GT2020-14927

MACHINE LEARNING FOR THE DEVELOPMENT OF DATA DRIVEN TURBULENCE CLOSURES IN COOLANT SYSTEMS

James Hammond[*]
Francesco Montomoli
Marco Pietropaoli
UQLab, Department of Aeronautics
Imperial College London
London (UK), SW7 2AZ
Email: j.hammond18@imperial.ac.uk

Richard D. Sandberg
Department of Mechanical Engineering
Univeristy of Melbourne
Parkville, Victoria 3010
Australia

Vittorio Michelassi
Baker Hughes
Florence, 50127
Italy

ABSTRACT

This work shows the application of Gene Expression Programming to augment RANS turbulence closure modelling for flows through complex geometries, designed for additive manufacturing. Specifically, for the design of optimised internal cooling channels in turbine blades. One of the challenges in internal coolant design is the heat transfer accuracy of the RANS formulation in comparison to higher fidelity methods, which are still not used in design on account of their computational cost. However, high fidelity data can be extremely valuable for improving current lower fidelity models and this work shows the application of data driven approaches to develop turbulence closures for an internally ribbed duct. Different approaches are compared, and the results of the improved model are illustrated.

The work shows the potential of using data driven models for accurate heat transfer predictions even in non-conventional configurations.

NOMENCLATURE

a_{ij} Anisotropic component of Reynolds stress tensor.
I_n Scalar invariants.
k Turbulent kinetic energy.
k' Turbulent kinetic energy threshold.
S'_{ij} Deviatoric component of strain rate tensor.

u'_i Fluctuating component of velocity vector.
V^n_{ij} Tensor bases.
β_n Coefficient of tensor basis V^n_{ij}.
δ_{ij} Kronecker delta.
ε Turbulence dissipation rate.
ν_t Turbulent (eddy) viscosity.
$\tau_{s,w}$ Turbulent timescales.
Ω_{ij} Rotation rate tensor.
$\mathscr{J}$ Cost functional.

INTRODUCTION

The drive for improved cycle efficiency in modern day gas turbines is necessitating higher entry temperatures. Improved cooling solutions are required to deal with these rising temperatures and maintain turbine blade longevity, without jeopardizing the cycle efficiency improvement with the efficiency penalty associated with large cooling flows [1]. Until recently however, internal cooling channel designs have had their complexity limited by the absence of manufacturing techniques capable of fabricating complex internal structures. Now, additive manufacturing (AM) has emerged as a viable alternative to traditional manufacturing techniques in the the aerospace industry [2]. This has enabled design methods such as Topology Optimisation (TO) to exploit the revolutionary flexibility that AM techniques unlock for complex design geometries [3].

---

[*]Address all correspondence to this author.

Increased complexity in cooling channel geometry to increase heat removal without excessive head losses, inevitably leads to more complex cooling fluid flows [4]. These flows can exhibit strong rotation, streamline curvature and separation; all of which are known to be poorly predicted by Reynolds-Averaged Navier-Stokes (RANS) methods with linear constitutive laws [5]. Despite this, RANS methods are still prevalent in industrial design, as higher fidelity methods such as Direct Numerical and Large Eddy Simulations (DNS/LES) remain out of reach from the perspective of computational requirements [6].

Machine Learning (ML) methods are being recognised throughout the fluid mechanics community as a possible avenue to improve the prediction of transitional and turbulent flows. The quantity of high fidelity data is growing exponentially and ML methods aim to utilise this data to augment current models and even develop new ones. Many directions are being taken for the application of ML and recent reviews from Durbin [7] and Duraisamy [8] provide full accounts. However, in this work the focus will be turbulence closure modelling. Specifically, data from high fidelity simulation (in this case DES) is used to train RANS turbulence models to mimic the high fidelity behaviour, whilst retaining the lower computational cost.

As is true for classical closure modelling, data driven approaches focus on improving the prediction of the Reynolds stresses and turbulent heat transfer, which are modelled as functions of mean flow variables. Wang et al. [9] and subsequently Wu et al. [10] aimed to provide a comprehensive framework for physics informed machine learning approaches to turbulence models. Their approach uses the random forest method to obtain discrepancy functions between the high fidelity Reynolds stresses and RANS predictions. In predictive cases, these discrepancy functions could then be applied to the RANS models to correct the Reynolds stresses. In this sense they act as a corrective patch, rather than aiming to provide a suitable form for the Reynolds stresses *a priori*. Ling et al. [11] used deep neural networks (DNNs) to learn the Reynolds stresses directly with reasonable predictive performance. However DNNs represent a black box approach and gives no explicit expression for the Reynolds stresses as a result. In contrast, Weatheritt and Sandberg [12] combined a symbolic regression approach with the framework of algebraic stress models to learn explicit functional forms for the Reynolds stresses that can be implemented directly into RANS solvers. For each of the mentioned works predictive results on similar cases are promising, but a high level of uncertainty is introduced for simulations further removed from the training case. This is due to the difficulty of identifying training data-sets general and representative enough to cover an engineering relevant design space, as suggested by Akolekar et al [13].

In all of the above cases, as well as in much of the available literature (examples including [14–16]), ML methods are applied to standard two dimensional geometries with only tentative extensions to simple three dimensional geometries [17] and more recently [18, 19], the latter of which only investigates heat flux.

This paper looks to extend ML methods to non-trivial, three dimensional geometries by applying them to an internally ribbed square duct, which is the product of a previously conducted TO.

The data driven aspect will follow the symbolic regression framework laid out by Weatheritt and Sandberg [16] based on Gene Expression Programming (GEP). The aim is to develop explicit expressions for the anisotropic component of the Reynolds stresses, which lead to improved predictions over the standard k-ε RANS model.

METHODOLOGY
Algebraic Stress Model
Many of the issues associated with RANS modelling can be traced back to the Boussinesq hypothesis [5]. Boussinesq assumed a linear relationship between the anisotropic components of Reynolds stresses and the deviatoric strain-rate tensor. The relationship is realised by the introduction of a proportionality coefficient known as the turbulent (eddy) viscosity and forms the basis of eddy viscosity models (EVMs) for turbulence closures. Defining the non-dimensional anisotropy as

$$a_{ij} = \frac{\overline{u'_i u'_j}}{2k} - \frac{1}{3}\delta_{ij}, \tag{1}$$

this linear relationship is seen by $2ka_{ij} = -2\nu_t S'_{ij}$. In contrast to EVMs, algebraic stress models (ASMs) are a distinct class of turbulence closure, which attempt to develop expressions for the Reynolds stresses by making approximations to their exact differential equations. Pope [20] argued on dimensional grounds that anisotropy is purely a function of the strain and rotation rate tensors, and the turbulent kinetic energy and dissipation rate, $a_{ij} = a_{ij}(S'_{ij}, \Omega_{ij}, k, \varepsilon)$. By means of a weak equilibrium hypothesis [21] and application of the Cayleigh-Hamilton theorem, the most general form for the anisotropy can then be written as the finite sum of 10 linearly independent tensor bases V^n_{ij} for $n = \{1, 2, ..., 10\}$

$$a_{ij} = \sum_{n=1}^{10} \beta_n V^n_{ij}. \tag{2}$$

These tensor bases are each functions of the non-dimensional strain and rotation rate tensors $V_{ij} = V_{ij}(s_{ij}, w_{ij})$, where $s_{ij} = \tau_s S'_{ij}$ and $w_{ij} = \tau_w \Omega_{ij}$ and τ_s and τ_w are turbulent timescales defined in this work as the inverse of the Frobenius norms of the rate tensors

$$\tau_s = \frac{1}{\sqrt{S_{ij}S_{ij}}}, \quad \tau_w = \frac{1}{\sqrt{\Omega_{ij}\Omega_{ij}}}. \tag{3}$$

Defining s_{ij} and w_{ij} in this way ensures that V_{ij}^n is an integrity basis for a_{ij} which is necessary to force Galilean invariance [22].

For general three-dimensional flows a five term approximation to the full explicit sum of Eqn. (2) is sufficient to determine the independent components of anisotropy without over-constraint. To explain this, imagine the anisotropy is known; given it is a symmetric rank-2 tensor of known trace, each of the coefficients in a five term sum could be solved for exactly [23]. The terms considered then are simply the first five tensor bases

$$V_{ij}^1 = s_{ij} \qquad\qquad V_{ij}^2 = s_{ik}s_{kj} - \tfrac{1}{3}I_1\delta_{ij}$$
$$V_{ij}^3 = w_{ik}w_{kj} - \tfrac{1}{3}I_2\delta_{ij} \qquad V_{ij}^4 = s_{ik}w_{kj} - w_{ik}s_{kj}$$
$$V_{ij}^5 = s_{ik}s_{kl}w_{lj} - w_{ik}s_{kl}s_{lj}.$$

Higher order bases are neglected by setting coefficients $\beta_{6,...,10} = 0$. The variables I_1 and I_2 appearing above are two of five scalar invariants which also arise from the Cayleigh-Hamilton theorem. These are given by

$$I_1 = s_{mn}s_{nm} \qquad I_2 = w_{mn}w_{nm} \qquad I_3 = s_{mn}s_{nk}s_{km}$$
$$I_4 = s_{mn}w_{nk}w_{km} \quad I_5 = s_{mn}s_{nk}w_{kl}w_{lm}.$$

The β_n coefficients are themselves functions of these scalar invariants and explicit expressions for the anisotropy are discovered by the data driven framework (which will be discussed in the next section) by finding the form of these functions. Resulting expressions can then be inserted directly into the RANS equations as a non-linear extension of the Boussinesq approximation for simulation and testing.

Data Driven Technique

The data driven technique used to discover non-linear expressions for anisotropy is built upon the framework of Gene Expression Programming (GEP). GEP is an evolutionary algorithm which uses symbolic regression to discover explicit expressions in terms of a set of input variables, which minimise a predefined cost function. In this work the cost function is defined as the mean square error of the anisotropy between the high fidelity data and a candidate expression created by GEP

$$\mathscr{J}(a_{ij}^{\text{gep}}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{3}\sum_{j=1}^{3}(a_{ij} - a_{ij}^{\text{gep}})^2. \qquad (4)$$

The explicit expressions are implemented programmatically using chromosomes which are simply linearly encoded strings. These chromosomes can then be read as expression trees (ETs) which in turn are read as mathematical expressions. Take an example chromosome

$$\texttt{*-I1I3s|I2I5I1I4I3I2,}$$
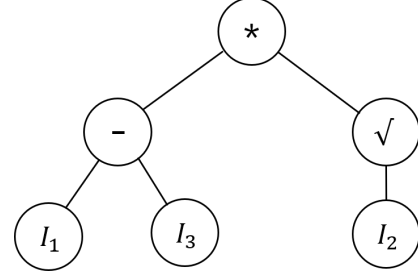
FIGURE 1: EXAMPLE EXPRESSION TREE (ET) FROM THE CHROMOSOME: `*-I1I3s|I2I5I1I4I3I2`.

when parsing this chromosome to create an ET, subsequent symbols are passed as arguments to preceding symbols. Variables and constants take no arguments and thus exist at the extremities of ETs. Two distinct regions of the chromosome exist; the head and the tail, separated by the '|' symbol. The head can consist of any operators, variables or constants whereas the tail is populated solely by symbols which take no arguments (variables and constants). This distinction guarantees that when reading the chromosome as an ET, there are enough variables and constants to fill the extremities of the tree and guarantee syntactic validity [24]. Reading the example chromosome from left to right the ET in Figure (1) can be constructed. Notice how symbols in the tail have been left unused but could be introduced if more symbols taking arguments appeared in the head.

The ET in Figure (1) can then be read recursively from top to bottom and from left to right to generate the mathematical expression represented by the example chromosome

$$(I_1 - I_3)\sqrt{I_2}.$$

To minimise the cost function of Eqn. (4), GEP simulates evolution over a number of generations. An initial population of N chromosomes is randomly generated and the fitness of each of the resulting mathematical expressions evaluated from the cost function. Chromosomes are then randomly selected (with replacement) to compete in tournaments of size $m << N$, simulating the process of natural selection. The fittest individual in each tournament takes one space in the following generation until it is populated with $N-1$ chromosomes. The final spot is reserved for the fittest individual in the previous generation to ensure it survives even if it has not been selected to compete in a tournament.

Following the tournament selection process, the surviving chromosomes are then stochastically subjected to imitations of genetic operations which introduce variation into the population. Two of the main operations are mutation and crossover. Mutation randomly changes one symbol in a chromosome, for example the 'I3' in the head of the example chromosome might be

changed to a '+'. The effect of this mutation on the resulting mathematical expression will have more weighting the closer the mutation occurs to the start of the chromosome. Crossover is a process which mimics reproduction. Portions of genetic material from two 'parent' chromosomes, are swapped randomly to create offspring with genetic material from both parents. Following application of genetic operations, the resulting population is then subjected to tournament selection and the evolutionary process continues, iterating over subsequent generations.

The result of the GEP algorithm is a guided search towards chromosomes which represent mathematical expressions that minimise $\mathcal{J}(a_{ij})$. In this way, new functional forms not specified by the user are uncovered by the evolutionary process. Evolution continues until either: a prescribed number of generations has occurred or; the cost function is below some specified tolerance. A full description of the algorithm can be found by Weatheritt and Sandberg [12] and in the book by Candida Ferreira who first introduced the algorithm [24].

Extension to Tensor Regression. The algorithm above applies only to problems involving scalar regression. In the current work tensorial regression is necessary to discover expressions for the anisotropy. Simply introducing tensor variables and constants to the standard algorithm can result in syntactically invalid, or at the least ambiguous, expressions. Take the example

$$T_{ij} = aA_{ij} + bB_{ij}$$
$$T_{ij} = aA_{ij} + bb,$$

where the tensor term B_{ij} mutates into a scalar b. Either $b \in \mathbb{R}$ which yields a dimensionally invalid expression, or the definition of multiplication can be made variable depending on its arguments. The term bb could be interpreted as $bb\delta_{ij}$ for example. However, this methodology is extremely inefficient because patches to fix such ambiguities require a large amount of overhead code.

Weatheritt and Sandberg introduced the concept of the plasmid as a natural way to perform tensorial regression, whilst guaranteeing syntactic validity of expressions. The plasmid methodology is also adopted in this work and will be explained here in brief, for a full description of the method the reader is referred to [12]. The full set of variables is split into a 'host' set, consisting only of the tensor variables present in the regression, and an additional symbol 'p' to denote the occurrence of a plasmid. This 'p' symbol acts as a reference to an expression formed from an entirely separate 'plasmid' population. The 'plasmid' population consists of the remaining scalar variables present in the regression as well as any constants. A host chromosome with
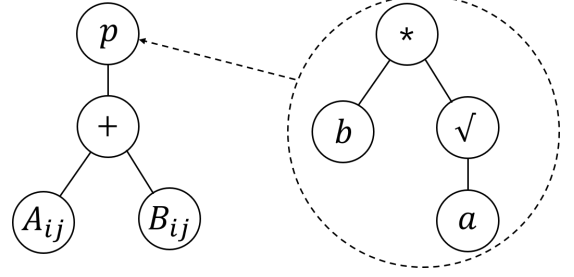
FIGURE 2: EXAMPLE HOST EXPRESSION TREE (ET) WITH ASSOCIATED PLASMID STRUCTURE.

one plasmid may, for example, take the form

```
p+AB*|BAABBB
*bsaa|aabbab.
```

Figure (2) shows the ET implementation where the plamsmid is shown in the black dotted circle. To evaluate this ET, the plasmid structure is placed where the 'p' symbol appears and multiplies itself by the argument. The mathematical expression of this ET can thus be written as

$$b\sqrt{a}(A_{ij} + B_{ij}).$$

Plasmids are subjected to the same genetic operation as the hosts. For every new occurrence of a 'p' symbol in the host, a plasmid is generated and tied to it. Equally, whenever a host is removed by natural selection or a 'p' removed by some genetic operation, the associated plasmid is removed.

Flow Configuration and Training Data

The geometry considered in this work is the result of a previously conducted TO within a square duct [3, 25]. The optimisation was aimed at maximising heat transfer to the duct walls whilst minimising pressure drop and the internal ribbed geometry produced is shown in Figure (3).

The high fidelity training data to be used in the GEP algorithm was extracted from DES. Figure (4) shows streamlines for flow through the duct coloured by the sense of the axial vorticity, resulting from the DES. These streamlines clearly highlight the pairs of counter rotating vortices created by the internal geometry and show the strong streamline curvature present in the flow. Streamline curvature is a well documented weakness of RANS calculations based on linear constitutive laws, making this a case well suited to the testing of data driven closures.

Whilst DNS/LES data would have been preferable to DES data for use as training data as they are more accurate representations of reality, it is considered that for a proof of concept the

FIGURE 3: TOPOLOGY OPTIMISED INTERNAL GEOMETRY FOR SQUARE DUCT SHOWING RIBBED STRUCTURE.
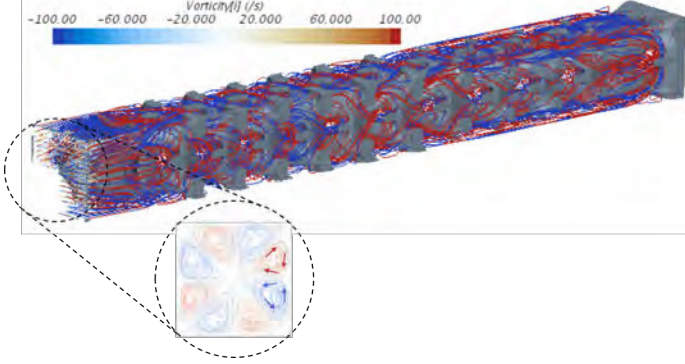

FIGURE 4: STREAMLINES AROUND INTERNAL RIBS FROM DES, COLOURED BY SENSE OF THE VORTICITY, HIGHLIGHTING COUNTER-ROTATING VORTICES GENERATED BY GEOMETRY.

exact nature of the training data is not critical. As long as the methodology can improve the approximation of the training data provided, then it can be considered successful regardless of how closely the training data approximates reality.

The flow considered for the DES has an inlet Reynolds number of 1×10^5 and the turbulence intensity was set to 5% with a viscosity ratio of 10.0. The inlet temperature is $600\,\text{K}$ with a constant wall temperature of $1000\,\text{K}$ and an adiabatic boundary condition on the outlet. The mesh is unstructured with polyhedral elements and a total of $5,643,681$ cells and the subgrid-scales are calculated using the k-ω SST model.

The mean and fluctuating components of velocity, U_i and u'_i were extracted directly from the DES and from this the anisotropy is determined using its definition in Eqn. (1). The symmetric strain rate S'_{ij} and anti-symmetric rotation rate Ω_{ij} tensors were also extracted directly from the DES allowing calculation of the full set of tensor bases and scalar invariants.

Producing data-driven closures in this way, using solely data from the high-fidelity simulation is termed passive learning. It ensures that the entire learning process can be performed *a priori* without the need to incorporate RANS calculations into the optimisation loop, which can dramatically reduce computational efficiency in non-trivial three-dimensional simulation cases. The
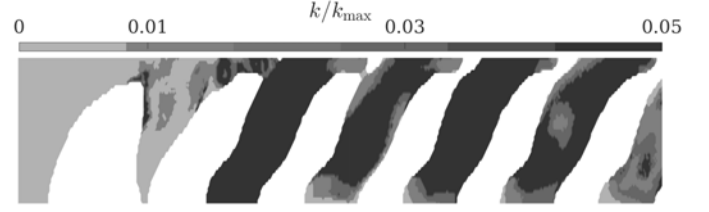
FIGURE 5: SECTION PLANE OF TURBULENCE THRESHOLDS OVER THE INLET REGION AND FIRST FOUR RIBS.

disadvantage of passive learning is that there is no guarantee of how the RANS model will react to the data driven closures, because there is no feedback implemented in the process.

To maintain consistency with the RANS fields, the anisotropy, tensor bases and scalar invariants from DES are interpolated onto the RANS grid. The RANS grid consists of 550 uniformly spaced elements in the *x*-direction and 36 elements in both the *y* and *z*-directions, with higher resolution close to the wall and stretching across the width of the domain ($712,800$ cells in total). Inevitably, some of these cells fall in regions of solid geometry so the total number of cells in the fluid region, and thus the number of data points available to train ASMs using the GEP algorithm is $365,702$.

As the data driven methodology is searching for suitable improvements to turbulence closures it is naturally detrimental to include laminar regions of the flow in the training process. To improve the quality of the training data the points considered are subjected to a threshold of turbulence intensity. Initially three thresholds were considered, each defined as a fraction of the maximum turbulent kinetic energy (TKE): $k' = k/k_{\max} = 0.01$; 0.03 and 0.05. Figure (5) shows these thresholds plotted on an section plane near the inlet region. Only points with k' above the selected threshold are then included in the training data. The inlet region, where the flow is close to laminar and thus has low turbulence intensity, is naturally omitted from the training region by the thresholding process.

Testing and Benchmark Cases

Before testing the methodology as a whole, it is first important to isolate the data driven aspect, to check that based solely on the training data, GEP is producing sensible solutions. This analysis is considered the *a priori* testing of the data driven closures and will be conducted by comparison against three benchmark cases: (1) the standard k-ε RANS turbulence model (which utilises the Boussinesq hypothesis); (2) a local linear regression where β_n coefficients are regressed at every point in the fluid domain and (3) a global linear regression which simply fits a constant value for each β_n coefficient, based on regression over the fluid domain as a whole.

Having determined the correct performance of the data

5

driven component it is then feasible to test the developed closures in the RANS calculations in what is considered the full *a posteriori* testing. For this level of testing to be considered successful, candidate closures must hold up in RANS calculations and provide improvements to the prediction of anisotropy over the standard k-ε model. The three same benchmark cases will also be carried over to *a posteriori* testing to compare the performance against simple regression techniques.

Finally, to test the predictive performance, the developed closures will be tested in an unseen 'out-of-sample' case which differs from the training data. The chosen case is the backward-facing step (BFS), as it is simple to evaluate and presents a very different flow situation to the training case. Further details of the set up of the case are given later in the section: Backward-Facing Step Test.

In order to quantify the performance of each of the cases, two performance measures are considered. The first is simply the cost function used in GEP given in Eqn. (4) which measures the square error between the anisotropy from the high fidelity data a_{ij} (considered to be exact), and the anisotropy from the GEP closures a_{ij}^{GEP}. The second is a measure of the mean alignment between the two tensors and is given by

$$-1 \leq \frac{1}{N}\sum_{n=1}^{N} \frac{a_{ij}a_{ji}^{\mathrm{gep}}}{\sqrt{a_{kl}a_{lk}a_{pq}^{\mathrm{gep}}a_{qp}^{\mathrm{gep}}}} \leq 1. \qquad (5)$$

where the sum is over all data points and N is the total number of points. A value of -1 indicates complete misalignment between the two tensors, 0 suggests that there is no alignment present (or complete orthogonality) and a value of 1 indicates that the two tensors are fully aligned. The problem of severe tensor misalignment was also scrutinised by Wissink et al. [26] and does not constitute a major problem for linear constitutive laws.

Preliminary tests were carried out to determine which turbulence threshold to take forward to the full study. The threshold for which the least performance gain is seen over the standard k-ε model will be chosen as this is the case in which GEP is working hardest to find improvements. Thus, this is considered the harshest level of testing that can be performed. Three runs of the GEP algorithm were performed for each threshold and Table (1) shows the *a priori* and *a posteriori* results of the fittest from each, as well as results from the k-ε model and the local and global linear regressions.

From Table (1) it is clear that the mean square error reduces for the k-ε model the larger the turbulence threshold. This occurs as when the turbulence intensity is higher, it is closer to isotropic and homogeneous in nature. It is under these conditions where the assumption of a linear relationship between anisotropy and mean strain rate is at its best. This also explains why the global regression and GEP expressions make the smallest im-

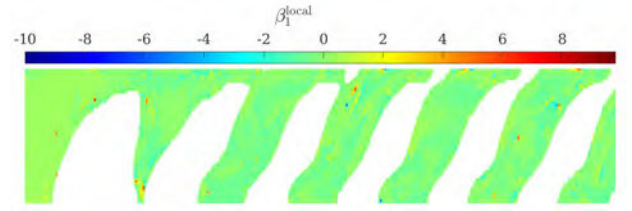|		Mean Square Error			
	k'	k-ε	Local	Global	GEP
a priori	0.00	0.3810	0.0	0.2378	0.2384
	0.01	0.2954	0.0	0.2288	0.2278
	0.03	0.2934	0.0	0.2360	0.2359
	0.05	0.2912	0.0	0.2369	0.2364
a posteriori	0.00	0.4761	-	0.2603	0.2582
	0.01	0.3766	-	0.2892	0.2650
	0.03	0.3752	-	0.3025	0.3039
	0.05	0.3733	-	0.3214	0.3196

TABLE 1: RESULTS FROM PRELIMINARY TESTS.

FIGURE 6: SECTION PLANE OF THE INLET REGION SHOWING LOCALISED SPOTS OF LARGE β_1-COEFFICIENT.

provements over the k-ε model for larger k'. As such, full testing is carried out at the $k' = 0.05$ threshold.

It is also clear from Table (1) that the local linear regression is able to recover the exact anisotropy *a priori*. This result is to be expected given the selection of a five term approximation, as coefficients which yield the exact anisotropy can be solved for uniquely at every point in the domain. However, when introduced to the RANS calculation for *a posteriori* testing the result of the local regression failed to converge. Further investigation revealed localised 'spots' in the domain where the β_n coefficients where more than two orders of magnitude higher than the background level. This is plotted for the β_1-coefficient in Figure (6). It is thought that these spots promote numerical instabilities in the simulation, but their cause is left to the subject of further study and for the time being no *a posteriori* results are available for the local regression.

RESULTS AND DISCUSSION

The GEP algorithm was run 100 times with threshold $k' = 0.05$ to produce a set of candidate equations to be tested. The *a priori* testing analyses the performance of these equations purely as regressions of the training data. Figure (7) shows the evolution of the fittest individual (best fitness) and the mean fitness of the population over 300 generations, taken from the first run of
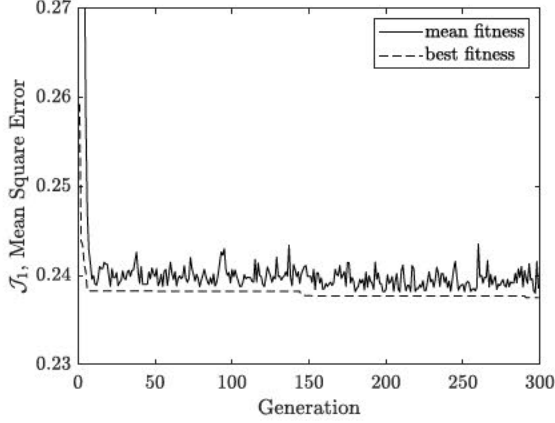
FIGURE 7: EVOLUTION OF THE POPULATION OVER ONE RUN OF THE GEP ALGORITHM.

the algorithm. Initially, there is large improvement in the fittest individual as the algorithm explores the optimisation landscape, which subsequently slows down as the exploration localises. The exploration of the optimisation landscape is highlighted by the high variation in the mean fitness of the population as the best fitness levels off and does not drop any further.

The mean square error and mean alignment for each of the 100 candidate equations are plotted on the histograms in Figure (8). The black line represents the performance of the k-ε model, the red line corresponds to the global linear regression and the blue line shows the performance of the equation obtained from ensemble averaging the full set of GEP solutions. An ensembled equation is used to avoid overfitting of the training data that may occur in individual candidates. The local linear regression is not included here as it has zero error and complete alignment at all points (a mean alignment of 1) so would detract from the clarity of the plot. The expression resulting from the global linear regression is presented below as a reference to analyse components of the GEP solutions

$$a_{ij}^{\text{global}} = -0.139V_{ij}^1 + 0.0194V_{ij}^2 - 0.0158V_{ij}^3$$
$$- 0.191V_{ij}^4 + 0.0223V_{ij}^5. \quad (6)$$

The improvement over the k-ε model is clear in Figure (8), but the similarity between the performance of the global regression and the ensembled GEP equation is interesting to note. In fact, although the ensembled equation is too long to present here, if the values of each of its coefficients $\beta_{1,..,5}$ are averaged across the domain, they are seen to be very similar to that of the global regression in Eqn. (6). These are presented in Table (2). The similarity between averaged coefficients of the ensembled GEP equation and coefficients of the global regression should be expected. The global regression attempts to find the best single
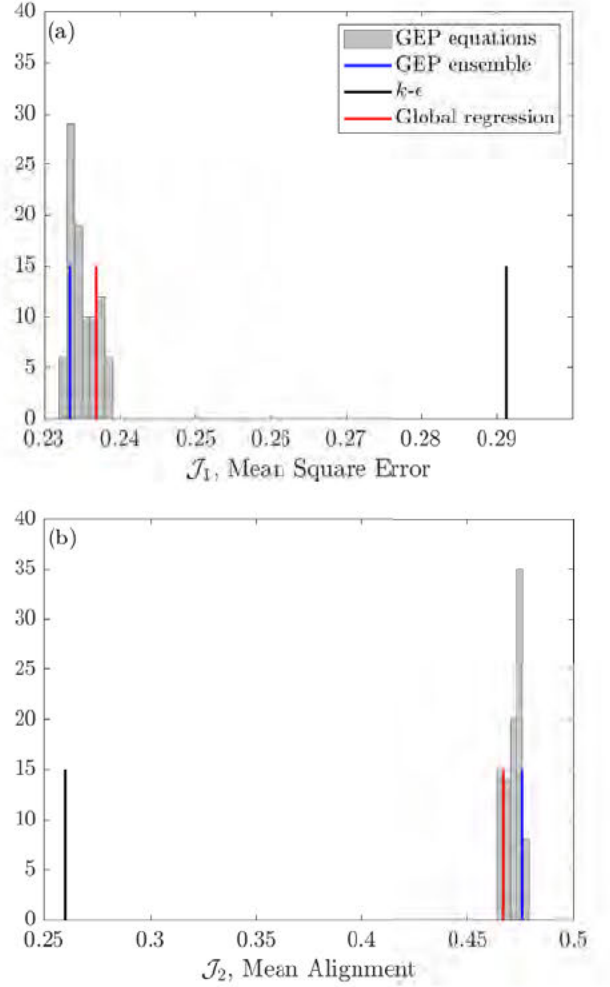
FIGURE 8: HISTOGRAM OF *A PRIORI* (a) MEAN SQUARE ERROR AND (b) MEAN ALIGNMENT OF GEP SOLUTIONS AND PERFORMANCE OF BENCHMARK CASES.

value to describe each coefficient across the whole domain, so it follows that these values should be similar to the average of the coefficient scalar fields discovered by GEP. Where GEP has its advantage over the simple global regression is that it can make use of the scalar invariants, of which the anisotropy is a function, to vary the coefficients across the domain and better fit the anisotropy locally. This explains the improvement of the ensembled GEP equation over the global regression seen in both mean square error and mean alignment. An example equation found by GEP on a single run is given by

$$a_{ij}^{\text{gep}} = 0.132(I_2 + I_4)V_{ij}^1 + (1 + I_5)^2(I_3 + I_4)I_1 I_5 V_{ij}^2$$
$$- (I_5^2 + 0.170I_4)V_{ij}^4. \quad (7)$$

7

β_n	Average Value
1	-0.1312
2	0.0147
3	-0.0052
4	-0.1765
5	0.0001

TABLE 2: AVERAGE β_n COEFFICIENTS FROM ENSEMBLE AVERAGED GEP.

The coefficient of the linear term is again, similar to that found for the global regression in Eqn. (6) and this trend is followed over all 100 individuals. However, coefficients of higher order bases are less predictable. It is likely that more complex expressions arise here to minimise errors introduced by inaccuracies in the lower order terms. This indicates that there is grounding in a baseline linear relationship between anisotropy and mean strain rate, upon which higher order non-linear terms can provide corrections.
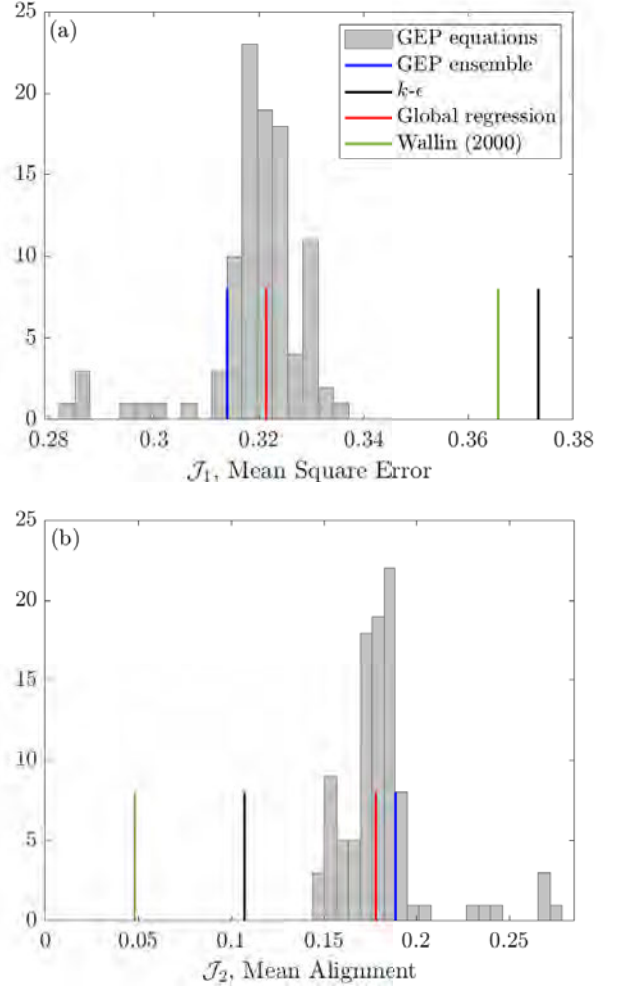
A Posteriori Performance

With the success of *a priori* testing, the set of 100 GEP solutions can be subjected to full *a posteriori* testing by inserting them into the general ASM and run in the RANS calculation. The results of the mean square error and mean anisotropy for each of the GEP solutions are shown on the histograms in Figure (9). Also plotted are the values for the k-ε model, the global regression and the ASM of Wallin and Johansson (2000) [27]. In direct comparison with the *a priori* results, the values of both cost functions for all models have deteriorated.

The mean square error of the k-ε model has risen from approximately 0.291 to 0.373 whilst the corresponding mean alignment has dropped from 0.260 to 0.107. Similarly for the ensembled GEP equation, the mean square error rose from 0.233 to 0.314 and the mean alignment dropped from 0.476 to 0.189. However, it is apparent that the improvement on the k-ε model observed in *a priori* testing, has largely been preserved in the transition to full testing of the GEP solutions. The average improvement on the k-ε model for mean square error dropped only slightly from 20 % to 16 % and the improvement in mean alignment dropped from 83 % to 76 %.

The results from the ASM of Wallin and Johansson (2000) are also included in this plot to benchmark against an ASM from literature. Whilst the ASM improves upon the mean square error achieved by the k-ε model, every one of the GEP solutions outperforms the ASM (and the k-ε model) in both mean square error and mean alignment, reinforcing the success of the data-driven methodology.

When considering the fittest equations as a result of *a posteriori* testing, it is interesting to note that a different set of GEP

FIGURE 9: HISTOGRAMS OF *A POSTERIORI* (a) MEAN SQUARE ERROR AND (b) MEAN ALIGNMENT, OF GEP SOLUTIONS AND PERFORMANCE OF BENCHMARK CASES.

candidates arise than for *a priori* testing and there is little correlation between *a priori* and *a posteriori* performance. This shows that the fittest individual *a priori* shouldn't be assumed to be the best model, justifying the use of an ensemble averaged GEP solution, which mitigates spurious overfitting introduced in individual candidates.

It has previously been shown that simply improving on the prediction of anisotropy does not guarantee improved velocity predictions [28], which in the case of heat transfer is a quantity of greater interest. Figure (10) shows the 3 components of velocity in the inlet region (plotted only where the turbulence threshold is met) for: (1) the DES data, (2) the k-ε model (3) the global linear regression and (4) the ensembled GEP equation. Other than the
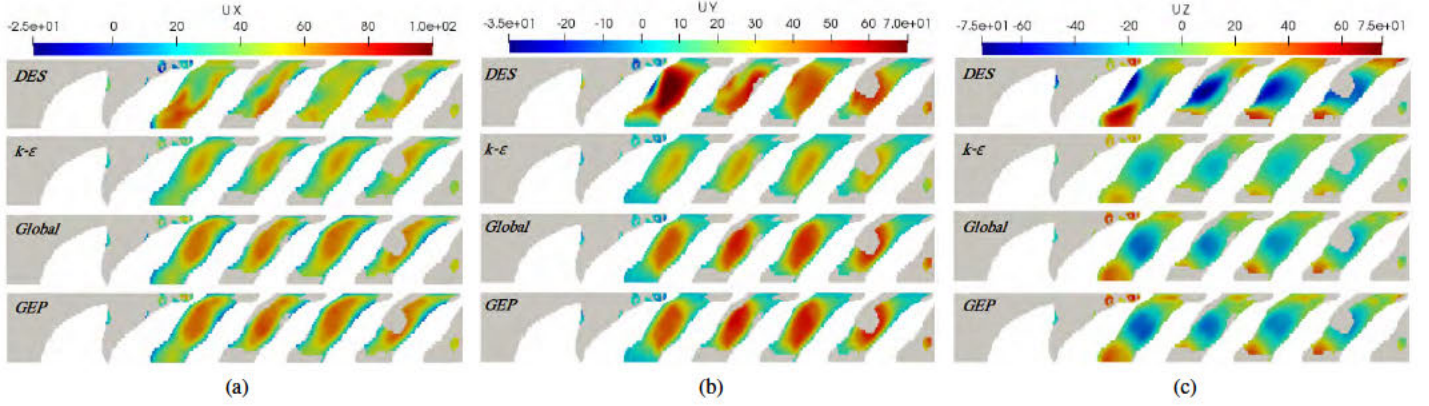
FIGURE 10: CONTOUR PLOTS OF (a) X, (b) Y AND (c) Z-VELOCITY IN THRESHELD AREAS OF INLET REGION.
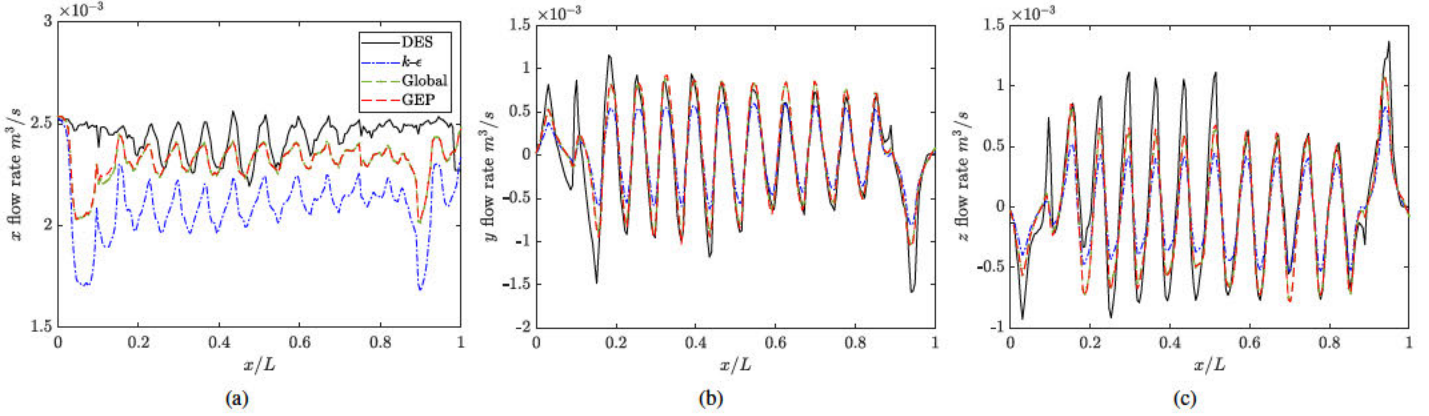

FIGURE 11: (a) X, (b) Y AND (c) Z-VELOCITY INTEGRATED OVER CROSS-SECTION AT EVERY STREAMWISE LOCATION.

x-component of velocity, for which the k-ε model visually seems to present a closer fit, the GEP solution clearly improves the velocity predictions of the y and z-components. The k-ε model appears to under predict the extremes of velocity, whereas the GEP solution (and the global regression) allows these extrema to be more closely matched.

These observations are confirmed by Figure (11). In these plots the three components of velocity are integrated over the cross sectional area at each streamwise location and the resulting volume flow rates plotted for the same four cases. In the y and z-plots the ensembled GEP equation can be seen to match the extrema of the DES data more closely than the k-ε model which appears damped. For the x-plot the ensembled GEP equation is also seen to more closely fit the DES data, despite the unclear conclusion from the contour plot. This is of particular importance as the plot shows the streamwise evolution of the axial velocity. Larger values indicate greater aerodynamic blockage, an essential feature in cooling channel design. The clear superior

prediction of GEP over the k-ε model should mitigate the risk of incorrect CFD performance predictions.

It is also apparent from both Figure (10) and Figure (11) that the velocity predictions of the global regression are very similar to those of the ensemble averaged GEP equation. It appears that the improved prediction in anisotropy of GEP has not been carried over to significantly improved velocity predictions. However, the main the advantage of GEP is expected to lie in its predictive performance. The scalar invariants in the coefficients contain information about the flow and should allow the equation to adapt when presented with a new flow case. The global regression on the other hand is unable to do so on account of its fixed coefficients.

Despite the similarity of GEP and global regression predictions, these results clearly demonstrate the success of the data driven methodology. Every candidate solution from the GEP algorithm was able to outperform the k-ε model, as well as an established ASM model from literature. Furthermore, this per-
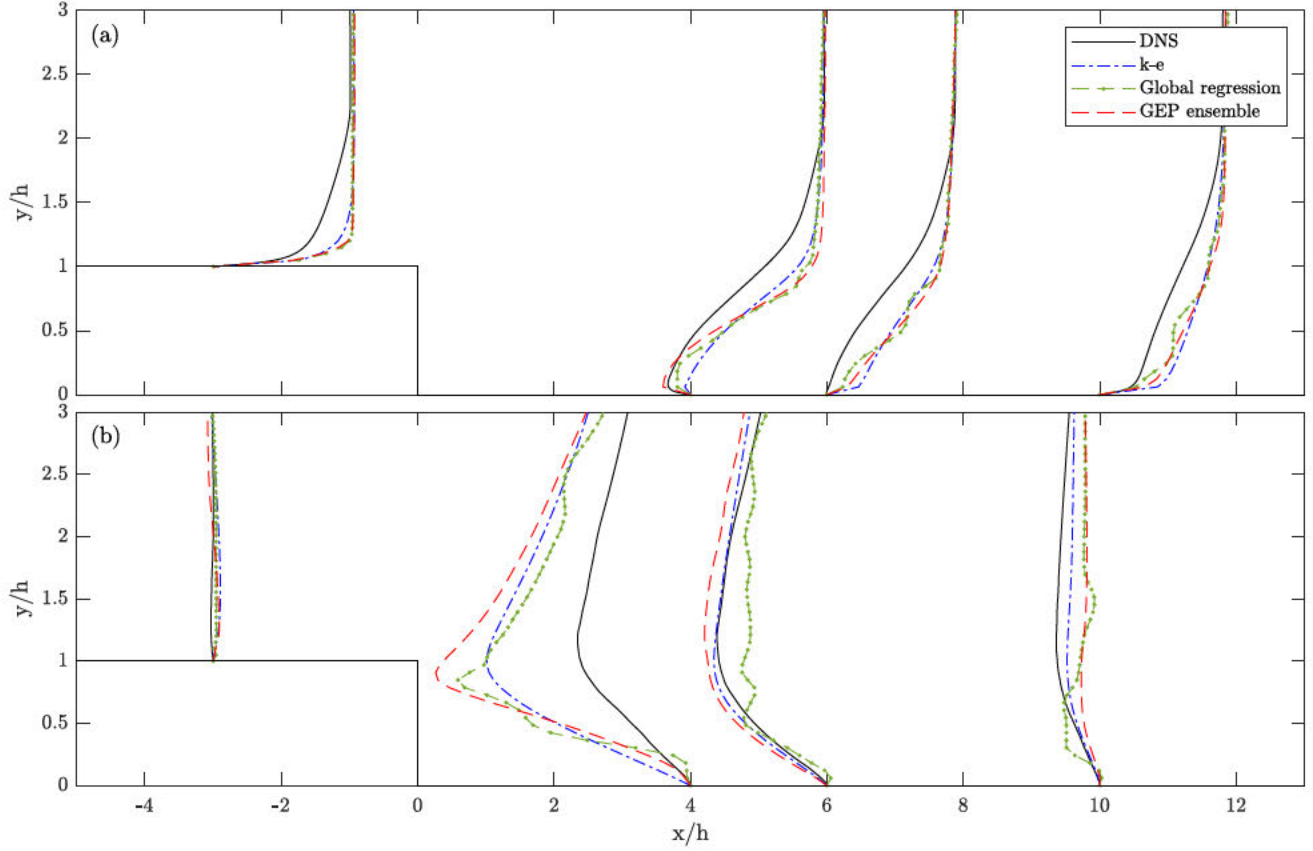
FIGURE 12: (a) *U* AND (b) *V* WALL-NORMAL PROFILES FOR BACK FACING STEP AT VARIOUS STREAMWISE LOCATIONS.

formance has come from methodology that can be considered 'implementation ready'. This means that no manual alterations were made and the equations resulting from the GEP algorithm, were inserted directly into the ASM.

Backward-Facing Step Test

Following *a posteriori* testing, the ensembled GEP equation is now applied to the BFS case to evaluate it's predictive performance. The Reynolds number based on the step height h and the free-stream velocity was $Re_h = 5100$, the channel height before the step is $5h$ and after the step is $6h$, corresponding to an expansion ratio of 1.20. The length of the entry section is $10h$ and the length of the post-expansion section is $20h$. Explain mesh. DNS data from Le, Moin and Kim [ref] is used as reference data. Figure (12) shows the wall normal profiles of streamwise and spanwise velocity at various streamwise locations for the DNS data, the k-ε model, the global regression and the ensemble averaged GEP model.

It is immediately obvious that whilst the GEP equation has successfully converged, the global regression has not. Now, the benefit of GEP over a simple global regression becomes truly apparent; the scalar invariants in the GEP coefficients have allowed the equation to adapt to the new scenario. In contrast, the fixed coefficients of the global regression cannot change when presented with a different flow case, which leads to the non-converging result.

Comparing the results of the ensembled GEP equation to the standard k-ε model, GEP matches the DNS profile more closely in the streamwise velocity profiles, especially in capturing the reversed flow in the recirculation region. However, this comes at the expense of deteriorated prediction in the spanwise profiles. Whilst the k-ε model over predicts the negative velocity in the spanwise direction, GEP predicts even more extreme values.

It should be noted that the quality of the mesh was also very important in achieving these results, the GEP case needed a much finer mesh than the standard k-ε model to produce a converged result so further work is required to make the method more robust. However, considering the vastly differing nature of the training and testing cases, the ability to achieve this result is still extremely promising.

CONCLUSIONS

In this work high fidelity data from DES was used with the GEP symbolic regression algorithm to learn explicit nonlinear expressions for the Reynolds stresses, to be used as turbulence closures in RANS calculations. This methodology was applied in the context of design for Additive Manufacturing of internal cooling channels in turbine blades. It is necessary to provide improved heat transfer predictions for complex 3D geometries, over those given by standard RANS models using linear constitutive laws.

The GEP algorithm was run 100 times to generate candidate equations for the anisotropic component of the Reynolds stresses and each of these solutions was found to outperform the original k-ε model. This superior performance was determined by considering the mean square error, for which there was an average improvement in the ensembled GEP equation of 16% over the k-ε model, and the mean alignment for which there was a 76% average improvement when tested *a posteriori*.

Local and global linear regressions on the anisotropy were also performed to benchmark GEP against simple variable regression techniques. The local linear regression was able to exactly recover the true anisotropy *a priori*, but this method is considered of little practical use as it cannot be used in a predictive situation. The *a posteriori* performance of the ensembled GEP equation showed a slight improvement over the global linear regression with mean square errors of 0.314 and 0.321, respectively, and mean alignments of 0.189 and 0.178, respectively. It was then shown by considering the BFS case, that the real advantage of GEP over the simple global regression comes from its predictive capabilities.

In comparison to the theoretically derived closed form ASM from Wallin and Johannson (2000), all of the GEP solutions performed better in both mean square error and mean alignment. This further attests to the success of the methodology, which can outperform not only standard linear models, but accepted nonlinear closures from literature as well.

Improved prediction of the Reynolds stresses is not a guarantee of the better prediction of the mean flow variables. However, in this case the mean velocity predictions were also shown to be superior to those obtained by the k-ε model. This improved accuracy of the mean flow field is also carried through to more accurate heat transfer prediction, the true quantity of interest for the design of internal cooling channels.

Whilst these results, and the ability of GEP to predict a vastly differing test case in the BFS are promising, it should be recognised that they have not yet been tested in a Topology Optimisation. To be considered useful in the context of design for AM, the data driven nonlinear closures should be able to provide more accurate flow predictions to fluid Topology Optimisation algorithms, in order for these to converge on better solution geometries than can be obtained with the standard k-ε model. It is therefore recommended that the developed closures are now tested in a fluid TO algorithm, to discern if the resulting geometry provides a superior design solution.

REFERENCES

[1] Chyu, M. K., and Siw, S. C., 2013. "Recent Advances of Internal Cooling Techniques for Gas Turbine Airfoils". *Journal of Thermal Science and Engineering Applications,* **5**(2).

[2] Tang, Y., and Zhao, Y. F., 2016. "A survey of the design methods for additive manufacturing to improve functional performance". *Rapid Prototyping Journal,* **22**(3), pp. 569–590.

[3] Pietropaoli, M., Montomoli, F., and Gaymann, A., 2018. "Three-dimensional fluid topology optimization for heat transfer". *Structural and Multidisciplinary Optimization,* **59**(3), pp. 801–812.

[4] Kunz, R. F., Hanson, D. R., Mcclain, S. T., Kinzel, M. P., Reutzel, E. W., Snyder, J. C., and Thole, K. A., 2018. "Direct Numerical Simulation , Up-scaled Measurement and RANS Analysis of Additively and Conventionally Manufactured Internal Turbine Cooling Passages". In 71st Annual Meeting of the APS Division of Fluid Dynamics.

[5] Wilcox, D. C., 2006. *Turbulence Modeling for CFD*, 3rd ed. DCW Industries, La Canada, California.

[6] Slotnick, J., Khodadoust, A., Alonso, J., and Darmofal, D., 2014. CFD Vision 2030 Study: A Path to Revolutionary Computational Aerosciences. Tech. rep., NASA Langley Research Center, Hampton.

[7] Durbin, P. A., 2018. "Some Recent Developments in Turbulence Closure Modeling". *Annual Review of Fluid Mechanics,* **50**(1), pp. 77–103.

[8] Duraisamy, K., Iaccarino, G., and Xiao, H., 2019. "Turbulence Modeling in the Age of Data". *Annual Review of Fluid Mechanics,* **51**(1), pp. 357–377.

[9] Wang, J. X., Wu, J. L., and Xiao, H., 2017. "Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data". *Physical Review Fluids,* **2**(3), pp. 1–22.

[10] Wu, J. L., Xiao, H., and Paterson, E., 2018. "Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework". *Physical Review Fluids,* **7**(3), pp. 1–28.

[11] Ling, J., Kurzawski, A., and Templeton, J., 2016. "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance". *Journal of Fluid Mechanics,* **807**, pp. 155–166.

[12] Weatheritt, J., and Sandberg, R., 2016. "A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship". *Journal of Computational Physics, 325*, pp. 22–37.

[13] Akolekar, H. D., Sandberg, R. D., Hutchins, N., Michelassi, V., and Laskowski, G., 2019. "Machine-Learnt Turbulence Closures for Low-Pressure Turbines With Unsteady Inflow Conditions". *Journal of Turbomachinery, 141*(10), pp. 1–11.

[14] Pichler, R., Sandberg, R. D., Michelassi, V., and Bhaskaran, R., 2016. "Investigation of the Accuracy of RANS Models to Predict the Flow Through a Low-Pressure Turbine". *Journal of Turbomachinery, 138*(12).

[15] Schoepplein, M., Weatheritt, J., Sandberg, R., Talei, M., and Klein, M., 2018. "Application of an evolutionary algorithm to LES modelling of turbulent transport in premixed flames". *Journal of Computational Physics, 374*, pp. 1166–1179.

[16] Weatheritt, J., Pichler, R., Sandberg, R. D., Laskowski, G., and Michelassi, V., 2017. "Machine Learning for Turbulence Model Development Using a High-Fidelity HPT Cascade Simulation". In Turbomachinery Technical Conference and Exposition.

[17] Weatheritt, J., and Sandberg, R. D., 2017. "The development of algebraic stress models using a novel evolutionary algorithm". *International Journal of Heat and Fluid Flow, 68*(November), pp. 298–318.

[18] Weatheritt, J., and Sandberg, R. D., 2019. "Improved Junction Body Flow Modeling Through Data-Driven Symbolic Regression". *Journal of Ship Research*.

[19] Weatheritt, J., Zhao, Y., Sandberg, R. D., Mizukami, S., and Tanimoto, K., 2019. "Data-driven scalar-flux model development with application to jet in cross flow". *International Journal of Heat and Mass Transfer*(xxxx).

[20] Pope, S. B., 1975. "A more general effective-viscosity hypothesis". *Journal of Fluid Mechanics, 72*, pp. 331–340.

[21] Rodi, W., 1976. "A New Algebraic Relation for Calculating the Reynolds Stresses". *Gesellschaft Angewandte Mathematik und Mechanik, 56*(3).

[22] Spencer, A. J., and Rivlin, R. S., 1958. "The theory of matrix polynomials and its application to the mechanics of isotropic continua". *Archive for Rational Mechanics and Analysis, 2*(1), pp. 309–336.

[23] Gatski, T. B., and Jongen, T., 2000. "Nonlinear eddy viscosity and algebraic stress models for solving complex turbulent flows". *Progress in Aerospace Sciences, 36*(8), pp. 655–682.

[24] Ferreira, C., 2001. "Gene Expression Programming: A New Adaptive Algorithm for Solving Problems". *Complex Systems, 13*(2), pp. 87–129.

[25] Pietropaoli, M., Gaymann, A., and Montomoli, F., 2020. "GT2020-14479 Three-Dimensional Fluid Topology Optimization and Validation of a Heat Exchanger With Turbulent Flow". In Turbomachinery Technical Conference and Exposition.

[26] Wissink, J. G., Michelassi, V., and Rodi, W., 2004. "Heat transfer in a laminar separation bubble affected by oscillating external flow". *International Journal of Heat and Fluid Flow, 25*(5), pp. 729–740.

[27] Wallin, S., and Johansson, A. V., 2000. "An explicit algebraic Reynolds stress model for incompressible and compressible turbulent flows". *Journal of Fluid Mechanics, 403*, pp. 89–132.

[28] Parneix, S., Laurence, D., and Durbin, P., 1996. "Second moment closure analysis of the backstep flow database". In Center for Turbulence Research Summer Program, pp. 47–62.

Author/s:
Hammond, J; Montomoli, F; Pietropaoli, M; Sandberg, R; Michelassi, V

Title:
MACHINE LEARNING FOR THE DEVELOPMENT OF DATA DRIVEN TURBULENCE
CLOSURES IN COOLANT SYSTEMS

Date:
2020-06-22

Citation:
Hammond, J., Montomoli, F., Pietropaoli, M., Sandberg, R. & Michelassi, V. (2020).
MACHINE LEARNING FOR THE DEVELOPMENT OF DATA DRIVEN TURBULENCE
CLOSURES IN COOLANT SYSTEMS. Proceedings of ASME Turbo Expo 2020
Turbomachinery Technical Conference and Exposition, 7A-2020, ASME: The American
Society of Mechanical Engineers. https://doi.org/10.1115/GT2020-15928.

Persistent Link:
http://hdl.handle.net/11343/241915

File Description:
Accepted version