



The Effect of Training Data on Predicting Turbulent Flow Through A Linear Cascade Using Physics-Informed Neural Networks

GT2025-153885

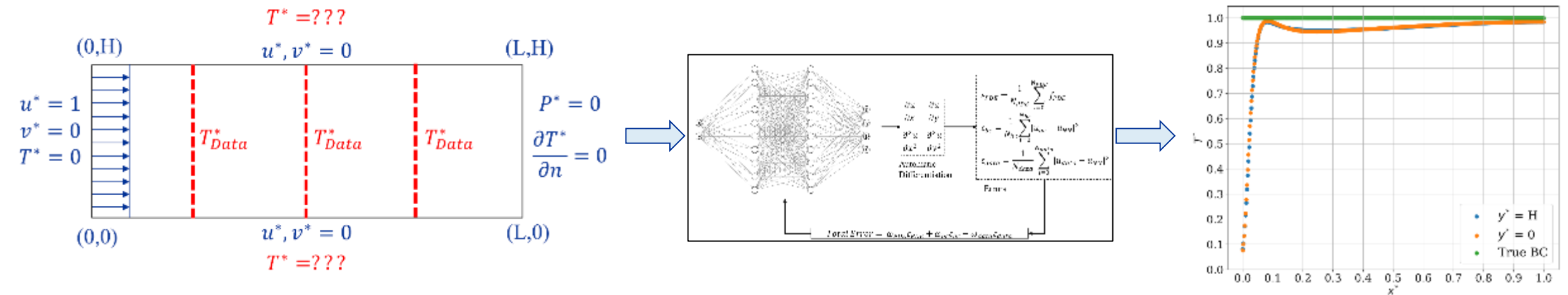
Ezra O. McNichols^{*1}, Jeffrey P. Bons²

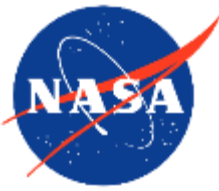
¹ NASA Glenn Research Center, Cleveland, Ohio 44135

² The Ohio State University, Columbus, Ohio 43210

Background

- What is a Physics-Informed Neural Network (PINN)?
 - PINN is a neural network that embed the governing physical laws into the training process
 - Instead of learning from data alone, PDE residuals are minimized at every point, enforcing the governing equations
 - Ideal for problems with sparse, incomplete, or noisy data, especially in applications like inverse problems or flow field reconstruction where full-field measurements are unavailable





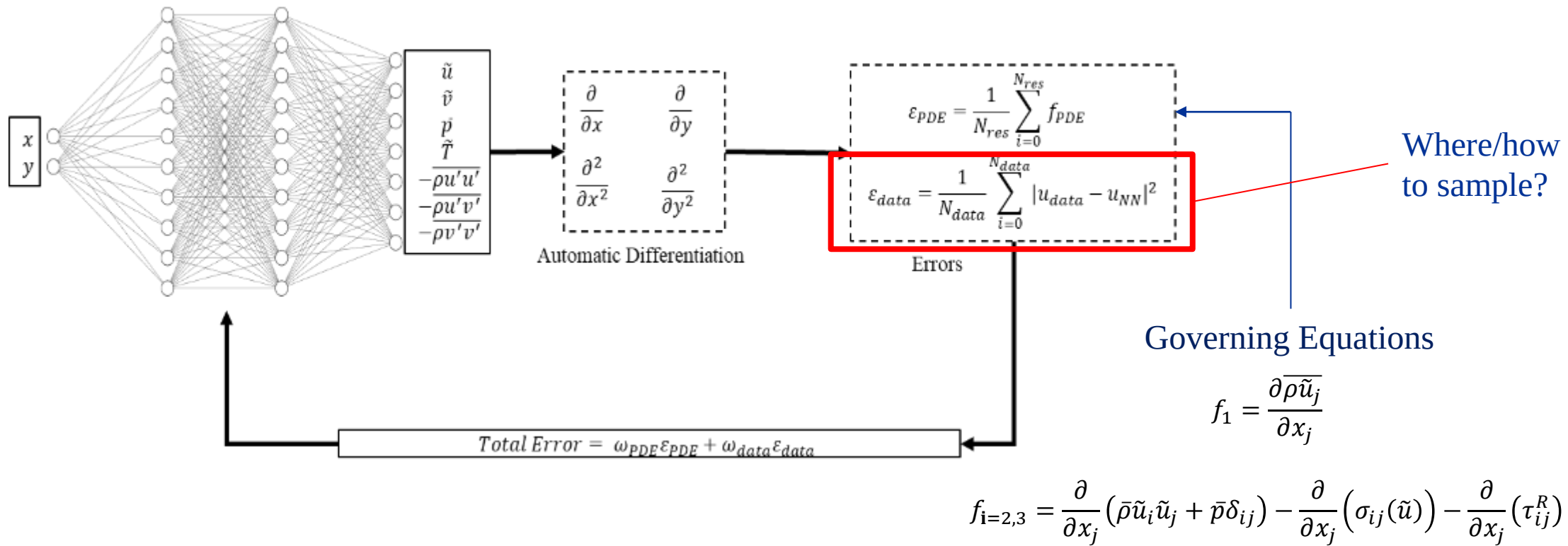
Motivation

- Wakes exhibit complex, anisotropic flow features that are difficult to resolve
- Experimental data are typically sparse and localized, hindering full-field comparisons
- PINNs provide a method to reconstruct critical flow properties from limited measurements by enforcing governing equations
- Primary End Use – Data assimilation for experimental methods to allow for field comparisons with CFD predictions

Questions to be studied:

- How does the placement of training data affect the accuracy and stability of PINNs in underdetermined flow regions?
- What spatial sampling strategies are most effective for capturing complex flow behavior – uniform, random, or physics-guided?

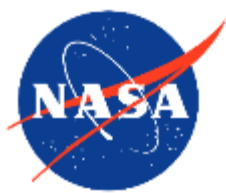
PINN Overview



Notes:

- This PINN structure does not include transport equations or models for Reynolds stresses
 - Underdetermined (4 equations, 7 unknowns)
- PINN was provided data for every variable at each training point
- Periodicity was encoded as a hard constraint through an input transform
- Output transforms were applied to: p , T , $\overline{\rho u' u'}$, $\overline{\rho v' v'}$ to ensure these were always positive during training

$$f_4 = \frac{\partial}{\partial x_j} \left(\tilde{u}_j \left(\overline{\rho C_v \tilde{T}} + \frac{1}{2} \overline{\rho \tilde{u}_i \tilde{u}_i} + \overline{\rho k} + \tilde{p} \right) \right) - \frac{\partial}{\partial x_j} (\tilde{u}_i \sigma_{ij}(\tilde{u})) + \tilde{u}_i \tau_{ij}^R - \frac{\partial}{\partial x_j} (\tilde{q}_j - \tilde{q}_j^R)$$



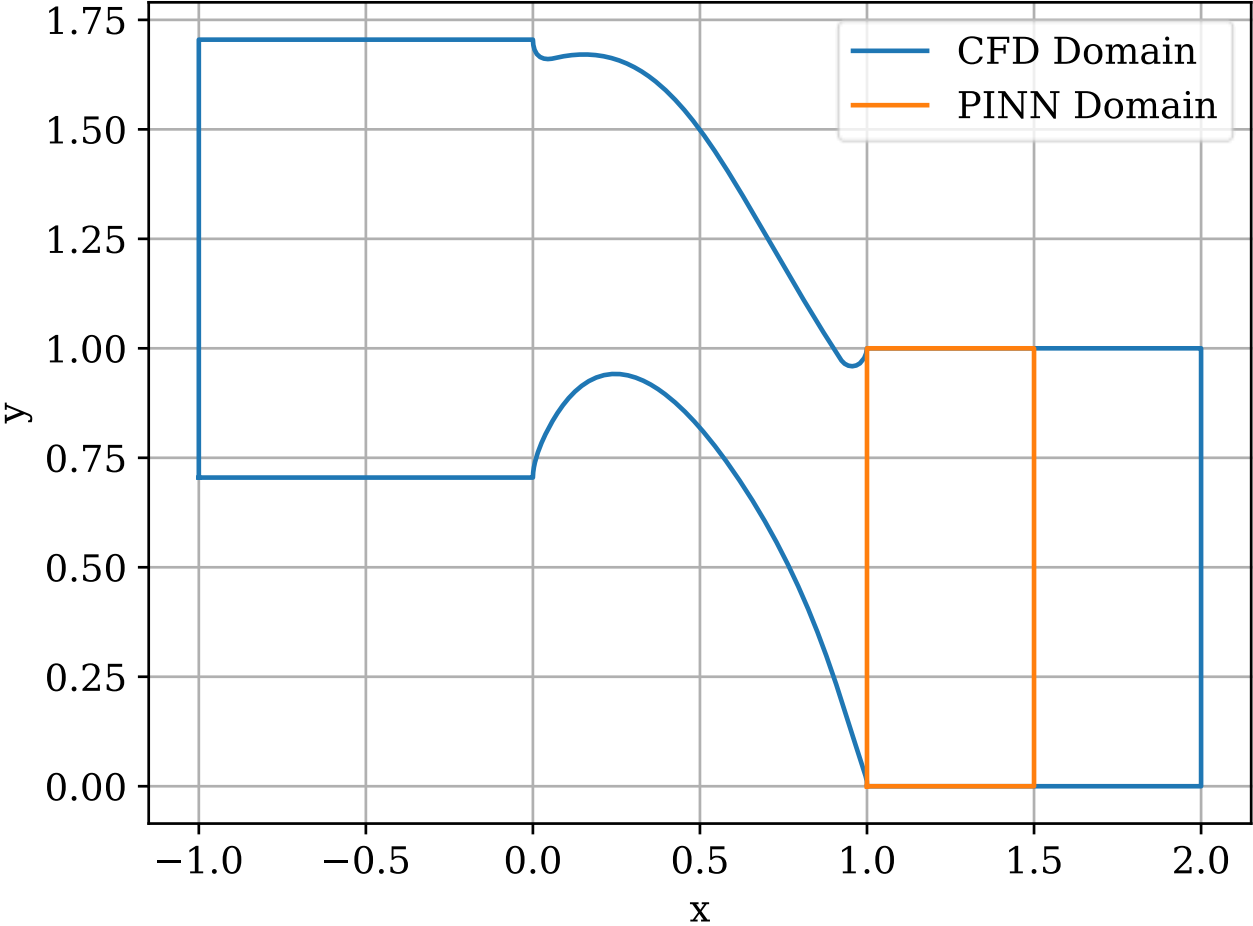
CFD Case for Training Data Generation

Table 1: CFD Boundary conditions from [1]

Variable	Value [1]
$p_{t,in}$ [Pa]	74,480.76
α_{in} [°]	38.8
Tu [%]	13
Turbulent Length Scale [m]	0.0026
$T_{t,in}$ [K]	295.6
p_{out} [Pa]	51,722.75
$M_{is,out}$	0.741
$Re_{Cx,out}$	1.24e6

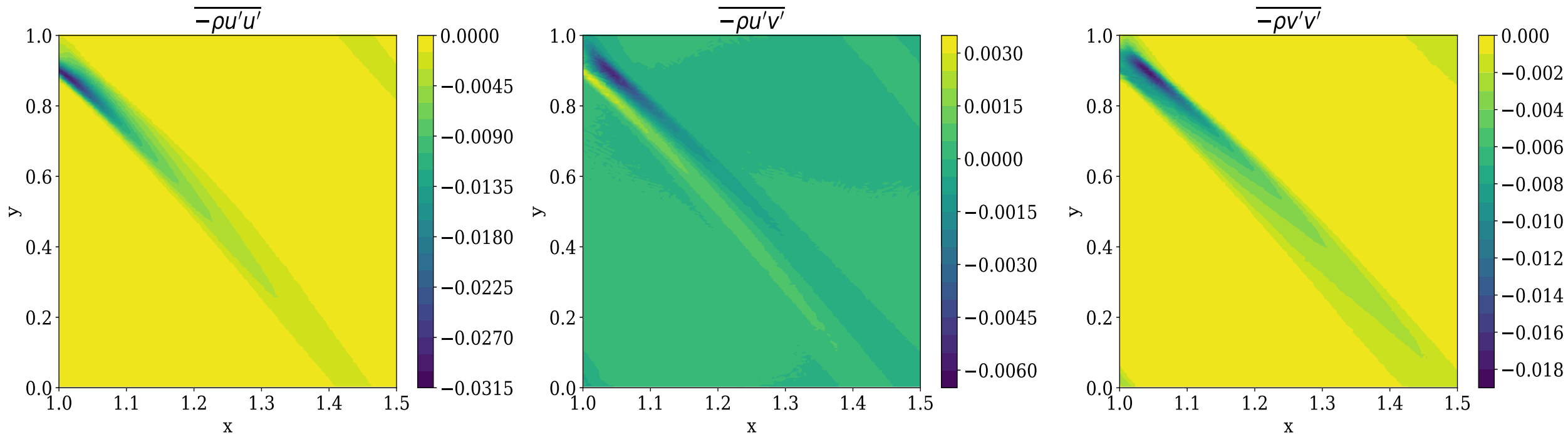
Table 2: CFD case details

Variable	Value
Turbulence Model	k- ω SST
Number of Cells	3.4 million
Continuity Residual	1e-3
X-Momentum Residual	1e-7
Y-Momentum Residual	1e-7
Energy Residual	1e-7
k and ω Residuals	1e-5



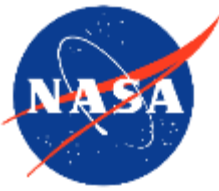
CMC9 geometry, simulated at conditions tested in (Giel et al., 2020)

Reynolds Stresses – CFD Results

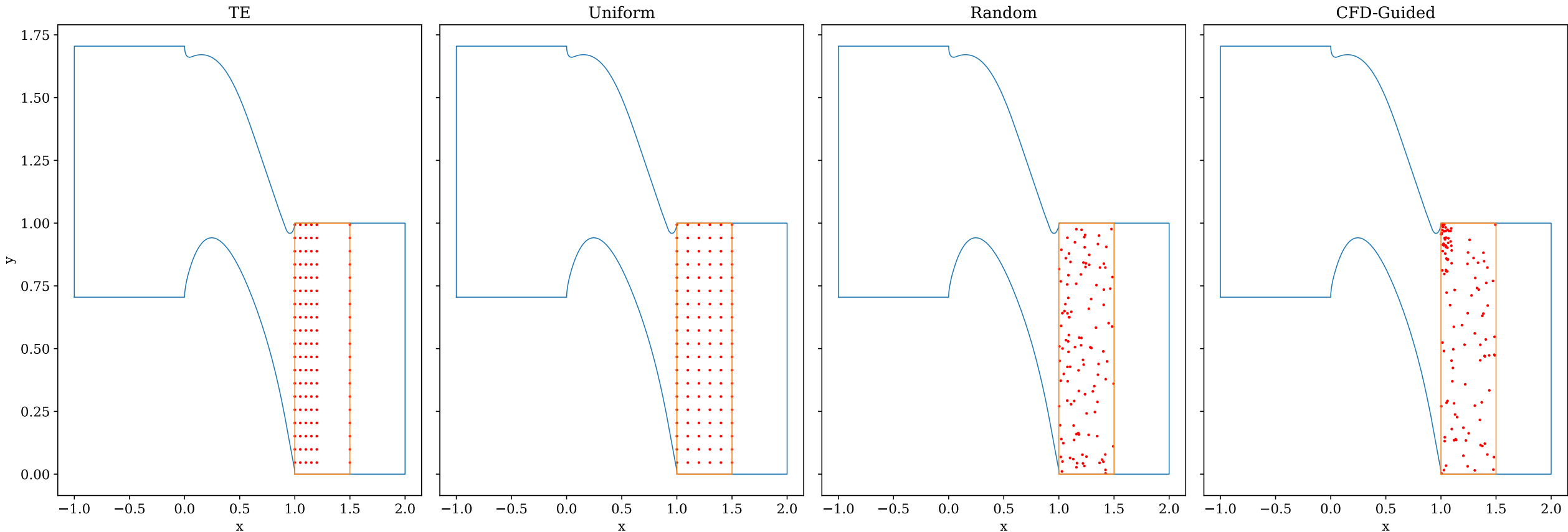




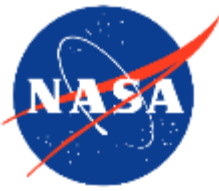
Effect of Spatial Sampling Strategy



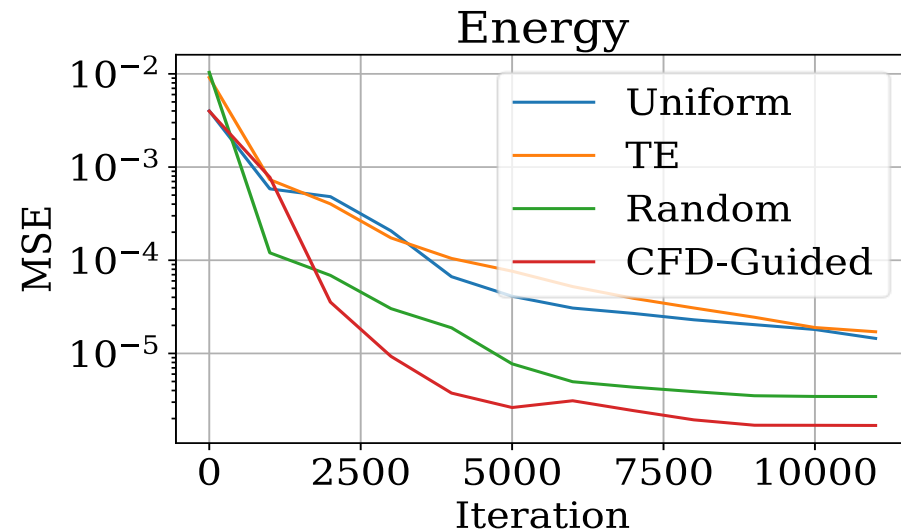
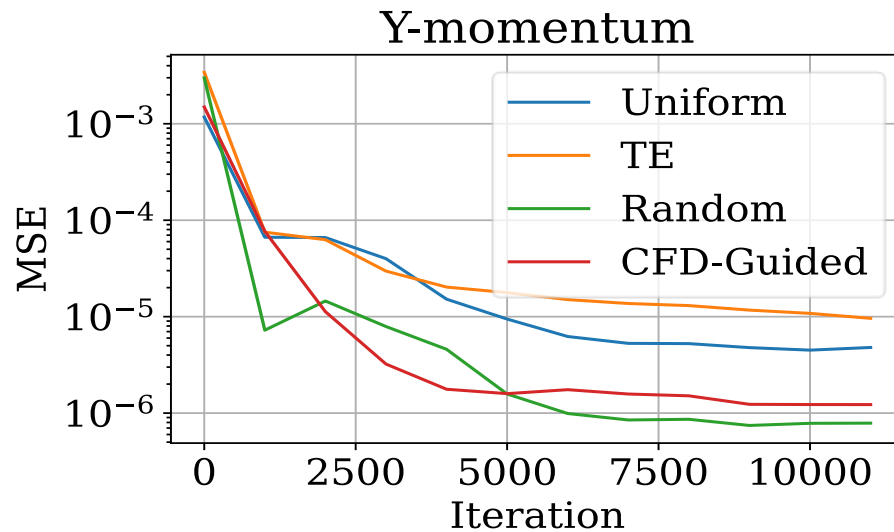
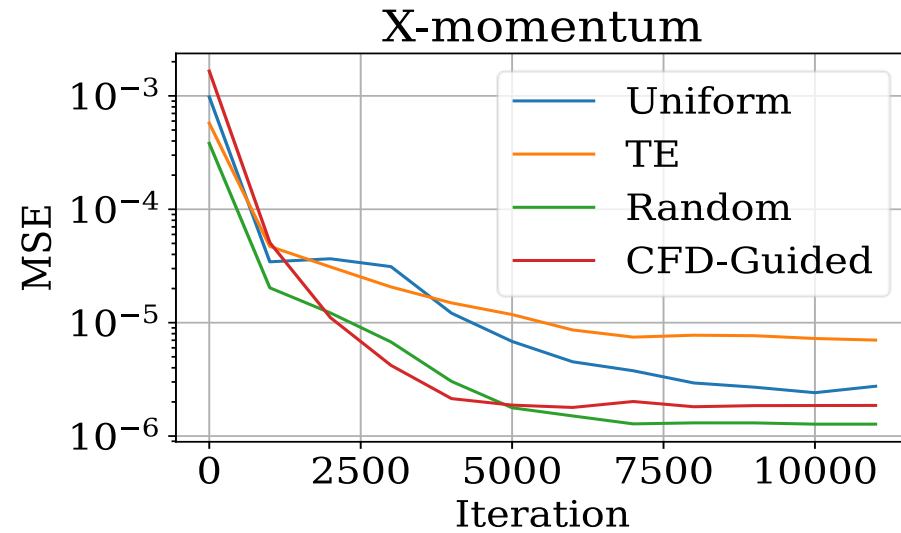
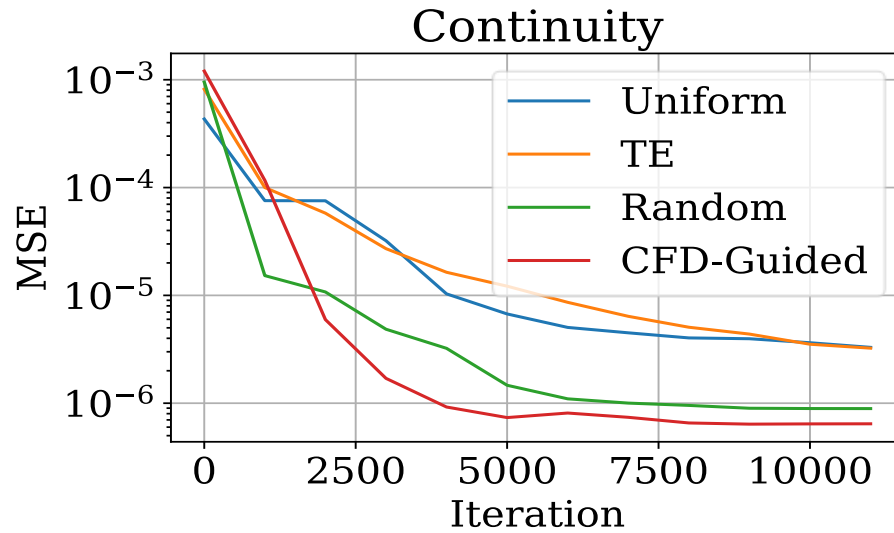
Training Data Locations



Markers represent locations where PINN was provided CFD solution data for training

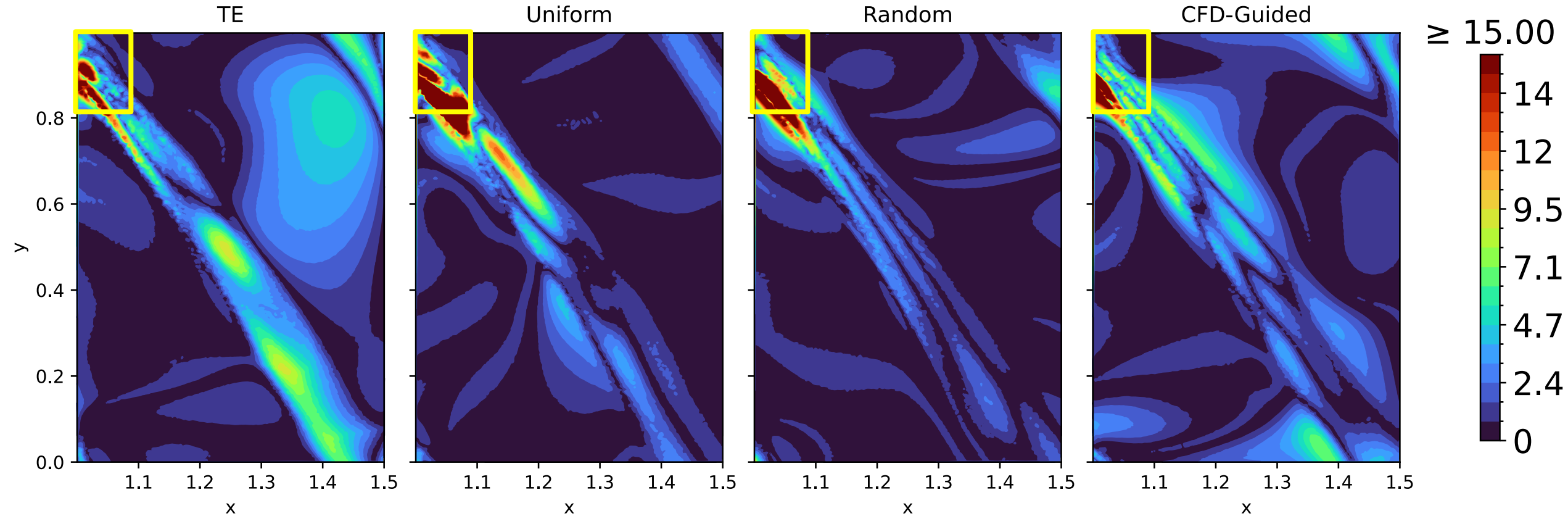


PINN Convergence History for Each Governing Equation

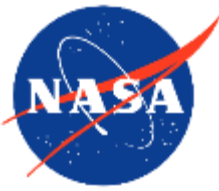




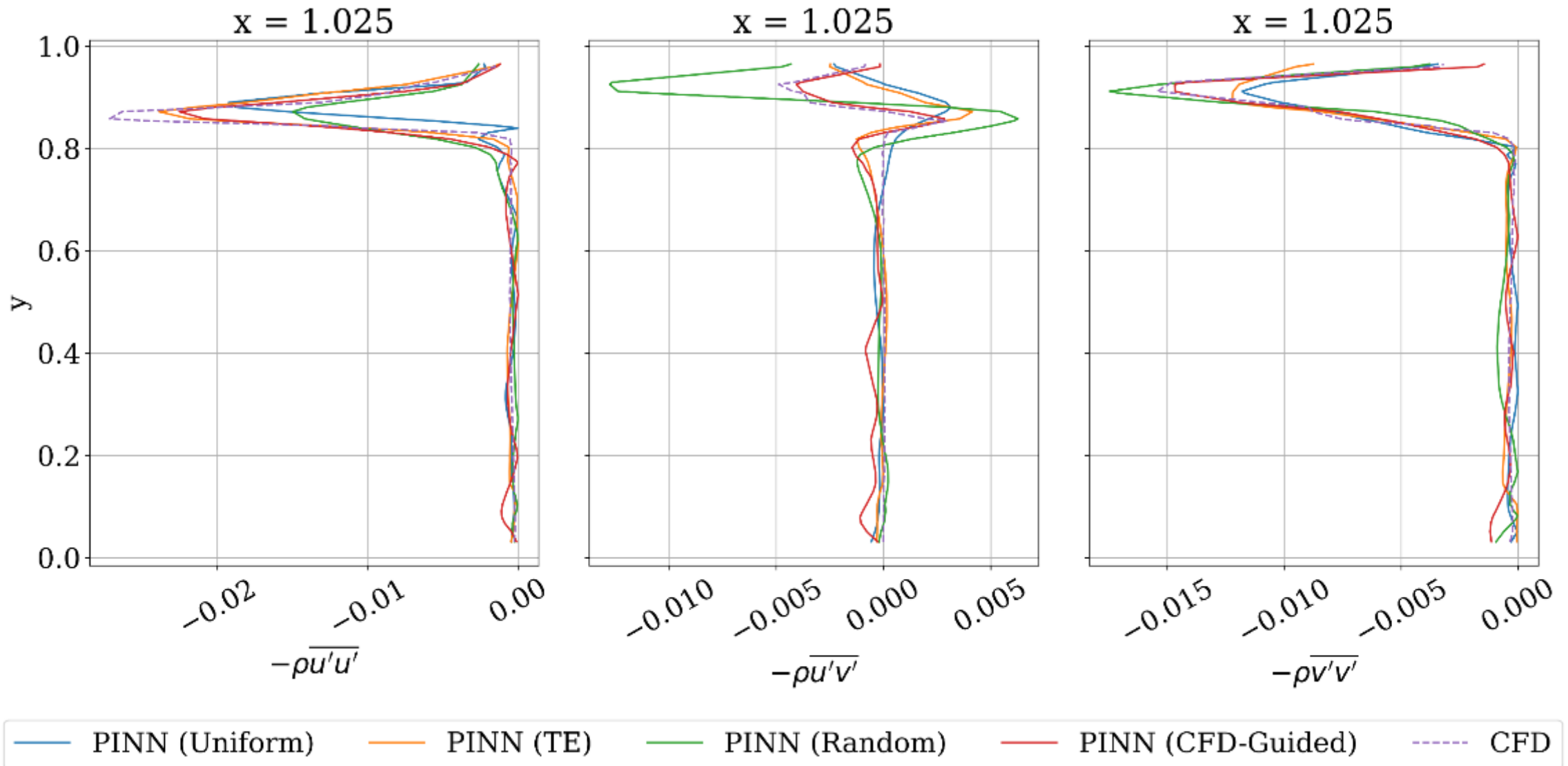
Percent Difference Between CFD and PINN: $\overline{\rho u' u'}$

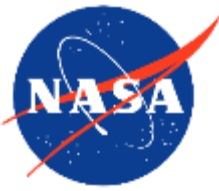


Adding points behind the TE made the predictions more accurate, at the expense of accuracy outside that region

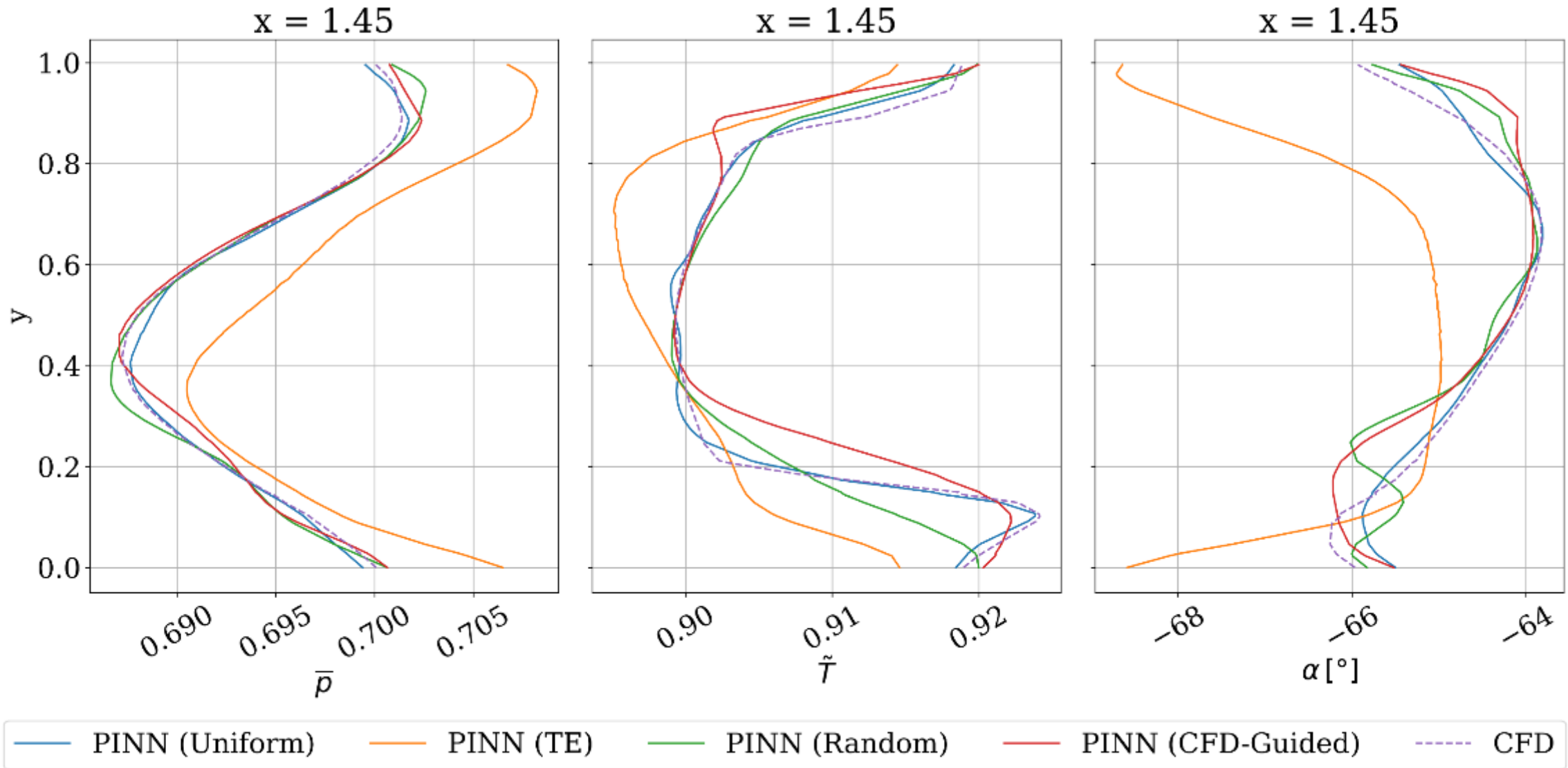


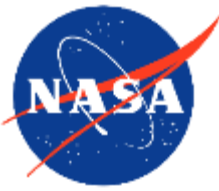
Pitchwise Surveys Behind Trailing Edge ($x = 1.025$)



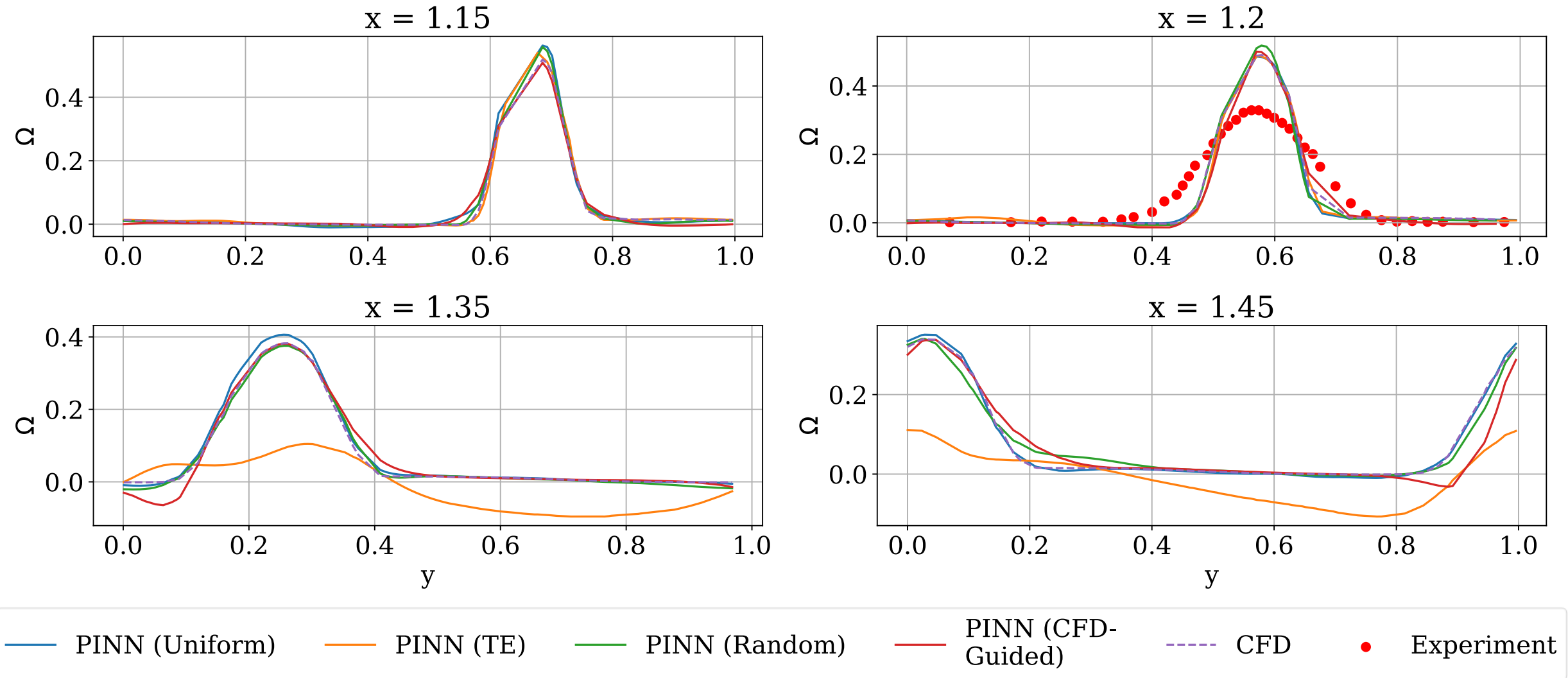


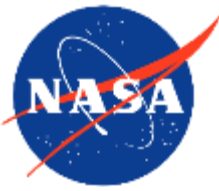
Pitchwise Surveys Near Outlet ($x = 1.45$)



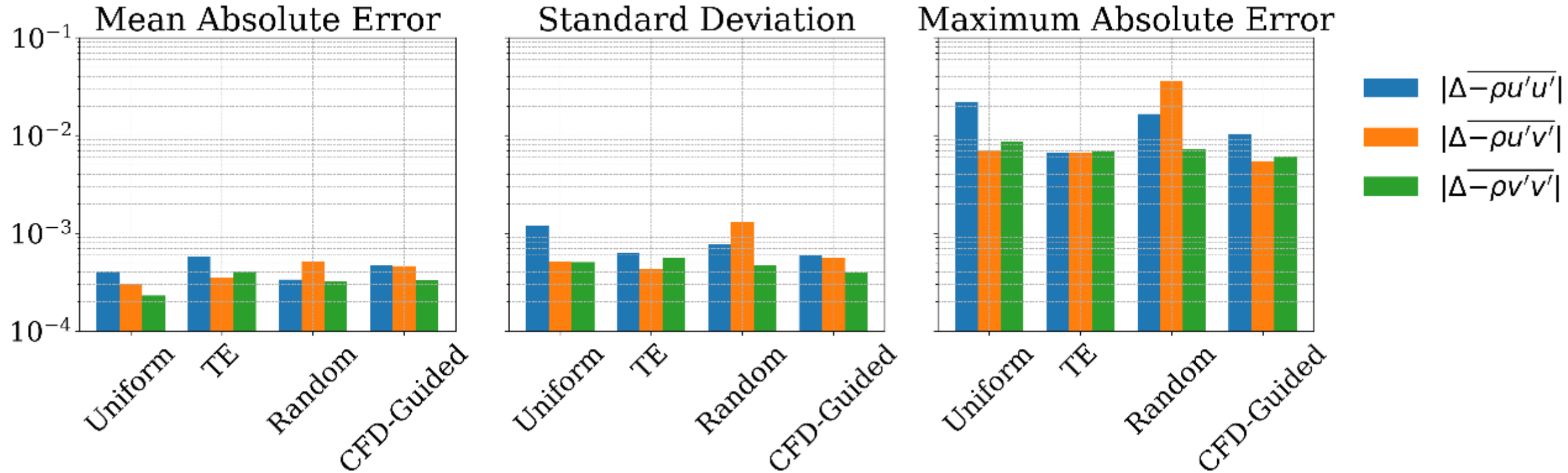


Wake Loss Surveys

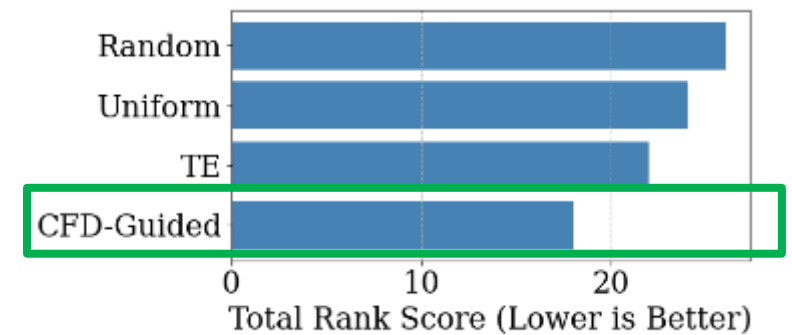




Overall Comparison of PINN Accuracy: Reynolds Stresses

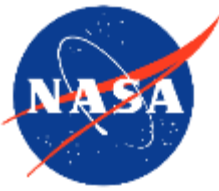


- For each Reynolds stress component, the mean, standard deviation, and maximum absolute error were computed
- For each metric and variable, sampling strategies are ranked from best (lowest error or deviation) to worst (highest)
- A strategy ranked 1st receives 1 point, 2nd receives 2 points, and so on
- Ranks are summed across all metrics and stress components to compute a total score
- Lower total score indicates better overall accuracy and consistency**

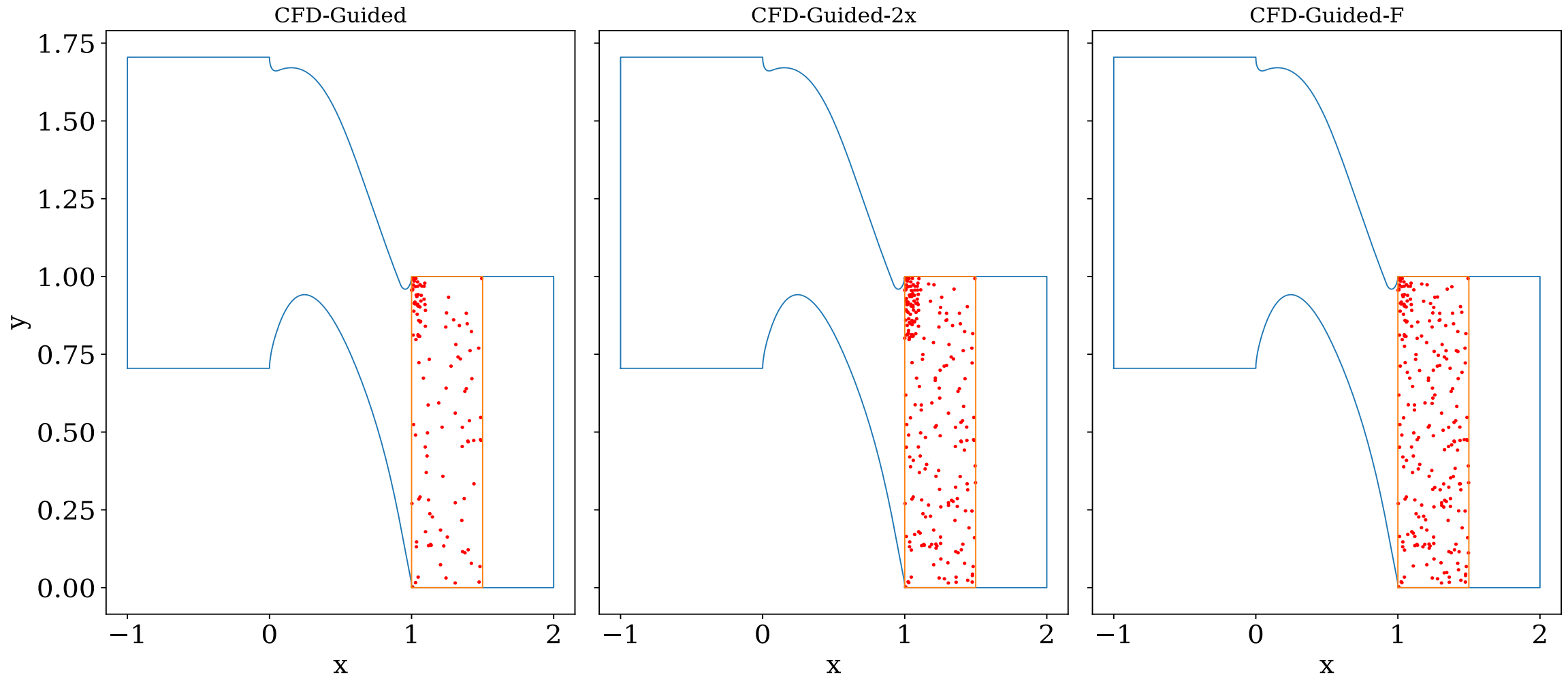




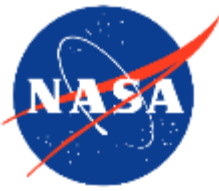
Effect of Sampling Density



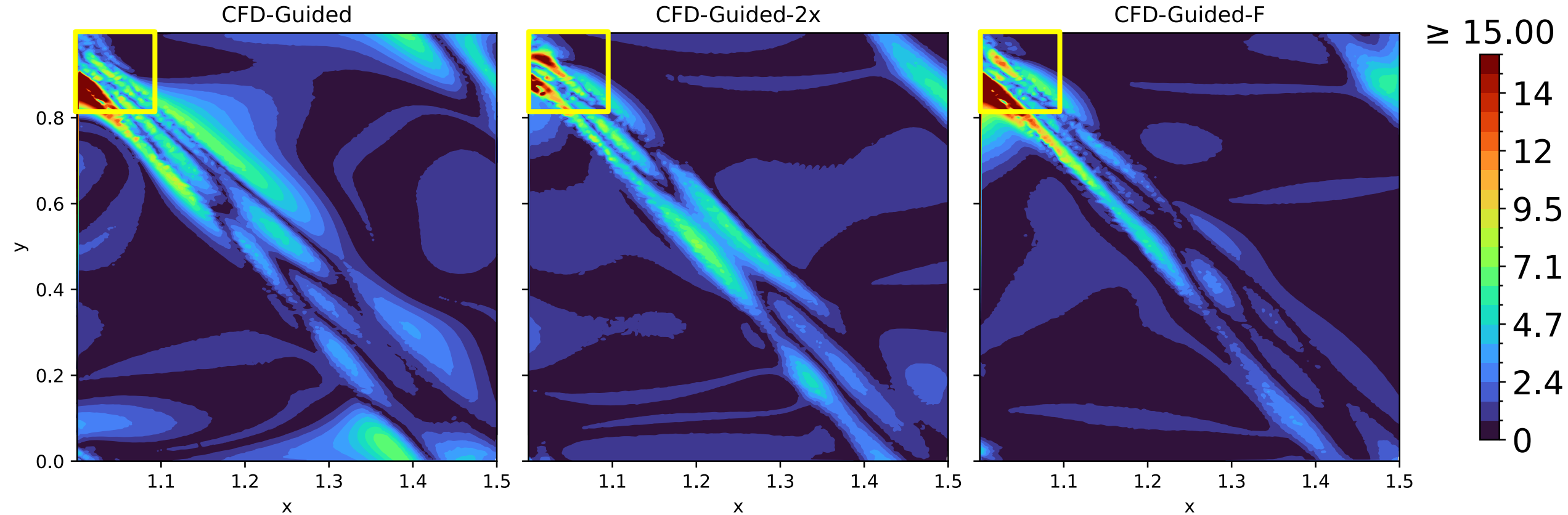
Training Data Locations

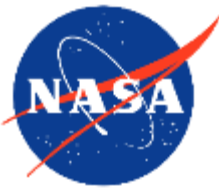


Markers represent locations where PINN was provided CFD solution data for training

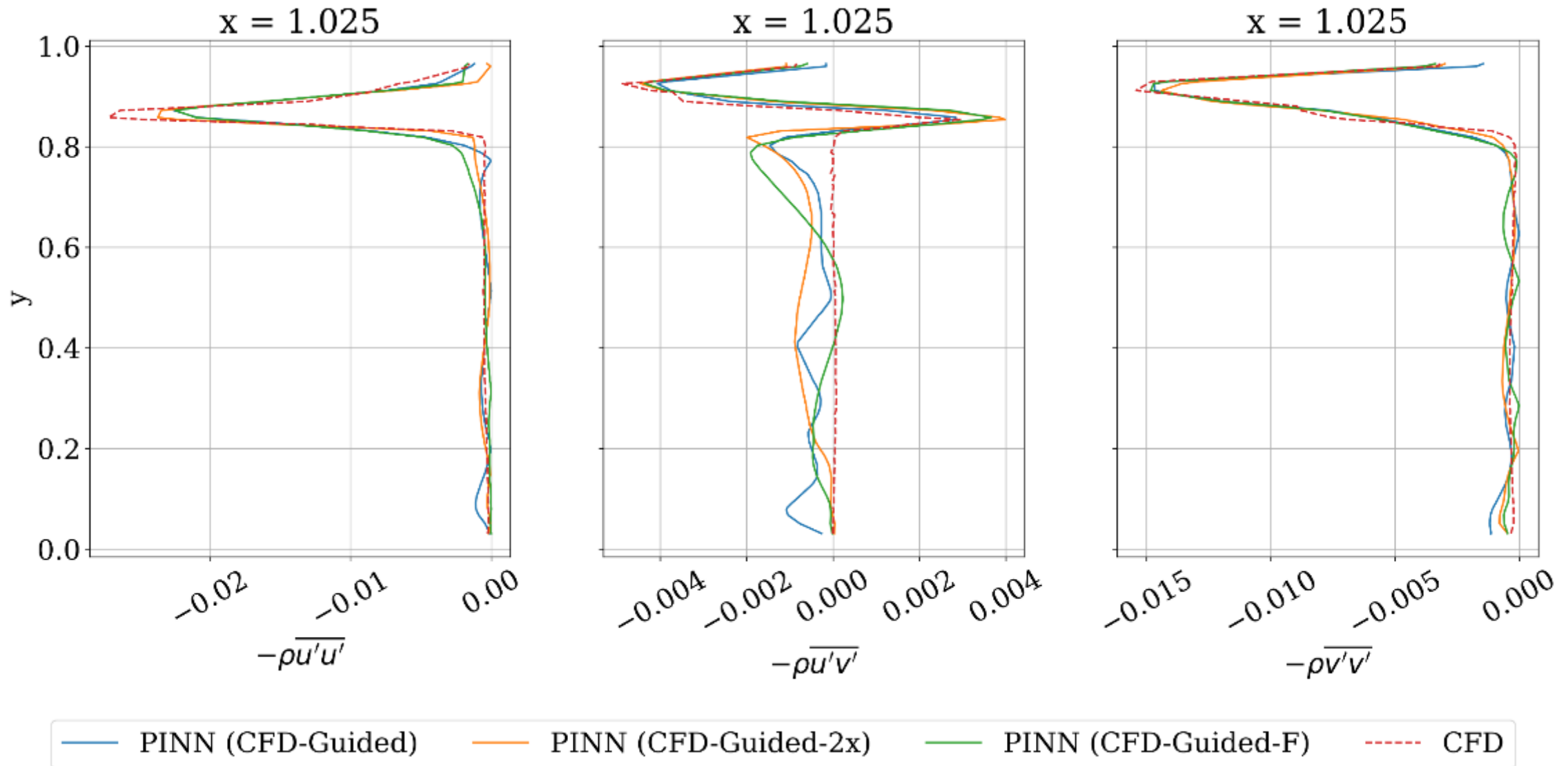


Percent Difference Between CFD and PINN: $\overline{\rho u' u'}$

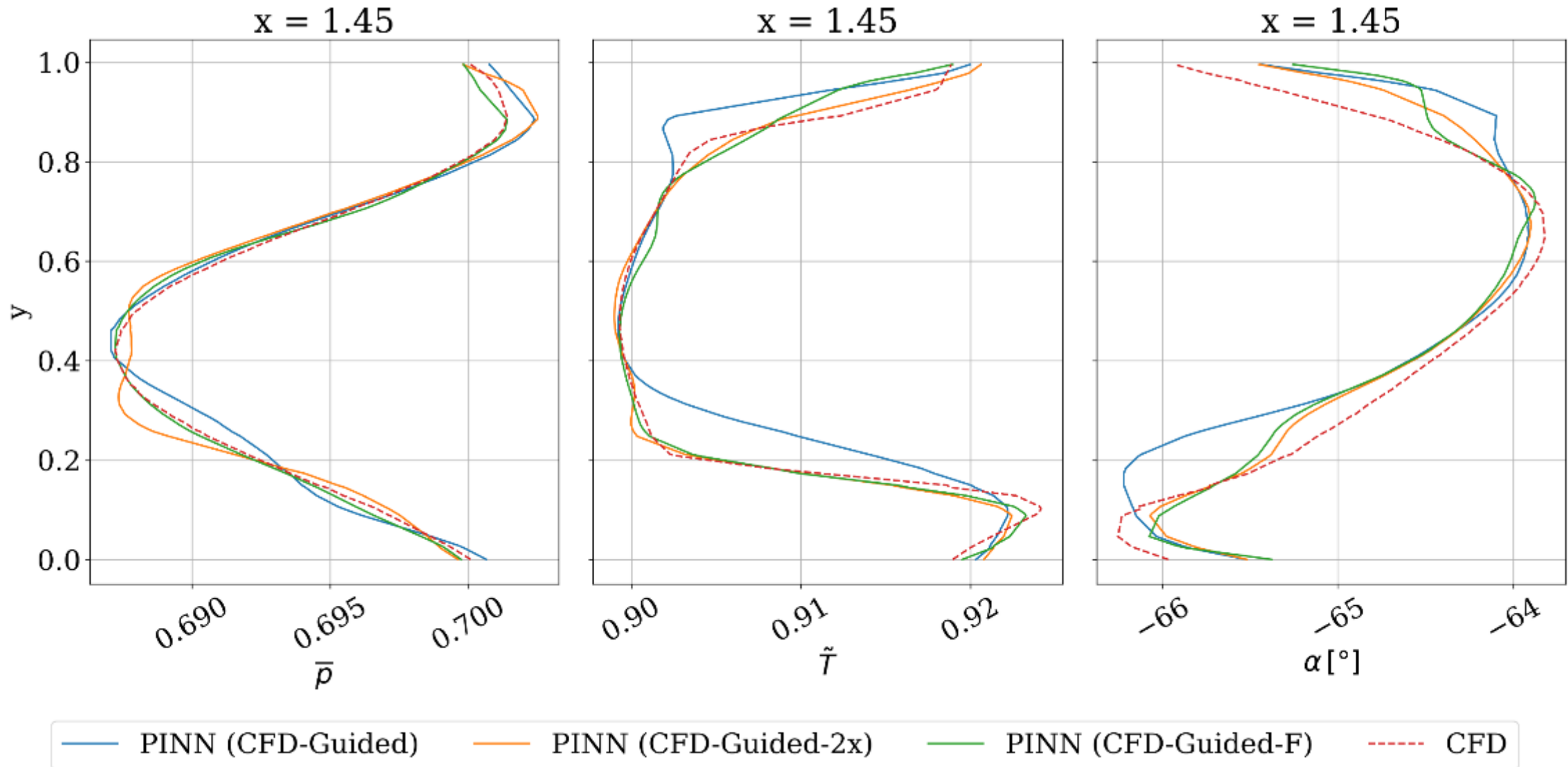


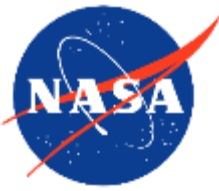


Pitchwise Surveys Behind Trailing Edge ($x = 1.025$)

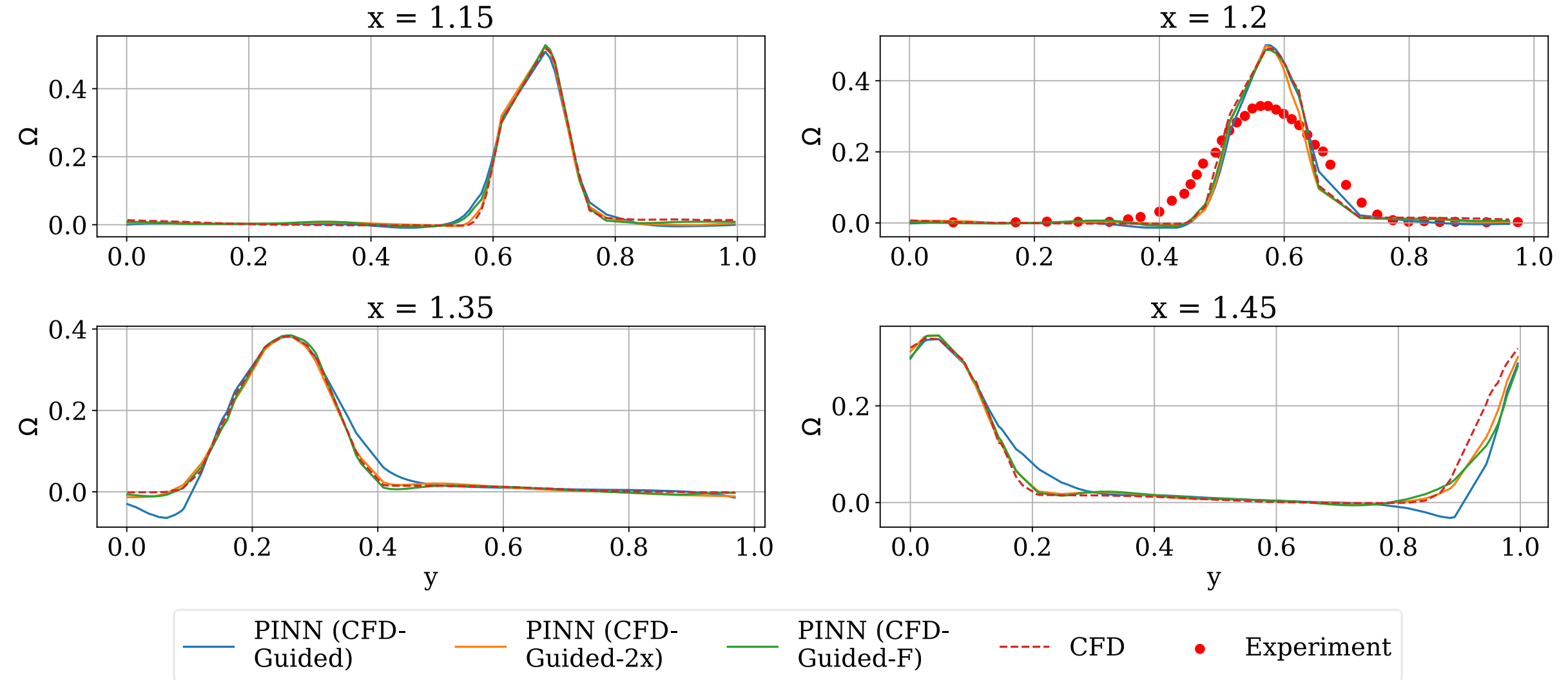


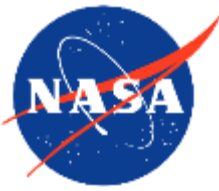
Pitchwise Surveys Near Outlet ($x = 1.45$)



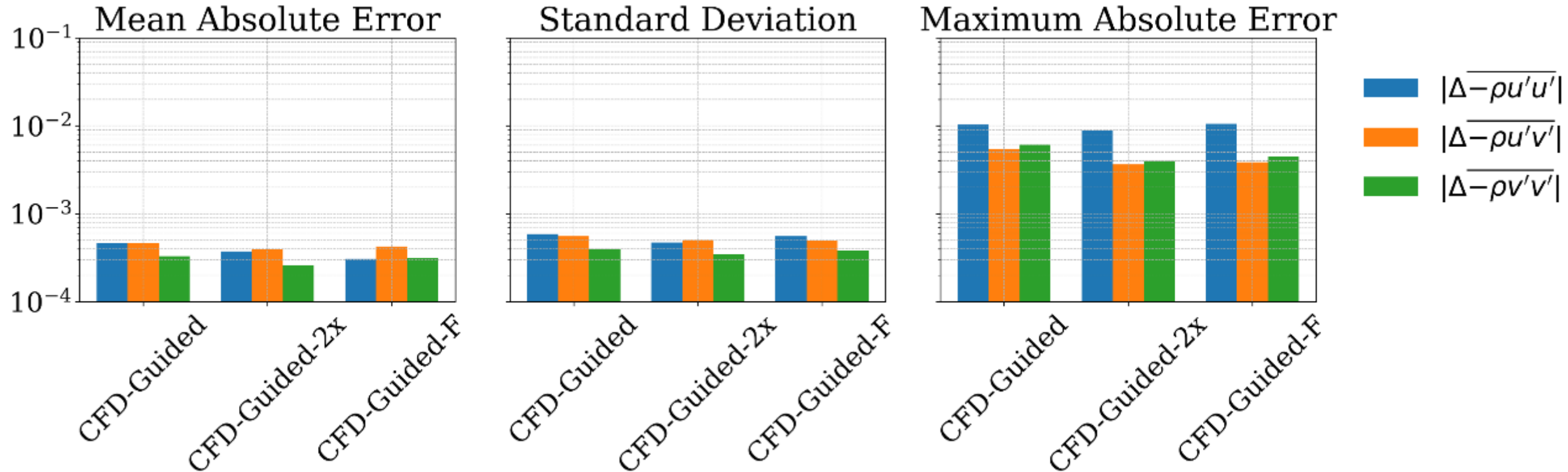


Wake Loss Surveys

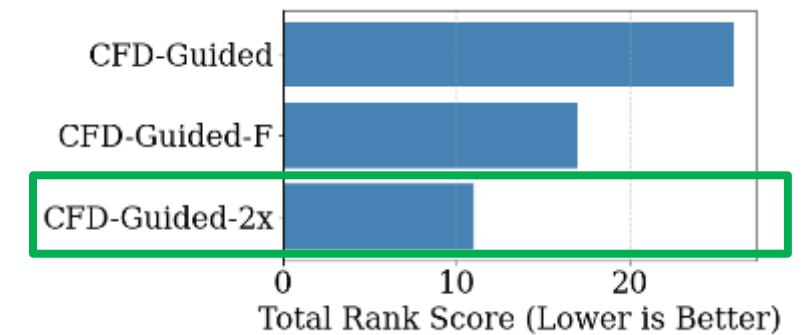


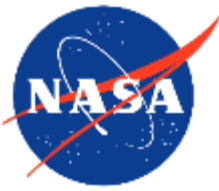


Overall Comparison of PINN Accuracy: Reynolds Stresses



- For each Reynolds stress component, the mean, standard deviation, and maximum absolute error were computed
- For each metric and component, sampling strategies are ranked from best (lowest error or deviation) to worst (highest)
- A strategy ranked 1st receives 1 point, 2nd receives 2 points, and so on
- Ranks are summed across all metrics and stress components to compute a total score
- Lower total score indicates better overall accuracy and consistency**





Conclusions and Future Work

Conclusions:

1. Location of training data affected PINN predictions
2. Random sampling cases converged faster, exhibited lower training errors, and captured sharp gradients better
3. CFD-Guided was the most robust strategy in this study
4. Doubling the training data size, against the baseline number, alleviated the deficiencies in the downstream predictions for the CFD-Guided cases
5. A threshold was seen where increasing the number of training points in the TE region resulted in diminishing returns

Future/Ongoing Work:

- Closing the gap to infer Reynolds stresses with minimal, or no, stress data
- More advanced architecture – replacing the multilayer perceptron (MLP) with a Kolmogorov-Arnold network (KAN)

Acknowledgements:

This work was funded by NASA's Advanced Air Transport Technology (AATT) Project



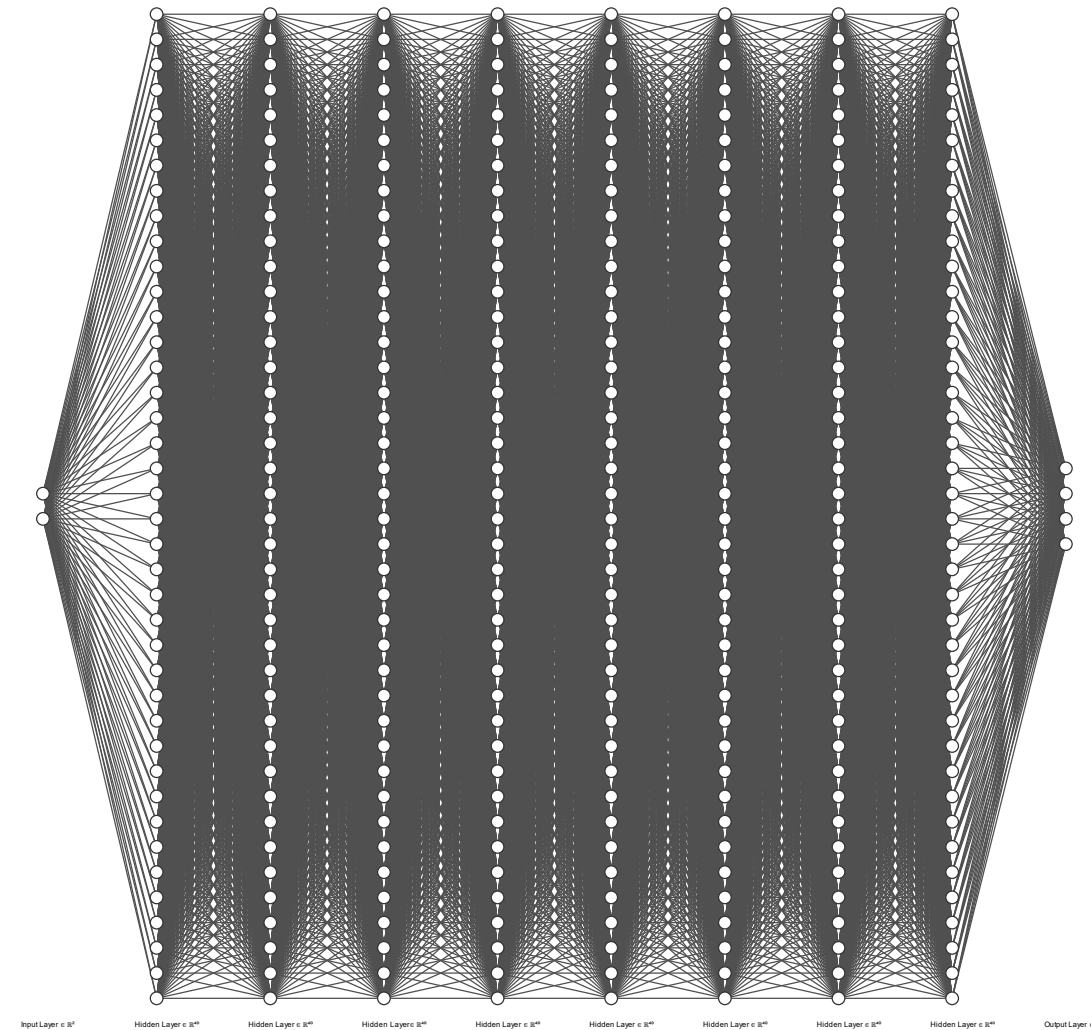
Thank you for your attention



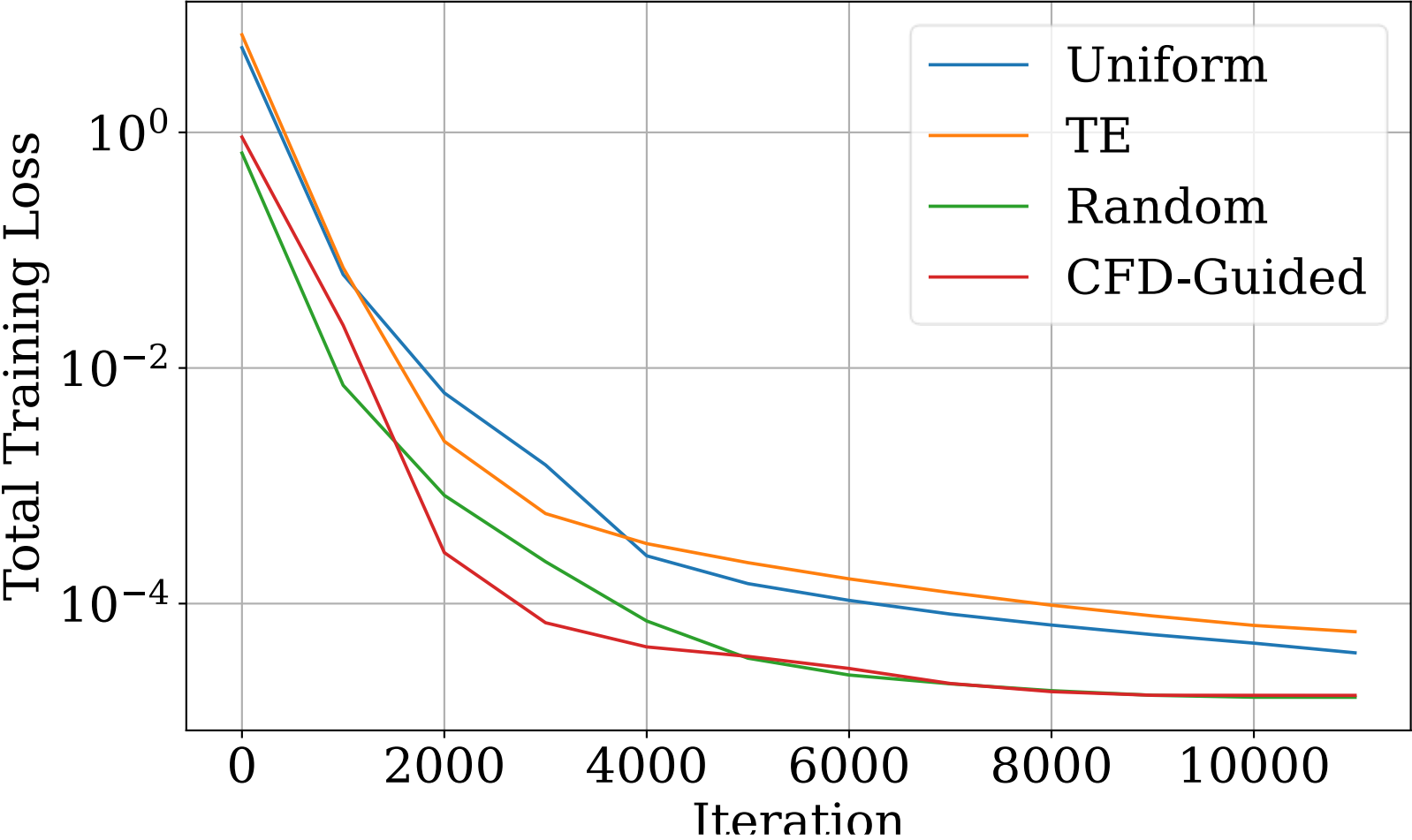
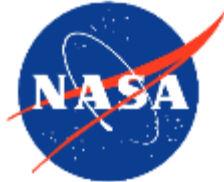
Appendix

Neural Network and Training

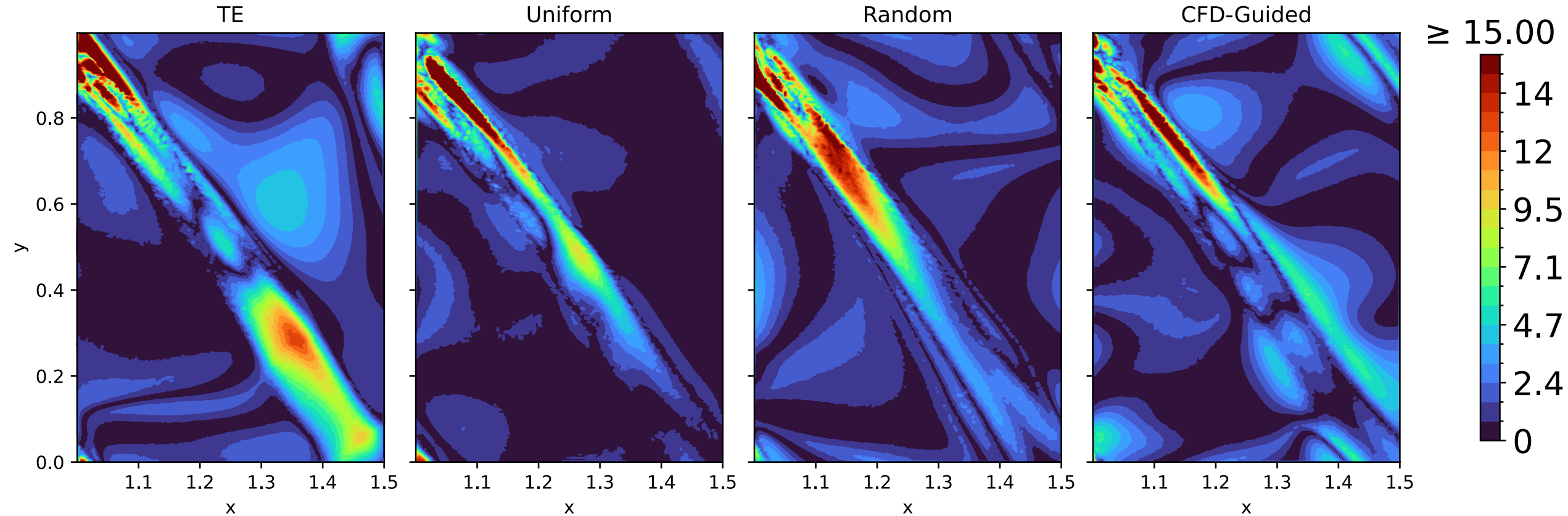
- 8 hidden layers
- 60 neurons per layer
- Activation function
 - tanh
- Output Transformation
 - Pressure and Temperature were transformed using an exponential function to ensure positive values
 - Normal stresses were forced to be positive to ensure physical consistency
- Residual Points
 - Cell-center coordinates from CFD mesh were used
 - 18,363 points in domain used for testing
- Training Process
 - Adam – 1000 iterations, learning rate = $1e-3$
 - L-BFGS – 10,000 iterations



Neural network architecture

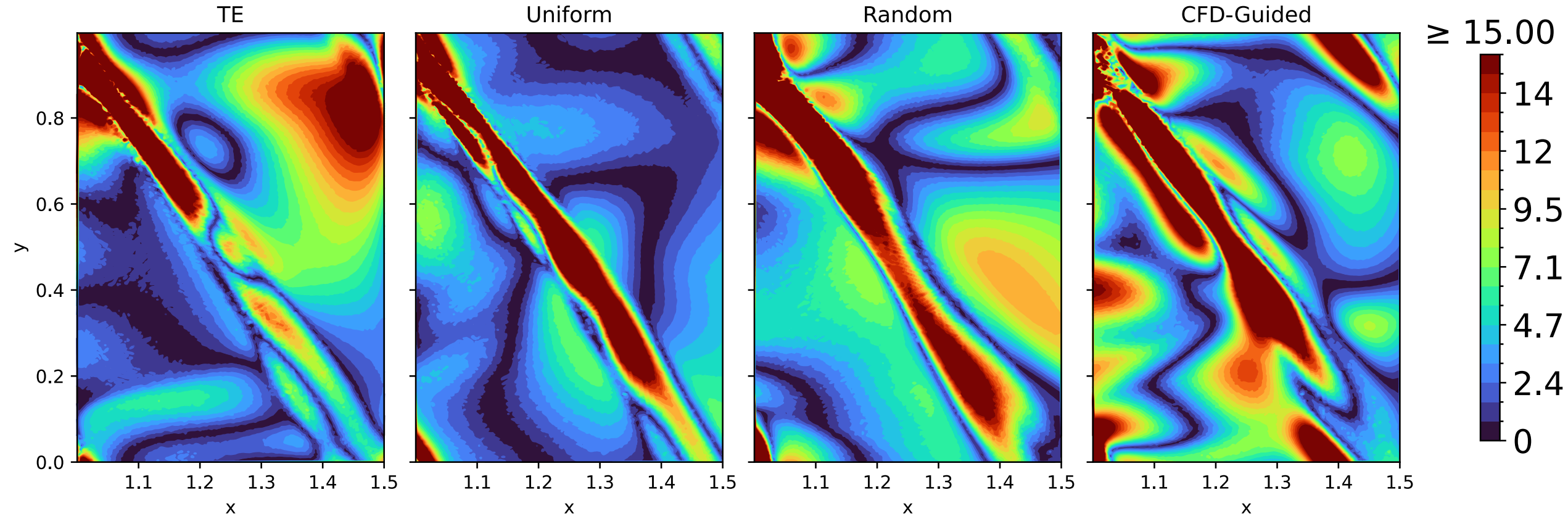


$$\Delta |-\overline{\rho v'v'}|$$

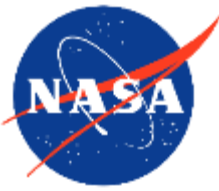


- CFD-Guided had smaller differences than any other method behind the TE
- At the expense of larger areas of difference downstream

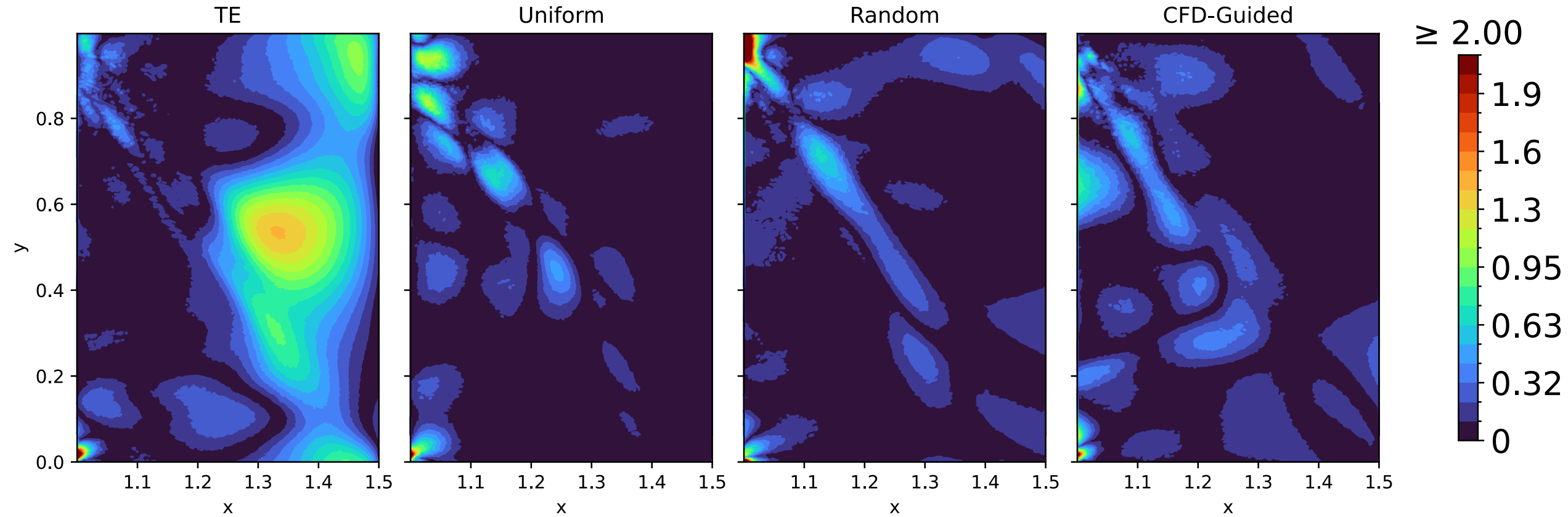
$$\Delta |-\overline{\rho u'v'}|$$

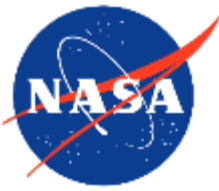


- CFD-Guided had smaller differences than any other method behind the TE
- At the expense of larger areas of difference downstream

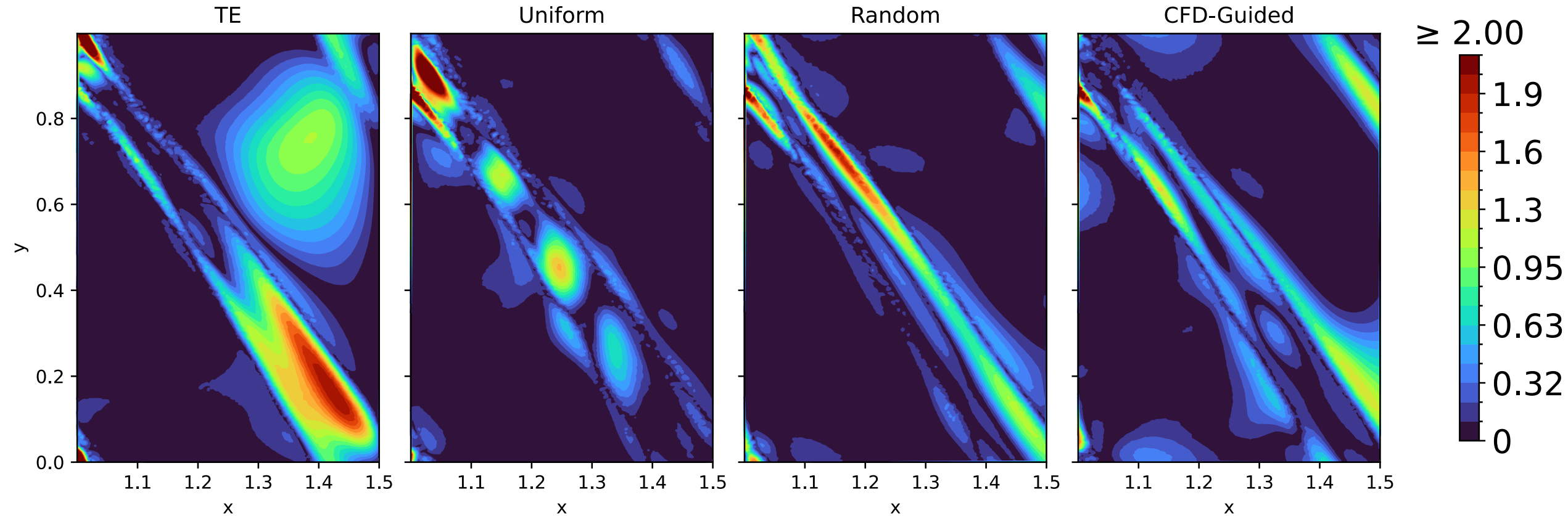


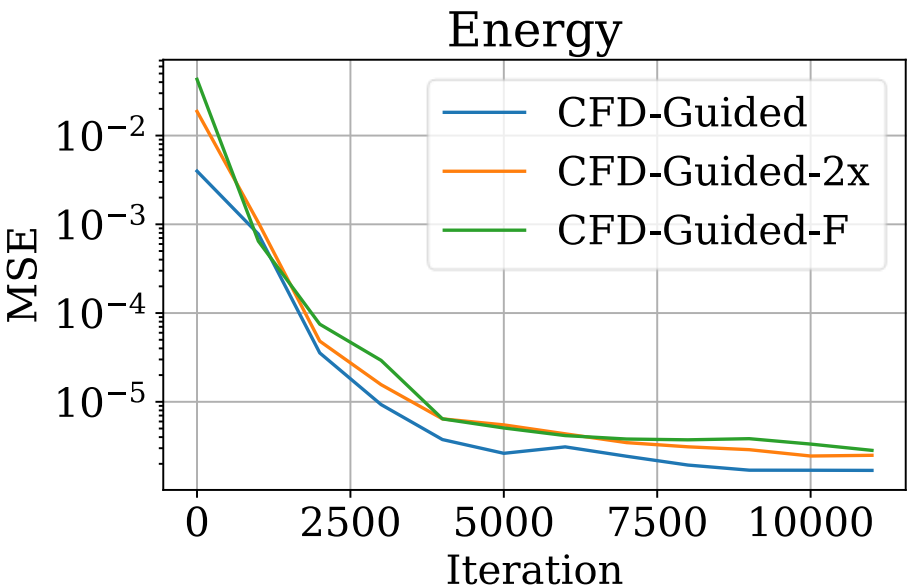
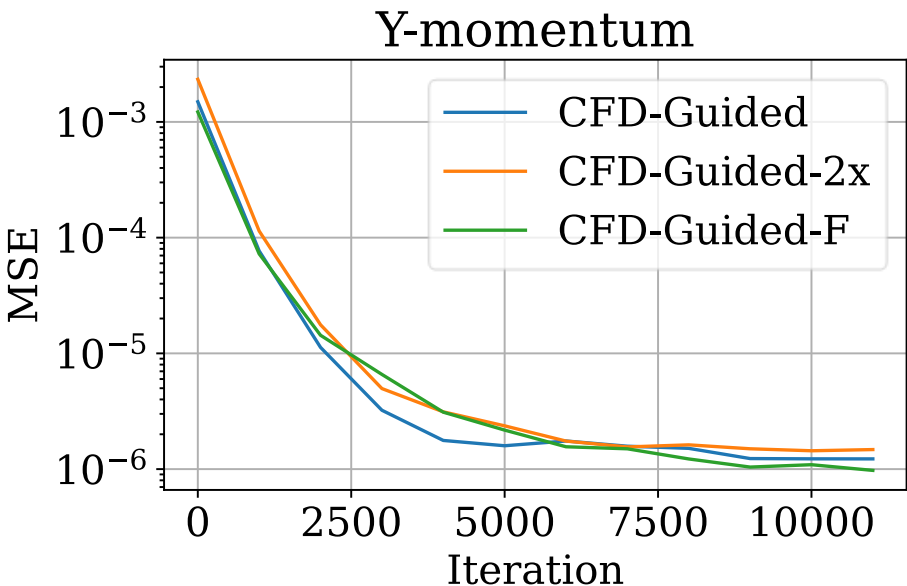
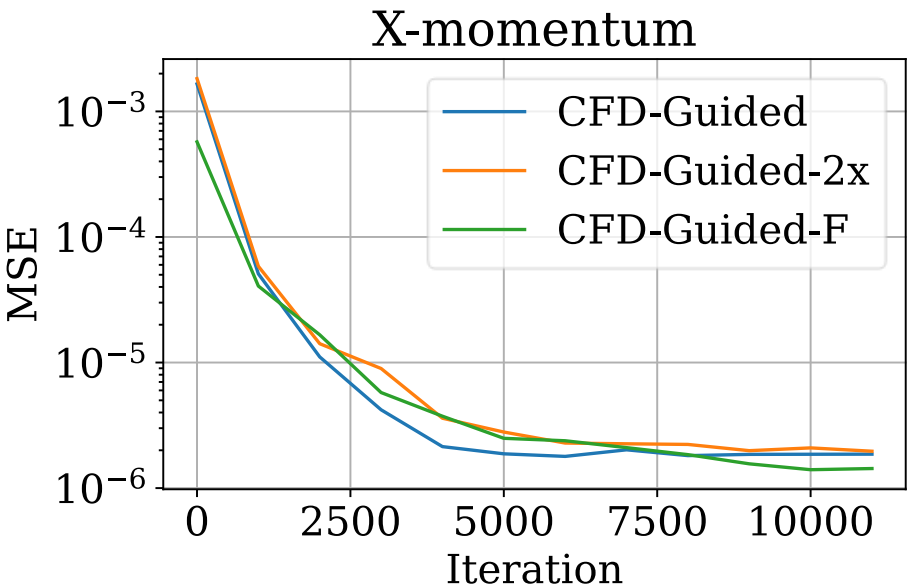
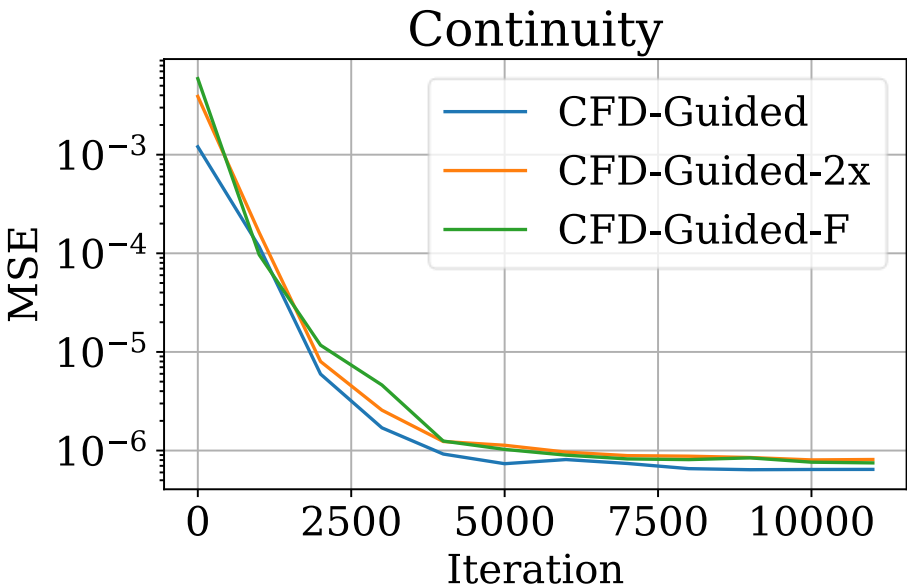
$$\Delta p$$





$$\Delta T$$



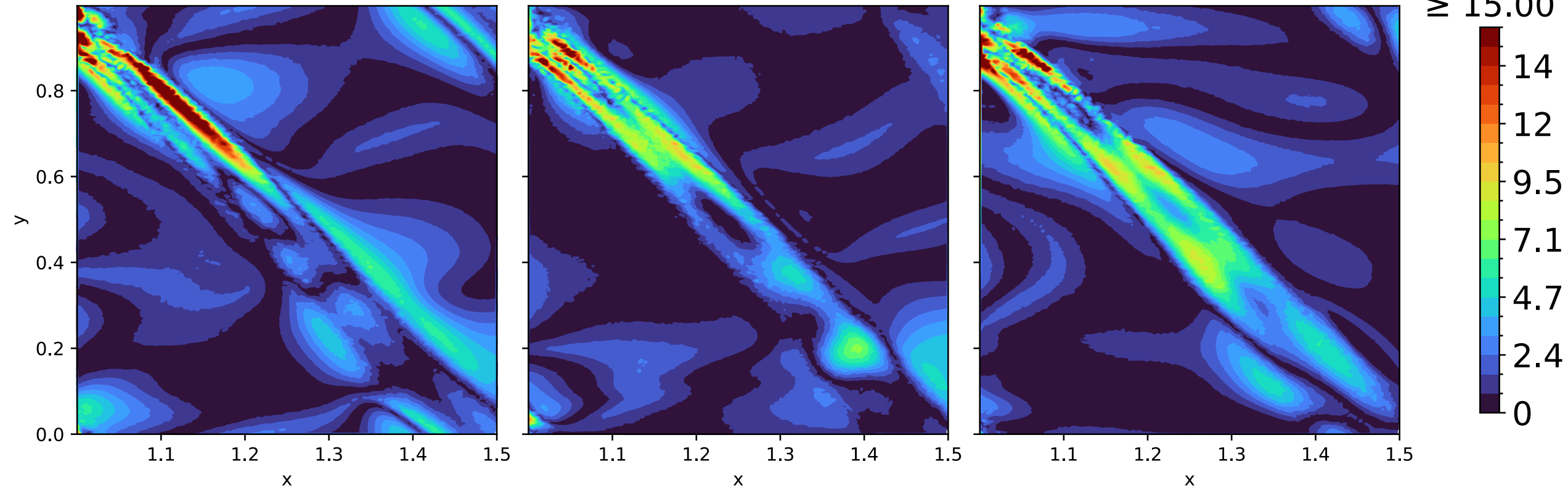


$$\Delta |-\overline{\rho v'v'}|$$

CFD-Guided

CFD-Guided-2x

CFD-Guided-F

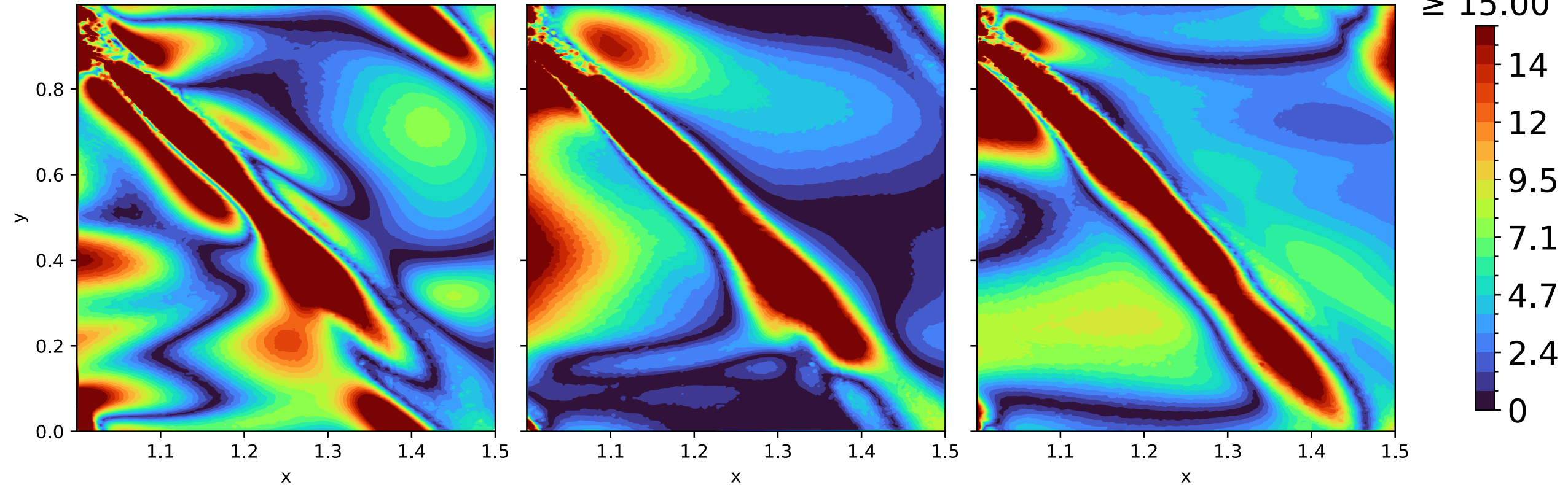


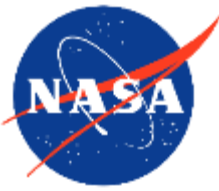
$$\Delta|-\overline{\rho u'v'}|$$

CFD-Guided

CFD-Guided-2x

CFD-Guided-F



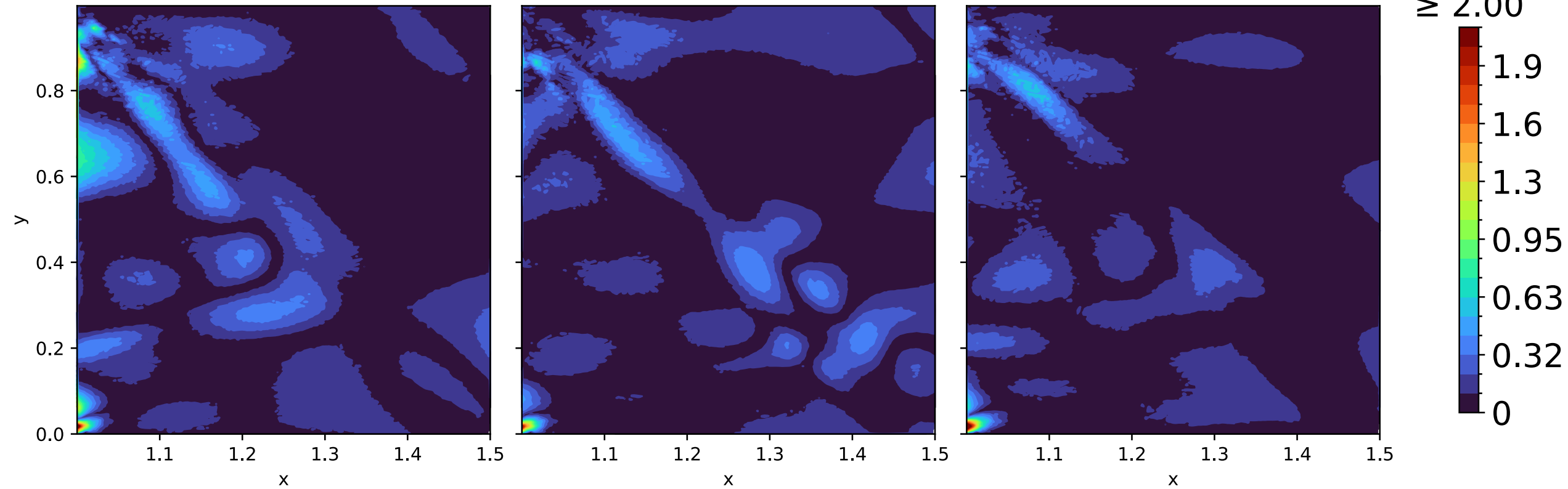


$$\Delta p$$

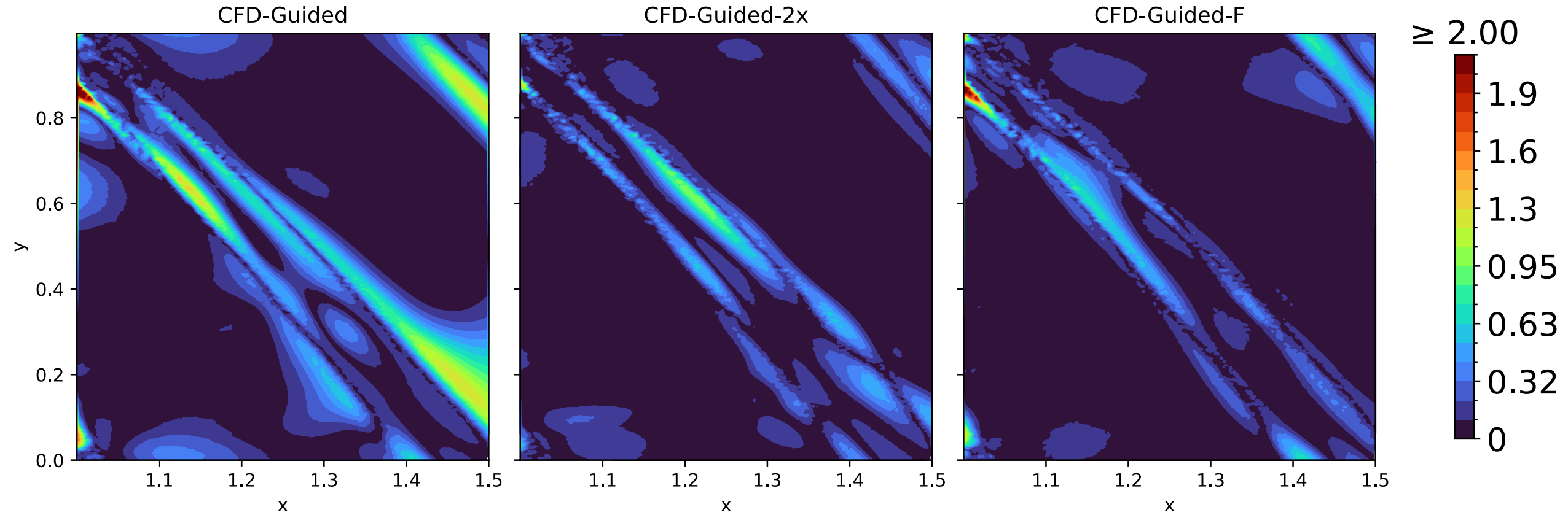
CFD-Guided

CFD-Guided-2x

CFD-Guided-F

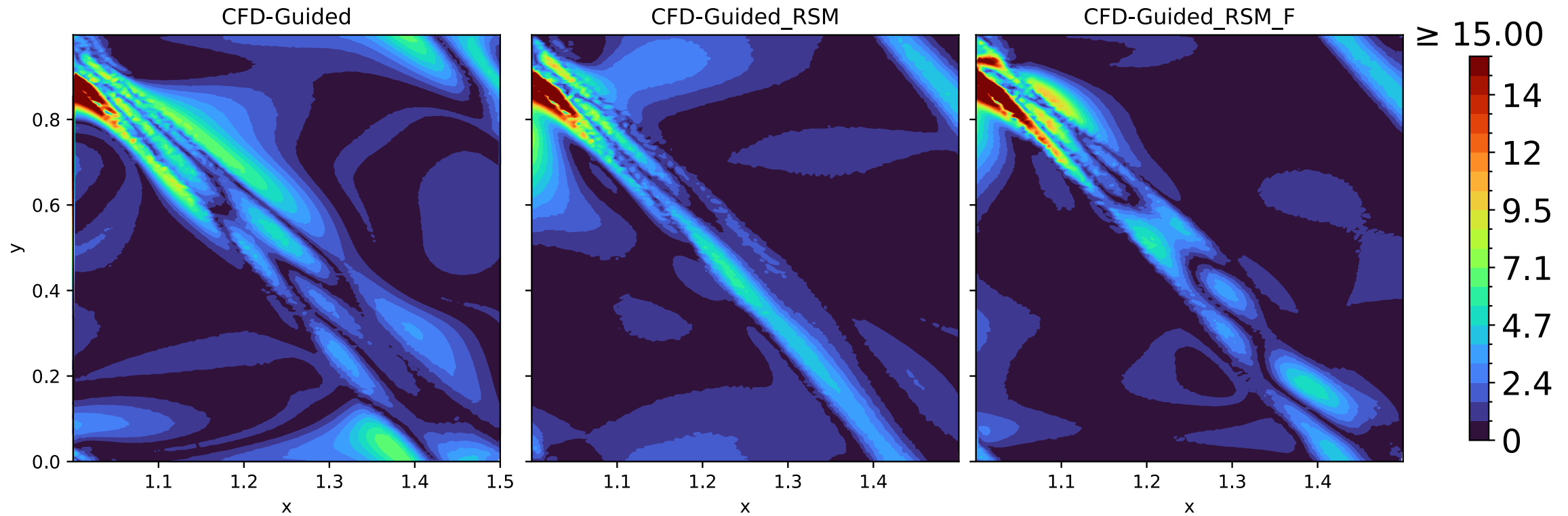


$$\Delta T$$





Future/Ongoing Work: PINN Model Comparisons



Note: “Simple” RSM equations to match assumptions used in $k-\omega$ SST