SAND2016-0030C

Exceptional service in the national interest

Sandia National Laboratories

SANDIA NATIONAL LABORATORIES

President Harry S. Truman Fellowship in National Security Science and Engineering

Machine Learning for Uncertainty Quantification in Turbulent Flow Simulations

Julia Ling, Jeremy Templeton

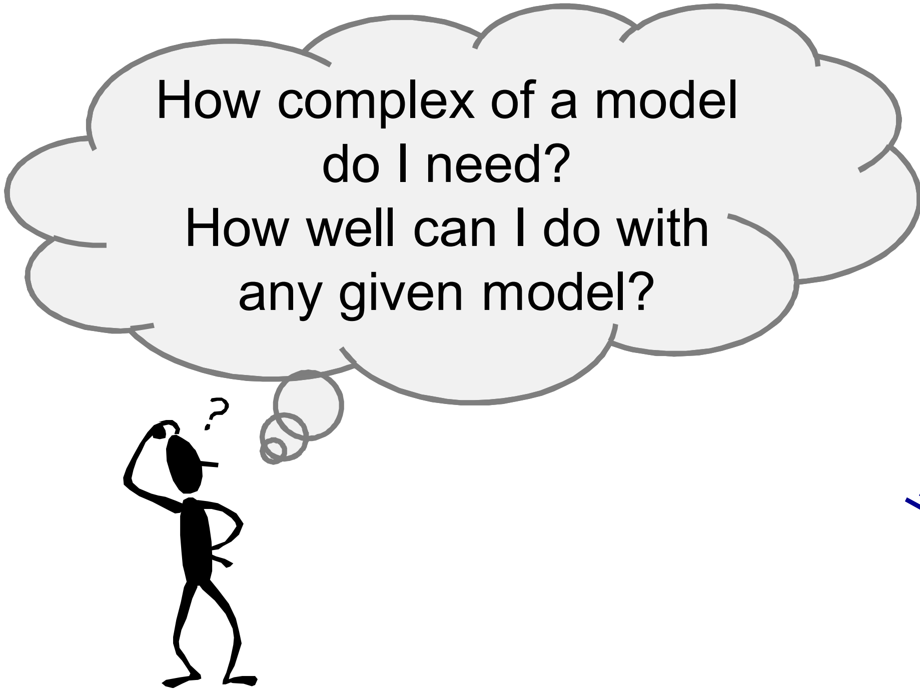
Apr 2016

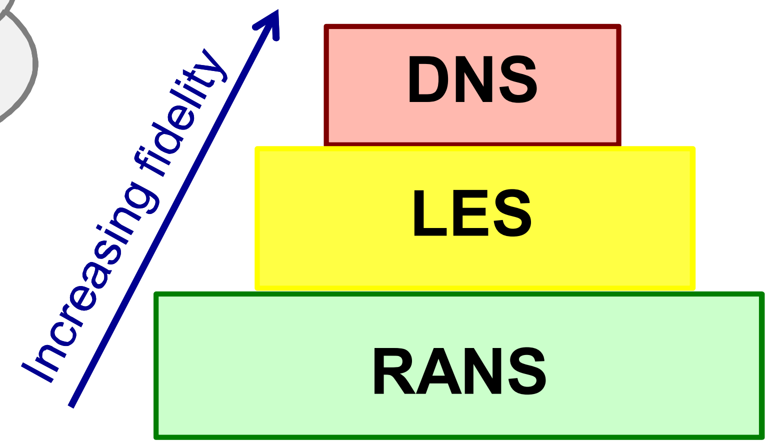
U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

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. SAND

Turbulence Simulations



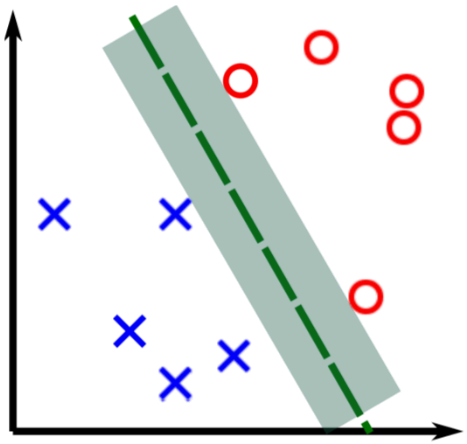


RANS

- Most widely used turbulence model
- Relies on modeling assumptions→ Model form uncertainty
- Sometimes accurate, sometimes inaccurate
  - Very difficult to assess model form uncertainty
- **Idea: Use machine learning to detect regions of high uncertainty based on when specific model assumptions are violated**

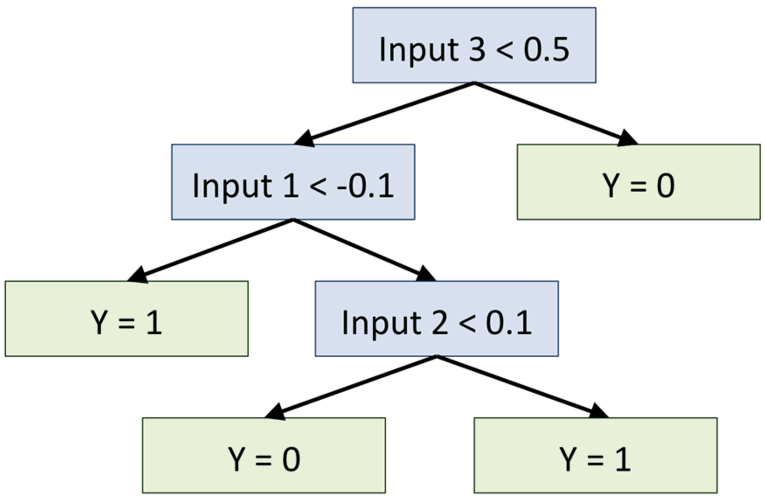
Machine Learning

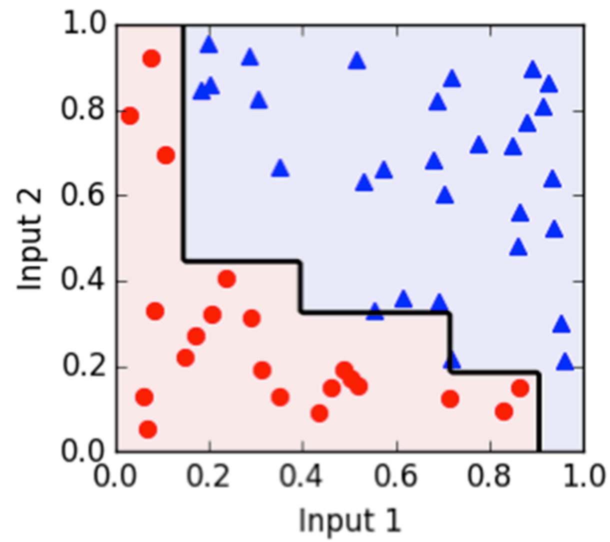
- Set of data-driven algorithms for regression, classification, clustering
- *E.g.*: linear regression, support vector machines, neural networks
- Have been broadly applied in finance, software engineering, retail
- Challenge: how to incorporate domain knowledge into machine learning algorithms
  - These techniques have a range of physics applications
- **For this application: use binary classifier to flag regions of high RANS uncertainty on a point-by-point basis**



Random Forests

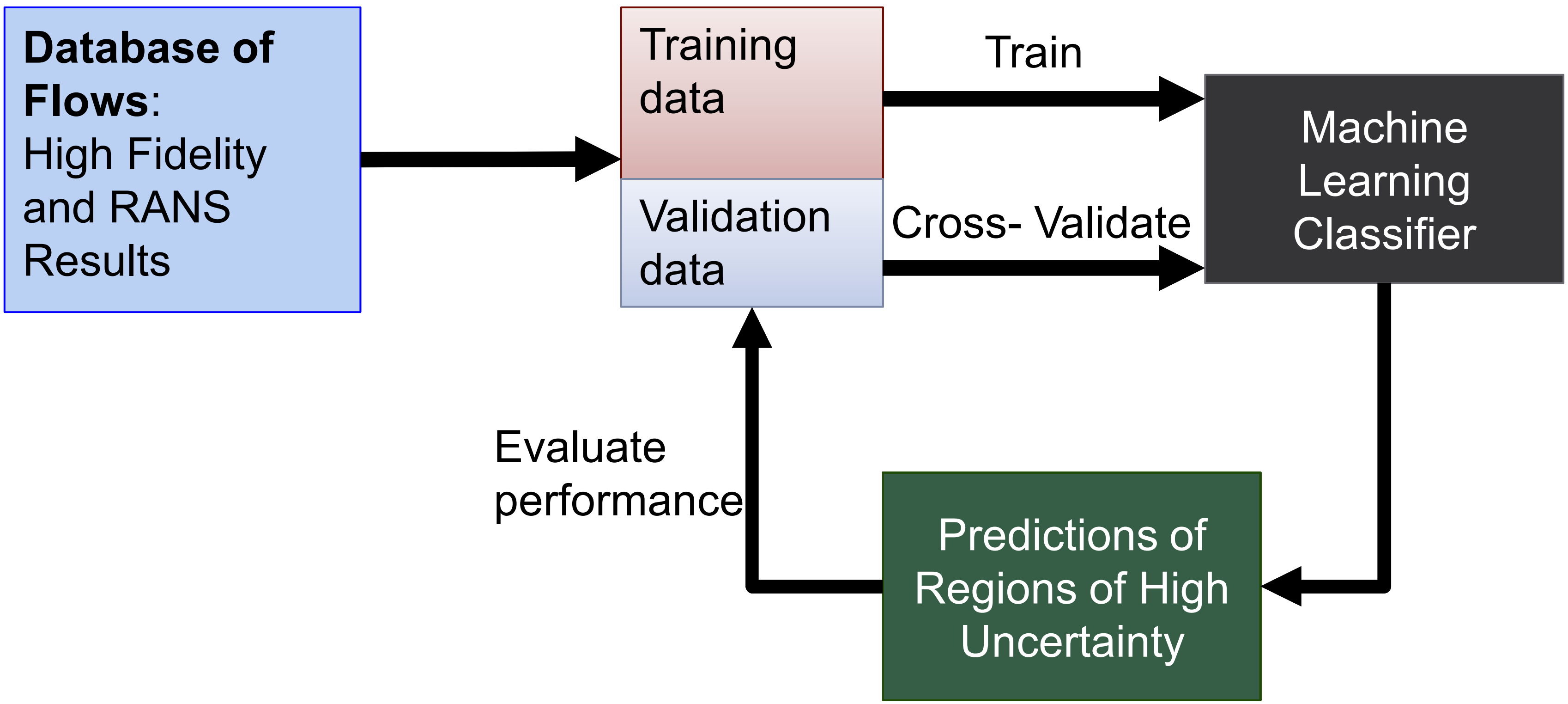
- Binary Decision Trees:
  - Simple, easy to understand and use
  - Tendency to overfit, poor performance with non-linear behavior





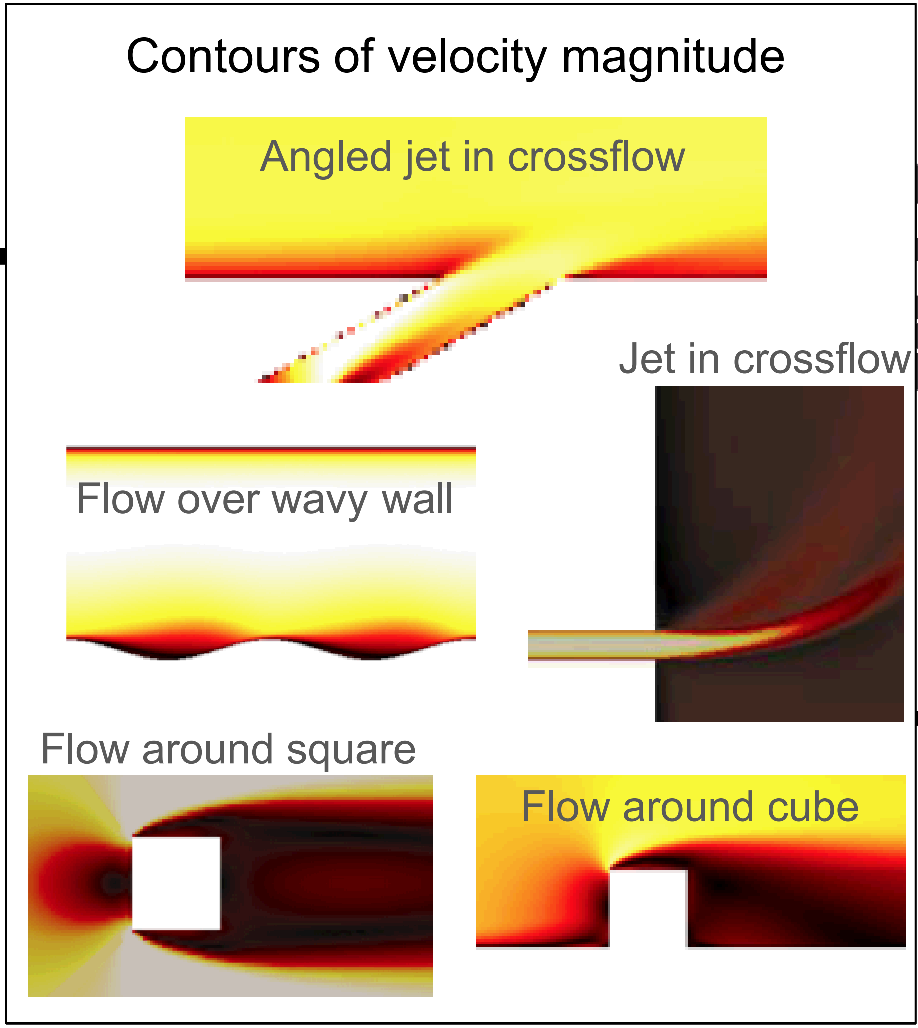
- Ensembles of Decision Trees:
  - Much more robust
  - Random Forests are a type of ensemble of decision trees

Classifier Development



Classifier Development

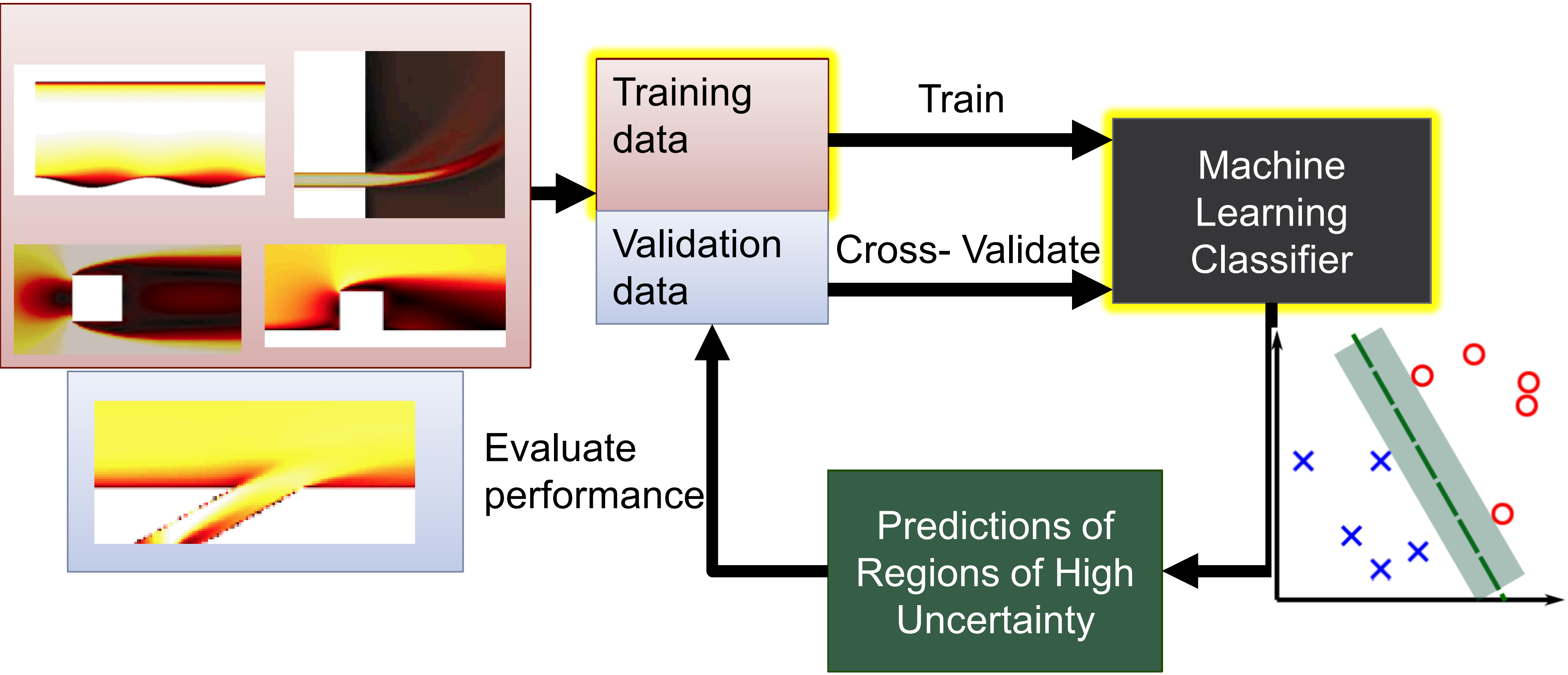
Database of Flows:
High Fidelity and RANS Results



Machine Learning Classifier

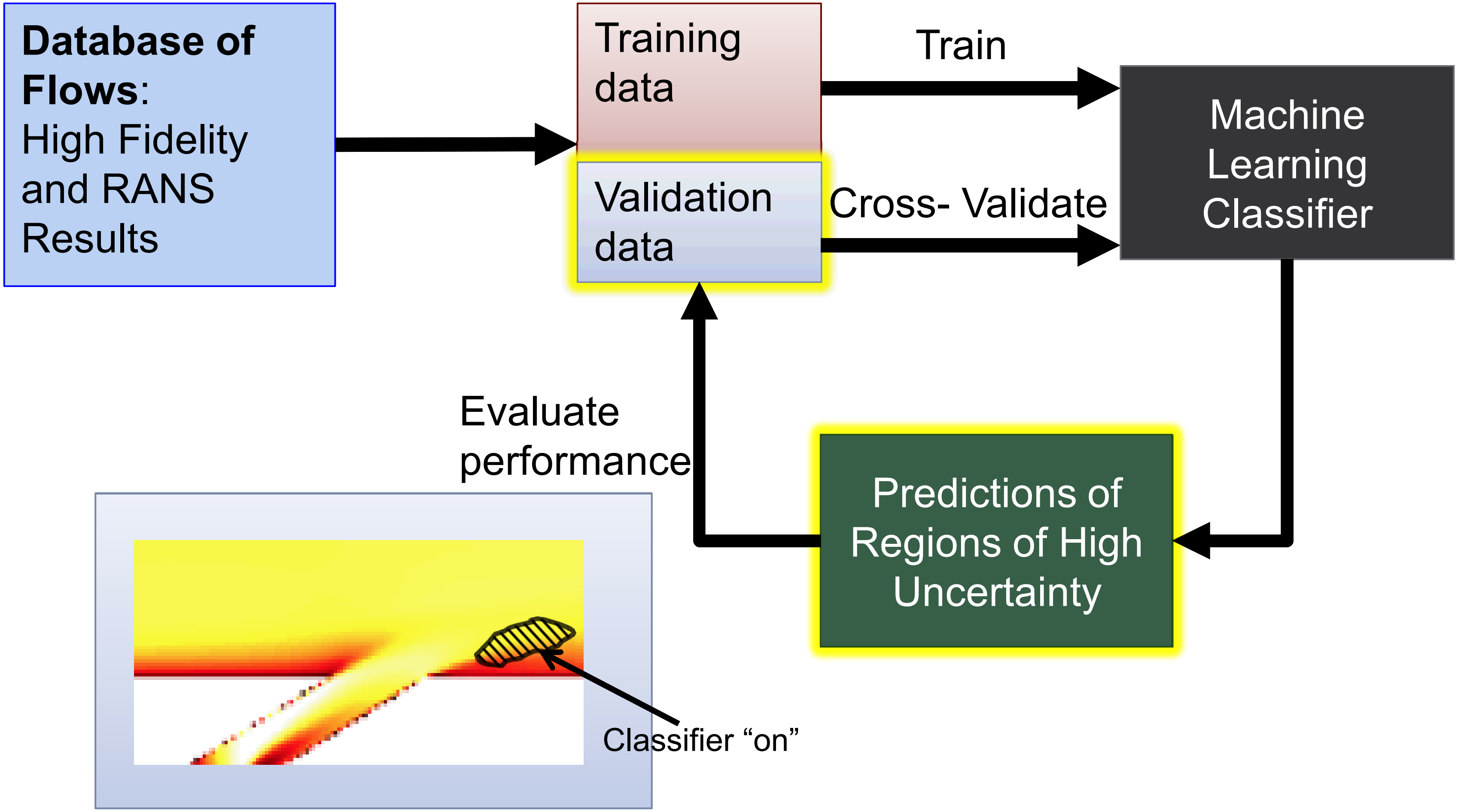
- Have database of canonical “building block” flows

Classifier Development



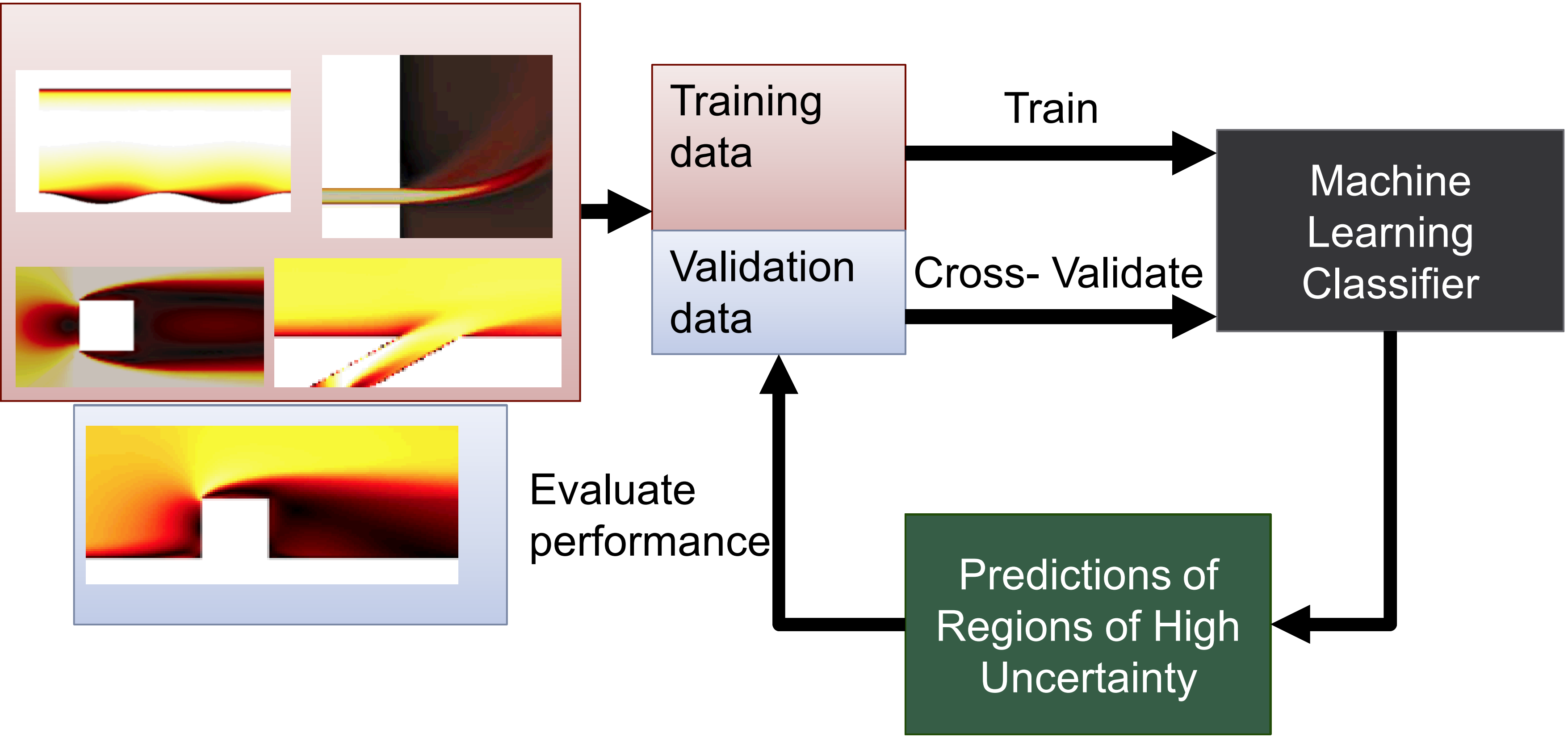
- Split data base into training and validation sets
- Train classifier
  - Input: Local flow variables from RANS
  - Output: Binary flag– “on” if RANS assumption violated, “off” otherwise

Classifier Development



- Use classifier to make predictions on validation set
- Evaluate classifier by comparing to high fidelity results

Classifier Development



- Cross-validate to ensure generalization

Assumptions Tested

$$\overline{u_i'u_j'} = \frac{2}{3}k\delta_{ij} - 2\nu_t S_{ij}$$

1. Non-negativity of eddy viscosity
   - Can extract eddy viscosity from LES/DNS
     $$\nu_t = \frac{-\overline{u_i'u_j'}S_{ij} + \frac{2}{3}k\delta_{ij}S_{ij}}{2S_{kl}S_{kl}}$$
   - Classifier should be "on" when LES/DNS eddy viscosity goes negative
2. Isotropy of Reynolds stresses
   - Classifier should be "on" when second invariant of anisotropy tensor exceeds a set threshold
3. Linearity of Boussinesq hypothesis
   - Extract linear and cubic eddy viscosity from LES/DNS data
   - If these values differ significantly, then uncertainty associated with linearity assumption is high

Inputs:

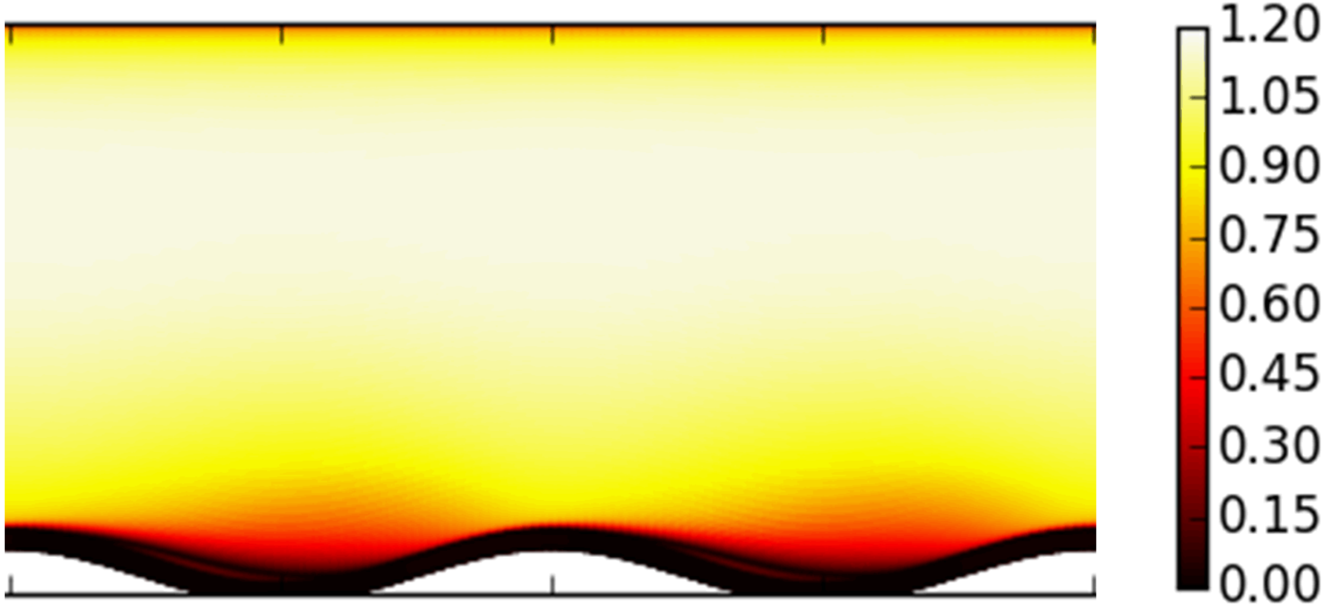
- Non-dimensional, rotationally invariant local flow variables from RANS

Applications of Classifiers

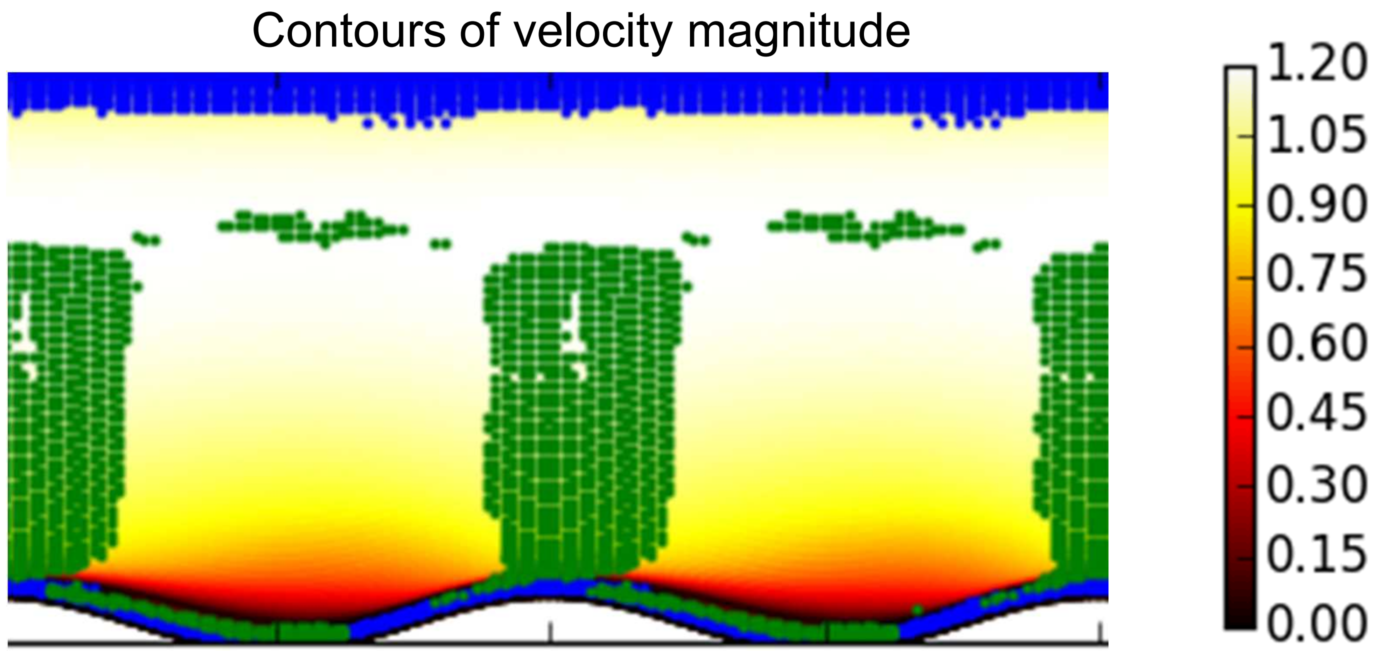
- Can quickly post-process RANS simulation to determine whether it's reliable in region of interest
  - Don't have to wait around for validation data set
  - Can determine what corrections to implement
- Can enable adaptive corrections during run time
- Experimental design
  - Design experiments to provide the strongest validation
- LES-RANS hybrids
  - Use classifiers to inform switching function

The Status Quo

Contours of velocity magnitude

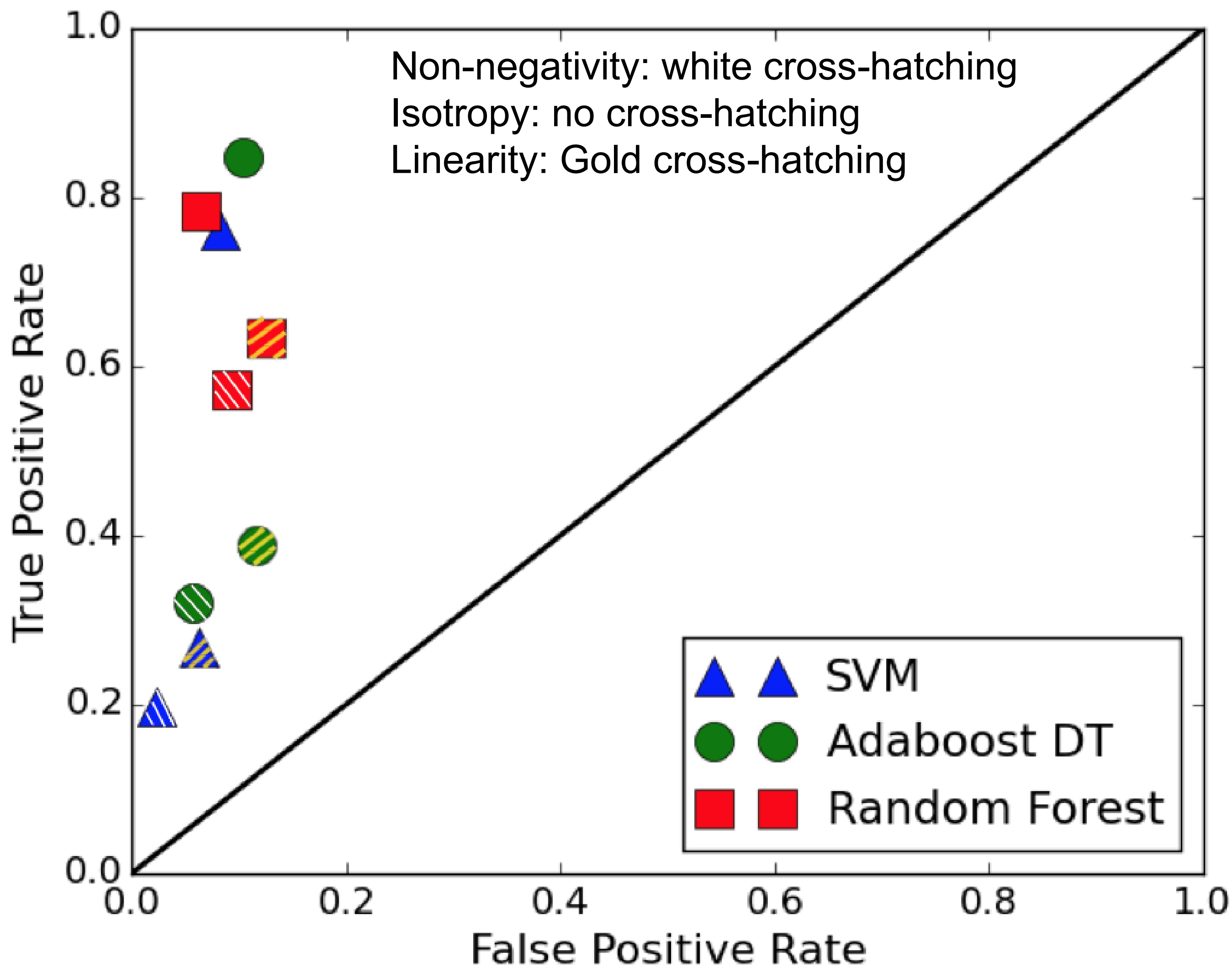


A Better Option

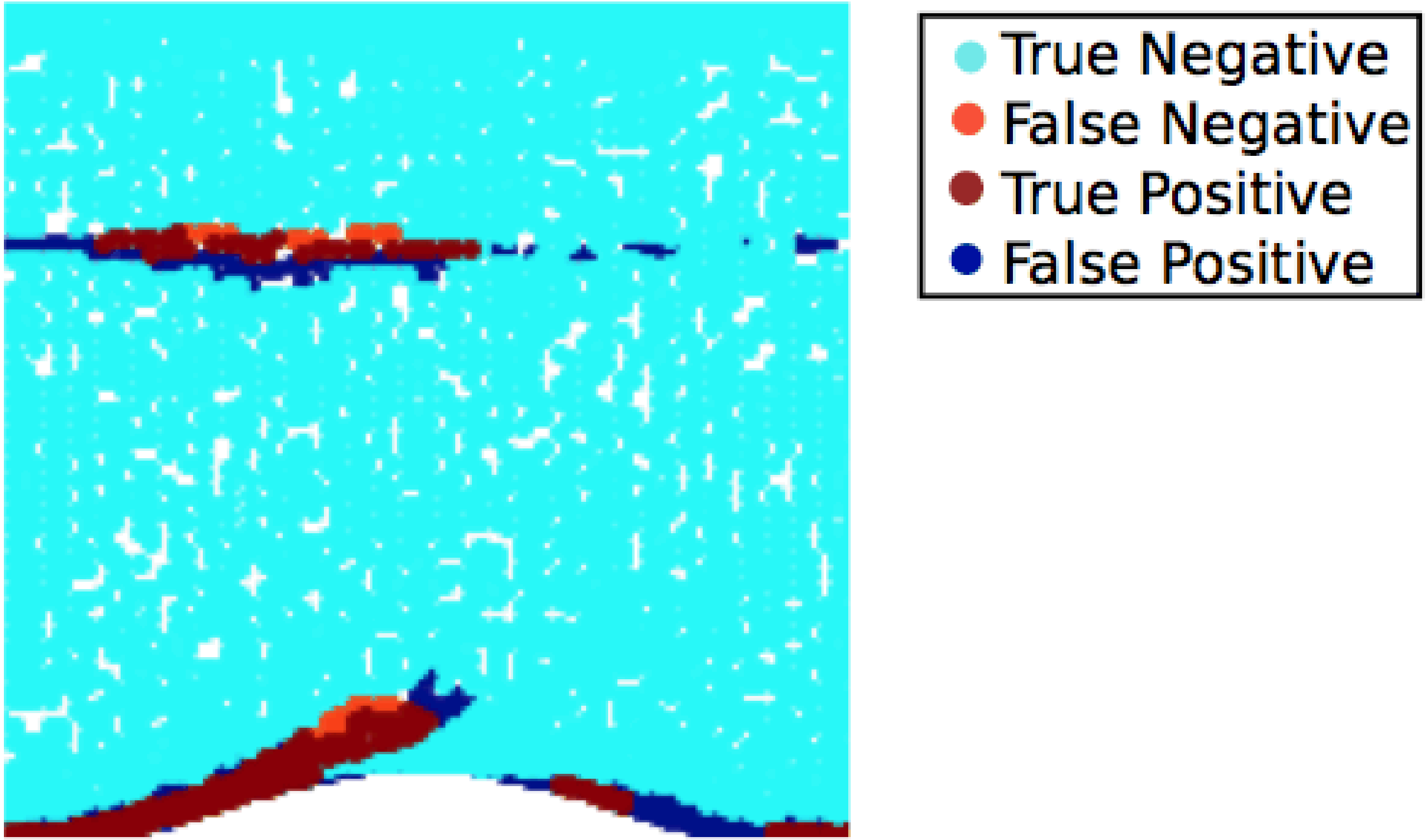


Blue: Regions where classifier predicts isotropy assumption violated
Green: Regions where classifier predicts linearity assumption violated

Classifier Performance



Classifier Performance



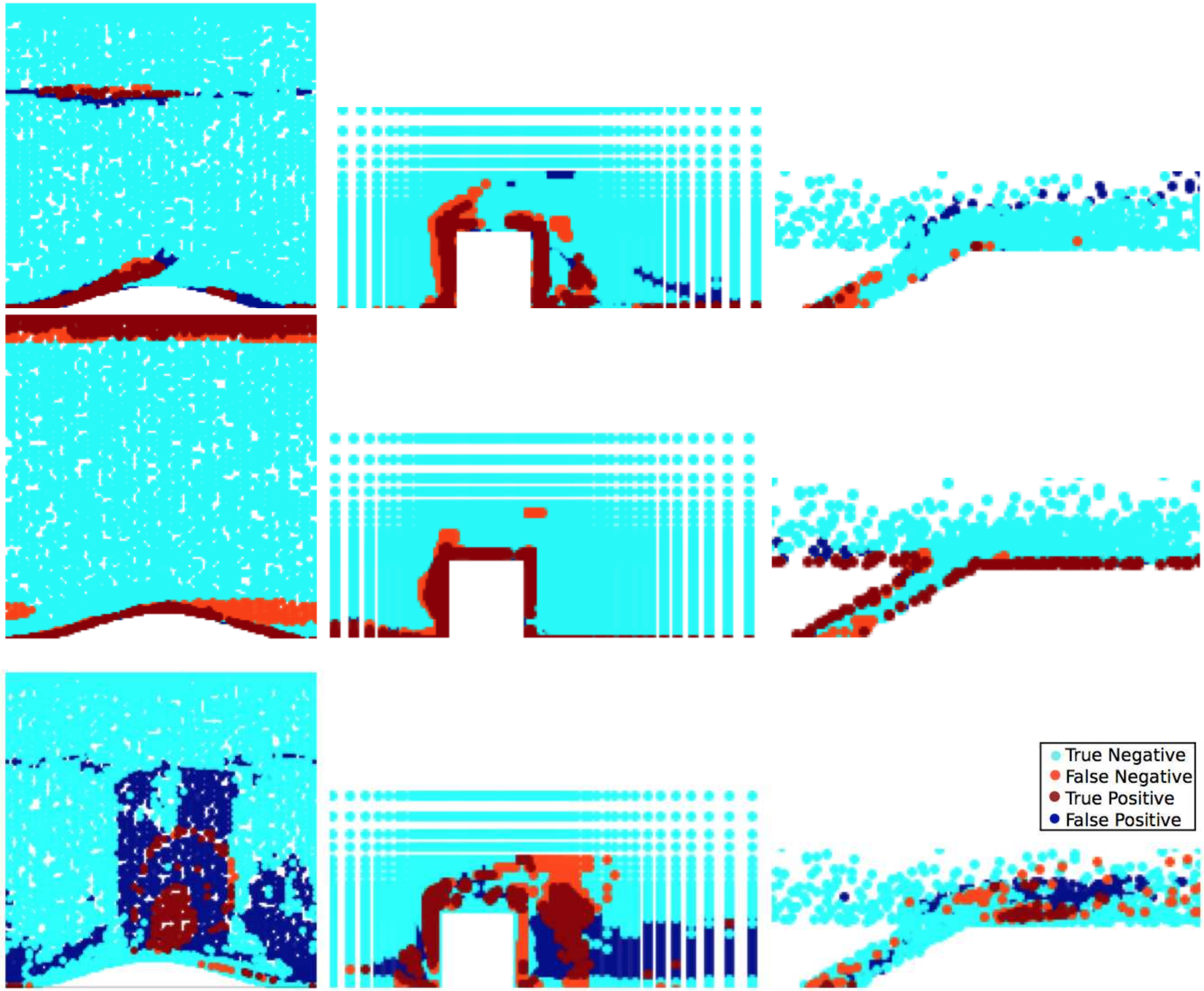
(a) Case 1, Marker 1:
Negative ν_t

Classifier Performance

Non-negativity assumption

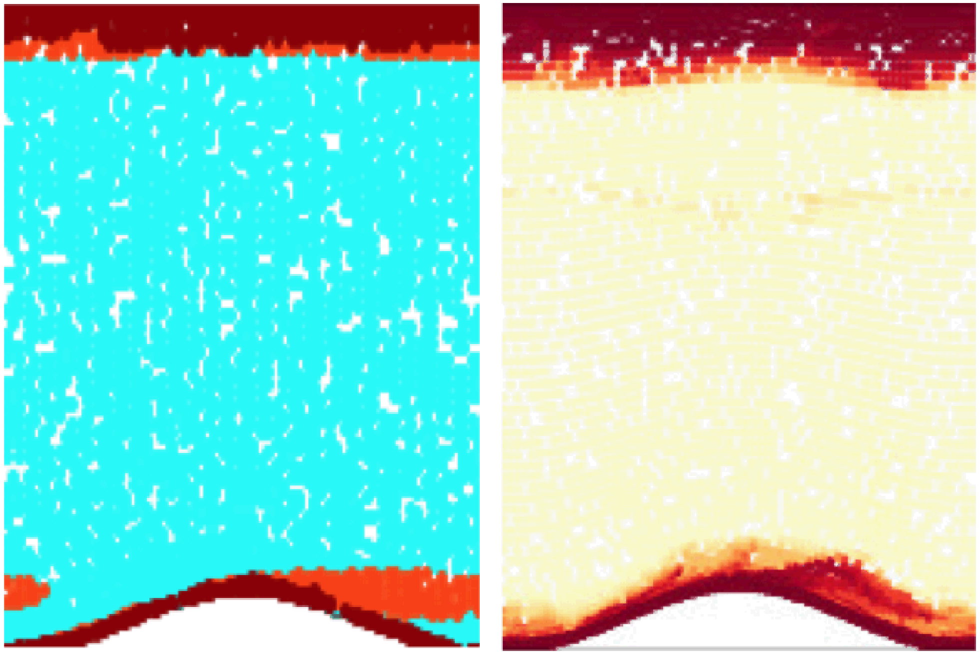
Isotropy assumption

Linearity assumption

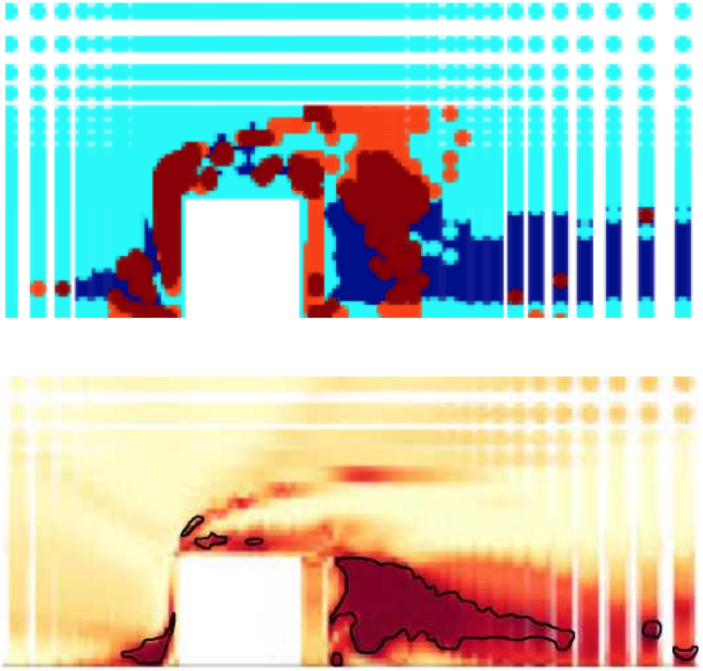


Classifier Confidence

Uncertain classifications

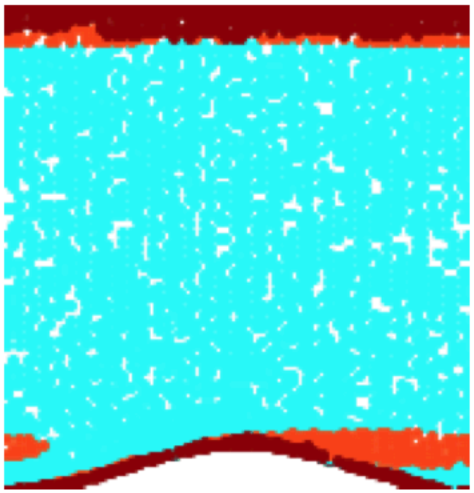


Extrapolation Detection



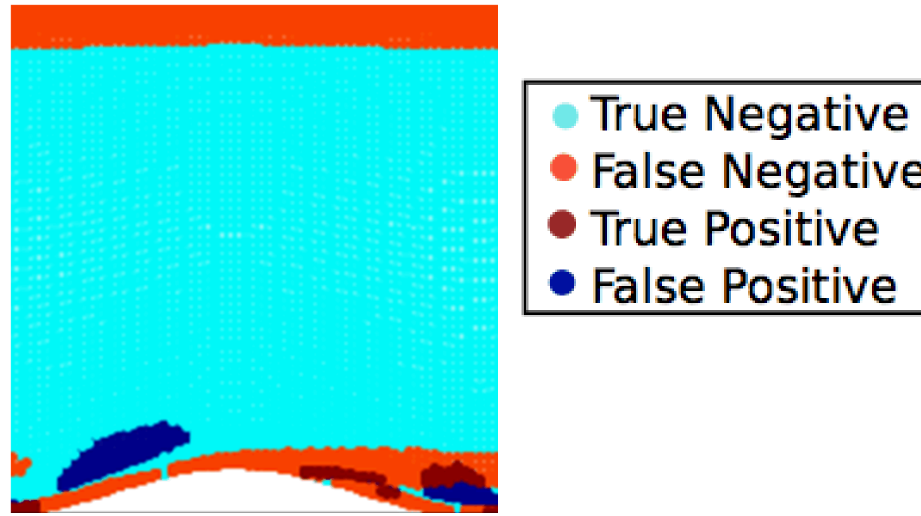
Comparison against State of the Art

Machine Learned Classifier



Cross-validation
Error rate: 11%

Physics-based Classifier of Gorle et al.



Cross-validation
Error rate: 33%

- Cross-validated classifier error rate: 11%
- 3 X more accurate than current state of the art physics-driven classifier of Gorle et al.
  - Gorle et al.'s classifier is used as an input to the ML classifier

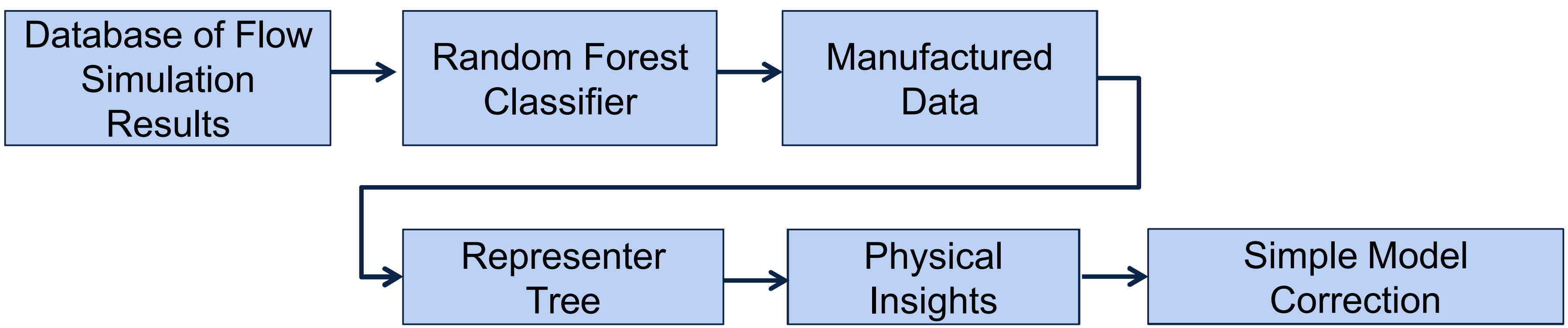
Impacts

- Classifiers for RANS model uncertainty can transform the way RANS results are post-processed and understood
  - Clarify when RANS simulations are predictive
  - Enable adaptive modeling corrections

- Develop techniques for using machine learning algorithms on physical systems
  - Leverage domain knowledge and physical constraints to develop smarter models
  - Use data-driven models to learn about the physical system

Rule Extraction

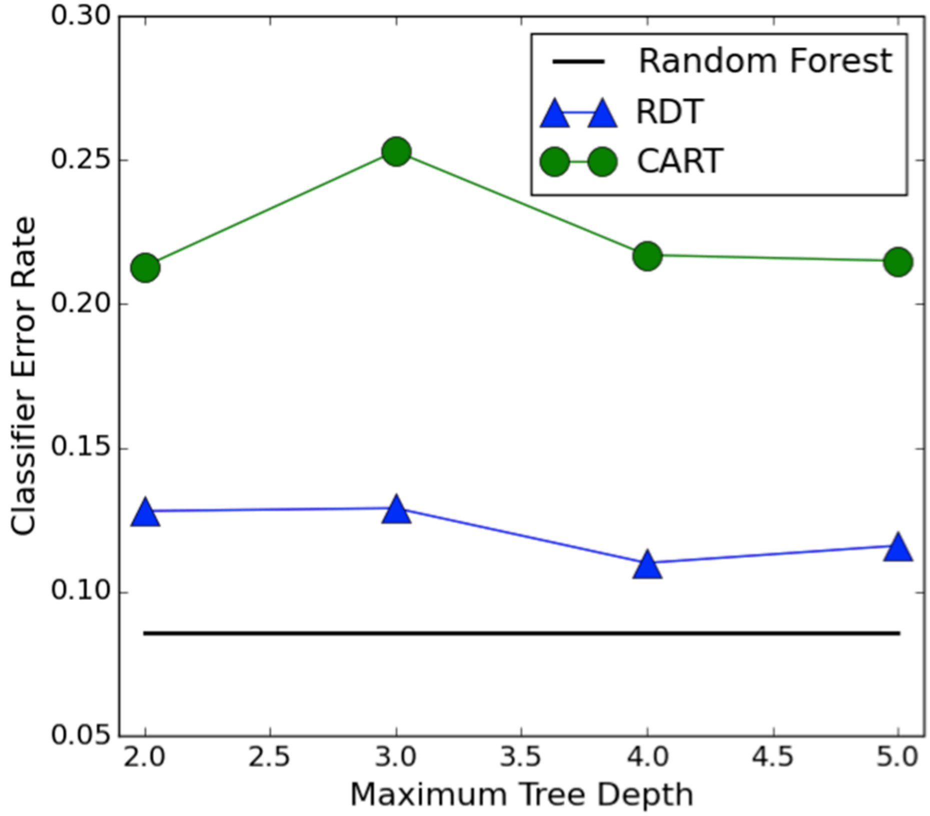
- Random Forests are much more robust and high-performance than single decision trees, but what have we lost?
  - Clarity—how can we understand these machine learned models?
- Representer Trees



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

Representer Decision Trees

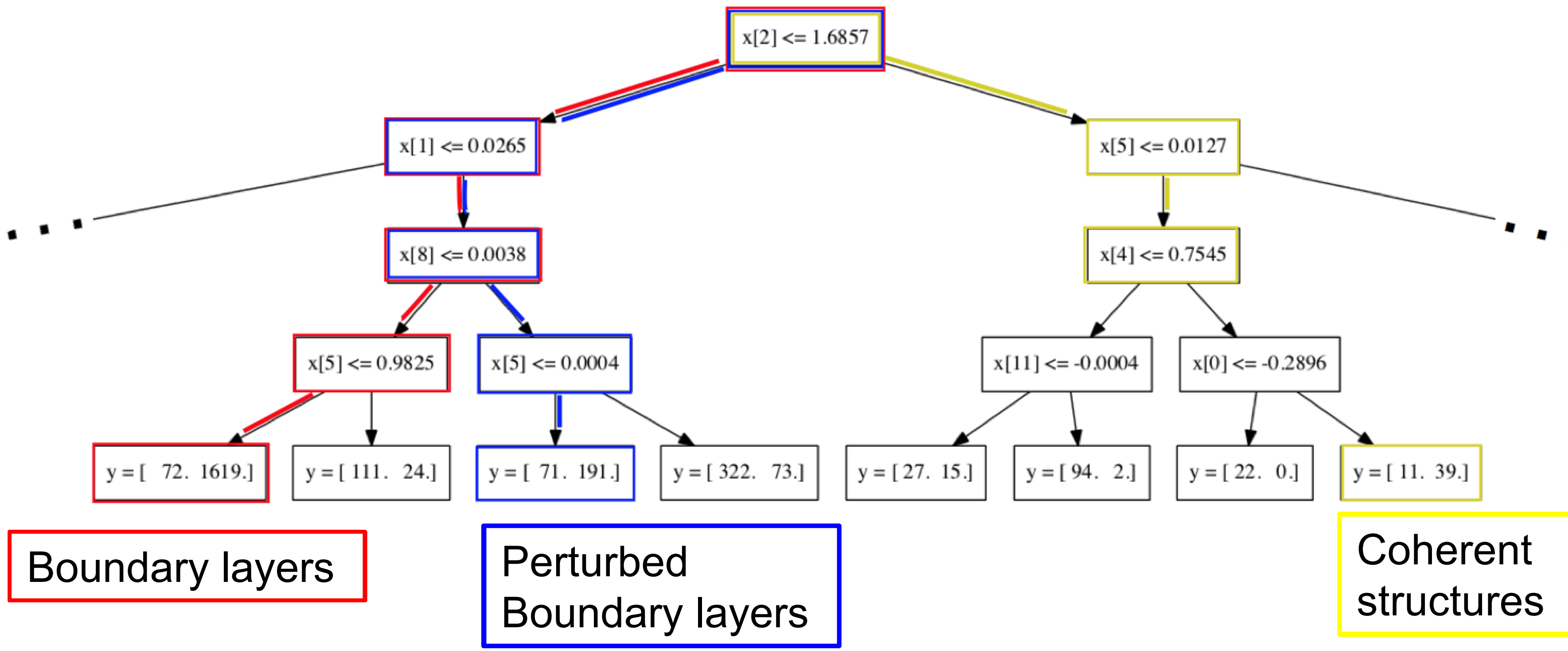
- Surprising result: the representer tree has better performance and is more stable than a tree trained on the original tra



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

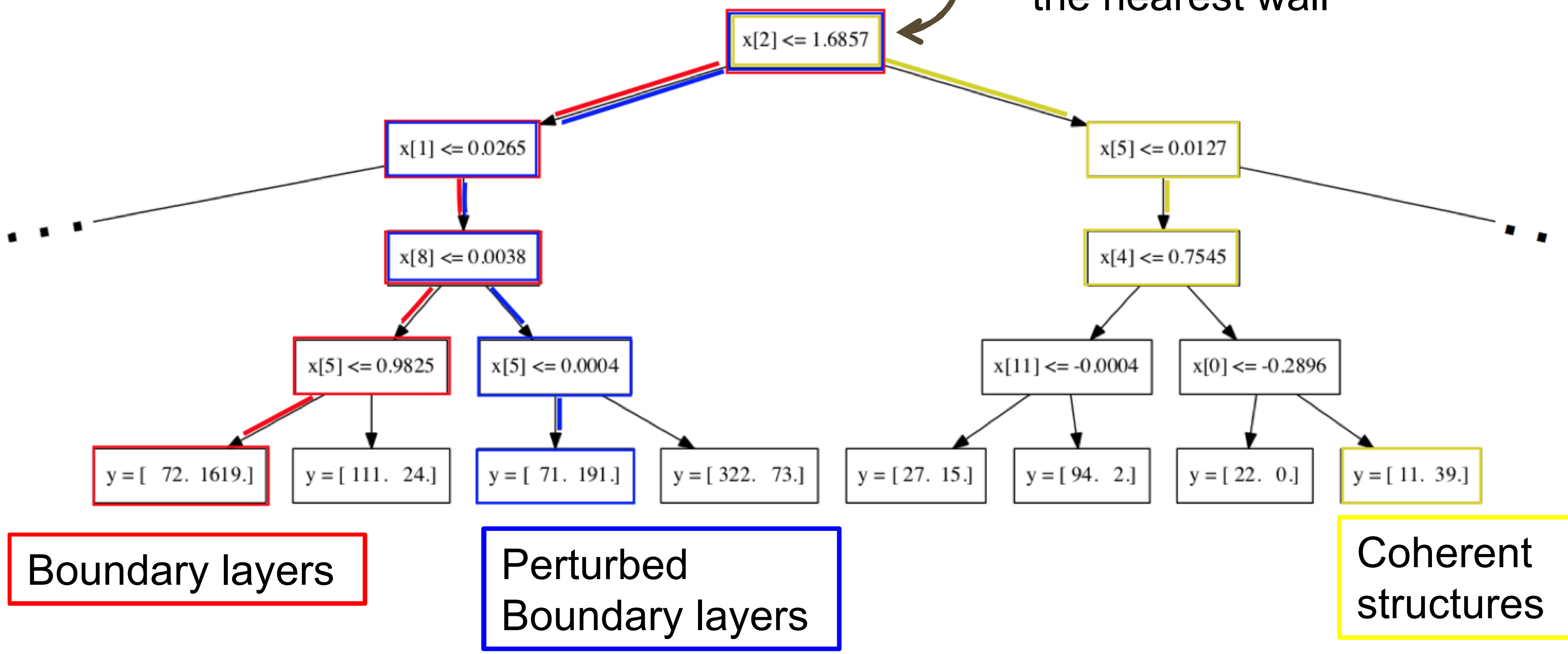
Analyzing the Representer Tree

- Look for consistent branches

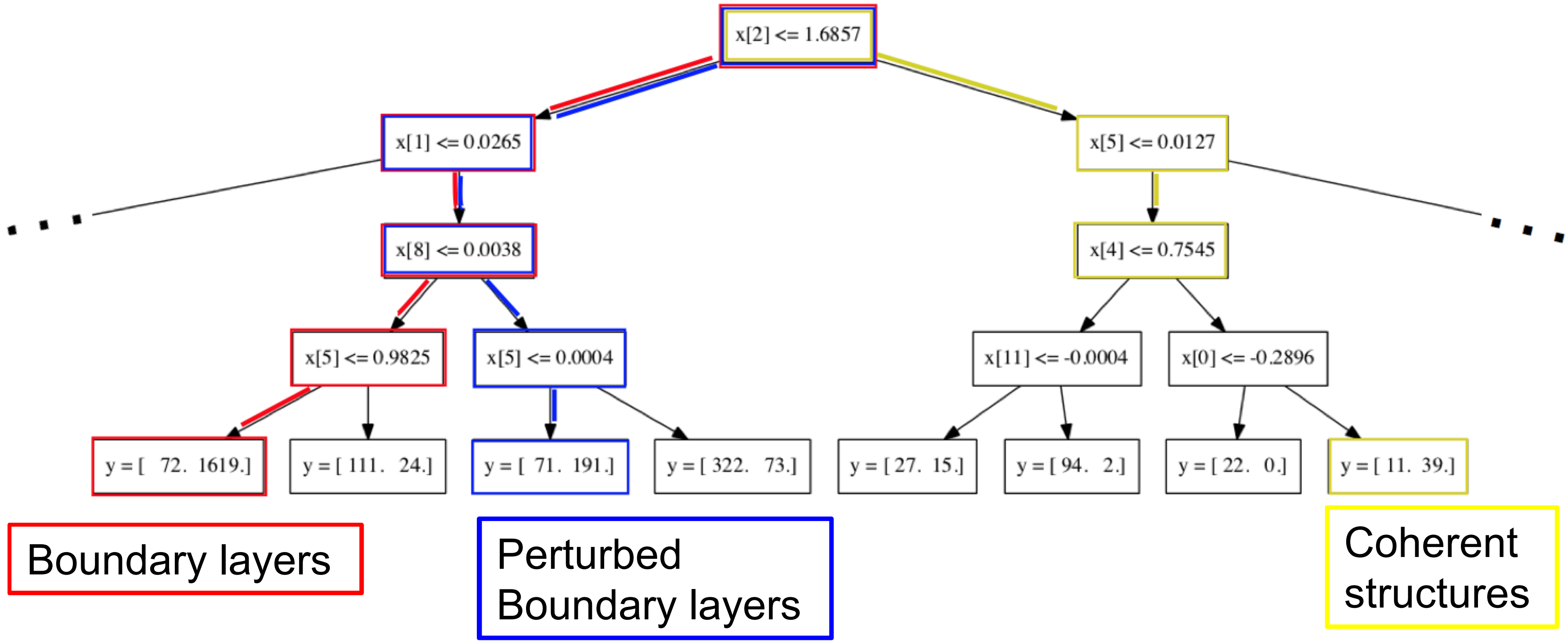


Analyzing the Representer Tree

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



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

Conclusions

- Machine learning was used to detect when RANS assumptions break down
- These data-driven methods achieved significantly improved classification accuracy by leveraging the high-dimensional data
- Rule extraction techniques were then used to regain physical intuition from the machine learning classifiers

Acknowledgments

Supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, 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. SAND

References

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,” ICMLA, (2015).

C. Gorle, J. Larsson, M. Emory, G. Iaccarino, “The deviation from parallel shear flow as an indicator of linear eddy viscosity model inaccuracy,” Physics of Fluids, (2014).