

**GT2017-63497**

**MACHINE LEARNING FOR TURBULENCE MODEL DEVELOPMENT USING A  
 HIGH-FIDELITY HPT CASCADE SIMULATION**

**Jack Weatheritt\***

**Richard Pichler**

**Richard D. Sandberg**

Department of Mechanical Engineering  
 University of Melbourne  
 Parkville, Victoria 3010  
 Australia

Email: jack.weatheritt@unimelb.edu.au

**Gregory Laskowski**

Thermal System Design  
 GE Aviation

Lynn, Massachusetts, 01905

**Vittorio Michelassi**

General Electric Oil & Gas  
 Florence, 50127  
 Italy

**ABSTRACT**

*The validity of the Boussinesq approximation in the wake behind a high-pressure turbine blade is explored. We probe the mathematical assumptions of such a relationship by employing a least-squares technique. Next, we use an evolutionary algorithm to modify the anisotropy tensor a priori using highly resolved LES data. In the latter case we build a non-linear stress-strain relationship. Results show that the standard eddy-viscosity assumption underpredicts turbulent diffusion and is theoretically invalid. By increasing the coefficient of the linear term, the far-wake prediction shows minor improvement. By using additional non-linear terms in the stress-strain coupling relationship, created by the evolutionary algorithm, the near-wake can also be improved upon. Terms created by the algorithm are scrutinized and the discussion is closed by suggesting a tentative non-linear expression for the Reynolds stress, suitable for the wake behind a high-pressure turbine blade.*

**NOMENCLATURE**

$a_{ij}$  Anisotropy tensor.  
 $k$  Turbulent kinetic energy.  
 $k'$  Normalized turbulent kinetic energy:  $k/k^{\max}$

$N$  Population size; number of training data points (context specific)  
 $M$  Number of GEP solutions  
 $\mathcal{P}_k$  Turbulent kinetic energy production:  $\tau_{ij}\partial_{x_j}U_i$   
 $S_{ij}$  Strain rate:  $\frac{1}{2}(\partial_{x_j}U_i + \partial_{x_i}U_j)$   
 $S'_{ij}$  Deviatoric component of strain rate:  $S_{ij} - \frac{1}{3}\delta_{ij}S_{kk}$ .  
 $t_t$  Turbulent time scale:  $1/\omega$ .  
 $U_i$  Time-averaged velocity vector  
 $s$  Intrinsic coordinate along the wake center line, normalized by axial chord.  
 $x, y$  Cartesian coordinates, normalized by axial chord.  
 $\beta$  Optimization parameter in least-squares regression.  
 $\partial_\phi$  Shorthand for  $\frac{\partial}{\partial\phi}$ . As a differential operator, it acts on everything to the right within a term.  
 $\varepsilon$  Least squares model error.  
 $\mu_t$  Eddy-viscosity.  
 $\rho$  Time-averaged density.  
 $\tau_{ij}$  Reynolds stress:  $\rho\overline{u'_i u'_j}$ .  
 $\omega$  Specific dissipation rate.  
 $\Omega_{ij}$  Rotation rate tensor:  $\frac{1}{2}(\partial_{x_j}U_i - \partial_{x_i}U_j)$

**INTRODUCTION**

For the design of gas turbines, the primary tools are Reynolds-Averaged Navier-Stokes (RANS) codes. This is be-

\*Address all correspondence to this author.

cause of the excessive computational effort required for Direct and Large Eddy Simulations (DNS/LES).

In the High-Pressure Turbine (HPT), the large Reynolds number and transonic regime of the flow are inimical to the predictions made utilizing such RANS solvers. On the blade itself, strong favorable pressure gradients act to keep the suction side boundary layer laminar. The freestream turbulence, inducing bypass transition, is therefore a major player in determining aerodynamic efficiency, loadings and heat transfer [1]. The prohibitive Reynolds number has delayed high-fidelity simulations until recently. Wheeler *et al.* [2] published the first DNS of a linear HPT cascade at transonic conditions. By varying the freestream turbulence, they showed the suction side flow to be highly unsteady — with the point of transition varying greatly over time. Such a sensitivity on and consequently prediction quality of varied freestream turbulence cases has been found lacking in RANS modeling [3]. This poor predictive capability continues downstream; wake loss profiles obtained with RANS exhibit excessive kurtosis than the fatter tailed reality [4].

For such RANS methods, a large form of model uncertainty is attributed to the Boussinesq hypothesis,

$$\tau_{ij} = \frac{2}{3}\rho k\delta_{ij} - 2\mu_t S'_{ij}. \quad (1)$$

This relates the anisotropy of the Reynolds stress to the deviatoric strain-rate tensor, via a coefficient of proportionality, which in non-dimensional form is interpreted as a turbulent time scale,

$$a_{ij} \equiv \frac{\tau_{ij}}{2\rho k} - \frac{1}{3}\delta_{ij} \quad (2a)$$

$$= -t_I S'_{ij} \quad (\text{linear}) \quad (2b)$$

$$= a_{ij}(V_{ij}^1, V_{ij}^2, \dots, I_1, I_2, \dots). \quad (\text{EASM}) \quad (2c)$$

Despite well documented shortfalls [5, 6, for example], Eq. 2b is by far the most popular choice in industry, due to robustness and its correct shear stress prediction in thin turbulent boundary layers [7]. A class of turbulence closures, known as Explicit Algebraic (Reynolds) Stress Models (EASMs) begin from the weak equilibrium hypothesis [8] to express a tensor basis  $V_{ij}^k$  and scalar invariants  $I_k$  for  $a_{ij}$  [9]. This basis and set of scalar invariants are functions of the non-dimensional strain and rotation rate tensors, denoted  $s_{ij} = t_I S'_{ij}$  and  $w_{ij} = t_I \Omega_{ij}$ . For two-dimensional flow, the basis and invariants are

$$\begin{aligned} V_{ij}^1 &= s_{ij}, & V_{ij}^2 &= s_{ik}w_{kj} - w_{ik}s_{kj}, \\ V_{ij}^3 &= s_{ik}s_{kj} - \frac{1}{3}\delta_{ij}s_{mn}s_{nm}, \\ I_1 &= s_{mn}s_{nm}, & I_2 &= w_{mn}w_{nm}. \end{aligned} \quad (3)$$

An EASM closes the system by assuming the functional form Eq. 2c and writing down algebraic expressions using Eq. 3. There are many such models found in the literature; for examples see Refs. [10, 11, 12].

Equation 2 is the focus of our discussion. We begin from high-fidelity reference data of the flow around an HPT blade [2] and inspect both models of the forms in Eq. 2b and Eq. 2c. First, we argue on statistical grounds that linear RANS models are inappropriate for such a case. Then, by optimizing a linear model, in an *a priori* least-squares sense, we show the modest upper bound in predictive performance possible. This process modifies the eddy-viscosity to improve turbulent diffusion, however such optimized linear models are unable to capture this process in the near-wake. Next, we discuss non-linear models by regressing EASM closures from the high-fidelity data. Thus, the approach is to define EASM models specific to HPT cascades. We perform symbolic regression via an Evolutionary Algorithm [13]. Our framework is an extension of Gene Expression Programming (GEP) [14] and comes as machine learning for turbulence modeling is gaining popularity [15, 16, 17, 18].

Such data-driven frameworks aim to reduce the number of assumptions made when producing solutions. Our GEP algorithm returns an algebraic tensor equation that can be inserted into the system (Eq. 2c). As the method is inexpensive and non-deterministic, we can repeatedly perform this to create many models and obtain statistical information such as algorithm efficacy and common features amongst solutions [15]. It is this last aspect of the framework that we focus on in this paper.

We begin, in the next section, with an outline of the HPT case used throughout. The following section scrutinizes the linear relationship in the wake on both theoretical and practical grounds. Then, the final sections describe and apply our evolutionary algorithm to further understand and improve upon the Boussinesq approximation for the wake behind an HPT blade. In particular, we ask what non-linear extensions are suitable — we do not propose a new model *per se*, rather components that assist in minimizing the error. In this sense, our work is *a priori* and is part of the ongoing development of machine learning methods that produce viable *a posteriori* turbulence closures [13, 15, 19].

## FLOW CONFIGURATION

The reference data used in this paper is part of a wide reaching, high-fidelity study into understanding the physics, particularly the loss mechanisms, in turbomachinery flows subject to a wide range of operating conditions. The current case is a highly resolved LES of the flow through a linear HPT cascade, initially proposed as a computational test case through an extensive experimental study [1]. The computational setup is similar to a previous DNS [2], which matched the original experiment. That is, exit Reynolds and Mach numbers of  $5.7 \times 10^5$  and 0.9 respectively with the blade treated as an isothermal, no-slip wall with a

temperature 3/4 the stagnation temperature.

As part of the wider study, the present flow has an increased inlet turbulent length scale at 5% axial chord and to accommodate this rise, the domain is thrice the DNS at 15% axial chord. The total number of grid points in the plane and span is approximately 834'000 and 386 respectively. The subgrid-scales were calculated with the Wall-Adapting Local Eddy-viscosity model (ubiquitously abbreviated to WALE) [20]. The simulation was carried out using a hybrid finite difference/spectral code purposely developed for compressible Navier-Stokes calculations on high-performance computers [21].

## AN INSPECTION OF THE LINEAR RELATIONSHIP

Many authors [22,23,24, for example] have utilized a ‘least-square sense’ optimized value of  $\mu_t$ ,

$$\frac{\mu_t}{\rho} = -\frac{ka_{mn}S'_{mn}}{S'_{ij}S'_{ij}}, \quad (4)$$

during *a priori* testing of RANS models. Here, we offer that Eq. 4 is the result of the scalar projection of  $a_{ij}$  into  $-S'_{ij}$ , such that the time-scale (not normed by  $\beta^* = 0.09$ ),

$$t_I = \frac{1}{\omega} = \frac{\mu_t}{k\rho} \quad (5)$$

is the magnitude of  $a_{ij}$  in this tensor space multiplied by  $\sqrt{S'_{ij}S'_{ij}}$ . We can use this relationship to test the assumption of linearity implied by Eq. 2b through the definition of a linear model,

$$a_{ij}S'_{ij} = -\beta \frac{1}{\omega} S'_{ij}S'_{ij} + \varepsilon. \quad (6)$$

$\beta$  is a free parameter, such that we build an  $\mu_t^{\text{eff}} = \mu_t\beta$ , and  $\varepsilon$  is the model error induced by the projection. For such a linear relationship to hold (and by extension Eq. 2b),  $\varepsilon$  must be normally distributed. This condition will form the bulk of discussions below. To minimize the error  $\varepsilon$ , we can define the cost function,

$$J(\beta) = \frac{1}{2} \sum_{n=1}^N (\varepsilon^{(n)})^2, \quad (7)$$

for the  $N$  data points in the LES. This cost function describes the total error at each data point squared and is a classical choice to solve linear regression problems. The global minima of this cost

function can be found simply, after  $J$  is quadratic in  $\beta$ , seen by inserting Eq. 6 into Eq. 7. Differentiation yields

$$\frac{\partial J}{\partial \beta} = 0 = \sum_{n=1}^N \left[ S'_{ij} \left( \frac{\beta}{\omega} S'_{ij} + a_{ij} \right) \right]^{(n)}, \quad (8)$$

which is a linear equation for the optimal turbulent viscosity for the considered domain. Rearranging Eq. 8 provides a closed expression for  $\beta$ ,

$$\beta = -\frac{\sum_{n=1}^N \left[ S'_{ij}S'_{ij} \right]^{(n)}}{\sum_{n=1}^N \left[ \frac{S'_{mn}S'_{mn}}{\omega} \right]^{(n)}}. \quad (9)$$

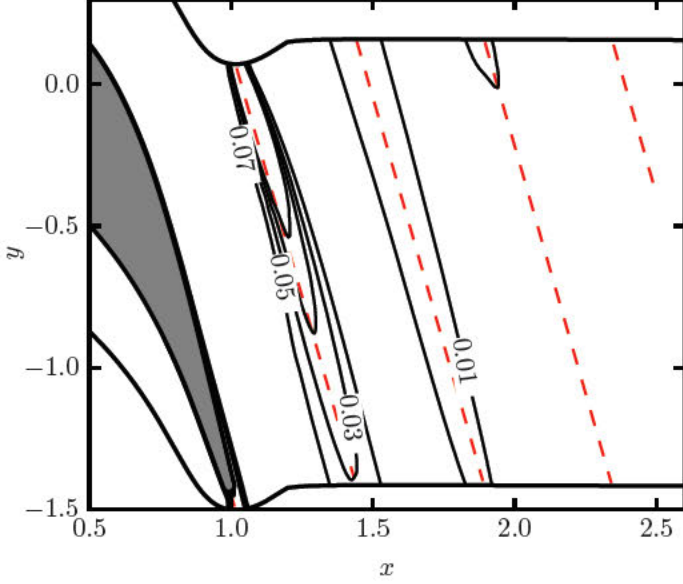
The values for  $a_{ij}$  and  $S'_{ij}$  are taken directly from the LES, whilst  $\omega$  is solved from its RANS transport equation [25],

$$\partial_{x_j} \rho U_j \omega = \frac{\gamma}{\mu_t} \mathcal{P}_k - \beta^w \rho \omega^2 + \partial_{x_j} (\mu + \sigma_1 \mu_t) \partial_{x_j} \omega + 2(1 - F_1) \frac{\rho \sigma_2}{\omega} (\partial_{x_j} k) (\partial_{x_j} \omega), \quad (10)$$

where the quantities not defined in the nomenclature are the standard constants and damping functions. Those quantities defined in the nomenclature are calculated from the LES and held constant, with the exception of  $\mu_t$  which depends on  $\omega$  (Eq. 5). This freezing [23,26] of flow quantities obtains the ‘correct’ RANS specific dissipation rate given the flow state  $(\rho, U_i, \tau_{ij})$ . With  $\omega$ , all variables in Eq. 8 can be defined and thus an effective eddy-viscosity  $\mu_t^{\text{eff}} = \beta \mu_t$  can be determined.

Four data sets are considered individually for minimizing Eq. 7. Each is a selection of the wake, depicted in Fig. 1 and determined by thresholding  $k' < k/k^{\text{max}}$  with  $k' = 0.01, 0.03, 0.05$  and  $0.07$ . A further constraint  $x > 1.07$  is enforced because, as shall be discussed, the linear assumption is entirely inadequate below this value.

Table 1 lists parameters for each regression. The optimized values  $\mu_t^{\text{eff}}/\mu_t \approx 1.8$  can explain approximately 90% of the variance, which is a modest improvement over the regular expression. A larger  $\mu_t^{\text{eff}}$  corresponds to an increase in turbulent diffusion and would likely improve the poor prediction of loss profiles in the wake. Figure 2 is a plot along the approximate wake center line [1] indicated with the dashed line (wrapped through the periodic boundaries) in Fig. 1.  $s$  is the coordinate along this line (normalized axial chord), with  $s = 0$  being the trailing edge and  $s = 5$  corresponding to the point  $(x, y) = (2.38, 0.034)$ . The line  $\varepsilon = 0$ , where the model is making no error, is also included



**FIGURE 1.** PLOT HIGHLIGHTING REGRESSION REGIONS (—) AND LOCATION OF DATA SAMPLING USED IN FIG 2 (---) THROUGHOUT THE COMPUTATIONAL DOMAIN.

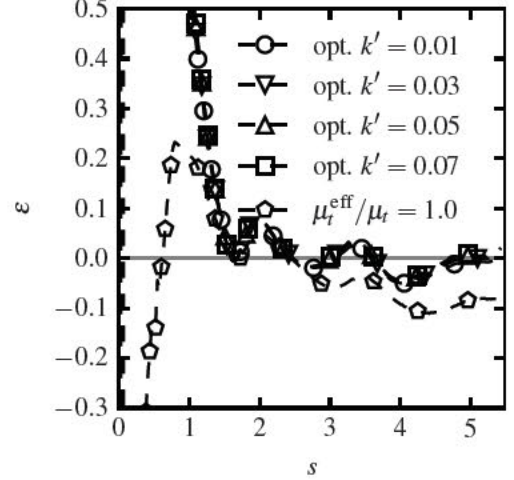
**TABLE 1.** SUMMARY OF REGRESSION PARAMETERS RELATIVE TO STANDARD MODEL. LAST ROW INDICATES THE SKEW AND KURTOSIS OF THE STANDARD GAUSSIAN DISTRIBUTION.

$k'$ thresh.	$\mu_t^{\text{eff}}/\mu_t$	$R^2$	skew	kurtosis
-	1.000	0.835	-0.234	0.452
0.010	1.736	0.922	0.668	17.081
0.030	1.7988	0.941	-0.197	3.546
0.050	1.8317	0.949	0.174	2.120
0.070	1.8557	0.958	0.501	2.046
Gaussian dist.	N/A	N/A	0	3

as a visual guide. Initially, close to the trailing edge, all models are highly in error. Then approximately two and half axial chords downstream, the optimized values report a low error, whilst the regular  $\mu_t$  underpredicts the projection of  $a_{ij}$  into  $-S'_{ij}$ .

To understand Fig. 2 more thoroughly, consider the quantity [6]

$$-1 \leq \frac{-a_{ij}S'_{ji}}{\sqrt{a_{mn}a_{nm}S'_{pq}S'_{qp}}} \leq 1, \quad (11)$$



**FIGURE 2.** MODEL ERROR  $\varepsilon$  ALONG THE WAKE CENTER LINE AS A FUNCTION OF ARC LENGTH COORDINATE  $s$ . SEE FIG. 1 FOR THE LINE THAT  $s$  DESCRIBES.

which measures the orthogonality of anisotropy and negative strain. This is a useful indicator of linear RANS model validity. Values close to one imply strong validity of the linear Boussinesq approximation, values close to minus one imply that Eq. 2b predicts the anisotropy oriented in the wrong direction whilst values approximately zero indicate there is zero correlation between the linear model and actual anisotropy (complete orthogonality). In Fig. 3, one can see that the model ( $-S'_{ij}$ ) for the anisotropy is highly misaligned in yellow regions, whilst in regions of blue, a linear model (with a modified coefficient) appears adequate. The interpretation of Fig. 2 is now clear. In the near-wake there is poorer alignment ( $< 1$ ), so linear models regardless of coefficient are in error. Then for a region of approximately two axial chords, the red line is close to (but not on) the center line, where  $-a_{ij}S'_{ji}/\sqrt{a_{mn}a_{nm}S'_{pq}S'_{qp}} \approx -1$  — i.e. strong misalignment — and a large coefficient acts to damage prediction. Note, on the center line itself, there is little or no strain yet tangential anisotropy. Then beyond three and a half axial chords (not shown in Fig. 3), the profile deviates from the center line into a region of good alignment ( $\approx 1$ ). In these regions, the adjusted coefficient is an improvement over the standard  $\mu_t$ .

The optimized coefficient is reasonably robust to the amount of data used from the wake, however not all data sets satisfy assumptions inherent in least-squares optimization. Reading from Table 1, the data set  $k' = 0.03$  appears most Gaussian, with a kurtosis of 3.5 and near zero skew. The largest data set,  $k' = 0.01$ , has an extreme kurtosis value that indicates a very heavily tailed distribution because very few data points are from the near-wake. Note, the less wake that is included in the regression, the more the ‘outliers’ (non-aligned regions in Fig. 3) can be felt and the coefficient rises. The quantile-quantile plot of Fig. 4 shows this



influence of the outliers. This plot splits each data set into equal quantiles and plots them against the quantiles of the standard Gaussian distribution. Only a very small portion of the optimized model error follows the theoretical Gaussian line. There is non-normality because we observe large portions of smaller-than-expected underpredictions — required because of our huge overpredictions near the blade edge. Counter this with the standardized residual of the regular model, the lower coefficient has resulted in less error when Eq. 11 evaluates to -1 (because the anisotropy is oriented incorrectly), skewing the data in Fig. 4 below the line. This non-normality is caused because of the failure of Eq. 1, seen via the systematic error induced along the wake line — emphasizing a non-linear relationship between stress and strain.

Note, the purpose of this section was not to find an optimized value of  $\beta$ ; rather - via the use of different wake regions — we have shown that the optimal value is dominated by the very near-wake.<sup>1</sup> This implies that the propagation of error from this region is likely catastrophic for prediction. This section confirms that the wake is neither well fit by optimizing the value of  $\beta$  nor do theoretical considerations indicate such a linear relationship. This gives us a mandate for more sophisticated approaches to stress-strain optimizations.

## EVOLUTIONARY METHODOLOGY

The evolutionary framework [13] — GEP — an extension to the original algorithm suitable for tensor regression, along with the ‘frozen’ methodology for obtaining  $\omega$  described in the section above, are used to build non-linear stress-strain relationships of the form Eq. 2c. The algorithm aims to minimize the cost function

$$J(a_{ij}^{\text{gep}}) = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^6 \sum_{j \leq i} w_{ij} |a_{ij} - a_{ij}^{\text{gep}}|, \quad (12)$$

where  $a_{ij}^{\text{gep}}$  is polynomial in the variables defined in Eq. 3.  $w_{ij}$  are weights that give import to components of the anisotropy tensor during the optimization. For this study, we set the four non-zero components to equal preference in the optimization process  $w_{11} = w_{12} = w_{22} = w_{33} = 0.25$ .

The non-linearity in a candidate solution  $a_{ij}^{\text{gep}}$  during the optimization is built naturally via the ‘evolution’ of Expression Trees (ETs) (known as Parse Trees in Genetic Programming [27]). This type of algorithm is a *symbolic* process; instead of tuning parameters (as with  $\beta$  above), new functional forms not specified by the user are revealed by the process. For example,

<sup>1</sup>Were one to fit an optimal value, one would discount the near-wake entirely. Further, this value is unlikely to be constant across the entire far-wake.

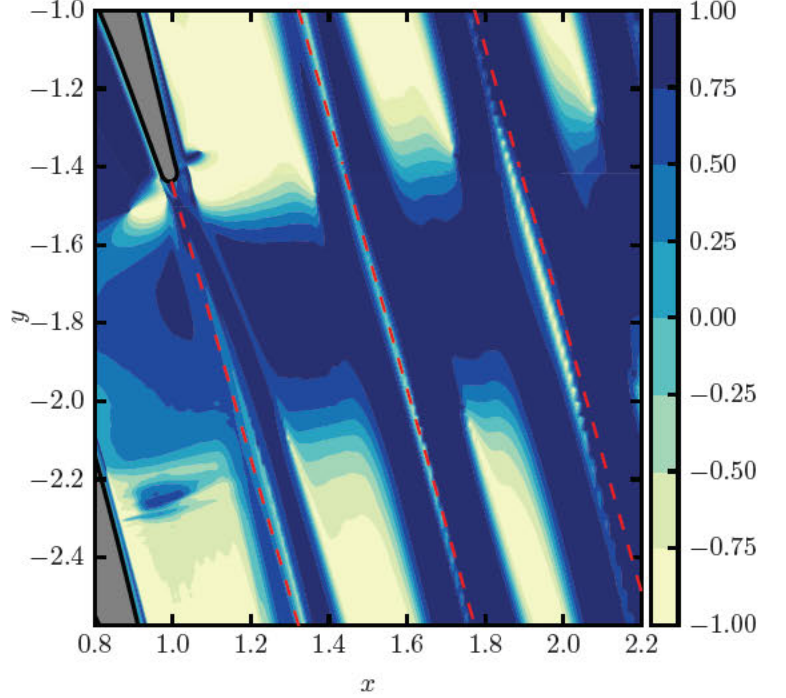


FIGURE 3. ALIGNMENT OF STRESS ANISOTROPY AND NEGATIVE STRAIN.

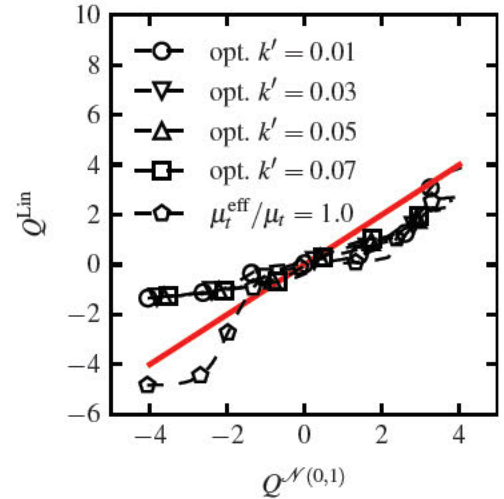


FIGURE 4.  $Q$ - $Q$  PLOT FOR DIFFERENT LINEAR FITS.

an algebraic expression for the anisotropy may be,

$$a_{ij}^{\text{gep}} = V_{ij}^1 - (V_{ij}^8 + (I_1 \sqrt{I_2} V_{ij}^2)). \quad (13)$$

The ET of this candidate expression for  $a_{ij}$ , shown in Fig. 5, is decoded by traversing top to bottom with child nodes as argu-

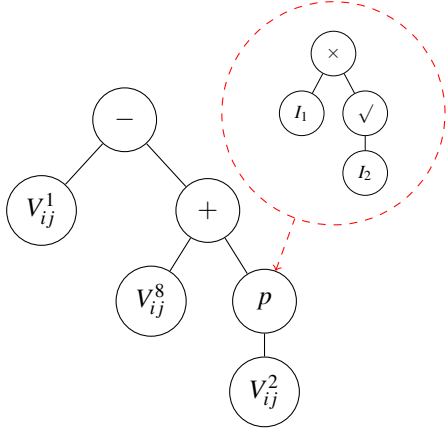


FIGURE 5. EXAMPLE ET.

ments to their parents. Symbols that take no arguments (variables and constants) therefore exist at the extremities of the tree. In GEP, such ETs are constructed from ‘chromosomes’. The current example consists of two,

$$\begin{aligned} & -V1+V8p|V2V3V2V4V2V1 \\ & *I3Q|I2I1I2I5 \end{aligned}$$

one for the scalar field and one for the matrix arguments. These chromosomes are read left to right, with each representing a tree in Fig. 5. The “|” denotes a split between the head and tail of the chromosome and Q is the square root function. The head may consist of operators, variables and constants, but the tail may only contain variables and constants. This distinction guarantees mathematical syntax [14]. These chromosomes are read recursively to build the ET of Fig. 5, until the trees are full (all extremities are filled with terminals). For this example the chromosomes are read up to  $V2$  and  $I2$ , with the remaining code left unread. This redundancy in the tail is how syntax is guaranteed, provided its length is greater than the head’s less one. The  $p$  symbol, known as a plasmid, is the joining point of the two chromosomes. The introduction of the plasmid symbol [13] has made it possible to express tensor (or more generally mixed dimension) expressions naturally.

In order to minimize Eq. 12, an evolutionary analogy is adopted. A ‘population’ of  $N$  randomly created candidate chromosomes are iteratively (each known as a ‘generation’) subject to natural selection, a process where better models survive to the next generation based on Eq. 12. Natural selection is mimicked through tournament selection [27]; small groups of individuals are randomly chosen (with replacement) for  $N-1$  spaces in the next generation. The fittest individual of each group — or tournament — takes one of the  $N-1$  places. The last spot in the generation is filled by the fittest individual in the population, thereby ensuring its survival (because it may not be cho-

sen for tournament selection). The survivors of each generation then randomly ‘mutate’ and ‘breed’ — two processes that change and share portions of chromosomes [14]. Mutation is a random change in one position of the chromosome, say changing the  $+$  to an  $I_1$  in the third position of the above chromosome. Breeding, known as crossover or recombination [27], involves randomly picking two individuals, cutting them at some random point and then exchanging the portions of their chromosome to the right. This process therefore shares genetic material between two parents to produce offspring. The product of natural selection and genetic variation is a gradual drift towards local minima of  $J(a_{ij}^{\text{gEP}})$ . The evolution of the population ceases when a stopping criteria is met; either a sufficiently local minimum has been found or the maximum number of generations is reached. Note, we do not claim *global* minima — GEP searches the space of all tensor functions, which contain an infinite number of minima<sup>2</sup> and as a result is a non-convex surface which is non-trivial to traverse. For a full description of the algorithm see Ref. [13] and for further information regarding specifics of that outlined above and implementations of a wide range of evolutionary algorithms, see Ref. [28].

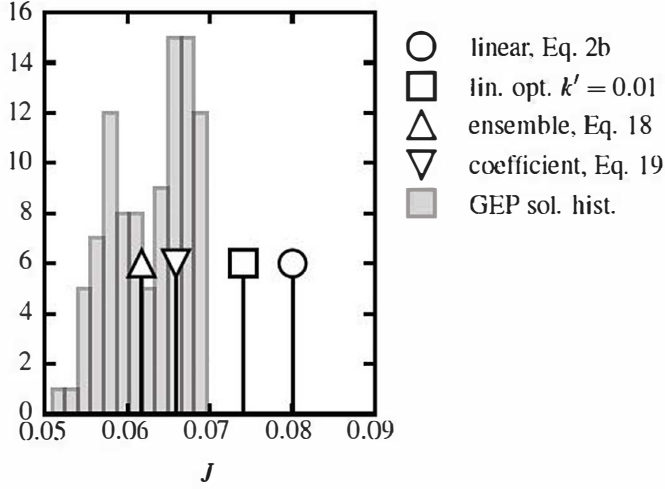
The data used is the same as the  $k' = 0.01$  linear optimization above, which is 2,711 data points. With a population of 300 individuals lasting 300 generations, a single optimization is on the order of 10 core minutes. As repetitions of the algorithm produce different solutions  $a_{ij}^{\text{gEP}}$ , due to the randomness induced by initial conditions, mutations and the highly convex cost function, we have run the algorithm  $M = 100$  times. From this class of solutions, formed by taking the best individual from each population after 300 generations, statistical information can be obtained about the process. This is discussed in the next section.

## RESULTS FROM THE EVOLUTIONARY ALGORITHM

The distribution of raw fitness values, Eq. 12, for each of the trained models  $a_{ij}^{\text{gEP}}$  is displayed in the histogram of Fig. 6. This fitness value  $J$  is a measure of the average component-wise distance from the actual anisotropy  $a_{ij}$ . One can also see from Fig. 6 that every single  $a_{ij}^{\text{gEP}}$  has a considerably better fitness than the regular (Eq. 2b) and optimized ( $\mu_r^{\text{eff}}/\mu_t = 1.736$ ) linear models. These 100 GEP solutions can be ensemble averaged to give a model  $\widetilde{a_{ij}^{\text{gEP}}}$ . This ensemble is the mean prediction of all models, an example of such a model contributing to this average is,

$$\begin{aligned} a_{ij}^{\text{gEP}} = & ((I_2 - 2I_1 + 1)(I_2 + 0.107) + 0.157)(V_{ij}^1 + V_{ij}^3) + \\ & ((I_2 - 2I_1 + 10)(I_2 + 0.107) + 0.157)(V_{ij}^2 + V_{ij}^1) - 2V_{ij}^1. \end{aligned} \quad (14)$$

<sup>2</sup>In practice this space is truncated by the length of the chromosome, however the argument being made still holds.



**FIGURE 6.** HISTOGRAM OF GEP SOLUTION FITNESS  $J$ . INCLUDED AS VISUAL REFERENCE ARE THE FITNESS VALUES OF OTHER DISCUSSED MODELS.

**TABLE 2.** SUMMARY OF MEAN SQUARE ERROR IN WAKE REGION.

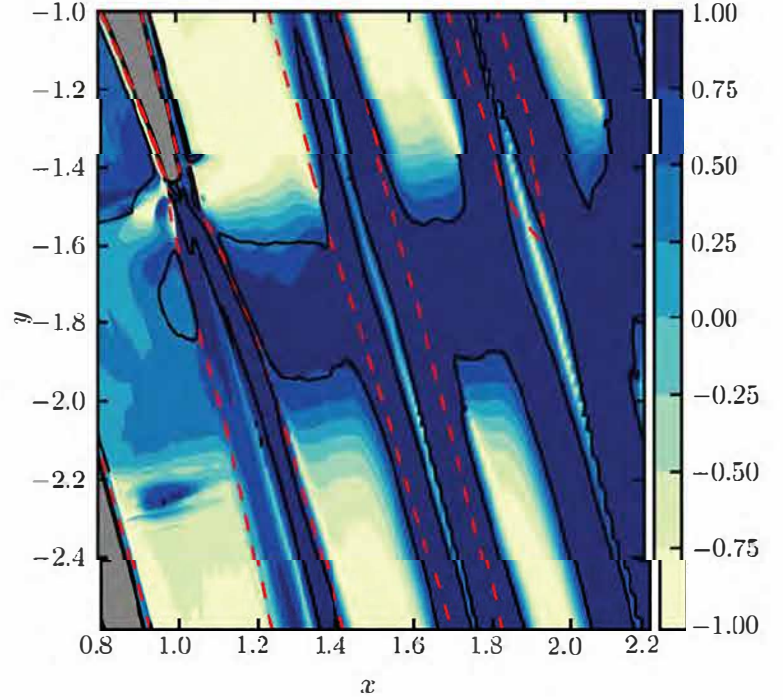
$-\frac{\overline{s_{ij}}}{\omega}$	$-1.736\frac{\overline{s_{ij}}}{\omega}$	$\overline{a_{ij}^{gep}}$
108.05	101.218	74.62

One can immediately observe structural similarities between the two scalar field expressions. This is a result of the evolutionary process re-using and modifying useful ‘genetic code,’ much like regular Darwinism. This behavior allows us to observe useful expressions that have emerged from the algorithm and potentially use them (or variants of) in non-linear expressions for the anisotropy tensor. Some features of the ensembled functional form are discussed below, after first advancing the discussion on the quantification of prediction capability.

Table 2 is a summary of the mean squared error

$$\sum_{n=1}^N \sum_{i=1}^6 \sum_{j \geq i} |a_{ij} - a_{ij}^{gep}|^2, \quad (15)$$

in the training region. This further quantifies the added bonus of including trained non-linear terms highlighted in Fig 1. By optimizing the parameter  $\beta$ , one can obtain a 6% improvement — but with the optimized EASM, one achieves 32%. This is a moderate improvement and pleasing considering the reasonable alignment of stress and strain in the far-wake for the linear mod-



**FIGURE 7.** ALIGNMENT OF STRESS ANISOTROPY AND NON-LINEAR PREDICTION (—), TRAINING REGION (---) AND A LINEAR ALIGNMENT VALUE OF 0.75 (—).

els.

Figure 7 is analogous to Fig. 3 for the ensemble averaged prediction, calculated via

$$\frac{\overline{a_{ij} a_{ji}^{gep}}}{\sqrt{\overline{a_{mn} a_{nm}^{gep} a_{nd}^{gep} a_{dn}^{gep}}}}, \quad (16)$$

Figure 7 also demarks the training data region, within which the non-linear models successfully minimize the alignment error. In particular, by comparing Fig. 3 with Fig. 7, one can see that the non-linear models better reproduce the near-wake emanating from the pressure side. This initial alignment drops below 0.75, but recovers much quicker than the linear prediction. Crucially, this optimized near-wake performance has not come at the cost of far-wake modeling. The non-linear corrections marginally outperform the original Boussinesq approximation — which already was very good in this region. Note, the yellow stripe of misalignment has not been eradicated, this is because of an almost symmetrical velocity distribution in the principal strain direction; at maximal velocity deficit  $\partial_{s\perp} U_{\parallel} \approx 0$ , the basis matrices  $V_{ij}^k$  are therefore no longer well aligned with  $a_{ij}$ . This results in a poorly aligned model  $a_{ij}^{gep}$ . This behavior is an innate problem



of using functions of velocity gradient and is perhaps best seen in the prediction of anisotropy in a turbulent channel [11].

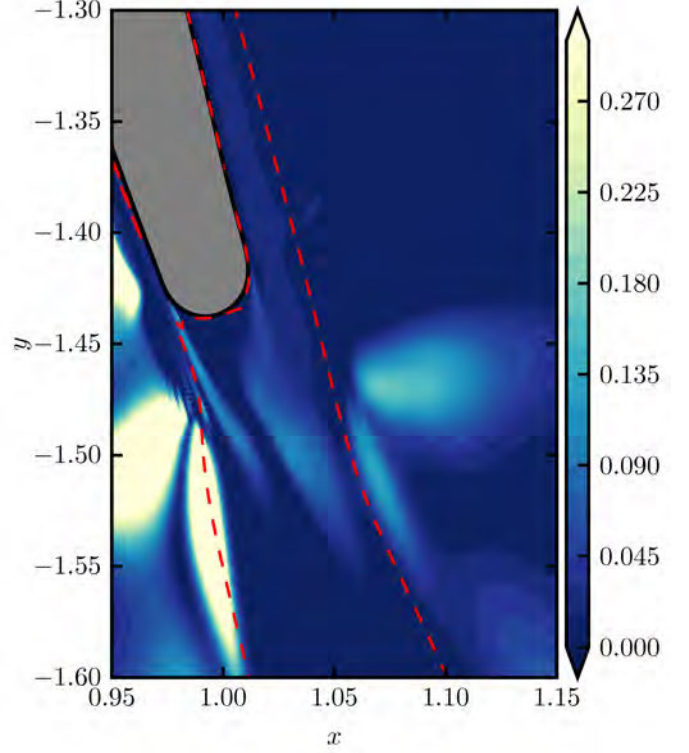
Outside of the demarcated region, the non-linear alignment is worse than that obtained using a linear model, most noticeably in the blade passage. This is an example of overfitting, i.e. the trained closures are specialist wake models and should only be trusted in regions sufficiently similar to the training conditions. Detection functions that activate modifications only in the wake could help in this regard [29, for example].

In Fig. 6, there is a clear distribution around the ensembled fitness  $J(a_{ij}^{\text{gEP}})$ . Interestingly, it appears bimodal which would allude to two functional families consistently being found by the algorithm, but with random variations in each individual. This has not been explored in the interest of space, but further decomposing the ensemble solution would possibly yield two candidate models. This distribution of fitness comes from spatial variation in each of the  $M = 100$  model predictions and can be expressed through the variance in alignment value,

$$\text{Var}\left(\frac{a_{ij}^{\text{gEP}}}{\sqrt{a_{mn}^{\text{gEP}}a_{nm}^{\text{gEP}}a_{pq}^{\text{gEP}}a_{qp}^{\text{gEP}}}}\right)(x,y) = \frac{1}{M} \sum_{k=1}^M \left( \frac{a_{ij}^{\text{gEP}}(k)}{\sqrt{a_{mn}^{\text{gEP}}(k)a_{nm}^{\text{gEP}}(k)a_{pq}^{\text{gEP}}(k)a_{qp}^{\text{gEP}}(k)}} - \left\langle \frac{a_{ij}^{\text{gEP}}}{\sqrt{a_{mn}^{\text{gEP}}a_{nm}^{\text{gEP}}a_{pq}^{\text{gEP}}a_{qp}^{\text{gEP}}}} \right\rangle \right)^2 \quad (17)$$

where  $a_{ij}^{\text{gEP}}$  denotes the vector of  $M = 100$  GEP solutions and  $(k)$  is the model index. Note that the second term is the mean of the alignments, not the alignment of the mean, so as to be consistent with the definition of variance. Equation 17 is therefore the variance of all the GEP models' alignment prediction for each point in space. This quantity is plotted in Fig. 8 for the GEP solutions. Within the marked training region, there is excellent consistency in prediction and consequently the majority of the wake is not shown in Fig. 8. The largest variation comes in the very near-wake region ( $y > -1.5$ ); the flow from each side of the blade — where the anisotropy is at its strongest — causes an individual model  $a_{ij}^{\text{gEP}}$  to be on average 0.09 away from the mean alignment. This is very encouraging considering the high level of anisotropy in this region, where the flow still 'remembers' the attached boundary layer. Outside of the training region, where the flow is not turbulent, the variance rises sharply. These regions, where the levels of GEP model anisotropy are very unpredictable, can cause numerical instabilities when they are used as turbulence models in CFD codes. For this reason, it is desirable to take the ensemble solution to use as a predictive model. This prevents overfitting the training region at the price of other parts of the flow.

We now move onto analysis of the functional form itself. Table 3 is a summary of the frequency of symbol selection amongst



**FIGURE 8.** VARIANCE OF THE GEP SOLUTION ALIGNMENT AS A FUNCTION OF SPACE.

**TABLE 3.** SUMMARY OF SYMBOL SELECTION FREQUENCIES.

basis:	$V_{ij}^1$	0.387	invariants:	$I_1$	0.514
	$V_{ij}^2$	0.286		$I_2$	0.486
	$V_{ij}^3$	0.328			

the 100 models. Note, each function has undergone a basic level of simplification that evaluates and cancels terms. We can see that the basis  $V_{ij}^1$  is picked more often than random, whilst  $V_{ij}^2$  is picked less often than random. The variable  $V_{ij}^3$  is picked about randomly. In terms of the invariants there is a slight tendency to pick  $I_1$ , yet this does not seem significant.

The implication is that the algorithm has found gains more readily by using more expressions in terms of  $V_{ij}^1$ ,  $I_1$  and  $I_2$ . There are less terms involving  $V_{ij}^2$  and approximately one third of the terms contain  $V_{ij}^3$ . Note, it cannot be argued that terms involving  $V_{ij}^1$  are more *complicated*, rather a weaker conclusion that the algorithm has generated and shared fragments containing  $V_{ij}^1$  more



frequently can be maintained. Whether these terms are complex or not, it implies that the linear eddy-viscosity component has proven effective during the optimization.

The coefficients to these basis tensors are functions of  $I_1$  and  $I_2$ . These coefficients, respectively of  $-V_{ij}^1$ ,  $V_{ij}^2$  and  $V_{ij}^3$ , of the ensemble model  $a_{ij}^{\text{sep}}$  are provided:

$$1.334 - 0.438I_1 - 2.635I_2 - 0.0102I_1^2 + 1.021I_2^2 - 12.280I_1I_2 - 102.047I_1I_2^2, \quad (18a)$$

$$0.573 - 1.096I_1 + 8.985I_2 - 1.102I_1^2 + 2.876I_2^2 + 90.633I_1I_2 - 0.0949I_1^3 + 11.020I_1I_2^2 - 0.142I_2^3, \quad (18b)$$

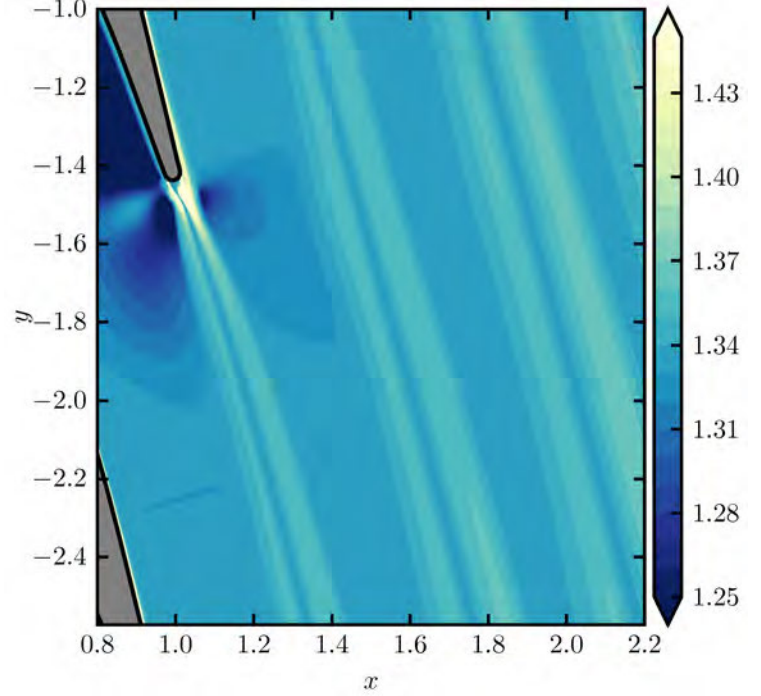
$$12.861 - 25.094I_1 + 6.449I_2 + 1.020I_1^2 - 71.898I_1I_2^2 + 304.979I_1I_2 - 184.519I_2^2 + 10097.273I_2^3 - 10204.0816I_1^4 + 10204.0816I_1^3I_2 - 255.647I_1I_2^3. \quad (18c)$$

By taking the constants from Eq. 18, we have a reduced model

$$a_{ij}^{\text{coeff}} = -1.334V_{ij}^1 + 0.573V_{ij}^2 + 12.861V_{ij}^3. \quad (19)$$

The addition of the scalar invariant terms in Eq. 18 produces a model with effective coefficients, that change with  $I_1$  and  $I_2$ . This full model is not likely to be a fit-for-purpose stress-strain relationship. For example, the elaborate nature of Eq. 18c does not indicate a complex expression in reality, rather the opposite — the 100 ensemble averaged solutions have not produced a reliable, coherent picture in polynomials of  $I_1$  and  $I_2$ . This implies that each GEP model has a very different  $V_{ij}^3$  coefficient such that there is no systemic dependence on  $V_{ij}^3$ . The coefficient and ensemble models have been included in Fig. 6. Both models are considerably better than the baseline linear model, and the coefficient model is only marginally worse than the ensemble.

In Figs. 9-11 the effective coefficients Eqs. 18a-18c are displayed for the wake region. The linear part of the ensembled model is  $\mu_t^{\text{eff}} = 1.334\mu_t$ . This is lower than the linear optimization above ( $\approx 1.8$ ). However one can see near the blade tip, the effect of Eq. 18a, where this coefficient is raised to approximately 1.65. This is the catharsis to the skew in error and linearity violations observed when assuming a constant  $\beta$  (see Fig. 2 and Fig. 4). Indeed, because of the high alignment of  $a_{ij}$  and



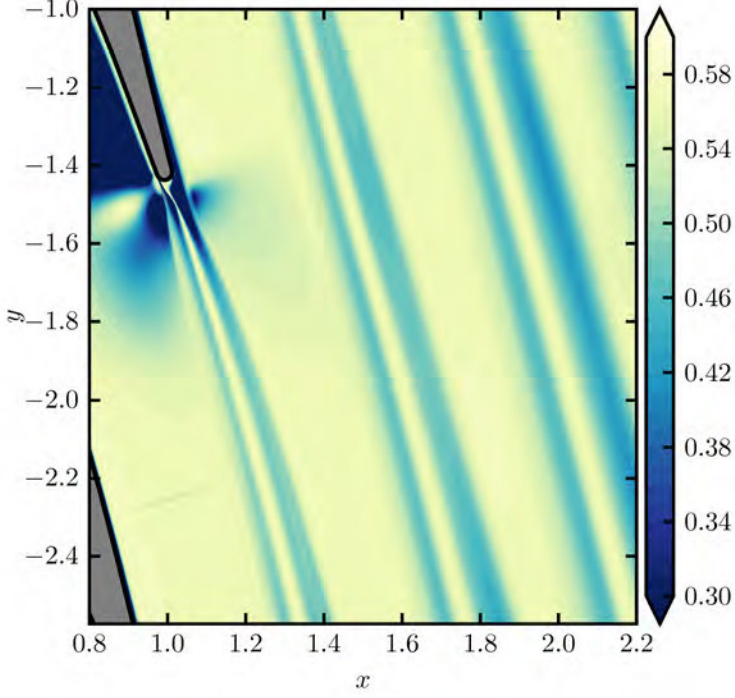
**FIGURE 9.** COEFFICIENT OF  $-V_{ij}^1$  IN THE ENSEMBLED NON-LINEAR SOLUTION.

$-S'_{ij}$  (Fig. 3) this is a fruitful addition through the assumption of Eq. 2c.

Of the remaining coefficients, Fig. 10 and Fig. 11 both show a similar structure, yet very different variation in magnitude. The  $V_{ij}^3$  coefficient is approximately constant around 12.861, whilst that of  $V_{ij}^2$  shows more relative variation below the coefficient 0.573 (Eq. 19). Note these coefficients vary by two orders of magnitude, a reflection of the basis tensor magnitude. This near constant behavior for  $V_{ij}^3$ , in the training region, confirms the discussion above — that there appears to be no systemic functional dependence on  $V_{ij}^3$  (that the algorithm can find.) Rather, this term has been used, differently in each optimization, to counter mistakes and overshoots induced by the use of the other terms. Further, despite the lower-than-random selection of  $V_{ij}^2$ , we can see more meaningful variation. This implies that simpler terms have been easy to include for modest gains, yet more difficult terms have been harder to find.

## CONCLUSIONS AND IMPLICATIONS FOR MODEL IMPROVEMENT

Using highly resolved LES data, we began this discussion by showing that in the wake of an HPT blade, the standard turbulent diffusion in the linear model is too low. An *a priori* optimized effective coefficient, approximately  $\mu_t^{\text{eff}} \approx 1.8\mu_t$  when including

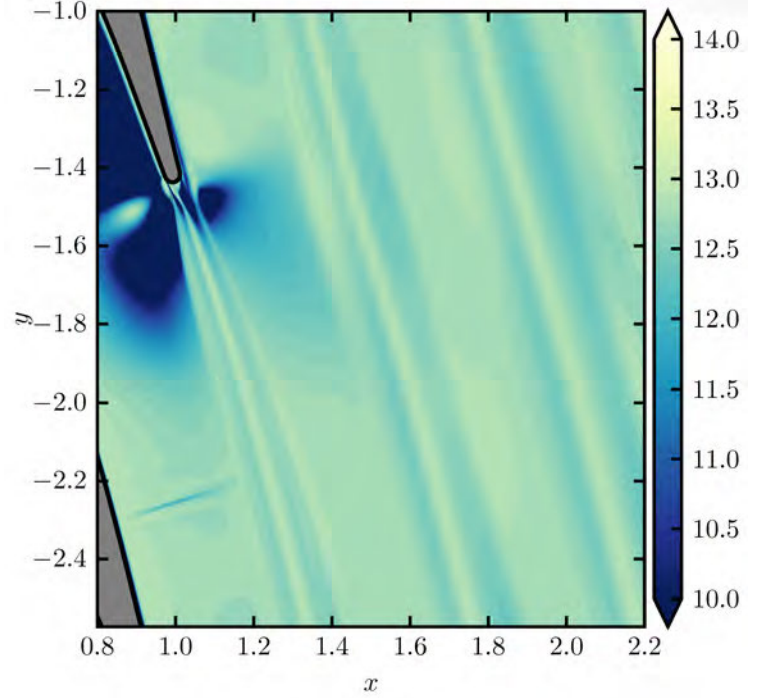


**FIGURE 10.** COEFFICIENT OF  $V_{ij}^2$  IN THE ENSEMBLED NON-LINEAR SOLUTION.

the near-wake in the training data, can improve the mean square error by approximately 6%. This coefficient was dominated by the near-wake, the inclusion of more or less far-wake data did not yield much variation in value. Indeed, the assumptions of least-squares regression were progressively more violated (see the kurtosis in Table 1). This analysis shows that a linear model will not be able to reproduce the anisotropy for this case and theoretically a linear relationship is invalid.

Using the same highly-resolved LES data, we lowered this error further by the inclusion of non-linear terms. Through the analysis of  $\widehat{a_{ij}^{lep}}$ , an ensembled solution consisting of 100 minimizations using our GEP framework, we showed that a non-constant coefficient for the  $-V_{ij}^1$  term was able to adjust to the near-wake. Coefficients for  $V_{ij}^2, V_{ij}^3$  appear to vary less tangential to the flow direction.

This study has been a useful exercise in employing a symbolic regression algorithm in an unusual way. By focussing largely on the physical equations produced, we have utilized one of the hidden benefits of a symbolic algorithm — that we can see ‘under the hood’ and understand the terms that best model the physics. We also can see some of the downsides; the ensembled model contains terms that are unlikely to provide stable/robust/accurate predictions in a CFD environment. Consider the



**FIGURE 11.** COEFFICIENT OF  $V_{ij}^3$  IN THE ENSEMBLED NON-LINEAR SOLUTION.

fragment in Eq. 18c,

$$-10204.0816I_1^4 + 10204.0816I_1^3I_2. \quad (20)$$

This term is likely an anathema to a CFD solver; for this case  $I_1 \approx I_2$ , however in general this is not necessarily true. Such large coefficients would produce extremely non-realizable and unstable contributions to the Reynolds stress. Eliminating terms such as this automatically is the primary obstacle for the production of CFD ready anisotropy functions obtained directly from machine learning frameworks.

Finally, in terms of proposing model changes from this study, we suggest (tentatively) a stress-strain relationship,

$$\begin{aligned} \tau_{ij} = & \frac{2}{3}\rho k\delta_{ij} - \\ & 2(1.334 - 0.438I_1 - 2.635I_2 - 0.0102I_1^2 + 1.021I_2^2 - \\ & 12.280I_1I_2 - 102.047I_1I_2^2)\mu_t S'_{ij} + \\ & 1.142\rho kV_{ij}^2 + 25.722\rho kV_{ij}^3, \quad (21) \end{aligned}$$

which is the inclusion of Eq. 18a and the constant from Eq. 18b. One may also start from just the linear term and ignore  $V_{ij}^2$  for marginal improvements over the optimized linear model. This

equation is likely requiring limiters, based on the relationship between Eq. 18a and 0.09, perhaps taking the maximum of these two quantities. Evident from Fig. 6, one may also attempt to use the model Eq. 19, this model is only quadratic in the velocity gradient and therefore is more robust than Eq. 21, whilst only sacrificing some accuracy.

## ACKNOWLEDGMENT

We acknowledge the funding provided by veski. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. This work was also supported by resources provided by The Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

## REFERENCES

- [1] Arts, T., Lambertderouvroit, M., and Rutherford, A., 1990. "Aero-thermal investigation of a highly loaded transonic linear turbine guide vane cascade. a test case for inviscid and viscous flow computations". *NASA STI/Recon Technical Report N*, **91**, p. 23437.
- [2] Wheeler, A. P., Sandberg, R. D., Sandham, N. D., Pichler, R., Michelassi, V., and Laskowski, G., 2016. "Direct numerical simulations of a high-pressure turbine vane". *Journal of Turbomachinery*, **138**(7), p. 071003.
- [3] Dunn, M. G., 2001. "Convective heat transfer and aerodynamics in axial flow turbines". In ASME Turbo Expo 2001: Power for Land, Sea, and Air, V004T06A001, American Society of Mechanical Engineers.
- [4] Pichler, R., Kopriva, J., Laskowski, G., Michelassi, V., and Sandberg, R., 2016. "Highly resolved LES of a linear HPT vane cascade using structured and unstructured codes". In ASME Turbo Expo 2016, V02CT39A041, American Society of Mechanical Engineers.
- [5] Hunt, J., and Savill, A., 2005. "Guidelines and criteria for the use of turbulence models in complex flows". *Prediction of Turbulent Flows*, pp. 291–343.
- [6] Schmitt, F. G., 2007. "About Boussinesq's turbulent viscosity hypothesis: historical remarks and a direct evaluation of its validity". *Comptes Rendus Mécanique*, **335**(9), pp. 617–627.
- [7] Leschziner, M., 2015. *Statistical turbulence modelling for fluid dynamics — demystified: an introductory text for graduate engineering students*. World Scientific.
- [8] Rodi, W., 1976. "A new algebraic relation for calculating the Reynolds stresses". In Gesellschaft Angewandte Mathematik und Mechanik Workshop Paris France, Vol. 56, p. 219.
- [9] Pope, S., 1975. "A more general effective-viscosity hypothesis". *Journal of Fluid Mechanics*, **72**(02), pp. 331–340.
- [10] Gatski, T., and Speziale, C., 1993. "On explicit algebraic stress models for complex turbulent flows". *Journal of Fluid Mechanics*, **254**, pp. 59–78.
- [11] 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.
- [12] Abe, K., Jang, Y.-J., and Leschziner, M. A., 2003. "An investigation of wall-anisotropy expressions and length-scale equations for non-linear eddy-viscosity models". *International Journal of Heat and Fluid Flow*, **24**(2), pp. 181–198.
- [13] Weatheritt, J., and Sandberg, R. D., 2016. "A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship". *Journal of Computational Physics*, **325**, pp. 22–37.
- [14] Ferreira, C., 2001. "Gene expression programming: A new adaptive algorithm for solving problems". *Complex Systems*, **13**(2), pp. 87–129.
- [15] Weatheritt, J., and Sandberg, R., 2015. "A new Reynolds stress damping function for hybrid RANS/LES with an evolved functional form". *Advances in Computation, Modeling and Control of Transitional and Turbulent Flows*, p. 330.
- [16] Tracey, B., Duraisamy, K., and Alonso, J. J., 2015. "A machine learning strategy to assist turbulence model development". *AIAA Paper*, **1287**, p. 2015.
- [17] Ling, J., and Templeton, J., 2015. "Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty". *Physics of Fluids (1994-present)*, **27**(8), p. 085103.
- [18] Parish, E. J., and Duraisamy, K., 2016. "A paradigm for data-driven predictive modeling using field inversion and machine learning". *Journal of Computational Physics*, **305**, pp. 758–774.
- [19] Weatheritt, J., and Sandberg, R. D., 2016. "The development of algebraic stress models using a novel evolutionary algorithm". In 11<sup>th</sup> International ERCOFTAC Symposium on Engineering Turbulence Modelling and Measurements.
- [20] Nicoud, F., and Ducros, F., 1999. "Subgrid-scale stress modelling based on the square of the velocity gradient tensor". *Flow, turbulence and Combustion*, **62**(3), pp. 183–200.
- [21] Sandberg, R. D., Michelassi, V., Pichler, R., Chen, L., and Johnstone, R., 2015. "Compressible direct numerical simulation of low-pressure turbines — part I: Methodology". *Journal of Turbomachinery*, **137**(5), p. 051011.
- [22] Muldoon, F., and Acharya, S., 2006. "Analysis of k and epsilon budgets for film cooling using direct numerical simulation". *AIAA journal*, **44**(12), pp. 3010–3021.
- [23] Spalart, P. R., Shur, M. L., Strelets, M. K., and Travin, A., 1997. "A cubic law for the near-wall turbulent velocity profile". *Journal of Fluid Mechanics*, **361**, pp. 75–105.

- A. K., 2015. “Direct simulation and RANS modelling of a vortex generator flow”. *Flow, Turbulence and Combustion*, **95**(2-3), pp. 335–350.
- [24] Pichler, R., Sandberg, R., 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), p. 121009.
- [25] Menter, F. R., 1994. “Two-equation eddy-viscosity turbulence models for engineering applications”. *AIAA journal*, **32**(8), pp. 1598–1605.
- [26] Parneix, S., Laurence, D., and Durbin, P., 1998. “A procedure for using DNS databases”. *Journal of fluids engineering*, **120**(1), pp. 40–47.
- [27] Koza, J. R., 1992. *Genetic programming: on the programming of computers by means of natural selection*, Vol. 1. MIT press.
- [28] Steeb, W.-H., 2014. *The nonlinear workbook: chaos, fractals, cellular automata, genetic algorithms, gene expression programming, support vector machine, wavelets, hidden Markov models, fuzzy logic with C++, Java and SymbolicC++ programs*. World Scientific Publishing Co Inc.
- [29] Bode, C., Aufderheide, T., Kožulović, D., and Friedrichs, J., 2014. “The effects of turbulence length scale on turbulence and transition prediction in turbomachinery flows”. In ASME Turbo Expo 2014, V02BT39A042, American Society of Mechanical Engineers.





Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Weatheritt, J; Pichler, R; Sandberg, RD; Laskowski, G; Michelassi, V

**Title:**

Machine Learning for Turbulence Model Development Using a High-Fidelity HPT Cascade Simulation

**Date:**

2017-01-01

**Citation:**

Weatheritt, J., Pichler, R., Sandberg, R. D., Laskowski, G. & Michelassi, V. (2017). Machine Learning for Turbulence Model Development Using a High-Fidelity HPT Cascade Simulation. Proceedings of the ASME Turbo Expo: Turbine Technical Conference and Exposition, 2017, VOL 2B, 2B-2017, American Society of Mechanical Engineers.  
<https://doi.org/10.1115/GT2017-63497>.

**Persistent Link:**

<http://hdl.handle.net/11343/241901>

**File Description:**

Accepted version