



LES informed data-driven models for RANS simulations of single-hole cooling flows

Christopher D. Ellis^{a,*}, Hao Xia^b

^a Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, University Park Campus, Nottingham, NG7 2TQ, Nottinghamshire, UK

^b National Centre for Combustion and Aerothermal Technology, Loughborough University, West Park, Loughborough, LE11 3GR, Leicestershire, UK

ARTICLE INFO

Keywords:

Film cooling
Jet flows
Gas turbines
Turbulence modelling
Data-driven
Machine learning
Neural networks
Random forests

ABSTRACT

A LES-informed data-driven approach for improved predictions of the turbulent heat flux vector has been sought for film and effusion cooling flow applications. Random forest and shallow neural networks have been used to train a spatially varying coefficient for the Higher-Order Generalised Gradient Diffusion Hypothesis (HOGGDH) turbulent heat flux closure model. *a priori* results of the turbulent heat flux magnitude showed significant improvements over the standard HOGGDH model. The random forest model was implemented into OpenFOAM with a previously published data-driven turbulent anisotropy model. The random forest model provided modest improvements to both low and high-blowing ratio film cooling cases along centreline and spanwise distributions. Large cooling effectiveness improvements (up to 82%) were found when compared to the Gradient Diffusion Hypothesis (GDH) model and marginal improvements were shown when compared to the HOGGDH model with its standard coefficient of 0.6.

1. Introduction

Data-driven modelling using machine learning tools has begun to provide functional models that improve upon the current state-of-the-art in Computational Fluid Dynamics (CFD). It has been used as a constructive tool to tune coefficients [1,2], identify model-form uncertainty [3] and develop closure models [4,5]. Closure models appear in the Reynolds-Averaged Navier–Stokes (RANS) simulations due to the Reynolds averaging procedure on the non-linear terms of the Navier–Stokes equations. RANS simulations form a suitable framework providing affordable solutions for engineering-based flows. The steady RANS equations for continuity, momentum and energy are:

$$\frac{\partial}{\partial x_j} (\bar{\rho} \bar{u}_j) = 0 \quad (1)$$

$$\frac{\partial}{\partial x_j} (\bar{\rho} \bar{u}_i \bar{u}_j) = -\frac{\partial \bar{P}}{\partial x_i} + \frac{\partial}{\partial x_j} (2\mu \bar{S}_{ij}^* - \bar{\rho} \overline{u'_i u'_j}) \quad (2)$$

$$\frac{\partial}{\partial x_j} (\bar{\rho} \bar{u}_j \bar{h}_0) = \frac{\partial}{\partial x_j} \left(\bar{\rho} \alpha \frac{\partial \bar{T}}{\partial x_j} - \bar{\rho} c_p \overline{u'_j T'} \right) \quad (3)$$

respectively, where the terms that require closing are the Reynolds stress tensor, $\overline{u'_i u'_j}$, and the turbulent heat flux vector $\overline{u'_j T'}$. The Reynolds stress tensor is commonly closed with a linear eddy viscosity model and the Boussinesq hypothesis,

$$\overline{u'_i u'_j} = \frac{2}{3} \delta_{ij} k - 2\nu_t S_{ij}^* \quad (4)$$

and the turbulent heat flux vector is often closed with the Gradient Diffusion Hypothesis (GDH),

$$\overline{u'_j T'} = -\alpha_t \frac{\partial \bar{T}}{\partial x_j} = -\frac{\nu_t}{Pr_t} \frac{\partial \bar{T}}{\partial x_j} \quad (5)$$

The Boussinesq hypothesis assumes that the anisotropic stress is proportional to the deviatoric strain rate, S_{ij}^* . However, the stresses in real turbulent flows, are rarely proportional to the deviatoric strain rate. Likewise, the GDH assumes that the turbulent heat flux vector is aligned to the mean temperature gradient of the flow, however, Ellis and Xia [6] highlight deviations between the LES-resolved turbulent heat flux and the GDH model.

One group of flows that RANS simulations struggle to predict are film and effusion cooling flows. These flows are found across turbine and combustor components in modern gas turbine systems. The aerodynamic cooling provided by these flows protects the components from the hot combustion gases. Film and effusion cooling features a system of holes perforated through the component providing a film of cool air across the external surface. In the combustor, effusion cooling holes cool the combustor liner using an array of densely packed cooling holes, allowing air from the high-pressure compressor to protect the liner from the hot gases found in the combustor (Fig. 1). In turbine systems, film cooling holes are found in a sparse arrangement on the blades and

* Corresponding author.

E-mail address: chris.ellis1@nottingham.ac.uk (C.D. Ellis).

Nomenclature**Acronyms/Abbreviations**

ACE	Adiabatic Cooling Effectiveness
BR	Blowing Ratio
CFD	Computational Fluid Dynamics
DNS	Direct Numerical Simulation
DR	Density Ratio
GDH	Gradient Diffusion Hypothesis
LES	Large-Eddy Simulation
LiTT	LES-informed Turbulent Transport
ML	Machine Learning
NN	Neural Network
OOB	Out-of-bag
PISO	Pressure Implicit Split Operator
RANS	Reynolds-Averaged Navier–Stokes
RF	Random Forest
RMSE	Root-Mean Square Error
RSTM	Reynolds-Stress Transport Model
SST	Shear Stress Transport
TBNN	Tensor-Basis Neural Network
WALE	Wall-Adaptive Local Eddy-viscosity

Greek Letters

α	Turbulence diffusivity (m^2/s)
ν	Kinematic viscosity (m^2/s)
ρ	Fluid density (kg/m^3)
τ	Time scale (s)
β^*	Turbulence model constant
ω	Specific dissipation rate ($1/\text{s}$)
δ_{ij}	Dirac delta
ε	Dissipation rate (m^2/s^3)
η_{aw}	Adiabatic cooling effectiveness

Roman Letters

a_{ij}	Turbulence anisotropy
c_θ	Model tuning constant
c_p	Specific heat capacity ($\text{J}/\text{kg}/\text{K}$)
D	Hole diameter (m)
D_{ij}	Diffusivity Tensor (m^2/s)
h	Fluid enthalpy (J/kg)
k	Turbulent kinetic energy (m^2/s^2)
l	Hole length (m)
P	Pressure (Pa)
p	Hole pitch (m)
Pr	Prandtl number
R_{ij}	RANS Reynolds stress tensor (m^2/s^2)
S_{ij}	Strain-rate ($1/\text{s}$)
s_{ij}	Non-dimensional strain-rate
T	Fluid temperature (K)
T_*	Normalised temperature gradient
U	Streamwise velocity (m/s)
u	Velocity (m/s)
u_i	Velocity vector (m/s)
W_{ij}	Vorticity tensor ($1/\text{s}$)

w_{ij}	Non-dimensional vorticity tensor
X	Streamwise component(m)
x_i	Spatial component (m)
X	Wall-normal component (m)
y	Wall distance (m)
X	Spanwise component (m)

Sub/Superscript

\square'	Fluctuating property
\square^*	Traceless property
\square_0	Stagnation property
\square_{aw}	Adiabatic wall property
\square	Mean property
\square_c	Coolant property
\square_∞	Freestream property
i, j, k, l, m, n	Indexing components
\square_t	Turbulent property

endwall regions of the turbine. The ability to accurately model film and effusion cooling flows in CFD is important in understanding the performance and lifetime of the turbine and combustor components as well as the overall gas turbine efficiency. For component lifetimes, Han et al. [7] illustrate the importance of predicting component temperatures by showing that component lifetime can halve with just a 2% error in predicted metal temperature.

Previous research has consistently demonstrated the challenges associated with accurately predicting the cooling performance of film and effusion cooling flows using Reynolds-Averaged Navier–Stokes (RANS) simulations. Hoda and Acharya [8] evaluated various $k-\varepsilon$ turbulence model formulations, including a DNS-informed model, against experimental data. While the DNS-informed model exhibited reasonable agreement for recirculation zones, all models struggled to capture the downstream velocity field. Acharya et al. [9] further emphasised these limitations by reporting overpredictions of coolant jet penetration and underpredictions of lateral spreading rate across different turbulence models, including a Lam-Bremhost Reynolds-Stress Transport Model (RSTM).

Walters and Leylek [10,11] observed that $k-\varepsilon$ models overpredicted centreline cooling effectiveness at low blowing ratios but captured overall trends. However, at higher blowing ratios, these models failed to reproduce the experimentally observed jet lift-off phenomenon. This deficiency was attributed to the absence of non-linear anisotropy, which significantly influences coolant lateral spread. Subsequent studies by Harrison and Bogard [12] confirmed these challenges, with the $k-\omega$ model providing accurate lateral-averaged cooling effectiveness but overpredicting centreline values. The realisable $k-\varepsilon$ model exhibited the opposite trend. Azzi and Jubran [13] attempted to improve predictions at lower blowing ratios by incorporating an anisotropic near-wall model [14], but the approach proved ineffective at higher blowing ratios where jet lift-off and reattachment behaviour dominated.

Although RANS simulations have struggled to provide accurate cooling distributions, Large-Eddy Simulations (LES) have successfully replicated cooling effectiveness results in a range of numerical studies [15–19]. However, LES requires a fine mesh and time step to resolve the fine-scale turbulent structures. This results in high computational costs that have limited the use of LES to a small number of cases while RANS is relatively low-cost and affordable for wider use.

Data-driven models have emerged as an alternative approach for modelling film cooling flows. Milani et al. [20] employed a machine-learned turbulent diffusivity within the GDH turbulent heat flux closure, trained on high-fidelity data. This approach yielded qualitative

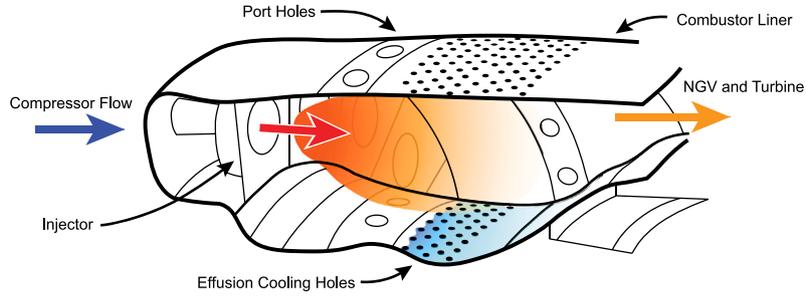


Fig. 1. Diagram of effusion cooling holes used in gas turbine combustors.

improvements in adiabatic cooling effectiveness (ACE) contours compared to traditional constant turbulent Prandtl number models. Subsequent work by the same authors [21] explored the use of Tensor-Basis Neural Networks (TBNN) [22] to model a tensor-based turbulent diffusivity, allowing for deviations from the GDH alignment assumption. While significant enhancements were observed in the centreline plane, improvements in the spanwise plane were less pronounced.

Ling et al. [23] proposed a random forest model to predict barycentric coordinates associated with turbulent anisotropy eigenvalues, demonstrating successful generalisation to different flow geometries. However, this approach was not integrated into a RANS framework for Reynolds stress closure. More recently, Ellis and Xia [24] developed a TBNN model for turbulence anisotropy using LES data, which exhibited improved *a posteriori* RANS results. When coupled with a HOGGDH turbulent heat flux model, this approach enhanced predictions of lateral coolant distributions, although the GDH model provided better agreement with experimental data in the initial jet development region.

In this article, a data-driven model for the HOGGDH coefficient extends the previous literature [20,25,26] and is then combined with a novel data-driven anisotropy model [24] in OpenFOAM [27], an open-source CFD framework. In the first section, the HOGGDH coefficient model is introduced and then the analysis extends previously published results with a test dataset isolated from the training procedure assessing its improved performance over existing models. In the second section, the implementation of the combined models in a RANS solver is discussed and then the results of the models are investigated on cooling cases to assess *a posteriori* performance of the models running in a steady RANS solver, compared to the time-averaged results of the LES solutions. The present article provides a novel contribution to data-driven modelling by investigating the combination of both models which will be discussed in detail.

2. Data-driven HOGGDH coefficient

2.1. Methodology

Previous studies [6] have shown that the HOGGDH closure improves the prediction of the turbulent heat flux vector in regions where it deviates from the direction of the temperature gradient. In the work of Ling et al. [28], the HOGGDH closure was used where its impact on the scalar field showed qualitative improvements in the early shear layer development. The HOGGDH closure,

$$\overline{u'_i T'} = -c_\theta \tau_i \frac{\overline{u'_i u'_k u'_k u'_j}}{k} \frac{\partial \overline{T}}{\partial x_j}, \quad (6)$$

features a coefficient (c_θ) that is similar to the turbulent Prandtl number. In the literature, heat transfer predictions have benefited from a spatially varying turbulent Prandtl number in the GDH closure [28,29]. A spatially varying diffusivity coefficient, c_θ , in the HOGGDH model is investigated in this article to identify its ability to improve heat transfer predictions in film cooling flows.

Table 1

High-fidelity simulation case parameters and dataset use.

Case	Training & Validation	Testing		BR	DR
		<i>a priori</i>	<i>a posteriori</i>		
<i>br05-dr20</i>	✓		✓	0.5	2.0
<i>br10-dr20</i>	✓		✓	1.0	2.0
<i>br06-dr12</i>		✓	✓	0.6	1.2

A data-driven model for c_θ is generated using two machine learning algorithms: random forests and shallow neural networks. The algorithms are trained on the time-averaged results of two LES datasets and then tested on a third, as shown in Table 1. The three LES cases featured in this article are single-row cooling configurations, described in Fig. 2, based on the experimental work of Sinha et al. [30]. The cases feature a single row of 35° inclined cylindrical cooling holes with a length-to-diameter ratio, l/D , of 1.75. The blowing and density ratios,

$$BR = \frac{\rho_c U_c}{\rho_\infty U_\infty} \quad (7)$$

and

$$DR = \frac{\rho_c}{\rho_\infty} \quad (8)$$

respectively, are varied between the cases. The lower blowing ratio jets will stay attached to the coolant plate and higher blowing ratio jets will separate and reattach to the coolant plate downstream, which results in significantly different coolant behaviour across the plate. The values of DR and BR chosen for the cases replicated those investigated by Sinha et al. [30] but are also within a suitable range of values feasible in a gas turbine combustor and turbine [31].

The critical dimensions of the numerical domain are provided in Fig. 2. To capture the row of cooling holes, a spanwise periodic boundary is applied to mimic the pitch, p , of the cooling holes. The mainstream inlet supplies the approaching turbulent boundary layer at a temperature and freestream velocity of 300K and 20 m/s respectively. The fluctuations of the turbulent boundary layer are provided to the inlet condition with a digital filtering technique [32] that was shown to replicate the approaching boundary layer in the work of Ellis and Xia [19]. Coolant is supplied to the base of the coolant plenum at a velocity and temperature that reproduces the case's blowing and density ratio.

OpenFOAM v1712 [27] has been used for the LES simulations using a pressure-based solver suitable for transient flows, *rhoPimpleFoam*. The Wall-Adaptive Local Eddy-viscosity (WALE) model [33] is used for the sub-grid scale stresses which provide the correct near-wall scaling without the requirement of a dynamic formulation. For the sub-grid scale heat fluxes, a sub-grid scale Prandtl number of 0.4 is recommended in the literature [34–36]. A structured, hexahedral meshing approach has been used to minimise non-orthogonality and skewness to ensure that numerical diffusivity is minimal across the jet mixing region. The resultant mesh for the investigated cases has a cell count of

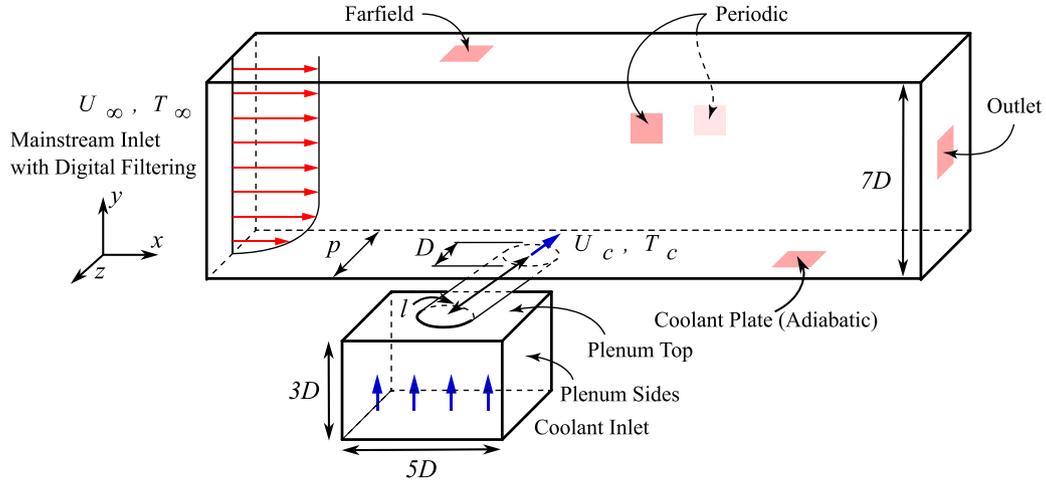


Fig. 2. Case geometric configuration with computational domain, boundary conditions and critical geometric parameters. Source: Adapted from Ellis and Xia [24].

20.6 million cells resolving more than 90% of the total turbulent kinetic energy. Additional mesh details, computational setup and validation are provided in the article of Ellis and Xia [19].

2.1.1. Machine learning features

Input and target features for the machine learnt c_θ models are extracted from the time-averaged LES results for each investigated case. The target c_θ for the HOGGDH model is obtained with,

$$c_\theta = -\frac{k}{\tau_t} \frac{\overline{u'_i T'}}{\overline{u'_l u'_m u'_n} \frac{\partial \overline{T}}{\partial x_n}} \left(\overline{u'_i u'_k u'_j} \frac{\partial \overline{T}}{\partial x_j} \right) \quad (9)$$

where each component of the turbulent heat flux is weighted by its defining component of the HOGGDH modelled turbulent heat flux vector. The turbulent time scale, $\tau_t = k/\varepsilon$, requires a RANS-suited dissipation rate, ε . Although the dissipation rate can be extracted from the LES data with a sufficiently fine mesh, the dissipation rate available in RANS simulations deviates from an LES-derived quantity. Therefore, to obtain a RANS-representative dissipation rate, an LES-informed Turbulent Transport (LiTT) process is defined. The LES time-averaged fields are frozen and a pseudo-simulation is solved using the transport equation for the k - ω SST specific turbulent dissipation rate. For the k - ω SST model, the turbulent dissipation rate is defined as $\varepsilon = \beta^* \omega k$ where $\beta^* = 0.09$.

The Reynolds stress tensor is present in the HOGGDH closure (Eq. (6)). The Boussinesq hypothesis is employed to calculate the Reynolds stresses for two-equation turbulence model approaches. Evaluation of c_θ from the LES also employs the same approach to ensure the model replicates the wider RANS model environment in which it can be used. The LiTT's calculation of ω enables the Boussinesq hypothesis to be evaluated with the mean strain rates from the frozen LES field.

The input features for the model are derived from an initial 19 features investigated in the preceding work [26]. The features are adopted from similar features investigated by Milani et al. [20] in their GDH model. These 19 features exhibit non-dimensional and Galilean invariance ensuring the machine-learned models of c_θ will not be biased by orientation or scale. Two features reflect the flow's physical properties: a wall-distanced Reynolds number and a viscosity ratio. The remaining features are the mathematically relevant independent invariants that can be found from the infinite polynomial of normalised strain-rate, vorticity and temperature gradient using the Cayley–Hamilton theorem. Normalisation of the strain-rate (S_{ij}) and vorticity (W_{ij}) tensors are achieved with the turbulent time scale, $\tau_t = k/\varepsilon$, so $s_{ij} = \tau_t S_{ij}$ and $w_{ij} = \tau_t W_{ij}$. The normalised temperature gradients,

$$T_* = \frac{k^{3/2}}{\varepsilon} \frac{\nabla T}{T_\infty - T_c} \quad (10)$$

Table 2

Feature prediction importance of the top five features with the random forest algorithm.

Feature	Feature Importance	
	OOB	Cumulative
$T_*^T s T_*$	10.2%	10.2%
v_i/ν	9.0%	19.2%
$T_*^T T_*$	8.2%	27.5%
$T_*^T s^2 T_*$	7.5%	35.0%
$y\sqrt{k}/\nu$	7.1%	42.0%

are normalised by the mainstream to coolant temperature difference and a turbulence length scale, $k^{3/2}/\varepsilon$.

In the present work, each model is trained on a reduced feature set dictated by the feature importance study conducted on the random forest model in Ellis et al. [26]. This simplifies the model for use in RANS frameworks. The feature importance is performed using an out-of-bag (OOB) predictor importance by permutation which highlights which parameters provide the largest change to the model when they are permuted within the input feature space. The reduced feature set used to train both the random forest and shallow neural network models is presented in Table 2. The top five features cumulatively make up 42% of the total relative feature importance from the original 19 features.

2.1.2. Machine learning algorithms

The machine learning algorithms: random forests and shallow neural networks were studied with the models that they provide. The random forests model [37–39] features an ensemble of decision tree learners where the collection of results from each tree is averaged in the present work to obtain the model output. The number of trees, maximum number of decision tree splits, minimum node size and number of randomly selected features at each node are hyperparameters that can be tuned to improve the model performance. The random forest machine learning method in turbulence modelling has shown success in previous works by Milani et al. [20] and Ling and Templeton [3]. The shallow neural networks algorithm features a feed-forward network with a single hidden layer between the input and output layers. Additional details on model hyperparameter selection are discussed by Ellis et al. [26].

In this work, the random forest model is studied with different numbers of trees and maximum numbers of splits to investigate increasing model complexity in the present case. For the minimum node size and number of randomly selected features at a node, Hastie et al. [40] recommends values of five and one-third of the total number of features

Table 3
Normalised error of machine learning predictions compared to the HOGGDH with $c_\theta = 0.6$.

	$RMSE/U_\infty(T_\infty - T_c) (\times 10^{-4})$		
	Training & Validation <i>br05-dr20</i> <i>br10-dr20</i>		Testing <i>br06-dr12</i>
HOGGDH ($c_\theta = 0.6$)	94.73	79.66	61.20
Random Forest	22.13	20.96	17.89
Shallow Neural Network	20.42	17.88	17.71

Table 4
Normalised error of the random forest models of differing complexities.

	Trees	Splits	$RMSE/U_\infty(T_\infty - T_c) (\times 10^{-4})$		
			Training & Validation <i>br05-dr20</i> <i>br10-dr20</i>		Testing <i>br06-dr12</i>
<i>RF Model 1</i>	25	100	22.13	20.96	17.89
<i>RF Model 2</i>	25	20,000	7.73	7.18	19.70
<i>RF Model 3</i>	1000	100	21.83	20.61	17.71
<i>RF Model 4</i>	1000	20,000	7.53	6.95	19.51

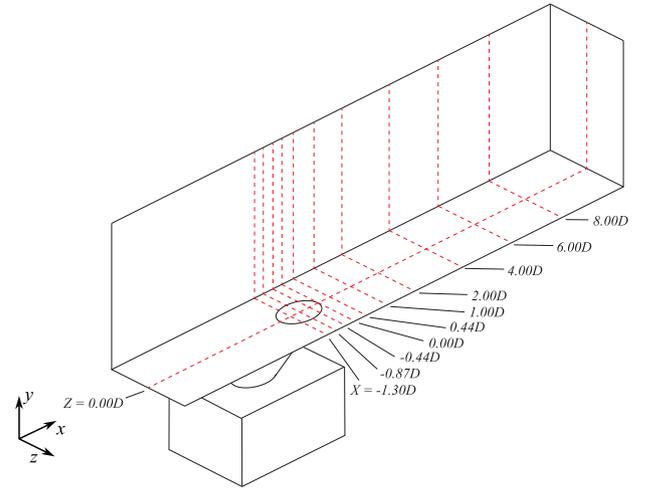


Fig. 3. Data slices for machine learning datasets.

respectively. For the present work, the number of randomly selected features at a node was rounded up to the nearest integer to give a value of two. For the shallow neural network, 20 hidden nodes were used. Additional layers were investigated for the present work but did not improve upon the model performance. A tan-sigmoid activation function is used for each node in the shallow neural network architecture.

2.2. Model a priori results

The two c_θ machine learning models, using the random forests and shallow neural network algorithms, are tested using the LES testing dataset (*br06-dr12*). Testing is performed on this unseen case, of a different density and blowing ratio, to confirm that the model can perform on a different cooling flow condition. Predictions of c_θ are made from the time-averaged, frozen LES field and compared with those extracted from the LES. The HOGGDH closure, with the machine learnt coefficient, is then used to evaluate the turbulent heat flux for direct comparison to those resolved by the LES. To assess the improvement of the machine learning model, the HOGGDH closure, with its standard coefficient of 0.6 [41], is used for comparison.

2.2.1. Error analysis

The constructed models are assessed by comparing the Root-Mean-Square Error (RMSE) between the machine learning models and the LES-resolved turbulent heat flux magnitude. The RMSE is defined in Eq. (11). The error is evaluated across domain data points located on the slices shown in Fig. 3. The slices show regions crossing the path of the cooling jet from the cooling hole and the jet's initial development to the downstream region.

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^3 \left(\overline{u_i' T'}_{model} - \overline{u_i' T'}_{LES} \right)^2} \quad (11)$$

Table 3 presents the RMSE results comparing the random forest and shallow neural network machine learning models with the advanced HOGGDH closure across the training and validation cases (*br05-dr20* and *br10-dr20*) and the testing case (*br06-dr12*). Both machine learning approaches show reductions in RMSE of more than 70% compared to the HOGGDH closure for both the training and validation cases as well as the testing case, showing the model can generalise to other blowing and density ratios not present in the model training. The shallow neural network outperforms the random forest model in all cases, but this reduction in error is much smaller in the testing case.

Table 4 shows the RMSE across both the training and validation datasets (*br05-dr20* and *br10-dr20*) and the testing dataset (*br06-dr12*)

to elaborate on the impact of model complexity. In the training and validation datasets, a 65% reduction in RMSE is achieved by increasing the maximum number of splits from 100 to the optimum value of 20,000. Increasing the number of trees in the ensemble, from 25 to 1000, provides minor improvements in the RMSE of about 1.5%. This performance is in line with the trends of Ellis et al. [26]. However, analysis of the model's response on the testing dataset (case *br06-dr12*), shows a different trend. For the testing dataset, increasing the maximum number of splits to 20,000 increases the RMSE by approximately 10%. This is evidence of model overfitting and shows the importance of performing testing on datasets that are completely isolated from the training data. Increasing the number of trees in the ensemble maintains the reduction of RMSE on the testing dataset. However, increases in the number of trees provide a linear increase in computational cost. The results of Table 4 justify the random forest hyperparameters chosen for further use (25 trees and 100 splits).

2.2.2. Test case coefficient prediction

Results of the machine learning models are presented across six profiles for the testing case (*br06-dr12*), covering the evolution of the coolant jet along the centreline ($Z = 0$) and a lateral location ($Z = 0.5D$) at locations $1D$, $4D$ and $8D$ downstream of the coolant hole trailing edge. Comparisons are made to the LES-resolved turbulent heat flux to validate the machine learning results. In addition, comparisons are made to the advanced HOGGDH closure to show the benefits of the machine learning model against a benchmark approach.

Fig. 4 presents profiles comparing the modelled c_θ and turbulent heat flux magnitude $|\overline{u_i' T'}|$. Both the random forest and the shallow neural network model show good agreement with the target c_θ profiles. However, Figs. 4(a), 4(b) and 4(c) show disagreements in the c_θ prediction to the LES in the near-wall region. Within this region, the turbulent heat flux magnitude is small compared to the shear layer. In Fig. 4(c), the LES indicates a large negative c_θ necessary to drive the turbulent heat flux magnitude indicating a region of strong counter-gradient transport. Despite these highlighted flaws, the magnitude of the turbulent heat flux is in close agreement with the LES. Within the shear layer on the jet centreline, both the neural network and random forest models improve the magnitude of the turbulent heat flux compared to the HOGGDH model in the three profiles progressing downstream. On the off-centreline profiles ($Z = 0.5D$), the turbulent heat flux magnitude and coefficients are well captured by the machine learning approaches as the jet develops downstream. At $X = 1D$, the off-centreline profile (Fig. 4(d)) shows the near-wall peak of the machine learning models matches the peak seen in the LES, while the

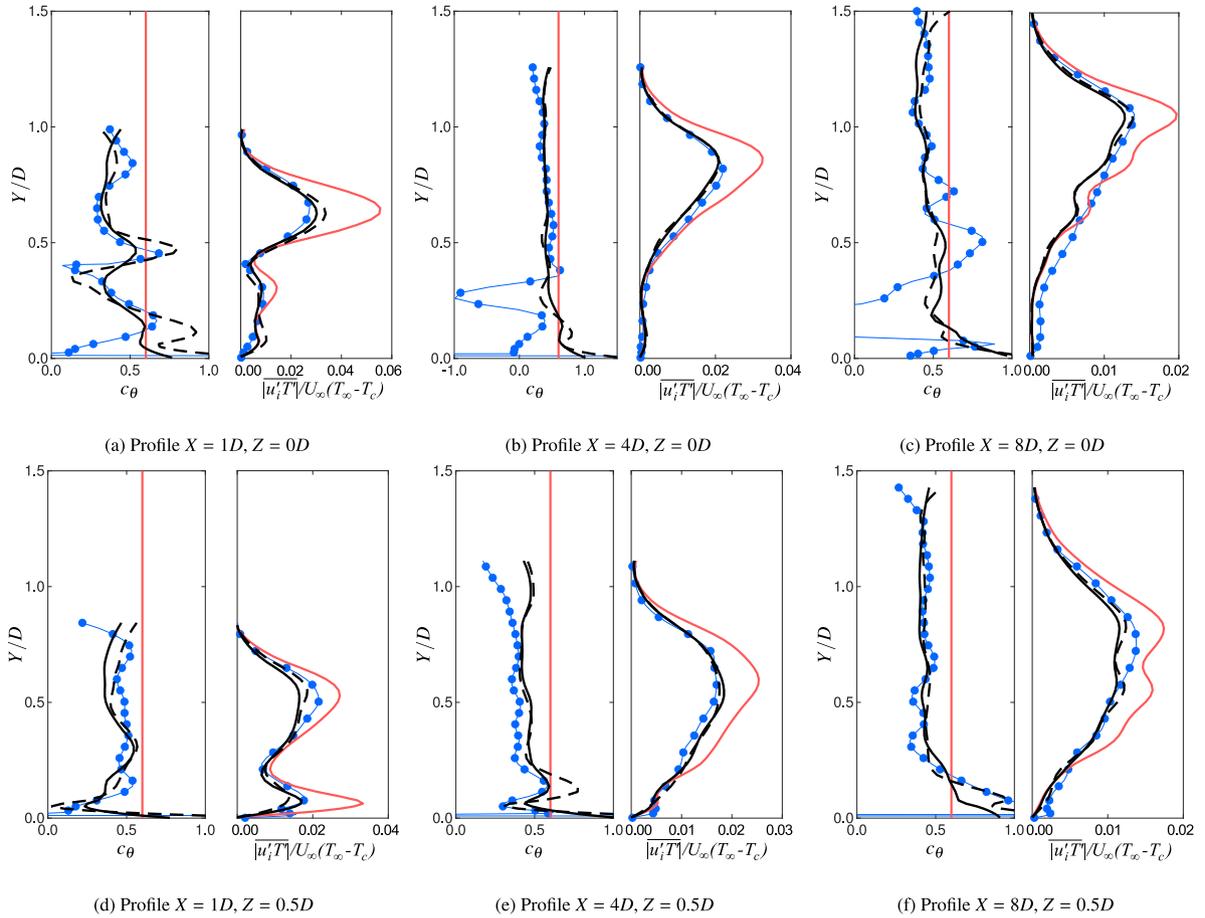


Fig. 4. Profiles of random forest [—] and neural network [---] modelled c_θ and turbulent heat flux magnitude compared to HOGGDH with $c_\theta = 0.6$ [—] and extracted and resolved LES [•]. Test case *br06-dr12*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

HOGGDH closure overpredicts this peak by a factor of two. Away from the wall, the shear layer is underpredicted and lacks the defined peak of the LES. This sharp peak is exhibited by the HOGGDH model, although it overpredicts the magnitude compared to the LES results.

Fig. 5 presents profiles of the streamwise, spanwise and normal components of the modelled turbulent heat flux. The c_θ models enable improvements to the turbulent heat flux magnitude but they do not dictate the distribution of each component. Recalling Eq. (6) shows that the component values are dictated by the inner product of the Reynolds stress tensor with itself and the temperature gradient vector. Fig. 5(a) demonstrates that this approach provides accurate turbulent heat flux components just downstream of the coolant hole trailing edge on the jet centreline. Further downstream, Fig. 5(c), shows the turbulent heat flux components deviate from the LES resolved turbulent heat flux, although the magnitude is in good agreement. The spanwise component is zero on the centreline profile where the coolant jet exhibits spanwise symmetry and has therefore been omitted.

In the off-centreline profiles, shown in Figs. 5(d) and 5(e), the turbulent heat flux exhibits deviations from the LES data but the magnitude was well-predicted (shown in Fig. 4). In the downstream profiles, this difference to the LES results is more prominent. The turbulent heat flux modelled by the machine learning approaches shows deviations in the turbulent heat flux components that present flaws in the functional form of the HOGGDH closure which dictates the angle of the modelled turbulent heat flux. With this approach, the LES resolved turbulent heat flux magnitude can be reproduced by a machine learning approach. However, the angle of the modelled turbulent heat flux cannot be recovered, unless the flow aligns with the direction of the HOGGDH closure.

3. Augmented RANS model

With the success of the model's *a priori* results, the behaviour of such models in a RANS environment is explored in this section. The *a priori* performance of both the shallow neural network and random forest models were similar and therefore the random forest model was taken forward due to ease of implementation. The random forest model of the HOGGDH coefficient was implemented and compiled into OpenFOAM. Within the present work, the model is used in conjunction with the turbulent anisotropy model investigated by Ellis and Xia [24], which showed significant improvement to the time-averaged coolant and velocity fields. The implementation is first discussed and then the results of the data-driven model combination on film and effusion cooling flows are presented.

3.1. Methodology

The two data-driven models are added into OpenFOAM's *rhoSimpleFoam* solver. The modified solver is described in Fig. 6. The two models are added to the code to provide improved data-driven Reynolds stresses and turbulent heat flux for the momentum and energy equations respectively which aim to improve the primitive variables in the simulation.

3.1.1. Turbulent anisotropy implementation

The implementation of the turbulence anisotropy model is discussed in detail by Ellis and Xia [24] and summarised here. The TBNN anisotropy model was also trained on case *br05-dr20* as well as two other cases covering cooling holes with different inclination angles

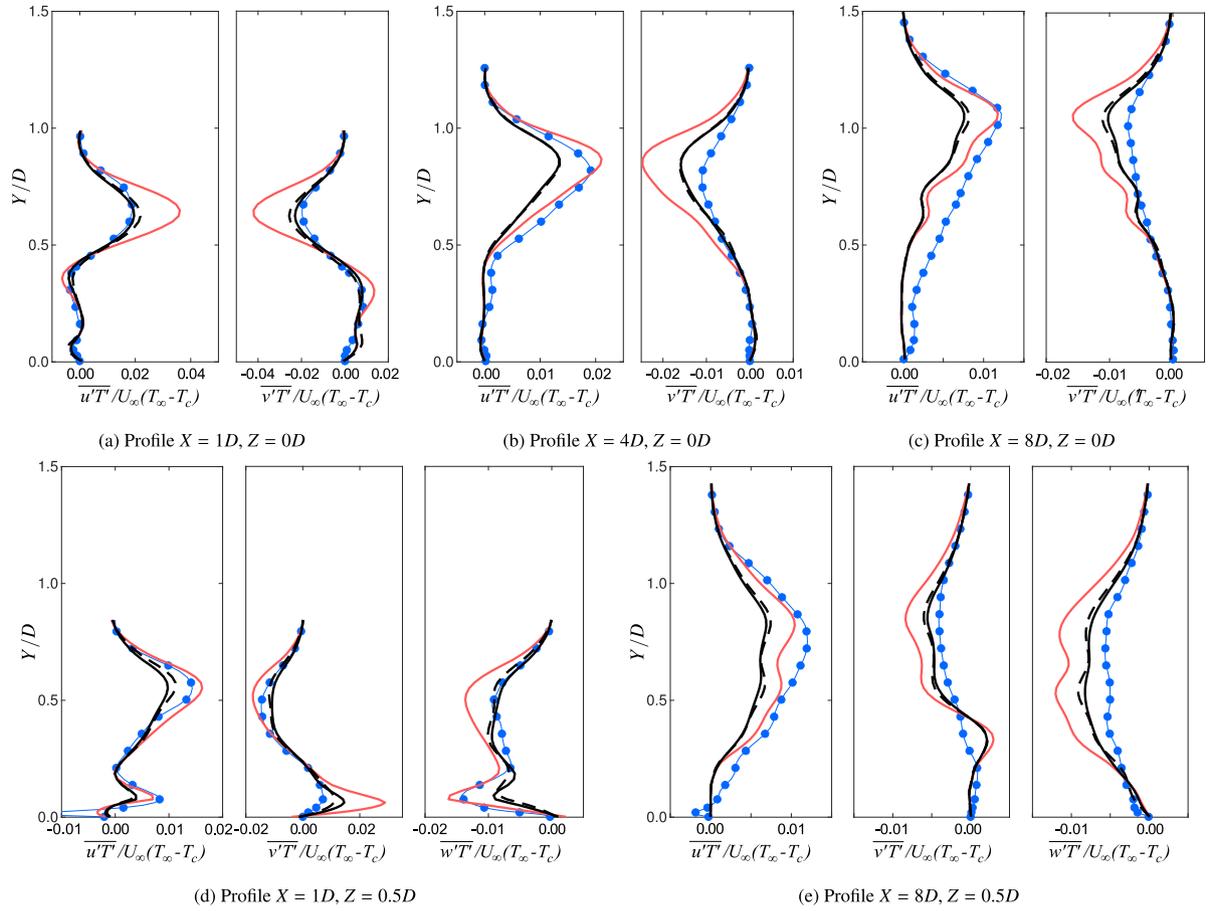


Fig. 5. Profiles of random forest [—] and neural network [---] modelled turbulent heat flux vector components compared to HOGGDH with $c_\theta = 0.6$ [—] and resolved LES [—●—]. Test case *br06-dr12*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and length-to-diameter ratios. The model employs a $k-\omega$ SST [42] turbulence model augmented with the turbulent anisotropy TBNN PyTorch model to improve the Boussinesq hypothesis Reynolds stresses available with a standard $k-\omega$ SST model [42]. After the turbulence equations are solved, the neural network model is evaluated to establish the turbulent anisotropy and then the Reynolds stresses for the next iteration, highlighted in Fig. 6. The neural network Reynolds stresses, R_{ij}^{NN} , are obtained from the anisotropy output by the neural network model,

$$R_{ij}^{NN} = 2k \left(a_{ij}^{NN} + \frac{1}{3} \delta_{ij} \right), \quad (12)$$

where the anisotropy directly output by the neural network is a_{ij}^{NN} .

The coupling of velocity, pressure and accurate Reynolds stresses within the momentum equation for high Reynolds number flows, relevant for engineering applications [43], can significantly impact numerical stability and model robustness [44]. To address model instability the modelled neural network Reynolds stress, R_{ij}^{NN} , is split into a linear and non-linear component where the linear component of the Reynolds stresses is equal to the Boussinesq hypothesis Reynolds stresses with an implicit and explicit part provided to the segregated momentum equations. The difference between the neural network model and the Boussinesq hypothesis is the non-linear component which is treated explicitly as an additional term within the momentum equation. This approach was used by Wu et al. [43] on simple turbulent channel flow, periodic hills and square duct flow.

Relaxation of the non-linear Reynolds stress term is applied across the domain with a value of 0.7. In the near-wall regions, where the anisotropy of the flow is high, additional relaxation is applied using a parabolic damping function that provides a smooth transition from

the wall to the mainstream flow. Finally, a box filter is used on the neural network anisotropy to provide a smooth field without spurious oscillations and discontinuities. The box filter is used with 10 passes to approximate a Gaussian filter with a standard deviation of 2.6 cells. Additional details and results are presented by Ellis and Xia [24] to justify the methodology.

3.1.2. Random forest implementation

The random forest model for the turbulent heat flux coefficient is compiled into OpenFOAM using Matlab's c++ code generator. This provided a method to translate the random forest model from Matlab to c++ to interface with OpenFOAM. The code was wrapped in a function that could be called from the desired OpenFOAM solver.

Local inputs are passed to the random forest model and the local coefficient, c_θ , is returned. Positivity is ensured in the coefficient within the solver to promote solver stability by constraining the resultant turbulent diffusivity tensor, D_{ij} , to be positive semi-definite. The closure,

$$\overline{u_i' T'} = -c_\theta \tau_t \frac{R_{ik} R_{kj}}{k} \frac{\partial T}{\partial x_j}, \quad (13)$$

provides the diffusive flux with a Laplacian scheme within the energy transport equation,

$$\frac{\partial}{\partial x_i} \left(\overline{u_i' T'} \right) = \frac{\partial}{\partial x_i} \left(-D_{ij} \frac{\partial T}{\partial x_j} \right) \text{ where, } D_{ij} = c_\theta \tau_t \frac{R_{ik} R_{kj}}{k} \quad (14)$$

Unlike the Reynolds stress augmentation, the c_θ coefficient from the RF model did not require additional treatment to ensure the solver's stability.

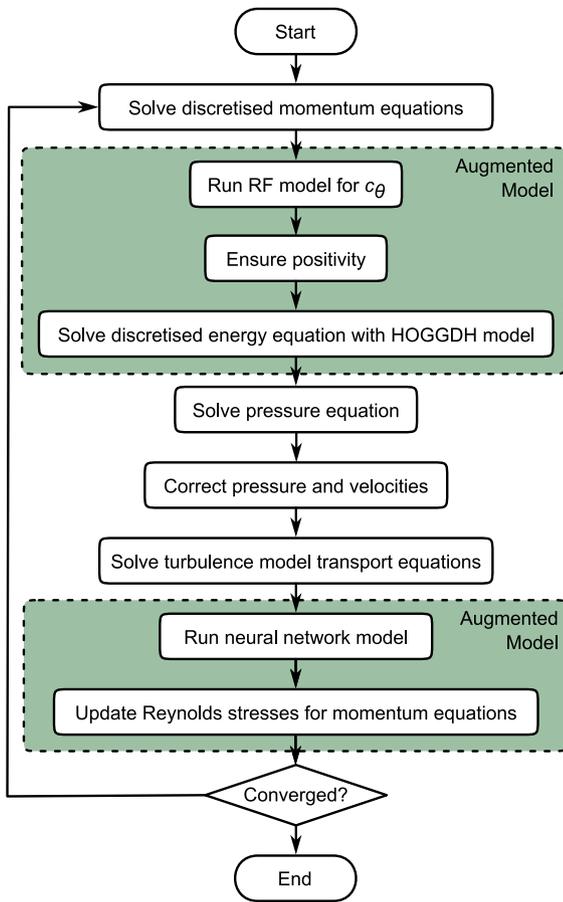


Fig. 6. OpenFOAM's *rhoSimpleFoam* algorithm with augmented machine learning models.

Table 5
Model names and descriptions for identification.

Case name	Turbulence model	Turbulent heat flux	
		Model	Coefficient
SST-GDH	Standard $k-\omega$ SST	GDH	$Pr_t = 0.85$
SST-HOGGDH		HOGGDH	$c_\theta = 0.6$
ML-GDH	SST + a_{ij}^{NN}	GDH	$Pr_t = 0.85$
ML-HOGGDH		HOGGDH	$c_\theta = 0.6$
ML-RF		Random Forest	Variable c_θ

3.2. Model a posteriori results

The predictive capability of the augmented turbulent heat flux coefficient using the random forest model is assessed to identify its improvements and behaviour in three single-hole cases defined in Table 1. The augmented SST model with the neural network anisotropy is used with some additional comparisons made to the standard $k-\omega$ SST turbulence model. Model names and descriptions are provided in Table 5 to aid identification. Standalone comparisons between the standard $k-\omega$ SST model and the model with the data-driven anisotropy are given in Ellis and Xia [24]. Results for ACE distributions are calculated with the adiabatic wall temperature as defined by,

$$\eta_{aw} = \frac{T_{aw} - T_\infty}{T_c - T_\infty}. \quad (15)$$

3.2.1. Case br05-dr20

First, the low blowing ratio case *br05-dr20* is shown in Fig. 7 comparing ACE distributions for the ML-GDH, ML-HOGGDH and ML-RF models. For this case and its conditions, the cooling jet demonstrates

fully-attached behaviour. Introducing the random forest model, where the c_θ parameter is controlled by the trained machine learning model, shows small improvements over the standard HOGGDH closure with a fixed c_θ of 0.6 (ML-HOGGDH). The centreline distribution is in good agreement with the LES dataset over a greater portion of the coolant plate. The lateral spreading of the coolant is underpredicted by the ML-HOGGDH and the ML-RF model but the ML-RF model is closer to the LES results. Both approaches predict the same trends observed in the LES and vastly improve the ACE predictions compared to the ML-GDH model. Introducing the HOGGDH approach provides coolant lateral spreading to a degree not observed by the GDH closure and controlling the c_θ coefficient with the random forest model translates to small improvements in surface coolant predictions.

Fig. 8 quantifies the improvement by introducing the random forest model. In the presented profiles, the random forest has the lowest RMSE. The ML-RF model shows a reduction in the RMSE from the ML-HOGGDH model by 33.4% across the centreline but less improvement is shown in the other profiles. Comparing the ML-RF model to the ML-GDH model shows that the GDH approach is only comparable in the spanwise average metric.

3.2.2. Case br10-dr20

Further analysis of the augmented random forest model is pursued with the higher blowing ratio case *br10-dr20*, where jet separation and reattachment play an important role in the cooling distribution. Comparisons of the modelled turbulent heat flux vector are made in Figs. 9 and 10 comparing the ML-GDH and ML-HOGGDH case with the ML-RF case. Centreline profiles of the turbulent heat flux vector show the random forest model improves both streamwise and normal turbulent heat flux components in the initial coolant jet development region ($X \leq 1.25D$), where the HOGGDH closure overpredicts the turbulent heat flux. The initial profile ($X = -1.25D$), immediately aft of the coolant hole leading edge, shows over predictions and the random forest model predictions of c_θ exceed the value needed to reproduce the LES resolved turbulent heat flux vector magnitude. Further attention to the initial development of the shear layer flow in the model training process could benefit these augmented models further.

In the mixed downstream regions, the normal component of the turbulent heat flux is in good agreement with the resolved LES but the streamwise component is underestimated. The deficit in this prediction, which was not present in the *a priori* analysis shown in Fig. 4, could be a result of the relaxed Reynolds stresses from the augmented Reynolds stress model that deviates from the full anisotropy output of the machine learning model. A potential improvement could be found by training a random forest model with the full or relaxed Reynolds stresses instead of the Boussinesq hypothesis stresses.

Off-centreline profiles of the turbulent heat flux vector are shown in Fig. 10. Small improvements are present in the predictions but a large underprediction in the augmented model is observed in the downstream streamwise turbulent heat flux peak associated with the shear layer. The spanwise component of turbulent heat flux with the augmented model improves the predicted profiles which is an indicator of the improved lateral mixing behaviour of the random forest model. Normal components are closer to those predicted by the LES but the model still overpredicts the values by approximately twice the resolved LES value in the near-wall region.

The propagation of the augmented turbulent heat flux to the temperature field is first shown by the non-dimensional temperature across centreline and off-centreline profiles in Fig. 11. In the centreline, the random forest augmented model shows an improved prediction of the early shear layer where the temperature gradient is particularly sharp. The improved prediction in this region reflects the reduced turbulent heat flux components shown in the centreline profiles of Fig. 9. Although the streamwise turbulent heat flux component is underpredicted in the downstream profiles the non-dimensional temperature profiles

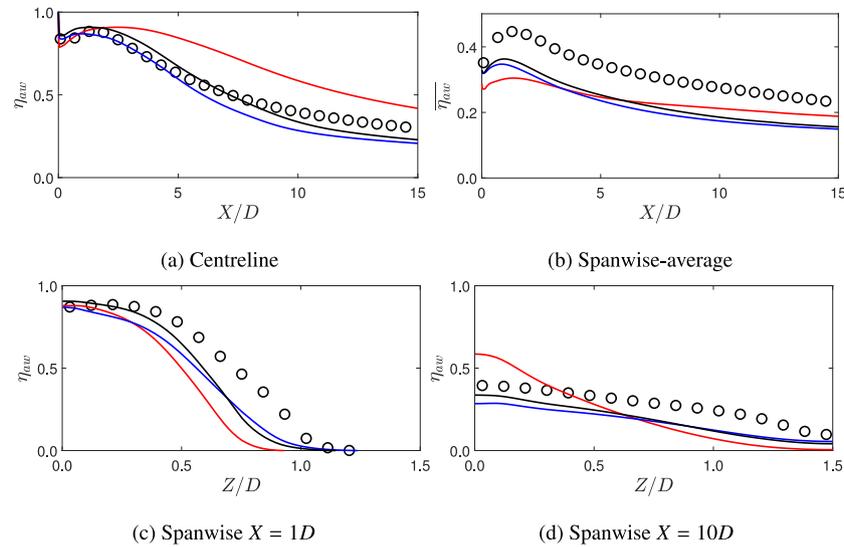


Fig. 7. ACE comparison between ML-GDH [—], ML-HOGGDH [—] and the ML-RF model [—] with LES results [•]. Case *br05-dr20*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

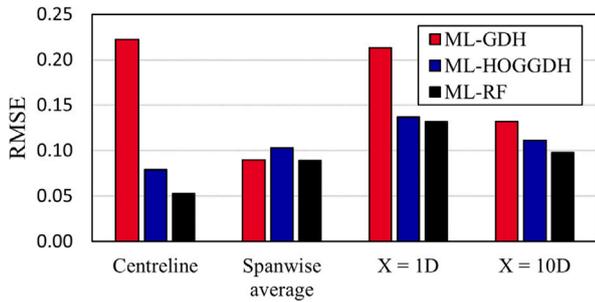


Fig. 8. RMSE of the ACE across the centreline, spanwise average and lateral profiles presented in Fig. 7. Case *br05-dr20*.

remain in good agreement with the LES for the ML-RF case. Off-centreline profiles (Fig. 11(b)) show the random forest model provides a distinct improvement over the standard HOGGDH closure in RANS simulations. Non-dimensional temperature profiles show good agreement to the LES results across the initial development of the cooling jet. This indicates the importance of a spatially varying diffusivity coefficient in film cooling predictions where lateral spreading rates are important in determining the flow's cooling ability.

Plots of ACE distributions are shown with comparisons to the LES results in Fig. 12. The ML-RF model, compared to the ML-GDH and ML-HOGGDH (all with the neural network anisotropy augmented SST model), presents wall ACE distributions that take the best of both the GDH and HOGGDH. In the centreline profile, the augmented random forest model predicts the cooling distribution in line with the GDH closure, where both provide a close representation of the LES results. Meanwhile, the lateral distributions at $X = 10D$ and $15D$ match the respective trends of the HOGGDH closure providing good agreement with the LES. The combination provides improved spanwise-averaged predictions.

Fig. 13 shows that the ML-RF model provides the lowest RMSE for the centreline and spanwise average profiles. The lateral distributions are contested by the ML-GDH model for the $X = 1D$ position and the ML-HOGGDH model for the $X = 10D$ but the ML-RF errors are low. These results indicate a distinct improvement for this high-blowing ratio case where jet separation and reattachment are present. The ML-RF model improves centreline and spanwise average distributions of the ACE over the ML-GDH and ML-HOGGDH models where the GDH

shows reasonable predictions in the centreline distribution and the ML-HOGGDH model improves the downstream predictions of coolant lateral spread.

Contour plots of ACE, shown in Fig. 14, further the analysis and discussion of augmented model improvements. Alongside the results with the neural network anisotropy augmented SST model, the standard $k-\omega$ SST model with GDH (SST-GDH) and HOGGDH (SST-HOGGDH) closures are shown. The addition of neural network anisotropy in the augmented SST model removes clear spanwise peaks that are present in both the SST-GDH and SST-HOGGDH models. The augmented SST model begins to improve the wall ACE with improved coolant spread and cooling distribution in the region just aft of the coolant hole.

The discussion surrounding the results of Fig. 12 is shown qualitatively in Fig. 14. Contours of the ML-RF model show improved predictions in the region just aft of the coolant hole, compared to the HOGGDH closure that over-predicts the cooling impact. Downstream, after some length of mixing with the mainstream gas, the results of the ML-RF model are in close alignment with the ML-HOGGDH with improved predictions of lateral coolant distributions that reflect the LES results. Although the ML-RF model makes improvements, these contours show that the data-driven turbulent anisotropy model makes the most significant improvement.

3.2.3. Case *br06-dr12*

A further case at a density ratio of 1.2 is investigated at an increased blowing ratio of 0.6. This case was used to test the random forest turbulent heat flux coefficient model but did not influence the construction of the model as a training or validation dataset would do. The results of the random forest model in Fig. 4 showed the turbulent heat flux magnitude closely matched the LES compared to the standard HOGGDH model. Components of the turbulent heat flux vector were shown to deviate in the downstream region because of the functional form of the base HOGGDH model. The results below compared ACE results between the three models to the LES predictions of coolant distribution.

Fig. 15 shows significant improvements using the HOGGDH and random forest model in the initial coolant jet development region ($X < 7D$). Within this region, the secondary peak associated with the coolant jet reattachment is observed in agreement with the LES at $X = 5D$, whereas the GDH model does not predict this peak until $10D$ downstream of the coolant trailing edge. Differences between the HOGGDH and random forest model are small but minor improvements are exhibited by the random forest model. Lateral spreading, predicted by the random forest model, is in better agreement with the LES in the

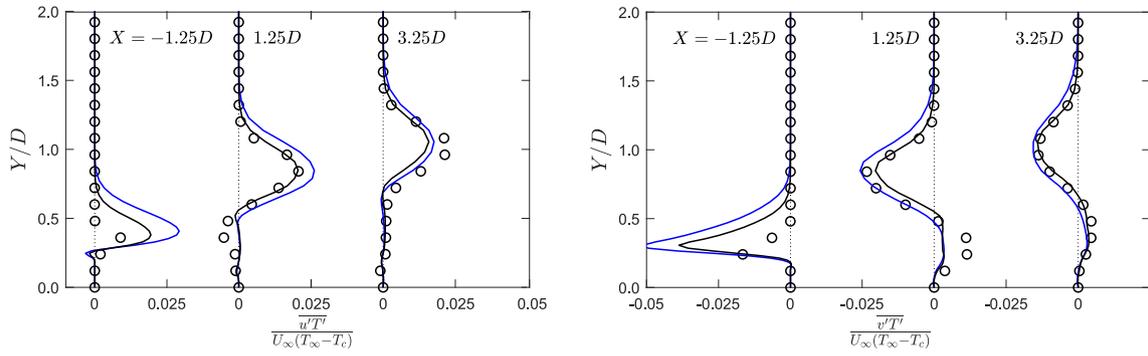


Fig. 9. Turbulent heat flux components of ML-HOGGDH [—] and ML-RF [—] augmented models across centreline profiles compared to LES resolved components [○]. Case *br10-dr20*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

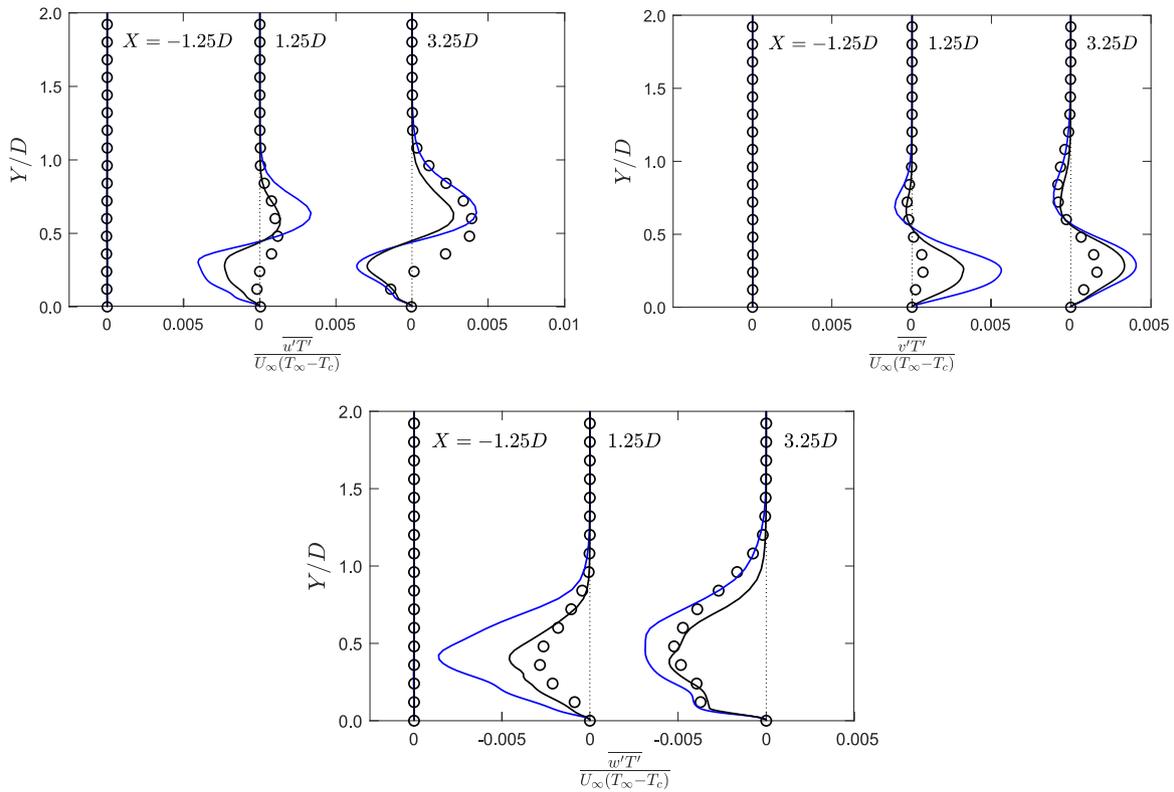


Fig. 10. Modelled turbulent heat flux vector components of ML-HOGGDH [—] and ML-RF [—] augmented models across off-centrelines ($Z = 1D$) compared to LES resolved components [○]. Case *br10-dr20*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

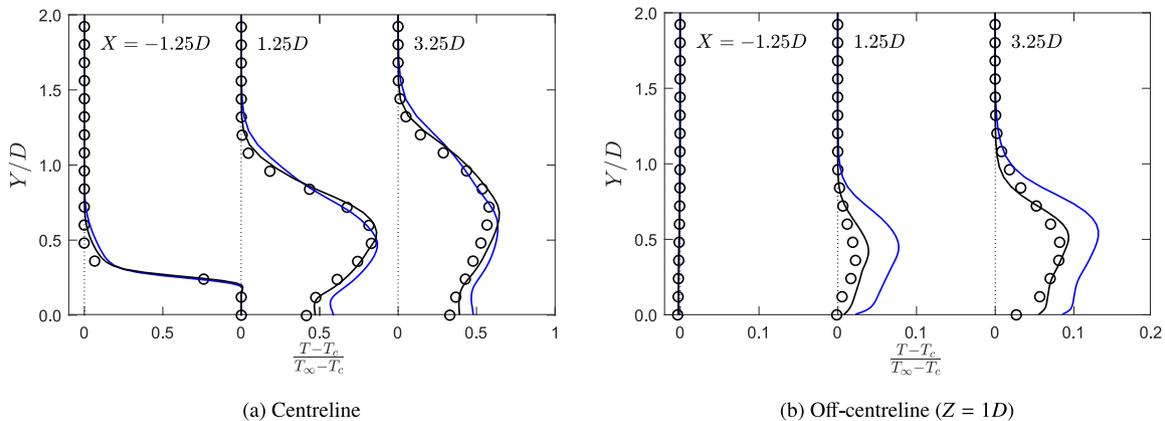


Fig. 11. Non-dimensional temperature profiles comparing ML-HOGGDH [—] and ML-RF [—] augmented models with the LES predictions [○]. Case *br10-dr20*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

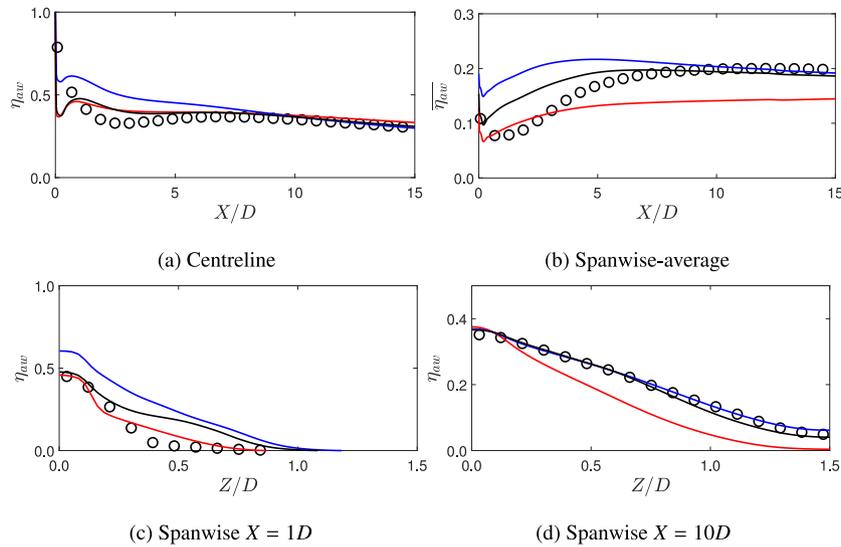


Fig. 12. ACE comparison of ML-GDH [—], ML-HOGGDH [—] and the ML-RF model [—] with LES data [•]. Case *br10-dr20*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

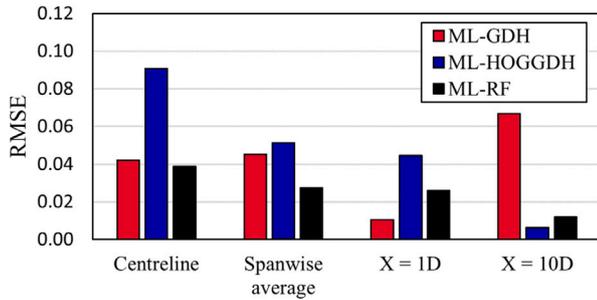


Fig. 13. RMSE of the ACE across the centreline, spanwise average and lateral profiles presented in Fig. 12. Case *br10-dr20*.

initial jet development region which is shown in the lateral distribution at $X = 1D$ and the spanwise-averaged ACE results. Fig. 16 summarises the small improvements of the combined model with the spanwise-averaged and lateral results demonstrating improvements over the standard HOGGDH model. However, the downstream regions show underestimated centreline ACE which in turn reduces the observed spanwise-averaged ACE than predicted by the LES flow, which is also illustrated by the root-mean-square error (Fig. 16). As shown in case *br05-dr20*, the difference between the ML-RF and ML-HOGGDH results is minimal but offers reassurance that the model does not adversely affect the flow field when used in novel cases.

4. Conclusions

A model has been developed for a spatially varying diffusivity coefficient, c_θ , for the HOGGDH closure of the turbulent heat flux using data-driven machine learning techniques: random forests and shallow neural networks. The model of c_θ was trained on two film cooling cases and tested upon a third case.

The random forest algorithm and the shallow neural network approach showed predictions with good agreement to the LES resolved data in the testing case, demonstrating its ability to generalise to this different coolant flow condition. Non-linearities in the profiles were well captured by the model demonstrating its ability to capture complexities in the flow. Predictions of c_θ provided improved profiles of the turbulent heat flux magnitude compared to the commonly used

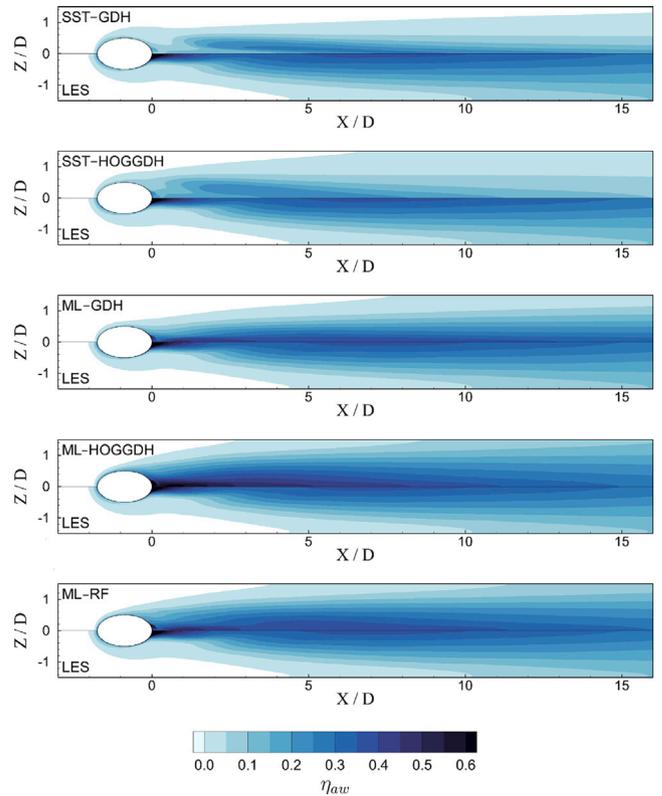


Fig. 14. Contours of ACE across the downstream plate comparing standard to augmented SST models with the GDH, HOGGDH published by Ellis and Xia [24] (reprinted under terms of the CC BY 4.0 license), and the combined augmented SST and RF model (ML-RF) for Case *br10-dr20*.

constant of $c_\theta = 0.6$ and indicated the advantage of a spatially varying coefficient.

Implementation of the random forest model was achieved in OpenFOAM and used alongside the data-driven turbulent anisotropy model developed by Ellis and Xia [24]. Results showed cooling effectiveness improvements when coupled together. The random forest model reduced the cooling effectiveness RMSE significantly in all but one cooling profile. Marginal improvements were made compared to the

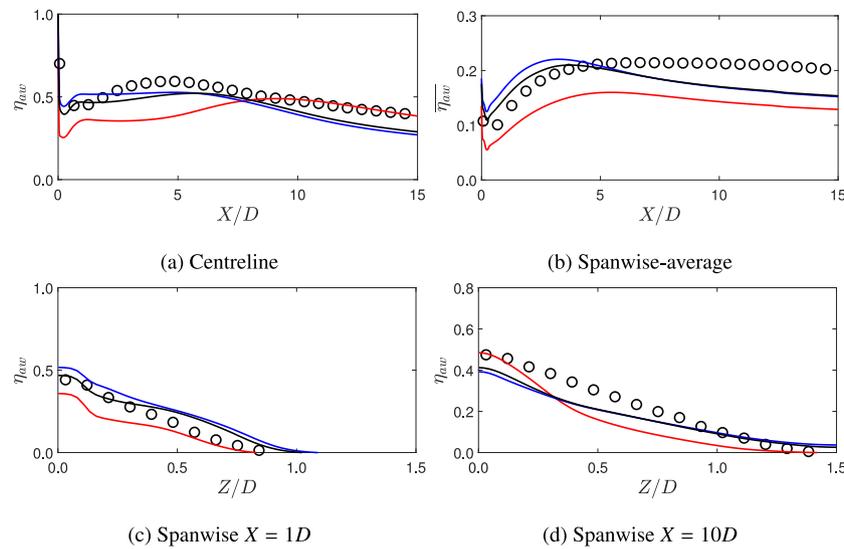


Fig. 15. ACE comparison of ML-GDH [—], ML-HOGGDH [—] and the ML-RF model [—] with LES data [●]. Case *br06-dr12*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

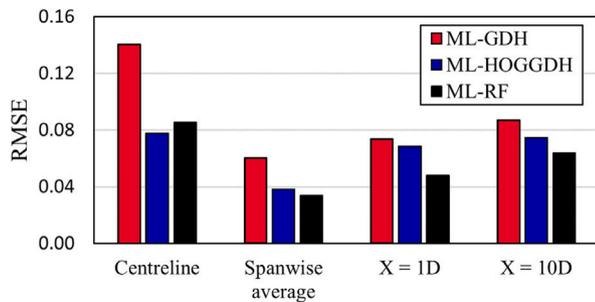


Fig. 16. RMSE of the ACE across the centreline, spanwise average and lateral profiles presented in Fig. 15. Case *br06-dr12*.

HOGGDH model which had the same closure structure without the machine learnt coefficient. Centreline and off-centreline profiles of non-dimensional temperature and turbulent heat flux showed that the random forest model combined with the data-driven turbulent anisotropy model was close to replicating the LES datasets. Although improvements were exhibited with the random forest model, the contours of cooling effectiveness showed the biggest improvement was made using the data-driven turbulent anisotropy model.

Following these studies, a variety of further work has been identified that would be beneficial for industrial usage of such approaches. First, future work should look to address a wider range of industrially relevant conditions improved by ML models. Second, new models on industrially relevant cases should investigate the incorporation of sparse experimental datasets. Finally, the model interpretability should be investigated to highlight their performance and future simplification that would aid industrial adoption.

CRedit authorship contribution statement

Christopher D. Ellis: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hao Xia:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Christopher D. Ellis reports financial support was provided by Rolls-Royce plc. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

C. D. Ellis is supported by the EPSRC Centre for Doctoral Training in Gas Turbine Aerodynamics (Grant No. EP/L015943/1). The authors would like to acknowledge the use of the HPCMidlands+ HPC facility, which was funded by the EPSRC, UK (Grant No. EP/P020232/1).

References

- [1] L. Bertolotti, R. Jefferson-Loveday, S. Ambrose, E. Korsukova, High-fidelity CFD-trained machine learning to inform RANS-modelled interfacial turbulence, *J. Glob. Power Propuls. Soc.* 7 (2023).
- [2] Y. Frey Marioni, E.A. de Toledo Ortiz, A. Cassinelli, F. Montomoli, P. Adami, R. Vazquez, A machine learning approach to improve turbulence modelling from dns data using neural networks, *Int. J. Turbomach. Propuls. Power* 6 (2) (2021) 17.
- [3] J. Ling, J. Templeton, Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier-Stokes uncertainty, *Phys. Fluids* 27 (085103) (2015).
- [4] J. Weatheritt, R. Sandberg, A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship, *J. Comput. Phys.* 325 (2016) 22–37.
- [5] J. Weatheritt, R.D. Sandberg, The development of algebraic stress models using a novel evolutionary algorithm, *Int. J. Heat Fluid Flow* 68 (September) (2017) 298–318.
- [6] C.D. Ellis, H. Xia, Turbulent closure analysis in heated separated and reattached flow using eddy-resolving data, *Phys. Fluids* 32 (4) (2020) 045115.
- [7] J.-C. Han, S. Dutta, S. Ekkad, *Gas Turbine Heat Transfer and Cooling Technology*, CRC Press, 2012.
- [8] A. Hoda, S. Acharya, Predictions of a film coolant jet in crossflow with different turbulence models, *J. Turbomach.* 122 (1) (2000) 558–569.
- [9] S. Acharya, M. Tyagi, A. Hoda, Flow and heat transfer predictions for film cooling, *Ann. New York Acad. Sci.* 934 (1) (2001) 110–125.

- [10] D.K. Walters, J.H. Leylek, A systematic computational methodology applied to a three dimensional film-cooling flowfield, in: International Gas Turbine & Aeroengine Congress & Exhibition, 1996.
- [11] D.K. Walters, J.H. Leylek, A detailed analysis of film-cooling physics Part 1: Streamwise injection with cylindrical holes, in: International Gas Turbine & Aeroengine Congress & Exhibition, 1997.
- [12] K.L. Harrison, D.G. Bogard, Comparison of RANS turbulence models for prediction of film cooling performance, in: Turbo Expo: Power for Land, Sea and Air, Vol. 43147, 2008, pp. 1187–1196.
- [13] A. Azzi, B.A. Jubran, Numerical modeling of film cooling from short length stream-wise injection holes, *Heat Mass Transf.* 39 (2003) 345–353.
- [14] G. Bergeles, A.D. Gosman, B.E. Launder, G. Bergeles, The turbulent jet in a cross stream at low injection rates : A three-dimensional numerical treatment, *Numer. Heat Transfer* 1 (1978) 217–242.
- [15] M. Tyagi, S. Acharya, Large eddy simulation of film cooling flow from an inclined cylindrical jet, *J. Turbomach.* 125 (October 2003) (2003) 734–742.
- [16] P. Renze, W. Schroder, M. Meinke, Large-eddy simulation of film cooling flows with variable density jets, *Flow Turbul. Combust.* 80 (2008) 119–132.
- [17] P. Renze, W. Schroder, M. Meinke, Large-eddy simulation of film cooling flows at density gradients, *Int. J. Heat Fluid Flow* 29 (2008) 18–34.
- [18] J. Bodart, F. Coletti, I. Bermejo-Moreno, J.K. Eaton, High-fidelity simulation of a turbulent inclined jet in a crossflow, *Ctr.Stanford.Edu* (2013) 263–275.
- [19] C. Ellis, H. Xia, Impact of inflow turbulence on large-eddy simulation of film cooling flows, *Int. J. Heat Mass Transfer* 195 (2022) 123172.
- [20] P.M. Milani, J. Ling, G. Saez-Mischlich, J. Bodart, J.K. Eaton, A machine learning approach for determining the turbulent diffusivity in film cooling flows, *J. Turbomach.* 140 (2) (2017) 021006.
- [21] P.M. Milani, J. Ling, J.K. Eaton, Turbulent scalar flux in inclined jets in crossflow: Counter gradient transport and deep learning modelling, 2020, arXiv preprint arXiv:2001.04600.
- [22] J. Ling, A. Kurzawski, J. Templeton, Reynolds averaged turbulence modelling using deep neural networks with embedded invariance, *J. Fluid Mech.* (2016).
- [23] J. Ling, A. Ruiz, G. Lacaze, J. Oefelein, Uncertainty analysis and data-driven model advances for a jet-in-crossflow, *J. Turbomach.* (2016).
- [24] C.D. Ellis, H. Xia, Data-driven turbulence anisotropy in film and effusion cooling flows, *Phys. Fluids* 35 (10) (2023) 105114, <http://dx.doi.org/10.1063/5.0166685>.
- [25] P.M. Milani, J. Ling, J.K. Eaton, Physical interpretation of machine learning models applied to film cooling flows, *J. Turbomach.* 141 (1) (2019) 011004.
- [26] C.D. Ellis, H. Xia, G.J. Page, LES informed data-driven modelling of a spatially varying turbulent diffusivity coefficient in film cooling flows, in: Turbo Expo: Power for Land, Sea, and Air, Vol. 84171, American Society of Mechanical Engineers, 2020, V07BT12A029.
- [27] OpenCFD Ltd 2019, OpenFOAM v17.12, 2017, <https://www.openfoam.com/releases/openfoam-v1712/>.
- [28] J. Ling, K.J. Ryan, J. Bodart, J.K. Eaton, Analysis of turbulent scalar flux models for a discrete hole film cooling flow, *J. Turbomach.* 138 (1) (2015).
- [29] D.C. Wilcox, et al., DCW industries La Canada, CA, 1998.
- [30] A.K. Sinha, D.G. Bogard, M.E. Crawford, Film-cooling effectiveness downstream of a single row of holes with variable density ratio, *J. Turbomach.* 113 (1991) 442–449.
- [31] L. Andrei, A. Andreini, C. Bianchini, G. Caciolli, L. Mazzei, A. Picchi, F. Turrini, Effusion cooling plates for combustor liners : experimental and numerical investigations on the effect of density ratio, *Energy Procedia* 45 (2014) 1402–1411.
- [32] M.C. Immer, Time-resolved measurement and simulation of local scale turbulent urban flow (Ph.D. thesis), ETH Zurich, 2016.
- [33] F. Nicoud, F. Ducros, Subgrid-scale stress modelling based on the square of the velocity gradient tensor, *Flow Turbul. Combust.* 62 (3) (1999) 183–200.
- [34] A. Schindler, B.A. Younis, B. Weigand, Large-eddy simulations of turbulent flow through a heated square duct, *Int. J. Therm. Sci.* 135 (May 2018) (2019) 302–318.
- [35] T.M. Eidson, Numerical simulation of the turbulent Rayleigh-Benard problem using subgrid modeling, *J. Fluid Mech.* 158 (1985) 245–268.
- [36] G. Grötzbach, M. Wörner, Direct numerical and large eddy simulations in nuclear applications, *Int. J. Heat Fluid Flow* 20 (3) (1999) 222–240.
- [37] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [38] L. Breiman, Classification and Regression Trees, Routledge, 2017.
- [39] G. Louppe, Understanding Random Forests, Cornell University Library, 2014, pp. 1–225, arXiv:arXiv:1407.7502v2.
- [40] T. Hastie, R. Tibshirani, J. Friedman, J. Franklin, The elements of statistical learning: data mining, inference and prediction, *Math. Intell.* 27 (2) (2005) 83–85.
- [41] K. Abe, K. Suga, Towards the development of a Reynolds-averaged algebraic turbulent scalar-flux model, *Int. J. Heat Fluid Flow* 22 (1) (2001) 19–29.
- [42] F.R. Menter, M. Kuntz, R. Langtry, Ten years of industrial experience with the SST turbulence model, *Turbul. Heat Mass Transfer* 4 (1) (2003) 625–632.
- [43] J. Wu, H. Xiao, R. Sun, Q. Wang, Reynolds-averaged Navier-Stokes equations with explicit data-driven Reynolds stress closure can be ill-conditioned, *J. Fluid Mech.* 869 (2019) 553–586, arXiv:arXiv:1803.05581v3.
- [44] B. Basara, S. Jakirlic, A new hybrid turbulence modelling strategy for industrial CFD, *Int. J. Numer. Methods Fluids* 42 (1) (2003) 89–116.