
Extending turbulence model uncertainty quantification using machine learning

Marcel Matha and Christian Morsbach

Institute of Propulsion Technology, German Aerospace Center (DLR)
Linder Hhe, 51147 Cologne, Germany
marcel.matha@dlr.de and christian.morsbach@dlr.de

Abstract

In order to achieve a more virtual design and certification process of jet engines in aviation industry, the uncertainty bounds for computational fluid dynamics have to be known. This work shows the application of a machine learning methodology to quantify the epistemic uncertainties of turbulence models. The underlying method in order to estimate the uncertainty bounds is based on an eigenspace perturbation of the Reynolds stress tensor in combination with random forests.

1 Introduction

As a compromise between computational time and accuracy, Reynolds-averaged Navier-Stokes (RANS) simulation is the workhorse in the industrial design process of turbomachinery. The derivation of the RANS equations reveals an unclosed term, called the Reynolds stress tensor. This tensor has to be approximated in Computational-Fluid-Dynamics (CFD) simulations by applying turbulence models. The prediction accuracy of the simulation is highly dependent on this kind of models, which mostly accomplished transport equations, mimicking main turbulence flow physics. Although RANS-based models, such as linear eddy viscosity models, are widely used for complex engineering flows, they suffer from the inability to replicate fundamental turbulent processes. Turbulence models are one of the main limitations in striving for reliable, environmental-friendly designs, due to general simplifying assumptions during formulation of closure models. These simplifications are the result of data observation and physical intuition, leading to a significant degree of epistemic uncertainty.

In recent years, the interest in quantifying these uncertainties, leading to more reliable simulative results, has grown. The group of Iaccarino proposed an eigenspace perturbation framework, which is based on the inability of common linear eddy viscosity models to deal with Reynolds stress tensor anisotropy [4] [6]. The emergence of machine learning strategies guided the path towards data driven approaches also for the turbulence modelling community [3]. Heyse et al. recently enhanced the uncertainty estimation based on the eigenspace perturbation approach by adding a data-driven method [5].

In order to obtain more sophisticated and trustworthy simulative results, we investigate data-driven enhancements of the Reynolds stress perturbation approach with DLR's CFD solver suite TRACE. TRACE is being developed by the Institute of Propulsion Technology with focus on turbomachinery and offers a parallelized, multi-block CFD solver for compressible RANS equations.¹

¹TRACE User Guide, trace-portal.de/userguide

2 Eigenspace perturbation framework

2.1 Data-free approach

RANS turbulence models are utilized in order to determine the Reynolds stress tensor $\tau_{ij} = \overline{u'_i u'_j}$ in terms of mean flow quantities, whereas u' is the fluctuating part of the velocity ($u = \bar{u} + u'$). The overbar indicates the time averaging of the flow quantities. The symmetric Reynolds stress tensor can be expressed by applying an eigenspace decomposition as

$$\tau_{ij} = k \left(a_{ij} + \frac{2}{3} \delta_{ij} \right) = k \left(v_{in} \Lambda_{nl} v_{jl} + \frac{2}{3} \delta_{ij} \right). \quad (1)$$

Equation 1 includes the split into the anisotropy tensor a_{ij} and the isotropic part of τ_{ij} , while $k = \frac{1}{2} \overline{u'_i u'_i}$ being the turbulent kinetic energy. The eigenspace decomposition provides the eigenvector matrix v and the diagonal eigenvalue matrix Λ . Iaccarino and co-workers proposed a strategy to perturb the eigenvalues and eigenvectors in Equation 1, resulting in a perturbed state of the Reynolds stress tensor

$$\tau_{ij}^* = k \left(v_{in}^* \Lambda_{nl}^* v_{jl}^* + \frac{2}{3} \delta_{ij} \right). \quad (2)$$

Every physical, realizable state of the Reynolds stress tensor, meaning to be positive semidefinite, can be mapped onto barycentric coordinates

$$\mathbf{x} = \mathbf{x}_{1c} \frac{1}{2} (\lambda_1 - \lambda_2) + \mathbf{x}_{2c} (\lambda_2 - \lambda_3) + \mathbf{x}_{3c} \frac{1}{2} (3\lambda_3 + 1) \quad \text{with} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3, \quad (3)$$

which is essentially a linear transform according to $\mathbf{x} = \mathbf{B}\lambda$ within the triangle defined by the corners \mathbf{x}_{1c} , \mathbf{x}_{2c} and \mathbf{x}_{3c} (see 1) [1].

The eigenvalue perturbation makes use of the fact, that the Reynolds stress tensor is limited between the one, two and three component corner of the barycentric triangle, where the limiting states reflect the number of non-zero eigenvalues of the tensor. As shown in Figure 1(a), the eigenvalue perturbation is defined as a shift in barycentric coordinates towards each of the limiting states to location \mathbf{x}^* , according to

$$\mathbf{x}^* = \mathbf{x} + \Delta_B (\mathbf{x}_{(t)} - \mathbf{x}). \quad (4)$$

The relative distance Δ_B controls the magnitude of eigenvalue perturbation towards the corner state $\mathbf{x}_{(t)} \in \{\mathbf{x}_{1c}, \mathbf{x}_{2c}, \mathbf{x}_{3c}\}$. The perturbed eigenvalues λ_i^* can be remapped afterwards.

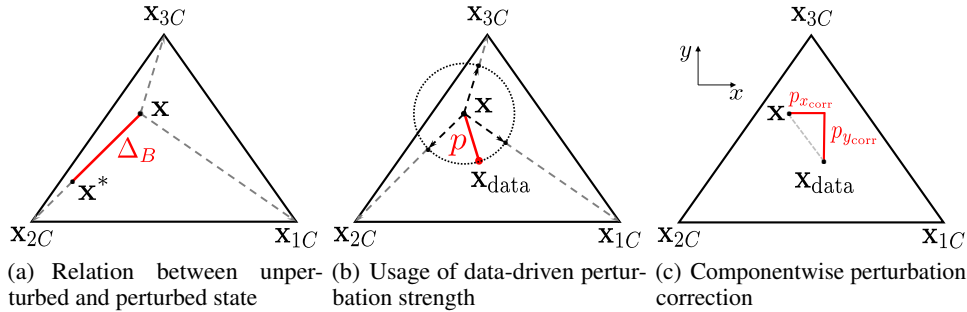


Figure 1: Schematic representation of the eigenvalue perturbation approach

The creation of the perturbed eigenvector matrix Λ^* is purely motivated on the manipulation of the turbulent production term $P_k = -\tau_{ij} \frac{\partial u_i}{\partial x_j}$ and is not discussed further during this extended abstract. Taking only the eigenvalue perturbation into account, a designer needs to run three distinct RANS simulations in order to get the full estimate of the turbulence model uncertainty bounds, since the methodology seeks to sample from the three extremal states in the barycentric map.

2.2 Data-driven approach

The data-free approach is a purely physics-based methodology, aiming for understandable uncertainty bounds for the turbulence modelling community. One of the major drawbacks of the proposed

method is the fact, that a user has to choose the degree of uncertainty by selecting Δ_B before each investigation. The perturbation amplitude, closely related to the degree of uncertainty, has only to become significant in flow regime, which contravene the assumptions during formulation of the turbulence model. Thus, enabling a spatially varying perturbation of the Reynolds stress tensor seems to be worthwhile.

The prediction of the correct anisotropy was investigated in an a priori study using random forests by Wang et al. [11]. Heyse et al. propose a strategy to combine a random forest model with the forward eigenvalue perturbation approach [5]. Physical flow features are extracted to train a random forest in order to predict the local perturbation strength $p = |\mathbf{x}_{\text{Data}} - \mathbf{x}_{\text{RANS}}| = |\mathbf{x}_{\text{RANS}}^* - \mathbf{x}_{\text{RANS}}|$ (see Figure 1(b)). Forward propagating CFD simulations follow training the model, where the predicted perturbation strength is used to modify the Reynolds stress towards the same three limiting states as in the data-free approach.

For the sake of interpretability and usability, an ensemble learning random regression forest is chosen in the presented work as well [2]. The python library *scikit-learn* is used to train the random forest and evaluate its prediction. We will investigate the proposed combination of a machine learning model and the perturbation approach within TRACE. The aim is to go one step further and merge the work of Wang et al. [11] and Heyse et al. [5].

As a first step a machine learning model should predict the componentwise perturbation strength $\mathbf{p}_{\text{corr}} = (p_{x_{\text{corr}}}, p_{y_{\text{corr}}})$ with respect to data, as illustrated in Figure 1(c). The effect of applying corrected eigenvalues of the Reynolds stresses can be observed by propagating them through the solver.

Furthermore, the orientation of the Reynolds stress tensor in terms of mutually orthogonal eigenvectors, can be described as a rigid body rotation with three degrees of freedom. RANS results show a discrepancy to the orientation of the eigenvectors obtained from high-fidelity data, e.g. scale resolving simulations. Consequently, intrinsic Tait-Bryan angles α, β, γ in the $z - y' - x''$ convention according to Rothmayr and Hodges [10] will be used to describe the eigenvector rotation from RANS results towards the data. The combination of corrected eigenvalues and eigenvectors completes a predictive framework to model anisotropy discrepancy and propagate the predicted Reynolds stresses to mean flow quantities. In terms of selected flow features, we extended physical quantities to the proposed exhaustive invariant feature list in the work of Ling et al. [8] and Wang et al. [11].

3 Channel flow results

The fully developed turbulent channel flow offers a possibility to demonstrate the coupling of the uncertainty quantification (UQ) perturbation approach with a machine learning model. The discrepancies between RANS and direct numerical simulations (DNS) of Lee and Moser [7] at different Reynolds numbers Re_τ serve as training and testing data for the random forests.

Table 1: Selected hyperparameters for the random forest regressors

Hyperparameter	Target quantities		
	p	\mathbf{p}_{corr}	$\mathbf{p}_{\text{corr}} + \alpha, \beta, \gamma$
max tree depth	6	9	9
min sample count	6	4	4
max active features	3	3	3
number of trees	30	15	30

Before the actual training of the model was conducted, the impact of four different hyperparameters on the accuracy of the random regression forest model is evaluated: the maximum tree depth, the minimum sample count, the maximum number of active features and the number of trees. The selected parameters for the respective target quantities are based on hyperparameter studies (see Table 1). The random forest regression model is trained on $Re_\tau \in \{180, 550, 2000, 5200\}$ using the mean squared error to determine the quality of each split, whereas the models are evaluated in forward CFD-simulations at $Re_\tau = 1000$. The discrepancies with respect to barycentric coordinates are moderate in the channel center and start to increase close to the wall due to strong anisotropy of turbulence. The two-equation, linear eddy viscosity Menter SST $k-\omega$ turbulence model is considered as the baseline

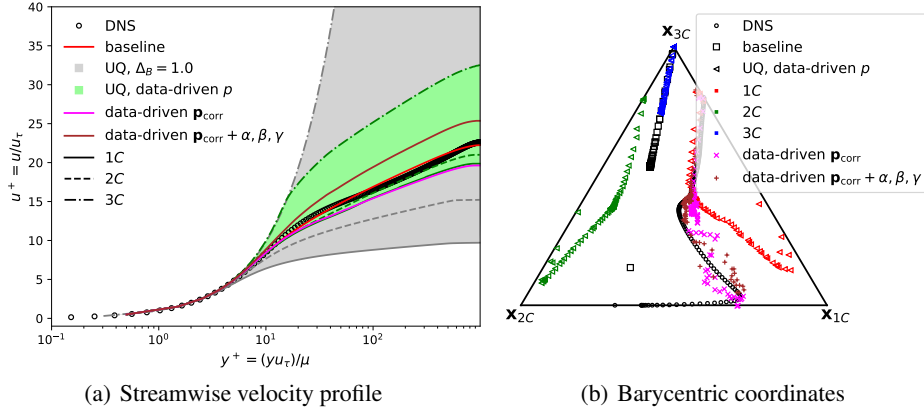


Figure 2: Performance of random forest models in combination with the perturbation uncertainty quantification approach at channel flow simulations of $Re_\tau = 1000$

model in the current study [9]. The advantage of the data-driven UQ approach, determining a relative perturbation strength p can be seen in Figure 4. The predicted uncertainty bounds in Figure 2(a) are significantly less conservative compared to the data-free method with a chosen $\Delta_B = 1.0$, while the traces of the barycentric coordinates look plausible for each of the perturbed solutions (see Figure 2(b)). As expected, the baseline model aligns itself with the plane strain line, whereas the perturbed solutions tend to the limiting states of turbulence in the corners. Additionally, the random forest model, predicting the perturbation direction vector \mathbf{p}_{CORR} , was evaluated in forward CFD solver mode. Although it struggles to predict the x -component of the perturbation, the resulting eigenvalues of the Reynolds stress tensor in the simulation run are satisfying, as illustrated in Figure 2(b). Unfortunately, guaranteeing eigenvalues close to the truth does not seem to be sufficient in order to obtain the correct quantity of interest, in this case the velocity profile.

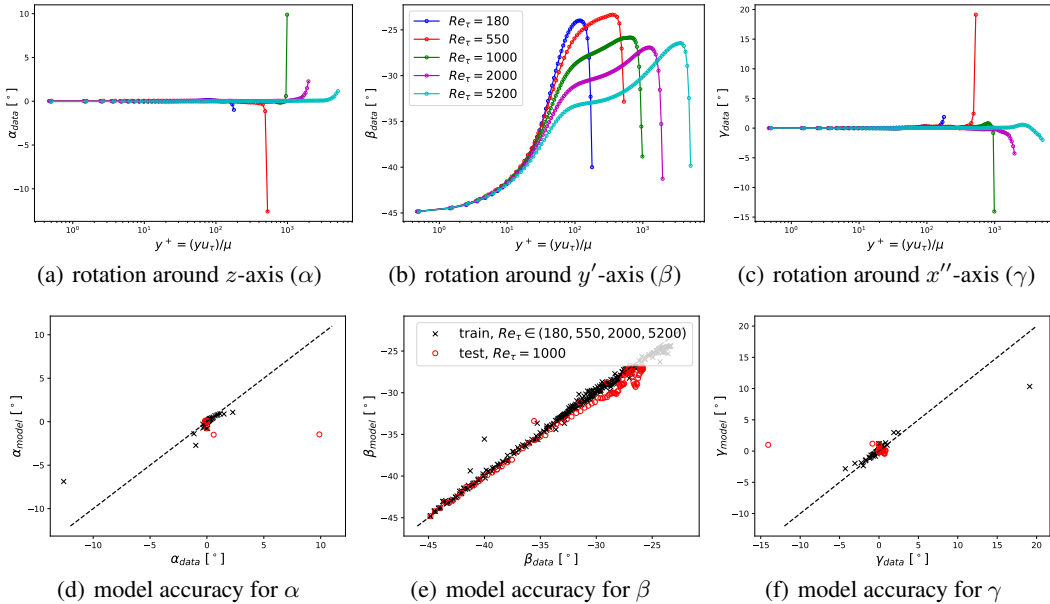


Figure 3: Overview of the eigenvector rotation (RANS \rightarrow DNS) for channel flow

Moreover, we start to explore the discrepancy field regarding the eigenvector orientation of the RANS and the DNS data in Figure 3(a), 3(b) and 3(c). Unlike the differences in barycentric coordinates, the eigenvector misalignment seems to be significant in the log-law region. A random forest regressor

was trained to predict the Tait-Bryan angles, as discussed above. Except for some outliers, the model is able to predict the correct rotation angles within an acceptable range (see Figure 3(d), 3(e) and 3(f)). A combination of propagating true eigenvalues and eigenvectors ($\mathbf{p}_{\text{corr}} + \alpha, \beta, \gamma$) in the forward CFD application led to a an improved prediction of the Reynolds stresses, as illustrated in Figure 4(a), although actually only the anisotropy tensor was corrected without modifying the turbulent kinetic energy explicitly (see Equation 2). Due to the continual response of the equations to the changed Reynolds stresses, the turbulence model produces more accurate turbulent kinetic energy compared with the DNS data (see Figure 4(c)).

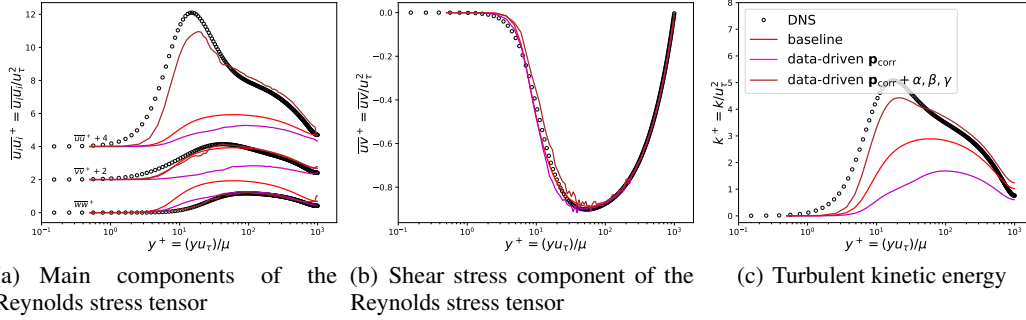


Figure 4: Evaluation of model predicted turbulent quantities at channel flow of $Re_\tau = 1000$

It might be possible, that the non-smoothness of the model prediction causes an issue on the derivatives of the Reynolds stresses, entering the RANS equations, leading to a discrepancy of the flow field [11]. In order to answer the research question, whether a propagation of corrected Reynolds stresses is able to improve the accuracy of quantities of interests, we propagated interpolated Reynolds stresses of the DNS through the RANS equations and observed a perfect match for the velocity profile at $Re_\tau = 1000$. Additionally, adding random noise to these interpolated Reynolds stresses and not noticing significant change in the velocity profile (not shown here), refute the non-smoothness theory. Therefore, the different behaviour of the shear stress component between $10 \leq y^+ \leq 40$ in Figure 4(b) is accountable for the misalignment of the velocity profile for the simulations containing a machine learned correction of the eigenvalues (\mathbf{p}_{corr}) and of the entire anisotropy tensor ($\mathbf{p}_{\text{corr}} + \alpha, \beta, \gamma$). Further work needs to deal with an additional correction for the turbulent kinetic energy based on a machine learning model, which potentially lead to more accurate quantities of interest.

4 Conclusion

This work presents the possibility to estimate data-driven uncertainty bounds for turbulence model with state-of-the-art methods in DLR's CFD solver TRACE. Moreover, recent approaches in literature acted as thought-provoking impulse to train machine learning models on parameters of the initial perturbation approach. The existing concerns were confirmed, that propagating the Reynolds stress corrections is challenging and might not easily lead to correct mean flow quantities of interest. Nevertheless, getting to know the turbulence model uncertainties and especially localize flow regions, where those are present, will help making design exploration and certification more reliable.

Impact statement Complex engineering designs always feature turbulent flows. Although computational power is increasing massively during the past years, scale resolving simulations for design optimization studies seems to be still a long way off. RANS simulation, whose accuracy is heavily dependent on turbulence models, will remain state-of the art in the upcoming years. The present work aims to consolidate the arisen methods in the field of turbulence model uncertainty quantification in combination with machine learning methods. Additionally, designing and implementing a framework to easily conduct uncertainty estimation for turbulence models was a major goal of this work. The authors do not see any ethical concern regarding the present work.

References

- [1] S. Banerjee, R. Krahl, F. Durst, and C. Zenger. Presentation of anisotropy properties of turbulence, invariants versus eigenvalue approaches. *Journal of Turbulence*, 8:N32, 2007. doi: 10.1080/14685240701506896. URL <https://doi.org/10.1080/14685240701506896>.
- [2] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001. doi: 10.1023/A:1010950718922.
- [3] K. Duraisamy, G. Iaccarino, and H. Xiao. Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics*, 51(1):357377, Jan 2019. ISSN 1545-4479. doi: 10.1146/annurev-fluid-010518-040547. URL <http://dx.doi.org/10.1146/annurev-fluid-010518-040547>.
- [4] M. Emory, J. Larsson, and G. Iaccarino. Modeling of structural uncertainties in reynolds-averaged navier-stokes closures. *Physics of Fluids*, 25(11):110822, 2013. doi: 10.1063/1.4824659. URL <https://doi.org/10.1063/1.4824659>.
- [5] J. Heyse, A. Mishra, and G. Iaccarino. Estimating rans model uncertainty using machine learning. *Journal of the Global Power and Propulsion Society*, pages 1–14, 05 2021. doi: 10.33737/jgpps/134643.
- [6] G. Iaccarino, A. Mishra, and S. Ghili. Eigenspace perturbations for uncertainty estimation of single-point turbulence closures. *Physical Review Fluids*, 2, 02 2017. doi: 10.1103/PhysRevFluids.2.024605.
- [7] M. Lee and R. D. Moser. Direct numerical simulation of turbulent channel flow up to $Re_\tau \approx 5200$. *Journal of Fluid Mechanics*, 774:395415, 2015. doi: 10.1017/jfm.2015.268.
- [8] J. Ling, R. E. Jones, and J. Templeton. Machine learning strategies for systems with invariance properties. *J. Comput. Phys.*, 318:22–35, 2016.
- [9] F. Menter, M. Kuntz, and R. Langtry. Ten years of industrial experience with the sst turbulence model. *Heat and Mass Transfer*, 4, 01 2003.
- [10] C. Roithmayr and D. Hodges. Dynamics: Theory and application of kane’s method. 2016.
- [11] J.-X. Wang, J.-L. Wu, H. Xiao, and E. Paterson. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids*, 3 (7), Jul 2018. ISSN 2469-990X. doi: 10.1103/physrevfluids.3.074602. URL <http://dx.doi.org/10.1103/PhysRevFluids.3.074602>.

Checklist

1. For all authors...
 - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? **[Yes]** See Section 4 and the impact statement.
 - (b) Did you describe the limitations of your work? **[Yes]** See Section 4
 - (c) Did you discuss any potential negative societal impacts of your work? **[Yes]** There is no negative societal impact, see impact statement.
 - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? **[Yes]** We acknowledge these guideline
2. If you are including theoretical results...
 - (a) Did you state the full set of assumptions of all theoretical results? **[N/A]**
 - (b) Did you include complete proofs of all theoretical results? **[N/A]**
3. If you ran experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **[N/A]**

- (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
- (a) If your work uses existing assets, did you cite the creators? [Yes] See list of references.
 - (b) Did you mention the license of the assets? [N/A]
 - (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [Yes] The only data, we used, was published and open source accessible DNS data.
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
- (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
 - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
 - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]