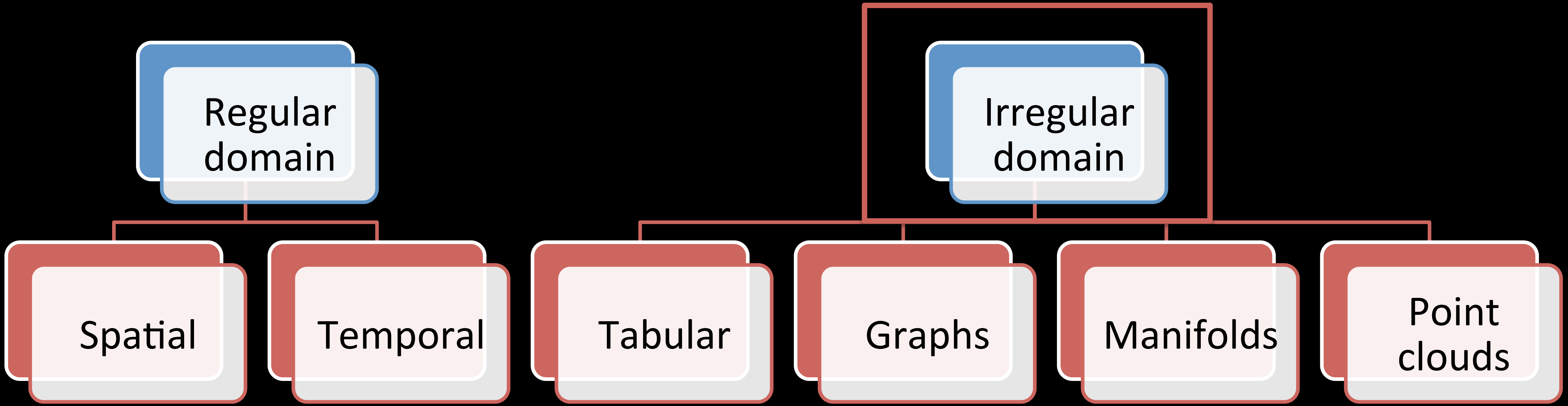
Scientific Domain-Informed Machine Learning

Prasanna Balaprakash
Computer Scientist
Mathematics and Computer Science Division and Leadership Computing Facility
Argonne National Laboratory

Joint work with P. Malakar, V. Vishwanath, K. Kumaran, V. Morozov, J. Wang, R. Kotamarthi

ALCF Simulation, Data, and Learning Workshop
October 3, 2018

DOE scientific data for machine learning



Regular domain

Spatial

Temporal

Irregular domain

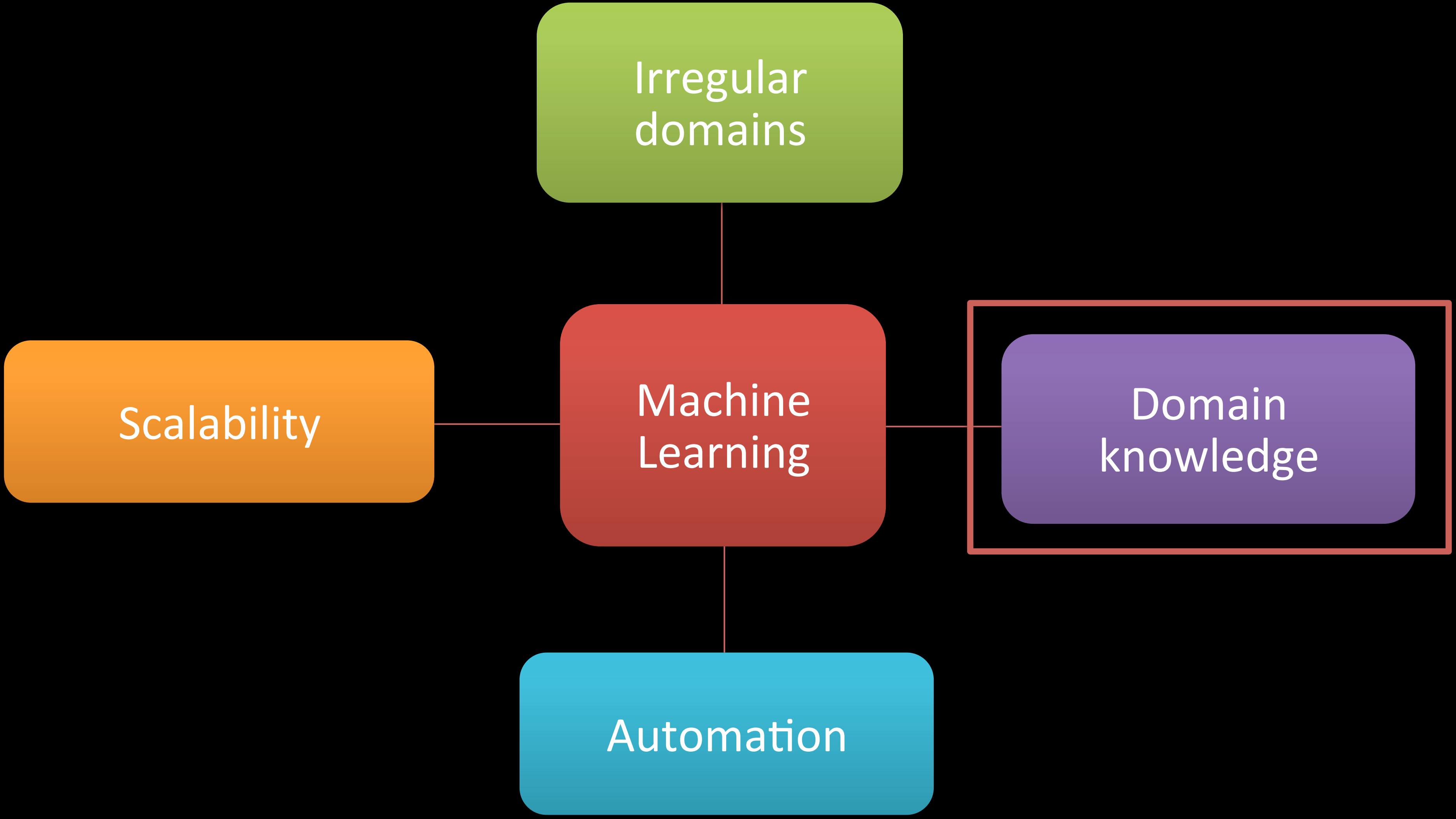
Tabular

Graphs

Manifolds

Point clouds

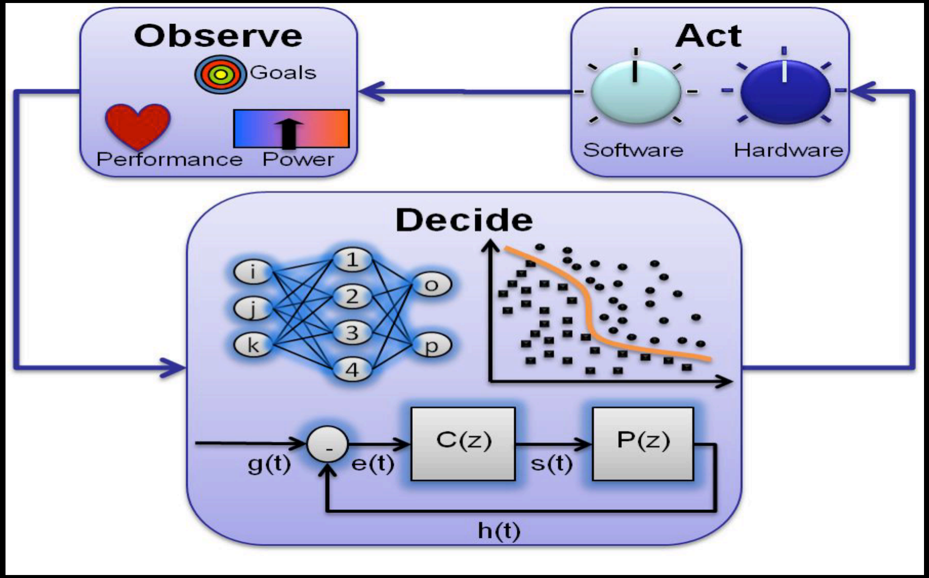
Challenges for irregular domains



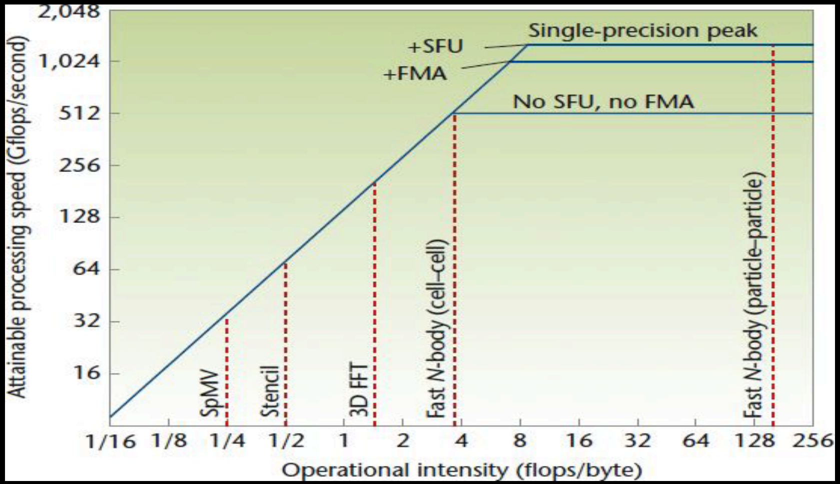
Case studies

- Scientific application performance modeling
- Surrogate modeling in weather simulation

Predictive models in HPC applications



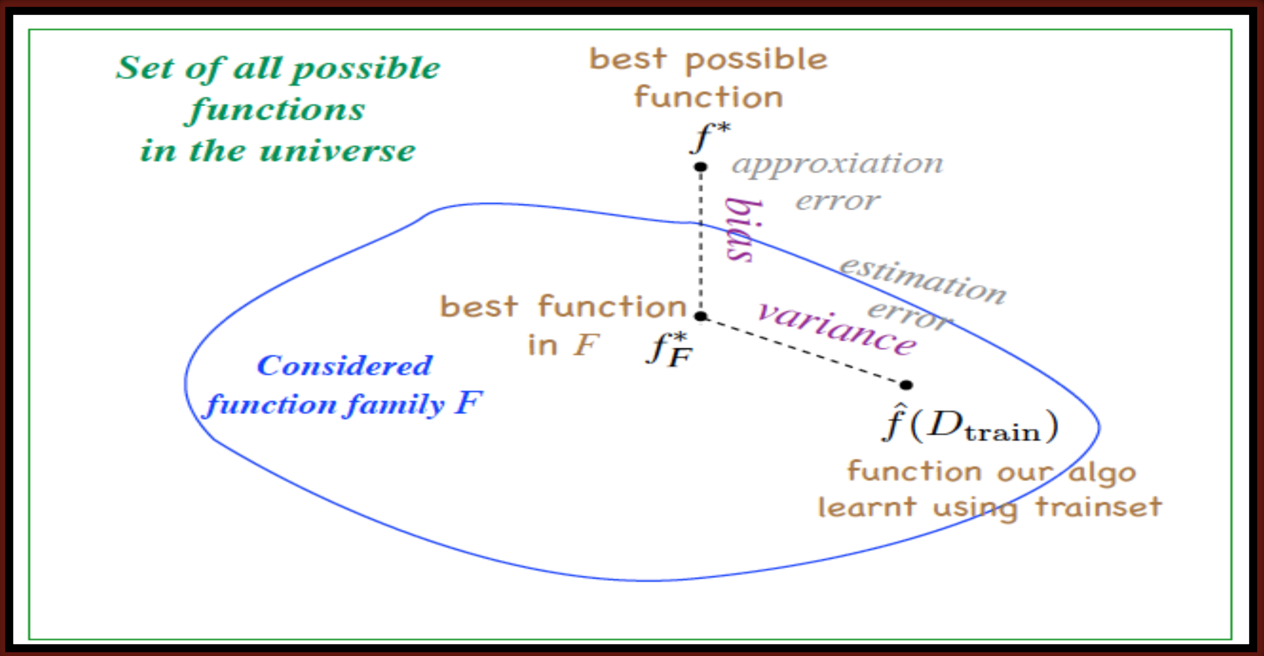
[H. Hoffmann, World Changing Ideas, SA 2009]



[S. Williams et al., ACM 2009]

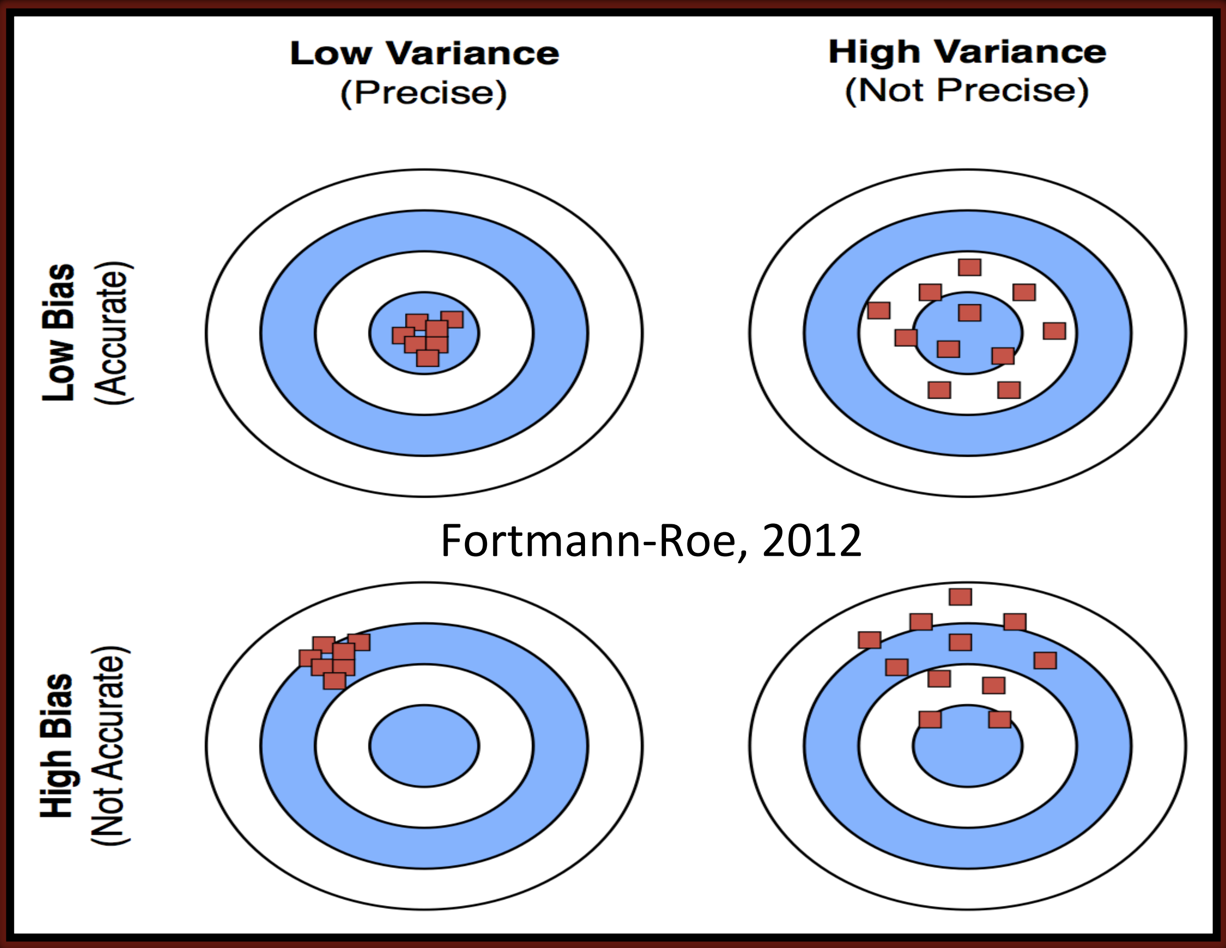
- Performance (run time) prediction still challenging
- ML-based performance modeling to bridge the gap
- Insights on important knobs that impacts performance
- Help prune large search spaces in performance tuning

Bias variance tradeoff



Set of all possible functions in the universe
best possible function
f^*
approximation error
bias
estimation error
variance
best function in F
f^*_F
Considered function family F
$\hat{f}(D_{\text{train}})$
function our algo learnt using trainset

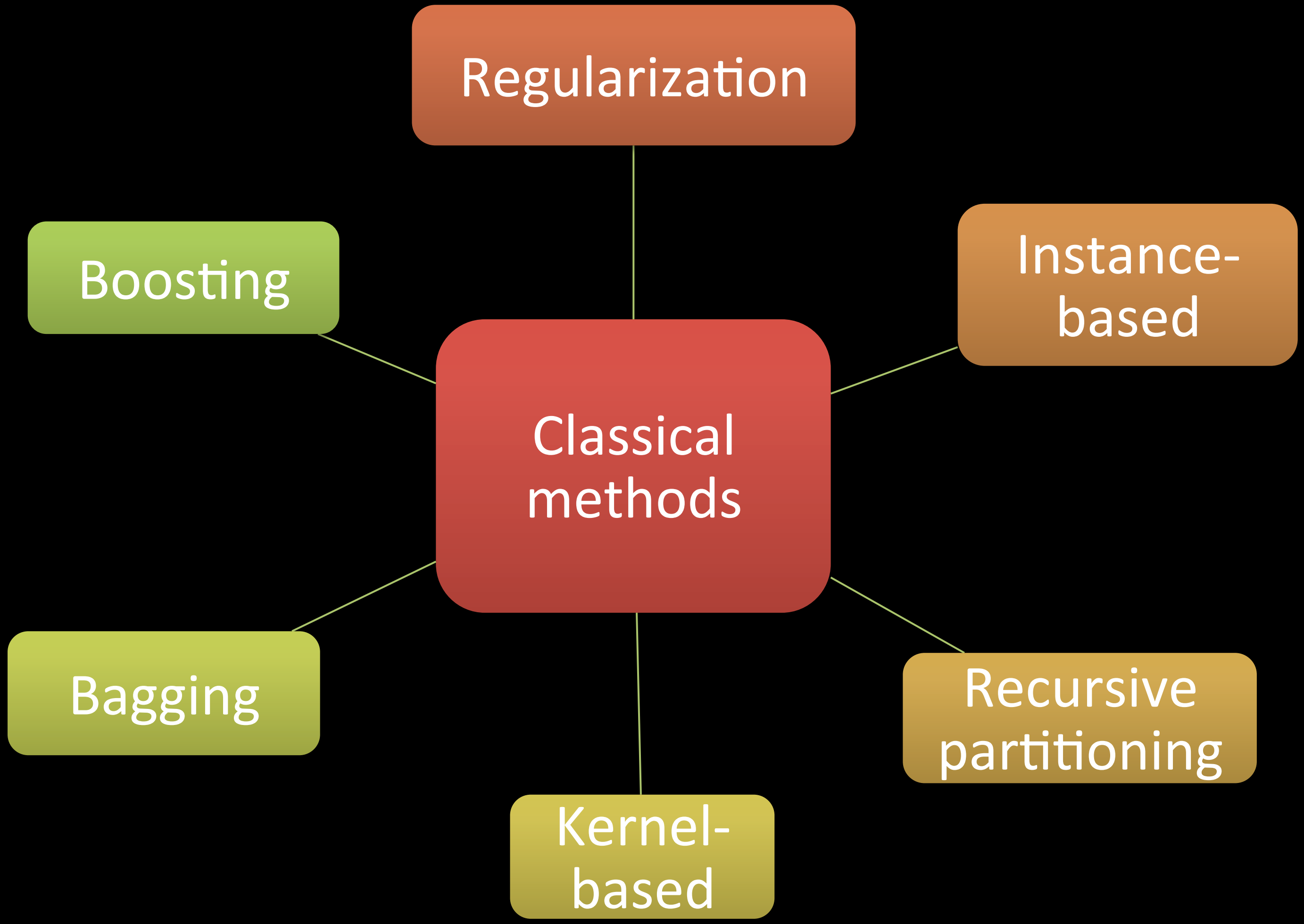
Deep learning summer school lecture, CIFAR, 2016



Fortmann-Roe, 2012

- ***No free lunch***: no single method will work well on all data set
- All supervised learning algorithms ***seek to reduce bias and variance*** in a different way

Supervised learning methods



Regularization

Boosting

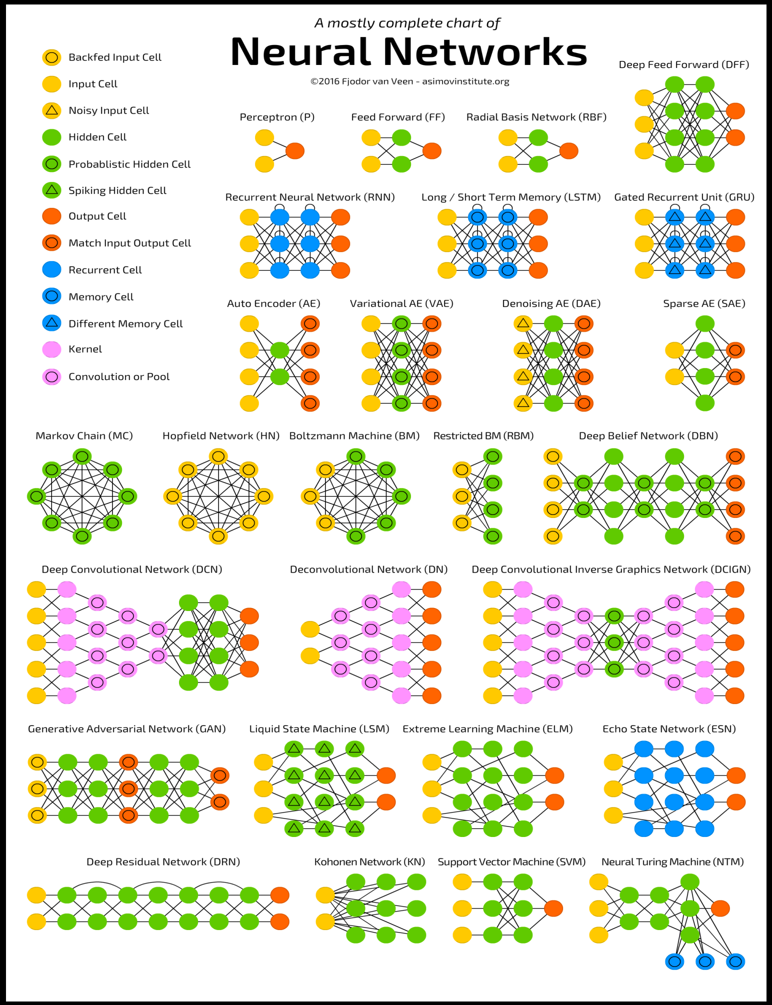
Instance-based

Classical methods

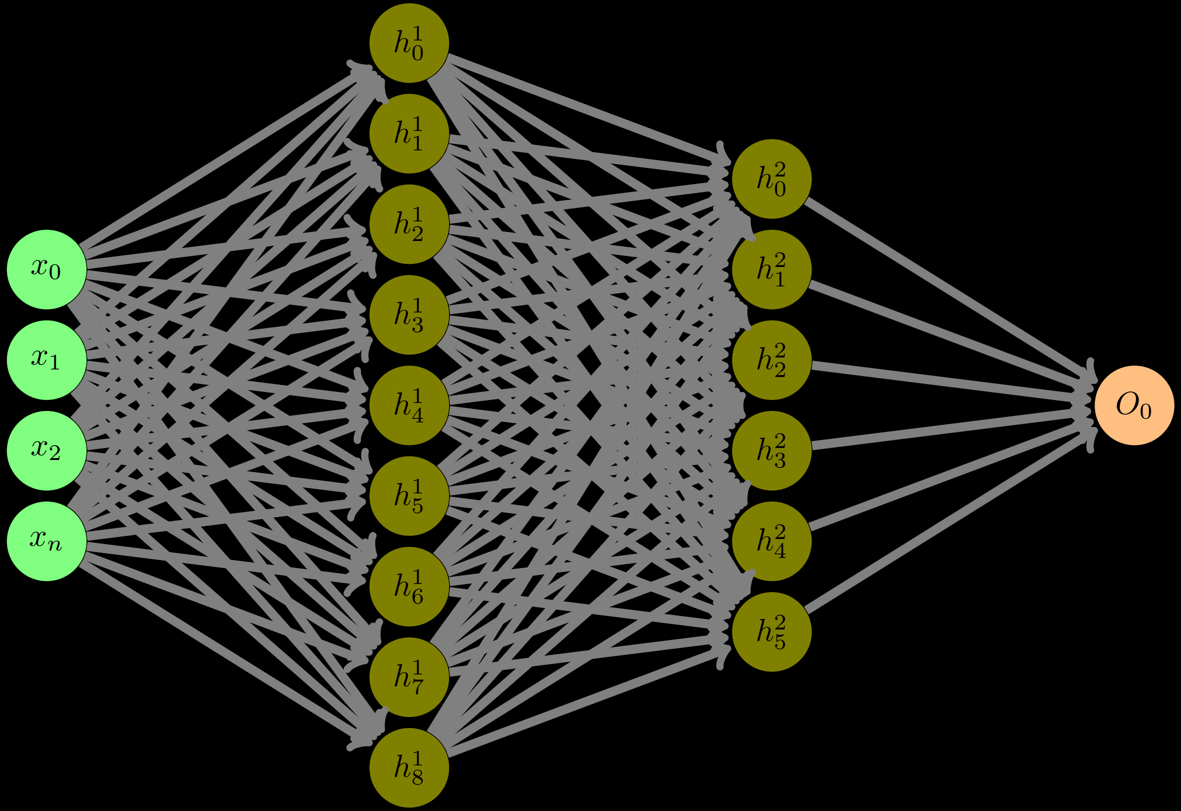
Bagging

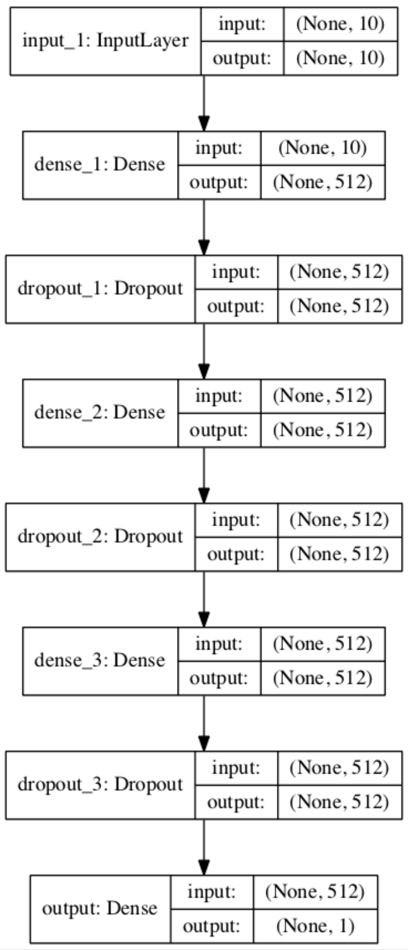
Recursive partitioning

Kernel-based


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Kohonen Network (KN)
Support Vector Machine (SVM)
Neural Turing Machine (NTM)

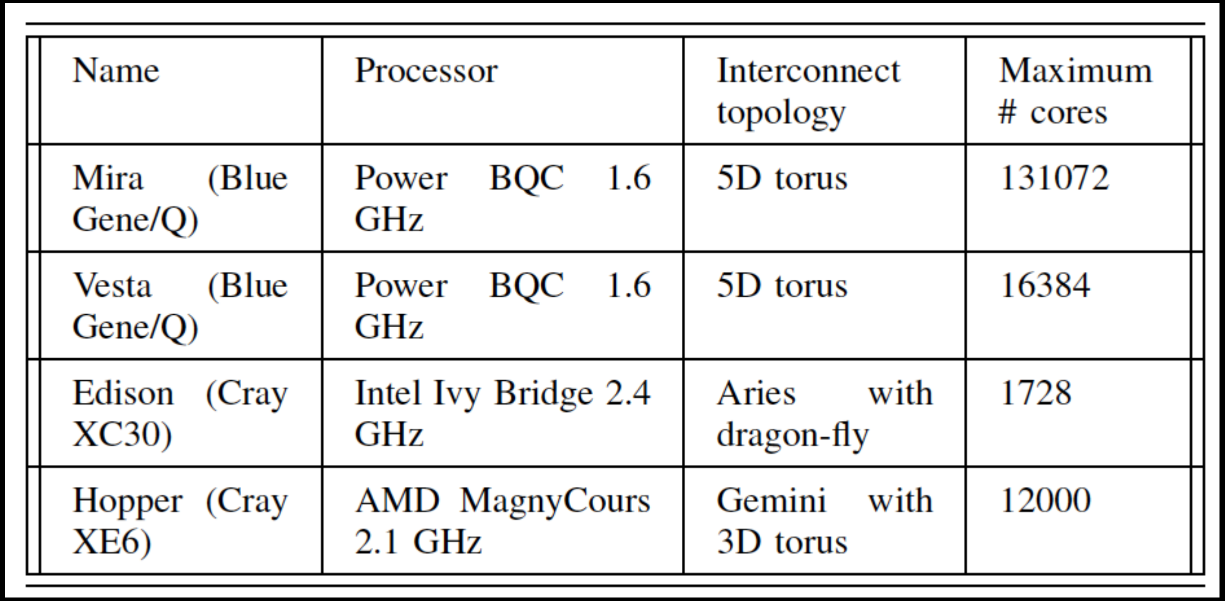
Deep neural networks





input_1: InputLayer	input:	(None, 10)
	output:	(None, 10)
dense_1: Dense	input:	(None, 10)
	output:	(None, 512)
dropout_1: Dropout	input:	(None, 512)
	output:	(None, 512)
dense_2: Dense	input:	(None, 512)
	output:	(None, 512)
dropout_2: Dropout	input:	(None, 512)
	output:	(None, 512)
dense_3: Dense	input:	(None, 512)
	output:	(None, 512)
dropout_3: Dropout	input:	(None, 512)
	output:	(None, 512)
output: Dense	input:	(None, 512)
	output:	(None, 1)

Applications and platforms

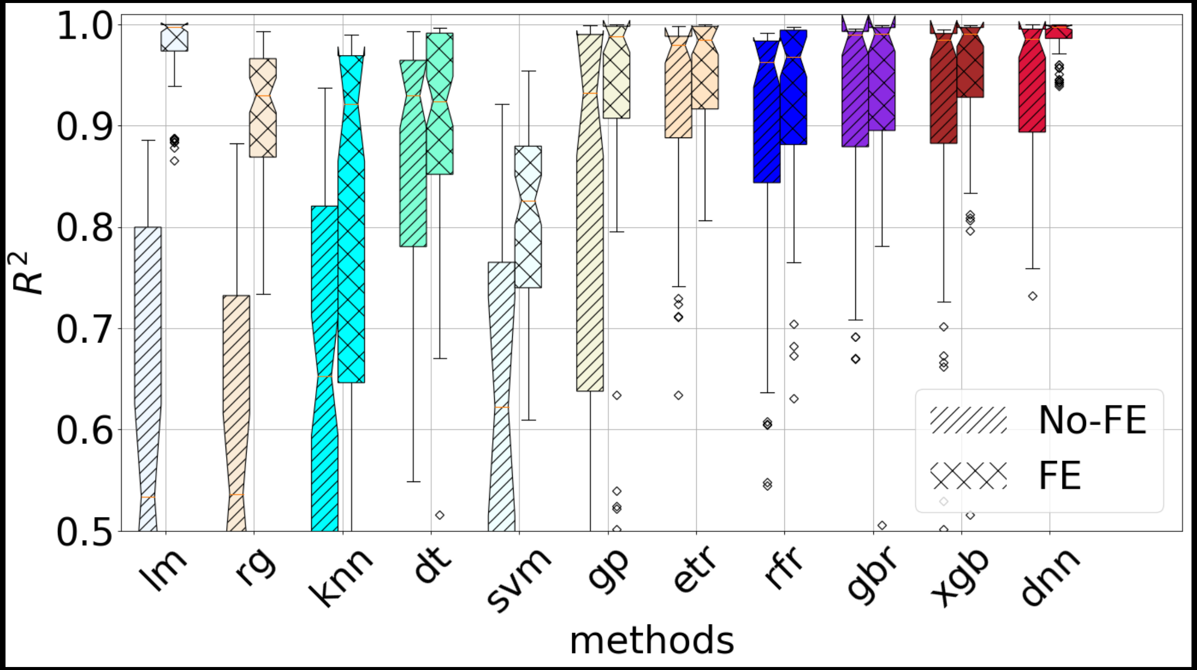


Name	Processor	Interconnect topology	Maximum # cores
Mira (Blue Gene/Q)	Power BQC 1.6 GHz	5D torus	131072
Vesta (Blue Gene/Q)	Power BQC 1.6 GHz	5D torus	16384
Edison (Cray XC30)	Intel Ivy Bridge 2.4 GHz	Aries with dragon-fly	1728
Hopper (Cray XE6)	AMD MagnyCours 2.1 GHz	Gemini with 3D torus	12000

- Miniapps (# no of data points):
  - miniMD (< 2K); O(1024) nodes
  - miniAMR (< 1K); O(4096) nodes
  - miniFE (6K to 15K); O(8192) node
  - LAMMPS (< 1K); O(1024) nodes

Impact of feature engineering

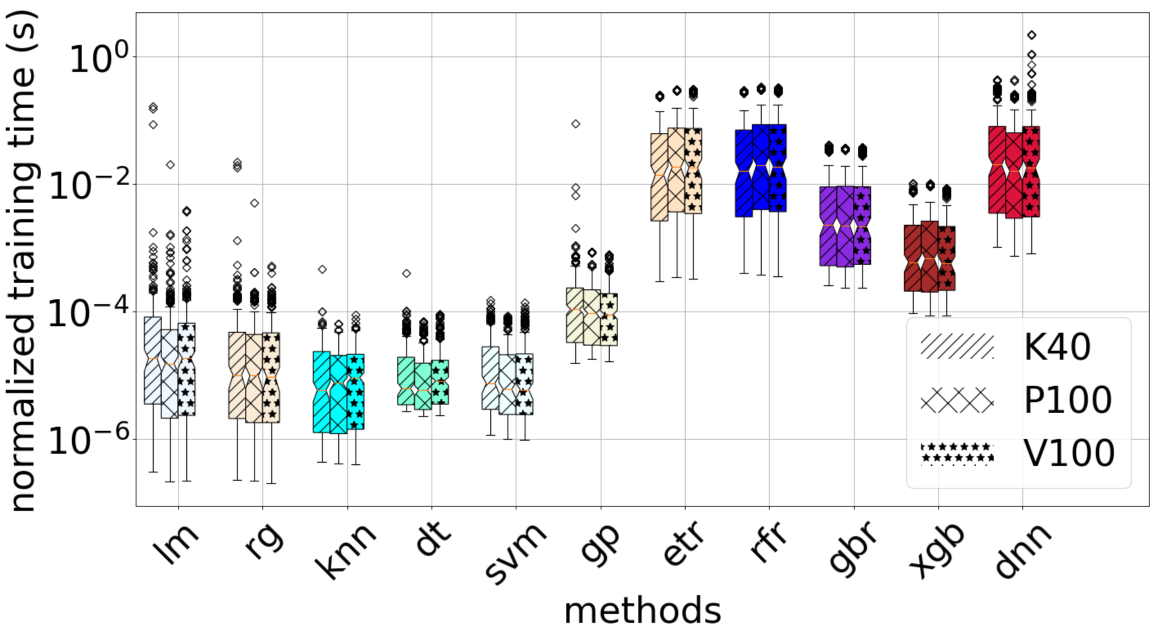
- No Feature Engineering (No-FE)
  - application input parameters
- Feature Engineering (FE)
  - application input parameters
  - ratio of the application problem size and the number of number of processes
  - inverse of the number of processes
  - binary logarithm of number of processes

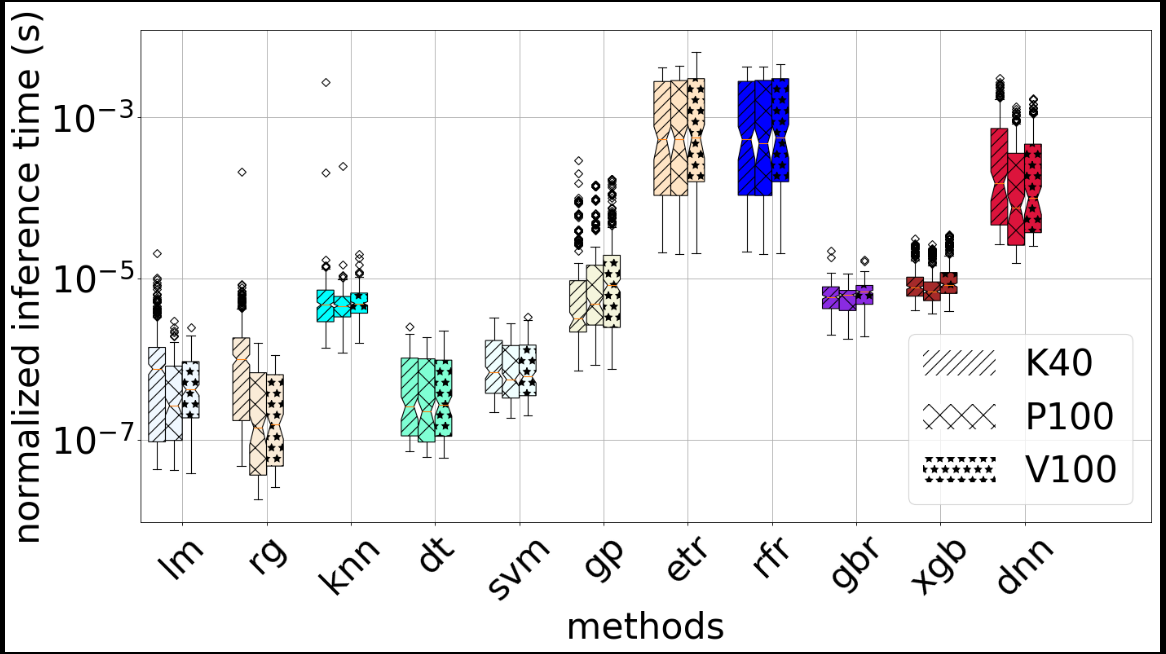


10 X 20:80 cross validation

Feature engineering has a significant impact on the accuracy

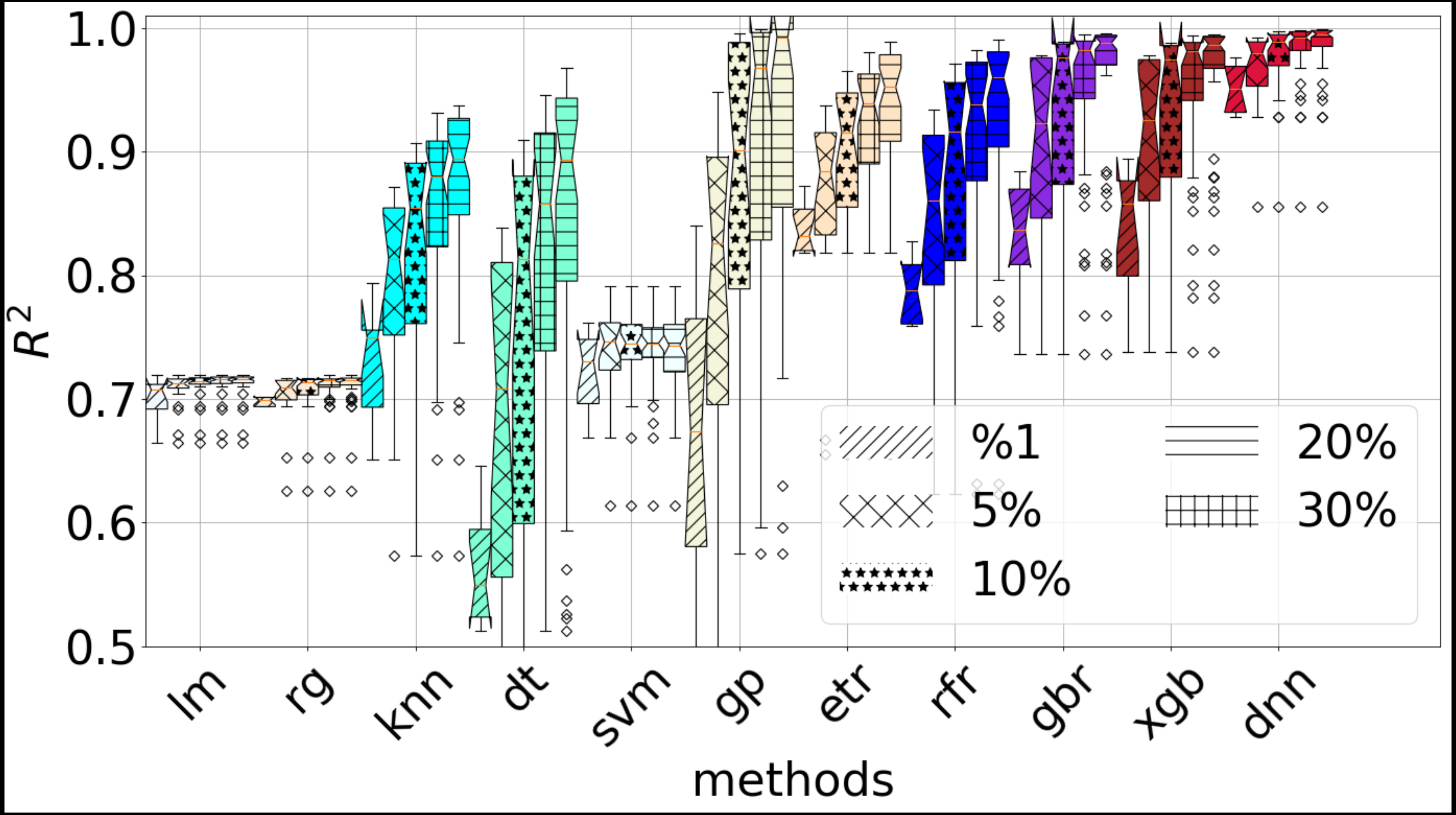
Impact of hardware platforms





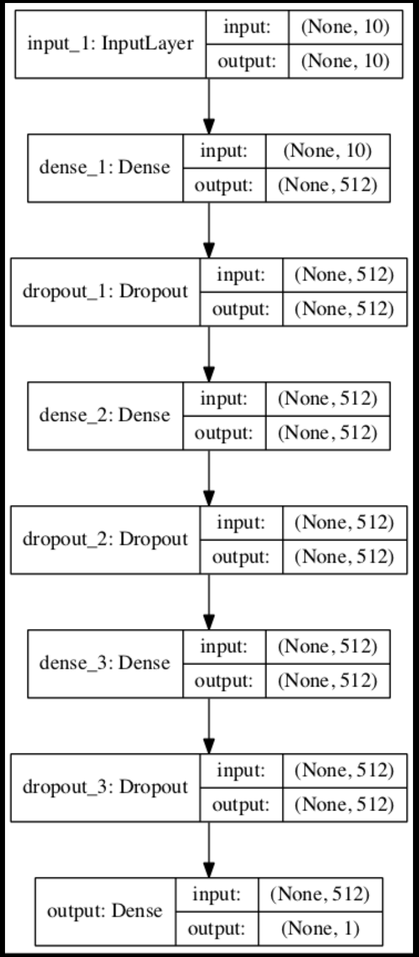
Algorithmic complexity has more impact than hardware platforms

Impact of training data size on accuracy



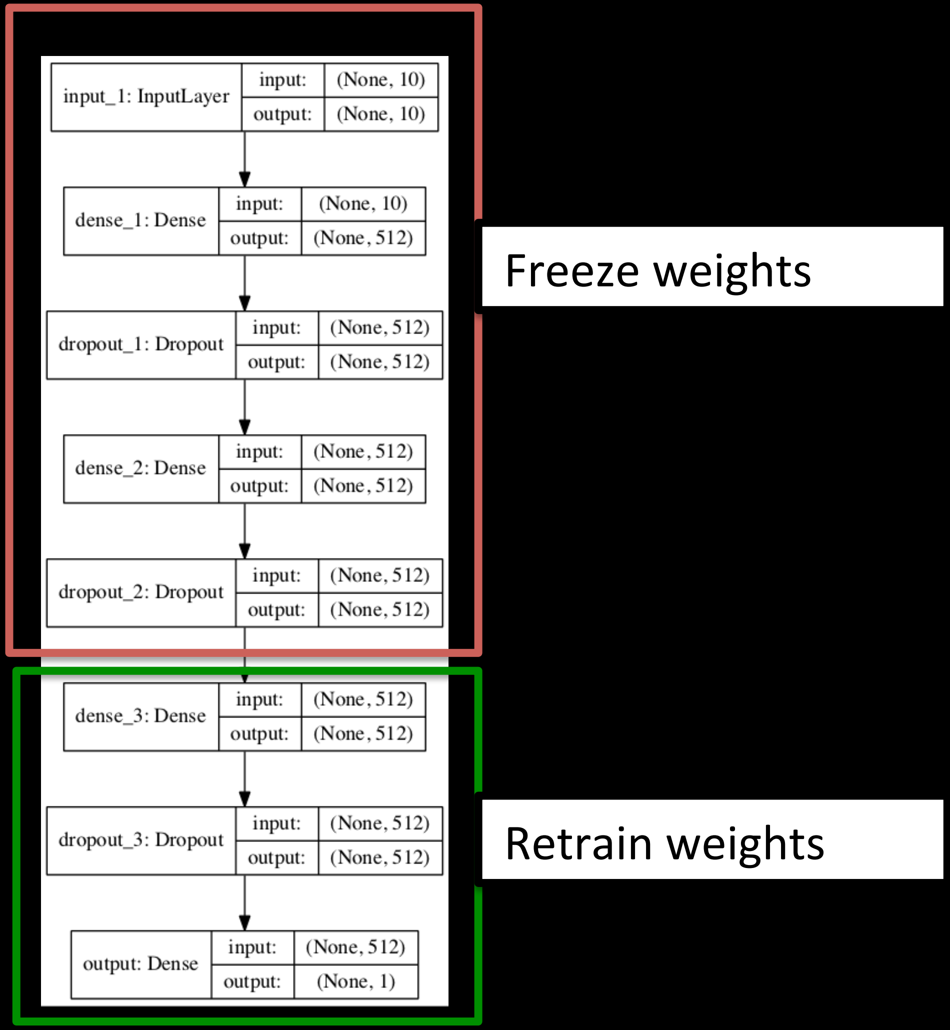
Nonlinear methods leverage large training data size

Transfer learning



input_1: InputLayer | input: (None, 10) | output: (None, 10)
dense_1: Dense | input: (None, 10) | output: (None, 512)
dropout_1: Dropout | input: (None, 512) | output: (None, 512)
dense_2: Dense | input: (None, 512) | output: (None, 512)
dropout_2: Dropout | input: (None, 512) | output: (None, 512)
dense_3: Dense | input: (None, 512) | output: (None, 512)
dropout_3: Dropout | input: (None, 512) | output: (None, 512)
output: Dense | input: (None, 512) | output: (None, 1)

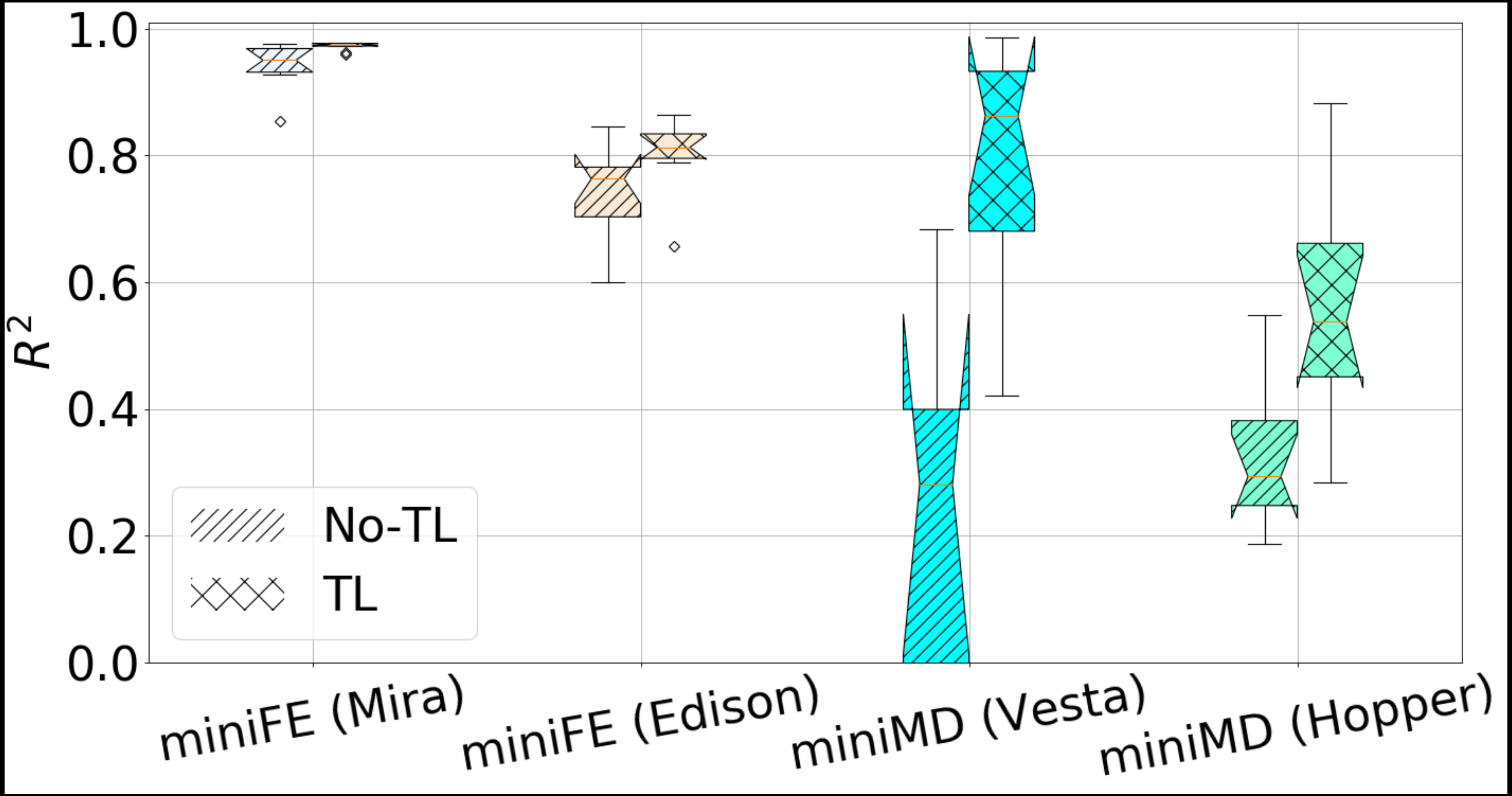
Platform 1



input_1: InputLayer | input: (None, 10) | output: (None, 10)
dense_1: Dense | input: (None, 10) | output: (None, 512)
dropout_1: Dropout | input: (None, 512) | output: (None, 512)
dense_2: Dense | input: (None, 512) | output: (None, 512)
dropout_2: Dropout | input: (None, 512) | output: (None, 512)
dense_3: Dense | input: (None, 512) | output: (None, 512)
dropout_3: Dropout | input: (None, 512) | output: (None, 512)
output: Dense | input: (None, 512) | output: (None, 1)
Freeze weights
Retrain weights

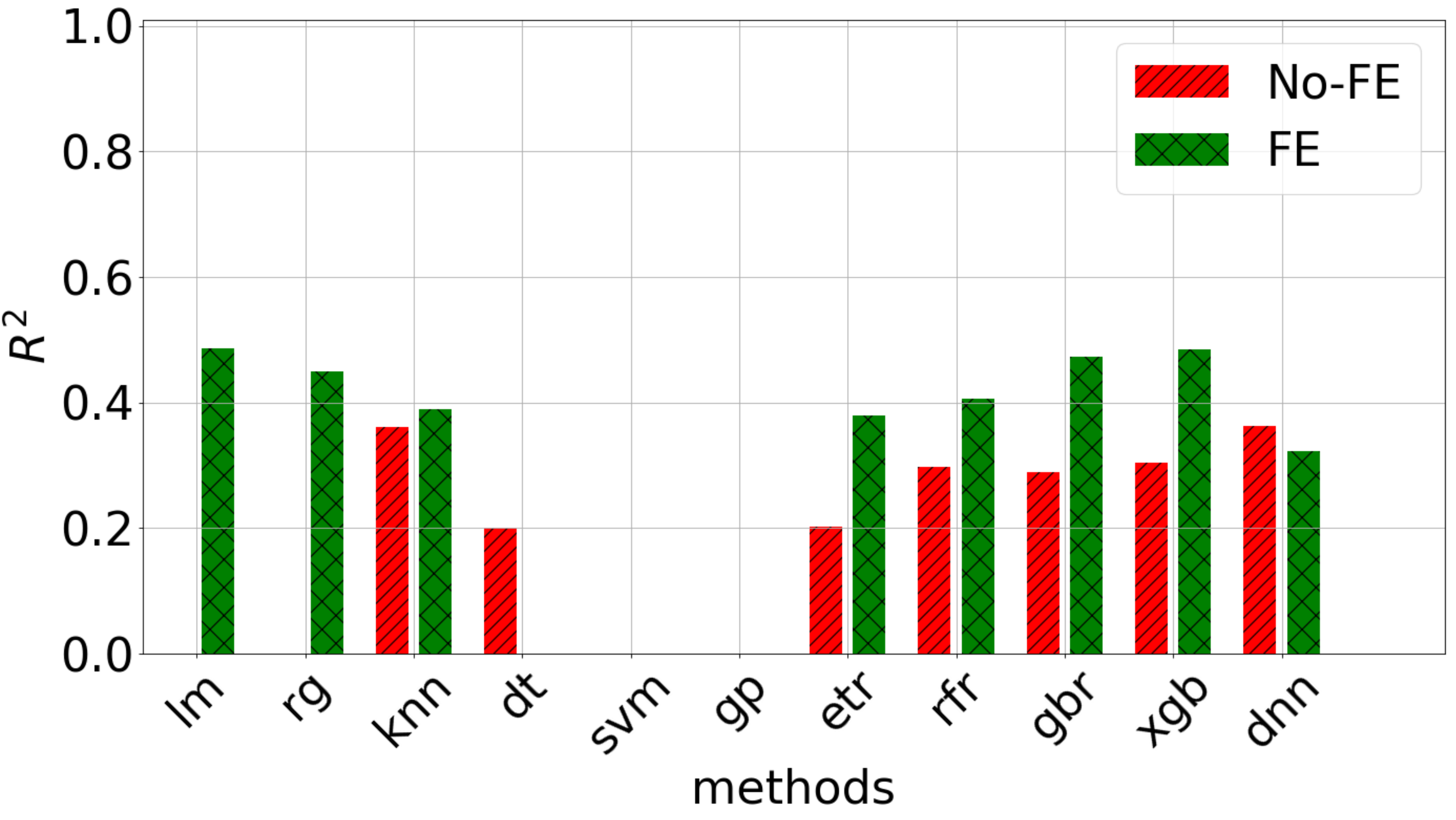
Platform 2

Transfer learning



Transfer learning significantly improves prediction accuracy

Extrapolation
from small to large problem sizes



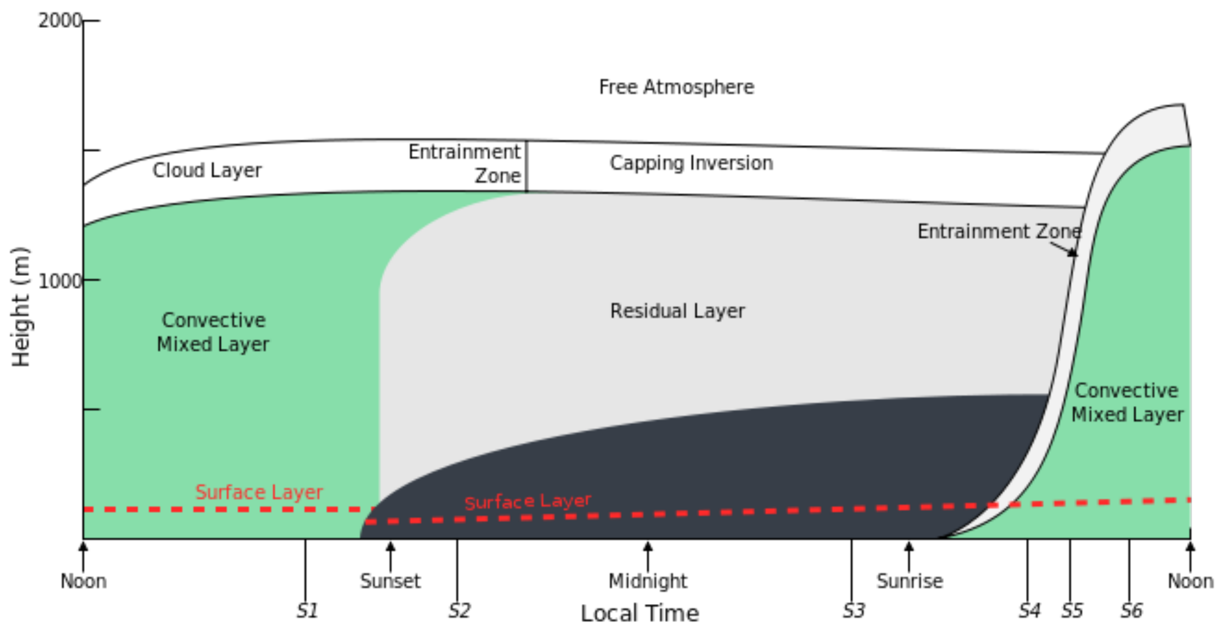
Incorporating domain knowledge helps in exploration

Case studies

- Scientific application performance modeling
- Surrogate modeling in weather simulation

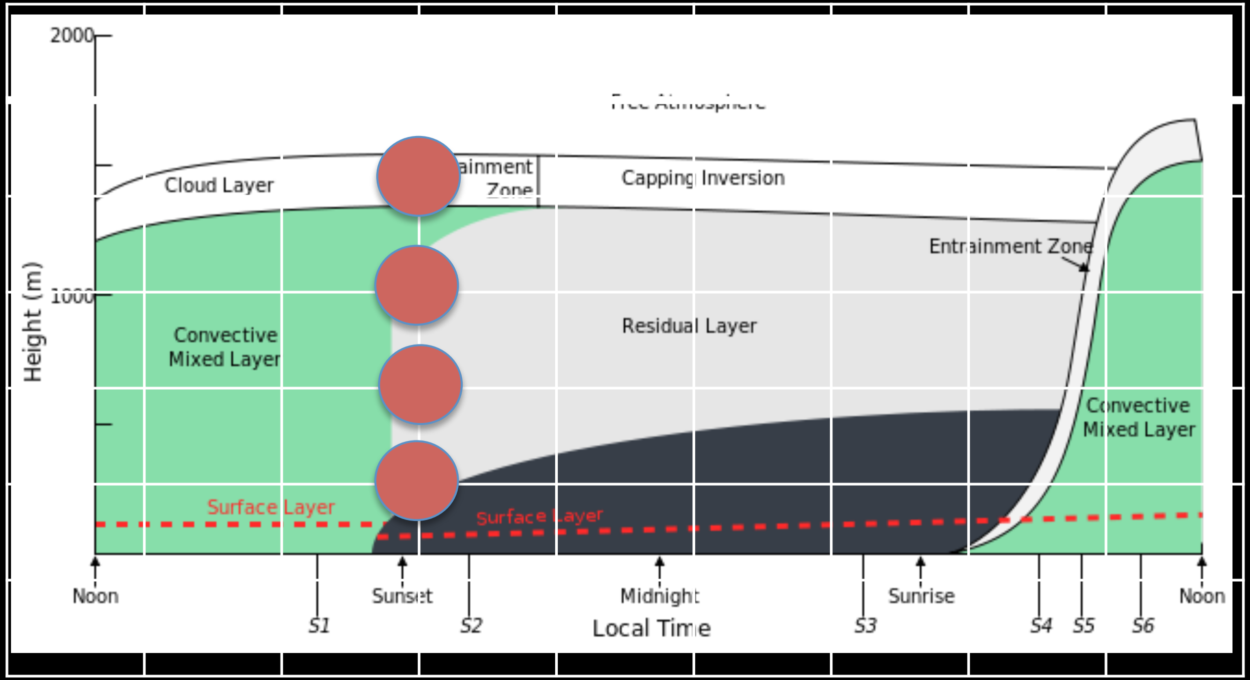
Planetary boundary layer

- Planetary boundary layer (PBL)
  - lowest part of the atmosphere
  - directly influenced by its contact with a planetary surface
  - responds to changes in surface radiative forcing
  - flow velocity, temperature, moisture, etc., display rapid fluctuations
  - computationally expensive in weather research and forecasting model



https://en.wikipedia.org/wiki/Planetary_boundary_layer

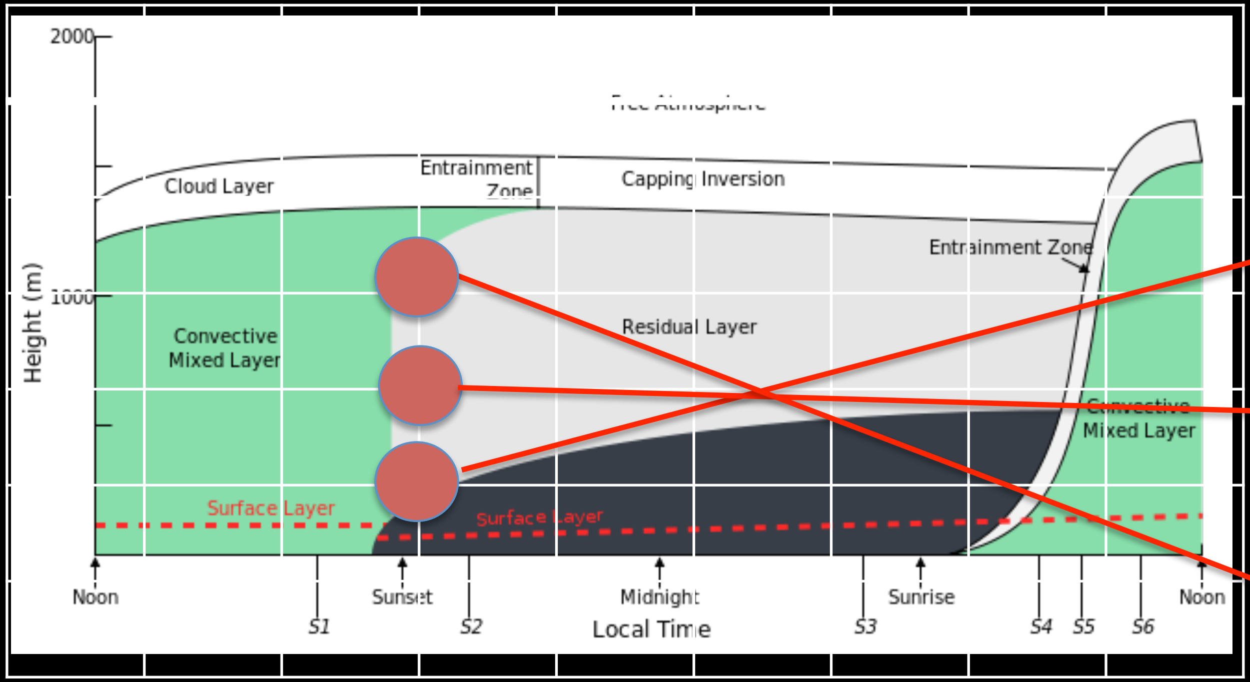
Planetary boundary layer

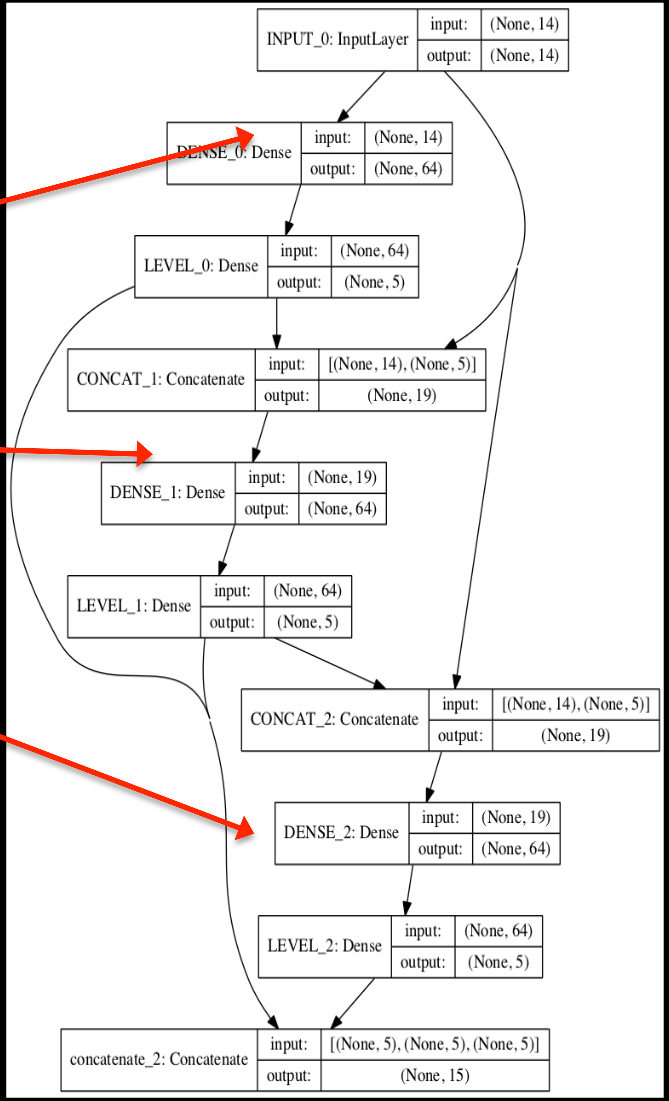


Output: profiles of wind, temperature, moisture

Input: Surface properties, fluxes, and ground temperature

Surrogate neural network





INPUT_0: InputLayer | input: (None, 14) | output: (None, 14)

DENSE_0: Dense | input: (None, 14) | output: (None, 64)

LEVEL_0: Dense | input: (None, 64) | output: (None, 5)

CONCAT_1: Concatenate | input: [(None, 14), (None, 5)] | output: (None, 19)

DENSE_1: Dense | input: (None, 19) | output: (None, 64)

LEVEL_1: Dense | input: (None, 64) | output: (None, 5)

CONCAT_2: Concatenate | input: [(None, 14), (None, 5)] | output: (None, 19)

DENSE_2: Dense | input: (None, 19) | output: (None, 64)

LEVEL_2: Dense | input: (None, 64) | output: (None, 5)

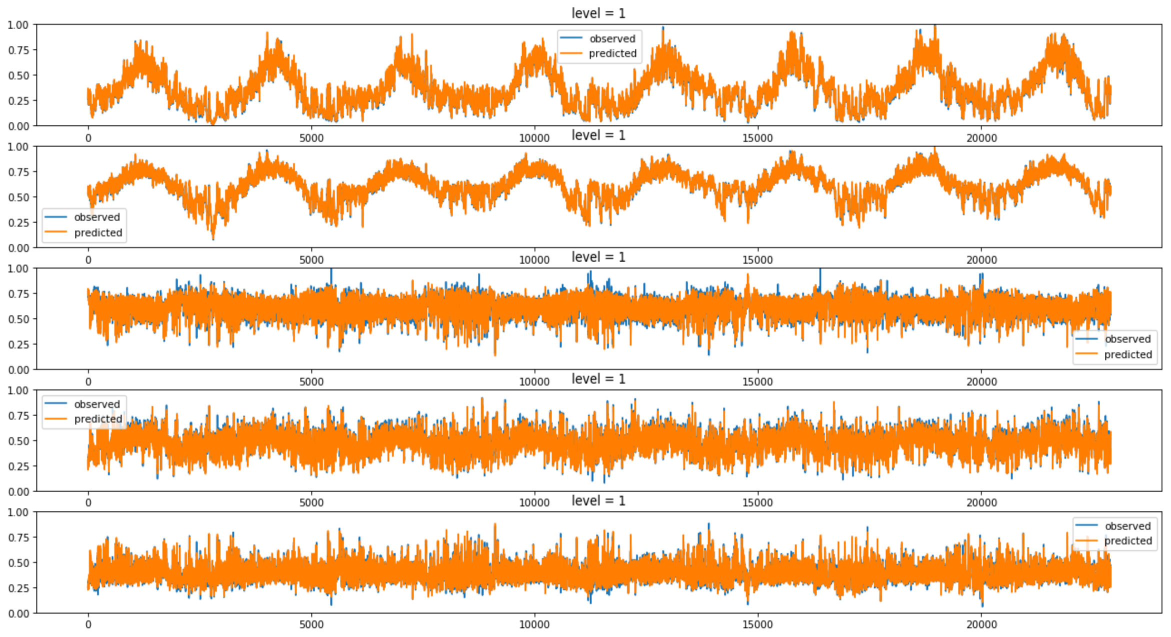
concatenate_2: Concatenate | input: [(None, 5), (None, 5), (None, 5)] | output: (None, 15)

Input: Surface properties, fluxes, and ground temperature

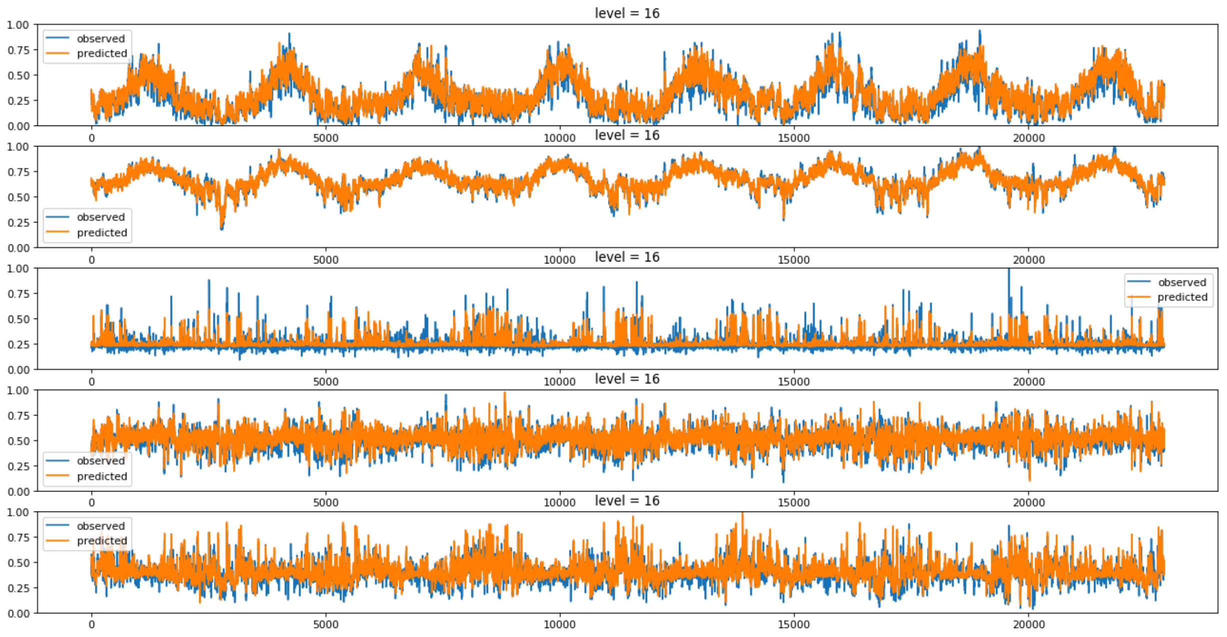
Output: profiles of wind, temperature, moisture

Prediction results

(level 1: close to surface)

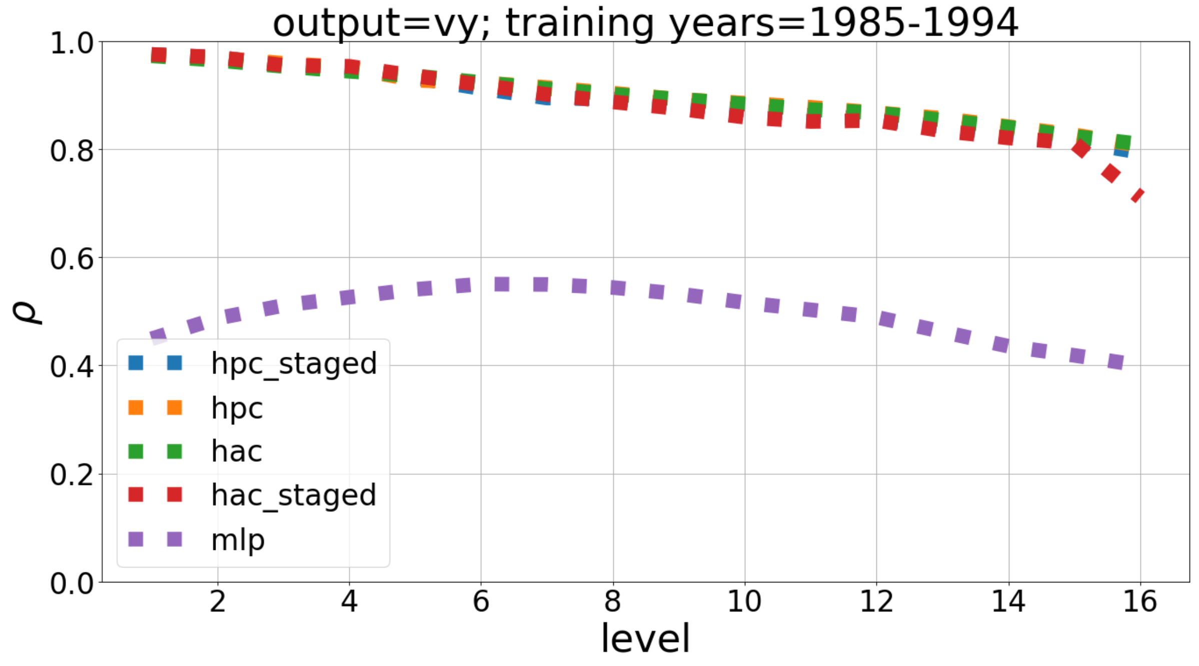


(level 16: ~2 km)

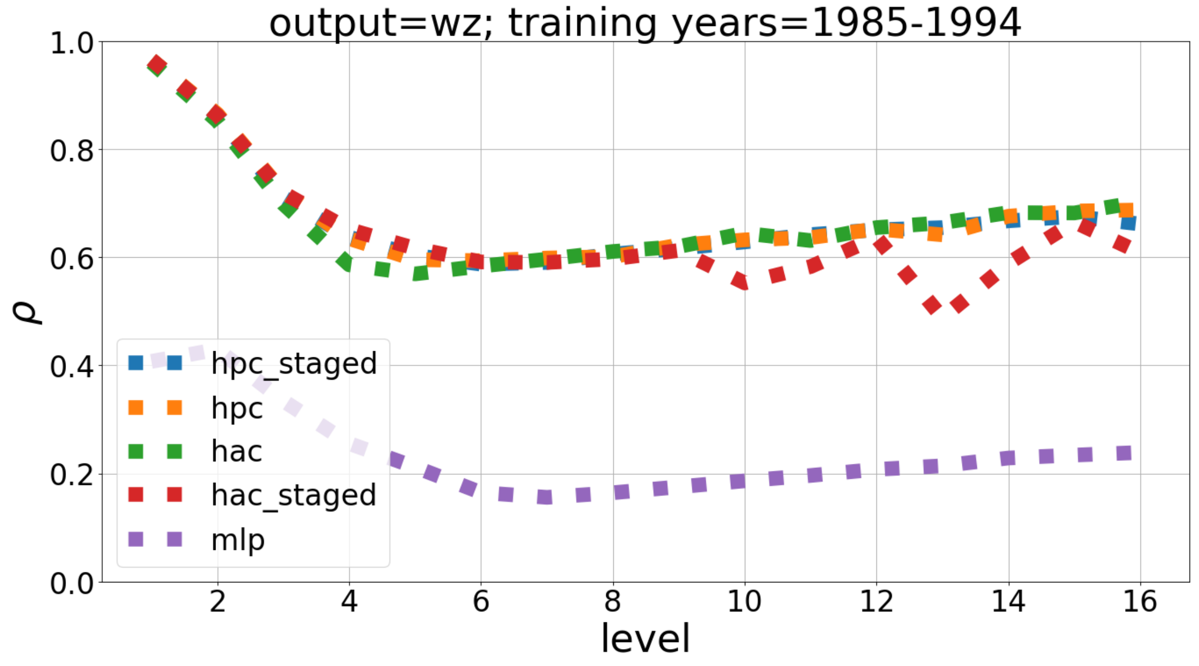


Prediction results

meridional wind (south-north wind)

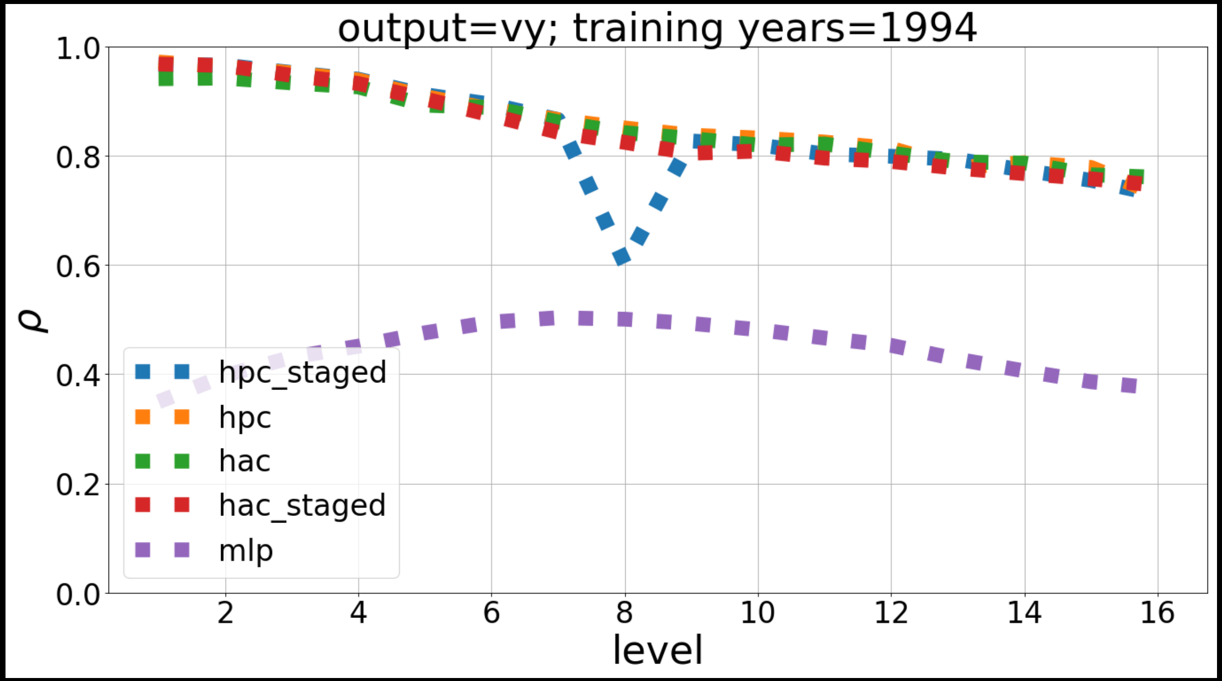


vertical velocity (up and down motions)

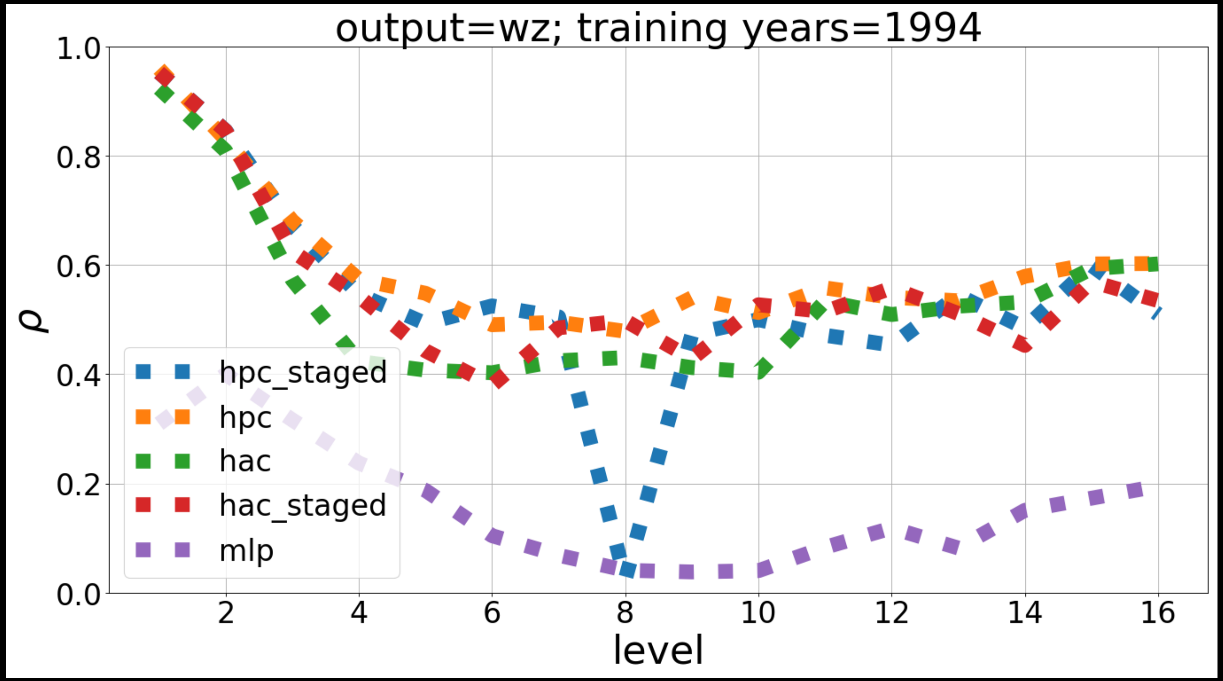


Prediction results

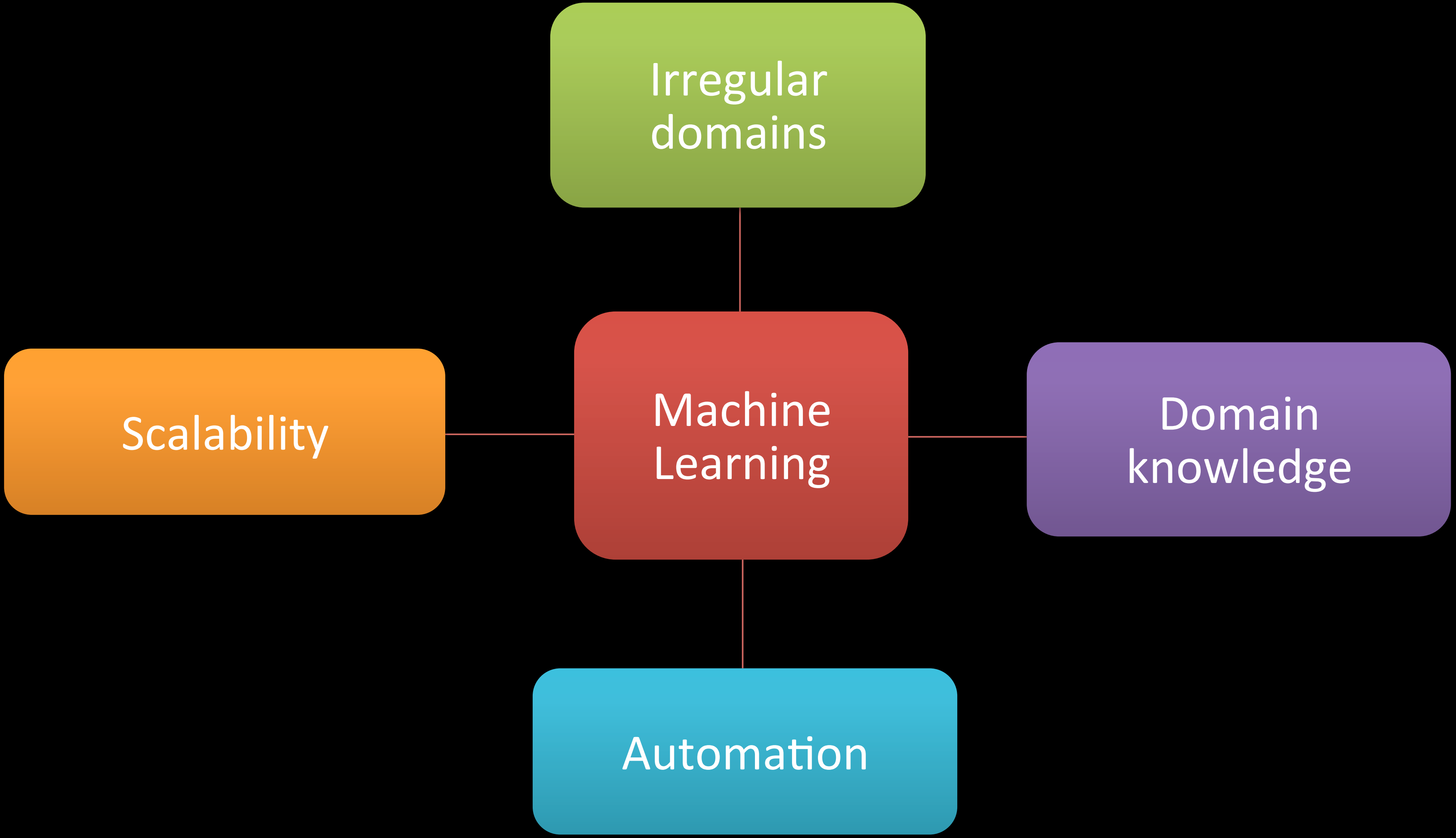
meridional wind (south-north wind)



vertical velocity (up and down motions)



Challenges for irregular domains



Irregular domains

Machine Learning

Scalability

Domain knowledge

Automation

Acknowledgements

- U.S. DOE
  - ANL LDRD
  - ALCF
  - ASCR, Early Career Research Program