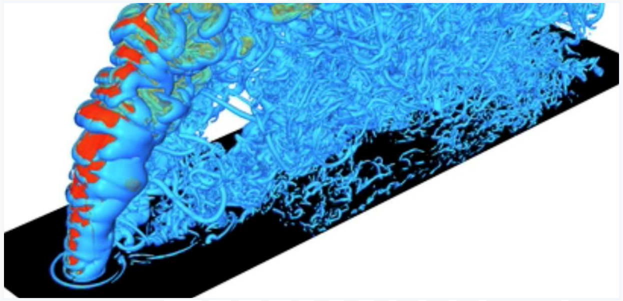
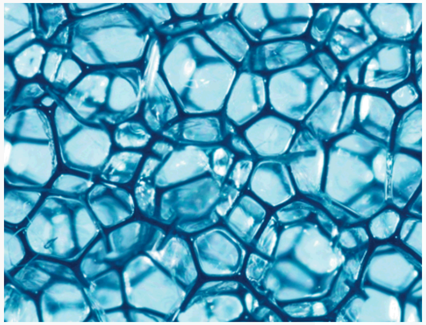
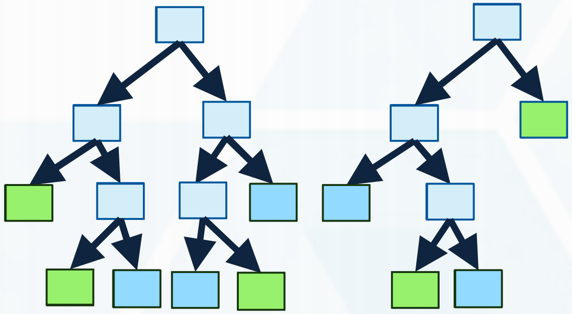
SAND2017-2142PE







Machine Learning + Physics

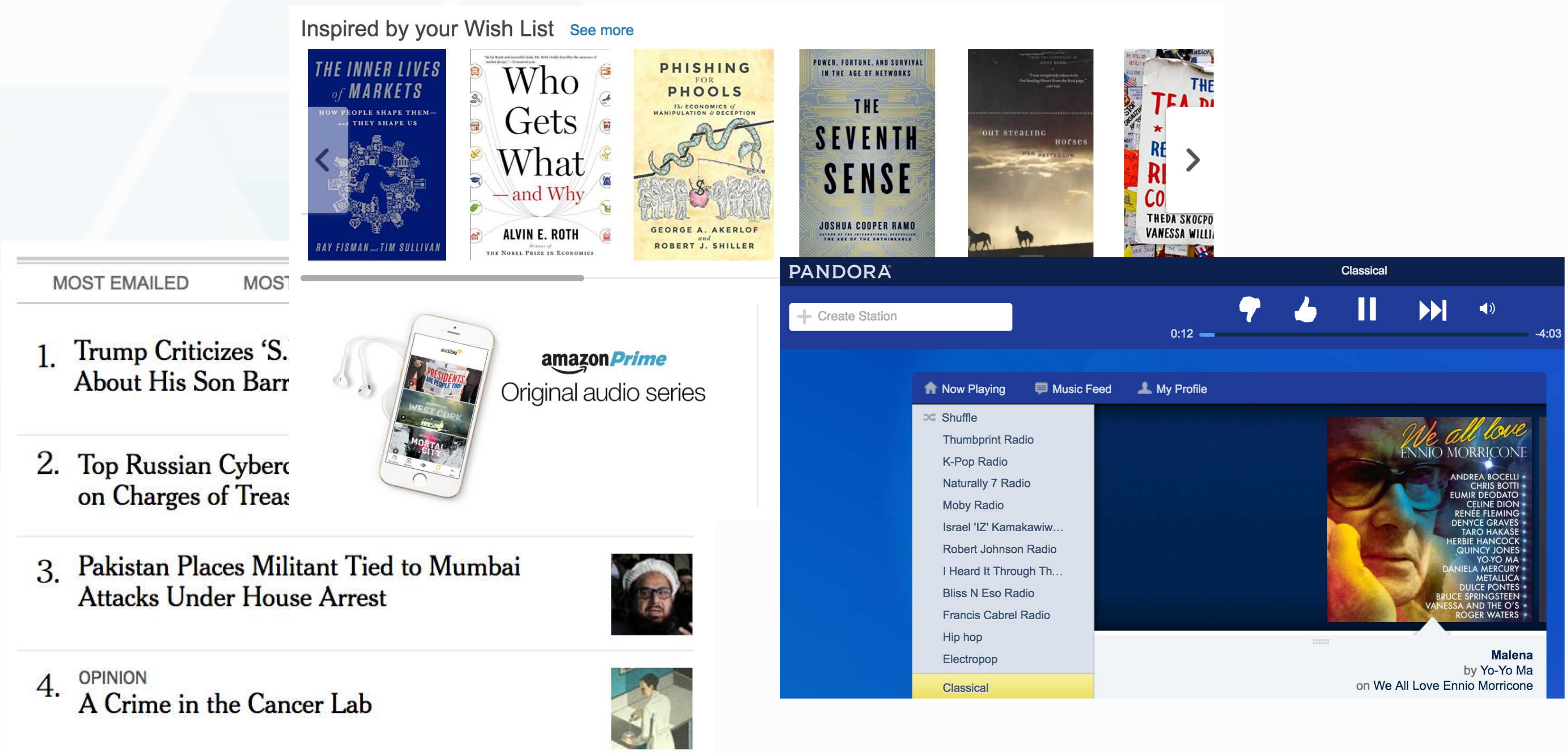
Julia Ling

May 2017

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

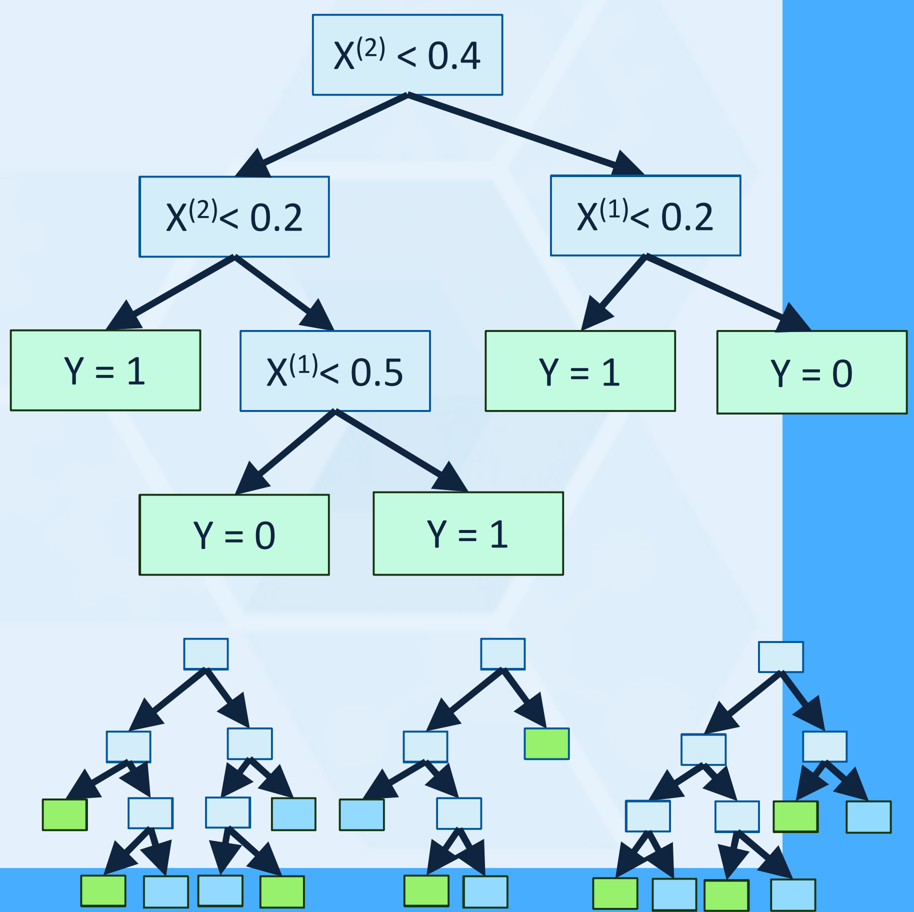
What is Machine Learning?

- Data-driven algorithms to discern patterns and make predictions on big, high-dimensional data
- Linear regression, support vector machines, neural networks

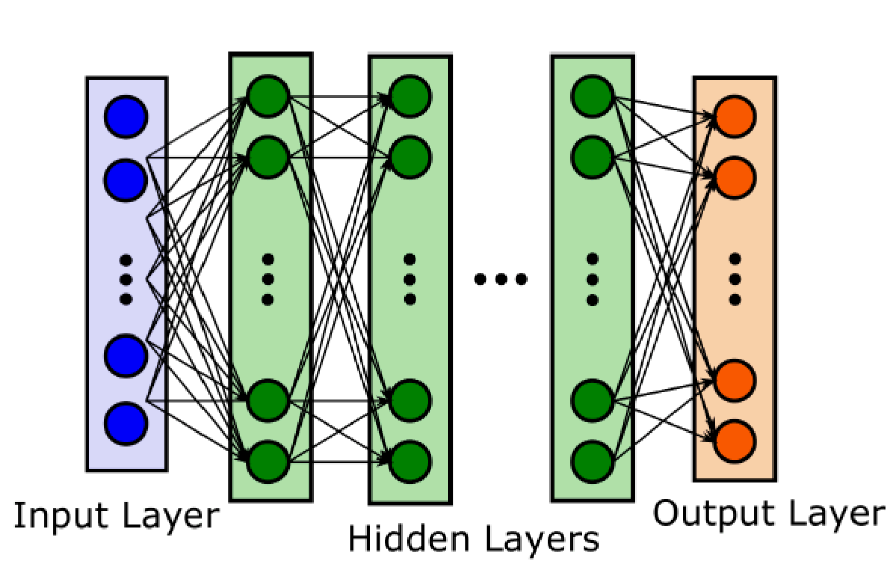


Machine Learning Algorithms

Random Forest



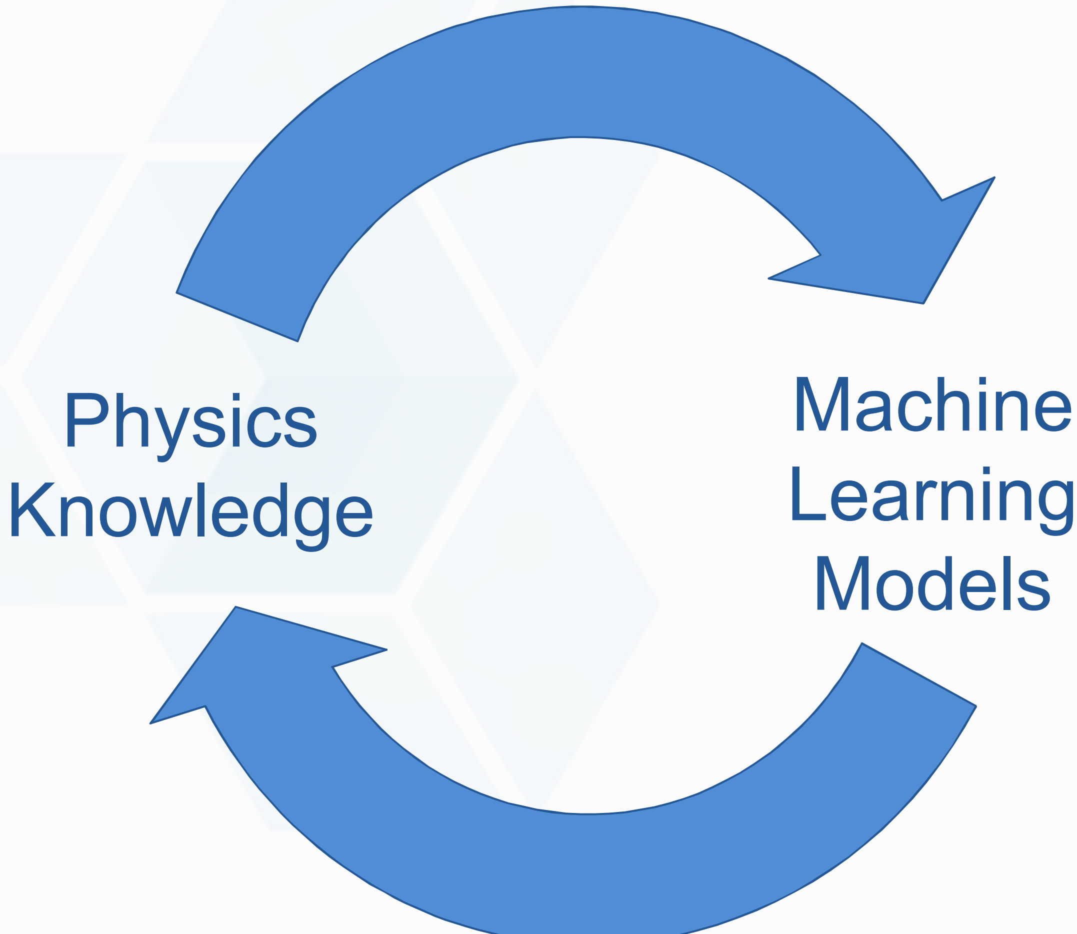
Neural Network

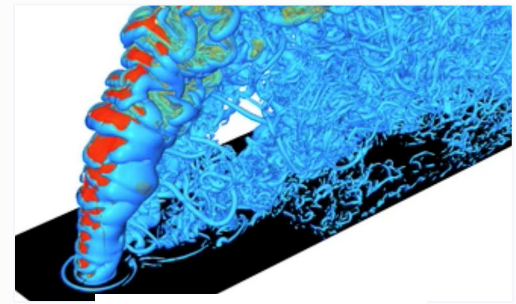


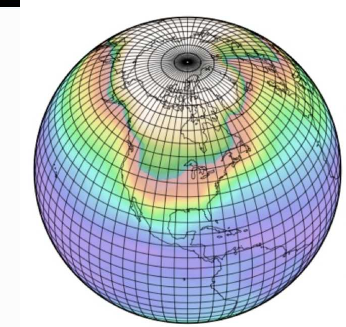
$$y = f(w^T x)$$

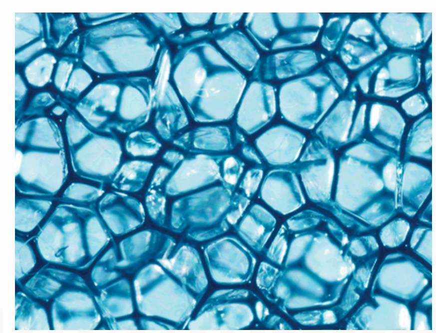
Machine Learning on Engineering Systems

- How should scientific domain knowledge interact with data-driven models?









Outline

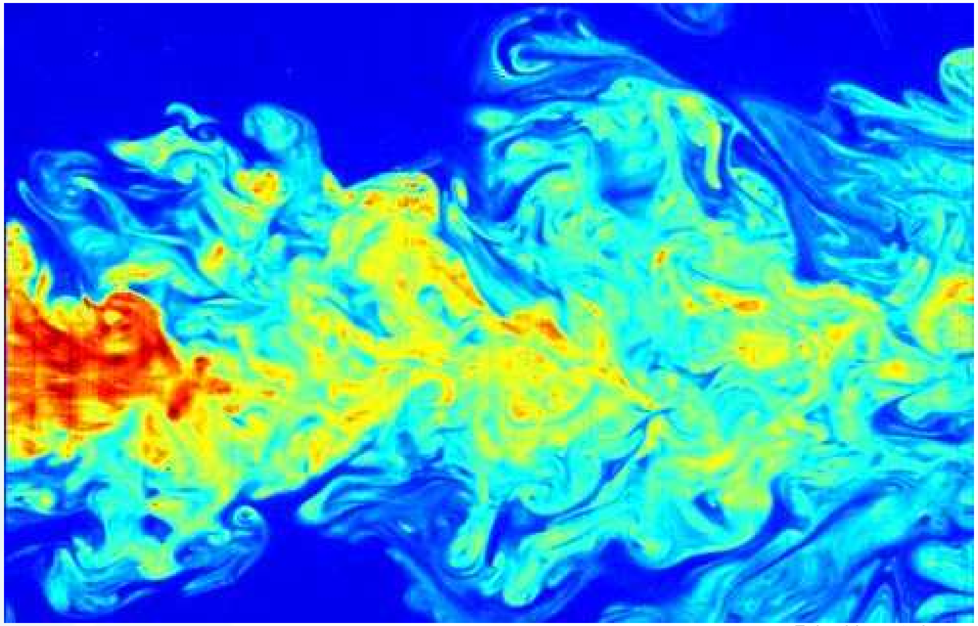
1) Machine learning for turbulence uncertainty quantification

1) Rule extraction to gain physical intuition

1) Deep learning with embedded invariances

Turbulence

Chaotic 3-D fluid motion at a continuum of scales



Fukushima et al.

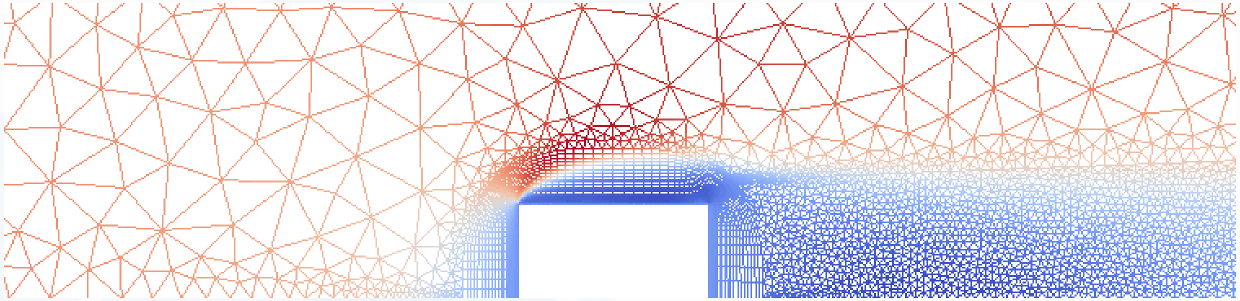
Hokusai (c 1830)

http://www.windturbinesyndrome.com/2011/wind-turbine-turbulence-what-are-the-micro-climate-effects/

https://brilliant.org/wiki/rocket-physics/

Turbulence Simulations

"When I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic." –Horace Lamb



Direct Numerical Simulation (DNS)

- Hundreds of millions of core hours for even simple flows
- Exact

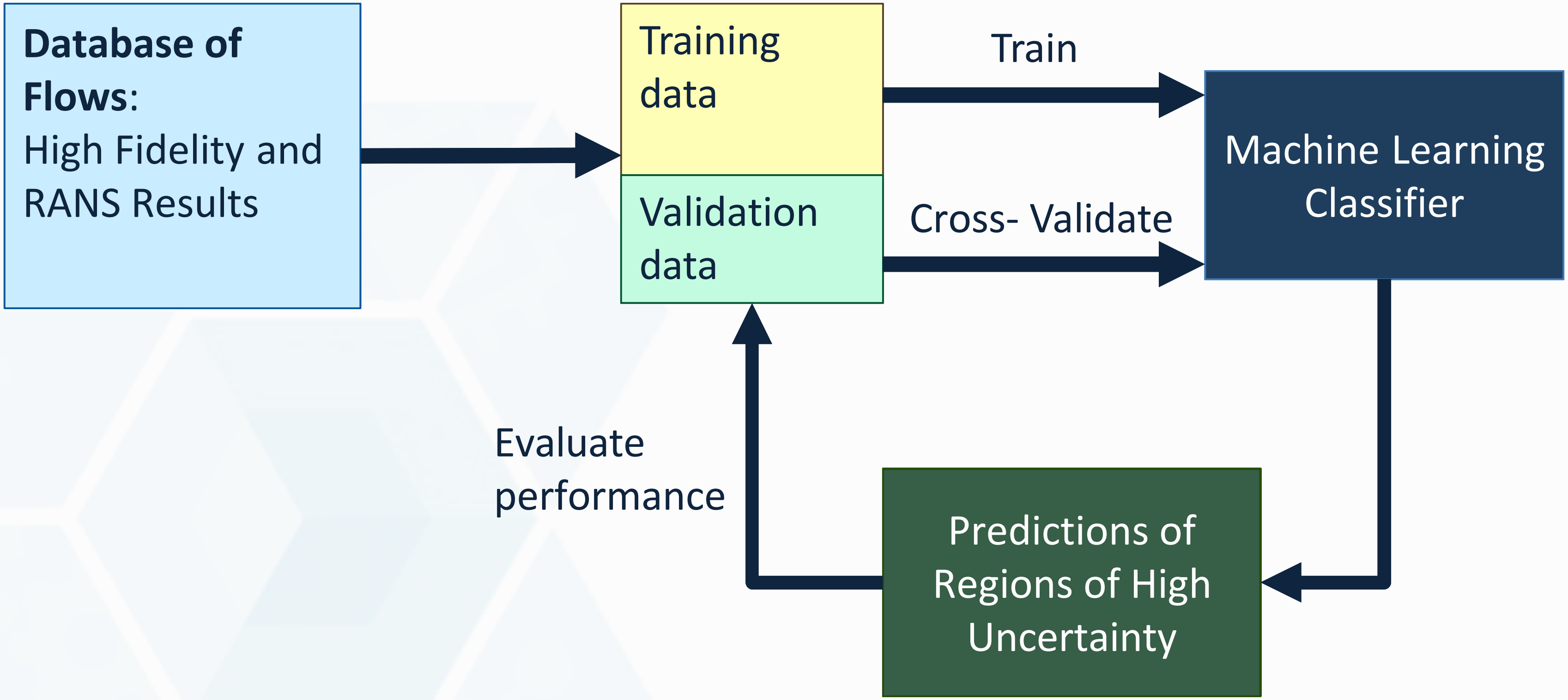
Reynolds Averaged Navier Stokes (RANS)

- Orders of magnitude less computationally expensive
- Approximate

Machine Learning Problem Set-Up

- Assemble a database of flows for which both DNS and RANS results are available
- Use the DNS as "truth" data for supervised machine learning
- Build a binary classifier to predict when RANS has high uncertainty based on when specific model assumptions are violated
- Inputs are local flow variables from RANS (velocity, pressure, density, …)
    - Each point in the mesh represents a separate data point

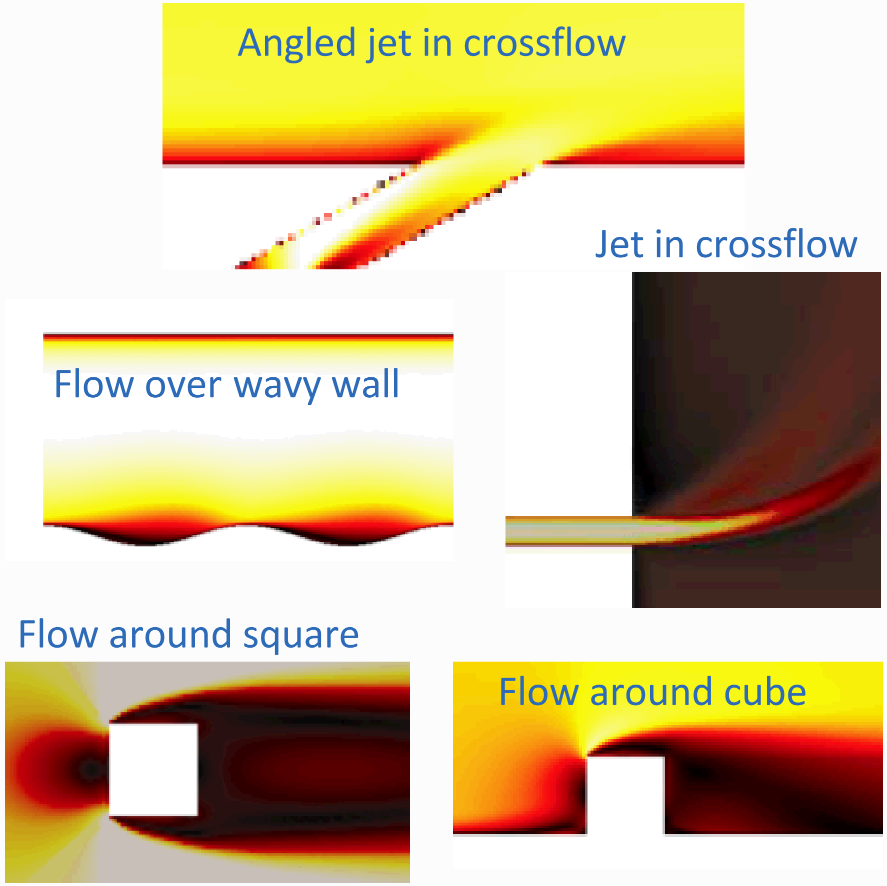
Classifier Development



Classifier Development

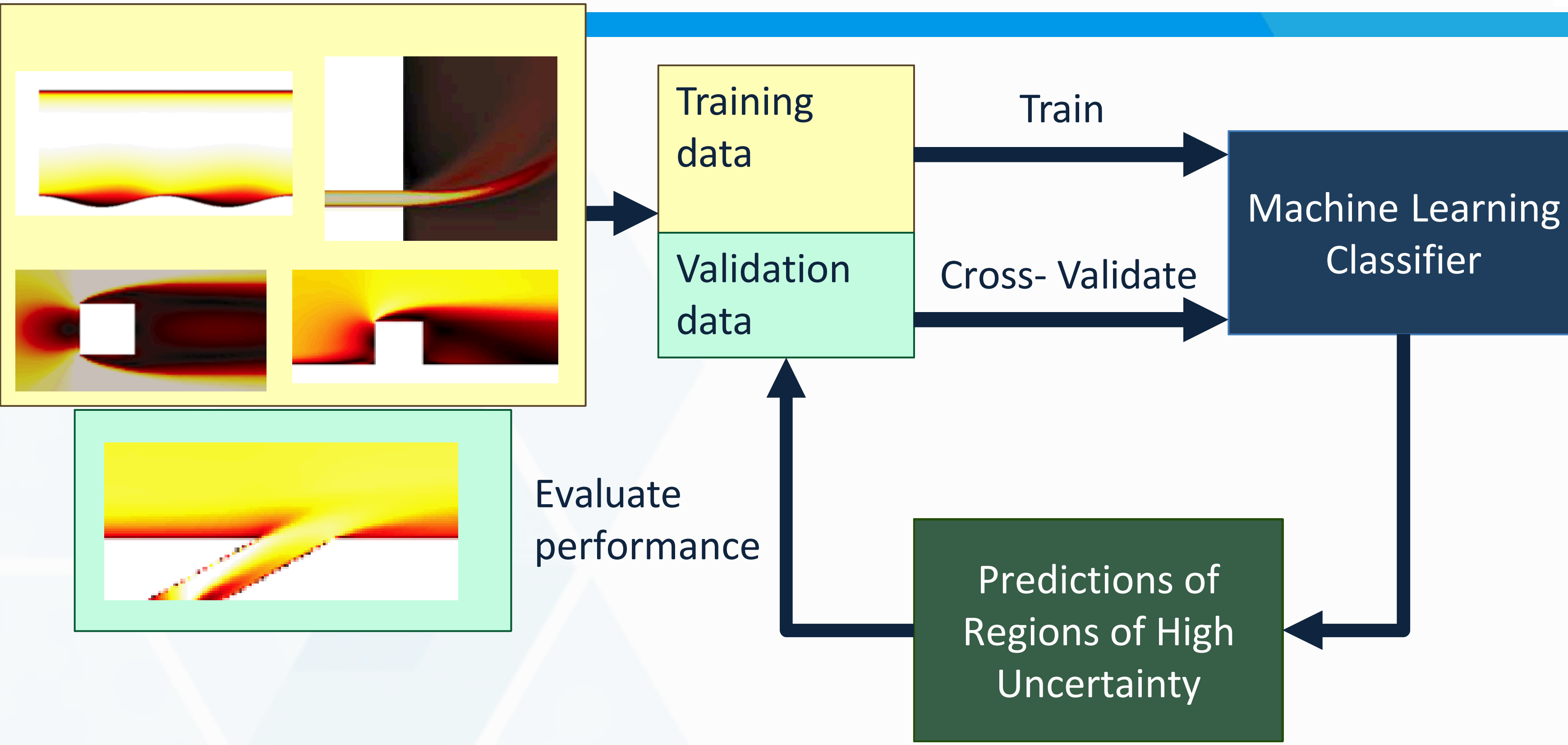
Database of Flows:
High Fidelity and RANS Results

Contours of velocity magnitude

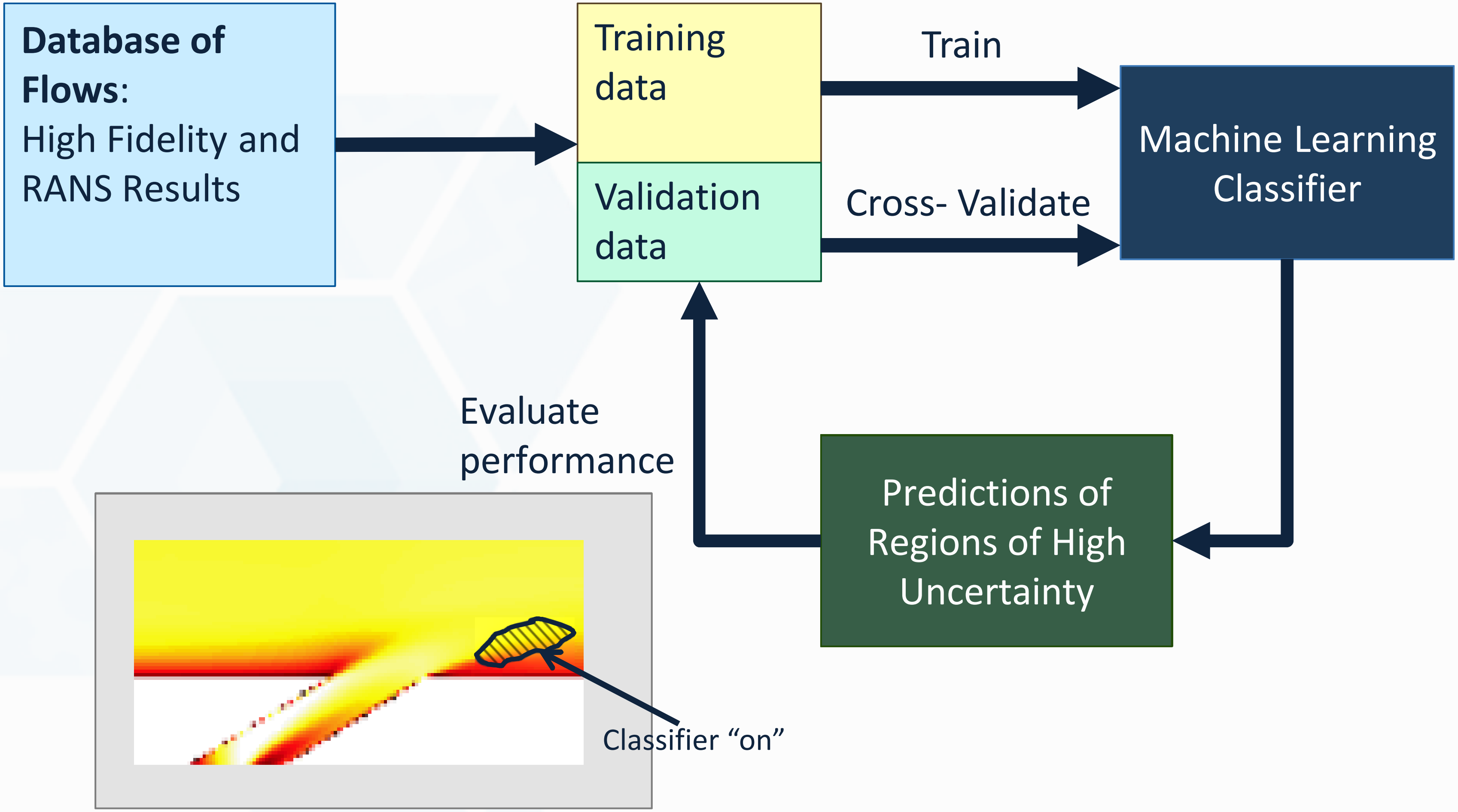


hine Learning Classifier

Classifier Development

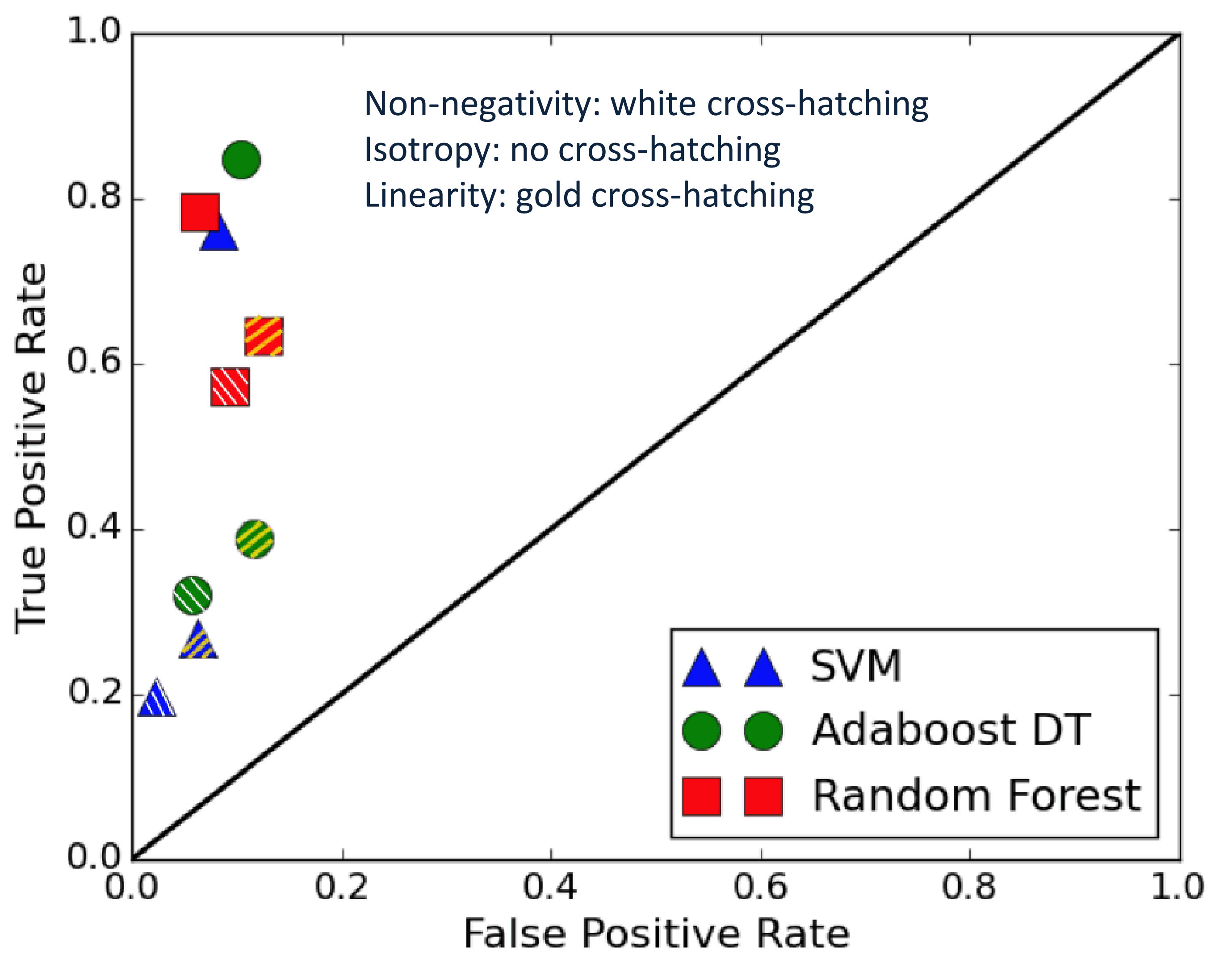


Classifier Development

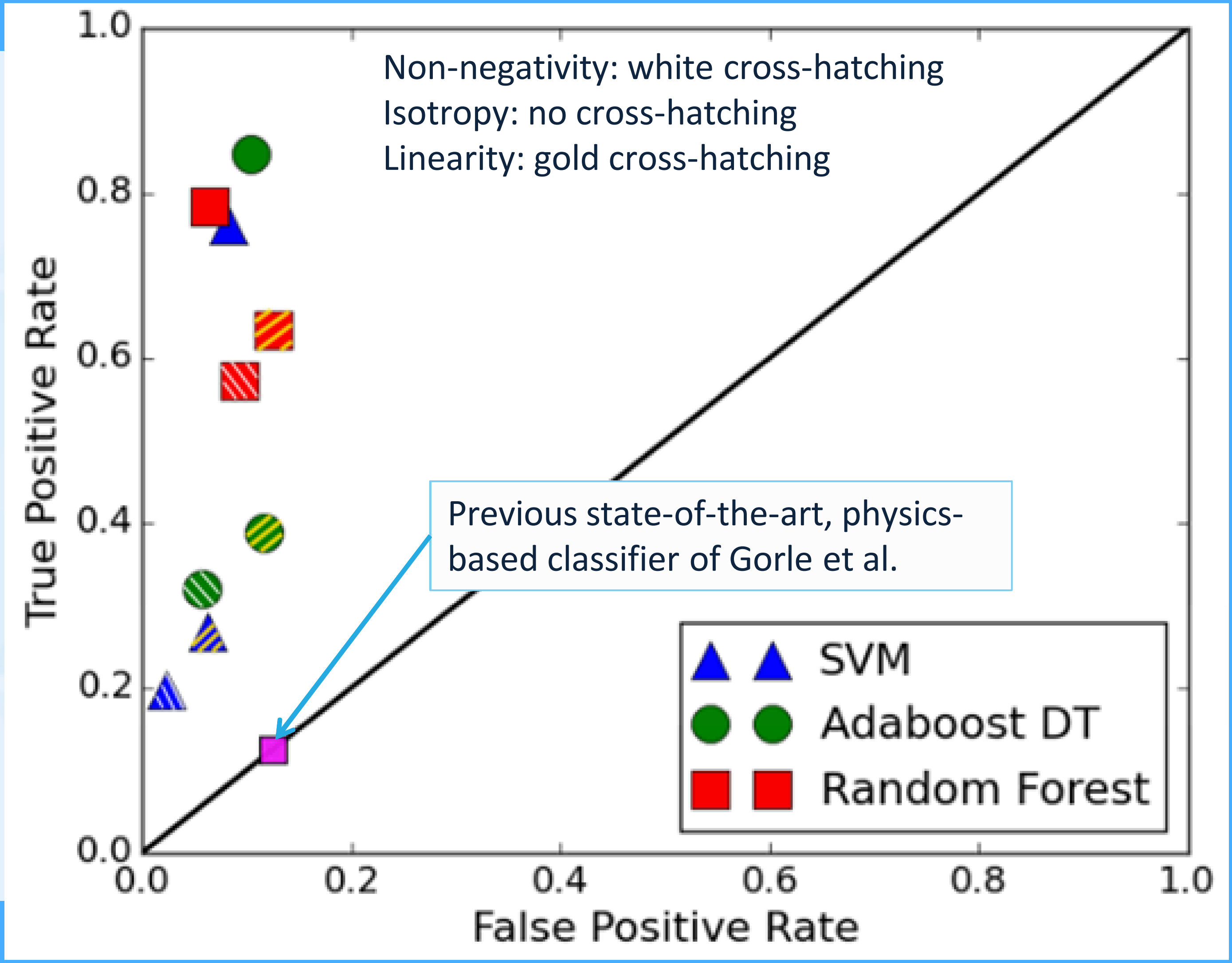


- Use classifier to make predictions on validation set
- Evaluate classifier by comparing to high fidelity results
- Leave-one-out cross-validation

Classifier Performance



Classifier Performance



Outline

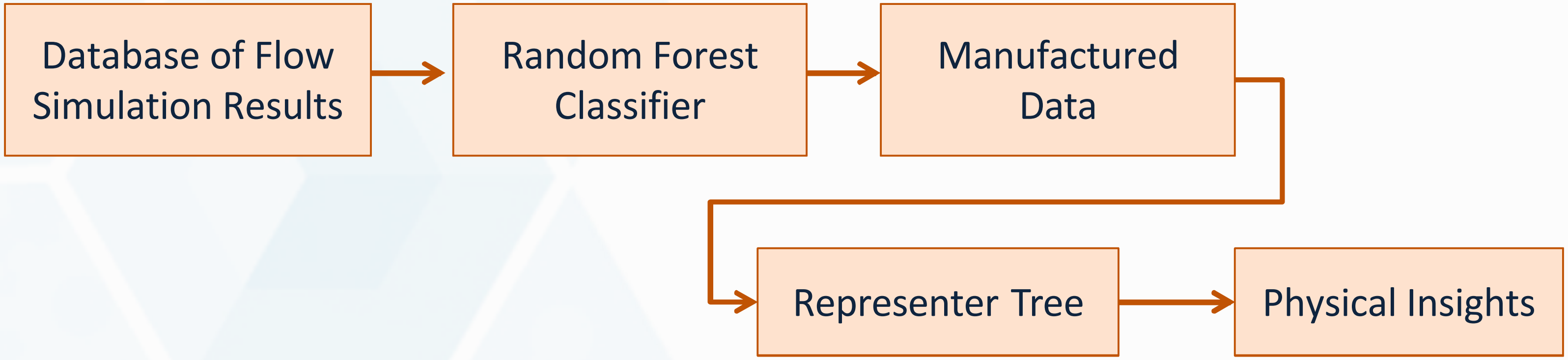
1) Machine learning for turbulence model form uncertainty quantification
   - Achieved 3X more accurate error detection than previous state of the art

2) Rule extraction to gain physical intuition

1) Deep learning with embedded invariances

Rule Extraction

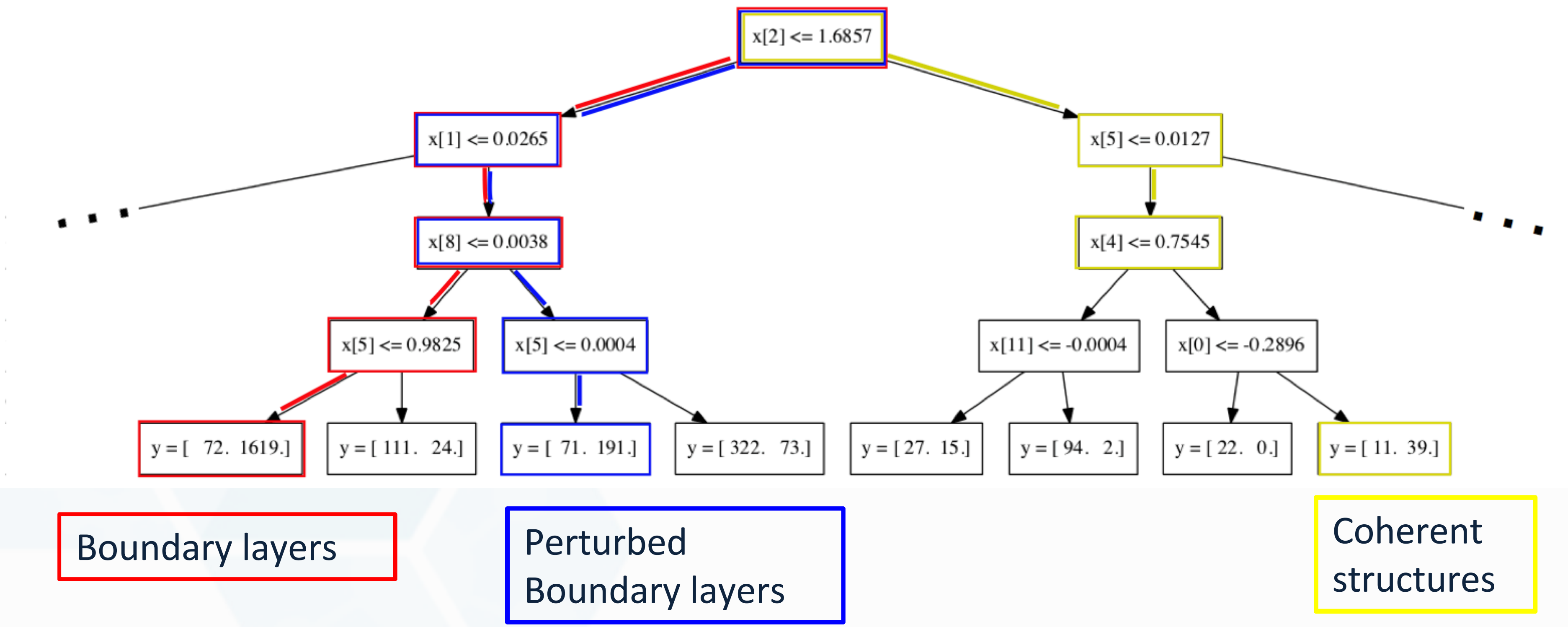
- Random Forests are much more robust and high-performance than single decision trees, but what have we lost?
  - Clarity—how can we understand these machine learned models?
- Representer Trees



- Trained a representer decision tree based on Random Forest that predicted when the RANS isotropy assumption was invalid

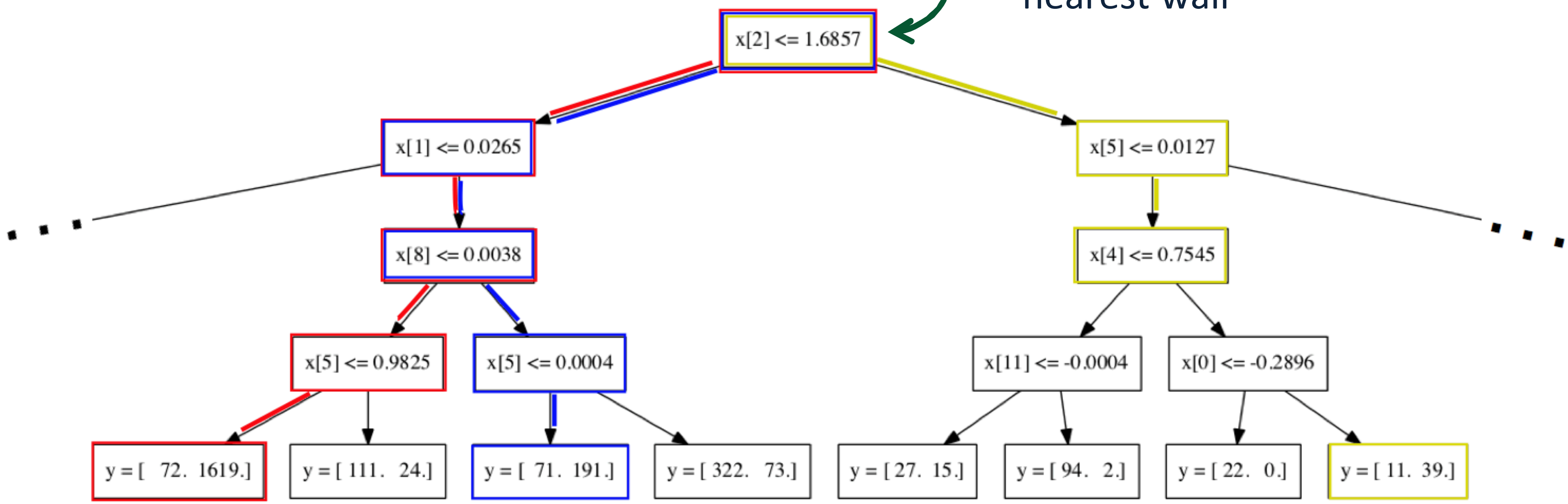
Analyzing the Representer Tree

- Look for consistent branches



Analyzing the Representer Tree

First split is on x[2], which is non-dimensional distance to the nearest wall

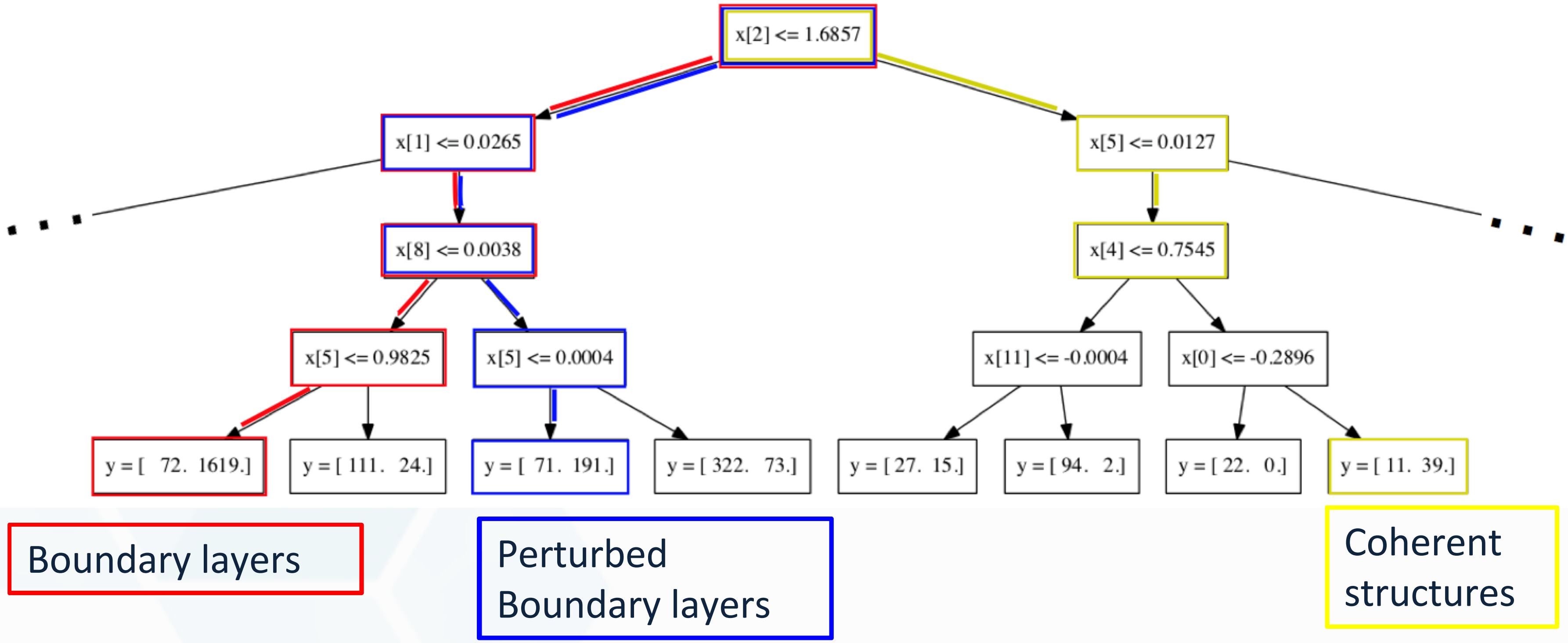


Boundary layers

Perturbed Boundary layers

Coherent structures

Analyzing the Representer Tree



- Can determine physical regimes where assumptions are violated
- Can see that different mechanisms cause assumption to break down in near wall region than in free stream

Outline

1) Machine learning for turbulence model form uncertainty quantification
   - Achieved 3X more accurate error detection than previous state of the art

2) Rule extraction to gain physical intuition
   - Used rule extraction to understand how the complex machine learning algorithm was making its predictions
   - Increases confidence in ML predictions
   - Provides feedback to turbulence modeler

3) Deep learning with embedded invariances

Deep Learning for Turbulence Modeling

- In RANS, use simplifying assumptions to get computational efficiency
  - Need model for unknown term: the Reynolds stress anisotropy tensor **A**
- Default model: Linear Eddy Viscosity Model
  - Based on theory + sparse experimental data
- Our approach: Deep neural network
- Inputs: Mean strain rate tensor **S**, mean rotation rate tensor **R**
- Outputs: Reynolds stress anisotropy **A**

W. P. JONES and B. E. LAUNDER

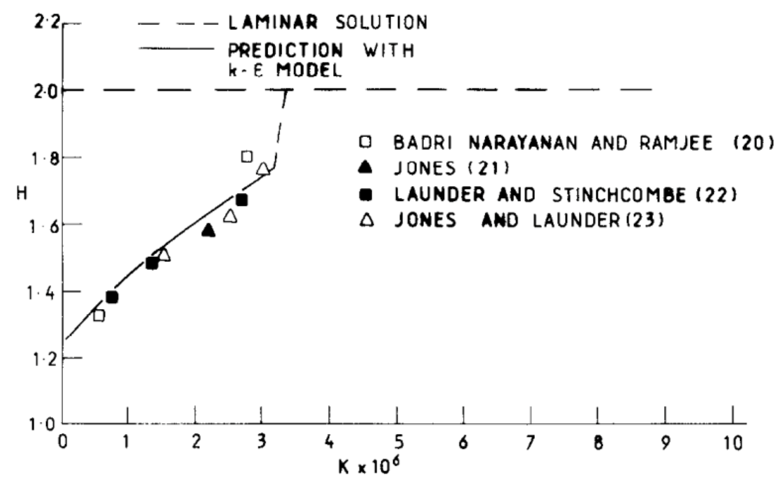


FIG. 2. Sink flow turbulent boundary layers.

(1973)

Deep Learning for Turbulence Modeling

- Inputs: Tensors **S**, **R**
- Output: Tensor **A**

- Would like to enforce Galilean invariance
  - Invariance to inertial coordinate frame transformations
  - Borrow some ideas from group theory, representation theory
  - All Galilean invariant tensors that are a function of **S** and **R** lie on a tensor basis: the *integrity basis* of **S** and **R** for the orthogonal group

$$\mathbf{A}(\mathbf{Q}\mathbf{S}\mathbf{Q}^T, \mathbf{Q}\mathbf{R}\mathbf{Q}^T) = \mathbf{Q}\mathbf{A}(\mathbf{S}, \mathbf{R})\mathbf{Q}^T$$

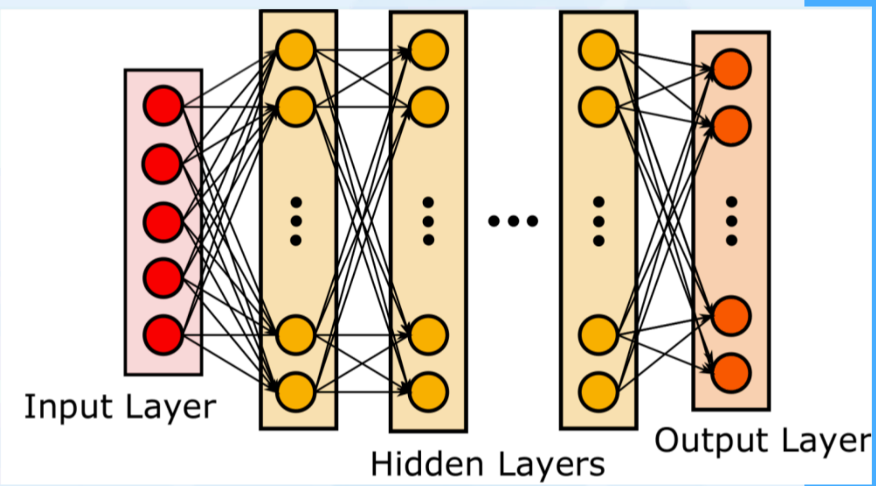
$$\mathbf{A} = \sum_{n=1}^{10} f^{(n)} \mathbf{B}^{(n)}$$

Unknown coefficients

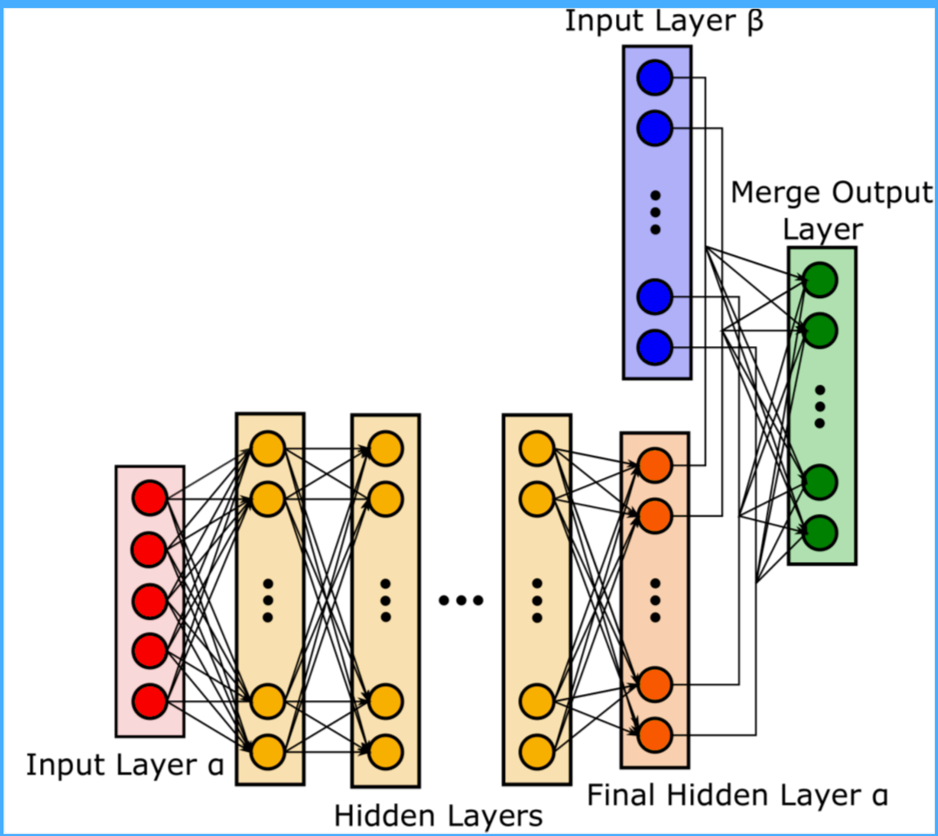
Known Tensor Basis

Embedding Galilean Invariance

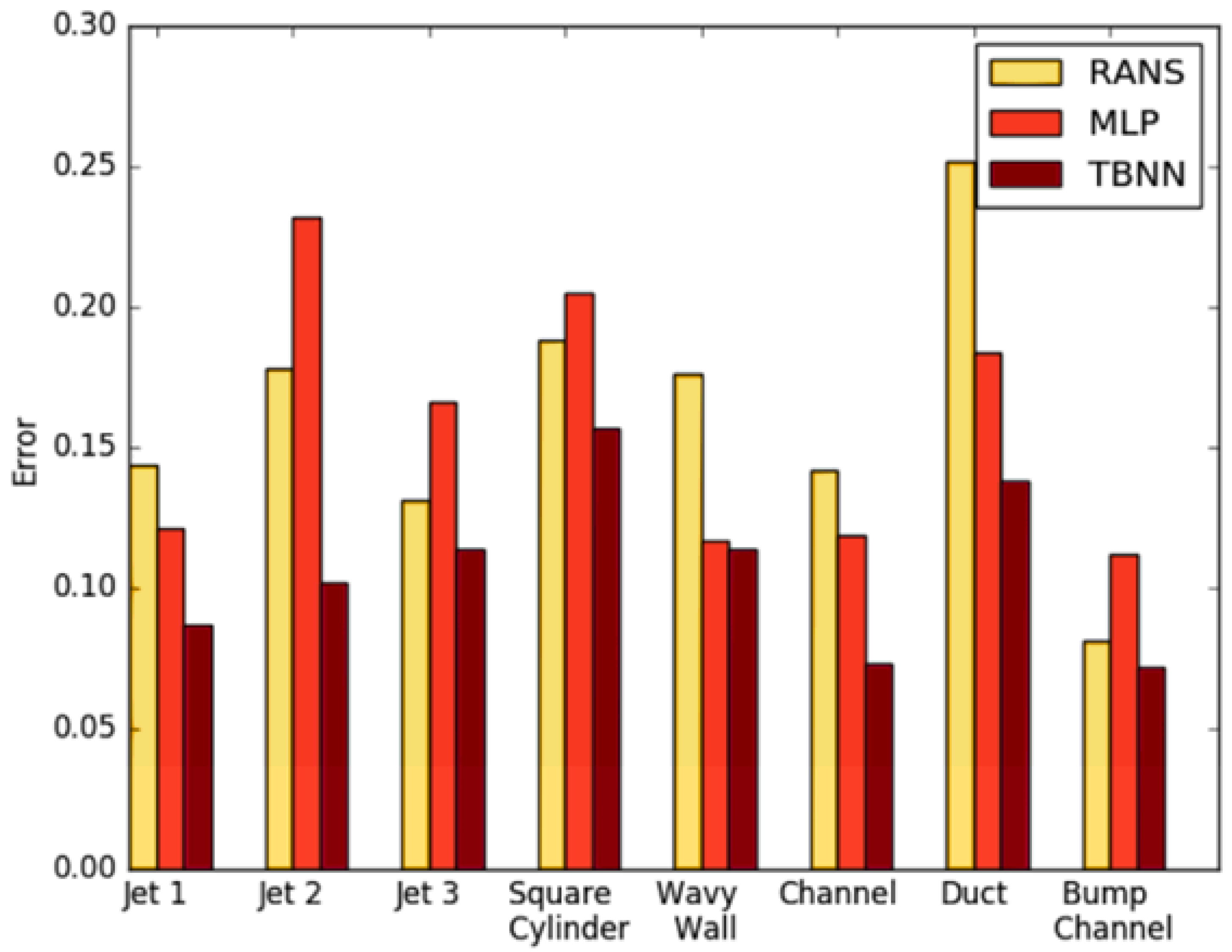
Multi-Layer Perceptron (MLP)



Tensor Basis Neural Network (TBNN)



Deep Learning for Turbulence Modeling



Outline

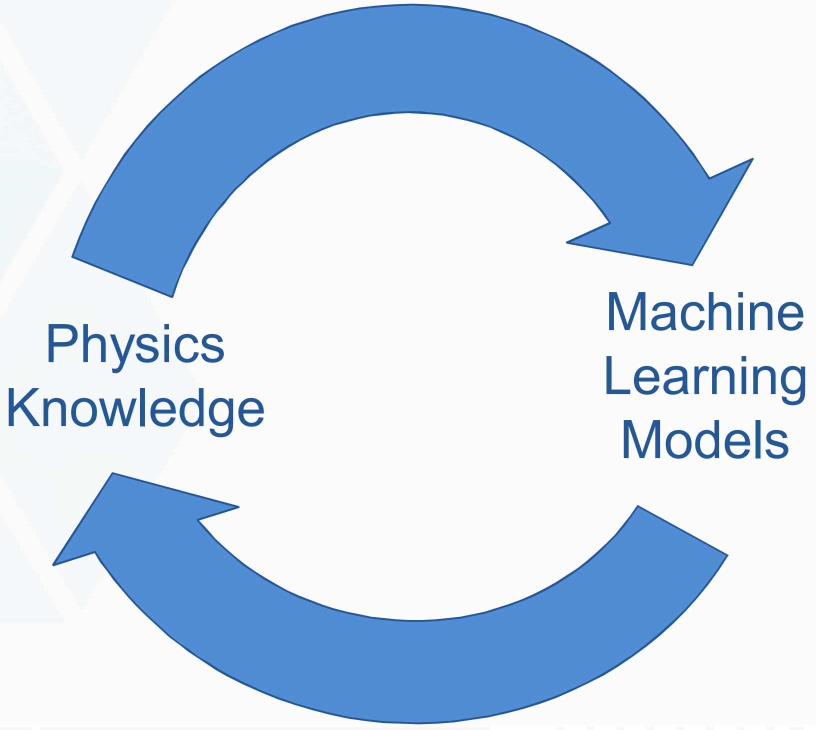
1) Machine learning for turbulence model form uncertainty quantification
   - Achieved 3X more accurate error detection than previous state of the art

2) Rule extraction to gain physical intuition
   - Used rule extraction to understand how the complex machine learning algorithm was making its predictions
   - Increases confidence in ML predictions
   - Provides feedback to turbulence modeler

3) Deep learning with embedded invariances
   - Embedded known invariance property into network architecture
   - Achieved significantly improved Reynolds stress anisotropy predictions
   - First application of deep learning to turbulence modeling

Big Picture

1. Directly embedding scientific domain knowledge into machine learning models can give improved performance, especially in data-limited scenarios
2. Rule extraction on machine learning models can give new physical understanding



References

- **J. Ling**, A. Kurzawski, and J. Templeton, “Reynolds Averaged Turbulence Modeling using Deep Neural Networks with Embedded Invariance,” Journal of Fluid Mechanics, (2016).
- **J. Ling** and J. Templeton, “Evaluation of machine learning algorithms for prediction of regions of high Reynolds averaged Navier Stokes uncertainty,” Physics of Fluids, (2015).
- **J. Ling**, “Using Machine Learning to Understand and Mitigate Model Form Uncertainty in Turbulence Models,” IEEE ICMLA, (2015).
- **J. Ling**, A. Ruiz, G. Lacaze, and J. Oefelein, “Uncertainty analysis and data-driven model advances for a Jet-in-Crossflow,” Journal of Turbomachinery, (2016).
- **J. Ling**, R. Jones, and J. Templeton, “Machine Learning Strategies for Systems with Invariance Properties,” J. Comp. Phys., (2016).
- **J. Ling**, F. Coletti, S. Yapa, J. Eaton, “Experimentally informed optimization of turbulent diffusivity for a discrete hole film cooling geometry,” International Journal of Heat and Fluid Flow, (2013).

Acknowledgements

- Collaborators at Sandia: Jeremy Templeton, Andrew Kurzawski, Reese Jones

Questions