SAND2024-05923C

Credibility Assessment of Machine Learning-based Surrogate Model Prediction on NACA 0012 Airfoil Flow

Jared Kirsch[1,2]
Nima Fathi[2]
Bill Rider[1]

05/16/2024

[1]Sandia National Laboratories
[2]Texas A&M University

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

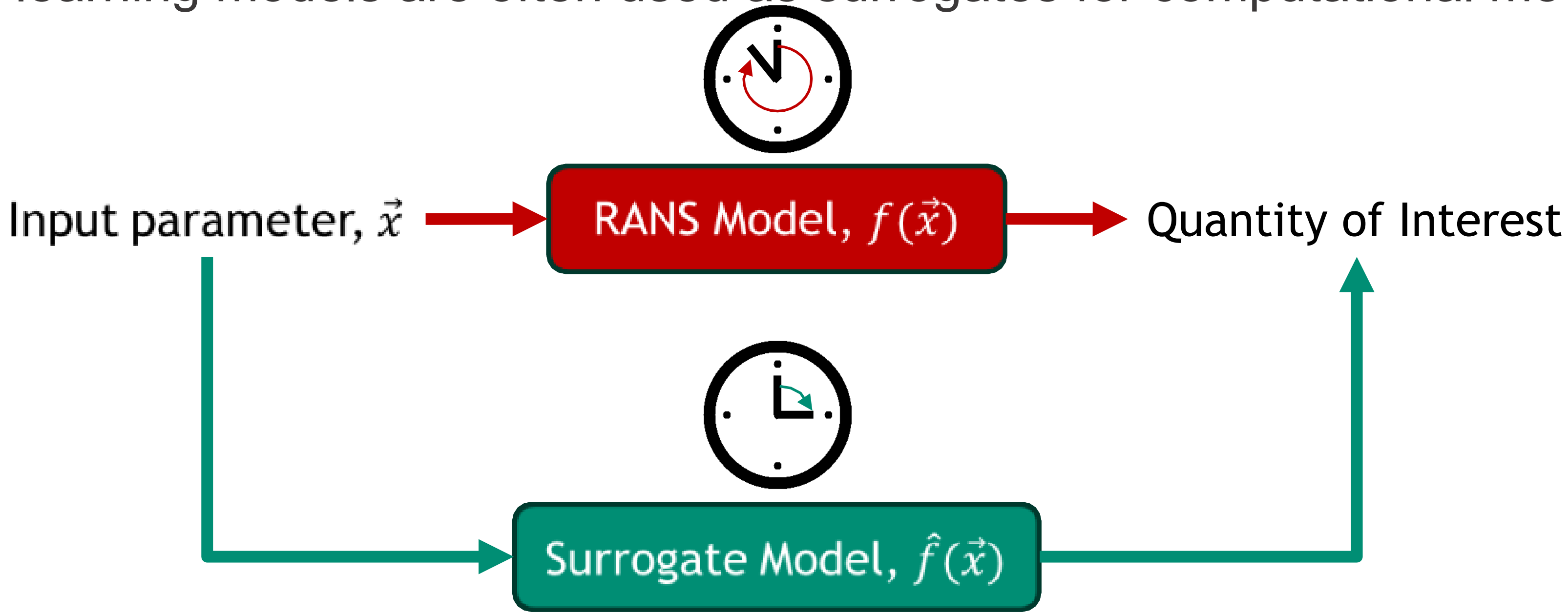
Context & Motivation

- Machine Learning (ML) is a fast-growing field with many engineering applications
- One application of ML is surrogate modeling in lieu of computational mechanics modeling for design optimization and uncertainty quantification
- Best practices for surrogate model credibility assessment are not yet established
- Several useful concepts from studies not directly involved with the VVUQ community are applicable. How do we bring these in to form a comprehensive picture?

"Artificial intelligence will create entirely new ways to do computational science: In just the last few years, artificial intelligence (AI) and machine learning (ML) have begun to transform broad swaths of commerce and society. These technologies are beginning to have major benefits for science and engineering as well, but the field is still young. AI is being used to accelerate simulations, to combine experiments with simulations, to automate workflows, to propose new hypotheses, and much more. This rapidly developing area will be a major driver of scientific progress for the foreseeable future, but only if investments are made to ensure that existing or new AI technologies are appropriately reliable and trustworthy for scientific and engineering applications." - 2024 SIAM Task Force Report: The Future of Computational Science.

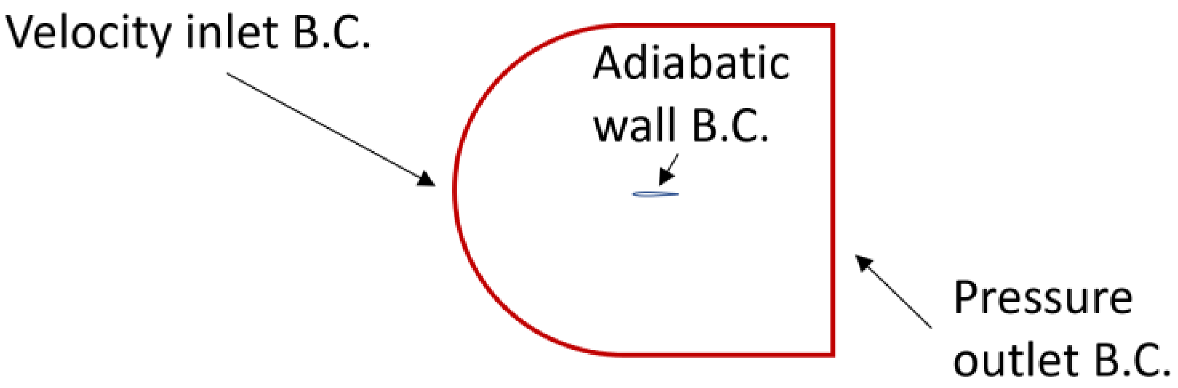
What is a Surrogate Model?

- A surrogate model is an approximate model that is used instead of a computational physics model in order to reduce the computational cost of an engineering analysis.
- Forrester: “Educated guesses as to what an engineering function might look like, based on a few points in space where we can afford to measure the function values. While these glimpses alone would not tell us much, they become very useful if we build a number of assumptions into the surrogate based on our experience of what such functions tend to look like (Forrester, 2008).”
- Machine learning models are often used as surrogates for computational mechanics models

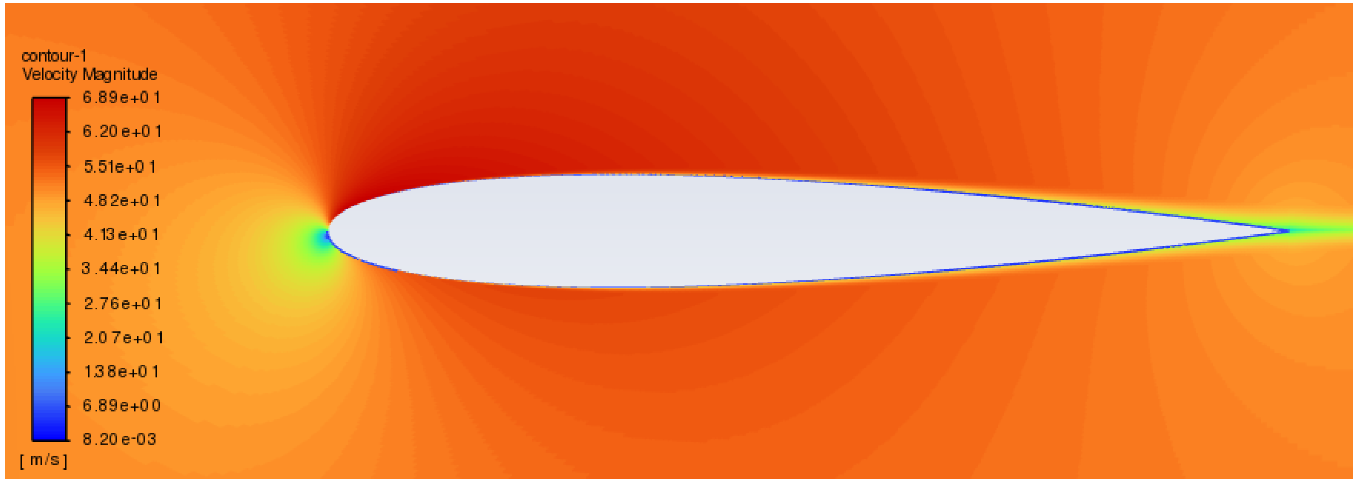


Physics Scenario

- 2D flow over NACA 0012 airfoil (Rumsey, 2022)
- Subsonic (Mach number = 0.15)
- Fully turbulent
- Separation at high angles of attack
- Reynolds number: $2 \times 10^6, 6 \times 10^6, 8.95 \times 10^6$

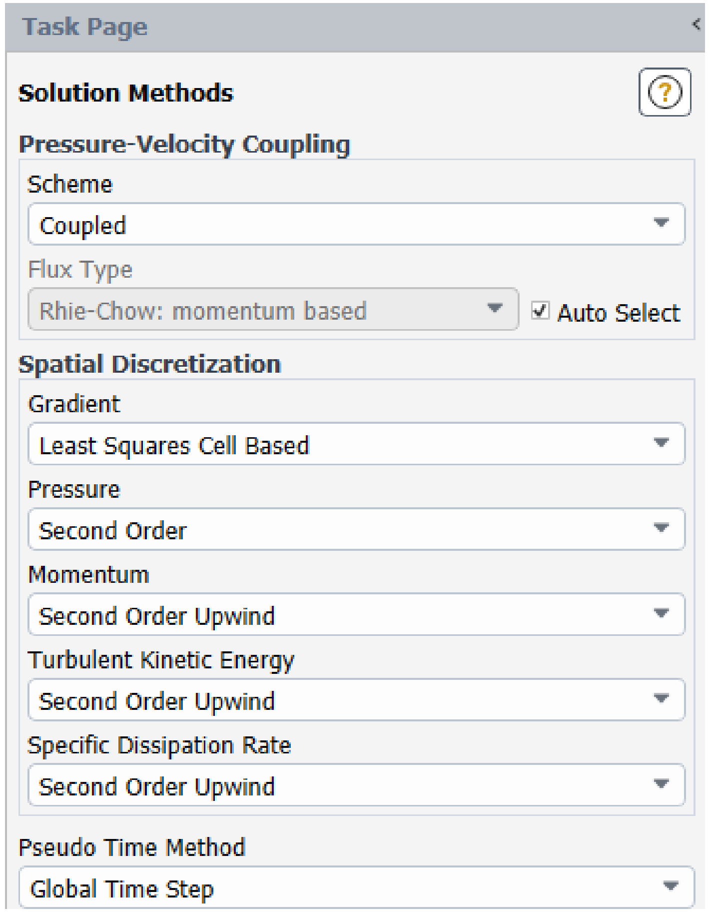


Quantity	Value
T_{ref}	300 K
p_{ref}	0 Pa
ρ_{ref}	1.177 kg/m^3
A_{ref}	1 m^2
U_{ref}	U_{inlet}
μ_{ref}	μ_{Re}



Simulations

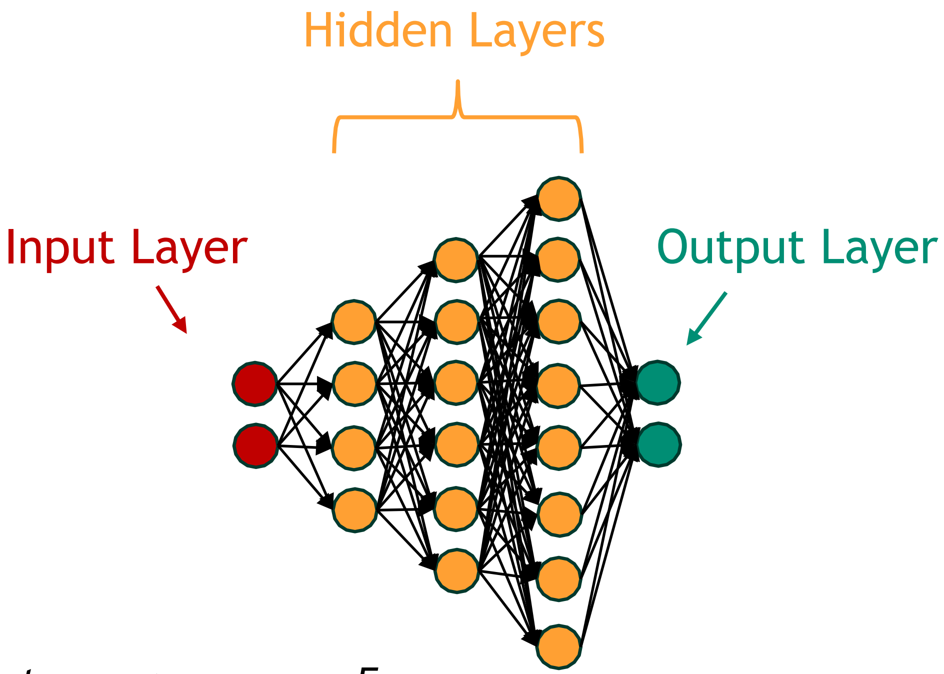
- Meshing and simulation done in Ansys
- Standard $k-\omega$ model used as turbulence model after comparison to Spallart—Allmaras on a sample case
- 2^{nd}-order discretization chosen
- Simulations run using Texas A&M's High Performance Research Computing (HPRC) Clusters
- Post-processing done using Python scripts



Case	Viscosity Updated?	Case and Journal Files Uploaded?	Simulations Run?	Data Extracted?	Data Post-processed?
Re = 2×10^6	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X
Re = 6×10^6	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X
Re = 8.95×10^6	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X	1X, 2X, 4X

Surrogate Model Description

- Surrogate model built using machine learning algorithms. Specifically, Deep Neural Network model implemented in Keras and SciKit Learn.
- Architecture:
  - Two inputs: Reynolds number and angle of attack
  - Three hidden layers with 28, 128, and 256 nodes respectively*
  - Two outputs: C_d and C_l
- Hyperparameters:
  - Kernel initializer: he_uniform
  - Activation function: Rectified Linear Unit (ReLu)
  - Optimizer: Adam
  - Loss function: mean squared error

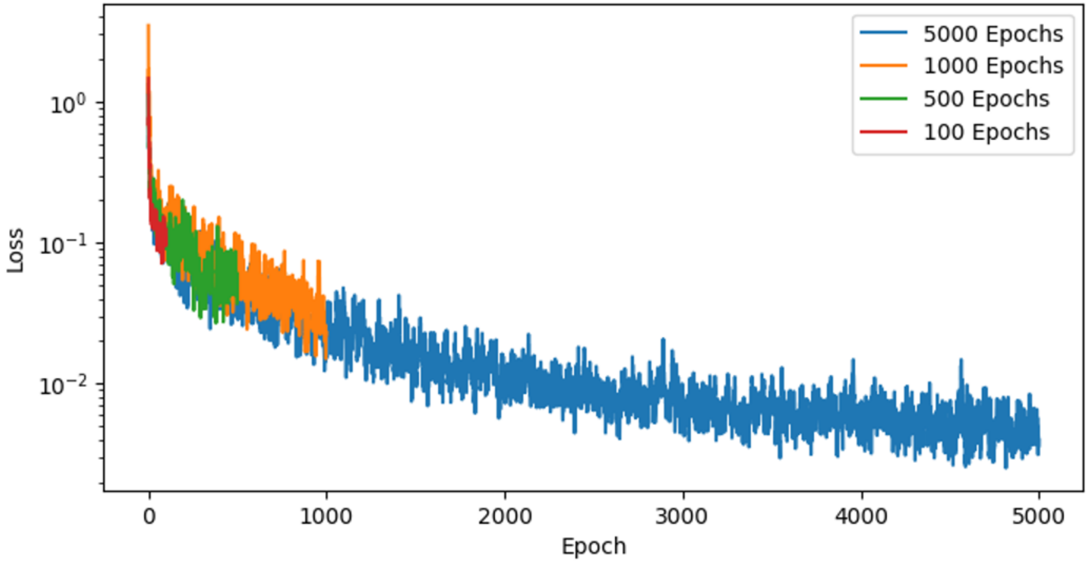


This is admittedly somewhat large. This network architecture was found to yield the highest average accuracy. For a larger dataset but otherwise similar problem, the network size would be reduced to reduce training/testing time.

Surrogate Model Training and Testing

- The data was split 90/10 between training and testing after evaluating using 80/20 and 70/30 splits
- Model is trained using 5000 epochs

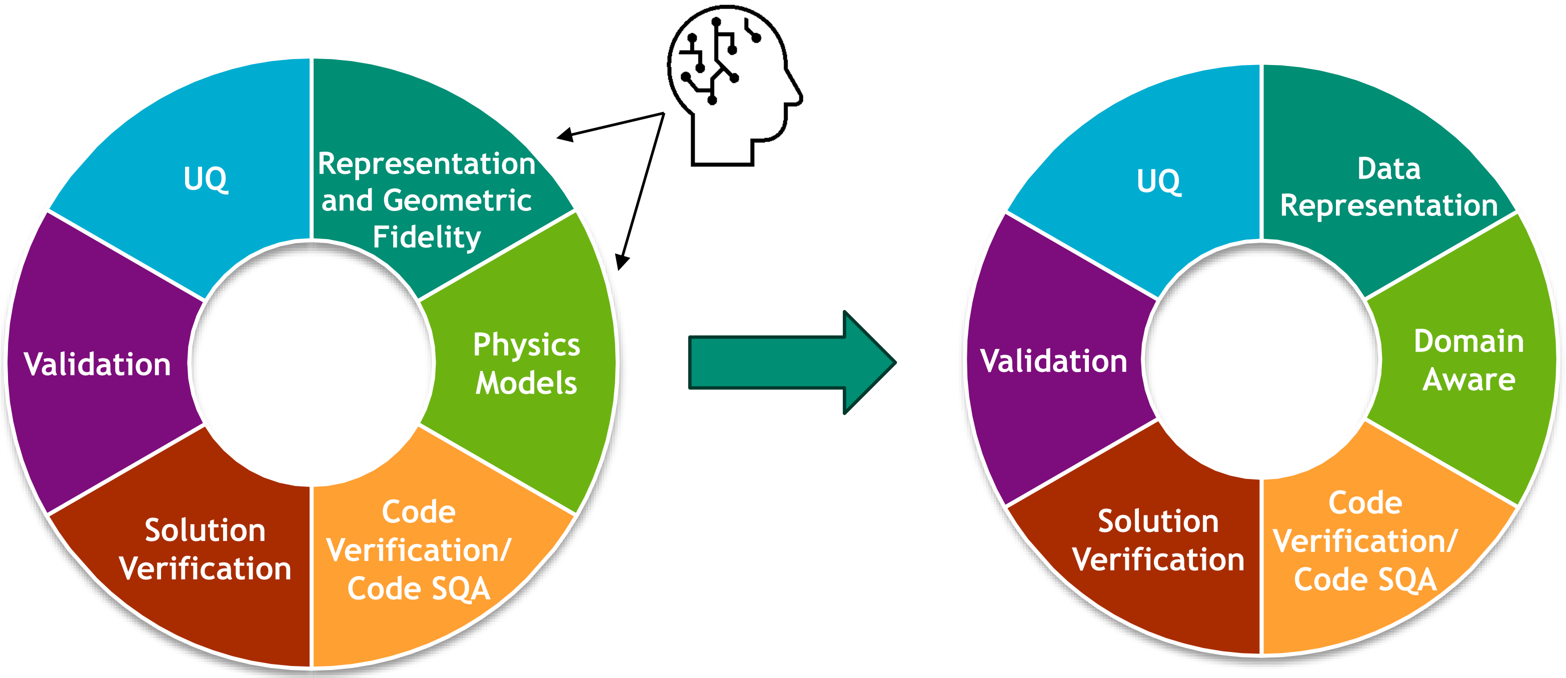
- Evaluation is done using repeated k-fold cross validation (10 splits, 3 repeats)
- Average root mean square error (RMSE) for both QoIs was on the order of 1×10^{-4}



Modified PCMM

- The Predictive Capability Maturity Model describes the necessary elements for assessing the level of maturity of computational modeling and simulation efforts (Oberkampf, 2007)
- This model was developed for use with computational physics models
- Acquesta et al. proposed a modified version for ML models, which is used in this study (Acquesta, 2022)

- Geometric fidelity → data representation (does the data provide a representative population for training/testing/validating?)
- Physics models → domain aware training (what physics phenomena need to be preserved in the model?)
- Verification is achieved differently for ML models. Explainability and interpretability play into this as well.

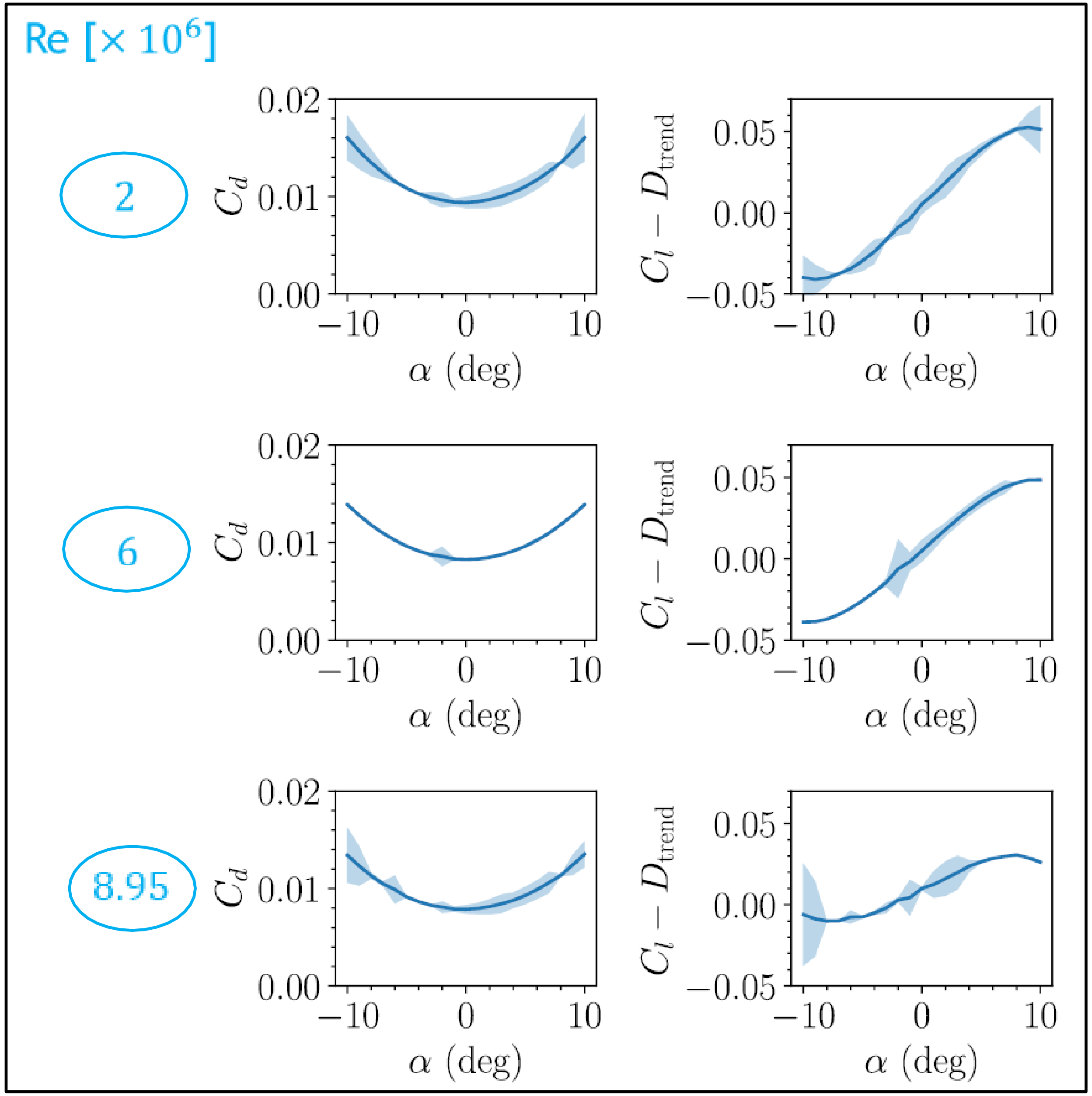


Results & Analysis

Solution Verification on Parent CFD Model

- Simulations were run on three meshes of different refinement and numerical uncertainty was quantified using the grid convergence index (GCI)
- Numerical uncertainty is generally reasonably small
- Expansion of uncertainty at the edges of the parameter space

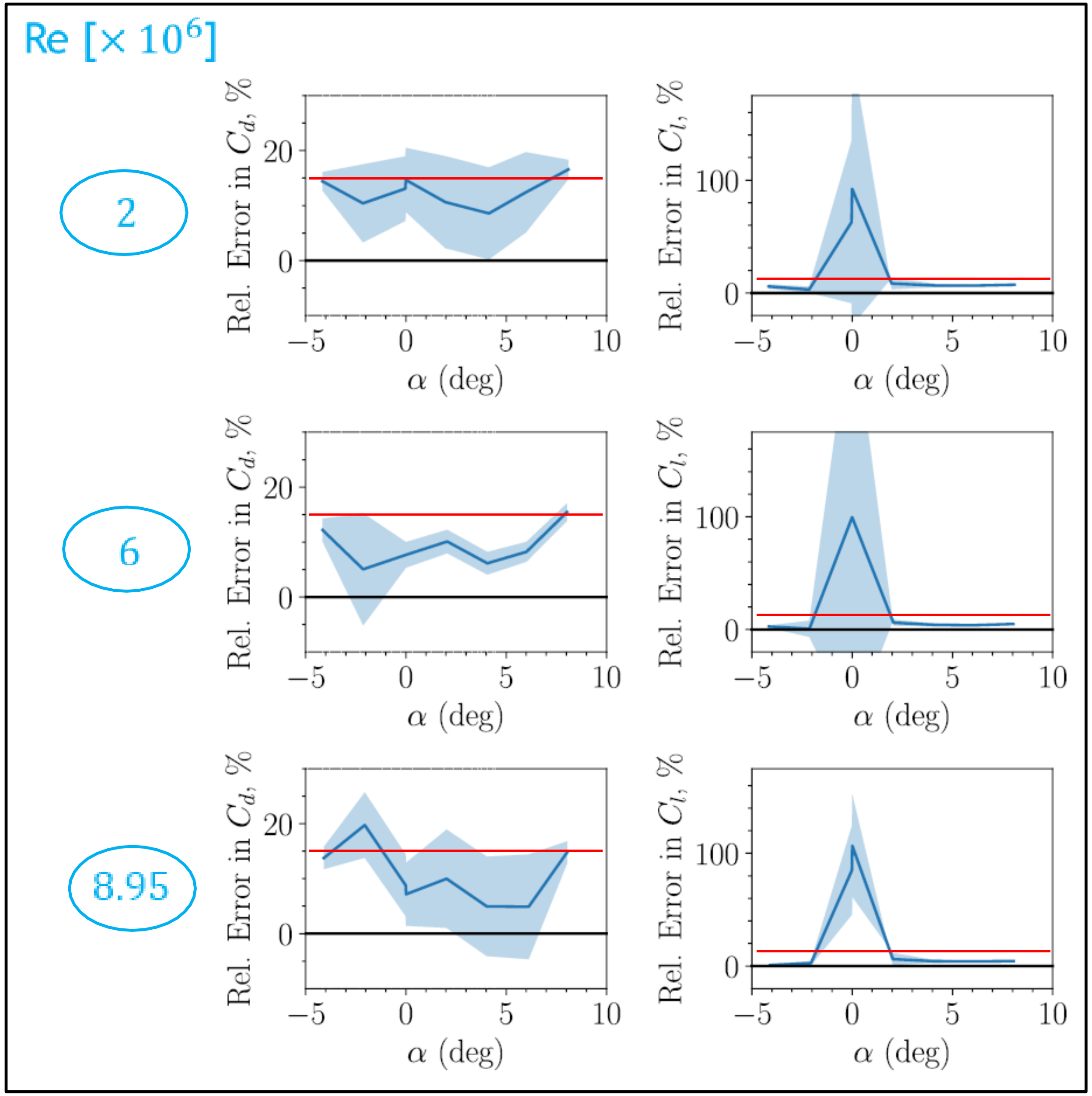
- C_l results transformed by subtracting linear fit to experimental data



Validation Analysis on Parent CFD Model

- Experimental data from Ladson, 1988
- Experimental uncertainty estimated from rough estimate of drag coefficient and normal-force coefficient

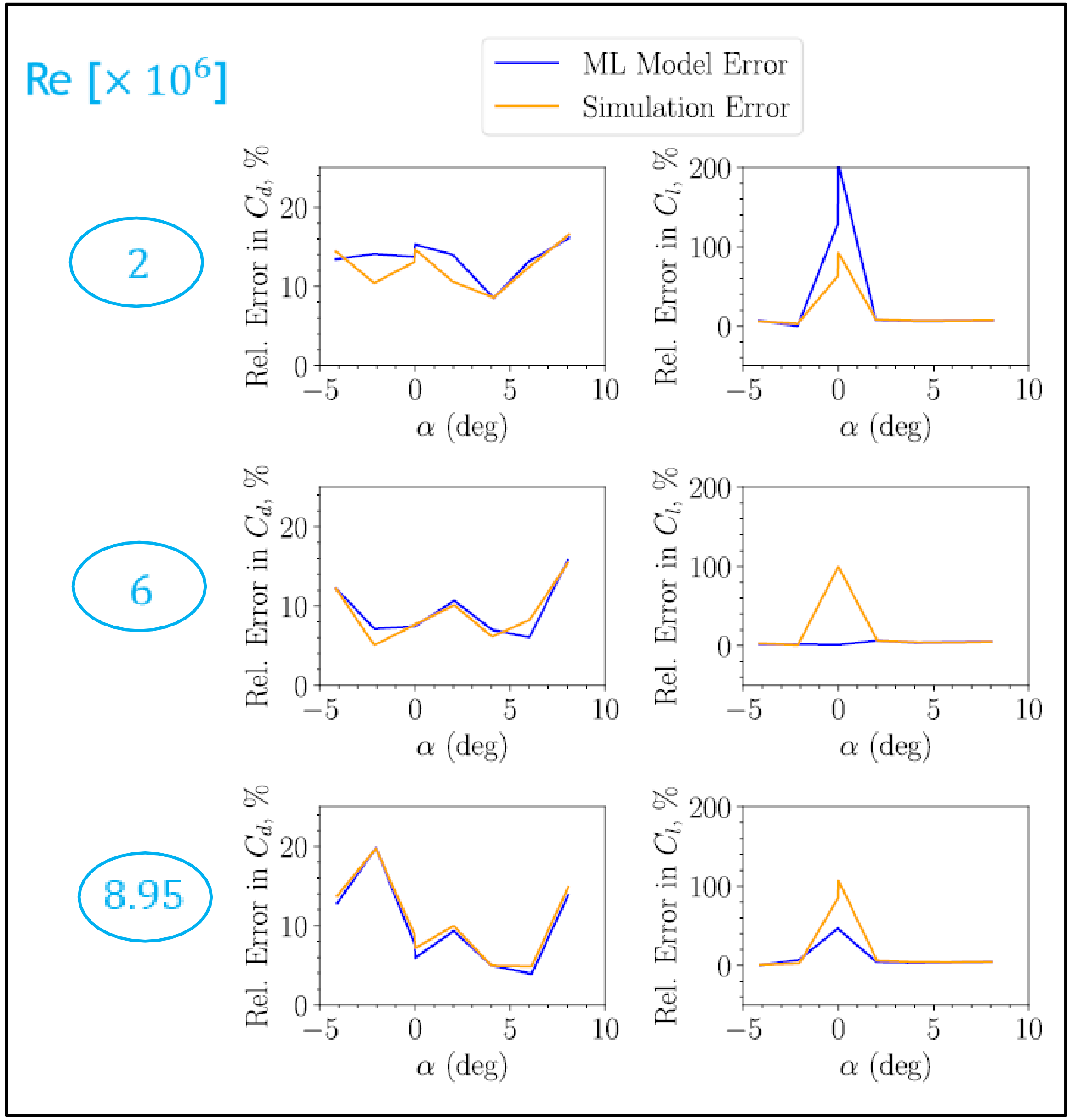
- Relative error in both quantities significant but generally less than 15%
- Much higher relative error near zero for C_l due to low magnitude of QoI there. Absolute error was not high near zero.



Validation Analysis on Surrogate Model

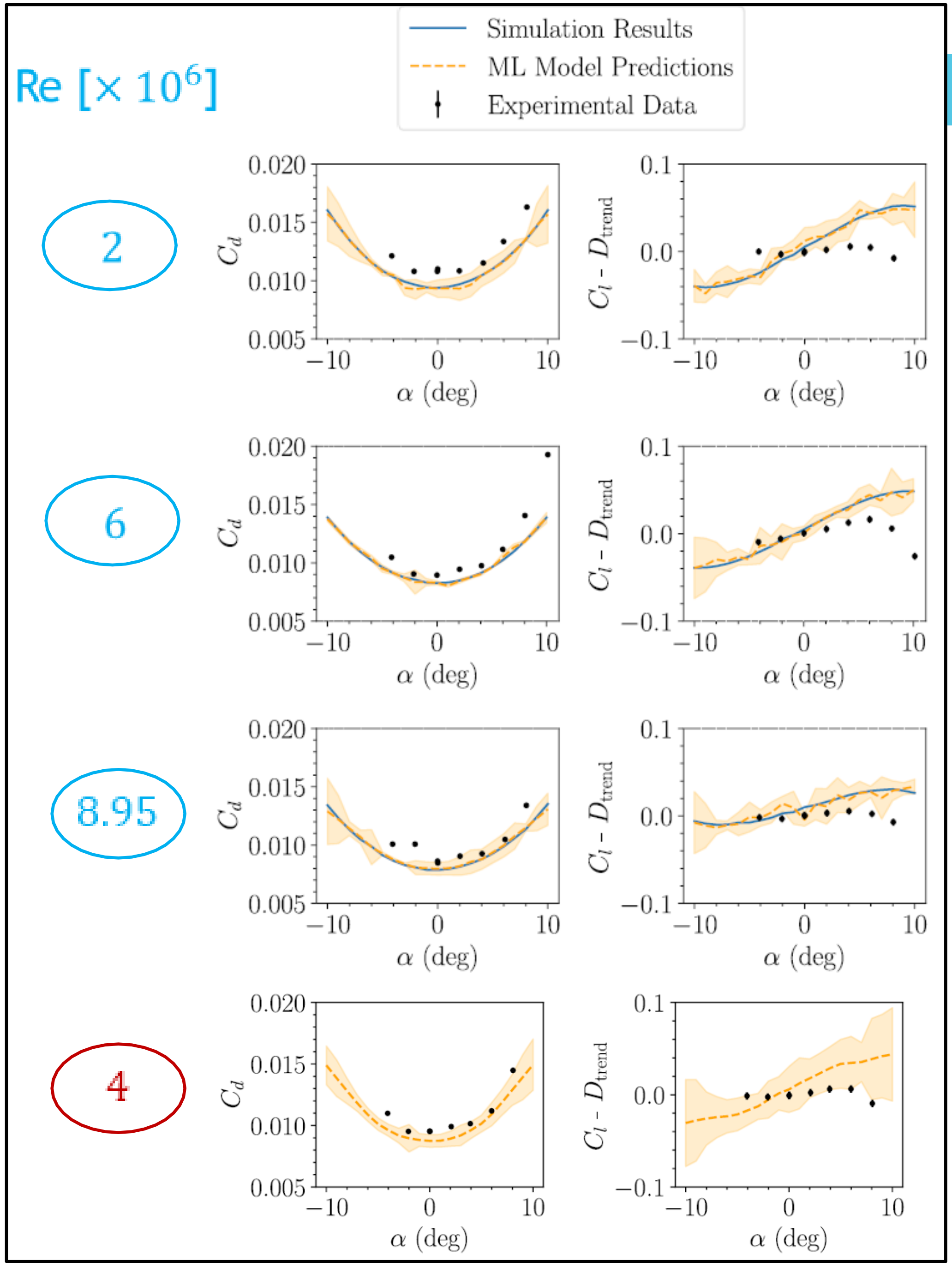
- Validation comparison error calculated for surrogate model predictions at experimental data locations
- Error trends for surrogate model very close to those of parent CFD model
- Most error in surrogate model appears to come from parent CFD model

- Run-to-run variability of surrogate model used to estimate numerical uncertainty



Summary Plots

- Surrogate model trends follow parent CFD model trends closely
- Primary source of surrogate model error is parent CFD model error
- Systematic errors/biases seen for both coefficients → potential to add bias correction to surrogate model
- Accuracy for Re = 4×10^6 similar to that of other Reynolds numbers



Explainability and Datasheet

- Datasheet created for training/testing dataset using template and guidance in “Datasheets for Datasets” (Bebru, 2021)
- Defines pedigree of data
- Enhances explainability and reduces reliance on SME/developer for insight into data
- Makes limitations, sensitivities, accessibility, purpose, and strengths of dataset apparent

Motivation	Composition
For what purpose was the dataset created? The dataset was created for training and testing of the surrogate model. Comparable datasets for this airfoil and flow scenario have been created for model validation.	**What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** Coefficients of lift and drag of the airfoil at flow conditions described by Reynolds number and angle of attack.
Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)? The primary researcher on this project (Jared Kirsch).	**How many instances are there in total (of each type, if appropriate)?** There are 310 instances of each QoI, in order to inform solution verification. 63 instances are used in surrogate model training and evaluation.
Who funded the creation of the dataset? Sandia National Laboratories, through contract with Texas A&M University.	**Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** The training and testing data is a sample of the larger set of 93 data points. The dataset was reduced to eliminate data at high and low angles of attack, where flow characteristics change significantly.
Any other comments? No.	**What data does each instance consist of?** Coefficients of lift and drag of the airfoil at flow conditions described by Reynolds number and angle of attack.
	Is there a label or target associated with each instance? Yes, the coefficients of lift and drag.
	Is any information missing from individual instances? No.

continued • • •

Conclusions

- An ML-based surrogate model was used to predict coefficients of lift and drag for a NACA 0012 airfoil at various angles of attack
- Modified PCMM framework was used to assess credibility
- Uncertainty quantification was done, excluding input uncertainty
- Validation analysis showed moderate error with respect to experimental data
- Most error was inherited from parent computational physics model
- Datasheet created to enhance explainability of dataset used in training/testing of surrogate model

- Potential extensions of work include bias correction in surrogate model, increased UQ efforts, and evaluation of surrogate model accuracy at more points

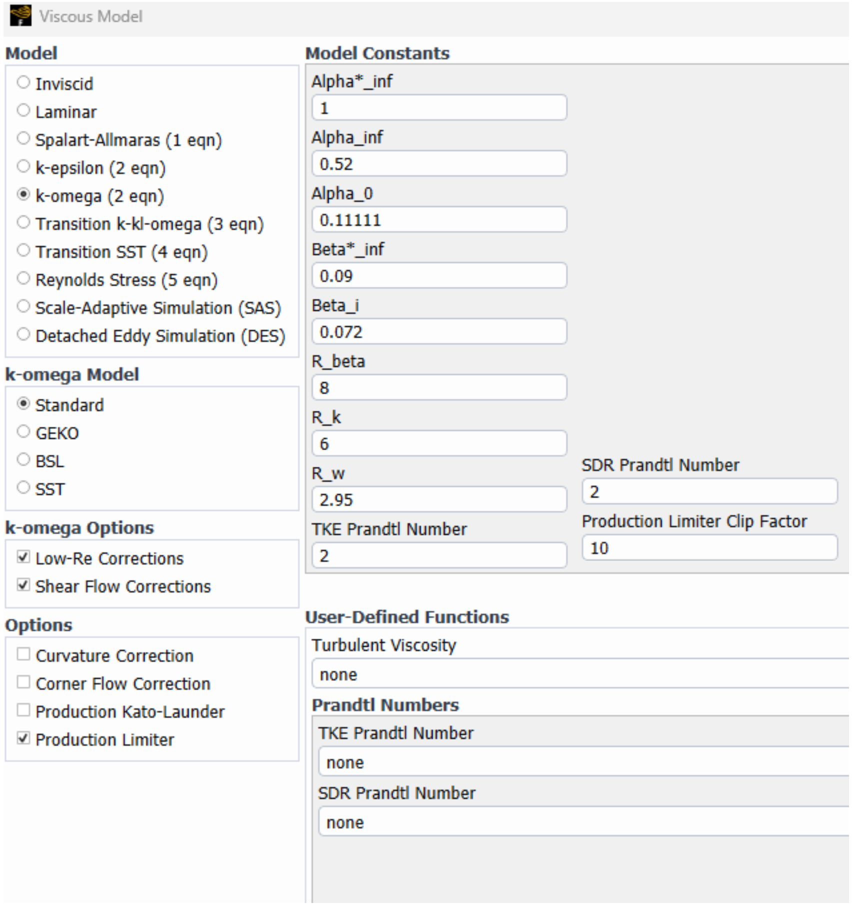
References

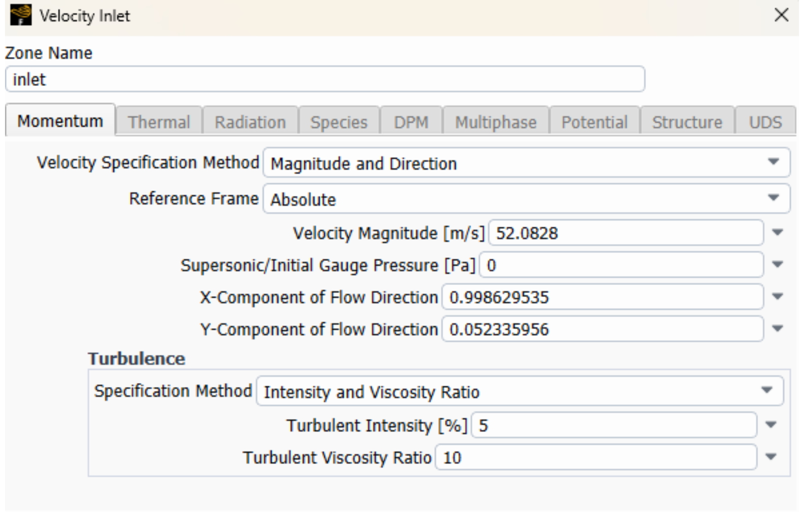
- “SIAM Task Force Report: The Future of Computational Science,” Society for Industrial and Applied Mathematics, 2024.
- A. I. J. Forrester, A. Sóbester, and A. J. Keane, “Engineering Design via Surrogate Modeling,” University of Southampton, UK: Wiley, 2008.
- Rumsey, C., “2DN00: 2D NACA 0012 Airfoil Validation Case,” NASA Turbulence Modeling Resource, 2022.
- W.L. Oberkampf, M. Pilch, T. G. Trucano, “Predictive Capability Maturity Model for Computational Modeling and Simulation,” Sandia Report No. SAND2007-5948, 2007.
- E. Acquesta, “Adapting Verification and Validation Principles to a Credibility Process for Scientific Machine Learning,” SIAM Conference on Uncertainty Quantification, 2022.
- C. L. Ladson, “Effects of Independent Variation of Mach and Reynolds Numbers on the Low-Speed Aerodynamic Characteristics of the NACA 0012 Airfoil Section,” NASA Technical Memorandum 4074, 1988.
- T. Bebru, J. Morgenstern, B. Vecchione, J.W. Vaughan, H.Wallach, H. Daume, and K. Crawford, “Datasheets for datasets,” Communications of the ACM, Dec. 2021.

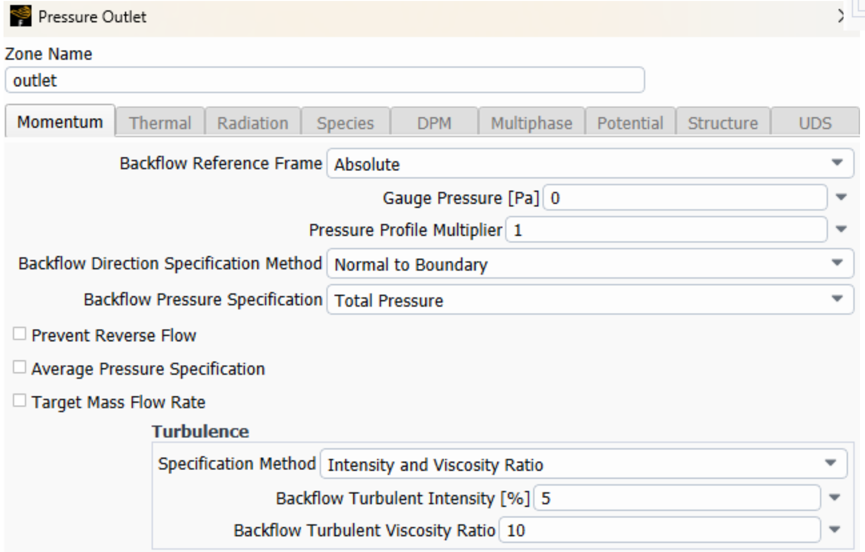
Backup Slides

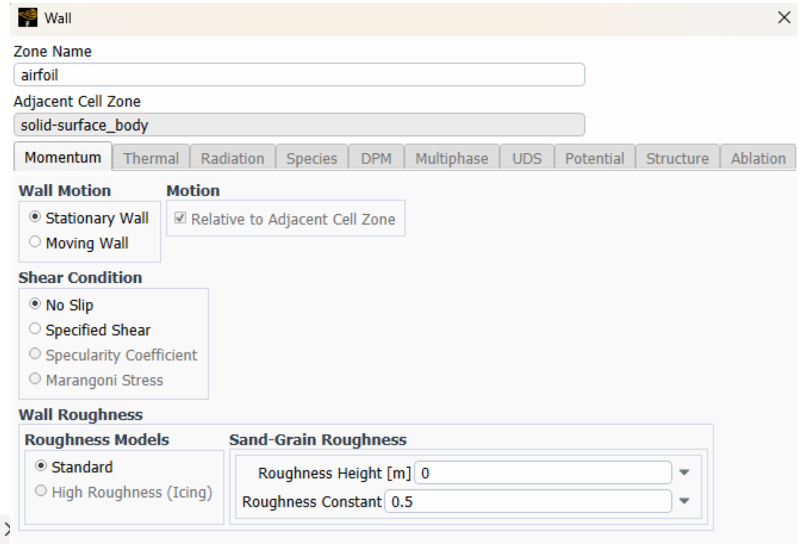
Backup Slide – Simulation Settings

- The









Ansys Fluent Standard vs. SST $k - \omega$ Models

"Both models have similar forms, with transport equations for k and ω. The major ways in which the SST model differs from the standard model are as follows:

- Gradual change from the standard $k - \omega$ model in the inner region of the boundary layer to a high-Reynolds-number version of the $k - \epsilon$ model in the outer part of the boundary layer
- Modified turbulent viscosity formulation to account for the transport effects of the principal turbulent shear stress"

Standard: https://www.afs.enea.it/project/neptunius/docs/fluent/html/th/node66.htm
SST: https://www.afs.enea.it/project/neptunius/docs/fluent/html/th/node67.htm

Backup Slide – Modifying the Reynolds Number

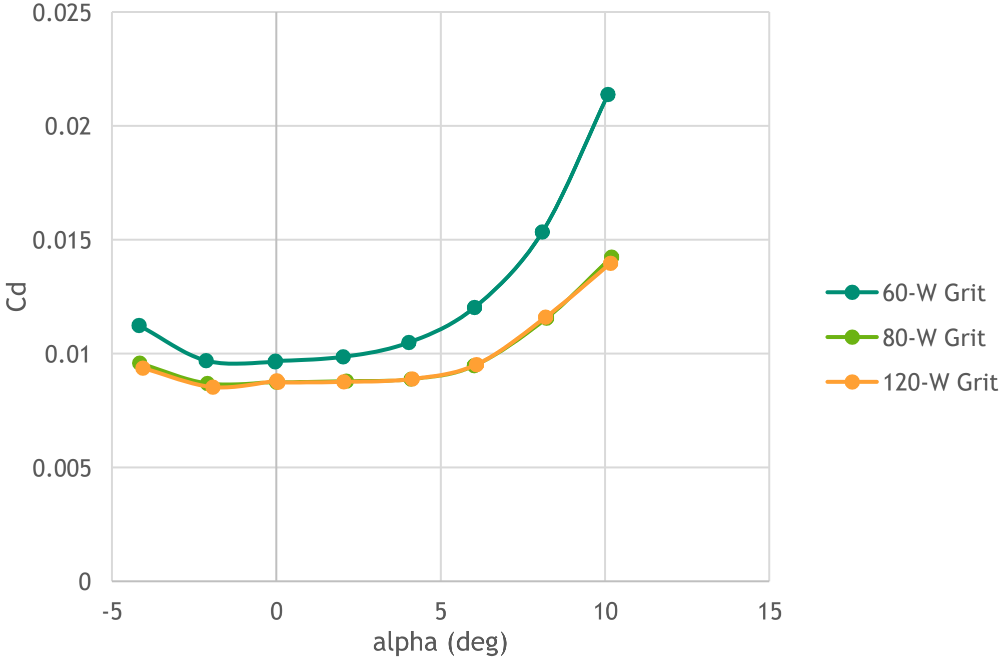
From the experimental data, we know the Mach and Reynolds number.
We assume T = 300 K and corresponding γ, R. ρ comes from engineering toolbox. Then we can calculate

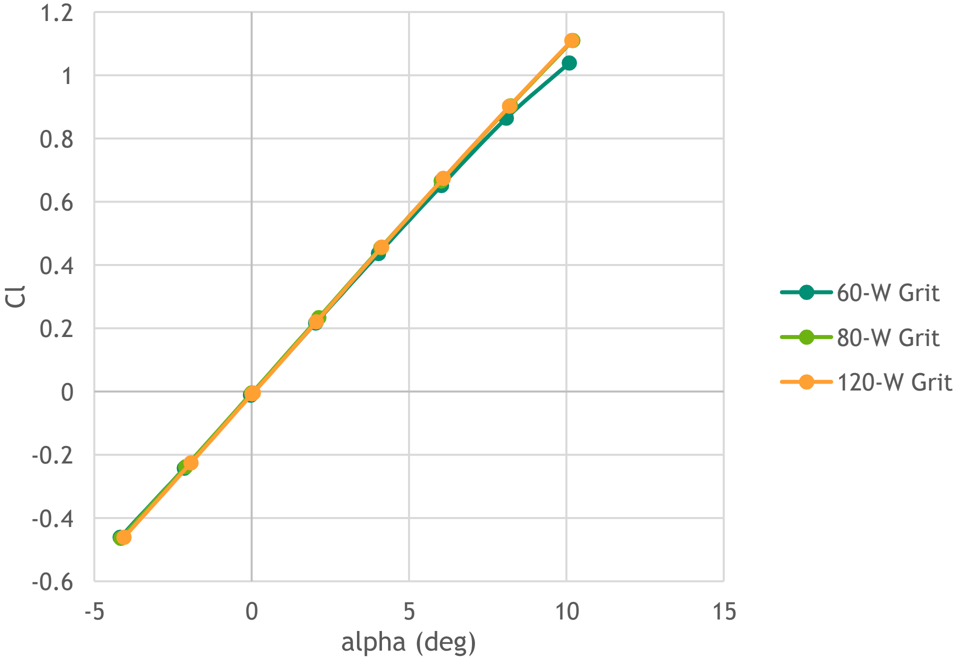
$$V = velocity\ magnitude = M * \sqrt{\gamma RT} = M * \sqrt{(1.4)\left(287.05\frac{kJ}{kg}\cdot K\right)(300\ K)}$$

$$\mu = dynamic\ viscosity = \frac{\rho VD}{Re}$$

Investigating the Effects of Using Different Grit to Trip in Ladson Report

- Uses 60-W, 80-W, and 120-W (possibly others as well) to trip flow
- Do these different grits cause an appreciable difference in the QoIs?
- Investigated for the three grits above, and plots are shown below
- Very good agreement for 80-W and 120-W grit. Significant disagreement for 60-W grit. 60-W grit corresponds to a wrap-around configuration.
- In the report, it says that the 60-W grit produced a large decrease in max. Cl and a large increase in min. Cd. This is consistent with the plots below.
- “The use of wrap-around grit produces a 20- to 30-percent decrease in max. lift-to-drag ratio when compared with that produced for the 0.05c transition location.”





Mesh Sizes

Growth rate: 1.2

Inflation option: smooth transition

Transition ratio: 0.272

Finest two meshes have y+ less than 1

4X size:
- Nodes: 3,844,800
- Elements: 3,840,000

2X size:
- Nodes: 962,400
- Elements: 960,000

1X size:
- Nodes: 241,200
- Elements: 240,000

0.5X size:
- Nodes: 60,600
- Elements: 60,000

Data Scaling

Divided Reynolds number by 1×10^6

Multiplied C_d by 100

Converted back to actual values in post-processing

Tried automated scaling, but no significant improvement over manual scaling seen

DNN Hyperparameters

Adam

Adam class

```
keras.optimizers.Adam(
    learning_rate=0.001,
    beta_1=0.9,
    beta_2=0.999,
    epsilon=1e-07,
    amsgrad=False,
    weight_decay=None,
    clipnorm=None,
    clipvalue=None,
    global_clipnorm=None,
    use_ema=False,
    ema_momentum=0.99,
    ema_overwrite_frequency=None,
    loss_scale_factor=None,
    gradient_accumulation_steps=None,
    name="adam",
    **kwargs
)
```

ReLU layer

ReLU class

[source]

```
keras.layers.ReLU(max_value=None, negative_slope=0.0, threshold=0.0, **kwargs)
```

```
def get_model(n_inputs, n_outputs):
    model = Sequential()
    model.add(Dense(28, input_dim=n_inputs, kernel_initializer='he_uniform', activation='relu'))
    model.add(Dense(128, activation='relu'))
    model.add(Dense(256, activation='relu'))
#   model.add(Dense(256, activation='relu'))
    model.add(Dense(n_outputs, kernel_initializer='he_uniform'))
    model.compile(optimizer='adam', loss='mean_squared_error', metrics=root_mean_squared_error)
    return model
```

Optimizing the DNN

Training Time vs. Network Size

5000 epochs still, network has 28, 56, 112 neurons in the 3 layers

```
Max(Cl_pd):   3.3405164104675644
Mean(Cl_pd):   1.3706956042615057
Max(Cd_pd):   3.9826643093702234
Mean(Cd_pd):   1.8428690758694888
```

```
Max(Cl_pd):   22.99418098038267
Mean(Cl_pd):   6.364033205022988
Max(Cd_pd):   5.763473969083305
Mean(Cd_pd):   2.8063858466954947
```

```
Max(Cl_pd):   181.29380279227897
Mean(Cl_pd):   37.82371000515967
Max(Cd_pd):   4.965601754040919
Mean(Cd_pd):   2.181177276139553
```

Training Time	Total (Training + Prediction) Time
13.612683899933472	13.678318799939007
16.71096729999408	16.775100099970587
19.806038399925455	19.87224779999815
13.281809599953704	13.342205699998885
14.011337500065565	14.076991100097075

Network with 28, 128, 256

Training Time	Total (Training + Prediction) Time
18.734906499972567	18.80263019993435
16.201080900034867	16.263148800004274
15.208082800032571	15.267421799944714
17.521413500071503	17.58225279999897
16.179354000021704	16.248251299955882

Estimating Uncertainty from Run-to-run Variability of Surrogate Model

Ran DNN model 100 times, which yielded 3 – 7 instances for each angle of attack.

Used average value as nominal

Used max. deviations to form uncertainty (conservative vs. using 2σ). This was done largely because the number of evaluations at each angle of attack was relatively small.

Domain Awareness

Separation occurs at large angles of attack – simulated values kept to $\alpha \in [-10°, 10°]$

Fully turbulent flow – no transition