



Data-driven augmentation of turbulence models for adverse pressure gradient flows

Anand Pratap Singh*, Racheet Matai[†], Karthik Duraisamy[‡], Paul Durbin[§]

Turbulent flows involving adverse pressure gradients, curvature and mild separation are analyzed and data-driven augmentations are developed for predictive models. Large eddy simulations are performed over a set of parametric flow configurations and the impact of varying curvature, boundary layer thickness and Reynolds number is assessed. The first step in the data-driven methodology involves the determination of functional discrepancies in existing models using inverse modeling. The inferred discrepancy is reconstructed as a function of locally non-dimensional flow features using a machine learning algorithm. This machine-learned discrepancy is embedded within a $k-\omega$ turbulence model. The impact of the choice of data used for the inverse modeling on the predicted velocity field and Reynolds stresses is analyzed in detail. The data-augmented turbulence model, trained using a very small subset of flows and limited data is shown to yield much-improved predictions of flow properties, over the entire set of configurations. This work represents a key step toward the development of more general data-augmented turbulence models.

I. Introduction

Inaccuracies in prediction of turbulent flows with adverse pressure gradients and separation has been long identified as a major stumbling block for the broader use of computational fluid dynamics in complex problems. In the aerospace industry there are still many problems of significance that cannot be properly tackled without an ability to better predict the afore-mentioned effects. Examples include high-lift systems, the off-design operating envelope of new vehicles, aeroacoustics of airframe noise, turbomachinery flows, combustors, etc.

There is a long-standing need to improve RANS-based closure models. Some recent developments suggest that data can be used more comprehensively to achieve this goal :

- We are now able to perform significant numbers of highly-accurate DNS and LES experiments, that result in massive amounts of high quality data,
- The capability to extract information from data has advanced in recent years,
- Machine learning methods are being explored for improving predictive capabilities in a number of disciplines.

The prevailing idea in the data science community is that if sufficient observations of a natural phenomenon are available, patterns and causality can be extracted and a predictive model can be created that bypasses the need for describing the detailed physics of the problem. We do not subscribe to this as a viable alternative for turbulence modeling; but, rather, we assume that high-fidelity data, combined with, and informed by, a detailed knowledge of the physical constraints on the evolution of turbulence is more likely to be a successful approach. Consonant with that philosophy, over the past 5 years, we, and others, have investigated data-driven techniques for model improvement. In Ref. 1, we applied neural networks to large eddy simulation data to learn the functional form of the discrepancy in the Reynolds stress and injected these functional forms into a predictive simulation. Since then, a number of groups²⁻⁷ have explored various ideas of using data to characterize and improve turbulence models.

While most of the cited works have focused on machine learning to reconstruct model discrepancies and constitutive relationships, the paradigm of field inversion and machine learning (FIML)²⁻⁴ **enforces consistency between the**

*PhD Candidate, Aerospace Engineering, Univ of Michigan

[†]PhD candidate, Aerospace Engineering, Iowa State Univ

[‡]Associate Professor, Aerospace Engineering, Univ of Michigan

[§]Professor, Aerospace Engineering, Iowa State Univ

learning and the model by introducing an inference step. In this procedure, instead of learning from the DNS data fields, the RANS model is first improved (via field inference) and learning is performed on the RANS model itself. An additional advantage is that **field inversion does not demand large data sets; it can work with sparse data from experiments.** The FIML approach consists of three steps : a) Infer the spatial (non-parametric) distribution of the model discrepancy in a number of problems using Bayesian inversion, b) Transform the spatial distribution into a function of flow variables, using machine learning, and c) Embed the functional form in a predictive setting. Applications of the FIML paradigm can be found in Refs. 8,9.

In the present work, we focus on a parameterized set of flows to carefully assess how one may improve the modeling of turbulent flows under adverse pressure gradients and curvature leading to flow separation. The implications of using sparse surface data (skin-friction) on the prediction of field quantities (velocity and Reynolds stress) is also analyzed in detail.

II. Field Inversion and Machine Learning (FIML) for Data-augmented Modeling

A schematic of the FIML^{4,8} methodology is shown in Fig. 1. Given an existing turbulence model, a spatially-varying correction field $\beta(\mathbf{x})$ is introduced as a model coefficient. Starting with some high fidelity or experimental data \mathbf{G}_d , (an appropriately regularized form of) $\|\mathbf{G}_d - \mathbf{G}_\beta\|$ is minimized, where \mathbf{G}_β is a RANS solution for a given correction field. Note that the intent is not to infer a limited set of parameters that characterizes the correction, but rather to extract the optimal field $\beta(\mathbf{x})$. Thus we are directly addressing structural errors and uncertainties in the model.

The correction fields are obtained for an ensemble of problems (on different data sets $\mathbf{G}_d^1, \mathbf{G}_d^2, \dots$ as shown in the figure) representative of the underlying physics. To be useful in predictive modeling, the spatial dependence of $\beta(\mathbf{x})$ is then transformed into feature space using machine learning^a. Machine learning effectively constructs a functional form $\beta(\boldsymbol{\eta})$, where $\boldsymbol{\eta}(\mathbf{U}(\mathbf{x}))$ are derived from flow and turbulence variables \mathbf{U} . The elements of the feature vector $\boldsymbol{\eta}$ have to be non-dimensional quantities^b such that the functional corrections can be translated across different problems.

Inversion and machine learning are *pre-processing* steps as shown in the red box, figure 1. During the predictive simulation, the solver will pass feature vectors $\hat{\boldsymbol{\eta}}$ to the ML evaluation routine and receive appropriate correction factors $\hat{\beta}$ for the turbulence model coefficients.

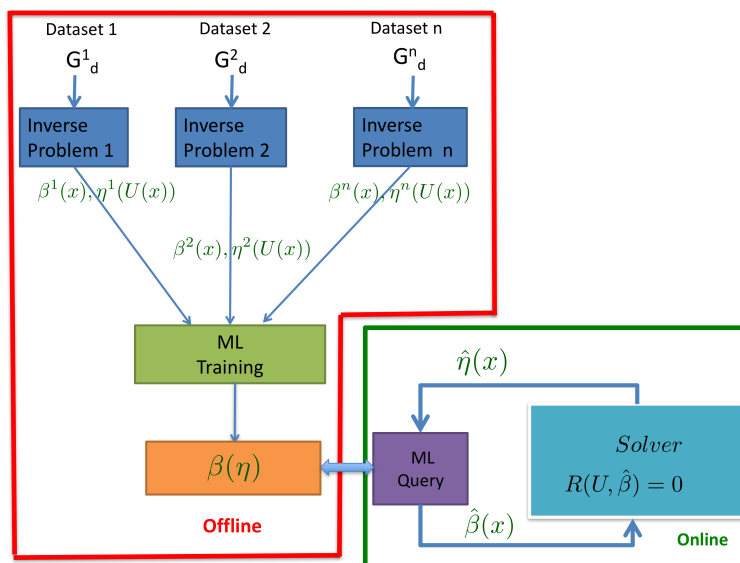


Figure 1: Schematic of field inversion and machine learning framework for data-augmented turbulence modeling

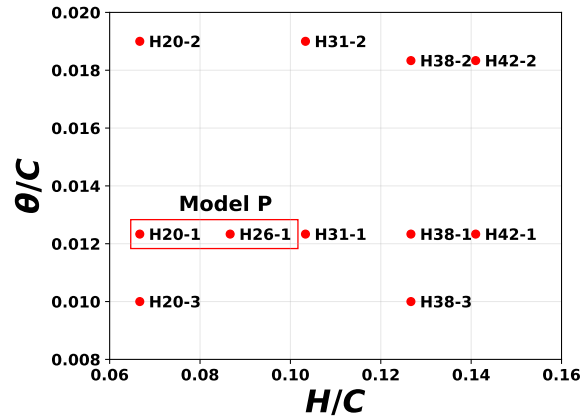


Figure 2: Labels for all the cases, with various bump heights and inlet momentum thicknesses, used for this work. The red box denotes the data-set used to train a model used for augmentation.

III. Data generation via Large Eddy Simulations

A. Solver and Validations

Turbulent flows over a parameterized set of bumps were simulated via wall-resolved, large eddy simulations (LES). The simulations were carried out with the OpenFOAM, PISO solver, invoking the dynamic Smagorinsky subgrid model.¹⁰ The baseline geometry is that of Webster, et al.¹¹ Their data were used for validation. Then the bump height and inlet momentum thickness were varied to create a series of datasets by high fidelity numerical simulation.

Fig. 3 is a schematic of the mesh and geometry. The crest of the baseline bump is 20mm high and, thus, this bump is labeled as H20. Upstream of the bump, an uncoupled, simultaneous, flat plate simulation generates a fully developed turbulent boundary layer that is fed into the main simulation domain. In the upstream region, the flow is recycled and rescaled to generate a fully developed turbulent boundary layer, as described by Arolla.¹² The inlet profile has a momentum thickness of $0.0121 C$ and $Re_\theta = 2500$, where the chord $C = 0.305m$.

Fig. 4 confirms that the simulation is in good agreement with experimental measurements.¹¹ The LES data extends closer to the wall than the laboratory data. It will be seen that this is crucial to the present topic of turbulence closure modeling. To ascertain whether the numerical resolution is adequate, the subgrid viscosity is compared to the kinematic viscosity in Fig. 5. The ratio is less than 0.5 in most of the domain, indicating close to DNS grid resolution.¹³

B. Parametric Studies

Four additional bumps were created by successively increasing the height of the crest. These are labeled as H26, H31, H38, H42 with the number denoting the height of the crest in mm. The flow over the original bump, H20, did not separate; H26 is on the verge of separation; and H31-42 involve separation, as seen in the skin friction coefficient (C_f) plots in Fig. 6a. As the bump height is increased, C_f over the bump also increased which is attributable to greater flow acceleration (Fig. 6b) over the crest. Fig. 8 shows the progressive change in velocity as the crest heights are increased. A similar progression in the turbulent shear stresses can be seen in Fig. 9. H42 involves a large (greater than 3-fold) increase in the shear stress compared to the original geometry at the end of the bump.

Additional data sets were created by changing the inlet momentum thickness. Two additional non dimensional momentum thicknesses equal to 0.019 (momentum thickness 2) and 0.0098 (momentum thickness 3) were prescribed at the inlet. Momentum thickness 2 is used to generate flow over H20, H31, H38, H42 and momentum thickness 3 is used with H20 and H38. The effect of the change in momentum thickness is shown in Fig. 5. The qualitative shape is determined by the pressure gradient, which depends on bump height. This can be seen in Fig. 7. The change in inlet momentum thickness did not have a significant impact on the flow behaviour and Fig. 10 shows that in non-dimensional

^aThe machine learning step may be avoided by clever trial and error model development

^bThey also should be coordinate invariant. $|S|\tau$ is an example of an acceptable feature, where $|S|$ is the magnitude of the strain-rate tensor and τ is a turbulent time-scale.

form, the velocity profiles overlap almost perfectly, except in the wake region. The turbulent shear stresses are also qualitatively similar as shown in Fig. 11.

Fig. 12b shows that the TKE is suppressed in the favorable pressure gradient region of the H20 bump and enhanced at the end. A similar trend can be seen over the H42 bump (Fig. 13b) but with a much higher TKE level in the recovery region. RANS fails to correctly capture these high TKE regions. Also, Fig. 13a clearly shows a region of backflow indicating separation for H42 which is absent in the H20 case (Fig. 12a).

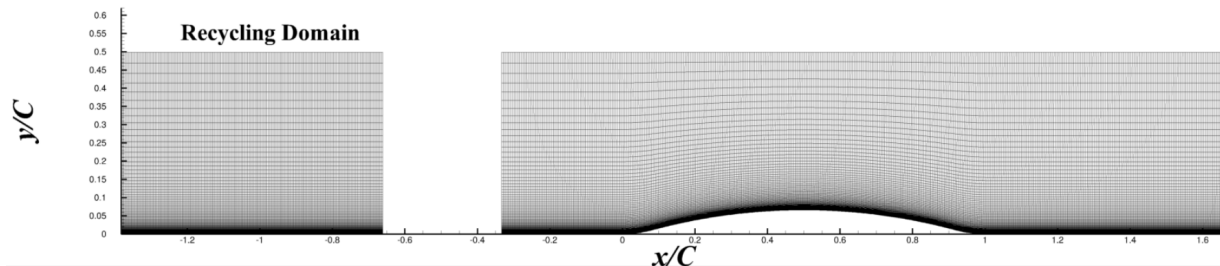


Figure 3: Mesh used for the LES simulation. Note the disconnected recycling domain for generating the inlet flow ($C = 0.305\text{m}$)

Re_θ	Δy^+	Δx^+	Δz^+	Number of cells	N_x, N_y, N_z
2500	<0.5	41	17	7 million	584,109,110

Table 1: Grid characteristics

IV. Inverse Modeling

Existing turbulence models perform satisfactorily for thin attached boundary layers with mild pressure gradients and curvature. Further, most of the popular turbulence models can recreate well-studied features such as the log-layer. Therefore it is desirable to leverage data to further improve such models rather than develop models from scratch. In the present work, we start with an eddy viscosity based two equation turbulence model – Wilcox’s $k - \omega^{14}$ – as our baseline model. The $k - \omega$ model utilizes transport equations for turbulent kinetic energy (k) and specific dissipation (ω) to estimate the eddy viscosity. For the baseline $k - \omega$ model, the transport equation for ω has the following form,

$$\frac{D\omega}{Dt} = P(k, \omega, \mathbf{U}) - D(k, \omega, \mathbf{U}) + T(k, \omega, \mathbf{U}), \quad (1)$$

where \mathbf{U} represents the Reynolds averaged conserved flow variables, $P(k, \omega, \mathbf{U})$, $D(k, \omega, \mathbf{U})$, and $T(k, \omega, \mathbf{U})$ represent the production, destruction and transport terms respectively. These source terms are modeled and typically hand-calibrated using canonical flows. The modeling involves discrepancies which are functional (rather than parametric) in nature. While we have explored many variants⁹ of introducing the discrepancy in eddy viscosity-based models³ and Reynolds stress-based models,¹⁵ we restrict attention to the scale equation in this work. Thus we introduce a spatially varying discrepancy term, $\beta(\mathbf{x})$, to the transport equation of the form

$$\frac{D\omega}{Dt} = \beta(\mathbf{x})P(k, \omega, \mathbf{U}) - D(k, \omega, \mathbf{U}) + T(k, \omega, \mathbf{U}). \quad (2)$$

The first step in the present formulation involves inferring the discrepancy field $\beta(\mathbf{x})$ — as discussed next.

Assume a mesh consisting of N_m control volumes. Given N_d data points (such as wall pressure, skin-friction, etc) $G_{j,d}$, we define the following inverse problem to extract the optimal field $\beta \equiv \beta(x_n) : 1 \leq n \leq N_m$:

$$\min_{\beta} J_1 + \lambda J_2 \equiv \min_{\beta} \sum_{j=1}^{N_d} [G_{j,d} - G_j(\beta)]^2 + \lambda \sum_{n=1}^{N_m} [\beta(x_n) - 1]^2, \quad (3)$$

where $G_j(\beta)$ is the output of the RANS computation. The optimization problem is solved using gradient based optimization algorithm as discussed in Ref. 3. The solution of an inverse problem is sensitive to the regularization

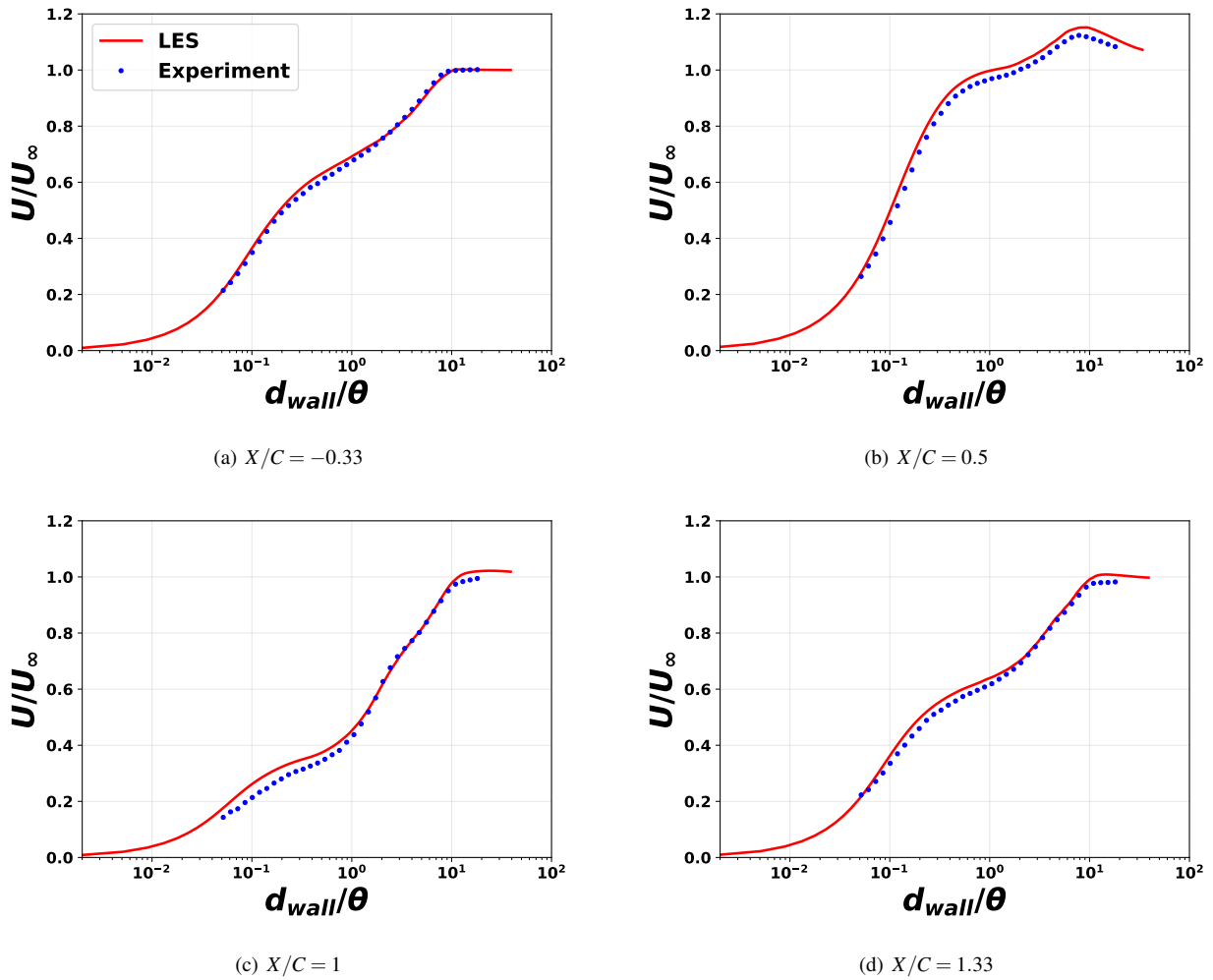


Figure 4: LES results compared with experimental measurements at different streamwise locations in the domain.

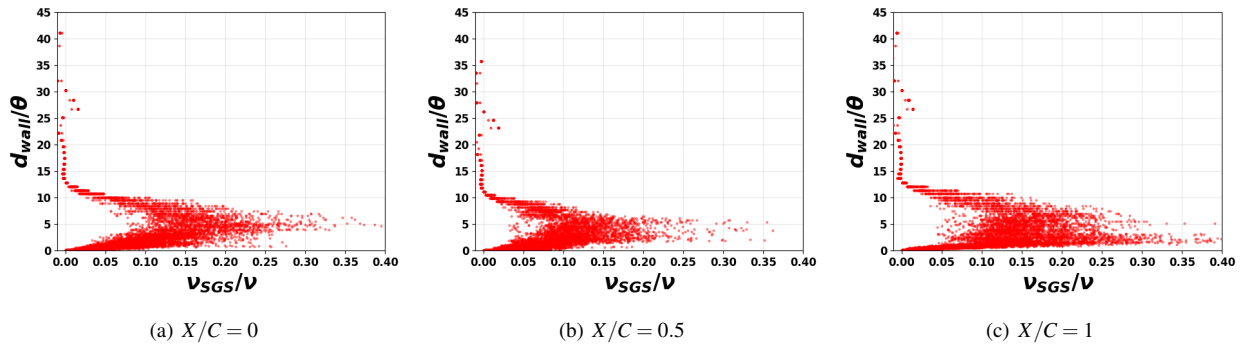
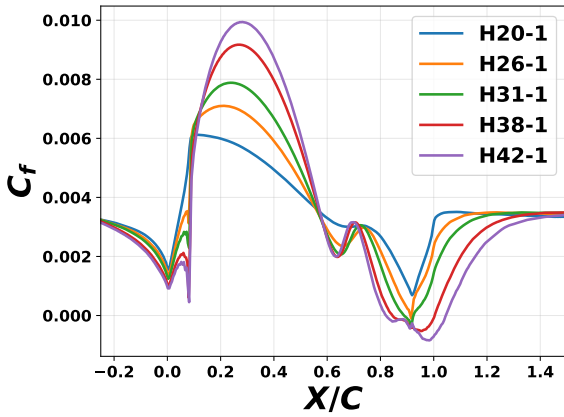
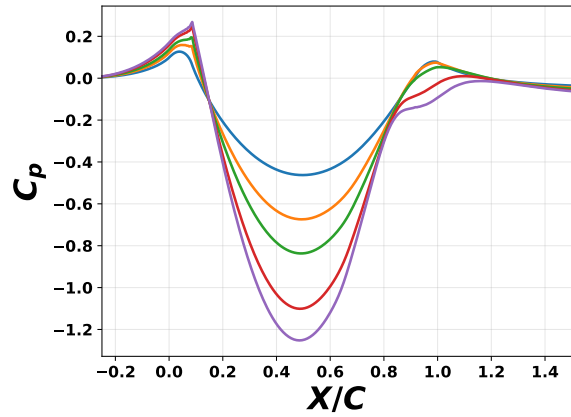


Figure 5: Ratio of subgrid viscosity to kinematic viscosity at three locations over the bump.

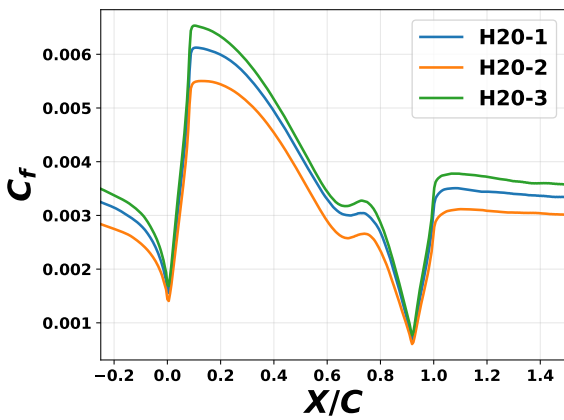


(a) Skin friction

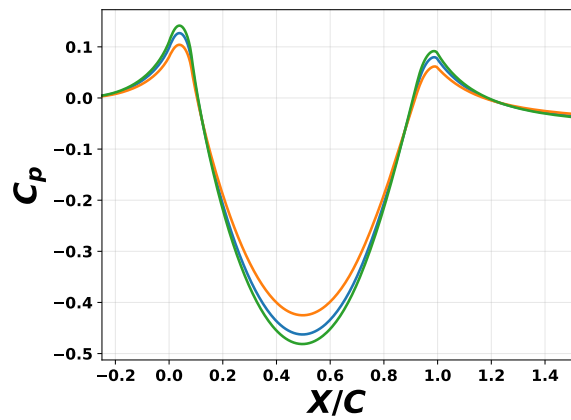


(b) Surface pressure

Figure 6: Comparison of skin friction coefficient and pressure coefficient over bumps with different bump heights.



(a) Skin friction



(b) Surface pressure

Figure 7: Comparison of C_f and C_p over the bump for 3 different inlet momentum thicknesses.

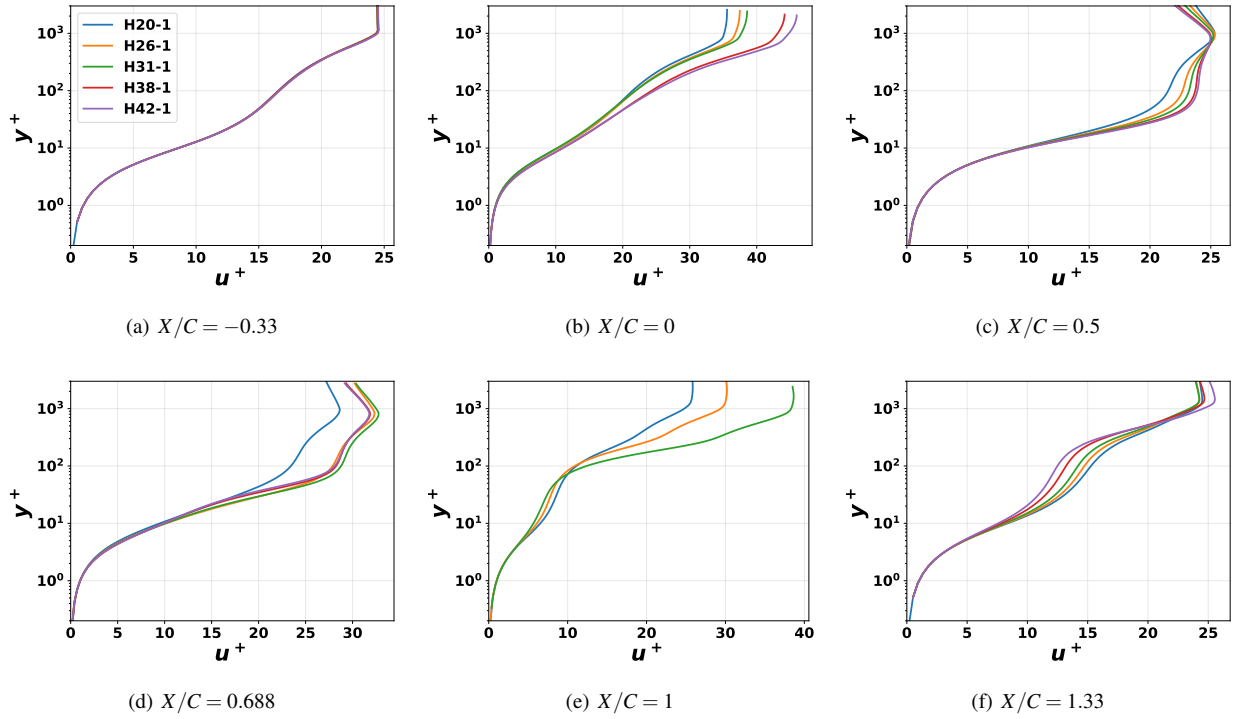


Figure 8: Comparison of streamwise velocity at multiple streamwise locations for different bumps.

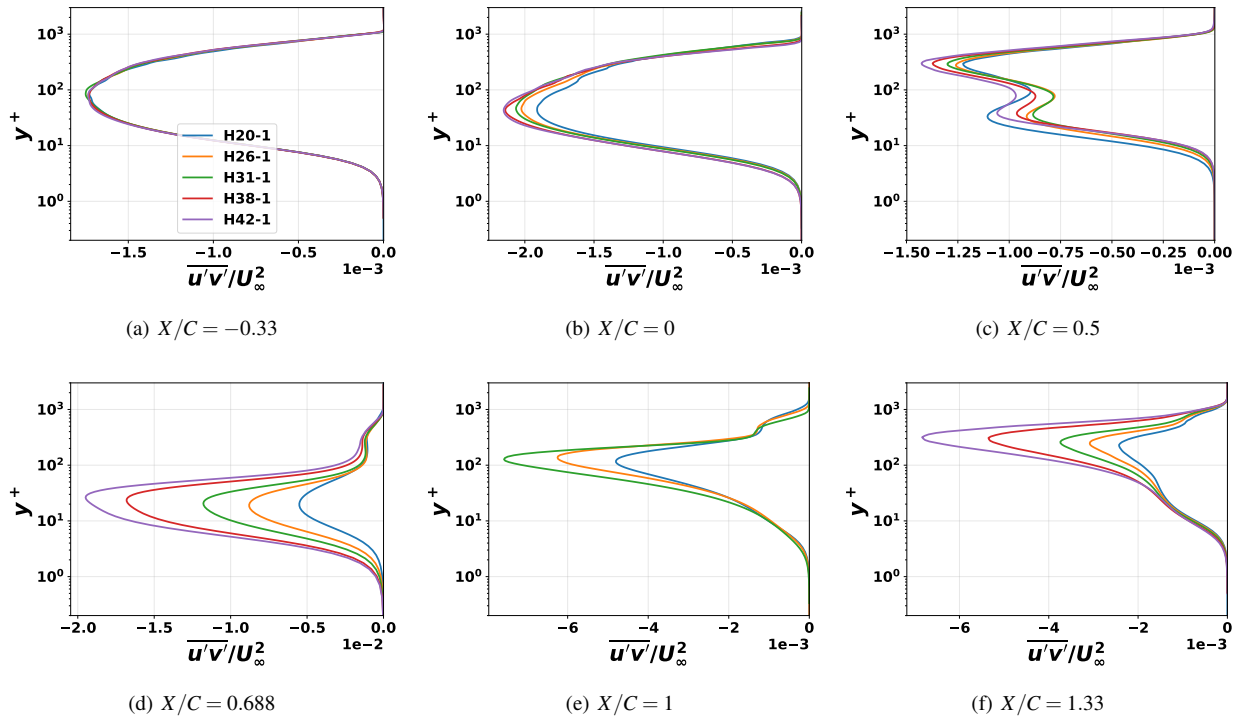


Figure 9: Comparison of turbulent shear stress at multiple streamwise locations for different bumps.

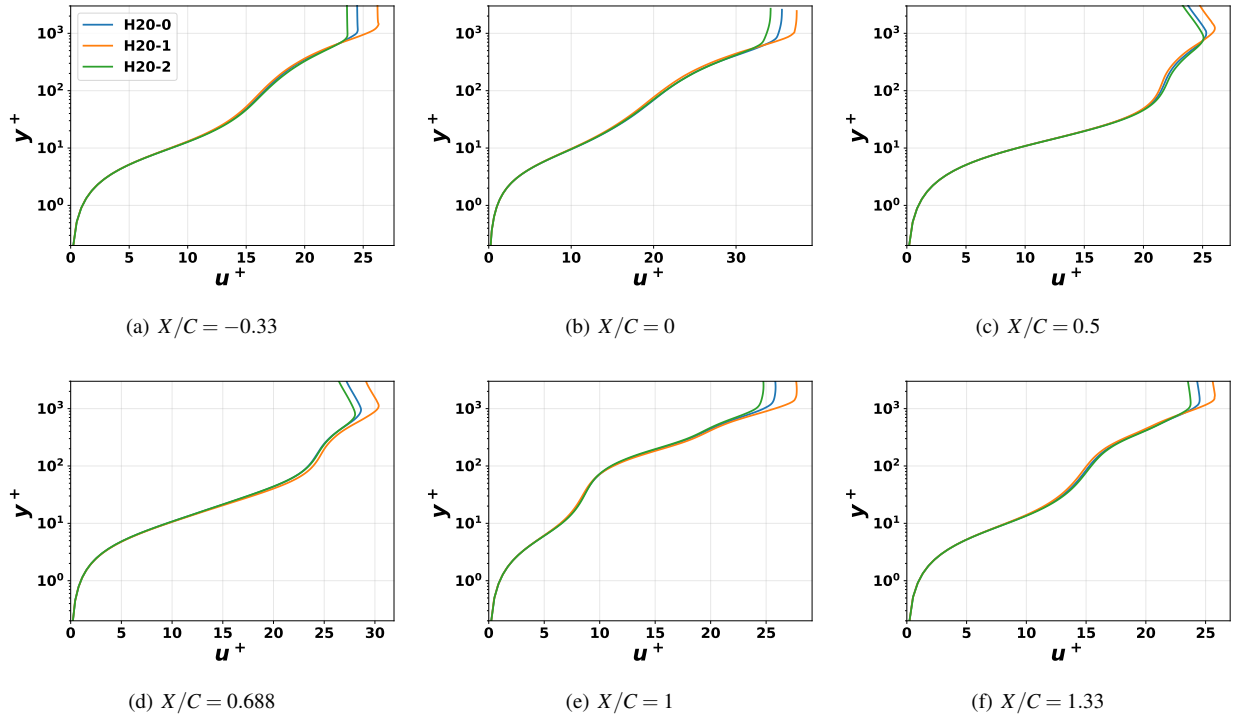


Figure 10: Comparison of streamwise velocity at multiple streamwise locations for three inlet momentum thicknesses.

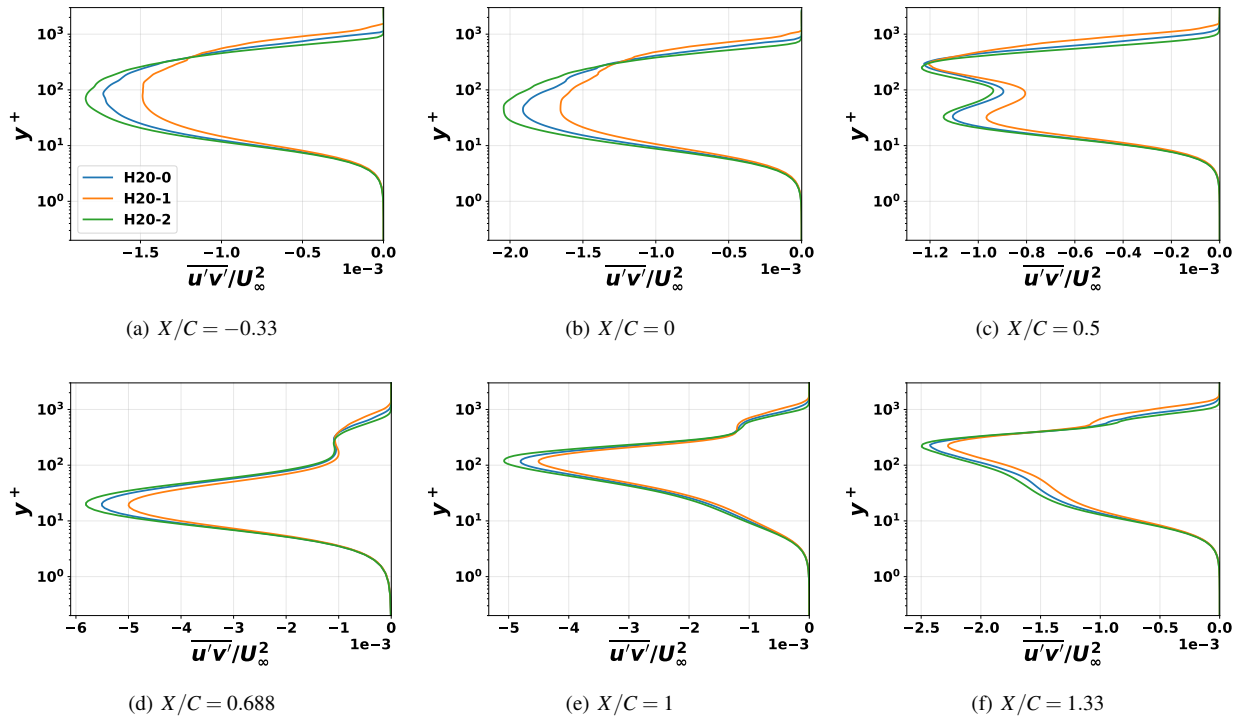


Figure 11: Comparison of turbulent shear stress at multiple streamwise locations for three inlet momentum thicknesses.

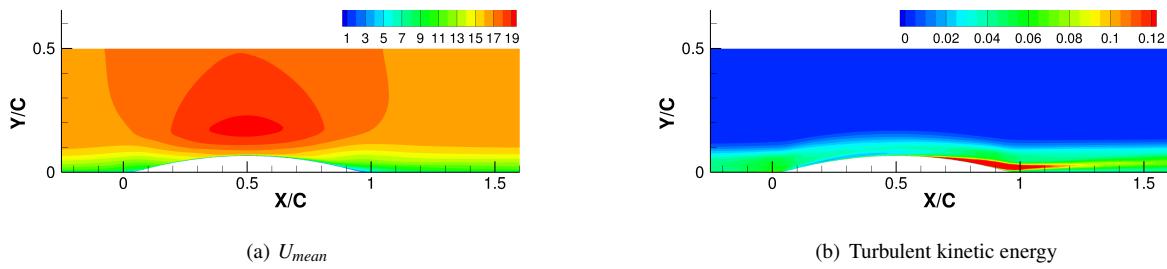


Figure 12: Contour plot of mean velocity in streamwise direction and TKE for H20-1.

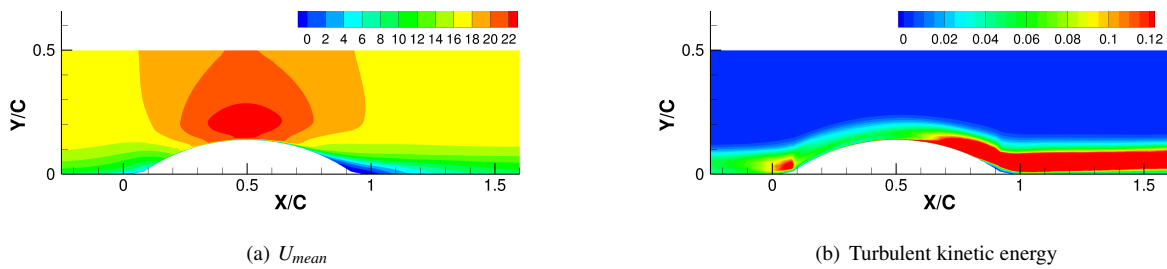


Figure 13: Contour plot of mean velocity in streamwise direction and TKE for H42-1.

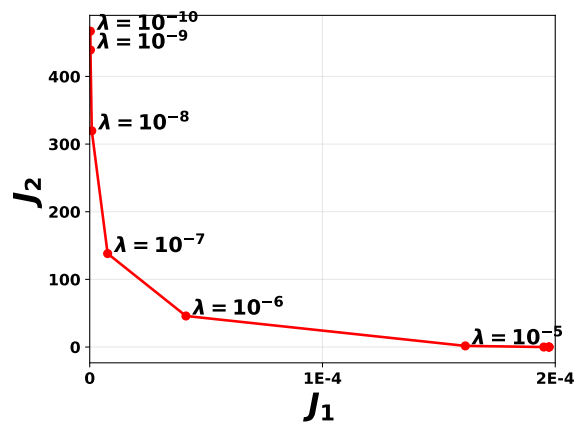


Figure 14: x-y plot of the two parts of the objective function, J_1 and J_2 , is used to fix the regularization constant. For this representative case $\lambda_{elbow} \approx 10^{-6}$.

constant λ . A very low value of λ leads to ill-conditioning (and ill-posedness) as well as over-fitting, while a high λ limits the improvement in the discrepancy term.

If the focus of this work were inference from a single dataset, and if we have detailed information about the error in the data and the model, Bayesian inversion could be used.⁴ In this work, however, the goal is improved modeling over a range of datasets. Therefore, a somewhat *ad hoc* decision has to be made regarding the amount of regularization. A value of λ is selected by obtaining the inverse solution for a set of λ 's followed by plotting the discrepancy term (J_1) and the regularization term (J_2) of the objective function and selecting the value corresponding to the elbow. Fig. 14 illustrates a representative J_1 vs J_2 curve which results in $\lambda_{elbow} = 10^{-6}$.

As discussed in the previous section, LES data are available for 11 different cases. Inverse problems are solved for each of these cases by utilizing skin friction at the lower wall for $G_{j,d}$ in eq. 3 and $\lambda = \lambda_{elbow}$. The resulting discrepancy field, $\beta(\mathbf{x})$, for each of these cases is shown in Fig. 16 and the corresponding skin friction is shown in Fig. 15.

The upstream half of the bump results in a favorable pressure gradient, which stabilizes the turbulent flow. The inverse solution promotes this stabilization by increasing the production of ω (regions with $\beta(\mathbf{x}) > 1$). The downstream half is dominated by an increase in turbulence with $\beta(\mathbf{x}) < 1$. The resulting skin friction is almost identical to the LES data and displays the same degree of separation as the LES. The shaded region in Fig. 15 shows the effect of selecting different values of the regularization constant λ .

To investigate the effect of the type of data used for inversion, the inverse problem was solved for H20-1 the full velocity field as $G_{j,d}$. Fig. 17 compares the effect of using skin friction to using the full field velocity on the skin friction predicted by the inferred model. Figs. 18-23 compares velocity, Reynolds stress and the eddy viscosity profiles at different stream-wise locations. The two inverse solutions are markedly similar and hence using (sparse) skin friction data for inversion is justified. Improvements in Reynolds stresses and eddy viscosity, which are not used for inversion, also shows that the inverse modeling does not just overfit the model but is able to infer the *actual* discrepancy. The inverse solution for such variables can highlight the nature of discrepancies present in turbulence models, and can be used, in their own right, for model improvement. These solutions also can be used to improve the models via machine learning, as is described in the next section.

V. Machine Learning and Predictions

The optimization problem produces spatial fields $\beta(\mathbf{x})$. These are specific to the particular problem solved to obtain them and cannot be utilized for predictions. The inferred discrepancies, $\{\beta(\mathbf{x})_1, \beta(\mathbf{x})_2, \dots, \beta(\mathbf{x})_n\}$, need to be correlated as functions of flow variables, η . This is done using machine learning algorithms.

In particular, *adaptive boosting* (Adaboost¹⁶) has been adopted for that purpose. Other machine learning algorithms, including neural networks and Gaussian processes, have been used by the authors previously.^{8,17} Adaboost iteratively applies weak learners to improve the machine learning model for the data points that previous learners failed to fit. For each iteration, points that are mistakenly predicted by the current learners are marked with higher weights, so that the next iteration perceives them as being more important.

Given training data $(\boldsymbol{\eta}_i, \beta_i)_{i=1,N}$, the steps in Adaboost regression are as follows:¹⁶

1. Assign an initial weight, $w_i = 1/N$, to each data point.
2. Choose N samples from the data (with replacement). Probability of choosing a data point is based on its weight.
3. Construct a predictor (e.g., decision tree regressor with a maximum depth and number of features) that minimize the error on current sample of the weighted training data.
4. Compute a measurement of the confidence of the predictor in step 3 based on a chosen loss function.
5. Update the weights w_i based on the confidence and point-wise loss function.
6. Repeat from Step 2 until the desired total number of predictors is reached.
7. The regression result is computed as a weighted median of all predecessor predictors based on the confidence of each predictor and learning rate.

The selection of the flow features, η_i , is critical for the quality of the model. These flow features must be non dimensional. Ideally, they are local and coordinate invariant. For this work, the features that were selected are

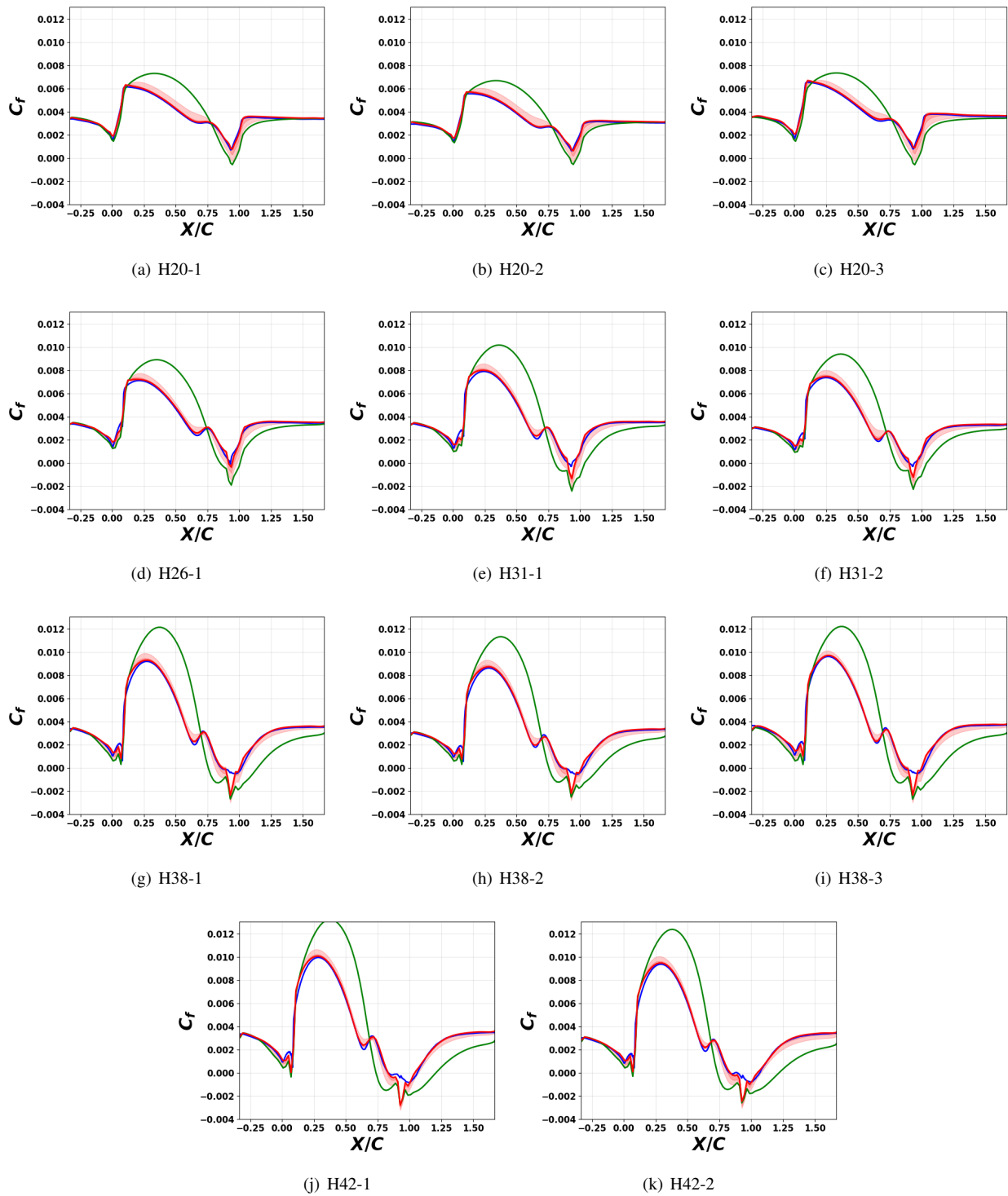


Figure 15: Skin friction for all the cases used in this work. Legend: LES, base $k - \omega$, inverse $k - \omega$ using C_f . Shaded red region contains inverse solutions for various $10^{-10} < \lambda < 10^{-6}$.

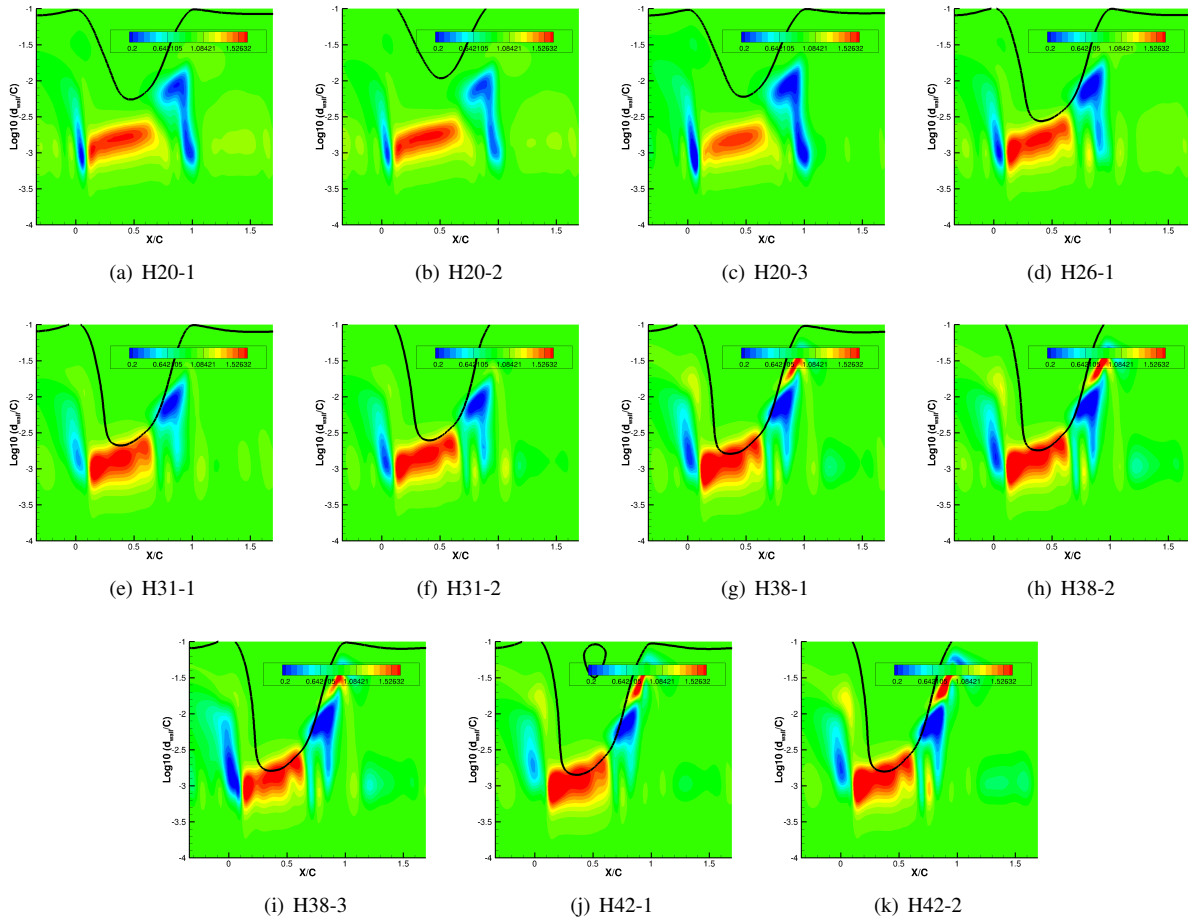


Figure 16: Inferred spatial discrepancy field, $\beta(\mathbf{x})$ using data for skin friction. Thick black line marks approximate location of the boundary layer edge.

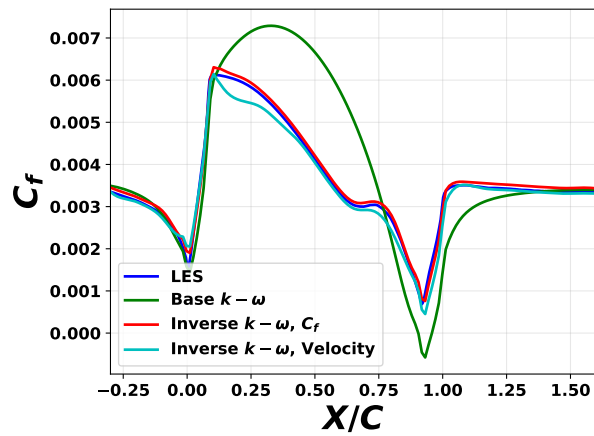


Figure 17: Skin friction obtained after inference using data for the skin friction and the full field velocity.

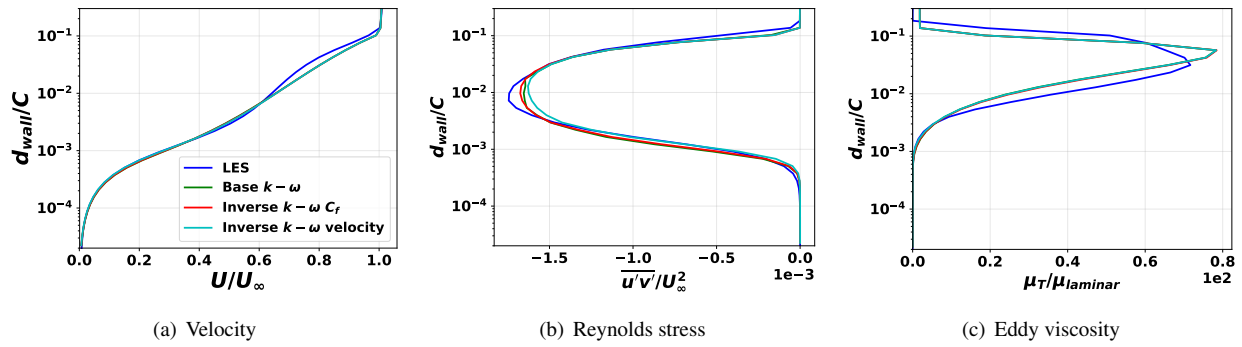


Figure 18: Flow solution at $X/C = -0.16$ obtained after inference using data for the skin friction and the full field velocity.

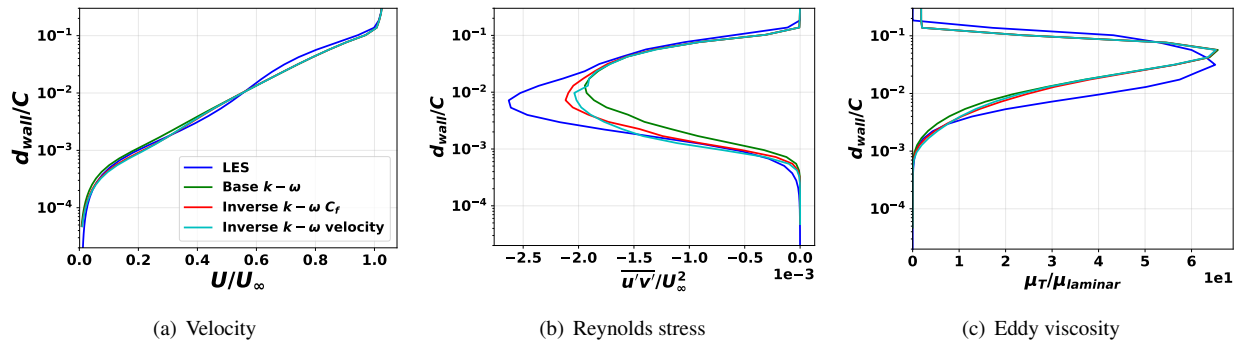


Figure 19: Flow solution at $X/C = 0$ obtained after inference using data for the skin friction and the full field velocity.

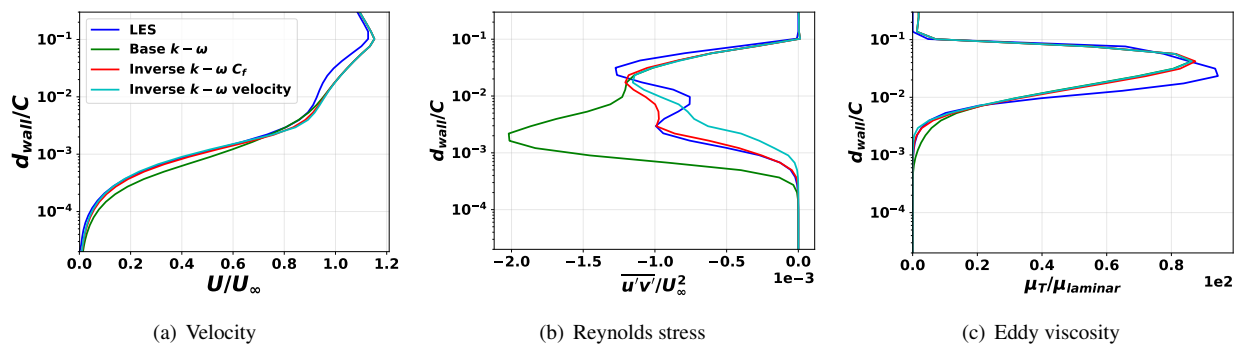


Figure 20: Flow solution at $X/C = 0.33$ obtained after inference using data for the skin friction and the full field velocity.

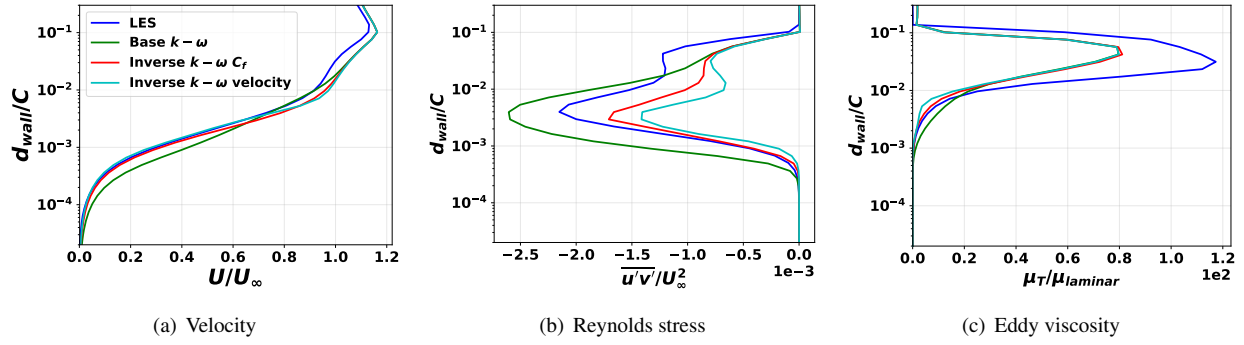


Figure 21: Flow solution at $X/C = 0.66$ obtained after inference using data for the skin friction and the full field velocity.

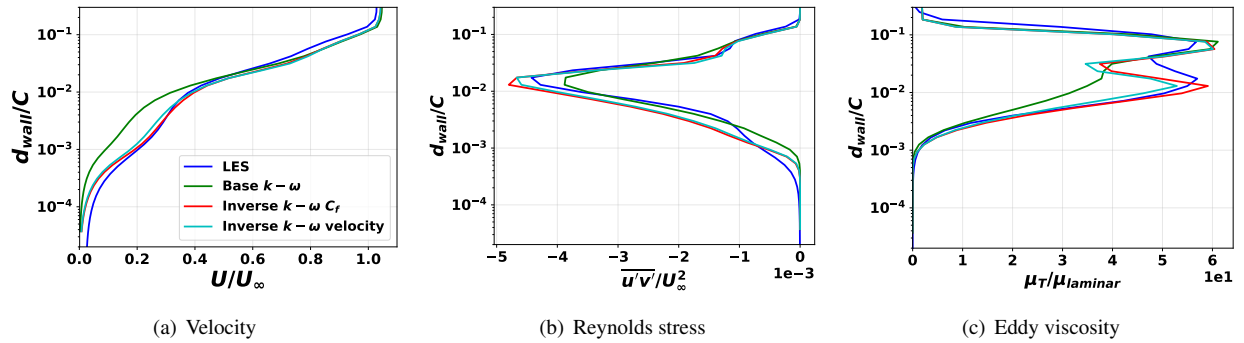


Figure 22: Flow solution at $X/C = 0.98$ obtained after inference using data for the skin friction and the full field velocity.

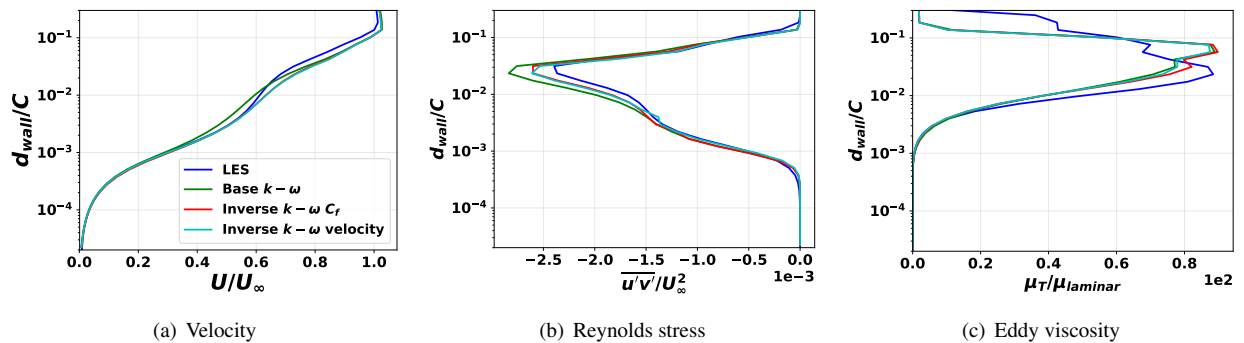


Figure 23: Flow solution at $X/C = 1.31$ obtained after inference using data for the skin friction and the full field velocity.

$\{\rho|S|d_{wall}^2/\mu, \mu_T|S|/\tau_{wall}, d_{wall}/C\}$, where $|S|$, ρ , d_{wall} , τ_{wall} represent the strain-rate magnitude, density, wall distance, and, the wall shear stress, respectively. These features are not definitive, but were found to be sufficient for the cases presented in this paper and others. τ_{wall} is not local. Wall distance is effective for correlating the corrections, because they are most important near the surface, as seen in fig. 16.

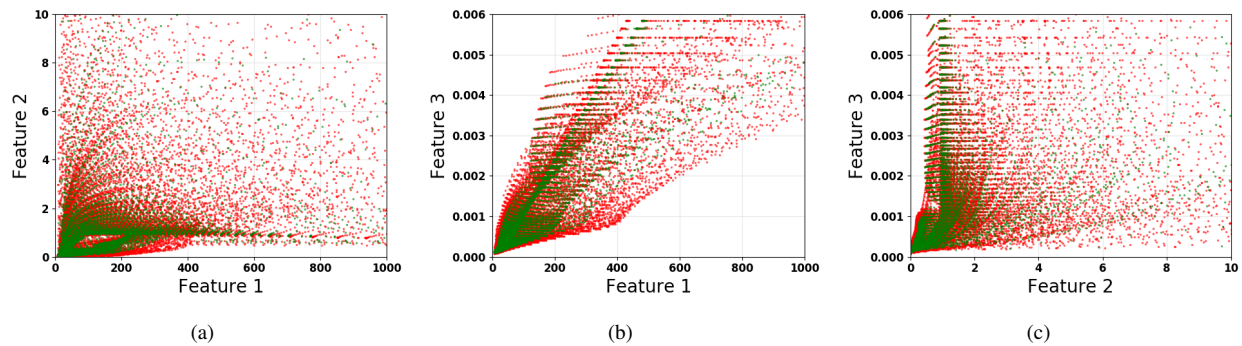


Figure 24: Pairwise scatter plots of the flow features used for training. Features for all the cases are shown in red and the features used to train model **P** are shown in green.

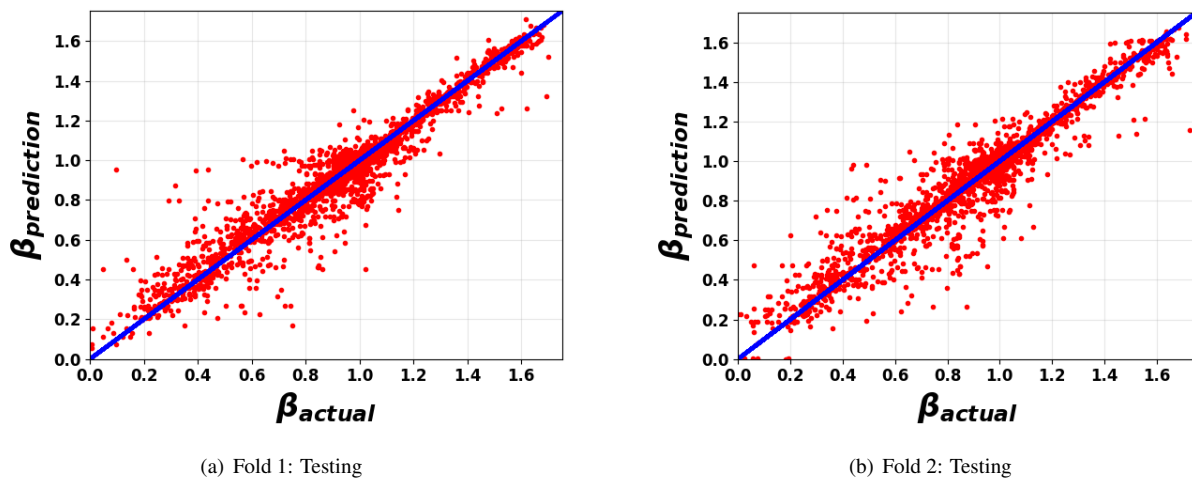


Figure 25: Training plots for the model **P** using Adaboost algorithm.

A number of combinations of the 11 datasets were used to train an ensemble of models. These models were then used in the RANS solver for predictions. To train the models, 2-folds cross validation is performed. The data is randomly partitioned into two equally-sized groups. One of those groups is used as a validation group and the other is used for training, thus resulting in a total of two test sets. Fig. 25 shows the training plot for a model trained on inverse solutions for the H20-1 and the H26-1 case (model **P** as marked in Fig. 2). The coefficients of determination for testing for the two folds was 0.92 and 0.91.

The predicted skin friction for all the cases are shown in Figs. 26 and 15. Figs. 28-33 show the velocity, Reynolds stress, and eddy viscosity at different streamwise locations. The solution with model **P** results in good prediction over all the cases. It is notable that just by using the skin friction data in two configurations, improvements are observed in the full field quantities using the machine learning augmentation over the entire set of cases. Predictions using other combinations vary in quality but display consistent improvement over the baseline model. While a more rigorous analysis is required to apriori determine the quality of a given combination, some insight can be gained by looking at the pairwise scatter plots of the flow features (Fig. 24). The points in green, which are used to train model **P**, are a good proxy of the entire data-set. Reduction in the number of data points also leads to better training of the machine learning model resulting in a better augmentation.

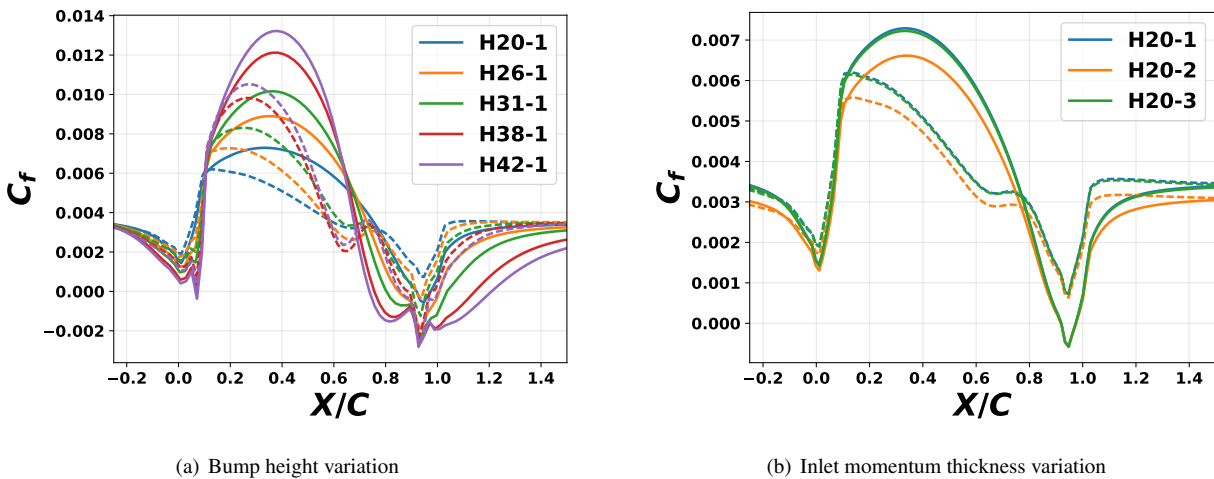


Figure 26: Skin friction prediction using the baseline $k - \omega$ (solid line) and the Adaboost-augmented model P (dashed line).

VI. Summary

Large Eddy Simulation data were generated for a series of bumps of increasing height. They provide a pressure distribution with increasingly large favorable and adverse gradients. A separation bubble forms behind the higher bumps. Using limited data from the LES, field inference and machine learning were used to augment Wilcox's $k - \omega$ turbulence model. The augmented model was then used to predict flow fields. The observations are summarized as follows:

1. Changing the bump height has a significant impact on the turbulent shear stress and velocity profiles. Increasing the bump height also increases the TKE downstream of the bump. These flow characteristics are not captured correctly by existing RANS models.
2. Spatially varying discrepancies are introduced in the ω equation. The inverse problem is solved by optimal fit to the LES skin friction data. This formulation provides a discrepancy field is consistent with the RANS model.
3. Inference is performed with either a) the surface skin friction, or b) the full velocity field. They yield similar velocity and Reynolds shear stress predictions, in agreement with the LES. The improvement in the full flow field using just the skin friction illustrates the possibility of using sparse experimental data for model development.
4. The inferred discrepancies are reconstructed, as a single function of non dimensional flow features, using a machine learning algorithm. A number of combinations of the inferred discrepancies are used for reconstructions resulting in ensemble of augmented models.
5. When used for predictions, all the augmented models result in better skin friction, velocity and Reynolds stresses than the baseline model, including cases that were not used for training.

The work presented here is a part of larger effort of augmenting physics-based models using data. The overarching goal is to extract and reconstruct common patterns in the discrepancy fields ($\beta(\boldsymbol{\eta})$) for a class of wall-bounded turbulent flows. These data-augmented models could then be employed in predictions similar to popular models in use today. Our view is that the loss of information in Reynolds averaging is so large that there can never be a perfect turbulence model. However, an optimal model, conditioned on available data will prove to be useful for classes of problems.

VII. Acknowledgments

APS and KD were supported by DARPA under the EQUiPS project (technical monitor: Dr Fariba Fahroo). RM and PD were supported by the National Science Foundation via Grant No. 1507928 (Technical monitor: Ron Joslin).

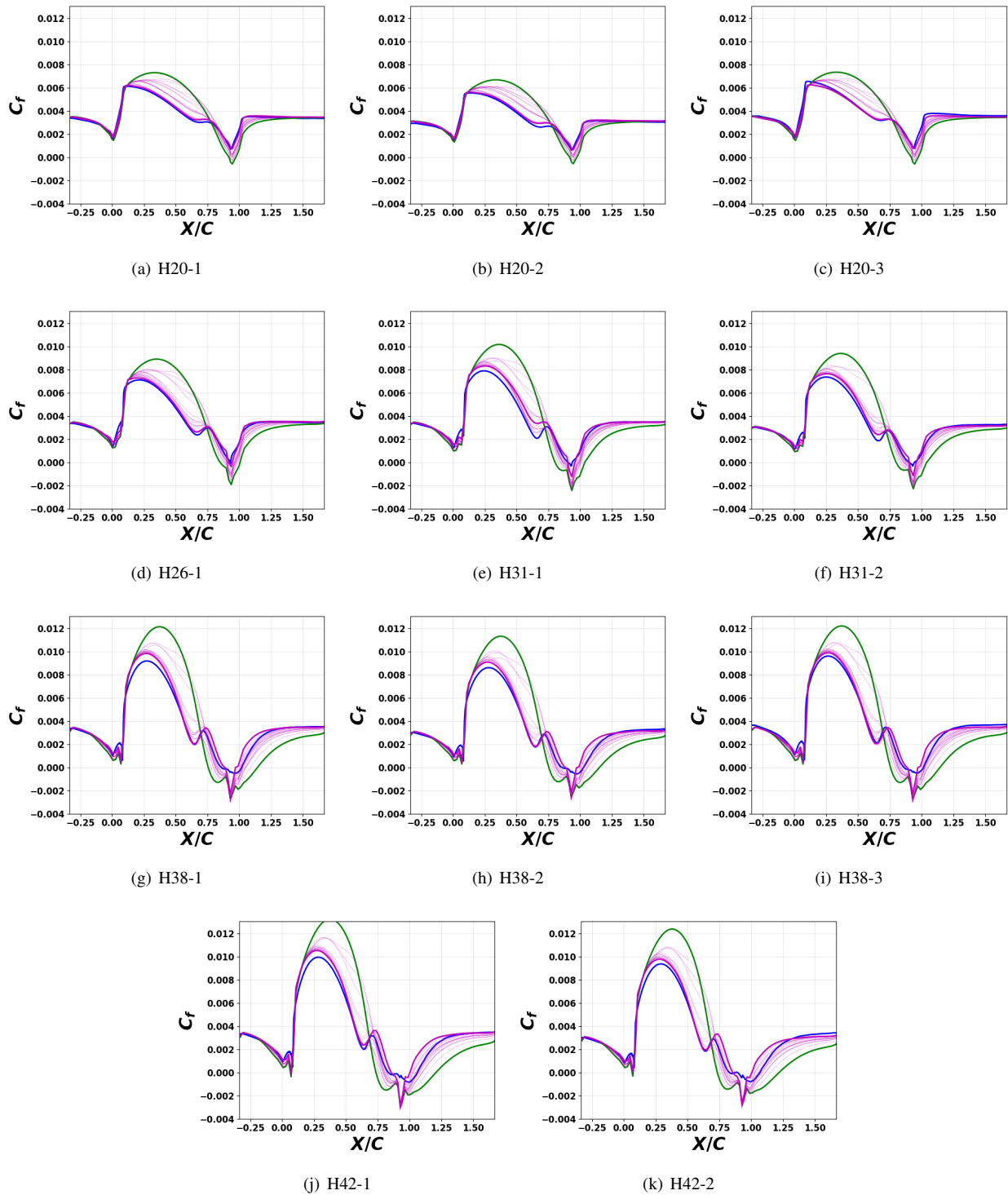


Figure 27: Skin friction prediction for all the cases used in this work. Legend: LES, base $k - \omega$, Adaboost augmented $k - \omega$ using model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

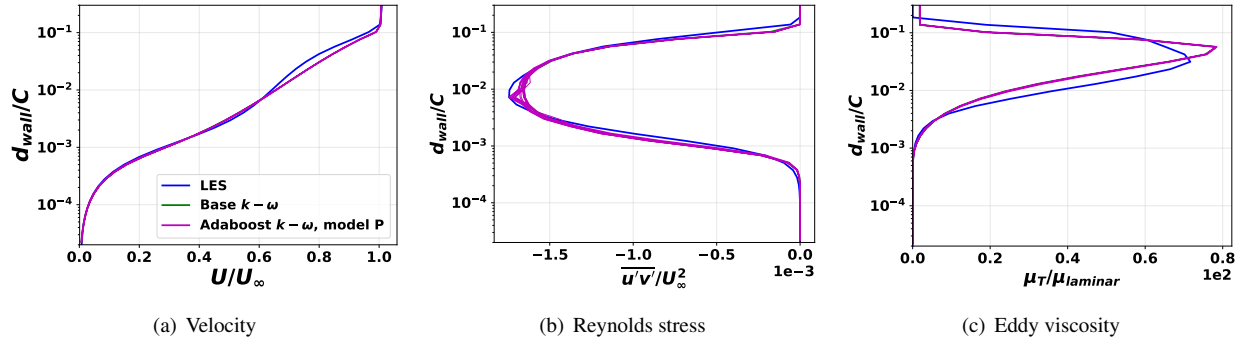


Figure 28: Flow solution at $X/C = -0.16$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

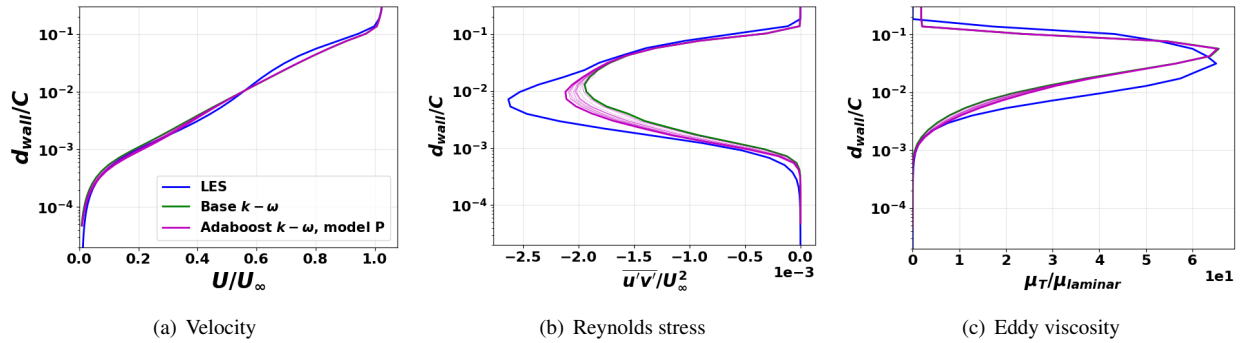


Figure 29: Flow solution at $X/C = 0$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

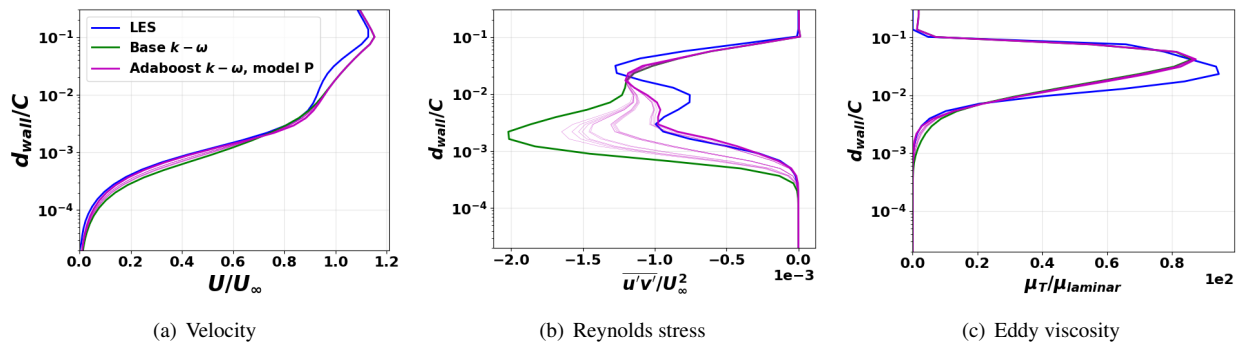


Figure 30: Flow solution at $X/C = 0.33$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

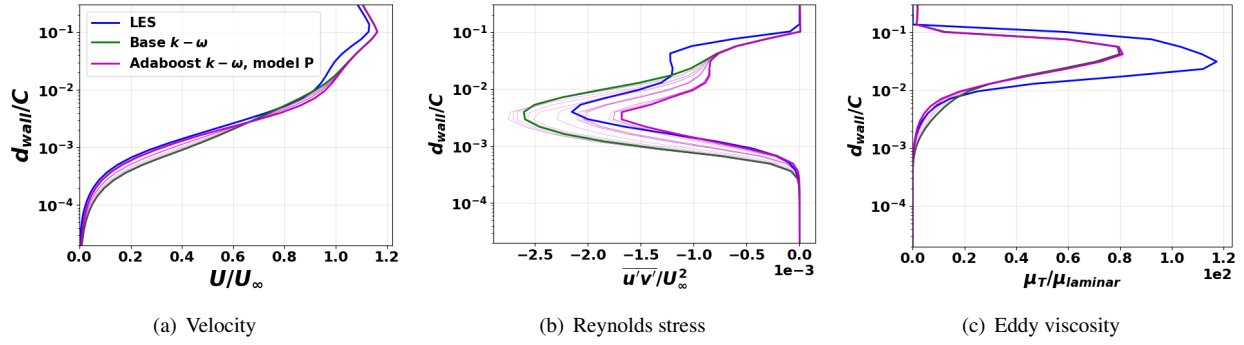


Figure 31: Flow solution at $X/C = 0.66$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

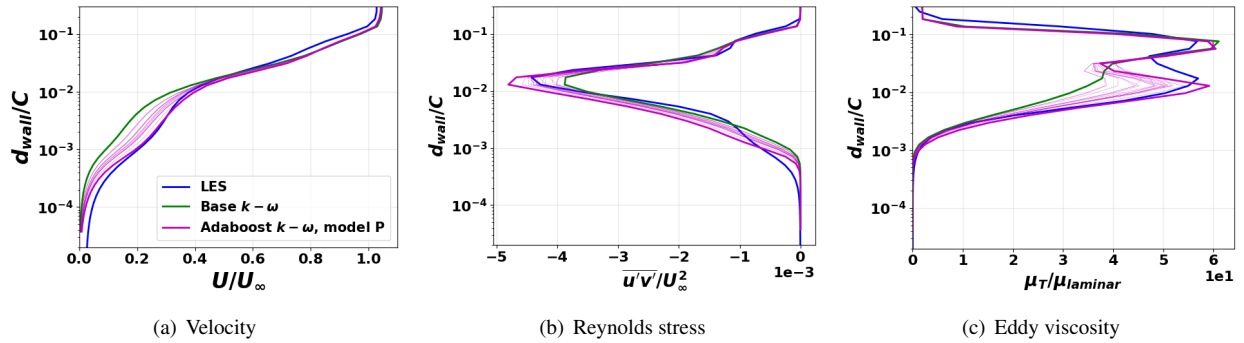


Figure 32: Flow solution at $X/C = 0.98$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

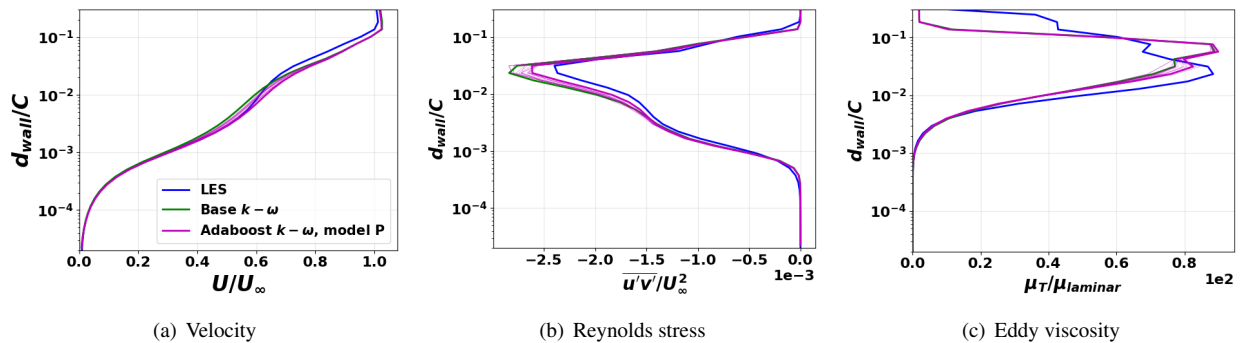


Figure 33: Flow solution at $X/C = 1.31$ using Adaboost augmented model P. Thin magenta lines show prediction using ensemble of machine learned models trained on different combinations of the inverse solutions.

Computing resources were provided by the NSF via grant MRI: Acquisition of Conflux, A Novel Platform for Data-Driven Computational Physics (Tech. Monitor: Ed Walker) and by the Extreme Science and Engineering Discovery Environment (XSEDE).

References

- ¹Tracey, B., Duraisamy, K., and Alonso, J., "Application of Supervised Learning to Quantify Uncertainties in Turbulence and Combustion Modeling," *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition Grapevine (Dallas/Ft. Worth Region), Texas*, Jan 2013.
- ²Duraisamy, K., Zhang, Z. J., and Singh, A. P., "New Approaches in Turbulence and Transition Modeling Using Data-driven Techniques," *53rd AIAA Aerospace Sciences Meeting, AIAA SciTech, Kissimmee, Florida*, Jan 2015.
- ³Singh, A. P. and Duraisamy, K., "Using field inversion to quantify functional errors in turbulence closures," *Physics of Fluids*, Vol. 28, No. 4, 2016, pp. 045110.
- ⁴Parish, E. J. and Duraisamy, K., "A paradigm for data-driven predictive modeling using field inversion and machine learning," *J. Comput. Physics*, Vol. 305, 2016, pp. 758–774.
- ⁵Xiao, H., Wu, J.-L., Wang, J.-X., Sun, R., and Roy, C., "Quantifying and Reducing Model-Form Uncertainties in Reynolds-Averaged Navier-Stokes Equations: An Open-Box, Physics-Based, Bayesian Approach," *arXiv preprint arXiv:1508.06315*, 2015.
- ⁶Weatheritt, J., *The development of data driven approaches to further turbulence closures*, Ph.D. thesis, University of Southampton, 2015.
- ⁷Ling, J., Kurzwski, A., and Templeton, J., "Reynolds averaged turbulence modelling using deep neural networks with embedded invariance," *Journal of Fluid Mechanics*, Vol. 807, 2016, pp. 155–166.
- ⁸Singh, A. P., Medida, S., and Duraisamy, K., "Machine-Learning-Augmented Predictive Modeling of Turbulent Separated Flows over Airfoils," *AIAA Journal*, Apr 2017, pp. 1–13.
- ⁹Singh, A. P., Duraisamy, K., and Zhang, Z. J., "Augmentation of Turbulence Models Using Field Inversion and Machine Learning," *AIAA SciTech Forum, Dallas, TX*, Jan 2017.
- ¹⁰Lilly, D. K., "A proposed modification of the Germano subgrid-scale closure method," *Physics of Fluids A: Fluid Dynamics*, Vol. 4, No. 3, 1992, pp. 633–635.
- ¹¹Webster, D., DeGraaff, D., and Eaton, J., "Turbulence characteristics of a boundary layer over a two-dimensional bump," *Journal of Fluid Mechanics*, Vol. 320, 1996, pp. 53–69.
- ¹²Arolla, S. K., "Inflow turbulence generation for eddy-resolving simulations of turbomachinery flows," *Journal of Fluids Engineering*, Vol. 138, No. 3, 2016, pp. 031201.
- ¹³Durbin, P. A. and Reif, B. P., *Statistical theory and modeling for turbulent flows*, John Wiley & Sons, 2011.
- ¹⁴Wilcox, D. C., "Formulation of the k-w Turbulence Model Revisited," *AIAA Journal*, Vol. 46, No. 11, Nov 2008, pp. 2823–2838.
- ¹⁵Parish, E. and Duraisamy, K., "Quantification of Turbulence Modeling Uncertainties Using Full Field Inversion," *AIAA Aviation Forum, Dallas, TX*, Jun 2015.
- ¹⁶Drucker, H., "Improving regressors using boosting techniques," *ICML*, Vol. 97, 1997, pp. 107–115.
- ¹⁷Zhang, Z., Duraisamy, K., and Gumerov, N., "Efficient Multiscale Gaussian Process Regression using Hierarchical Clustering," *arXiv preprint arXiv:1511.02258v2*, 2016.