Estimating performance bounds of machine-learning Reynolds-stress models via optimal tensor basis expansions

By A. J. Banko AND J. K. Eaton

1. Motivation and objectives

Engineering design relies on numerical simulations of turbulent flows to perform analysis and optimize components. Realistic turn-around times for the ensembles of simulations needed for design iterations or uncertainty quantification require the use of the Reynolds-Averaged Navier-Stokes (RANS) equations for the mean flow with simplified models to represent the effects of turbulence. RANS-type equations are also the basis of the near-wall treatment in wall-modeled large-eddy simulation (LES), which will likely see increased application due to recent advancements in mesh generation and computer architectures (Lozano-Duran *et al.* 2020). A major limitation of all RANS models is an inability to reliably predict three-dimensional (3D) separated flows (Slotnick *et al.* 2014). This is because the models were historically tuned with limited datasets to predict statistically one- and two-dimensional flows including homogeneous shear flow, flat plate boundary layers, backward facing step flows, and curved channel flows (Craft *et al.* 1996).

The availability of highly accurate direct numerical simulation and wall-resolved LES of increasingly realistic geometries with 3D flow separation has been leveraged with modern machine learning algorithms to produce a new generation of data-driven RANS models that have already demonstrated improvements over conventional models (Duraisamy *et al.* 2019). The machine learning component supplants a term or terms in the governing equations by using these large datasets to learn the complex functional relationships between the local mean flow quantities and the modeled term.

Most approaches model the Reynolds stress anisotropy tensor defined as

$$\mathbf{b} = \frac{\overline{\mathbf{u}' \otimes \mathbf{u}'}}{2k} - \frac{1}{3}\mathbf{I}, \tag{1.1}$$

where $\overline{\mathbf{u}' \otimes \mathbf{u}'}$ is the Reynolds stress tensor, k is the turbulent kinetic energy (TKE), and $\mathbf{I}$ is the identity tensor. This is because the linear-eddy viscosity hypothesis used by industry standard RANS models is inaccurate in non-equilibrium flows with large streamline curvature and strong pressure gradients. The anisotropy tensor is approximated using a tensor basis expansion in terms of the mean strain and rotation rate tensors, similar to previous non-linear algebraic stress models (Pope 1975). In contrast to previous models, the machine learning approaches then construct the coefficients as complex functions of the strain and rotation rate tensor invariants (and possibly other mean flow scalars) using neural networks, gene expression programming, or random forests (Ling *et al.* 2016; Weatheritt & Sandberg 2017; Taghizadeh *et al.* 2020; Kaandorp & Dwight 2020). The integration of the machine learning architecture within the tensor basis expansion and the use of scalar invariant input quantities are critical to obtain coordinate-independent

models. The machine learning algebraic stress models are then coupled to the auxiliary transport equations for the TKE and dissipation rate (or other turbulence length and timescale equations) in order to predict the mean flow field. These models have demonstrated improved predictions of the flow over a wavy wall, turbulence-generated secondary flows in a square duct, and separated flow in an asymmetric diffuser, among others.

Despite initial success, data-driven RANS models are still in their infancy and significant challenges remain. The reader is referred to Kutz (2017) for a discussion of several open research topics. This work is motivated by a central difficulty of all machine learning Reynolds stress anisotropy models. Specifically, it is often unclear how to improve a trained model if the anisotropy or mean flow predictions are not accurate enough because model form error is difficult to separate from machine learning regression error. In other words, is the choice of tensor basis expansion (e.g., model form) a limiting factor, or does the choice of input quantities and algorithm hyperparameters used for regression need to be revised?

This work proposes the concept of optimal tensor basis expansions as a deterministic method to separate errors in the chosen tensor basis from the input feature selection and machine learning regression steps. The optimal expansion represents the performance bound of any machine learning model using a specific tensor basis expansion, and it determines the coefficients that the model must predict. We further propose applying these optimal approximations in simulations aided by LES data in order to isolate the effect of the anisotropy on the mean flow field. In this way, a tensor basis expansion can be deemed sufficient or insufficient prior to the machine learning step.

This brief is organized as follows. Section 2 describes the 3D bump flow used as a test case. Section 3 details the optimal tensor basis expansion and compares the optimal expansion to the prediction of a tensor basis neural network. Section 4 presents a methodology to simulate the flow field corresponding to the optimal tensor basis approximation of the anisotropy tensor, and to isolate model form errors associated with the Reynolds stress anisotropy and auxiliary turbulence equations. Results are discussed for a three-term optimal tensor basis expansion. Conclusions are given in Section 5.

2. Three-dimensional bump flow

Results are demonstrated for the 3D bump flow described in Ching & Eaton (2020), and future work will explore a wider variety of flows including a family of jet-in-crossflows. The bump geometry is illustrated in Figure 1(a) and has an elliptical footprint when viewed from above with a 4/3 major to minor axis ratio, and cosine cross-sections when view from the side. The bump is mounted in a rectangular channel and the minor axis is skewed by 10 degrees with respect to the streamwise direction. The flow is incompressible and the Reynolds number based on the bulk flow velocity and bump height is 16,000.

Wall-resolved LES data were obtained using the code Vida from Cascade Technologies. Grid-converged solutions were verified through a grid refinement study. Complete details of the calculation and validation against magnetic resonance imaging experiments are given in Ching & Eaton (2020).

RANS simulations using the realizable k-ϵ (RKE) model were conducted using the Fluent commercial software. The near-wall grid cells were refined to have y^+ values less than unity, and enhanced wall-treatment was used. The steady-state, incompressible flow was obtained using an iterative scheme with second-order discretization of the spatial deriva-

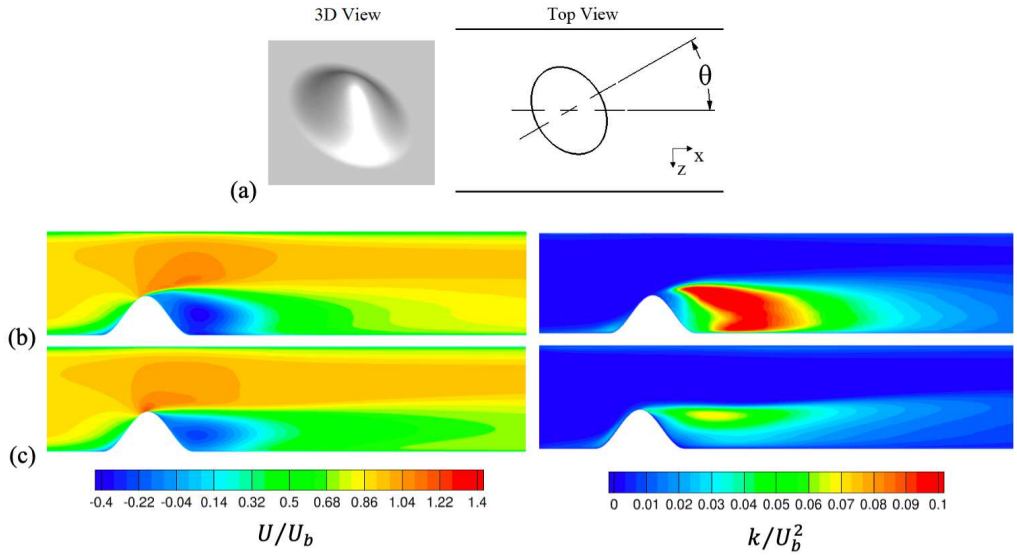

FIGURE 1. Overview of the 10 degree bump flow showing (a) isometric and top-down views of the geometry, (b) contours of LES mean streamwise velocity and turbulent kinetic energy on the bump centerplane normalized by the bulk velocity, and (c) contours of streamwise velocity and turbulent kinetic energy predicted by the RKE model on the bump centerplane.

tives, upwinding for the advective term, and the SIMPLEC algorithm for determination of the pressure.

Figures 1(b,c) compare results for the streamwise velocity and TKE on the centerplane of the bump. The LES exhibits a large separation bubble that reattaches approximately one bump height downstream. The RKE model underpredicts the size and strength of reversed flow in the separation bubble, and overpredicts the momentum deficit and recovery length of the wake downstream of reattachment. The TKE in the wake of the bump is also substantially underpredicted. Therefore, the 3D bump flow provides a challenging test case for RANS models.

3. Optimal tensor basis expansions

3.1. *Tensor basis expansions*

Pope (1975) introduced the tensor basis expansion for representing algebraic Reynolds stress anisotropy models in terms of the mean strain and rotation rate tensors. This framework automatically enforces the correct symmetric tensor structure and transformation properties, and has been widely adopted in both the classical and machine-learning Reynolds stress modeling literature. Note that additional tensors can also be included, such as objective time derivatives of the strain rate tensor in order to include flow history effects (Speziale 1987). The present work focuses exclusively on the mean strain and rotation rate tensors, but the methods developed are applicable to more general expansions.

The N-term tensor basis expansion for the anisotropy tensor is

$$\mathbf{b}^{(N)} = \sum_{i=1}^{N} g_i(\lambda_k)\mathbf{T}_i(\mathbf{S}^*, \mathbf{R}^*). \tag{3.1}$$

Here, $\mathbf{b}^{(N)}$ is N-term approximation of the anisotropy tensor, $\mathbf{T}_i$ are the basis tensors, and g_i are the scalar coefficients. $\mathbf{S}^* = (k/\epsilon)\mathbf{S}$ and $\mathbf{R}^* = (k/\epsilon)\mathbf{R}$ are the strain and rotation rate tensors made non-dimensional using the local TKE and dissipation rate. The λ_k are tensor invariants of $\mathbf{S}^*$ and $\mathbf{R}^*$. One could also include additional Galilean-invariant scalars formed from other mean flow and turbulence quantities (Kaandorp & Dwight 2020).

For statistically 3D flows, there are ten independent basis tensors that can be formed from $\mathbf{S}^*$ and $\mathbf{R}^*$, and five independent tensor invariants. The tensor invariants are

$$\lambda_1 = \mathrm{tr}\left(\mathbf{S}^{*2}\right), \quad \lambda_2 = \mathrm{tr}\left(\mathbf{R}^{*2}\right), \quad \lambda_3 = \mathrm{tr}\left(\mathbf{S}^{*3}\right),$$
$$\lambda_4 = \mathrm{tr}\left(\mathbf{R}^{*2}\mathbf{S}^*\right), \quad \text{and} \quad \lambda_5 = \mathrm{tr}\left(\mathbf{R}^{*2}\mathbf{S}^{*2}\right). \tag{3.2}$$

The reader is referred to Pope (1975) for a complete list of the basis elements. Therefore, $N \leq 10$ where $N = 10$ would make Eq. (3.1) exact if the true anisotropy depends only on $\mathbf{S}^*$ and $\mathbf{R}^*$, and $N < 10$ is in general an approximation.

Note that linear eddy viscosity models fall under the category of a one-term tensor basis expansion. Namely,

$$\mathbf{b}^{LEVM} = -\frac{\nu_T \epsilon}{k^2}\mathbf{S}^* \tag{3.3}$$

for the RKE model where ν_T is the eddy viscosity.

3.2. *Tensor basis neural network*

The tensor basis neural network (TBNN) was introduced by Ling *et al.* (2016) as a data-driven approach to determine algebraic anisotropy models that automatically embed the correct tensor structure and transformation properties. A neural network with densely connected layers is used to construct the functions $g_i(\lambda_k)$. Specifically, the input layer is the set of invariants, λ_k, followed by a specified number of hidden layers and terminating in the output layer whose nodes represent the g_i. This layer then multiplies the tensor basis elements to form $\mathbf{b}^{(N),TBNN}$.

The network weights and biases are found by minimizing a loss function over the entire training data (i.e., the entire flow field) simultaneously. It is common to consider the L2 loss function quantifying the error between the predicted anisotropy tensor and the LES data. Specifically, the network parameters $\tilde{p}$ are determined by

$$\tilde{p} = \mathrm{argmin}_p \sum_{m=1}^{M} \left|\left| \mathbf{b}_m^{LES} - \sum_{j=1}^{N} g_j(\lambda_k, p)\mathbf{T}_j \right|\right|_2^2. \tag{3.4}$$

Here, $\mathbf{b}^{LES}$ is the LES anisotropy that the TBNN is designed to approximate, $||\cdot||_2$ denotes the L2 norm, and summation is over all points in the training data where M is the total number of points. Then, the TBNN prediction of the anisotropy tensor is

$$\mathbf{b}^{(N),TBNN} = \sum_{j=1}^{N} g_j(\lambda_k, \tilde{p})\mathbf{T}_j. \tag{3.5}$$

3.3. *Optimal tensor basis*

The optimal tensor basis expansion determines the coefficients that produce the best fit to the LES anisotropy tensor at each point in the flow (Gatski & Jongen 2000). Specifically,

$$\{\tilde{g}_i\} = \text{argmin}_{\{g_i\}} \left|\left| \mathbf{b}^{LES} - \sum_{j=1}^{N} g_j \mathbf{T}_j \right|\right|_2^2. \tag{3.6}$$

Then, the optimal tensor basis representation of the anisotropy tensor is

$$\mathbf{b}^{(N),opt} = \sum_{j=1}^{N} \tilde{g}_j \mathbf{T}_j. \tag{3.7}$$

Since the anisotropy tensor is symmetric and traceless, and therefore has five independent components, any tensor basis expansion with $N \geq 5$ can exactly represent b^{LES} except in certain degenerate cases (Gatski & Jongen 2000). For $N \leq 4$, the optimal tensor basis expansion is an approximation for statistically 3D flows.

The TBNN and the optimal expansion are different in two important ways. First, Eq. (3.4) defines a minimization over the entire flow field simultaneously, whereas Eq. (3.6) minimizes the argument at each point individually. Second, Eq. (3.4) determines g_i as functions of the specified λ_k, whereas Eq. (3.6) does not require the $\tilde{g}_i$ to be learnable functions.

Equation (3.6) is an optimal approximation in the sense that

$$\left|\left|\mathbf{b}^{LES} - \mathbf{b}^{(N),opt}\right|\right|_2 \leq \left|\left|\mathbf{b}^{LES} - \mathbf{b}^{(N)}\right|\right|_2, \tag{3.8}$$

where $\mathbf{b}^{(N)}$ is any other expansion in terms of the same N tensor basis elements. Therefore, the optimal expansion establishes the best possible performance of the TBNN (or any similar machine learning model), and it determines the coefficients that the TBNN must predict if it were to achieve the lowest possible loss.

Equation (3.6) can be solved analytically. The optimal one-term expansion is simply the least-squares solution for a linear eddy-viscosity model. Specifically,

$$\mathbf{b}^{(1),opt} = \tilde{g}_1 \mathbf{S}^* \;\text{ where }\; \tilde{g}_1 = \frac{\text{tr}\left(\mathbf{b}^{LES}\mathbf{S}^*\right)}{\lambda_1}. \tag{3.9}$$

Gatski & Jongen (2000) also provide a solution for the three- and five-term expansions. Because of space considerations, we consider only the three-term case. The expansion is

$$\mathbf{b}^{(3),opt} = \tilde{g}_1 \mathbf{S}^* + \tilde{g}_2 \left(\mathbf{S}^*\mathbf{R}^* - \mathbf{R}^*\mathbf{S}^*\right) + \tilde{g}_3 \left(\mathbf{S}^{*2} - \frac{1}{3}\text{tr}\left(\mathbf{S}^{*2}\right)\mathbf{I}\right), \tag{3.10}$$

with coefficients

$$\tilde{g}_1 = \frac{\lambda_1^2 \text{tr}\left(\mathbf{b}^{LES}\mathbf{S}^*\right) - 6\lambda_3 \text{tr}\left(\mathbf{b}^{LES}\mathbf{S}^{*2}\right)}{\lambda_1^3 - 6\lambda_3^2},\; \tilde{g}_2 = \frac{-2\text{tr}\left(\mathbf{b}^{LES}\mathbf{R}^*\mathbf{S}^*\right)}{\lambda_1\lambda_2 - 6\lambda_5^2},\; \text{and}$$
$$\tilde{g}_3 = 6\frac{\lambda_1 \text{tr}\left(\mathbf{b}^{LES}\mathbf{S}^{*2}\right) - \lambda_3 \text{tr}\left(\mathbf{b}^{LES}\mathbf{S}^*\right)}{\lambda_1^3 - 6\lambda_3^2}. \tag{3.11}$$

3.4. *A priori results*

The optimal coefficients were computed using the three-term tensor basis expansion for the bump flow. The corresponding anisotropy tensor is compared to the LES data, the

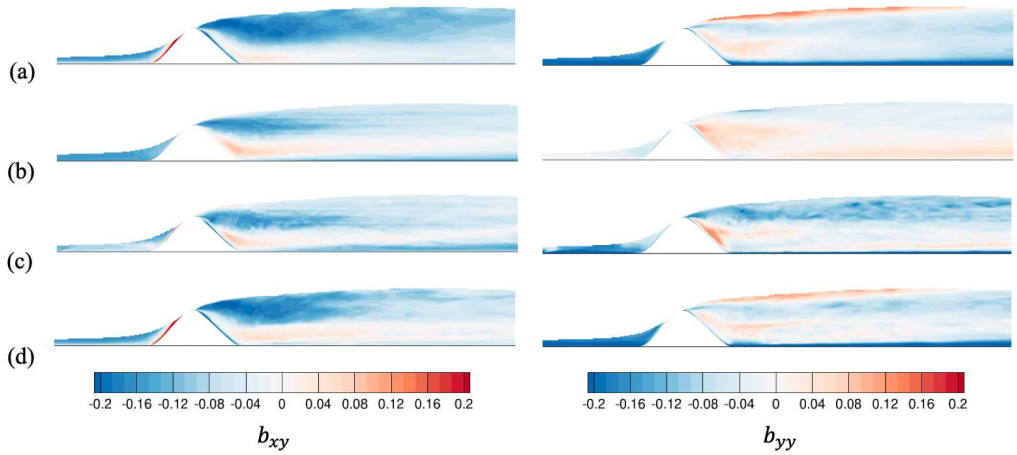

FIGURE 2. Shear stress and wall-normal (vertical) anisotropy components for (a) the LES data, (b) the RKE model, (c) the three-term TBNN, and (d) the three-term optimal tensor basis expansion.

RKE model, and a three-term TBNN. A network with five layers and ten nodes per layer was trained using the Lasagne Python package with rectified linear unit activation functions, the Adam optimizer, and a learning rate of 10^{-5}. The five input features listed in Eq. (3.2) were used, with the mean velocity gradient tensor and TKE taken from the LES data. The dissipation rate was obtained by solving its RANS transport equation with LES data for the mean velocity and TKE (Weatheritt & Sandberg 2017). This method of training produces stand-alone turbulence models that do not depend on a baseline RANS solution, which can be inaccurate in separated flows.

Figure 2 presents comparisons of the shear stress and wall-normal components of the anisotropy tensor. The shear stress component is predicted reasonably well by the RKE and TBNN. This result is partly expected because conventional RANS models were tuned primarily to capture this component. The TBNN does not significantly degrade or improve the result. The optimal expansion is in closest agreement with the LES data. In fact, an interesting consequence of the optimal expansion is that the turbulent kinetic energy production term, which is dominated by this stress component in the shear layer, is exactly captured. This result can be verified by direct computation using the coefficients in Eq. (3.11).

Greater differences between models are observed in the wall-normal stress component of the anisotropy. The RKE model entirely misses the near-wall damping of the wall-normal fluctuations. The TBNN captures the damping very close to the wall. However, the anisotropy further from the wall in the wake of the bump is only moderately improved. The optimal expansion accurately captures this component of anisotropy throughout the flow field.

Figure 3 plots contours of the L2 error for the RKE, TBNN, and optimal basis with respect to LES data. The conclusions accounting for all of the tensor components are similar to those described above. The RKE model shows high error throughout the flow field and especially near the wall due to the misprediction of the normal components of the Reynolds stress. The TBNN improves the near-wall points but does not improve the L2 error elsewhere to any significant degree. The optimal expansion shows substantially

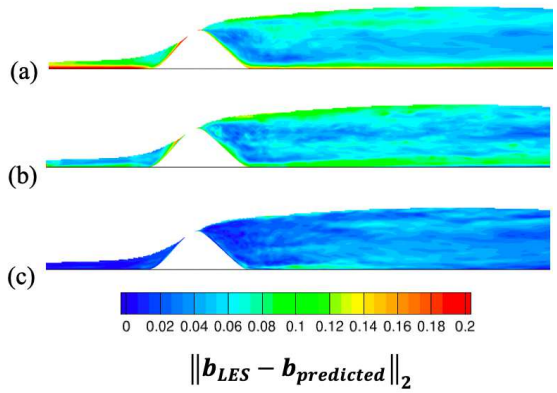

FIGURE 3. Error in the predicted anisotropy tensor with respect to LES data, measured using the 2-norm, for (a) the RKE model, (b) the three-term TBNN, and (c) the three-term optimal tensor basis expansion.

lower error than the TBNN throughout the entire flow. Some residual error remains near the reattachment point downstream of the bump.

These results show that the three-term tensor basis expansion can accurately approximate the anisotropy tensor in a 3D flow, despite the fact that it has fewer degrees of freedom than are required to exactly determine all five independent components of the tensor. Therefore, performance of the TBNN is likely not limited by this choice of reduced basis. Instead, the number of input quantities may not be sufficient to predict the optimal coefficients.

The analytical solution of optimal expansion provides insight on how to improve the TBNN. Equation (3.11) shows that any model based on a three-term expansion must include λ_1, λ_2, λ_3, and λ_5 as inputs. Additionally, the scalar invariants formed by products of $\mathbf{b}^{LES}$, $\mathbf{S}^*$ and $\mathbf{R}^*$ must be predicted by a finite set of input quantities. In future work, these scalars will be analyzed using feature selection techniques to determine an improved set of input quantities. Applying these techniques to the scalar invariants instead of the anisotropy tensor components is preferred because the tensor components are coordinate-dependent, while conclusions drawn from the scalar invariants will be coordinate-independent.

Analyzing these scalar quantities instead of tensor components is preferred for developing improved models because the results will be coordinate-independent.

4. Simulation methodology using LES data

4.1. *Overview*

The results described in the previous section suggest that three tensor basis terms may be sufficient for modeling the bump flow despite its complex 3D flow topology. However, this conclusion is based on the error in the anisotropy tensor components and does not guarantee that it would lead to an accurate mean flow field. This is because the L2 error does not enforce that the divergence of the Reynolds stresses is accurately represented.

In order to evaluate the impact of the optimal basis anisotropy on the mean flow field, we use LES data to reconstruct the corresponding Reynolds stresses and apply these stresses in the RANS momentum equations. The RANS equations contain auxiliary

transport equations for the TKE and dissipation rate. The velocity field may also be affected by model form errors in the auxiliary equation. Therefore, we perform two sets of simulations in order to isolate the errors associated with the anisotropy tensor and the auxiliary equations.

The first simulation uses the LES TKE and the optimal tensor basis expansion to construct the Reynolds stresses. By assuming the correct TKE distribution, we remove compounding errors in the auxiliary equations. Specifically, the Reynolds stress tensor is

$$\overline{\mathbf{u}' \otimes \mathbf{u}'} = 2k^{LES}\left(\mathbf{b}^{(N),opt} + \frac{1}{3}\mathbf{I}\right). \tag{4.1}$$

The second simulation uses the TKE predicted by the auxiliary equations and the LES anisotropy tensor. The LES anisotropy is also included in the TKE production term for consistency. By assuming the correct anisotropy distribution, we remove compounding errors in the anisotropy model. Specifically, the Reynolds stress tensor is

$$\overline{\mathbf{u}' \otimes \mathbf{u}'} = 2k\left(\mathbf{b}^{LES} + \frac{1}{3}\mathbf{I}\right), \tag{4.2}$$

and the TKE production term is

$$P_k = -2k\mathrm{tr}\left(\mathbf{b}^{LES}\mathbf{S}\right). \tag{4.3}$$

This decomposition could be taken further in future work to combine the effects of an anisotropy model or optimal tensor basis representation with the auxiliary transport equations.

4.2. *Simulation stability and implementation*

Directly inserting the Reynolds stress tensor in the RANS momentum equations is unstable for two reasons. First, the system of equations is ill-conditioned when the Reynolds stress is treated explicitly (Wu *et al.* 2019). In order to improve simulation stability, we adopt the semi-implicit method described by Wu *et al.* (2019), where the Reynolds stress is split into a linear-eddy viscosity component that is treated implicitly and a remainder that is treated explicitly. Rather than projecting the entire Reynolds stress tensor onto the linear strain term, we simply decompose the stress into the component predicted by the baseline RANS model and a deviation via

$$\overline{\mathbf{u}' \otimes \mathbf{u}'} = -2\nu_T\mathbf{S} + \overline{\mathbf{u}' \otimes \mathbf{u}'}^{\Delta}. \tag{4.4}$$

The simulations are also unstable due statistical convergence of the LES data used for the TKE, anisotropy, or optimal tensor basis. Since the momentum equations depend on the divergence of the Reynolds stress, statistical noise is amplified and prevents the system of equations from converging. A filtering procedure is used to smooth the data on the scale of the grid. The Reynolds stress is filtered using an elliptic differential filter given by

$$\widetilde{\overline{\mathbf{u}' \otimes \mathbf{u}'}} - \nabla \cdot \left(\sigma^2 \nabla \widetilde{\overline{\mathbf{u}' \otimes \mathbf{u}'}}\right) = \overline{\mathbf{u}' \otimes \mathbf{u}'}, \tag{4.5}$$

where $\widetilde{\overline{\mathbf{u}' \otimes \mathbf{u}'}}$ is the filtered stress. The filter length scale is defined to be $\sigma = CV^{1/3}$, where V is the local grid cell volume and C is a user-defined constant. This work uses $C = 4$, but ongoing research suggests that C can be reduced further.

Equations (4.1)–(4.5) were implemented in Fluent via user-defined functions, user-defined sources, and user-defined scalar equations on top of the RKE turbulence model.

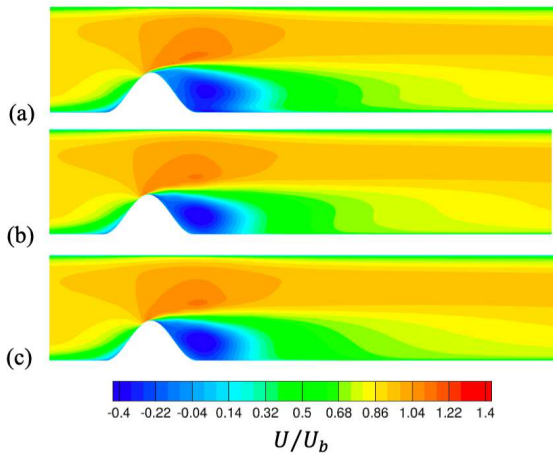

FIGURE 4. Contours of mean streamwise velocity on the bump centerplane for (a) the LES data, and the RANS momentum equations solved using filtered Reynolds stresses formed from the LES TKE and the anisotropy tensor for (b) the LES data, and (c) the three-term optimal tensor basis expansion.

The boundary conditions on Eq. (4.5) are $\widetilde{\overline{\mathbf{u}' \otimes \mathbf{u}'}} = 0$ at the walls and zero-gradient at the inlet and outlet. Figures 4(a,b) show contours of streamwise velocity on the bump centerplane for the LES and the simulated velocity field using the filtered LES stresses. The filtering operation is observed to have a negligible effect on the overall flow field and facilitates a stable solution.

4.3. *Isolation of the anisotropy model*

The three-term optimal tensor basis expansion was simulated using the LES TKE in order to assess the effect of the approximated anisotropy tensor on the mean flow field. Figure 4 compares the optimal tensor basis result to the LES data. The separation bubble size and wake recovery are significantly improved over the RKE model (c.f. Figure 1). Differences are observed in the detailed isocontour shapes. In particular, the LES contains a vertical layer of near-constant streamwise velocity in the wake while the optimal basis velocity gradually increases with increasing height. Nevertheless, the three-term expansion results in a substantially improved flow field in addition to improving the anisotropy components.

4.4. *Isolation of the auxiliary equations*

A three-term expansion appears sufficient, provided that the TKE is accurately predicted with respect to LES data. In order to understand the relative importance of the anisotropy model compared with the auxiliary equations, the mean flow field, TKE, and dissipation rate were solved for using the LES anisotropy as described above. Figures 5(a,b) show the contours of streamwise velocity and TKE on the bump centerplane for the LES data and the simulation using the LES anisotropy. Despite the correct distribution of anisotropy, Figure 5(b) shows that the TKE remains underpredicted and that the wake recovery is too slow. Additionally, the separation bubble strength and shape are incorrect. A comparison of Figures 1 and 5 indicates that a perfect anisotropy model does not improve the results over the baseline RKE model. Therefore, the greatest errors appear to reside in the auxiliary transport equations, at least in the context of the bump flow. It is hypothesized that this could be the case in other flows with large-scale separation.

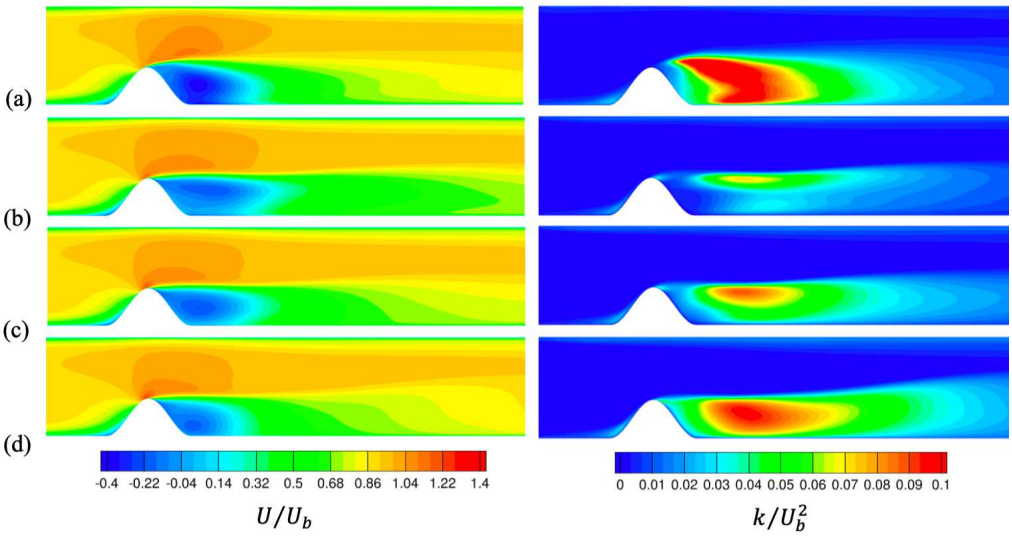

FIGURE 5. Contours of mean streamwise velocity and TKE on the bump centerplane for (a) the LES data, and the RANS momentum and auxiliary equations solved using filtered Reynolds stresses formed from the predicted TKE and the LES anisotropy tensor for (b) $\beta = 1$, (c) $\beta = 0.8$, and (d) $\beta = 0.6$.

Errors in the TKE transport equation were assessed by evaluating the overall balance of the equation using LES data, similar to the method employed by Ching *et al.* (2019). Specifically, LES data were used to evaluate each term in the equation and calculate the residual. The LES-implied dissipation rate was obtained as described in Section 3.4 from the method of Weatheritt & Sandberg (2017). The residual was nonzero, indicating that the LES fields do not represent a fixed point of the equation set. The sign of the residual was consistent with an overly dissipative model in the bump wake downstream of reattachment.

A dissipation rate correction was hypothesized in order to test the sensitivity of the solution to the balance of the TKE transport equation. The dissipation rate term, ϵ, was modified to be $\beta\epsilon$ where β is a constant. This modifies the magnitude of the dissipation rate term in the TKE transport equation. The modification can be justified by the fact that dissipation plays multiple roles in the RANS context: it determines the turbulence time and length scales, and it is responsible for the dissipation of TKE. These roles could be distinct in non-equilibrium flows. Additionally, the RANS transport equation for dissipation rate is entirely heuristic. Therefore, the dissipation rate found directly from the transport equation and used in the eddy viscosity may not be the same dissipation as needed in the TKE equation.

LES data were used to determine that β in the range of 0.6 to 0.8 approximately balances the TKE equation in the wake of the bump. The dissipation rate correction was implemented in Fluent and Figures 5(c,d) show the results of simulations with different β values. By reducing the role of dissipation in the TKE transport equation, the TKE downstream of the bump increases and the wake accelerates in better agreement with the LES. Thus, using the overall balance of the TKE equation may be one method for determining model improvements. However, a balance of terms may be a necessary but not sufficient condition for the solution to converge close to the true flow field.

5. Conclusions

Many data-driven approaches to RANS modeling seek to improve the Reynolds stress anisotropy tensor because the linear-eddy viscosity hypothesis is known to be inaccurate in non-equlibrium turbulent flows and separated flows of practical interest. These models are typically based on a tensor basis expansion, and machine learning regression algorithms are used to determine the coefficients as functions of other mean flow quantities. Despite demonstrated success in a variety of statistically 2D and 3D flows, it remains unclear if further improvement is limited by the choice of tensor basis, input features, or algorithm hyperparameters.

This work introduces optimal tensor basis expansions as a means to estimate the performance bounds of data-driven anisotropy models. The optimal expansion determines the best fit of a given tensor basis to the true anisotropy tensor at each point in the flow field. Any model using the same expansion will have greater error than that of the optimal approximation. A method to simulate the flow field produced by the optimal expansion was also described. The methods presented offer a deterministic framework to assess the utility of different tensor bases prior to any machine learning step. In other words, the optimal tensor basis expansion allows one to separate model form error from machine learning regression error. Additionally, analytical solutions for the optimal coefficients provided further insight into necessary functional dependencies. These included invariants of the mean strain and rotation rate tensors that must be provided to the machine learning algorithm. It also identified scalars derived from products of the true anisotropy and basis tensors that should be the focus of future analysis.

The method was applied to a three-term tensor basis expansion in terms of the strain and rotation rate tensors and tested on a 3D bump flow with large-scale separation. Comparisons of the optimal anisotropy approximation to LES data, a baseline RANS model, and a tensor basis neural network showed that the three-term expansion was sufficient in terms of both the predicted anisotropy components and the resulting mean flow field. The TBNN had larger error than the optimal expansion, and future research will investigate the choice of input features to the neural network in order to reduce this error. Simulations using LES data to isolate errors in the anisotropy model and the auxiliary transport equations showed that far greater errors reside in the TKE transport equation equation compared to the anisotropy model. A simple dissipation rate correction designed to improve the overall balance of the TKE transport equation brought the turbulent kinetic energy and mean velocity fields into better agreement with the LES data. Future research will investigate spatially dependent corrections and their generality across a wider variety of flows.

Finally, the optimal tensor basis methods described here focused on expansions in terms of the mean strain and rotation rate tensors. However, the method is general and other tensors can be considered. The method could also be extended to other transport physics that use tensor basis expansions, such as the turbulent scalar flux in turbulent mixing applications (Milani & Eaton 2020).

Acknowledgments

This investigation was funded by Bosch Corporation and the United States Office of Naval Research under Grant # N000141912075.

REFERENCES

Ching, D. S., Banko, A. J., Milani, P. M., & Eaton, J. K. 2019 Machine learning modeling for RANS turbulent kinetic energy transport in 3D separated flows. *Proceedings of the 11th International Symposium on Turbulence and Shear Flow Phenomena*.

Ching, D. S. & Eaton, J. K. 2020 Large-eddy simulation study of unsteady wake dynamics and geometric sensitivity on a skewed bump. *J. Fluid Mech.* **885**, A22.

Craft, T. J., Launder, B. E., & Suga, K. 1996 Development and application of a cubic eddy-viscosity model of turbulence. *Int. J. Heat Fluid Flow* **17**, 108–115.

Duraisamy, K., Iaccarino, G., & Xiao, H. 2019 Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377.

Gatski, T. B. & Jongen, T. 2000 Non-linear eddy viscosity and algebraic stress models for solving complex turbulent flows. *Prog. Aerosp. Sci.* **36** 655–682.

Kaandorp, M. L. A. & Dwight, R. P. 2020 Data-driven modeling of the Reynolds stress tensor using random forests with invariance. *Comput. Fluids* **202**, 104497.

Kutz, J. N. 2017 Deep learning in fluid dynamics. *J. Fluid Mech.* **814**, 1–4.

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

Lozano-Duran, A., Bose, S. J., & Moin, P. M. 2020 Prediction of trailing edge separation on the NASA juncture flow using wall-modeled LES. *AIAA* 2020-1776.

Milani, P. M. & Eaton, J. K. 2020 Turbulent scalar flux in inclined jets in crossflow: counter gradient transport and deep learning modeling. *arXiv*. 2001.04600.

Pope, S. B. 1975 A more general effective-viscosity hypothesis. *J. Fluid Mech.* **72**, 331–40.

Slotnick, J., Khodadoust, A., Alonso, J., Darmofal, D., Gropp, W., Lurie, E., & Mavriplis, D. 2014 CFD Vision 2030 study: a path to revolutionary computational aerosciences. NASA Technical Report.

Speziale, C. G. 1987 On non-linear $k-l$ and $k-\epsilon$ models of turbulence. *J. Fluid Mech.* **178**, 459–475.

Taghizadeh, S., Witherden, F. D., & Girimaji, S. S. 2020 Turbulence closure modeling with data-driven techniques: physical compatibility and consistency conditions. *arXiv*. 2004.03031.

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

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