity source term, which was explicitly written in the Reynolds stress transport equations; and (3) the pressure normal stress and normal shear stress effect on the eddy viscosity source term, which has not attracted much attention in previous research. The methods and the knowledge obtained from this work provide useful guidance for data-driven turbulence model developers, and they are transferable to future ML turbulence model developments.

1. Introduction

Due to the prohibitive cost of large-eddy simulation (LES) and direct numerical simulation (DNS), RANS simulations will continue to be the standard approach for predicting flows across the majority of aerospace product categories (Slotnick et al., 2014). However, limited by the human decisions (i.e., empiricism) made under observing small data of simple canonical flows, conventional RANS turbulence models often fail to predict complex flows accurately, including shock-boundary layer (BL) interaction, 3D separation, etc. The underlying uncertainties of a RANS turbulence model can be categorized into four levels during the derivation of the model (Duraisamy et al., 2019): the ensemble-average process in deriving the RANS equations (L1), the constitutive relation linking the macroscopic and the microscopic state (e.g., Boussinesq hypothesis) (L2), the functional forms of the transport equations of turbulence quantities (e.g., k, ϵ, etc.) (L3) and the model parameters in the transport equations (L4). With the advancement of ML algorithms and the accumulation of high-fidelity flow databases, there is a growing interest in developing ML-assisted turbulence models to alleviate the empiricism at L2, L3 and L4.

Regarding the model-form uncertainty in the Boussinesq hypothesis at L2, a pioneering work from Ling and Templeton (2015) sets up an ML classification problem to identify regions where the Boussinesq hypothesis violates the DNS data. In their work, an abundant list of input features was provided, and different ML algorithms including random forest (RF) and adaptive boosting were compared using a wide range of LES/DNS databases. Later research by Ling et al. (2016) solved an ML regression problem to construct a non-linear constitutive equation of the Reynolds stress tensor. The resulting ML model based on a deep neural network demonstrated a higher accuracy compared to existing linear/non-linear eddy viscosity models. Wang et al. (2017) and Wu et al. (2018) proposed an ML regressor to calculate the discrepancies between the Reynolds stress tensor predicted

* Corresponding author.
E-mail address: xiao.he2014@imperial.ac.uk (X. He).

https://doi.org/10.1016/j.ijheatfluidflow.2022.109038
Received 31 January 2022; Received in revised form 23 June 2022; Accepted 28 July 2022
Available online 12 August 2022
0142-727X/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

APG	Adverse Pressure Gradient
BL	Boundary Layer
CV	Cross-Validation
DNS	Direct Numerical Simulation
DT	Decision Tree
FPG	Favorable Pressure Gradient
Hi-Fi	High-Fidelity
LES	Large Eddy Simulation
Lo-Fi	Low-Fidelity
ML	Machine Learning
RANS	Reynolds-Averaged Navier–Stokes
RF	Random Forest
SA	Spalart–Allmaras
SHAP	Shapley Additive Explanations
SST	Shear Stress Transport
c	bump chord length (m)
C_p	pressure coefficient
M	Mach number; number of input features of RF
N	number of input features of DT
P	number of training samples of RF
q_i	input feature
r	radial coordinate (m)
R^2	coefficient of determination
Re	Reynolds number
S_{ij}	strain rate tensor (s^{-1})
T	number of trees of RF
W_{ij}	rotation rate tensor (s^{-1})
x	axial coordinate (m)
ϵ_{L1}	L1-type error
ϵ_{L2}	L2-type error
ν	laminar kinematic viscosity (m^2/s)
ν_t	turbulent kinematic viscosity (m^2/s)
ϕ_i	SHAP value
$\phi_{i,j}$	SHAP interaction value

by established turbulence models and that derived from high-fidelity data. The proposed approach based on an RF regressor demonstrated a better prediction compared to the baseline two-equation model and the baseline Reynolds stress model. Based on the same approach, success was also achieved in a broader range of flows (Wang et al., 2019; Xiao et al., 2020). While the aforementioned ML methods may be argued to be black boxes, Weatheritt and Sandberg (2016, 2017) demonstrated a certain success in the application of symbolic regression. In their work, the gene expression programming algorithm was used to select from a dictionary of mathematical expressions for an algebraic Reynolds stress model. Another example of explainable ML methods is from Schmelzer et al. (2020), where the elastic net-regularization technique was used to promote sparsity of candidate model terms.

Regarding the model-form uncertainty in the transport equations at L3, Singh and Duraisamy (2016) proposed a field inversion approach to estimate the functional terms of a transport equation via Bayesian inference. This approach sets up a cornerstone for ML-assisted correction of L3 uncertainty as it provides means of estimating the functional terms, which cannot be derived from the LES/DNS data explicitly. Later research combined the field inversion and an ML regressor to correct the functional terms of existing models, and improved prediction accuracy was achieved by tuning the production term of the $k-\omega$ model (Parish and Duraisamy, 2016) and the SA model (Singh et al., 2017). The same ideal also applies to the correction of L4 uncertainty. Edeling et al. (2014) performed Bayesian calibration on the model parameters of $k-\epsilon$ model using measured BL data under different pressure gradients. Matai and Durbin (2019) adopted the field inversion method and the RF algorithm to assign different parameter sets of the $k-\omega$ model to regions with different flow features.

The literature review above demonstrated the feasibility of ML-assisted turbulence modeling in a range of canonical flows. However, its further development is constrained by the lack of explainability. From the perspective of academia, the explainability of an ML model is indispensable for identifying new physics. From the perspective of industry, human users can hardly trust an ML-assisted turbulence model unless it is explainable to some extent.

This paper discusses ways to improve the explainability of an ML-assisted turbulence model, and it is organized as follows. In Section 2, the investigated ML-assisted turbulence modeling framework is described in detail, including the flow solver, the training and testing data, the input and output formulation and the random forest model. In Section 3, the intrinsic method and the post-hoc method that improve the explainability of an ML model are introduced. Finally, the results of the intrinsic method and the post-hoc method will be presented in Sections 4 and 5, respectively.

2. Machine-learning-assisted turbulence modeling framework

2.1. Revisit the RANS closure problem

By applying the Favre average to the three-dimensional unsteady Navier–Stokes equation, the RANS equation is obtained as Eq. (1):

$$\frac{\partial \bar{\rho}\tilde{u}_i}{\partial t}+\frac{\partial \bar{\rho}\tilde{u}_i\tilde{u}_j}{\partial x_j}=-\frac{\partial \bar{p}}{\partial x_i}+\frac{\partial \bar{\tau}_{ij}}{\partial x_j}+\frac{\partial(-\overline{\rho u_i' u_j'})}{\partial x_j} \tag{1}$$

where $\bar{(\cdot)}$ and $\tilde{(\cdot)}$ denotes the ensemble average and the density-weighted ensemble average, respectively.[1] The symmetric Reynolds stress tensor $\tau_{ij}^{RANS} \equiv -\overline{\rho u_i' u_j'}$ represents the effect of turbulent fluctuations on the mean flow, which needs to be closed via a turbulence model. An underlying assumption of a RANS turbulence closure is that τ_{ij}^{RANS} can be uniquely determined by the mean flow quantities without the information of the microscopic state. Such an assumption introduces the L1 uncertainty that is fundamentally irrecoverable.

An efficient and widely applied approach to close τ_{ij}^{RANS} is based on the Boussinesq assumption shown in Eq. (2):

$$\tau_{ij}^{RANS}=2\rho\nu_t S_{ij}^{*}-\frac{2}{3}\rho k\delta_{ij} \tag{2}$$

where $S_{ij}^{*}=S_{ij}-\mathrm{tr}(S_{ij})$ is the traceless strain rate tensor, ν_t is eddy viscosity and k is the turbulence kinetic energy. This equation reduces the closure task of the τ_{ij}^{RANS} tensor to the two scalars of ν_t and k. However, such a linear constitutive relation has long been known deficient in 3D separated flows (Wilcox, 2006). A more general constitutive relation can be formulated by a linear combination of ten basis tensors, and the coefficients are functions of five tensor invariants (Pope, 1975). By using the leading two basis tensors, the quadratic constitutive relation (QCR) (Spalart, 2000) is obtained:

$$\tau_{ij}^{RANS}=2\rho\nu_t[S_{ij}^{*}-c_{cr1}(O_{ik}S_{jk}^{*}+O_{jk}S_{ik}^{*})]-\frac{2}{3}\rho k\delta_{ij} \tag{3}$$

where $O_{ij}=2W_{ij}/\sqrt{\partial_n u_m \partial_n u_m}$ represents the antisymmetric normalized rotation rate tensor, and the coefficient $c_{cr1}=0.3$ was calibrated in a BL flow (Spalart, 2000). Note that the constitutive relations of Eq. (2) and (3) are not exact; uncertainty introduced at this stage is referred to as the L2 uncertainty.

[1] For simplicity, the symbols $\bar{(\cdot)}$ and $\tilde{(\cdot)}$ of mean flow quantities are removed in the following discussion.

The closure of ν_t and k involves algebraic or transport equations. For the one-equation SA model (Spalart and Allmaras, 1994) investigated in this work, it solves the transport equation of Eq. (4):

$$\frac{\partial \hat{\nu}}{\partial t} + u_j \frac{\partial \hat{\nu}}{\partial x_j} = \underbrace{c_{b1}\hat{S}\hat{\nu}}_{\text{production}} \underbrace{-c_{w1} f_w \left(\frac{\hat{\nu}}{d}\right)^2}_{\text{destruction}} + \underbrace{\frac{1}{\sigma}\left[\frac{\partial}{\partial x_j}\left((\nu+\hat{\nu})\frac{\partial \hat{\nu}}{\partial x_j}\right) + c_{b2}\frac{\partial \hat{\nu}}{\partial x_i}\frac{\partial \hat{\nu}}{\partial x_i}\right]}_{\text{diffusion}} \tag{4}$$

where the eddy viscosity $\nu_t = \hat{\nu} f_{v1}$; d is the wall distance; f_{v1}, $\hat{S}$ and f_w are additional algebraic equations; c_{b1}, c_{w1}, σ and c_{b2} are model coefficients that were calibrated in free shear flows and BL flows (Spalart and Allmaras, 1994). The k term is usually ignored in the standard version of the SA model, but it can be approximated via Eq. (5) (Rumsey et al., 2020):

$$k = 3 c_{cr2} \nu_t \sqrt{2 W_{ij} W_{ij}} \tag{5}$$

where the constant $c_{cr2} = 1.25$. Note that the transport equation of Eq. (4) was not supported by any physical conservation law; the formulation was proposed for equilibrium (i.e., production equals destruction) turbulence of attached BL flows, but it will introduce L3 uncertainty in detached flows. In previous research, many empirical fixes to reduce the L3 uncertainty in specific types of flow have been proposed, and certain degree of success has been achieved (Dacles-Mariani et al., 1995; Spalart and Shur, 1997; Zhang and Yang, 2013; He et al., 2022).

Finally, parametric uncertainty or L4 uncertainty arises due to the nature of calibration. From the previous uncertainty quantification studies, however, tuning the coefficients within the calibrated ranges does not appear useful for complex academia flows or industrial flows (He et al., 2020, 2021).

The discussion above indicates that data-driven modifications at L2 or L3 appear effective. Considering L3 modification is better conditioned and more robust than the L2 modification (Wu et al., 2019), this paper aims to reduce the L3 model-form uncertainty in the SA-QCR transport equation by an ML model.

2.2. Framework overview

The flowchart of the ML-assisted turbulence model is illustrated in Fig. 1, which generally follows the method of Ref. Wu et al. (2018). The idea is to introduce a data-driven correction term of eddy viscosity $\Delta\nu_t$, i.e., $\nu_t^{\mathrm{ML}} = \nu_t + \Delta\nu_t$. This $\Delta\nu_t$ term can be interpreted as an additional source term calculated by ML, which reduces the deficiency of the SA-QCR model in complex flows. The complete framework consists of two phases: (1) offline training and test; (2) online test and prediction. Here, the difference between online and offline lies in whether the ML model participates in the iterative process of the CFD solver.

During the offline phase, the first step is to select a range of learning flow cases considering the flow physics of interest and the availability of data. The selected cases are simulated by both high-fidelity (Hi-Fi) and low-fidelity (Lo-Fi) CFD to generate a database. This database will be divided into a training set and a test set, where the training set is known by the ML model for training purposes, whereas the test set is unseen by the ML model in order to measure its prediction accuracy. The input of the ML model is the non-dimensional flow features calculated from the Lo-Fi flow solutions, which inform the local flow physics to the model. The output of the ML model is the difference between the Hi-Fi and the Lo-Fi turbulence quantities, which is the anisotropy of the Reynolds stress tensor for L2 correction, the eddy viscosity for L3 correction (adopted in this work) and the model coefficients for L4 correction. Finally, the behavior of the trained ML model will be explained to build trust in the ML model.

For the online test and prediction phase, the trained ML algorithm will be implemented in the CFD solver to predict unseen application flows. During each CFD iteration, the ML model will be called to correct the turbulence quantities predicted by an established turbulence model. The outcome of the online test includes not only the turbulence quantities but the mean flow field.

This paper focuses on the model explanation and discusses the results from the offline training/test phase only.

2.3. Learning flow cases

The case studied in this work is the axisymmetric transonic bump flow featured by shock-BL interaction and shock-induced separation, as illustrated in Fig. 2. The operating conditions of the bump are at the freestream temperature T_∞ of 300 K, the Reynolds number $Re_c = U_\infty c/\nu_\infty$ of 2.763×10^6, and six different freestream Mach numbers $M_\infty = U_\infty/a_\infty$ at 0.6, 0.8, 0.85, 0.875, 0.9 and 0.925, respectively. These settings are identical to the experiment (Johnson, 1987).

The geometry and the boundary conditions of the bump are illustrated in Fig. 3(a), where values at the Riemann boundaries are calculated from the freestream T_∞, Re_c and M_∞. Two groups of grids are used to discretize the bump flow domain in this work, as summarized in Table 1. The first group of grids (ID 1, 2 and 3) is designed by NASA TMR (Rumsey, 2020) for RANS simulations. The grid refinement from the coarse, medium to fine grid is performed by doubling the number of grid nodes in all dimensions. These grids will be used for validation and verification of the flow solver and demonstration of the intrinsic method (Section 3.1) to improve ML model explainability. The second group of grids (ID 4 and 5) is obtained from a previous LES study (Uzun and Malik, 2019). A refined grid (ID 4) was used by LES to generate a turbulence database, but it is too dense for RANS simulations. Therefore, by picking every few grid nodes in each dimension of the original grid, a simplified grid (ID 5) is generated, which has a grid density between RANS(medium) and RANS(fine). The simplified grid will be used for demonstration of the post-hoc method (Section 3.2) to improve ML model explainability. A zoom-in view of the RANS(medium) grid is shown in Fig. 3(b).

2.4. Flow solver, validation and verification

The steady RANS solver with the Jameson–Schmidt–Turkel central scheme (Jameson et al., 1981) from the Stanford University Unstructured (SU2) tool suite (Version 7.1.0) (Economon et al., 2016) is used in this paper. To verify the implementation of turbulence models in SU2, the predictions of the wall pressure coefficient are compared against that of CFL3D in Fig. 4(a). It shows that the SU2 results almost overlap with that from CFL3D, validating the implementation of the SA (Spalart and Allmaras, 1994) and SST (Menter, 1994) models in SU2. The mesh independence is checked in Fig. 4(b) where the wall pressure coefficient distributions predicted by different grids are compared. Results show that the mesh independence is reached when using a grid density equal to or finer than the RANS(medium) grid. Mesh independence is also reached when using the SST model and comparing the skin friction distribution, which is not shown here for brevity. Therefore, the RANS(medium) grid and the LES(simplify) grid are considered adequately fine and can be used in the following work.

2.5. Training and testing data

The open-accessed Hi-Fi data for transonic separated flows are currently limited—Hi-Fi data from LES (Uzun and Malik, 2019) and Laser Doppler velocimetry measurements (Bachalo and Johnson, 1986) of the bump case are only available at $M_\infty = 0.875$. Meanwhile, it was shown previously in Fig. 4(a) that the SST model predicts the wall pressure distribution well at $M_\infty = 0.875$. Using the SST-QCR results to train the SA-QCR model in the bump case seems plausible. In the following, the fidelity of the SST-QCR result and its potential usage for the ML-assisted turbulence model are discussed in detail.

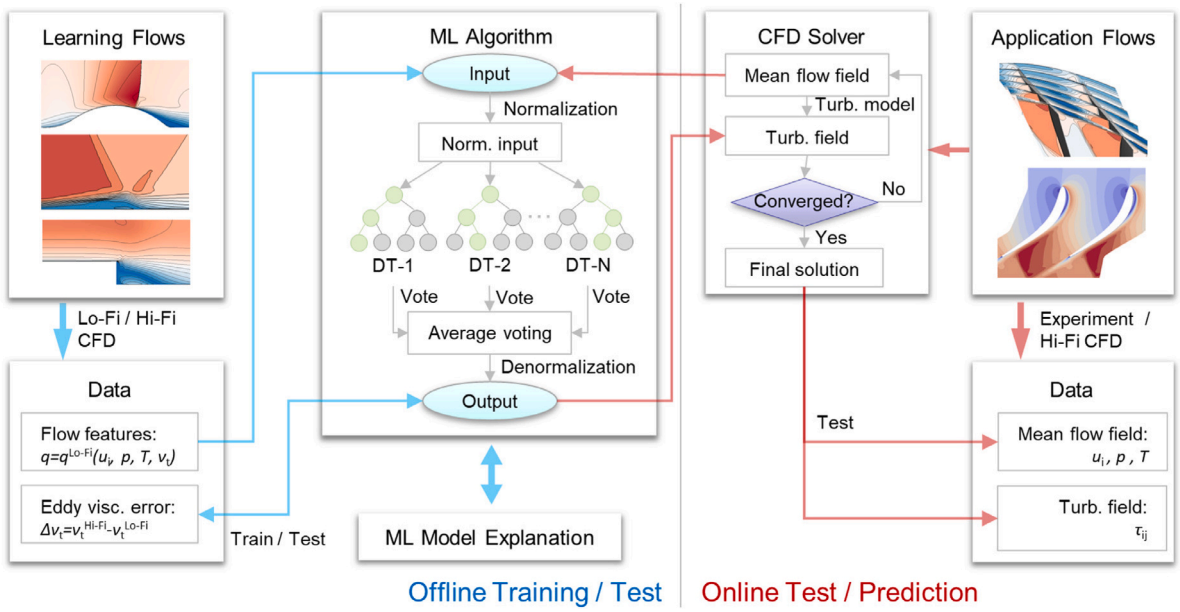


Fig. 1. Illustration of machine-learning-assisted turbulence modeling framework.

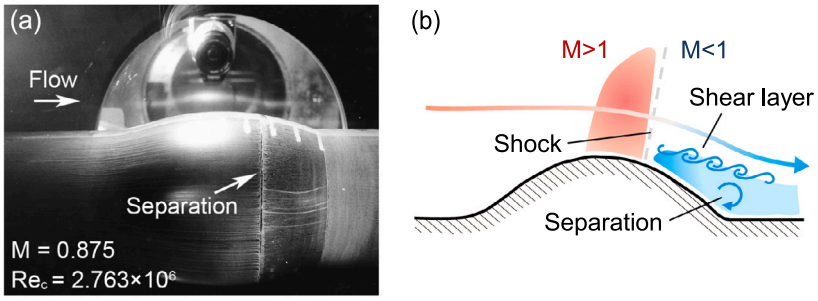


Fig. 2. Illustration of transonic bump: (a) surface oil flow visualization and (b) flow schematics.
Source: (a) courtesy of Dr. Dennis Johnson.

Table 1
Summary of grid details (coordinate system is consistent with Fig. 3).

Grids				Domain size (in chord length)				Node number	
ID	Name	Source	Usage	x_{min}	x_{max}	r_{min}	r_{max}	n_x	n_r
1	RANS(coarse)	NASA TMR (Rumsey, 2020)	Section 2	−3.2	4.4	0.375	4.0	91	41
2	RANS(medium)	NASA TMR (Rumsey, 2020)	Sections 2 and 4	−3.2	4.4	0.375	4.0	181	81
3	RANS(fine)	NASA TMR (Rumsey, 2020)	Section 2	−3.2	4.4	0.375	4.0	361	161
4	LES(original)	Uzun and Malik (2019)	–	−1.5	1.7	0.375	10.0	5921	489
5	LES(simplify)	Uzun and Malik (2019)	Section 5	−1.0	1.7	0.375	4.4	320	120

The fidelity of the SST-QCR model in predicting the mean flow field is examined in Fig. 5. In Fig. 5(a), the pressure coefficient distributions at different M_∞ conditions are presented. It shows that the SST-QCR model predicts the pressure distribution well at all Mach number conditions, especially at $M_\infty \geq 0.85$ where the shock front location and the downstream pressure plateau in the separation region are well-captured; the SA-QCR model predicts the pressure distribution well at low Mach numbers without shocks, but it predicts a delayed shock front at $M_\infty \geq 0.85$. In Fig. 5(b), the velocity profiles at $M_\infty = 0.875$ are shown. The SST-QCR results show close agreement with the LES results especially in the pre-shock BL ($x/c < 0.6$) and the flow reversal region ($0.7 < x/c < 1.2$); the SA-QCR model predicts a delayed shock-induced separation and deviates from the LES data the most.

The fidelity of the SST-QCR model in predicting the turbulence quantity field is examined in Fig. 6. In Fig. 6(a), the Reynolds shear stress profiles at $M_\infty = 0.875$ are presented. Results show that both the SA-QCR and SST-QCR model under-predict the eddy viscosity downstream of the shock, and the SST-QCR model outperforms the SA-QCR model slightly. In Fig. 6(b), the viscosity ratio contours at $M_\infty = 0.875$ are presented in log-scale. Here, the eddy viscosity v_t of LES is calculated from the turbulence statistics and the QCR, which will be introduced in detail later in Section 2.7. The SST-QCR model qualitatively captures the near-wall high eddy viscosity regions. Still, the details such as the contour shape and the eddy viscosity magnitude in these regions are different from the LES result.

Considering the limited availability of LES data and the fidelity of the SST-QCR data, the SST-QCR data will be used for tuning the structural parameters of the ML model (Section 4), and the LES data will be used for the final training (Section 5). Flow physics hidden beneath the LES data is expected to be unveiled by the ML model explanation.

Table 2
Input features of the ML-assisted turbulence model. $||\cdot||$ and $|\cdot|$ represent matrix norm and vector norm, respectively. All flow features are calculated in the reference frame attached to the bump.

Group	ID	Raw feature	Normalized feature
(A) pressure gradient	q_1	$U_k \frac{\partial p}{\partial x_k}$	$\frac{U_k \frac{\partial p}{\partial x_k}}{\sqrt{\frac{\partial p}{\partial x_j}\frac{\partial p}{\partial x_j} U_i U_i} + \lvert U_k \frac{\partial p}{\partial x_k} \rvert}$
(A) pressure gradient	q_2	$U_k \frac{\partial p}{\partial x_k}$	$\frac{U_k \frac{\partial p}{\partial x_k}}{10^{-5} \frac{\rho^2 \lvert U_i \rvert^3}{\mu} + \lvert U_k \frac{\partial p}{\partial x_k} \rvert}$
(A) pressure gradient	q_3	$\sqrt{\frac{\partial p}{\partial x_k}\frac{\partial p}{\partial x_k}}$	$\frac{\sqrt{\frac{\partial p}{\partial x_k}\frac{\partial p}{\partial x_k}}}{\sqrt{\frac{\partial p}{\partial x_k}\frac{\partial p}{\partial x_k}} + \lvert \frac{1}{2}\rho \frac{\partial U_k^2}{\partial x_k} \rvert}$
(B) strain versus rotation	q_4	$\frac{1}{2}\left(\lvert\lvert W_{ij}\rvert\rvert^2 - \lvert\lvert S_{ij}\rvert\rvert^2\right)$	$\frac{\lvert\lvert W_{ij}\rvert\rvert^2 - \lvert\lvert S_{ij}\rvert\rvert^2}{\lvert\lvert W_{ij}\rvert\rvert^2 + \lvert\lvert S_{ij}\rvert\rvert^2}$
(B) strain versus rotation	q_5	$\frac{\lvert\lvert S_{ij}\rvert\rvert}{\lvert\lvert W_{ij}\rvert\rvert}$	$\tanh\left(\frac{\lvert\lvert S_{ij}\rvert\rvert}{\lvert\lvert W_{ij}\rvert\rvert}\right)$
(B) strain versus rotation	q_6	$\frac{\lvert\lvert W_{ij}\rvert\rvert}{\lvert\lvert S_{ij}\rvert\rvert^2}\left(\lvert\lvert W_{ij}\rvert\rvert - \lvert\lvert S_{ij}\rvert\rvert\right)$	$\tanh\left(\frac{\lvert\lvert W_{ij}\rvert\rvert}{\lvert\lvert S_{ij}\rvert\rvert^2}\left(\lvert\lvert W_{ij}\rvert\rvert - \lvert\lvert S_{ij}\rvert\rvert\right)\right)$
(C) flow misalignment	q_7	$\lvert U_i U_j \frac{\partial U_i}{\partial x_j} \rvert$	$\frac{\lvert U_i U_j \frac{\partial U_i}{\partial x_j} \rvert}{\sqrt{U_i U_i U_i \frac{\partial U_i}{\partial x_j} U_k \frac{\partial U_k}{\partial x_j}} + \lvert U_i U_j \frac{\partial U_i}{\partial x_j} \rvert}$
(D) wall proximity	q_8	$r = \frac{\hat{\nu}}{S\kappa^2 d^2}$	$f_w = g(r)\left[\frac{1+c_{w3}^6}{g(r)^6 + c_{w3}^6}\right]^{1/6}$
(E) viscosity ratio	q_9	ν_t/ν	$\frac{\nu_t/\nu}{500 + \lvert\nu_t/\nu\rvert}$

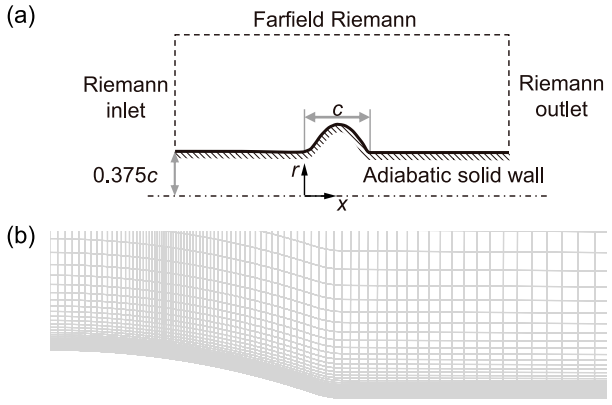


Fig. 3. Flow domain of transonic bump: (a) geometry and boundary conditions and (b) RANS(medium) grid (zoom-in view).

2.6. Input formulation

The input formulation for the ML model is crucial because it incorporates physical domain knowledge into the model and hence enables the model to extrapolate beyond training flows. In this work, an abundant set of input features are formulated in Table 2, partly following the work of Ling and Templeton (2015). In total, nine input features ($q_i, 1 \le i \le 9$) from five physics groups (group A, B, C, D and E) are investigated. The physical relevance of each feature is discussed in the following.

The group (A) pressure gradient features account for the observation that an adverse pressure gradient (APG) enhances turbulence while a favorable pressure gradient (FPG) diminishes it (Kline et al., 1967). In particular:

- q_1 measures the alignment between the pressure gradient vector and the velocity vector. An APG is indicated by $q_1 > 0$ while an FPG is represented by $q_1 < 0$. q_1 is bounded within $[-0.5, 0.5]$.
- q_2 measures the streamwise pressure gradient magnitude, which is non-dimensionalized by the local kinetic energy and the local Reynolds number per unit length following the K parameter (Kline et al., 1967). Its sign indicates APG/FPG as q_1 does, and its absolute value represents the streamwise pressure gradient magnitude. q_2 is bounded within $[-1.0, 1.0]$.
- q_3 measures the ratio of pressure normal stress to the normal shear stress. q_3 is bounded within $[0.0, 1.0]$. $q_3 < 0.5$ indicates the normal shear stress dominates over the pressure normal stress, and vice versa.

In addition to the features discussed above, another possible candidate is the vortical pressure gradient $\omega_i \frac{\partial p}{\partial x_i}$ (He et al., 2022). This feature is pronounced in industrial flows with a 3D swirl motion and a strong pressure gradient, but it remains zero in 2D flows and thus not considered here. In previous BL flow research (Nickels, 2004), the pressure gradient term can also be non-dimensionalized by using the wall shear stress τ_w and the BL thickness δ. However, when the investigated flow has flow separations or multiple walls, the physical relevance of using these wall quantities for non-dimensionalization becomes vague. Hence, they are not adopted for new feature development either.

The group (B) strain versus rotation features address the fact that a strong shear enhances turbulence intensity whereas a strong rotation reduces it (Spalart and Shur, 1997). Specifically:

- q_4 represents the dimensionless Q-criterion (Hunt et al., 1988) that measures the rotation-rate magnitude in excess of the strain-rate magnitude. It is bounded within $[-1.0, 1.0]$. When $q_4 > 0$, the rotation effect dominates over the shear effect, and vice versa.
- q_5 measures the ratio of strain-rate magnitude over rotation-rate magnitude. For better scaling, it is limited to the interval of $[0.0, 1.0]$ by a hyperbolic tangent function. In previous research, modification to the SA model based on q_5 (i.e., SA-R) showed improved predictions on the wingtip vortex (Dacles-Mariani et al., 1995).
- q_6 is the Richardson number (Hellsten, 1998) that measures the rotation-rate magnitude over strain-rate magnitude (i.e., opposite to q_5). It is also limited to the interval of $[-1.0, 1.0]$ by a hyperbolic tangent function. In previous research, implementation of q_6 in the SA model achieved certain improvements in predicting some flows with rotation and curvature (Zhang and Yang, 2013).

To avoid division by zero, a reference strain rate u_{ref}/l_{ref} is added to both the $\|W_{ij}\|$ and $\|S_{ij}\|$ terms, where u_{ref} and l_{ref} are characteristic velocity and length scale of the investigated case.

The group (C) flow misalignment feature considers the fact that RANS models are tuned accurately in 2D parallel shear flows but often fail in flows deviating from them (Gorlé et al., 2014). In particular, q_7 measures the non-orthogonality between the velocity vector and the gradient of the streamline-aligned velocity. It is bounded within $[0.0, 0.5]$. When $q_7 = 0$, it indicates either a perfectly parallel flow or the shear vanishes; in the case of any deviation from a parallel shear flow, $q_7 > 0$. Another candidate of this group is the vortex tilting measure (VTM) (Shur et al., 2015), which identifies quasi-2D flow regions from

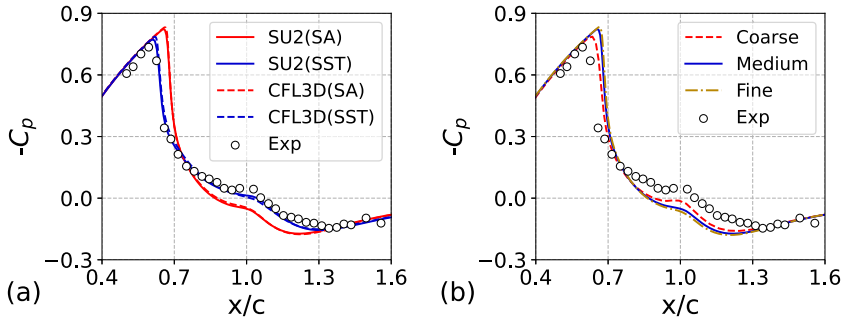


Fig. 4. Wall pressure coefficient distribution at $M_\infty = 0.875$: (a) verification of SU2 and (b) mesh independence (SA results).

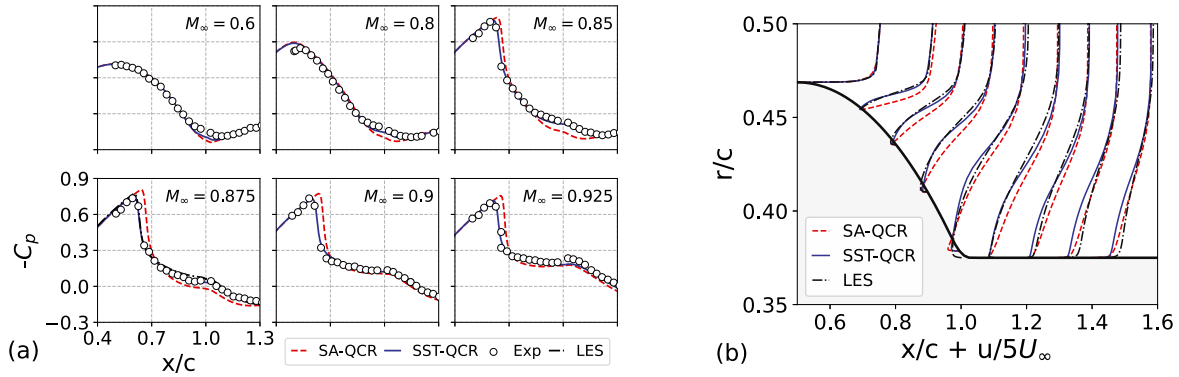


Fig. 5. Comparison between the SA-QCR and SST-QCR model on predicting the mean flow field: (a) wall pressure coefficient distribution at all M_∞ numbers (LES is only available at $M_\infty = 0.875$ and is overlapping with SST-QCR); (b) velocity profiles at $M_\infty = 0.875$.

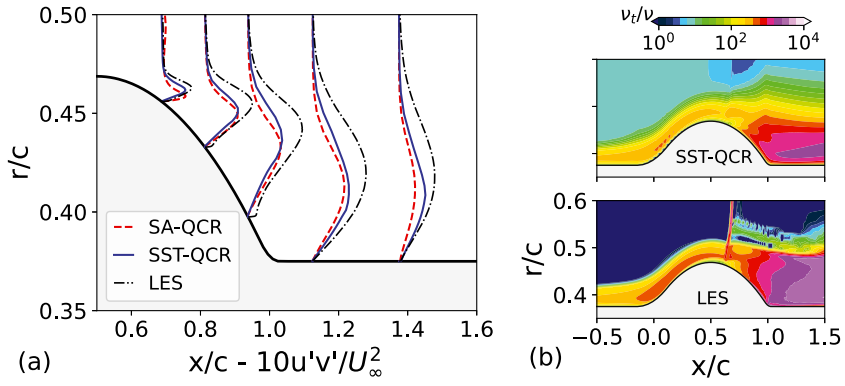


Fig. 6. Comparison between the SA-QCR and SST-QCR model on predicting the turbulence quantity field: (a) Reynolds shear stress profiles at $M_\infty = 0.875$; (b) viscosity ratio contours at $M_\infty = 0.875$.

developed 3D vortical regions. However, this feature remains zero in 2D flows and thus is not considered here.

The group (D) wall proximity feature considers that the turbulence length scale (i.e., mixing length) is proportional to the wall distance d in a classic log-law region of a BL. In the SA model, the destruction term scales with d^{-2} based on dimensional analysis (Spalart and Allmaras, 1994). Here, q_8 is based on the r parameter in the SA model, which represents the squared length scale ratio between the mixing length and the wall distance. It is bounded within $[0.0, 2.0]$ via the f_w function in the SA model, which is originally used to speed up turbulence destruction for better prediction of skin friction. The f_w function values and the correspond identified flow regions are summarized in Table 3. Another option to normalize r is via the shielding function f_d of DDES (Spalart et al., 2006). This form of wall proximity is mostly binary and has low resolution inside the BL. Thus, it is not considered here.

Table 3
Identification of flow regions by the f_w function.

f_w values	Flow region
$0.0 \le f_w < 0.4$	Free shear flow; viscous sublayer
$0.4 \le f_w < 0.8$	Outer part of BL
$0.8 \le f_w < 1.0$	Near-wall BL under FPG
$f_w = 1.0$	Log layer of BL
$1.0 < f_w \le 2.0$	Near-wall BL under APG

The group (E) viscosity ratio feature measures the ratio between the turbulent viscosity ν_t and the laminar viscosity ν, which is high in the near-wall region and the separation region. From dimensional analysis, the production term and the destruction term of the SA model scale with ν_t and ν_t^2, respectively (Spalart and Allmaras, 1994). Here, q_9 is a bounded version of the viscosity ratio, whose value falls

within [0.0, 1.0]. A scaling factor of 500 applies to reflect the magnitude variation in this case.

2.7. Output formulation

The output of the ML model is the normalized eddy viscosity difference (i.e., normalized $\Delta\nu_t = \nu_t^{\text{Hi-Fi}} - \nu_t^{\text{Lo-Fi}}$) with the same normalization scheme of q_9 shown in Table 2. Here, $\nu_t^{\text{Lo-Fi}}$ is calculated by solving the turbulence model transport equation(s) iteratively while keeping the mean flow field the same as the Hi-Fi results; $\nu_t^{\text{Hi-Fi}}$ is readily available for the SST-QCR data, but it needs to be approximated from the turbulence statistics of LES/DNS results.

The optimal approximation of $\nu_t^{\text{Hi-Fi}}$ minimizes $J = \left\| -\overline{\rho u_i' u_j'} - \tau_{ij}^{RANS} \right\|_F$. Therefore, the approximated ν_t value will be the root of $\partial J/\partial \nu_t = 0$. If τ_{ij}^{RANS} is calculated from the Boussinesq constitutive relation of Eq. (2), ν_t can be obtained as Eq. (6):

$$\nu_t = -\frac{\overline{u_i' u_j'} S_{ij} - \frac{1}{3}\overline{u_k' u_k'} S_{kk}}{2[S_{ij}S_{ij} - \frac{1}{3}(S_{kk})^2]} \tag{6}$$

which was firstly derived by Spalart et al. (2017). If τ_{ij}^{RANS} is calculated from the QCR of Eq. (3) and the approximated k of Eq. (5), ν_t can be obtained from Eq. (7):

$$\nu_t = -\frac{\overline{u_i' u_j'} X_{ij}}{2X_{ij}X_{ij}} \tag{7}$$

$$X_{ij} = S_{ij}^* - c_{cr1}(O_{ik}S_{jk}^* + O_{jk}S_{ik}^*) - c_{cr2}\sqrt{2W_{mn}W_{mn}}\delta_{ij} \tag{8}$$

Note that Eq. (7) becomes identical to Eq. (6) if $c_{cr1} = c_{cr2} = 0$.

When solving ν_t using Eq. (6) and (7), additional limiters may apply:

- Due to the deficiency of the linear/quadratic constitutive relation in predicting the transport of each Reynolds stress component, a negative ν_t value can be obtained. This not only violates the non-negative assumption used by the notion of eddy viscosity but also is disfavored from the perspective of numerical stability. Thus, the calculated ν_t value is truncated as $\max\{\nu_t, 0\}$.
- The denominator may be effectively small, resulting in an unrealistic high value of ν_t. This can be solved by limiting the absolute value of the denominator to be no less than a threshold S_{ref}^2.

The Boussinesq-based approach and the QCR-based approach with different S_{ref} values are investigated in detail in Appendix. It is found that the QCR-based approach with $S_{ref} = 10^3 s^{-1}$ keeps the near-wall regions of interest uncontaminated from the limiters whereas reducing the non-physical high ν_t/ν value above the shear layer. Thus, they are adopted in this work. To improve the quality of training data further, samples that are affected by the above limiters are excluded from the database used by ML.

Based on the above method, the normalized eddy viscosity difference between the LES and SA-QCR results at $M_\infty = 0.875$ are obtained, and they are presented as contours in Fig. 7(a). To help correlate the eddy viscosity difference with the flow field, contours of the local Mach number are also shown in Fig. 7(b). It shows that the SA-QCR model over-predicts eddy viscosity in the separation bubble but under-predicts it in the pre-shock BL, the shear layer, the reattached BL and the shock front. Among these regions, the eddy viscosity in the pre-shock BL determines the location of the shock-induced separation point; the eddy viscosity in the separation bubble and the shear layer determines the location of the reattachment point; the eddy viscosity near the shock front, however, are generated due to the shock buffet, which is not an effect of turbulence and thus is conceptually impossible to be modeled by steady RANS simulations.

2.8. Machine learning model: Random forest

In this work, the random forest (RF) algorithm (Breiman, 2001) is used to create a functional mapping between the input and the output, i.e., $\Delta\nu_{t,norm} = R(q_i)$ where $R(\cdot)$ denotes the RF model and q_i represents the input feature. An RF regressor is an ensemble of decision trees (DT), which takes the average prediction of all DTs as its output. An illustrative diagram of a decision tree is shown in Fig. 8, which produces its prediction by going through layers of if-then-else decision rules based on a subset of inputs. During the training of the RF model, the DT is allowed to grow to full depth without pruning to maximize its capability; the training data are weighted by the control volume so that the trained RF model is independent of the grid topology. The RF algorithm from the Python library Scikit-learn (Pedregosa et al., 2011) is adopted.

To quantitatively measure the global accuracy of the RF model, the coefficient of determination R^2, the $L1$-type error ϵ_{L1} and the $L2$-type error ϵ_{L2} are used, as defined in the following:

$$R^2 = 1 - \frac{\left\| y_t - y_p \right\|_F}{\left\| y_t - \bar{y}_t \right\|_F} \tag{9}$$

$$\epsilon_{L1} = \frac{\left| y_t - y_p \right|}{\left| y_t \right|} \tag{10}$$

$$\epsilon_{L2} = \frac{\left\| y_t - y_p \right\|_F}{\left\| y_t \right\|_F} \tag{11}$$

where y_t and y_p are the target value and the RF predicted value of the output, $\bar{(\cdot)}$ is the average of a quantity, $\|\cdot\|_F$ is the Frobenius norm and $|\cdot|$ is the $L1$ norm. A good accuracy is indicated by a high R^2 value and low ϵ_{L1} and ϵ_{L2} values.

3. Methods to improve explainability of machine learning model

The methods to improve the explainability of an ML model can be classified into two types: intrinsic methods and post-hoc methods (Molnar, 2021). The intrinsic methods improve the model explainability by restricting the complexity of the ML model. The rationale is that a simpler model is easier to understand for human users, and thus a higher level of explainability. An extreme example is to downgrade an RF to a DT, which is self-explained as illustrated in Fig. 8. However, a trade-off between accuracy and complexity often exists—a predictive ML model can be simplified by the intrinsic method to some extent, but it does not necessarily become self-explanatory. Thus, a post-hoc method is needed for further explanations. The post-hoc methods use model-agnostic tools to analyze the ML model behavior. They can help identify the causal link between the inputs and the outputs and the interaction among different inputs.

3.1. Intrinsic method

The time complexity to construct an RF model is estimated in Eq. (12) (Louppe, 2014):

$$\mathcal{O}(T N P \log_2 P) \tag{12}$$

where T is the number of trees, N is the maximum number of input features considered by each DT, and P is the number of samples. By reducing the time complexity, a simpler and thus more explainable RF model can be achieved. The $P\log_2 P$ term only depends on data and thus cannot be tuned. The rest term can be further divided into the TN/M term and the M term, where M is the total number of input features and N/M denotes the fraction of input features fed into each DT. Reducing T and N/M can be achieved by a hyperparameter study, which is a problem of ML and is independent of flow physics. Reducing M requires an input feature selection study, which involves knowledge of fluid mechanics. In this work, the hyperparameter study and the input feature study will be performed later in Sections 4.1 and 4.2, respectively.

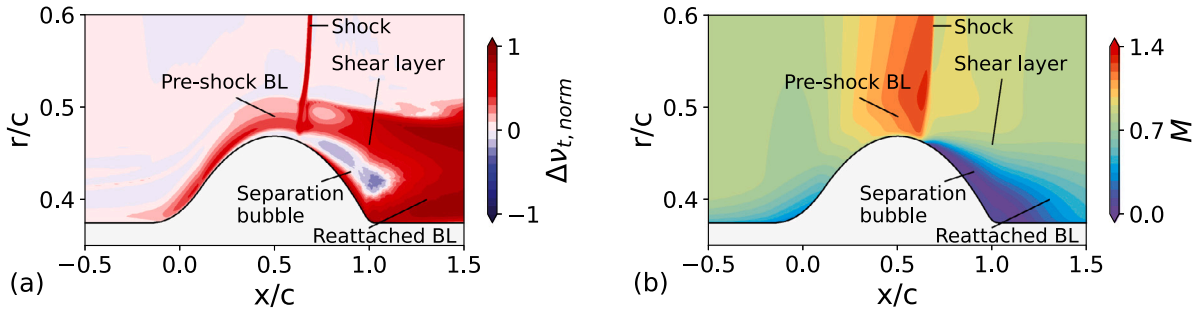


Fig. 7. Contours of (a) normalized eddy viscosity difference (LES minus SA-QCR) and (b) Mach number (LES result) at $M_\infty = 0.875$.

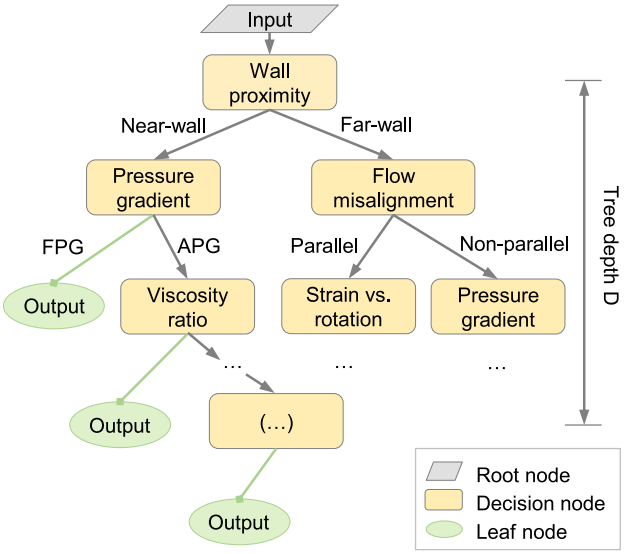


Fig. 8. Illustration of a decision tree structure.

3.2. Post-hoc method

The model-agnostic explanation method named Shapley additive explanations (SHAP) (Lundberg and Lee, 2017) is used to explain the trained RF model in this work. SHAP is based on the Shapley value (Shapley, 1953) from game theory, which assigns payouts to players depending on their contributions to a game. In the context of this work, the "game" is the prediction on $\Delta\nu_{t,norm}$; the "players" are the input features discussed in Section 2.6; the "payouts" are the contributions of each input feature to the predicted $\Delta\nu_{t,norm}$. Thus, the predicted $\Delta\nu_{t,norm}$ can be written as the sum of contributions from each feature, as denoted by Eq. (13):

$$\Delta\nu_{t,norm}(\boldsymbol{x}) = R(q_i(\boldsymbol{x})) = \phi_0 + \sum_{i=1}^{M} \phi_i(\boldsymbol{x}) \tag{13}$$

where $\boldsymbol{x}$ is the spatial coordinates, ϕ_i is the contribution term from q_i and is referred to as the SHAP value, and $\phi_0 = E(R(q_i))$ is the expectation of the RF prediction over ranges of all input features.

To calculate the SHAP value ϕ_i, contributions from q_i itself and all the possible coalitions between q_i and other features need to be considered. This is achieved by Eq. (14):

$$\phi_i = \sum_{S\subseteq Q\setminus\{q_i\}} \frac{|S|!(|Q|-|S|-1)!}{|Q|!}\delta_i(S) \tag{14}$$

$$\delta_i(S) = R_x(S\cup\{q_i\}) - R_x(S) \tag{15}$$

$$R_x(S) = E(R(q_k)|S) \tag{16}$$

where the sum extends over all subsets S of the full set of input features Q excluding $\{q_i\}$; the factorial term is the weight based on probability, in which $|\cdot|$ denotes the number of elements in a set, e.g., $|Q| = M$; $\delta_i(S)$ represents the marginal contribution of $\{q_i\}$ when added to the subset S; $E(R(q_k)|S)$ is the conditional expectation of the RF prediction, which can be efficiently calculated from the TreeSHAP algorithm (Lundberg et al., 2020).

The SHAP value ϕ_i provides global and local sensitivities for the RF model. The global significance of each input feature is ranked by the mean SHAP value $|\phi_i|$. By visualizing the ϕ_i field in the flow domain, it can be identified that whether q_i will increase or reduce the eddy viscosity (i.e., $\phi_i > 0$ or < 0) in a specific region of interest.

Following the concept of the SHAP value, the interaction between q_i and q_j $(i \neq j)$ can be quantified by the SHAP interaction value $\phi_{i,j}$ (Fujimoto et al., 2006) defined by Eq. (17):

$$\phi_{i,j} = \sum_{S\subseteq Q\setminus\{q_i,q_j\}} \frac{1}{2}\frac{|S|!(|Q|-|S|-2)!}{(|Q|-1)!}\delta_{i,j}(S)$$
$$\delta_{i,j}(S) = R_x(S\cup\{q_i,q_j\}) - R_x(S\cup\{q_i\}) - R_x(S\cup\{q_j\}) + R_x(S) \tag{17}$$

where the factorial term is the weight based on probability; $\delta_{i,j}(S)$ represents the marginal contribution of $\{q_i, q_j\}$ when added to the subset S; the total interaction effect is split equally to each feature, i.e., $\phi_{i,j} = \phi_{j,i}$, and thus the weight of $1/2$. By subtracting all the interaction effects from the SHAP value, the pure effect of q_i can be defined by Eq. (18):

$$\phi_{i,i} = \phi_i - \sum_{j\neq i}^{M} \phi_{i,j} \tag{18}$$

The SHAP interaction value $\phi_{i,j}$ provides global and local explanations for the interaction effects between different input features. The mean SHAP interaction value $|\phi_{i,j}|$ quantifies the global significance of the interaction effect. Local interaction effects can be visualized by plotting the $\phi_{i,j}$ field in the flow domain. Interactions among three (or more) input features $\phi_{i,j,k}$ can be derived in a similar fashion to Eqs. (14) and (17), but this is not pursued in this work.

In this work, the SHAP analysis and the SHAP interaction analysis will be performed later in Sections 5.1 and 5.2, respectively. These analyses are conducted using the open-source SHAP library in Python.[2]

4. Improving explainability by intrinsic method

In this section, the intrinsic methods including the hyperparameter study and the input feature selection are demonstrated. The Hi-Fi data used in this section are obtained by the SST-QCR model, as justified in Section 2.5. Data at $M_\infty = 0.6, 0.8, 0.85, 0.875$ conditions form the cross-validation (CV) set, which is used for tuning the hyperparameters

[2] Package retrieved from: https://github.com/slundberg/shap. [Accessed in May 2021].

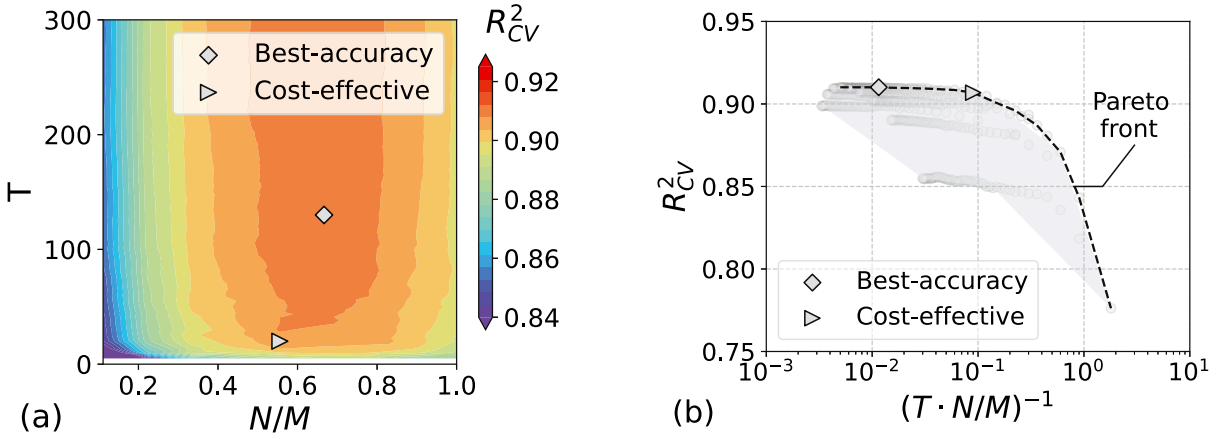


Fig. 9. Hyperparameter study: (a) contours of mean R^2 scores; (b) mean R^2 scores versus RF model complexity.

and selection of the input features. Afterward, two test scenarios are performed: the interpolation scenario takes data at $M_\infty = 0.9$ as the test set, bounded by the remaining conditions as the training set; the extrapolation scenario takes data at $M_\infty = 0.925$ as the test set, which is outside the remaining conditions used for training.

4.1. Hyperparameter study

The hyperparameters of interest in the RF model include the number of trees T and the fraction of the input features fed to each tree N/M. In searching for an optimal set of T and N/M, a grid search is conducted in the range of $0 < T \leq 300$ and $1/9 \leq N/M \leq 9/9$. During the hyperparameter study, the RF model uses all input features listed in Table 2 (i.e., $M = 9$) and all data in the CV set. Its performance with each set of hyperparameters is quantified by the mean R^2_{CV} score, which is calculated from the k-fold cross-validation with $k = 5$ considering the bias–variance trade-off (James et al., 2013).

In Fig. 9(a), contours of the R^2_{CV} score in variation of T and N/M are presented. In general, it shows that a complex RF model with large values of T and N/M is not necessarily predictive. Regarding T, the R^2_{CV} score increases rapidly with T when $T \leq 20$; however, increasing T further shows relatively small increment in the R^2_{CV} score. Regarding N/M, the R^2_{CV} score increases rapidly with N/M when $N/M \leq 4/9$; when $5/9 \leq N/M \leq 7/9$, a high R^2_{CV} score can be achieved given a sufficiently large value of T; increasing N/M beyond $8/9$ reduces the R^2_{CV} score. The best accuracy is found at $T = 130$ and $N/M = 6/9$, as represented by the diamond shape. However, hyperparameter values in the ranges of $5/9 \leq N/M \leq 7/9$ and $T \geq 20$ can also lead to an RF model with good accuracy. A cost-effect hyperparameter set within this range is identified at $T = 20$ and $N/M = 5/9$, as represented by the triangle shape.

In Fig. 9(b), the R^2_{CV} score is plotted against the inverse of the RF model complexity $(T \cdot N/M)^{-1}$, where the top-right corner represents an RF model with high accuracy and explainability. The data obtained in the grid search are represented by circles enveloped by a shaded area, and the top-right boundary of the shaded area represents the Pareto front. Any hyperparameter set on the Pareto front can be taken as the optimal set, depending on whether accuracy or explainability is of priority. In particular, the best-accuracy hyperparameter set represented by the diamond shape has $R^2_{CV} = 0.91011$ and $T \cdot N/M = 86.67$; the cost-effective hyperparameter set represented by the triangle shape has $R^2_{CV} = 0.90682$ and $T \cdot N/M = 11.11$. The complexity of the cost-effective set is about 1/8 of the best-accuracy set, but the reduction in the R^2_{CV} score is only 0.004.

4.2. Input feature selection

Within the input feature group (A) and (B) in Table 2, there is more than one feature sharing similar flow physics. To help find out the most informative feature(s) within the same group, a feature selection study is performed by using an additive method and a subtractive method. The additive method measures the performance improvement of the RF model by adding a feature to the input; here, the input refers to a subset of all the input features excluding the investigated feature group of interest. The subtractive method measures the performance deterioration of the RF model by removing a feature from the input; here, the input is the full set of input features. During the input feature selection, all data of the CV set and the best-accuracy hyperparameters are used by the RF model, and the mean R^2_{CV} score of the k-fold cross-validation procedure ($k = 5$) (James et al., 2013) is used to measure the RF performance.

The feature selection results of the group (A) is presented in Fig. 10(a). It shows that q_2 and q_3 are the most significant features, as adding either feature substantially increases the accuracy meanwhile subtracting either feature reduces the accuracy. On the contrary, q_1 seems redundant as subtracting q_1 does not reduce the RF accuracy but rather increases it. To visualize the flow regions identified by each feature, contours of q_1, q_2 and q_3 at two representative conditions $M_\infty = 0.6$ and 0.9 are shown in Fig. 10(b). It is observed that q_1 and q_2 identify similar flow regions: the APG and the FPG regions in the attached flow of $M_\infty = 0.6$; the shear layer and the separation bubble in the shock-induced separation flow of $M_\infty = 0.9$. However, q_2 is more informative, because it not only informs the direction of pressure gradient (as q_1 does) but the pressure gradient magnitude. Thus, q_2 is selected but q_1 is dropped. On the other hand, q_3 identified flow regions different from q_1 and q_2 because the flow physics behind these features are different: the former measures the total pressure gradient magnitude, but the latter measures the streamwise component of the pressure gradient. Therefore, q_3 is also selected.

The feature selection results of the group (B) is presented in Fig. 11(a). It illustrates that q_4, q_5 and q_6 have almost the same level of significance: including either one of the features to the subset of input feature, or removing either one of the features from the full set of input features, has almost the same RF accuracy. From the contour plots of q_4, q_5 and q_6 shown in Fig. 11(b), all the three features identify similar flow regions including the pre-shock BL near the curved surface and the shear layer downstream of the shock. Therefore, the group (B) features carry the same flow physics, although they were originally proposed for turbulence modeling by different authors. Either one of the group (B) features can be selected and the rest features can be dropped. In

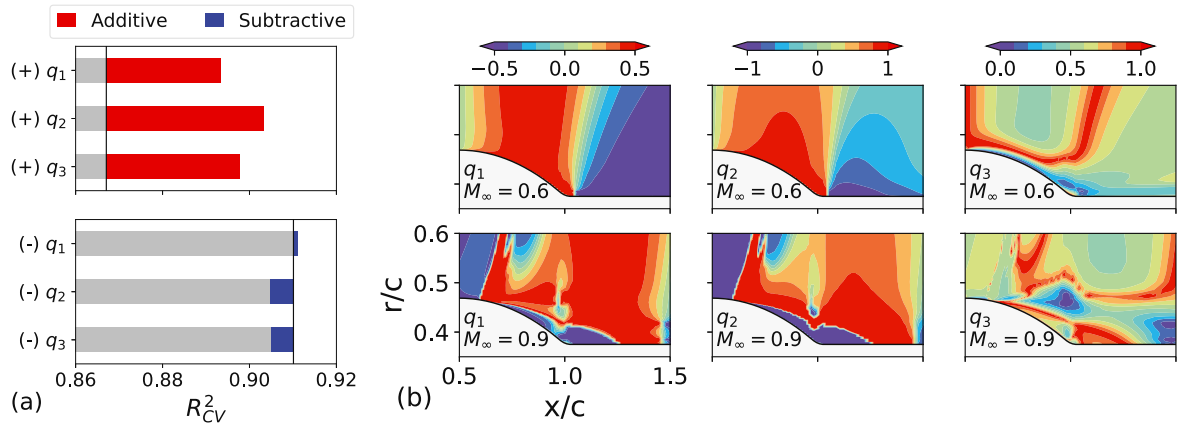


Fig. 10. Feature selection within the group (A) pressure gradient: (a) mean R^2_{CV} scores obtained with different feature subsets (the black vertical line represents the performance without q_1, q_2, and q_3 for the additive test and the performance with q_1, q_2, and q_3 for the subtractive test); (b) contours of normalized input features.

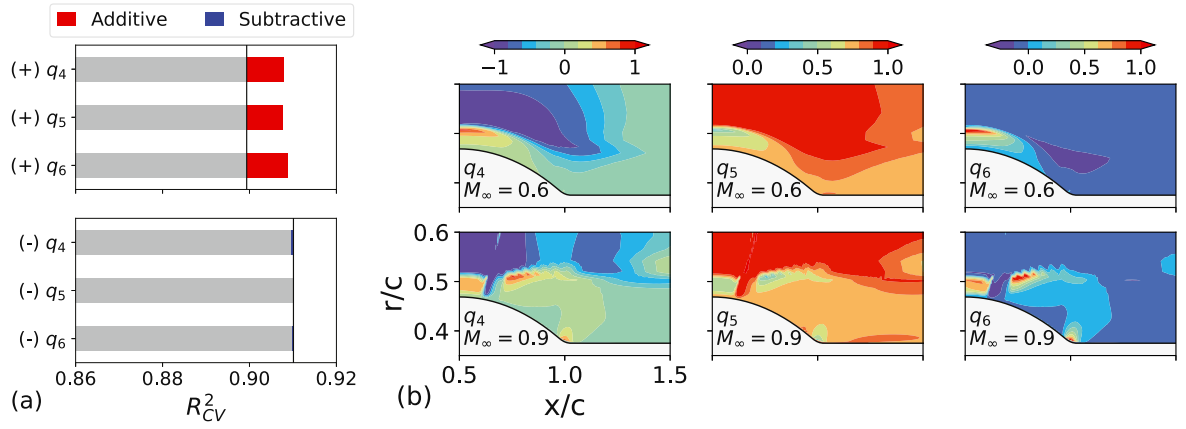


Fig. 11. Feature selection within the group (B) strain versus rotation: (a) mean R^2_{CV} scores obtained with different feature subsets (the black vertical line represents the performance without q_4, q_5, and q_6 for the additive test and the performance with q_4, q_5, and q_6 for the subtractive test); (b) contours of normalized input features.

this work, only q_4 based on the Q-criterion is selected due to its better dimensionless form.

4.3. Comparison of prediction accuracy

To quantify the effect of different hyperparameter sets and input features on the RF model accuracy, the RF model is tested in an interpolation and an extrapolation scenario where the test M_∞ is 0.9 and 0.925 respectively. Table 4 summarizes the prediction errors for different cases. In general, ϵ_{L2} for extrapolation scenario is 3%–4% higher than that of the interpolation scenario, which is similar to that reported in subsonic separated flows (Xiao et al., 2020). Regarding the hyperparameters, ϵ_{L2} of the cost-effective hyperparameter set is only 0.3–1.4% higher than the best-accuracy set. Considering the cost-effective set reduces the RF model complexity to 1/8 of the best-accuracy set, this loss of accuracy is acceptable. Regarding the input features, using the selected features instead of the full features increases ϵ_{L2} by 0.2–1.1% but reduces the RF model complexity further to 2/3.

For a direct comparison between the SA-QCR-ML results with the target SST-QCR results, the viscosity ratio profiles are presented in Fig. 12. In these figures, results from the SA-QCR-ML model using the best-accuracy hyperparameter set and the selected input features are represented by the circles; results obtained from all the four settings in Table 4 are enveloped by the error bars. For the interpolation scenario in Fig. 12(a), the SA-QCR model predicts the peak of the viscosity ratio profile at a higher radius than the SST-QCR model. The SA-QCR-ML model effectively reduces such differences and matches well with the SST-QCR model at all sections. For the extrapolation scenario in Fig. 12(b), similar deficiency of the SA-QCR model is observed, and it is alleviated from the SA-QCR-ML results. In addition, it is found that using different RF settings in Table 4 predicts effectively the same viscosity ratio. A simple and thus more explainable ML model is sufficient to reduce the turbulence model deficiency in the current case.

5. Improving explainability by post-hoc method

In this section, the post-hoc method SHAP (Lundberg and Lee, 2017) is demonstrated. The LES data at $M_\infty = 0.875$ from Uzun and Malik (2019) is adopted to train the RF model with the best-accuracy hyperparameter set and the selected input features. The task of the SHAP method is to unveil the causal link between the input features and the output eddy viscosity difference of the trained RF model.

5.1. SHAP analysis

The SHAP value ϕ_i, as introduced in Section 3.2, quantifies the contribution of input feature q_i to the RF output, i.e., $R(q_i(\boldsymbol{x})) = \sum_{i=0}^{M} \phi_i(\boldsymbol{x})$. The SHAP values of the trained RF model for turbulence modeling are calculated, and the results are summarized in Fig. 13.

Table 4
Effect of input features of the random forest model on the prediction accuracy.

Random forest model					Interpolation			Extrapolation		
Hyperparameter	Input feature	M	N	T	R^2	ϵ_{L1}	ϵ_{L2}	R^2	ϵ_{L1}	ϵ_{L2}
Best-accuracy	Full features	9	6	130	0.98682	7.0%	11.0%	0.97632	10.8%	14.7%
Best-accuracy	Selected features	6	4	130	0.98601	7.2%	11.3%	0.97551	10.9%	14.9%
Cost-effective	Full features	9	5	20	0.98610	7.4%	11.3%	0.97393	11.6%	15.4%
Cost-effective	Selected features	6	3	20	0.98325	7.9%	12.4%	0.97057	11.9%	16.3%

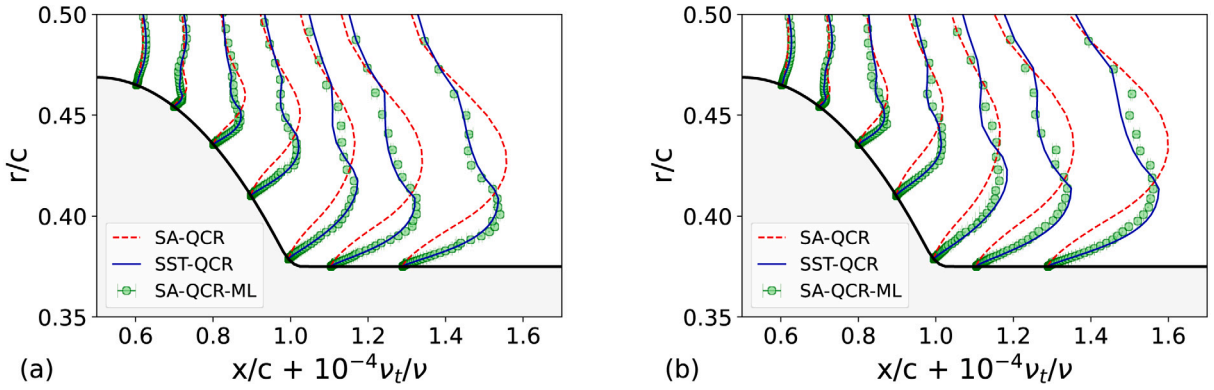


Fig. 12. Viscosity ratio profiles: (a) $M_\infty = 0.9$ (interpolation) and (b) $M_\infty = 0.925$ (extrapolation).

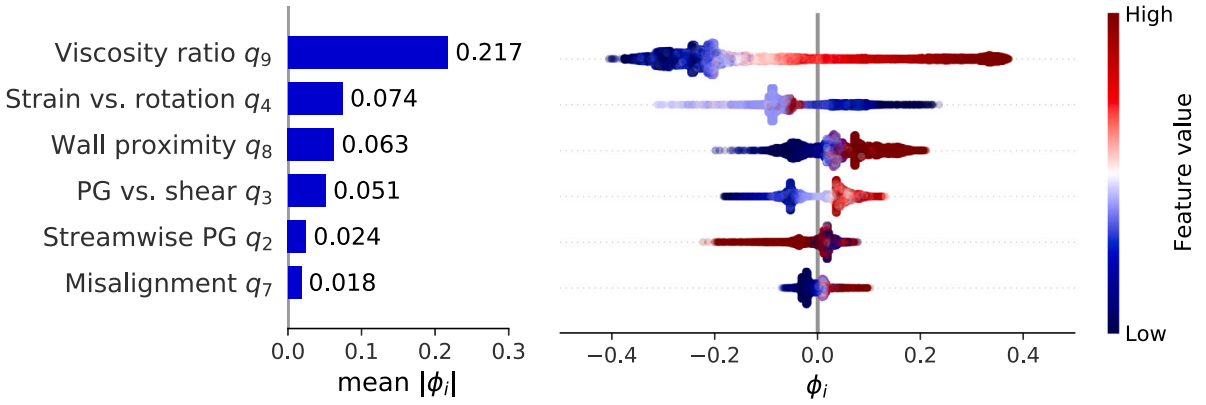


Fig. 13. Summary of SHAP analysis. Left: bar chart of the mean absolute SHAP value. Right: bee swarm plots of SHAP values. Each dot corresponds to a CFD grid point, with the horizontal location representing the SHAP value ϕ_i (impact of feature q_i on the RF prediction) and the color representing the feature value q_i at that grid point. When multiple points locate at the same horizontal position, they pile up in the vertical direction to show the occurrence probability density.

In the bar chart on the left of Fig. 13, the input features are ranked by their mean absolute SHAP values, which represent the global significance of each input feature to the output eddy viscosity difference. It shows that the contribution of viscosity ratio q_9 is significantly higher than the rest features, which was also observed in a previous research (Wang et al., 2017). This is because the RF output is essentially an extra production or destruction term in the transport equation of the SA model, which scales with ν_t or ν_t^2 as introduced earlier in Section 2.6. The rest of the features in descending order of significance are Q-criterion-based strain versus rotation q_4, wall proximity q_8, pressure normal stress versus normal shear stress q_3, streamwise pressure gradient magnitude q_2 and flow misalignment q_7.

In the bee swarm plot on the right of Fig. 13, the statistical distributions of the SHAP value of each input feature are presented. A general proportional relation between the SHAP value ϕ_i and the feature values q_i is found for the following features:

- $\phi_9 \propto q_9$, i.e., the eddy viscosity increases with the viscosity ratio. The rationale of q_9 is the scaling between $\Delta\nu_t$ and ν_t from dimensional analysis, as discussed in Section 2.6.
- $\phi_4 \propto -q_4$, i.e., the eddy viscosity is increased when the shear effect dominates over the rotation effect, and vice versa. Such an effect of q_4 has long been identified in the source term of Reynolds-stress models, as discussed in Section 2.6.
- $\phi_8 \propto q_8$, i.e., the eddy viscosity is increased near the identified BL log-law region and generally decreased in other regions. Such an observation indicates that the equilibrium turbulence assumption used by the original SA model fails in the transonic bump flow, which is featured by streamline curvature and strong pressure gradient.
- $\phi_3 \propto q_3$, i.e., the eddy viscosity is increased when the pressure normal stress dominates over the shear stress, and vice versa. The rationale of q_3 has not yet been reported and needs to be verified for future research.
- $\phi_7 \propto q_7$, i.e., the eddy viscosity is increased when the flow deviates from parallel shear flows. Note however this proportional relation is not as pronounced as the others because $|\phi_7|$ is very small.

The only exception is the streamwise pressure gradient magnitude q_2, which has a non-monotonic relationship with the SHAP value ϕ_2. It indicates that q_2 alone is incapable of determining the eddy viscosity difference, and interaction effects with other features may exist.

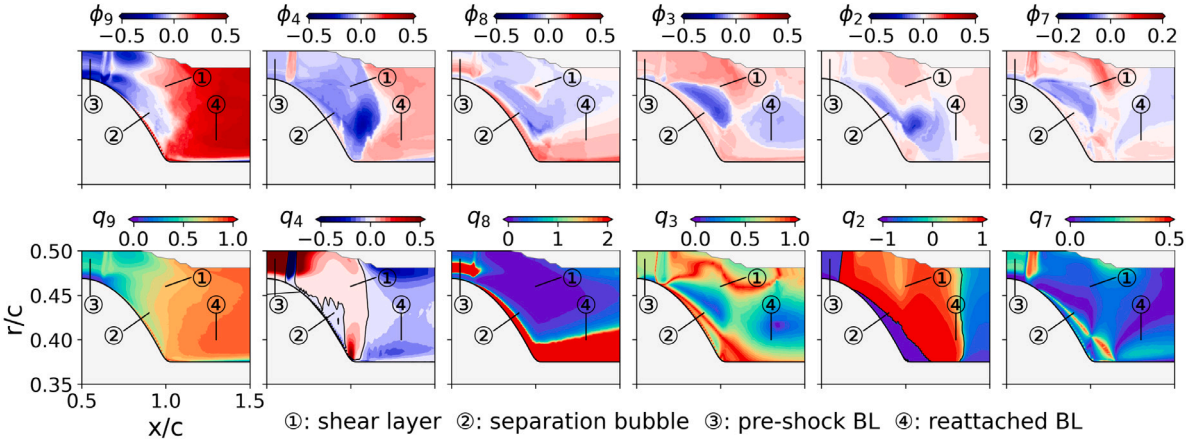


Fig. 14. Contours of SHAP values (top row) and input feature values (bottom row). Notations of flow regions are consistent with Fig. 7. The isoline of $q_i = 0$ is represented by a solid curve.

To correlate the SHAP values with the input feature values, contours of the SHAP values and the input feature values are presented in Fig. 14. In this figure, the top row and the bottom row represent the SHAP value and the input feature value, respectively; each column corresponds to an input feature, and it is arranged in descending order of feature significance from left to right.

Based on the observations of the input feature values and the SHAP values, the causal link between the input and the output is interpreted as follows:

- The viscosity ratio q_9 highlights regions with high eddy viscosity (e.g., the shear layer and the reattached BL) and increases the local eddy viscosity further ($\phi_9 > 0$). The change of eddy viscosity ϕ_9 is qualitatively proportional to the feature value q_9, which again shows the scaling between $\Delta\nu_t$ and ν_t.
- The dimensionless Q-criterion q_4 identifies the pre-shock BL and the shear layer by $q_4 > 0$, where the local rotation rate surpasses the strain rate. The eddy viscosity in these regions is decreased ($\phi_4 < 0$). On the other end, q_4 also identifies the reattached BL and the separation bubble by $q_4 < 0$, where the eddy viscosity in these regions is increased ($\phi_4 > 0$). It again shows the correlation that a strong shear enhances turbulence intensity whereas a strong rotation reduces it. Note however the shear layer and the separation bubble are almost indistinguishable by q_4. Additional features are needed to distinguish these regions.
- The wall proximity q_8 identifies the near-wall log-law region (e.g., pre-shock BL and reattached BL) by a large q_8 value, and the eddy viscosity in these regions is increased ($\phi_8 > 0$). For far-wall regions with small q_8 (e.g., shear layer and separation bubble), the eddy viscosity can either be increased or decreased, indicating an interaction effect between q_8 and other features.
- The pressure normal stress versus the normal shear stress q_3 identifies the shear layer by large q_3, where the local pressure normal stress dominates over the normal shear stress. The eddy viscosity in this region is increased ($\phi_3 > 0$). The separation bubble and the reattached BL are identified by small q_3, indicating the local normal shear stress dominates over the pressure normal stress. The eddy viscosity in these regions is decreased ($\phi_3 < 0$).
- The streamwise pressure gradient magnitude q_2 identifies the shear layer, the upper half of the separation bubble (with forwarding flow) and the front part of the reattached BL as regions experiencing an APG ($q_2 > 0$), whereas the lower half of the separation bubble (reverse flow) and the rear part of the reattached BL are experiencing an FPG ($q_2 < 0$). However, the relation between q_2 and ϕ_2 is non-monotonic, which is consistent with the observation in Fig. 13.
- The flow misalignment q_7 identifies the pre-shock BL, the lower half of the separation bubble and the reattached BL by $q_7 > 0$, where the local flow deviates from the parallel flow; the upper half of the separation bubble and the shear layer are identified by $q_7 \to 0$, where the shear vanishes or the local flow is parallel. Although the effect of q_7 on the eddy viscosity field is comparatively small, a positive correlation between q_7 and ϕ_7 is still visible, which agrees with the observation in Fig. 13.

To sum up the SHAP analysis, it provides a ranked list of input features according to their contribution to the output eddy viscosity difference. Correlations between q_i and ϕ_i are obtained qualitatively, but the pure effect of q_i and the interaction between q_i and other features remain unclear. These further details need to be obtained via SHAP interaction analysis.

5.2. SHAP interaction analysis

The SHAP value ϕ_i of input feature q_i can be further divided into a pure effect term $\phi_{i,i}$ and interaction terms $\phi_{i,j}$, i.e., $\phi_i(\boldsymbol{x}) = \sum_{j=1}^{M} \phi_{i,j}(\boldsymbol{x})$, as introduced in Section 3.2. The SHAP interaction values of the trained RF model for turbulence modeling are summarized in Fig. 15.

In Fig. 15(a), the mean absolute SHAP interaction values $\overline{|\phi_{i,j}|}$ are presented in a symmetric square matrix (note $\phi_{i,j} = \phi_{j,i}$), with the diagonal elements denoting the pure effect terms and the off-diagonal elements denoting the interaction effect terms; the rows and columns are ranked in descending order of the pure effect terms $\overline{|\phi_{i,i}|}$; the color and the number in each matrix element represent the corresponding value of $\overline{|\phi_{i,j}|}$, which measures the global significance of the pure/interaction effects. Results show that the pure effect of each feature is generally stronger than its interaction effects with other features. The ranked list of input features by $|\phi_{i,i}|$ is almost the same as that by $|\phi_i|$. The only exception is the wall proximity q_8 which has dropped one place. It implies that q_8 relies more on the interaction effects than other features, which is expected because q_8 serves better as a flow region classifier rather than a regressor.

In Fig. 15(b), the mean absolute SHAP interaction values are normalized by the corresponding pure effect term, i.e., $\overline{|\phi_{i,j}|}/\overline{|\phi_{i,i}|}$, which shows the relative importance of the interaction effect better. Among the interaction effects, the interaction terms $\phi_{9,4}$, $\phi_{8,4}$, $\phi_{9,3}$ and $\phi_{8,3}$ are pronounced and need further examination. Interaction terms of q_2 are also significant as expected, since the correlation between q_2 and ϕ_2 is non-monotonic. However, they are not examined in detail due to the low global significance of q_2.

To help identify the correlation between the feature value q_i and its pure effect on the eddy viscosity difference $\phi_{i,i}$, $\phi_{i,i}$ are plotted against q_i in Fig. 16. In these plots, each scatter point corresponds to a CFD grid

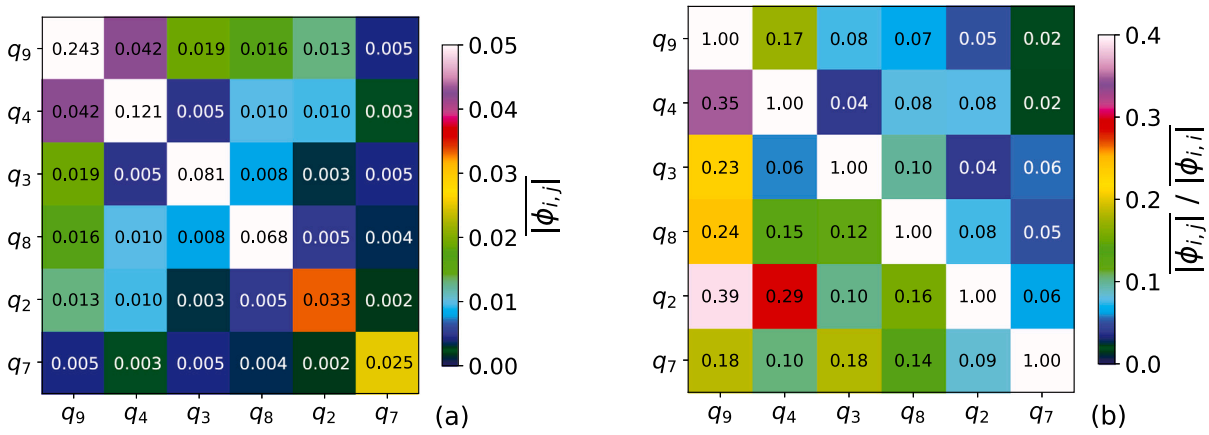


Fig. 15. Heatmap of (a) mean absolute SHAP interaction values and (b) normalized mean absolute SHAP interaction values.

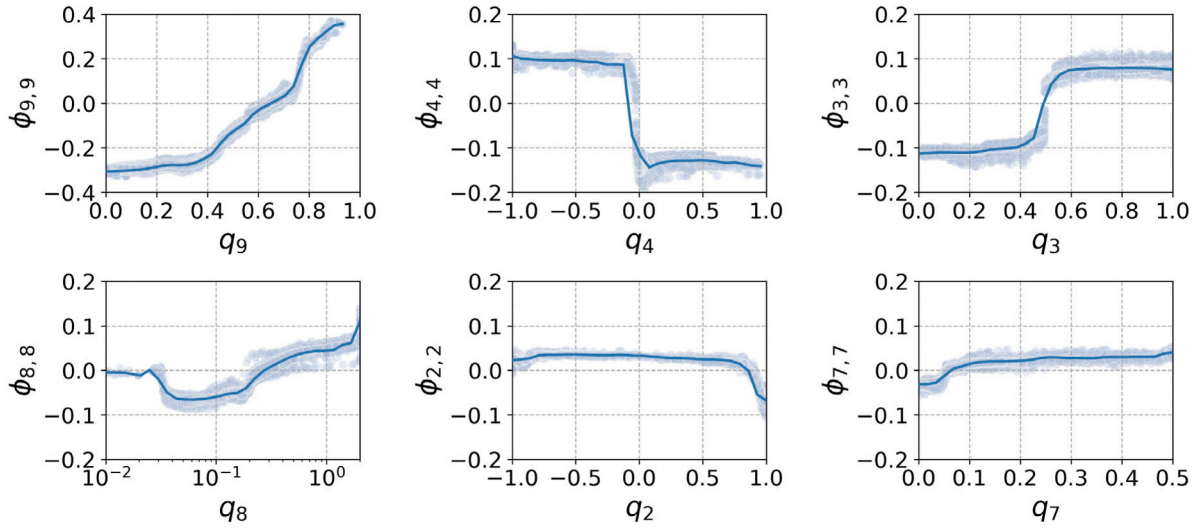


Fig. 16. Dependency plot of feature values versus diagonal SHAP interaction values.

point, whose horizontal and vertical coordinates represent the feature value q_i and the SHAP interaction value $\phi_{i,i}$ respectively; a solid curve is fitted through all the scatter points to indicate the general relation between q_i and $\phi_{i,i}$. These plots clearly show the monotonic relations between some of the q_i and $\phi_{i,i}$, i.e., $\phi_{9,9} \propto q_9$, $\phi_{4,4} \propto -q_4$, $\phi_{3,3} \propto q_3$ and $\phi_{8,8} \propto q_8$ when ($q_8 > 0.1$), which are consistent with the observations in Fig. 13. In particular:

- The viscosity ratio q_9 scales with its pure effect on the eddy viscosity difference $\phi_{9,9}$.
- The dimensionless Q-criterion q_4 reduces the eddy viscosity ($\phi_{4,4} < 0$) when the rotation rate surpasses the strain rate ($q_4 > 0$), and vice versa.
- The pressure normal stress versus normal shear stress q_3 increases the eddy viscosity ($\phi_{3,3} < 0$) when the pressure normal stress surpasses the normal shear stress ($q_3 > 0.5$), and vice versa.
- The wall proximity q_8 reduces the eddy viscosity ($\phi_{8,8} < 0$) in free shear flow and the viscous sublayer ($0 < q_8 < 0.3$), but increases it ($\phi_{8,8} > 0$) in the log-law region ($0.3 < q_8 < 0.7$) and the outer part of the BL ($0.7 < q_8 < 2.0$).

Regarding the rest features, the streamwise pressure gradient magnitude q_2 is sensitive only at strong pressure gradients ($|q_2| \to 1$), and the flow misalignment q_7 is sensitive only when the flow is close to a perfectly parallel flow ($|q_7| \to 0$). Their mean effects $\overline{|\phi_{i,i}|}$ are also relatively small. Thus, they are less informative than the other features in this case.

To help identify the interaction effect between two input features, the interaction effect terms $\phi_{i,j}$ in variation with the feature values q_i and q_j are presented in Fig. 17. In these plots, each scatter corresponds to a CFD grid point, whose horizontal coordinate, vertical coordinate and face color represent the feature value q_i, the SHAP interaction value $\phi_{i,j}$ and the feature value q_j respectively; a solid curve is fitted through the scatters where $q_j > q_{j,mean}$, and a dashed curve is fitted through the rest scatters with $q_j < q_{j,mean}$. Observations of the interaction effects are listed as follows:

- For the viscosity ratio q_9 and the dimensionless Q-criterion q_4, their interaction effect $\phi_{9,4}$ counteracts the pure effect $\phi_{4,4}$ when $q_9 < 0.8$ (i.e., $\nu_t/\nu < 2000$) whereas enhances it in the rest range of q_9.
- For the viscosity ratio q_9 and the pressure normal stress versus normal shear stress q_3, their interaction effect $\phi_{9,3}$ counteracts the pure effect $\phi_{3,3}$ when $q_9 < 0.5$ (i.e., $\nu_t/\nu < 500$) or $q_9 > 0.8$ (i.e., $\nu_t/\nu > 2000$) whereas enhances it when $0.5 < q_9 < 0.8$ (i.e., $500 < \nu_t/\nu < 2000$).
- For the wall proximity q_8 and the dimensionless Q-criterion q_4, their interaction effect $\phi_{8,4}$ counteracts the pure effect $\phi_{4,4}$ when $q_8 < 0.1$ (i.e., free shear flow or viscous sublayer) or $q_8 > 1.6$ (i.e., near-wall BL under APG) whereas enhances it when $0.1 < q_8 < 1.6$.
- For the wall proximity q_8 and the pressure normal stress versus normal shear stress q_3, their interaction effect $\phi_{8,3}$ counteracts the

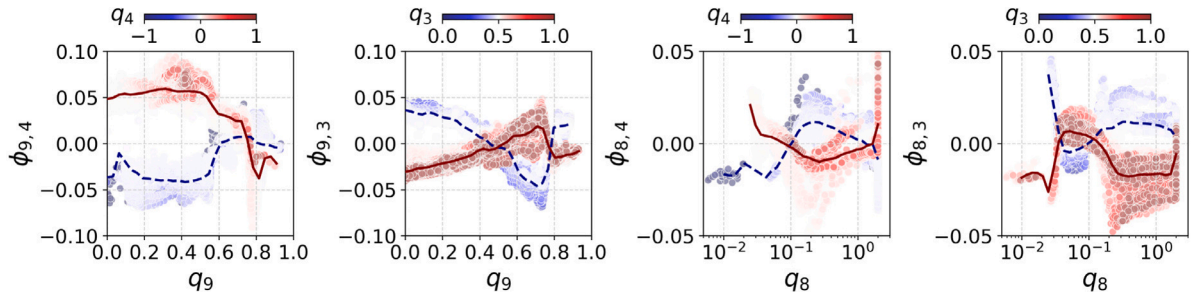


Fig. 17. Dependency plot of feature values versus off-diagonal SHAP interaction values.

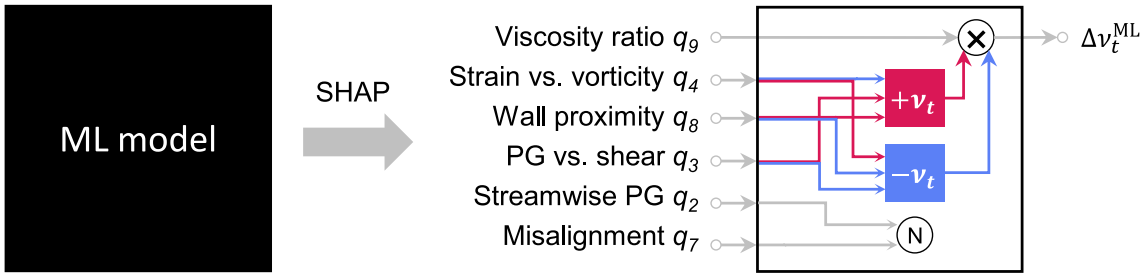


Fig. 18. Illustration of the explained random forest model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pure effect $\phi_{3,3}$ in most range of q_8 except the range of $0.04 < q_8 < 0.1$.

5.3. ML model explained

After the SHAP analysis and the SHAP interaction analysis, the black-box ML model that assists the SA turbulence model becomes partially transparent, as illustrated in Fig. 18. Here, the input features attached to the transparent box are ranked according to their global significance to the output; the arrowed lines in red or blue correspond to large or small values of the input feature. The major findings from the SHAP explanation are summarized as follows:

- The ML model has re-discovered the scaling between the eddy viscosity ν_t and its source term, which has already been found by human turbulence modelers via dimensional analysis.
- The ML model has re-discovered the shear versus rotation effect on the eddy viscosity source term (i.e., shear enhances turbulent fluctuation whereas rotation reduces it), which has already been found in Reynolds stress transport equations.
- The ML model suggests a possible correlation between the eddy viscosity source term and the ratio between the pressure normal stress and the normal shear stress (i.e., turbulent fluctuation is enhanced when the pressure normal stress dominates over the shear stress, and vice versa). Further studies in more flow cases are needed to verify this finding.

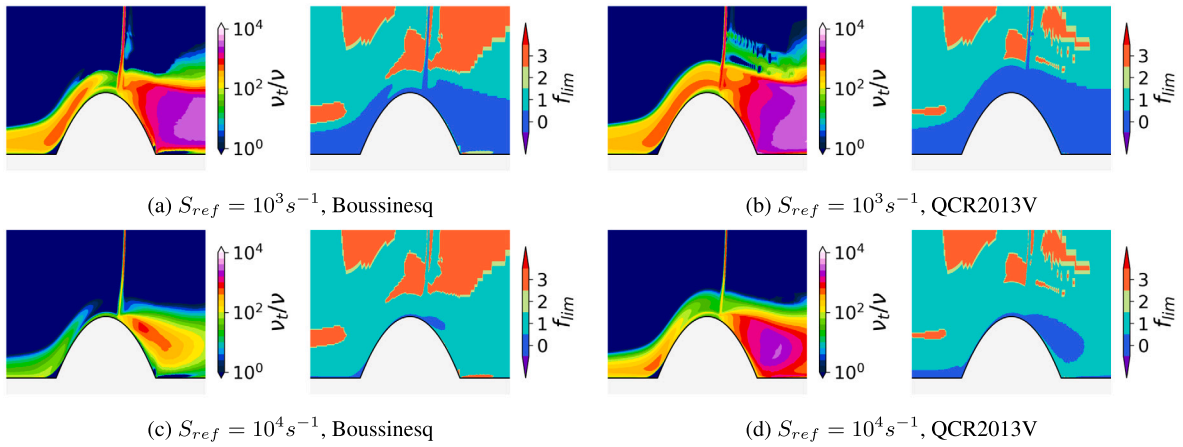
Despite the success of SHAP demonstrated above, limitations of the SHAP method do exist. Firstly, it should be reminded that the SHAP analysis is data-dependent. When training the ML model with a larger database, the input–output causal link identified by SHAP may become more complex than that demonstrated in this work. Secondly, SHAP only describes how the model responds to the input features without offering why. A physical domain explanation still needs to be conducted by human researchers based on the SHAP results. Lastly, since the current ML-assisted turbulence model is based on a physics-informed framework, any physics beyond the range prescribed by the input features cannot be discovered by SHAP. Searching for suitable input features for RANS turbulence modeling should be regarded as the focus of future research.

6. Conclusions

In this paper, two types of methods, i.e., the intrinsic methods and the post-hoc methods, were presented to improve the explainability of an ML-assisted turbulence model. The intrinsic methods improve the explainability by restricting the ML model complexity, which include a physics-free hyperparameter study and a physics-based input feature selection. The post-hoc methods help identify the causal link between the inputs and the outputs of the ML model, among which the state-of-the-art post-hoc method SHAP was applied. The investigated ML-assisted turbulence modeling framework aims to improve the SA model accuracy in transonic bump flows. A random forest model was trained, cross-validated and tested to correct the eddy viscosity predicted by the SA model. Several conclusions can be drawn as follows.

In the hyperparameter study, a Pareto front of model complexity and model accuracy was identified, indicating the trade-off between accuracy and complexity of the ML model. The ML model accuracy does not necessarily increase with the model complexity, and the best-accuracy hyperparameter set does exist. A cost-effective hyperparameter set was also identified, whose $L2$-type error is 1.4% higher and model complexity is 87.5% lower than the best-accuracy set. Such a cost-effective hyperparameter set is advantageous when the model explainability or numerical efficiency is of equal concern as the model accuracy.

In the input feature selection study, it was found that dropping redundant input features can reduce the complexity and thus improve the explainability of the ML model with limited sacrifice in accuracy (e.g., 1.1% higher $L2$-type error and 33.3% lower model complexity in this case). Among the feature group of the pressure gradient, the streamwise pressure gradient magnitude (q_2) and the pressure normal stress versus normal shear stress (q_3) were found most informative, whereas the feature regarding the alignment between the velocity and the pressure gradient (q_1) can be dropped. Among the feature group of strain versus rotation, using either one of the Q-criterion (q_4), the strain-to-vorticity ratio (q_5) or the Richardson number (q_6) achieved the same level of accuracy. These understandings provide guidance for input feature selection for the future development of turbulence models.



(a) $S_{ref} = 10^3 s^{-1}$, Boussinesq (b) $S_{ref} = 10^3 s^{-1}$, QCR2013V

(c) $S_{ref} = 10^4 s^{-1}$, Boussinesq (d) $S_{ref} = 10^4 s^{-1}$, QCR2013V

Fig. 19. Contours of eddy viscosity ratio ν_t/ν and limiter indicator f_{lim}.

By using the SHAP method, the ML model became partially explainable rather than a complete black-box system. The SHAP values provided a global significance rank of the input features, which in descending order reads viscosity ratio (q_9), Q-criterion (q_4), wall proximity (q_8), pressure normal stress versus shear stress (q_3), streamwise pressure gradient magnitude (q_2) and flow misalignment (q_7). In addition, the SHAP values also offered a local explanation of each input feature, e.g., q_4 reduced the eddy viscosity in the pre-shock BL but increases it in the reattached BL, etc. The SHAP interaction values not only showed the pure effects of input features, e.g., $\Delta\nu_t \propto q_9, -q_4, q_8, q_3$, but also unveiled the interaction effects between input features, especially the interaction with q_9 and q_8. These results together showed that the trained SA-QCR-ML model discovered: (1) the scaling between the eddy viscosity and its source term; (2) the effect of shear and rotation on the eddy viscosity source term; (3) the effect of pressure-to-shear normal stress ratio on the eddy viscosity source term.

The methods presented in this work are transferable to a general ML turbulence model and are recommended for future ML turbulence model developments.

CRediT authorship contribution statement

Xiao He: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Jianheng Tan:** Conceptualization, Data curation, Software, Validation, Visualization, Writing – original draft. **Georgios Rigas:** Supervision, Writing – review & editing. **Mehdi Vahdati:** Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xiao He reports financial support was provided by Imperial College London. Xiao He reports a relationship with Imperial College London that includes: employment.

Acknowledgments

Xiao He greatly acknowledges the Imperial College President's PhD Scholarship (ID: 01015425), the Henry Lester Trust and the Great Britain–China Educational Trust for the financial support. The constructive comments from the anonymous reviewers are greatly appreciated.

Appendix. Effect of constitutive relation and S_{ref} on the calculation of ν_t

In this section, the effect of the constitutive relation and the S_{ref} limiter value on the calculated eddy viscosity field is examined in detail. The contours of eddy viscosity ratio ν_t/ν and limiter indicator f_{lim} obtained from different methods are presented in Fig. 19. Here, the ν_t/ν contours are plotted in log-scale; the f_{lim} values of 0, 1, 2 and 3 correspond to regions without any limiter effect, with the S_{ref} limiter effect, with the $\nu_t \geq 0$ limiter effect and with both limiter effect, respectively.

When comparing the constitutive relation effect at the same S_{ref} value, the QCR2013V-based method identified a larger area free of any limiter effect than the Boussinesq-based method. It indicates that the QCR2013V-based method is less sensitive to the limiters and thus more reliable than the Boussinesq-based method.

When comparing the S_{ref} limiter effect at the same constitutive relation, a low S_{ref} leads to a non-physical high ν_t/ν value above the shear layer, meanwhile a high S_{ref} contaminates the ν_t/ν value in the shear layer, the separation bubble and the reattached BL. By trial-and-error, the value of $S_{ref} = 10^3 s^{-1}$ combined with the QCR2013V-based approach reaches a good balance between the near-wall flow and the shear layer flow. Therefore, they are adopted in this work.

The above findings can be reproduced via the open-accessed Python script TurbAna[3], which is partly developed for this work.

References

Bachalo, W.D., Johnson, D.A., 1986. Transonic, turbulent boundary-layer separation generated on an axisymmetric flow model. AIAA J. 24 (3), 437–443.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Dacles-Mariani, J., Zilliac, G.G., Chow, J.S., Bradshaw, P., 1995. Numerical/experimental study of a wingtip vortex in the near field. AIAA J. 33 (9), 1561–1568.
Duraisamy, K., Iaccarino, G., Xiao, H., 2019. Turbulence modeling in the age of data. Annu. Rev. Fluid Mech. 51, 357–377.
Economon, T.D., Palacios, F., Copeland, S.R., Lukaczyk, T.W., Alonso, J.J., 2016. SU2: An open-source suite for multiphysics simulation and design. AIAA J. 54 (3), 828–846.
Edeling, W.N., Cinnella, P., Dwight, R.P., Bijl, H., 2014. Bayesian estimates of parameter variability in the k-ϵ turbulence model. J. Comput. Phys. 258, 73–94.
Fujimoto, K., Kojadinovic, I., Marichal, J.-L., 2006. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. Games Econom. Behav. 55 (1), 72–99.
Gorlé, C., Larsson, J., Emory, M., Iaccarino, G., 2014. The deviation from parallel shear flow as an indicator of linear eddy-viscosity model inaccuracy. Phys. Fluids 26 (5).

[3] Package available from: https://github.com/HexFluid/TurbAna. [Accessed in June 2022].

He, X., Zhao, F., Vahdati, M., 2020. Uncertainty quantification of spalart-allmaras turbulence model coefficients for simplified compressor flow features. ASME J. Fluids Eng. 142 (9), 091501.

He, X., Zhao, F., Vahdati, M., 2021. Uncertainty quantification of spalart-allmaras turbulence model coefficients for compressor stall. ASME J. Turbomach. 143 (8), 081007.

He, X., Zhao, F., Vahdati, M., 2022. A turbo-oriented data-driven modification to the spalart-allmaras turbulence model. In: ASME Paper No. GT2022-80456.

Hellsten, A., 1998. Some improvements in menter's $k-\omega$ SST turbulence model. In: 29th AIAA Fluid Dynamics Conference. AIAA Paper No. 98-2554.

Hunt, J.C.R., Wray, A.A., Moin, P., 1988. Eddies, streams, and convergence zones in turbulent flows. In: Proceedings of Summer Program 1988. Center for Turbulent Research Report CTR-S88, pp. 193–208.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer, New York, USA.

Jameson, A., Schmidt, W., Turkel, E., 1981. Numerical solution of the Euler equations by finite volume methods using Runge Kutta time stepping schemes. In: 14th Fluid and Plasma Dynamics Conference. AIAA Paper No. 81-1259.

Johnson, D.A., 1987. Transonic separated flow predictions with an eddy-viscosity/Reynolds-stress closure model. AIAA J. 25 (2), 252–259.

Kline, S.J., Reynolds, W.C., Schraub, F.A., Runstadler, P.W., 1967. The structure of turbulent boundary layers. J. Fluid Mech. 30 (4), 741–773.

Ling, J., Kurzawski, A., Templeton, J., 2016. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. J. Fluid Mech. 807, 155–166.

Ling, J., Templeton, J., 2015. Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty. Phys. Fluids 27 (8).

Louppe, G., 2014. Understanding Random Forests: From Theory to Practice (Ph.D. thesis). University of Liège.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.

Matai, R., Durbin, P.A., 2019. Zonal eddy viscosity models based on machine learning. Flow Turbul. Combust. 103 (1), 93–109.

Menter, F.R., 1994. Two-equation eddy-viscosity turbulence models for engineering applications. AIAA J. 32 (8), 1598–1605.

Molnar, C., 2021. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Lulu.com, (Accessed March 2021).

Nickels, T.B., 2004. Inner scaling for wall-bounded flows subject to large pressure gradients. J. Fluid Mech. 521, 217–239.

Parish, E.J., Duraisamy, K., 2016. A paradigm for data-driven predictive modeling using field inversion and machine learning. J. Comput. Phys. 305, 758–774.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pope, S.B., 1975. A more general effective-viscosity hypothesis. J. Fluid Mech. 72 (2), 331–340.

Rumsey, C.L., 2020. Turbulence Modeling Resource. NASA Langley Research Center, Hampton, Virginia, USA, (Accessed January 2020).

Rumsey, C.L., Lee, H.C., Pulliam, T.H., 2020. Reynolds-averaged Navier-Stokes computations of the NASA juncture flow model using fun3D and overflow. In: AIAA Paper No. 2020-1304. p. 1304.

Schmelzer, M., Dwight, R.P., Cinnella, P., 2020. Discovery of algebraic Reynolds-stress models using sparse symbolic regression. Flow Turbul. Combust. 104 (2–3), 579–603.

Shapley, L.S., 1953. A value for n-person games. Contrib. Theory Games 2 (28), 307–317.

Shur, M.L., Spalart, P.R., Strelets, M.K., Travin, A.K., 2015. An enhanced version of DES with rapid transition from RANS to LES in separated flows. Flow Turbul. Combust. 95 (4), 709–737.

Singh, A.P., Duraisamy, K., 2016. Using field inversion to quantify functional errors in turbulence closures. Phys. Fluids 28 (4).

Singh, A.P., Medida, S., Duraisamy, K., 2017. Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. AIAA J. 55 (7), 2215–2227.

Slotnick, J., Khodadoust, A., Alonso, J., Darmofal, D., Gropp, W., Lurie, E., Mavriplis, D., 2014. CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences. Tech. Rep., NASA Langley Research Center, Hampton, Virginia, USA, NASA/CR-2014-218178.

Spalart, P.R., 2000. Strategies for turbulence modelling and simulations. Int. J. Heat Fluid Flow 21 (3), 252–263.

Spalart, P.R., Allmaras, S.R., 1994. One-equation turbulence model for aerodynamic flows. Rech. Aerosp. (1), 5–21.

Spalart, P.R., Belyaev, K.V., Garbaruk, A.V., Shur, M.L., Strelets, M.K., Travin, A.K., 2017. Large-eddy and direct numerical simulations of the Bachalo-Johnson flow with shock-induced separation. Flow Turbul. Combust. 99 (3–4), 865–885.

Spalart, P.R., Deck, S., Shur, M.L., Squires, K.D., Strelets, M.K., Travin, A., 2006. A new version of detached-eddy simulation, resistant to ambiguous grid densities. Theor. Comput. Fluid Dyn. 20 (3), 181–195.

Spalart, P.R., Shur, M., 1997. On the sensitization of turbulence models to rotation and curvature. Aerosp. Sci. Technol. 1 (5), 297–302.

Uzun, A., Malik, M.R., 2019. Wall-resolved large-eddy simulations of transonic shock-induced flow separation. AIAA J. 57 (5), 1–18.

Wang, J.X., Huang, J., Duan, L., Xiao, H., 2019. Prediction of Reynolds stresses in high-mach-number turbulent boundary layers using physics-informed machine learning. Theor. Comput. Fluid Dyn. 33 (1).

Wang, J.X., Wu, J.L., Xiao, H., 2017. Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. Phys. Rev. Fluids 2 (3).

Weatheritt, J., Sandberg, R., 2016. A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship. J. Comput. Phys. 325, 22–37.

Weatheritt, J., Sandberg, R.D., 2017. The development of algebraic stress models using a novel evolutionary algorithm. Int. J. Heat Fluid Flow 68, 298–318.

Wilcox, D.C., 2006. Turbulence Modeling for CFD, third ed. DCW Industries.

Wu, J.L., Xiao, H., Paterson, E., 2018. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. Phys. Rev. Fluids 3 (7), 074602.

Wu, J., Xiao, H., Sun, R., Wang, Q., 2019. Reynolds-averaged Navier-Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned. J. Fluid Mech. 869, 553–586.

Xiao, H., Wu, J.L., Laizet, S., Duan, L., 2020. Flows over periodic hills of parameterized geometries: A dataset for data-driven turbulence modeling from direct simulations. Comput. & Fluids 200, 104431.

Zhang, Q., Yang, Y., 2013. A new simpler rotation/curvature correction method for spalart-allmaras turbulence model. Chin. J. Aeronaut. 26 (2), 326–333.