

**GT2016-56191**

**UNCERTAINTY ANALYSIS AND DATA-DRIVEN MODEL ADVANCES FOR A JET-IN-CROSSFLOW**

**Julia Ling**

Thermal/Fluid Science and Engineering  
 Sandia National Laboratories  
 Livermore, CA 94551  
 Email: jlingle@sandia.gov

**Anthony Ruiz**

Reacting Flow Research  
 Sandia National Laboratories  
 Livermore, CA 94551

**Guilhem Lacaze**

Reacting Flow Research  
 Sandia National Laboratories  
 Livermore, CA 94551

**Joseph Oefelein**

Reacting Flow Research  
 Sandia National Laboratories  
 Livermore, CA 94551

**ABSTRACT**

For film cooling of combustor linings and turbine blades, it is critical to be able to accurately model jets-in-crossflow. Current Reynolds Averaged Navier Stokes (RANS) models often give unsatisfactory predictions in these flows, due in large part to model form error, which cannot be resolved through calibration or tuning of model coefficients. The Boussinesq hypothesis, upon which most two-equation RANS models rely, posits the existence of a non-negative scalar eddy viscosity, which gives a linear relation between the Reynolds stresses and the mean strain rate. This model is rigorously analyzed in the context of a jet-in-crossflow using the high fidelity Large Eddy Simulation data of Ruiz et al. (2015), as well as RANS  $k-\epsilon$  results for the same flow. It is shown that the RANS models fail to accurately represent the Reynolds stress anisotropy in the injection hole, along the wall, and on the lee side of the jet. Machine learning methods are developed to provide improved predictions of the Reynolds stress anisotropy in this flow.

**NOMENCLATURE**

$a_{ij}$  Reynolds stress anisotropy tensor

$d$  Hole diameter

$k$  Turbulent kinetic energy

$S_{ij}$  Mean strain rate tensor  $\frac{1}{2}(\frac{dU_i}{dx_j} + \frac{dU_j}{dx_i})$

$u'_i$   $i^{th}$  component of the velocity fluctuations

$U_i$   $i^{th}$  component of the mean velocity field

$U_{jet}$  Bulk jet velocity

$U_\infty$  Freestream crossflow velocity

$x_B, y_B$  Barycentric coordinates

$\epsilon$  Turbulent dissipation rate

$\Pi_a$  Second invariant of the Reynolds stress anisotropy tensor

$\lambda$  Eigenvalue of the Reynolds stress anisotropy tensor

$\rho$  Density

$\nu_t$  Eddy viscosity

**1 INTRODUCTION**

Jets-in-crossflow occur in multiple contexts in turbomachinery flows, including in the film cooling of turbine blades and the combustor lining, and in fuel injection. Multiple studies have shown that current Reynolds Averaged Navier Stokes (RANS) models are insufficient for accurate heat transfer and velocity

predictions in these flows [1–6]. Hoda and Acharya [1] evaluated several RANS models for prediction of a jet-in-crossflow, and reported that all models over-predicted the velocity on the lee-side of the jet. He et al. [2] reported that their  $k$ - $\varepsilon$  RANS simulation under-predicted the turbulence intensity in their jet-in-crossflow configuration. Muppudi and Mahesh [3] suggested that many RANS models would struggle with jets-in-crossflow because of the non-isotropic, non-equilibrium, three-dimensional nature of the turbulence in these flows. Coletti et al. [4] compared experimental results to realizable  $k$ - $\varepsilon$  RANS results for an inclined jet-in-crossflow and showed that RANS under-predicted the strength of the counter rotating vortex pair and significantly over-predicted the centerline film cooling effectiveness. Harrison and Bogard [5] compared heat transfer predictions for multiple RANS turbulence models for a film cooling flow and showed that no model was accurate in all regions of the flow. Ling et al. [7] analyzed Large Eddy Simulation (LES) results for an inclined jet-in-crossflow and showed that a fixed turbulent Prandtl number paired with the gradient diffusion hypothesis would not accurately predict the magnitude or direction of the turbulent heat fluxes. These results reflect the general consensus that current RANS models do not yield satisfactory jet-in-crossflow predictions.

Several efforts have been made to improve RANS predictions for jets-in-crossflow. Ray et al. [8, 9] used Bayesian methods to calibrate  $k$ - $\varepsilon$  model parameters for a jet-in-crossflow and showed improved accuracy with this calibrated model. However, even the calibrated model was not accurate in all regions of the flow, particularly at higher blowing ratios. Ling et al. [10] used experimental results to optimize the turbulent diffusivity for an inclined jet-in-crossflow, and demonstrated improved, but still imperfect, heat transfer predictions with this tuned diffusivity. However, simply tuning the model parameters does not address model form uncertainty, which is the uncertainty due to the underlying model assumptions. Therefore, the success of calibrated RANS models will continue to be limited in flows where the model assumptions are invalid.

More complex RANS models have also been investigated in an effort to achieve improved predictions. Hoda and Acharya [1] evaluated two non-linear eddy viscosity models for jet-in-crossflow simulations and reported that neither produced significantly improved accuracy. They attributed this unimproved accuracy to the fact that most non-linear models are calibrated to simple wall-bounded flows, quite unlike the jet-in-crossflow. Kaszeta and Simon used triple wire anemometry to measure the Reynolds stresses in a film cooling configuration and revealed significant anisotropy in the stresses [11]. Subsequent efforts to implement more sophisticated RANS models capable of more accurately modeling this anisotropy have met with some success. Rajabi-Zagarabadi and Bazdidi-Tehrani [12] demonstrated improved heat transfer predictions using their implicit algebraic heat transfer model. Similarly, Azzi and Lakehal [13] showed

that using anisotropic eddy viscosity and eddy diffusivity models led to improved predictions of the lateral spread of the film coolant. These studies show that understanding which assumptions are violated in key regions of the flow can lead to improved model selection and enhanced predictivity.

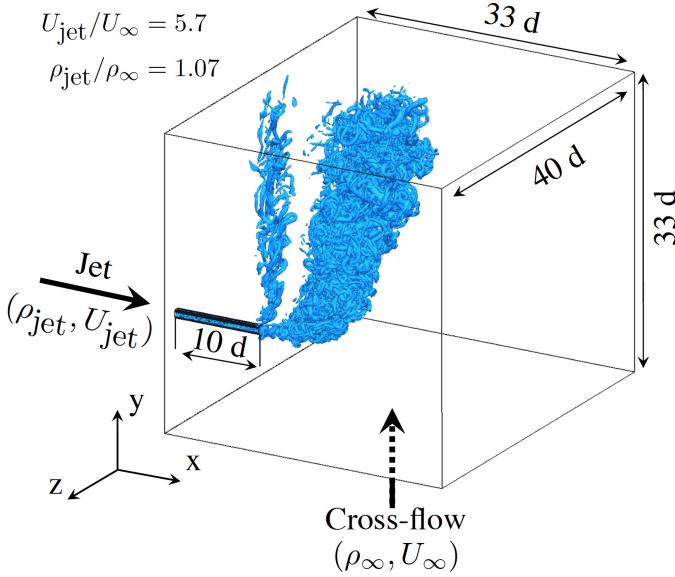
Given the observations above, it is useful to investigate in more detail the validity of RANS assumptions for the jet-in-crossflow. This paper will focus on RANS models that rely on the Boussinesq hypothesis of a linear eddy viscosity

$$\overline{u'_i u'_j} = -2\nu_t S_{ij} + \frac{2}{3}k\delta_{ij}. \quad (1)$$

In this equation,  $\overline{u'_i u'_j}$  is the Reynolds stress tensor,  $\nu_t$  is the eddy viscosity,  $k$  is the turbulent kinetic energy, and  $S_{ij}$  is the mean strain rate tensor. Implicit in Eq.(1) are two underlying assumptions: i) the eddy viscosity is non-negative, and ii) the mean strain rate tensor adequately captures the anisotropy of the Reynolds stresses. Once it has been determined where these model assumptions are violated, it would be desirable to correct the models to mitigate these sources of model form error. This paper will present machine learning methods that can be used to determine more accurate closures for the Reynolds stresses.

Machine learning encompasses a broad set of data-driven algorithms, including familiar methods such as linear regression as well as more advanced concepts such as neural networks, random forests, and support vector machines. These methods have been broadly applied in many fields, such as finance [14], marketing [15], and image recognition [16]. Machine learning methods have also recently been employed for several turbulence modeling applications. Tracey et al. [17] used non-parametric data-driven methods to model the Reynolds stress anisotropy in a converging diverging channel and a non-equilibrium boundary layer. These methods showed improved anisotropy predictions when tested on the same flow on which they were trained, but significant inaccuracy on other flows. Duraisamy et al. [18] used neural networks to predict an intermittency factor to improve turbulence transition simulations. Ling and Templeton [19] developed a suite of machine learning classifiers that can predict when different RANS modeling assumptions are violated. These classifiers were shown capable of generalizing to flows significantly different from those on which they were trained.

In this paper, the highly resolved LES jet-in-crossflow results of Ruiz et al. [20] and corresponding RANS results for the same flow are analyzed in depth. The objective is to determine in which regions of this flow the various RANS eddy viscosity assumptions are violated, and to explore the potential of machine learning techniques to provide improved models. Section 2 presents the flow configuration and computational set-up for the LES and RANS simulations. Section 3 presents analysis of the LES results that shows in which regions of the flow the RANS



**FIGURE 1.** Schematic of LES flow configuration showing instantaneous isosurfaces of the Q-criterion.

Boussinesq assumption is violated. In Section 4, the ability of Random Forests to improve the Reynolds stress anisotropy predictions is explored, and Section 5 presents the conclusions of these investigations.

## 2 Computational Methodology

### 2.1 Flow Configuration

The flow configuration is based on that of the experiments of Su and Mungal [21], and is shown in Figure 1. The jet is injected perpendicularly into the crossflow and the jet Reynolds number based on the jet bulk velocity  $U_{jet}$  and hole diameter  $d$  is 5000. The blowing ratio is  $\frac{\rho_{jet}U_{jet}}{\rho_{\infty}U_{\infty}} = 6.1$  and the density ratio is  $\frac{\rho_{jet}}{\rho_{\infty}} = 1.07$ . The flow is in a low-Mach, incompressible regime. In this configuration, the  $x$ -axis is aligned with the jet injection, the  $y$ -axis is aligned with the crossflow direction, and the  $z$ -axis is in the spanwise direction.

### 2.2 LES Set-up

Details of the LES computational methodology can be found in Ref. [20]. A brief synopsis will be provided here. The LES was performed using RAPTOR [22], an in-house solver that uses a finite-volume framework with non-dissipative numerical methods. A mixed dynamic Smagorinsky SGS model was employed to model the subgrid-scale (SGS) stresses.

A fine isotropic and uniform grid spacing ( $\Delta = d/15$ ) was used, which yielded a mesh containing 190 million cells. The

jet pipe was 10d long, and a fully developed velocity profile was prescribed at the jet inlet. The channel inlet was set 5d upstream of the injection hole, and a Blasius profile was used to prescribe the mean streamwise velocity at the channel inlet. The boundary layer thickness,  $\delta = 1.025d$ , was set to match the conditions of the experiments of Su and Mungal [21].

Figure 1 shows an instantaneous isosurface of the Q-criterion calculated by this LES. A wide range of structures were identified, including ring-vortices, v-shape secondary instabilities, the counter-rotating vortex-pair, hair-pin vortices at the wall, and far-field turbulence. The complex turbulence dynamics exhibited in this LES reinforce the difficulty associated with modeling this flow using RANS.

In Ref. [20], an extensive validation of the LES results with experimental data was performed, which demonstrated that turbulence is accurately resolved by this high-fidelity LES. Figure 2 presents profiles of mean velocity from the LES and from the experiments of Su and Mungal [21]. The mean relative error between the numerical results and the hot-wire measurements was 7% across the profile locations shown, confirming the good agreement between the LES and experiment. *An a posteriori* analysis also showed that on average 99% of the turbulent kinetic energy was directly resolved and only 1% of the turbulent kinetic energy was modeled via the SGS model, indicating that this was a well-resolved simulation. This high fidelity simulation data was critical in this analysis, since the analysis required knowledge of all components of the Reynolds stresses and velocity gradients throughout the flow; experimental acquisition of such a complete data set would have been infeasible.

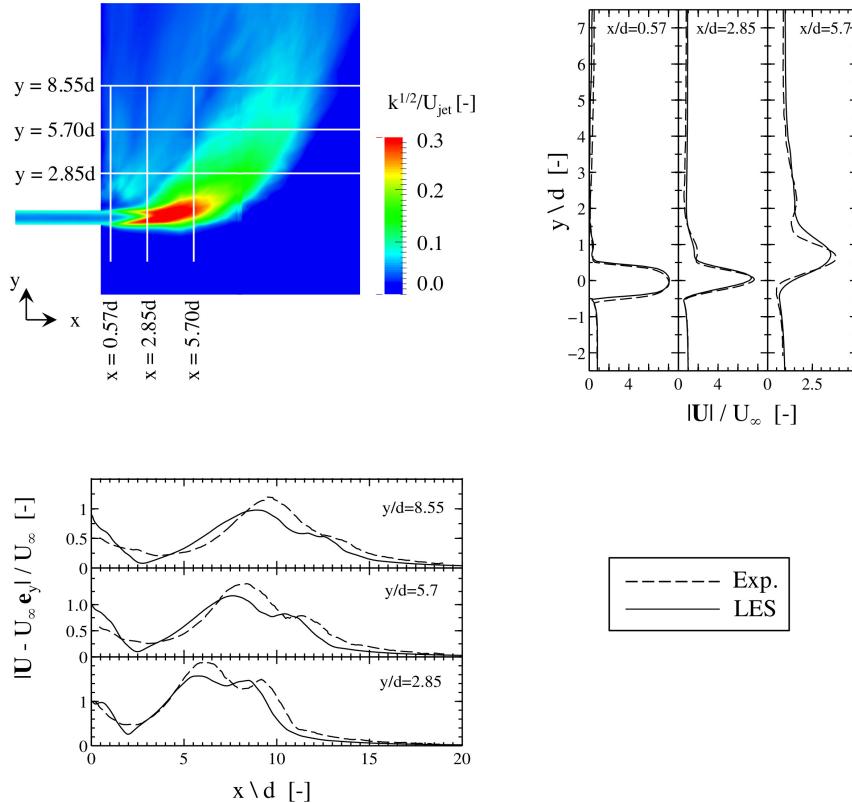
### 2.3 RANS Set-up

The RANS simulation has been previously reported in Ref. [19]. An in-house Sandia solver, Sierra Fuego, was used to run the RANS simulation using the  $k-\epsilon$  turbulence model. The same computational domain as in the LES was used, with the same mean flow boundary conditions. An unstructured mesh with 5-million hexahedral cells was used, as shown in Fig. 3. A mesh refinement study was conducted, and it was shown that when RANS simulations run on a 7.5-million cell mesh were compared to those run on the 5-million cell mesh, the turbulent kinetic energy and mean velocity field both differed by less than 3% between the two meshes. These results confirmed grid convergence on the 5-million cell mesh.

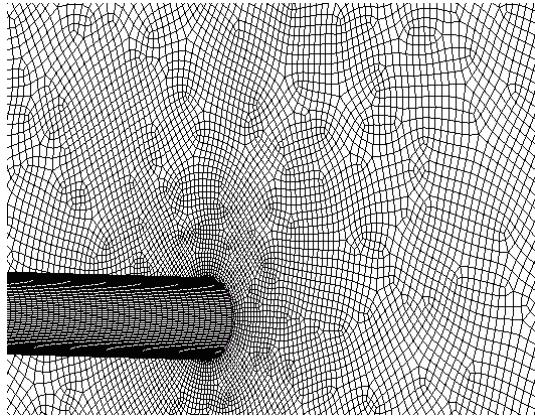
## 3 Analysis of Model Form Uncertainty

### 3.1 Comparison of RANS and LES Results

It is useful to begin by comparing the RANS and LES results to determine to what extent RANS is able to accurately capture the flow field and turbulence quantities. Figure 4 shows contours of x-velocity and turbulent kinetic energy  $k$  from LES and RANS



**FIGURE 2.** Time average profiles of the velocity magnitude in the jet center plane ( $z/d=0$ ). Reference data labeled “Exp” were extracted from the experiment of Su and Mungal [21].



**FIGURE 3.** Zoomed-in view of near-injection region of mesh used for RANS simulations

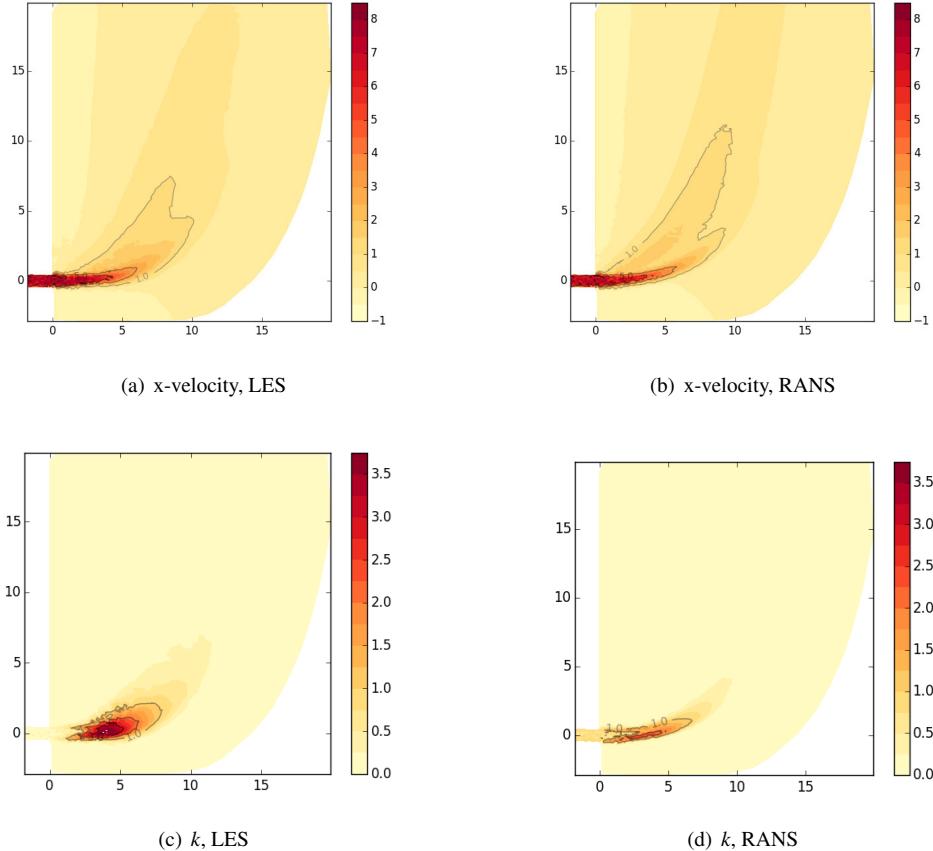
in a plane located at  $z = 0.25d$ . The contours are shown at this plane instead of the mid-plane at  $z = 0$  in order to avoid the singularities and atypical behavior that are often observed at symmetry

planes. As this figure shows, RANS over-predicts the penetration of the jet into the cross flow. This over-prediction could be due to under-prediction of the turbulent mixing in the near-injection region, as indicated by the severely under-predicted levels of turbulent kinetic energy in this region. These results are in agreement with those of He et al. [2], who also reported that RANS  $k-\varepsilon$  under-predicted the turbulence intensity in their jet-in-crossflow configuration. In the next section, the root causes of these inaccuracies are investigated.

### 3.2 Violation of Key RANS Assumptions

In order to determine the cause of the inaccurate RANS predictions, it is useful to analyze when different RANS model assumptions are violated. Many of these assumptions can be directly evaluated using the LES data.

One key assumption common to all eddy viscosity models is that the eddy viscosity is non-negative. Indeed, in  $k-\varepsilon$  models,  $\nu_t = \frac{C_\mu k^2}{\varepsilon}$ , where  $C_\mu$  is a model parameter, commonly set at a value of 0.09. In this formulation,  $\nu_t$  is guaranteed to be positive, since  $\varepsilon$  is restricted to be positive in most solvers. However, in



**FIGURE 4.** Contours of x-velocity (a,b) and turbulent kinetic energy (c,d) as predicted by LES (a,c) and RANS (b,d). Isocontour lines super-imposed in gray. The velocity and turbulent kinetic energy have been non-dimensionalized with the mean free stream velocity  $U_\infty$ . Contours shown in a spanwise plane at  $z = 0.25d$ .

real flows, it is possible for momentum to be transported counter to a velocity gradient, leading to negative eddy viscosities. An eddy viscosity field can be extracted from the LES as [7, 23]

$$\nu_t = \frac{-\overline{u'_i u'_j} S_{ij} + \frac{2}{3} k \delta_{ij} S_{ij}}{2 S_{kl} S_{kl}}, \quad (2)$$

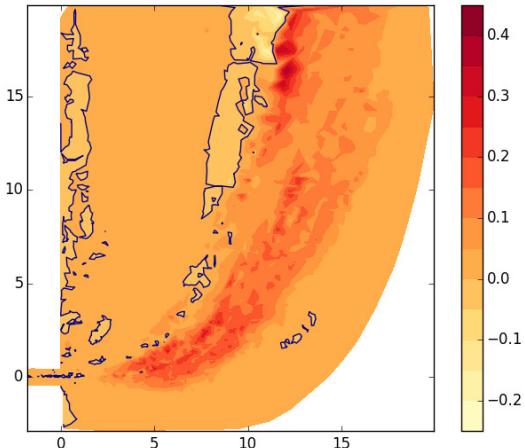
where

$$\overline{u'_i u'_j} = \overline{u'_i u'_{j,res}} + \frac{2}{3} k_{SGS} \delta_{ij} - 2 \nu_{t,SGS} S_{ij}. \quad (3)$$

In Eq.(3),  $\overline{u'_i u'_{j,res}}$  are the resolved Reynolds stresses, and  $k_{SGS}$  and  $\nu_{t,SGS}$  are the SGS turbulent kinetic energy and turbulent viscosity, respectively. Because the LES is highly resolved, the SGS contribution to the Reynolds stresses is minimal ( $\approx 1\%$ ).

Figure 5 shows contours of the extracted eddy viscosity. As shown in the figure, there are significant regions of this flow with negative eddy viscosities. The eddy viscosity is negative in the near wall region downstream of injection, as well as in the shear layer on the lee-side of the jet. These results show that in these regions, an eddy viscosity model will not accurately capture the turbulent transport. Analysis of the cause of the negative eddy viscosity on the lee-side of the jet suggests that it is due in part to anisotropy in the normal stresses in this region. The normal Reynolds stress in the x-direction  $\overline{u_1 u_1}$  is greater than  $\frac{2}{3} k$  and  $S_{11} = \frac{du}{dx}$  is greater than 0 on the lee-side of the jet, leading to the calculation of a negative eddy viscosity. This anisotropy is not unexpected given that previous researchers have noted the presence of large coherent vortices in jet-in-crossflow wakes [24].

A related assumption in the Boussinesq hypothesis (Eq.(1)) is that the mean strain rate tensor accurately captures the Reynolds stress anisotropy. The non-dimensional Reynolds



**FIGURE 5.** Contours of eddy viscosity, non-dimensionalized by  $U_\infty$  and  $d$ . Regions of negative eddy viscosity are outlined with solid lines.

stress anisotropy tensor  $a_{ij}$  is defined as

$$a_{ij} = \frac{\bar{u}_i \bar{u}_j - \frac{2}{3} k \delta_{ij}}{2k}. \quad (4)$$

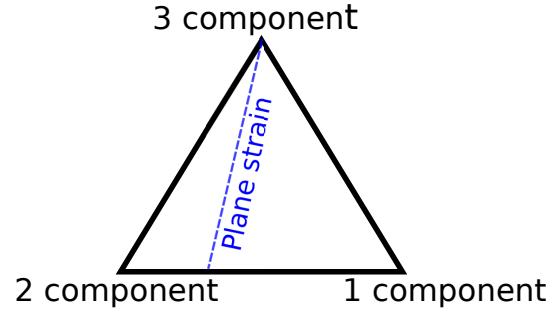
Based on the Boussinesq hypothesis, RANS predicts that

$$a_{ij, \text{RANS}} = \frac{-v_t S_{ij}}{k}. \quad (5)$$

One method of assessing the accuracy of the RANS anisotropy predictions is to visualize the anisotropy on a barycentric map [25]. The barycentric map plots the Reynolds stress anisotropy on a triangle that represents the realizable states of turbulence. The top corner of the triangle represents 3-component turbulence, the bottom left corner represents 2-component turbulence, and the bottom right corner represents 1-component turbulence. A schematic of this map is shown in Fig. 6. This figure also shows a dashed line, representing plane strain states of the Reynolds stresses. In 2-D flow, RANS would predict the Reynolds stresses to lie entirely along this dashed line.

Figure 7(a, b) shows the barycentric maps of the Reynolds stress anisotropy, both as calculated using LES, and as predicted by RANS, for randomly selected points in this flow. As these plots show, the RANS model does a poor job of predicting the Reynolds stress anisotropy in this flow, predicting far too many points lying near the 3-component limit at the top of the triangle, and not enough points in the 1-component and 2-component limits at the bottom corners of the triangle.

In order to visualize in which regions of the flow RANS pre-



**FIGURE 6.** Schematic of the barycentric map

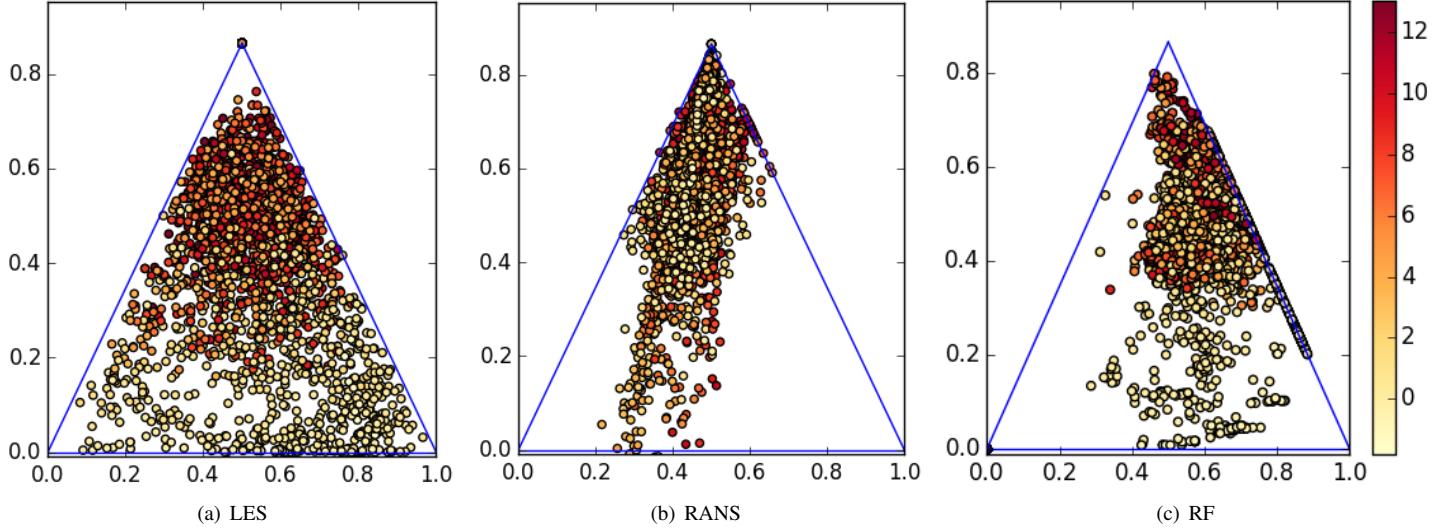
dicts the anisotropy poorly, Figure 8(a, b) shows contours of the second invariant of the anisotropy tensor  $\Pi_a = a_{ij} a_{ji}$  as predicted by RANS and LES. This invariant is a useful indicator of the degree of stress anisotropy: it ranges in value from 0 (indicating isotropic turbulence) up to 2/3 (indicating a high degree of anisotropy) [26,27]. As Fig. 8 demonstrates, RANS cannot satisfactorily model the Reynolds stress anisotropy. It misses several regions of high anisotropy along the wall, in the injection hole, upstream of the jet, and on the lee-side of the jet. It also predicts falsely high values of  $\Pi_a$  in the upstream shear layer. Based on these results, it is clear that there is significant room for improvement in the RANS predictions of the Reynolds stress anisotropy.

## 4 Machine Learning for Reynolds Stress Anisotropy Predictions

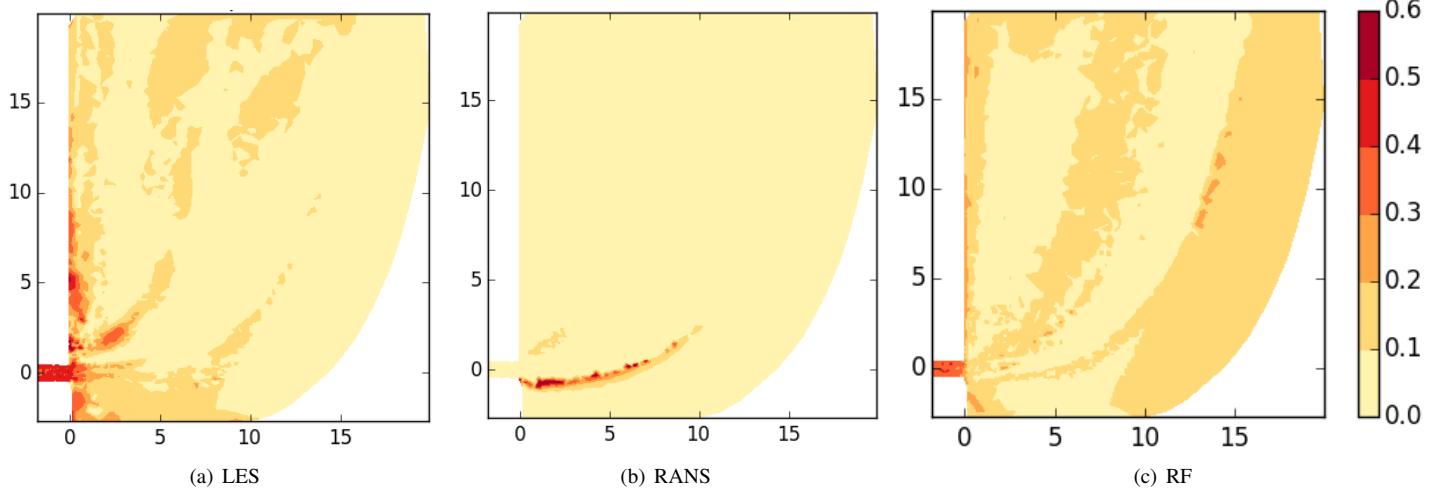
### 4.1 Machine Learning Algorithm

Machine learning algorithms are data-driven methods that can be applied for clustering, classification, and regression [16]. In the present study, Random Forest (RF) regressors were employed to predict the barycentric coordinates  $(x_B, y_B)$  of the Reynolds stress anisotropy. This algorithm uses *supervised learning*: the model is trained on data for which the correct answer is known. The training data used in this study are presented in Section 4.2. Random Forests are composed of an ensemble of binary decision trees. Each decision tree uses an if-then logic to categorize points based on a series of binary branches. While individual decision trees are susceptible to over-fitting, ensembles of multiple decision trees have been shown to be both robust and high-performing [28]. Random Forests are ensembles of decision trees where each tree is trained on a random subset of the training data and the random subset is sampled with replacement from the original training data in a strategy known as *bagging* [29].

Each tree in the ensemble was allowed to grow to its full depth, allowing for maximal tree diversity. Therefore, the only hyper-parameter for this algorithm was the number of trees in the ensemble. In general, the performance of the RF will improve as



**FIGURE 7.** Scatter plot of Reynolds stress anisotropy predictions on a barycentric map for (a) LES, (b) RANS, and (c) Random Forest. Points colored by x-coordinate normalized by hole diameter  $d$ .

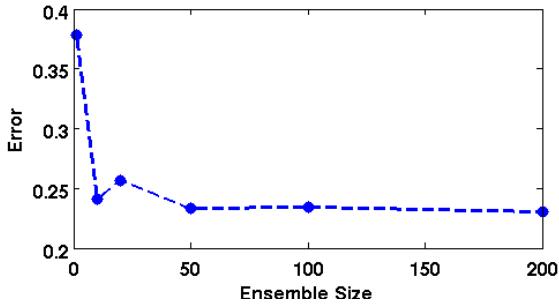


**FIGURE 8.** Contours of the second anisotropy invariant  $\text{II}_a$  from (a) LES, (b) RANS, and (c) Random Forest predictions in a plane at  $z = 0.25d$ .

the number of trees in the ensemble grows larger, but with diminishing returns. On the other hand, a larger ensemble size imposes larger computational cost and memory usage requirements. Figure 9 shows RF model error in predicting  $x_B$  as a function of ensemble size. As shown in the figure, there is a general trend of decrease in error as the ensemble size increases, but for ensemble sizes greater than 50, there is not a strong dependence of error on size. The ensemble size was therefore set to 100 to avoid any strong dependence of performance on ensemble size.

While there are many possible choices of regression algo-

rithms, RF regressors were chosen for this application because they are robust to noise, they do not require input feature pre-processing or feature selection, they have only one tunable hyper-parameter, and they can handle the non-linear decision boundaries that would be expected in turbulence modeling. In comparison, linear regression, while robust and simple, is limited in its ability to handle non-linear behavior. Neural networks, on the other hand, are well suited to non-linear decision boundaries, but have many tunable hyper-parameters such as network size, architecture, regularization scheme, and activation function. Addi-



**FIGURE 9.** Error in Random Forest predictions of  $x_B$  for varying Random Forest ensemble sizes.

tionally, neural networks can be very computationally intensive to train because training them requires iteratively optimizing a non-convex cost function. Ling and Templeton [19] showed that RF classifiers performed well in comparison to Adaboost Decision Trees and Support Vector Machines in detecting regions of high RANS uncertainty. RFs, therefore, represent an attractive balance between robustness, ease-of-implementation, and high performance that makes them suitable for this early effort at applying machine learning methods to turbulence modeling.

## 4.2 Training Data

If the machine learning model were trained on the same jet-in-crossflow configuration upon which it was tested, that would provide limited insight into the ability of the model to generalize to new flows for which high fidelity results may not be available. Therefore, the machine learning model was trained on data sets from two very different flows: flow around a wall-mounted cube and fully developed turbulent duct flow. These two data sets were chosen because there is Direct Numerical Simulation data available for both flows, and they contain many of the relevant flow regimes.

The flow around the wall-mounted cube data has been previously presented by Rossi et al. [30, 31]. The Reynolds number was 5,000, based on the cube height and mean free stream velocity. This data set has regions of stagnation and impingement on the upstream face of the cube, flow curvature, and separation and re-attachment on the leeward side of the cube. It therefore contains many of the challenging three-dimensional and anisotropic regimes that are encountered in jets-in-crossflow.

The duct flow DNS data was presented by Pinelli et al. [32], and represents a square duct at a Reynolds number of 3,500 based on the channel half-height and streamwise bulk velocity. This flow has stress-driven corner vortices that the linear Boussinesq hypothesis model fails to predict. This configuration is therefore a good example of a flow where Reynolds stress anisotropy plays a crucial role in determining key flow structures.

For both of these flows, DNS and RANS  $k - \varepsilon$  data were available. The training data were composed of 5,000 randomly sampled points from each of these two flows, for 10,000 total training points.

## 4.3 Machine Learning Inputs and Outputs

The inputs to the machine learning algorithm were local flow variables from the RANS simulations. The available RANS local flow variables included the mean velocity, the mean velocity gradient, the turbulent kinetic energy, the turbulent kinetic energy gradient, the turbulent dissipation rate, the density, the pressure gradient, the molecular and turbulent viscosities, and the distance to the nearest wall. While it would have been possible to use these raw flow variables as inputs to the machine learning model, the resulting model would have been unlikely to generalize well, since those raw variables are neither non-dimensional nor Galilean invariant.

Ling et al. [33] described a procedure for creating a basis of features that respect invariance properties using concepts from invariant theory and representation theory. This methodology was used to construct a basis of 49 rotationally invariant, translationally invariant, and non-dimensional input features based on the tensor, vector, and scalar raw local flow variables from RANS. While conventional eddy viscosity models predict the Reynolds stresses as a function of only  $k$ ,  $\varepsilon$  and the mean velocity gradient, the machine learning model had information on the mean pressure gradient and the turbulent kinetic energy gradient as well.

The output from the RF model was a real number prediction for the barycentric coordinates  $(x_B, y_B)$ . These coordinates fully define the Reynolds stress anisotropy eigenvalues  $\lambda_1, \lambda_2, \lambda_3$  by the following relations [25]:

$$x_B = \lambda_1 - \lambda_2 + \frac{3}{2}\lambda_3 + \frac{1}{2} \quad (6)$$

$$y_B = \frac{\sqrt{3}}{2}(3\lambda_3 + 1) \quad (7)$$

$$0 = \lambda_1 + \lambda_2 + \lambda_3 \quad (8)$$

The RF model predictions of the barycentric coordinates could therefore be used to modify the eigenvalues of the Reynolds stress anisotropy tensor in order to construct a more accurate Reynolds stress closure. At each point in the flow field, the machine learning model uses the RANS local flow variables to make a prediction about the local Reynolds stress anisotropy. In order to ensure realizability, all RF predictions were constrained to lie in the triangle that delineates the realizable region in the barycentric map. Predictions outside this triangle were moved to the closest point within the triangle.

The models were trained on the aforementioned data from

the cube in crossflow and duct flow data sets, and tested on the Ruiz et al. [20] jet-in-crossflow data set.

#### 4.4 Machine Learning Results

As was discussed in Section 3, the Boussinesq hypothesis leads to inaccurate predictions of the Reynolds stress anisotropy. Visualizations of the barycentric map in Fig. 7(b) showed that the RANS predictions significantly under-predicted the 1-component and 2-component regions of the flow. Figure 7(c) shows the corresponding RF predictions of the barycentric coordinates. It is clear that the RF is not able to completely replicate the LES results for the barycentric map. Nevertheless, the RF does successfully predict the tendency of some regions of the flow to have turbulence approaching the 1-component limit. In order to quantitatively compare the accuracy of the RANS and RF predictions for  $(x_B, y_B)$ , it is useful to evaluate the relative prediction errors  $E_{x_B}$  and  $E_{y_B}$ :

$$E_{x_B} = \frac{\sum_i |x_{B,\text{pred}_i} - x_{B,\text{LES}_i}|}{\sum_i |x_{B,\text{LES}_i}|} \quad (9)$$

$$E_{y_B} = \frac{\sum_i |y_{B,\text{pred}_i} - y_{B,\text{LES}_i}|}{\sum_i |y_{B,\text{LES}_i}|} \quad (10)$$

In Eq.(10) the sum is over all the points in the test data set and  $(x_{B,\text{pred}}, y_{B,\text{pred}})$  denote the predicted values of the barycentric coordinates, either from RANS or the RF. For the RANS predictions,  $E_{x_B} = 0.25$  and  $E_{y_B} = 0.53$ , reflecting the high error in the RANS predictions of the anisotropy. For the RF predictions,  $E_{x_B} = 0.23$  and  $E_{y_B} = 0.31$ . While these error levels are still relatively high, they represent a decrease in error relative to the nominal RANS performance, particularly in the prediction of  $y_B$ .

The anisotropy invariant  $\Pi_a$  can also be calculated from the barycentric coordinates, using the relations (6)-(8), along with the relation  $\Pi_a = 2(\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)$ . Figure 8(c) shows the RF predictions of  $\Pi_a$ . While the RF was trained to optimize its predictions of  $(x_B, y_B)$ , Fig. 8(c) shows that its improved predictions of the barycentric coordinates have also translated to improved predictions of  $\Pi_a$ . Unlike the default RANS model, the RF is able to correctly predict elevated levels of  $\Pi_a$  in the injection hole, along the wall, and in the shear layer on the lee side of the jet. These results demonstrate that even though the RF model was trained on two very different flows—flow over a cube and duct flow—it is able to make reasonable predictions of the Reynolds stress anisotropy on the jet-in-crossflow configuration that significantly surpass the default RANS model predictions in accuracy.

To further explore the generalization properties of this ma-

chine learned model, it was tested on two other canonical flow configurations for which high-fidelity data were available. The first was flow over a wavy wall at  $Re = 6850$ , for which DNS data have been presented by Rossi et al. [?, ?]. The second was flow around a square cylinder at  $Re = 21,400$ , for which a highly resolved LES was presented in References [?, 8]. A detailed discussion of these computations is beyond the scope of this paper, but they have been well documented in the referenced papers. The RF model was able to reduce  $E_{x_B}$  by 41% and  $E_{y_B}$  by 80% for the wavy wall case, and by 9% and 47% respectively for the square cylinder case. While it would be desirable to test the RF model across a much wider database of flows, these results across three challenging and disparate flow configurations suggest that the RF model could provide improved predictions across a broad class of flows.

However, this RF model is not expected to be a universal model across all turbulent flows. The machine learned model is only expected to be valid in flows dynamically similar to those on which it was trained: i.e. incompressible and non-reacting. In order to develop a more universal model, the algorithm would almost certainly need to be trained on a much broader class of flows. Nevertheless, these initial results are very encouraging, as they demonstrate consistently improved predictions, not only for the jet-in-crossflow case of interest in this study, but also for two other canonical flow cases.

It should also be noted that the RF model would not add significant computational cost to the RANS calculation if it were built into the solver. On a single CPU, the RF model requires 5 seconds to make Reynolds stress anisotropy predictions for 1 million points. Because the RF prediction process is completely parallelizable, including it in the RANS solver would add less than 1% to the RANS compute time per iteration. Overall, the Ruiz et al. [20] LES results required approximately  $2 \times 10^6$  processor hours to run the 20 flow through times required for statistical convergence. The RANS simulation used in this study required approximately  $2 \times 10^4$  processor hours to reach convergence, representing a decrease in computational cost of 2 orders of magnitude. The Reynolds stress anisotropy predictions provided by the RF model have the potential to improve the accuracy of the RANS predictions without sacrificing this significant computational costs savings.

## 5 Conclusions

The high-fidelity LES of Ruiz et al. [20] was rigorously analyzed to determine regions of the flow where the RANS Boussinesq hypothesis would be invalid. This flow configuration has relevancy to film cooling, dilution cooling, and fuel injection flows in the context of turbomachinery. Two different underlying assumptions were investigated: the non-negativity assumption for the eddy viscosity and the ability of the mean strain rate to capture the Reynolds stress anisotropy. It was shown

that anisotropy in the normal stresses caused the eddy viscosity to go negative in the leeward shear layer of the jet. It was also shown that the RANS models significantly under-predicted the Reynolds stress anisotropy in that shear layer, in the injection hole, and along the wall. This model form error is a root cause of the inaccuracy widely reported in RANS calculations of jet-in-crossflow and film cooling configurations, and cannot be eliminated through model calibration alone. More detailed model closures, which can represent the correct Reynolds stress anisotropy, are required to mitigate this source of uncertainty.

Machine learning models were investigated for their ability to more accurately capture this Reynolds stress anisotropy. Random Forest regressors were trained to predict the barycentric coordinates ( $x_B, y_B$ ). These regressors were trained on two different flows: a duct flow and flow around a wall-mounted cube. They were then tested on the jet-in-crossflow configuration. These models showed a remarkable ability to generalize across flows, and provided significantly improved Reynolds stress anisotropy predictions as compared to the default RANS predictions. These machine learning models therefore demonstrate the potential for giving improved anisotropy predictions. However, these results represent only the first step in improving RANS predictions for this flow. Future work will be aimed at integrating these machine learning models into the forward RANS simulation, instead of applying them *a posteriori* as was done here. It remains to be seen if they will pose challenges in simulation convergence, and to what extent they can improve predictions of quantities of interest such as wall stresses and heat fluxes. In order to further explore the use of data-driven methods for turbulence modeling, it would be useful to have an open-access database of high fidelity flow solutions available to all turbulence modeling researchers. This database would enable faster model development, broader generalization studies, and improved reproducibility.

## ACKNOWLEDGMENT

Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND

## REFERENCES

[1] Hoda, A., and Acharya, S., 1999. "Predictions of a film coolant jet in crossflow with different turbulence models". *ASME 1999 International Gas Turbine and Aeroengine Congress and Exhibition*.

[2] He, G., Guo, Y., and Hsu, A., 1999. "The effect of Schmidt number on turbulent scalar mixing in a jet-in-crossflow". *International Journal of Heat and Mass Transfer*, **42**, pp. 3727–3738.

[3] Muppidi, S., and Mahesh, K., 2007. "Direct numerical simulation of round turbulent jets in crossflow". *Journal of Fluid Mechanics*, **574**, pp. 59–84.

[4] Coletti, F., Benson, M., Ling, J., Elkins, C., and Eaton, J., 2013. "Turbulent transport in an inclined jet in crossflow". *International Journal of Heat and Fluid Flow*, **43**, pp. 149–160.

[5] Harrison, K., and Bogard, D., 2008. "Comparison of RANS turbulence models for prediction of film cooling performance". *ASME Turbo Expo 2008*, **4**, pp. 1187–1196.

[6] Ling, J., Yapa, S., Benson, M., Elkins, C., and Eaton, J., 2013. "3D velocity and scalar field measurements of an airfoil trailing edge with slot film cooling: The effect of an internal structure in the slot". *Journal of Turbomachinery*, **135**.

[7] Ling, J., Ryan, K., Bodart, J., and Eaton, J. "Analysis of turbulent scalar flux models for a discrete hole film cooling flow". *J. Turbomach*.

[8] Ray, J., Lefantzi, S., Arunajatesan, S., and Dechant, L., 2014. "Bayesian calibration of a  $k - \epsilon$  turbulence model for predictive jet-in-crossflow simulations".

[9] Ray, J., Lefantzi, S., Arunajatesan, S., and Dechant, L., 2015. "Bayesian calibration of a rans model with a complex response surface - a case study with jet-in-crossflow configuration". *45th AIAA Fluid Dynamics Conference*.

[10] Ling, J., Coletti, F., Yapa, S., and Eaton, J., 2013. "Experimentally informed optimization of turbulent diffusivity for a discrete hole film cooling geometry". *International Journal of Heat and Fluid Flow*, **44**, pp. 348–357.

[11] Kaszeta, R., and Simon, T., 2000. "Measurement of eddy diffusivity of momentum in film cooling flows with streamwise injection". *Journal of Turbomachinery*, **122**, January, pp. 178–183.

[12] Rajabi-Zargarabadi, M., and Bazdidi-Tehrani, F., 2010. "Implicit algebraic model for predicting turbulent heat flux". *International Journal for Numerical Methods in Fluids*, **64**, pp. 517–531.

[13] Azzi, A., and Lakehal, D., 2002. "Perspectives in modeling film cooling of turbine blades by transcending conventional two-equation turbulence models". *J. Turbomach.*, **124**.

[14] Bose, I., and Mahapatra, R., 2001. "Business data mining? a machine learning perspective". *Information and management*, **39**, pp. 211–225.

[15] Witten, I., and Frank, E., 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Inc, San Francisco.

[16] Bishop, C., 2006. *Pattern Recognition and Machine Learning*. Springer, New York, NY.

[17] Tracey, B., Duraisamy, K., and Alonso, J., 2013. "Application of supervised learning to quantify uncertainties in tur-

bulence and combustion modeling”. *AIAA Aerospace Sciences Meeting*.

[18] Duraisamy, K., Shang, Z., and Singh, A., 2015. “New approaches in turbulence and transition modeling using data-driven techniques”. *AIAA SciTech*.

[19] Ling, J., and Templeton, J., 2015. “Evaluation of machine learning algorithms for prediction of regions of high RANS uncertainty”. *Physics of Fluids*, **27**.

[20] Ruiz, A., Lacaze, G., and Oefelein, J., 2015. “Flow topologies and turbulence scales in a jet-in-cross-flow”. *Physics of Fluids*, **27**, p. 045101.

[21] Su, L., and Mungal, M., 2004. “Simultaneous measurements of scalar and velocity field evolution in turbulent crossflowing jets”. *Journal of Fluid Mechanics*, **513**, pp. 1–45.

[22] Oefelein, J., 2006. “Large eddy simulation of turbulent combustion processes in propulsion and power systems”. *Prog. Aerosp. Sci.*, **42**, pp. 2–37.

[23] Pichler, R., Sandberg, R., Michelassi, V., and Bhaskaran, R., 2015. “Investigation of the accuracy of RANS models to predict the flow through a low-pressure turbine”. *ASME Turbo Expo 2016*.

[24] Mahesh, K., 2013. “The interaction of jets with crossflow”. *Annual Review of Fluid Mechanics*, **45**, pp. 379–407.

[25] Banerjee, S., Ertunc, O., and Durst, F., 2008. “Anisotropy properties of turbulence”. *Proceedings of the 13th WSEAS International Conference on Applied Mathematics*.

[26] Pope, S., 2000. *Turbulent Flows*. Cambridge University Press.

[27] Radenkovic, D., Burazer, J., and Novkovic, D., 2014. “Anisotropy analysis of turbulent swirl flow”. *FME Transactions*, **42**, pp. 19–25.

[28] Banfield, R., Hall, L., Bowyer, K., Bhadaria, D., Kegelmeyer, W., and Eschrich, S., 2004. “A comparison of ensemble creation techniques”. *Multiple Classifier Systems*, pp. 223–232.

[29] Breiman, L., 2001. “Random forests”. *Machine Learning*, **45**, pp. 5–32.

[30] Rossi, R., Philips, D., and Iaccarino, G., 2010. “A numerical study of scalar dispersion downstream of a wall-mounted cube using direct simulations and algebraic flux models”. *International Journal of Heat and Fluid Flow*, **31**, May, pp. 805–819.

[31] Rossi, R., and Iaccarino, G., 2013. “Numerical analysis and modeling of a plume meandering in passive scalar dispersion downstream of a wall-mounted cube”. *International Journal of Heat and Fluid Flow*, **43**, October, pp. 137–148.

[32] Pinelli, A., Uhlmann, M., Sekimoto, A., and Kawahara, G., 2010. “Reynolds number dependence of mean flow structure in square duct turbulence”. *Journal of Fluid Mechanics*, **644**, pp. 107–122.

[33] Ling, J., Jones, R., and Templeton, J. “Machine learning strategies for systems with invariance properties”. *Journal of Computational Physics*, *submitted*.