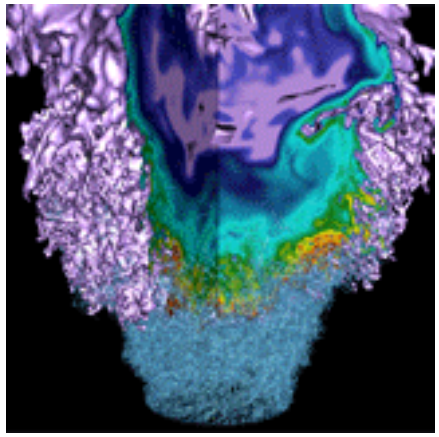
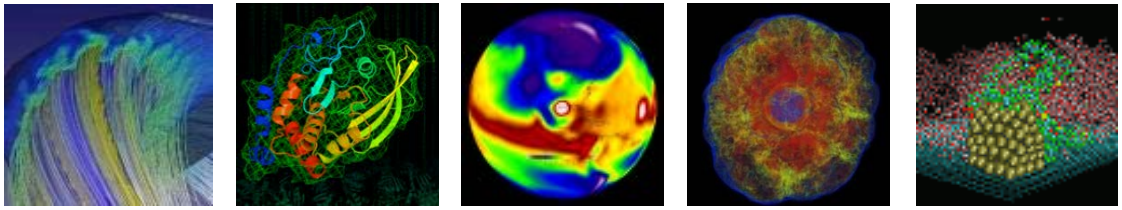
Deep Learning recognizes weather and climate patterns



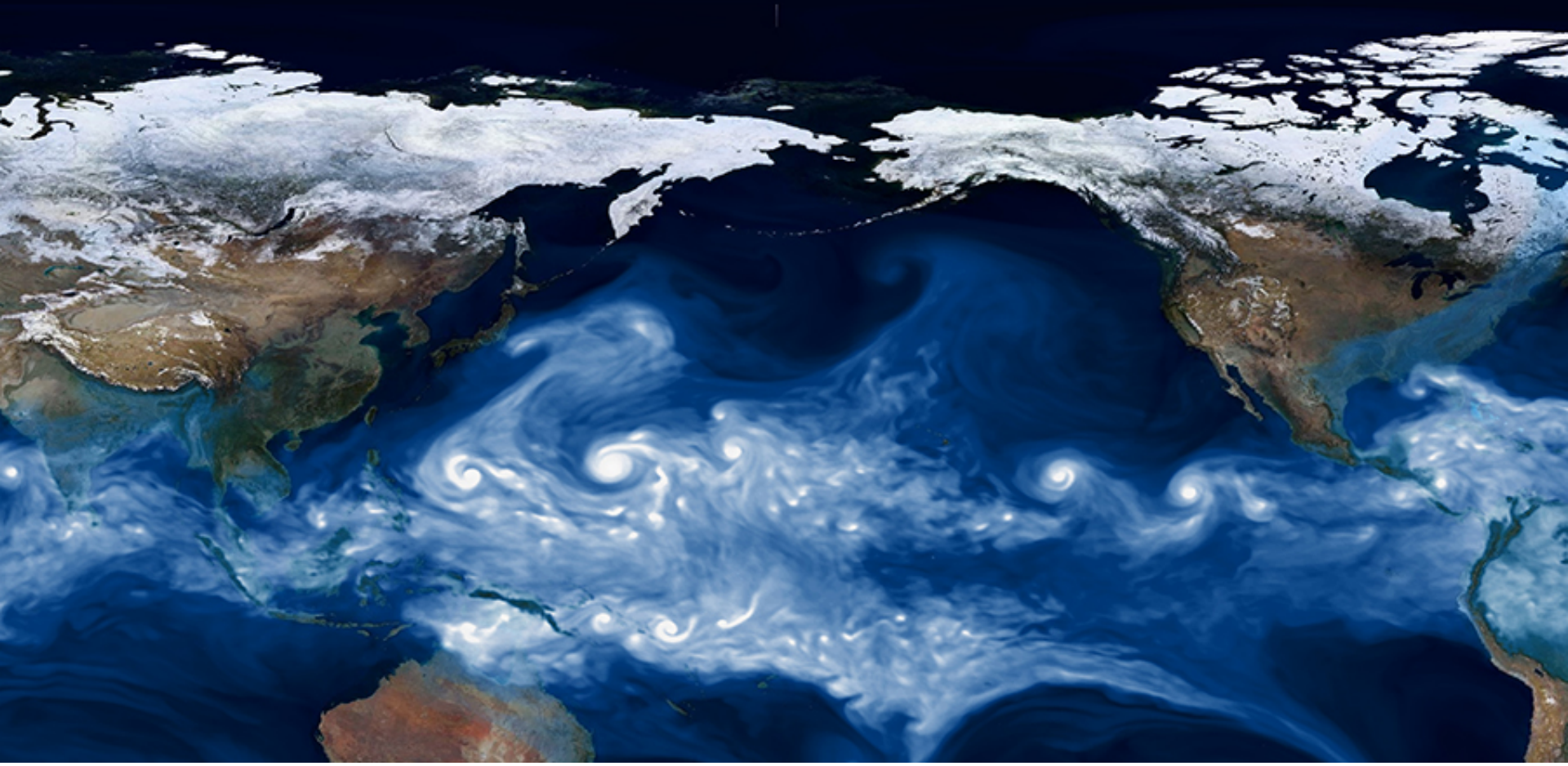


Karthik Kashinath, Prabhat, Mayur Mudigonda, Kevin Yang, Ankur Mahesh, Travis O'Brien, Michael Wehner, Bill Collins

Lawrence Berkeley National Laboratory

Collaborators: Benjamin Toms, Yunjie Liu, Evan Racah, Soo Kyung Kim, Samira Kahou, Christopher Beckham, Chris Pal, Tegan Maharaj, Jim Biard, Kenneth Kunkel, Dean Williams

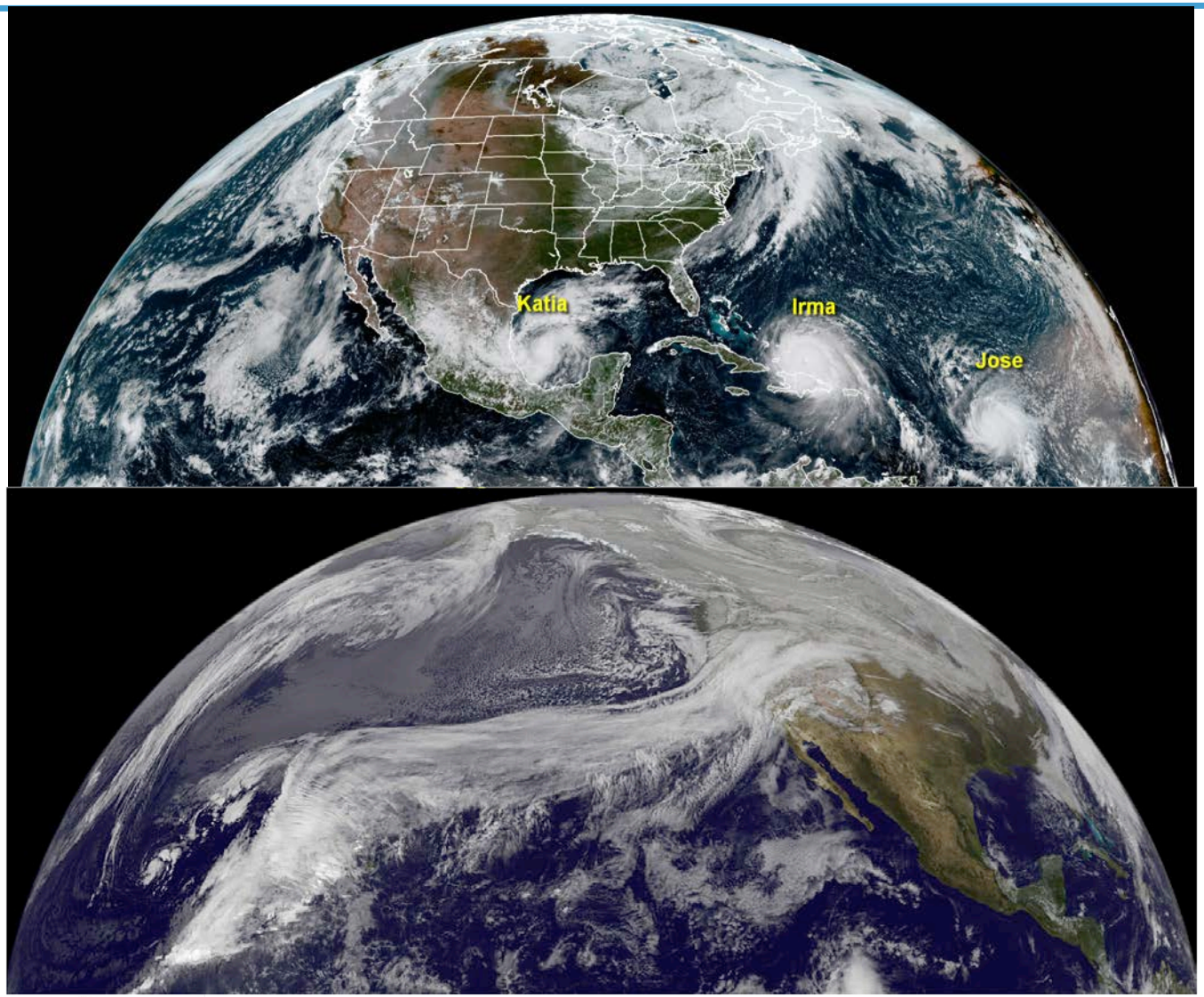
Characterizing Extreme Weather in a Changing Climate



U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

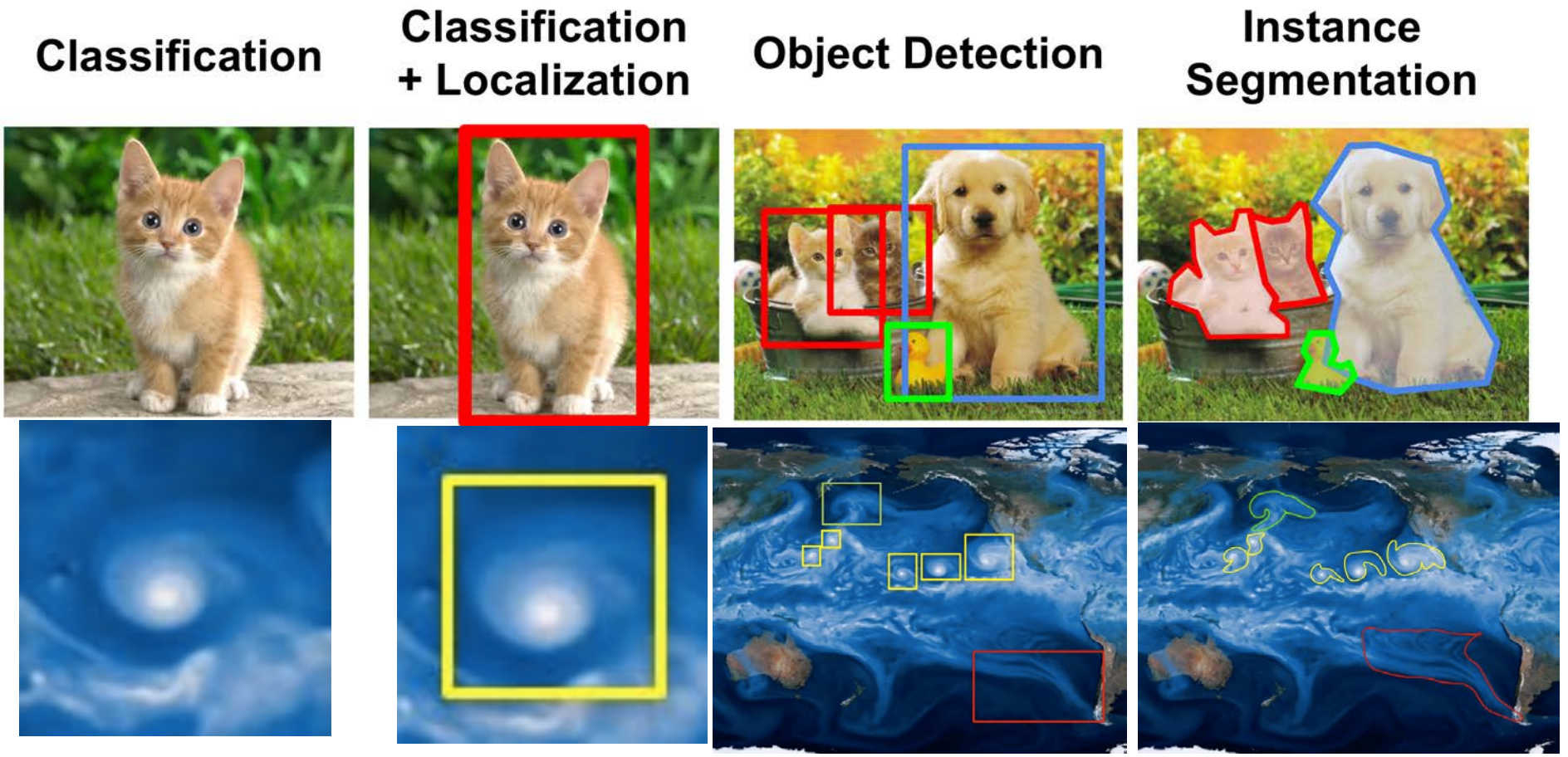
Characterizing Extreme Weather in a Changing Climate



Outline

- Deep Learning for classification, localization and pixel-wise segmentation of Tropical Cyclones, Atmospheric Rivers & Weather Fronts
- Deep Learning for detecting the MJO
- ClimateNet: a community-sourced expert-labeled dataset for training a unified DL model
- Challenges and Opportunities

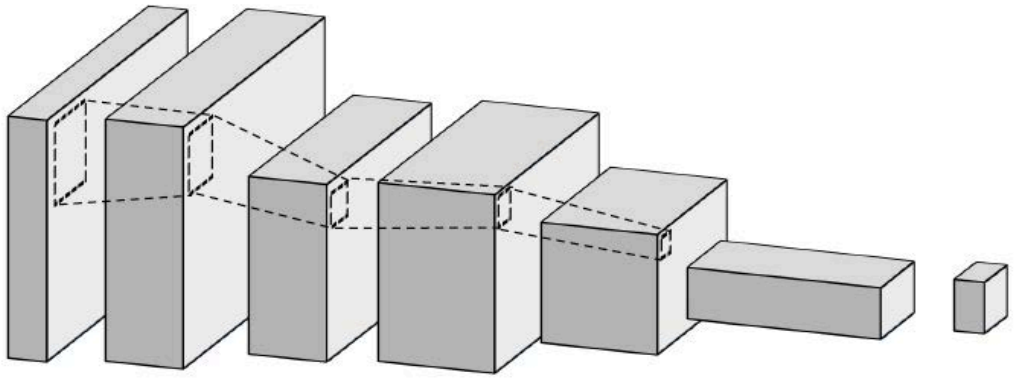
Computer Vision <> Climate Science



Key Differences: Multi-variate; Different & multiple spatio-temporal scales; Double precision floating point; Underlying statistics are different.

Supervised Classification Accuracy (NIPS 16)

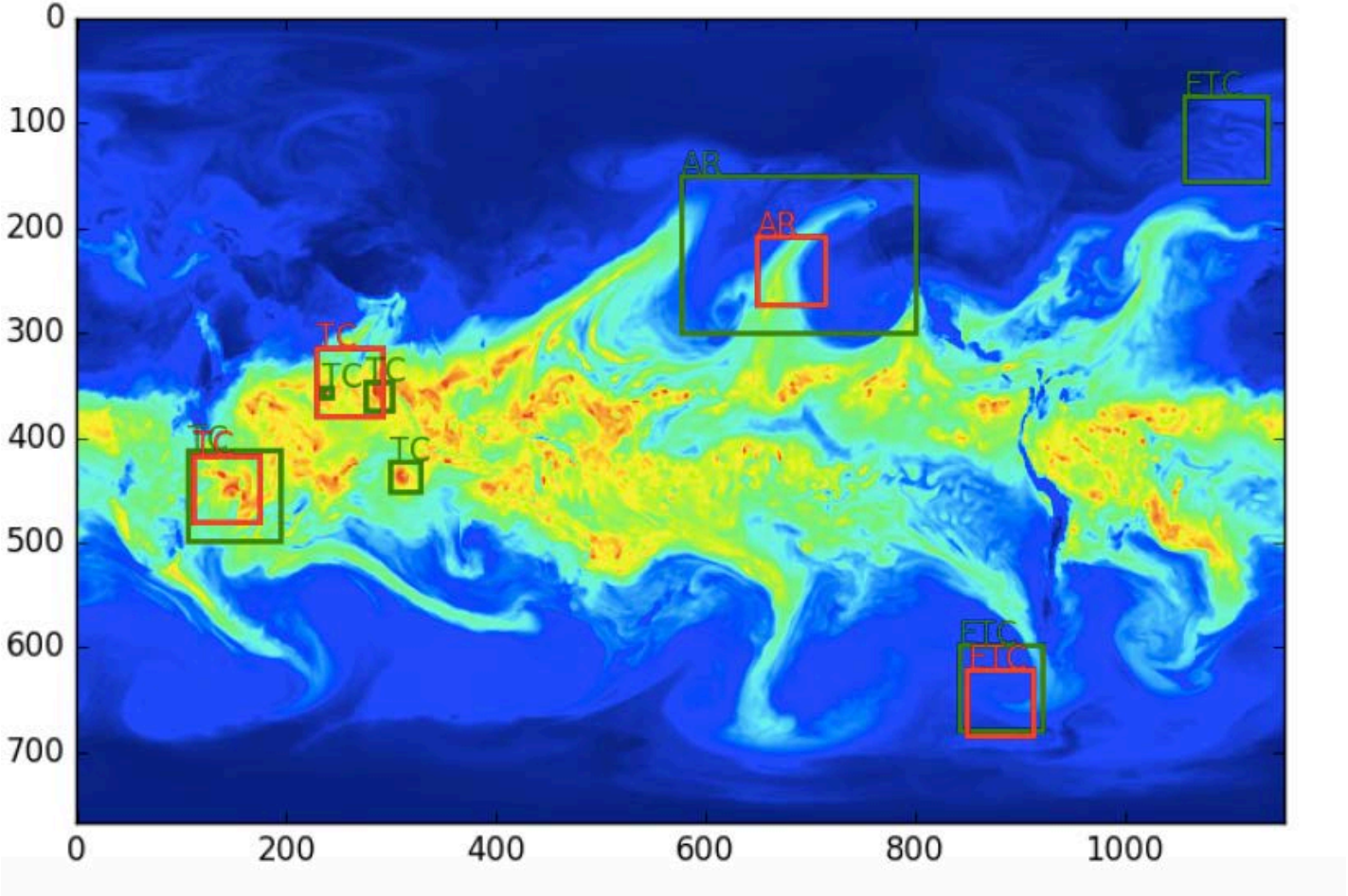
	Logistic Regression	K-Nearest Neighbor	Support Vector Machine	Random Forest	ConvNet
	Test	Test	Test	Test	Test
Tropical Cyclone	95.9	97.9	95.9	**99.4**	**99.1**
Atmospheric Rivers	82.7	81.7	83.0	88.4	**90.0**
Weather Fronts	89.8	76.45	**90.2**	87.5	**89.4**



Semi-Supervised Detection

- **Objectives:**
  - Create unified architecture for all weather patterns
  - Predict bounding box location for each weather pattern
  - Work with few/no labels for several weather patterns
  - **Discover new (hitherto unlabeled) patterns**

Classification + Detection (NIPS17)

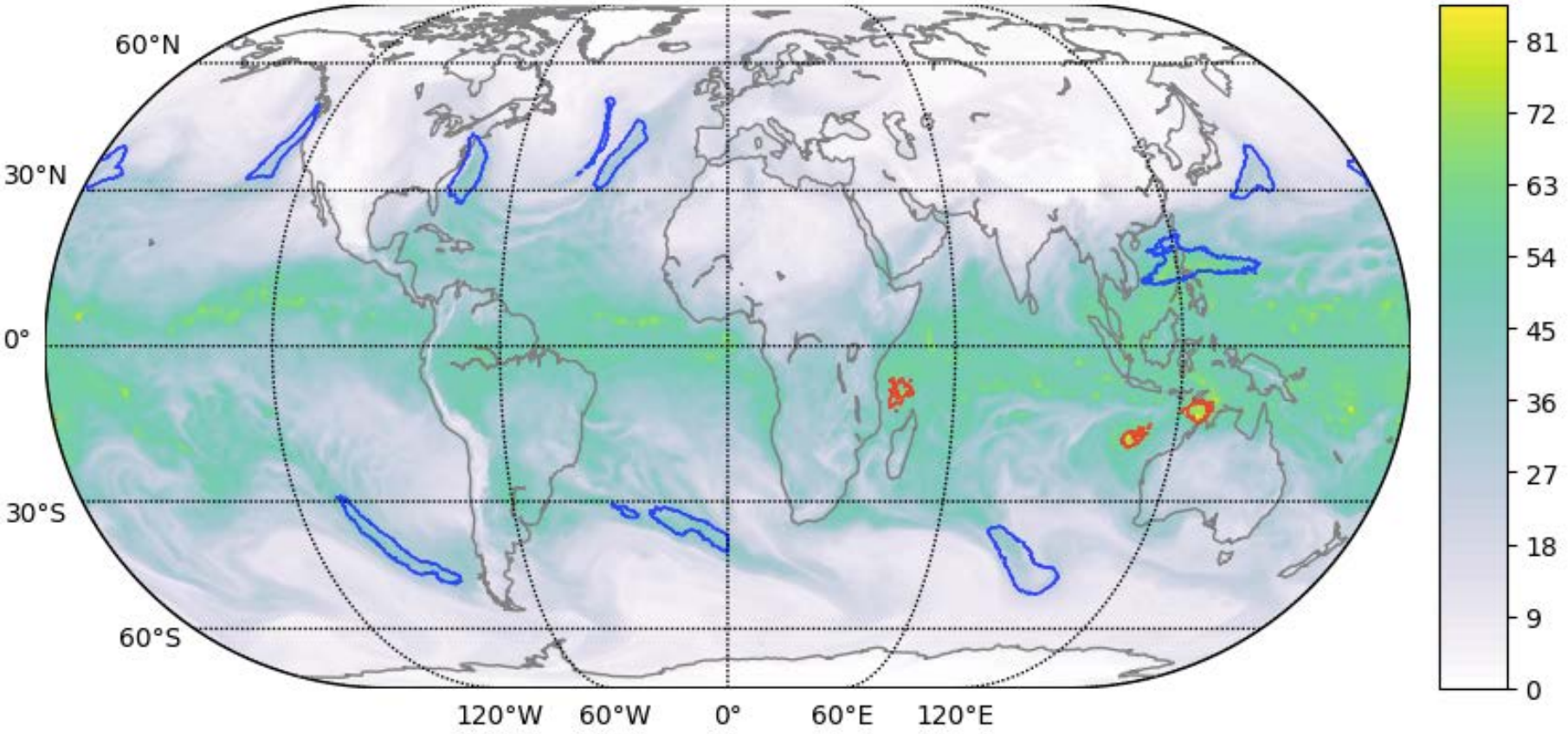


Ground Truth
Prediction

Contributors: Evan Racah, Chris Pal, Chris Beckham, Samira Kahou, Tegan Maharaj. *MILA*

Pixel-wise Segmentation: SC'18

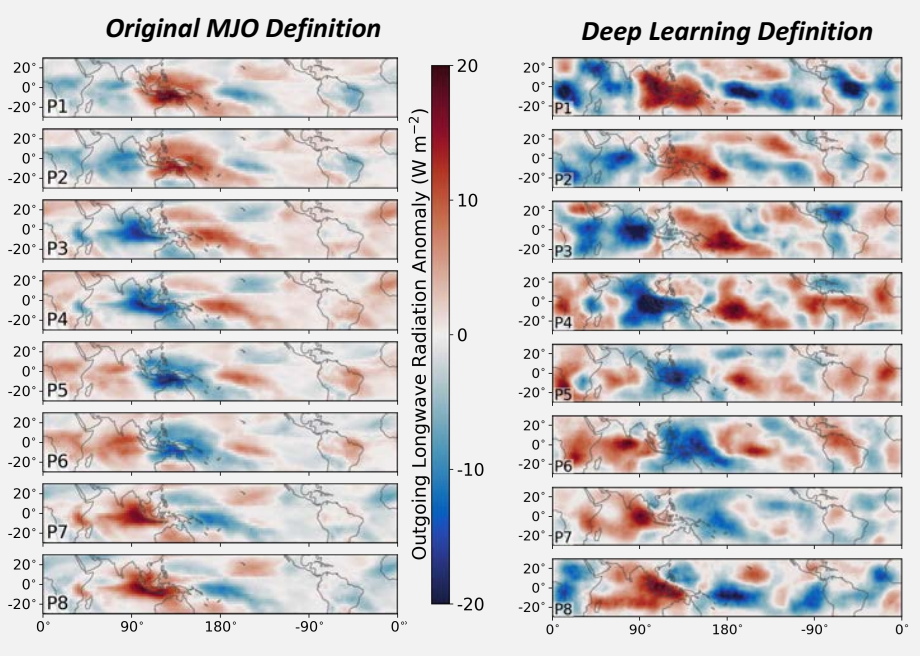
Gordon Bell Prize Winner! DL @ 1EF



Contributors: Thorsten Kurth, Sean Treichler, Joshua Romero, Mayur Mudigonda, Nathan Luehr, Everett Phillips, Ankur Mahesh, Michael Matheson, Jack Deslippe, Massimiliano Fatica, Michael Houston. *NERSC, NVIDIA, UC Berkeley, OLCF*

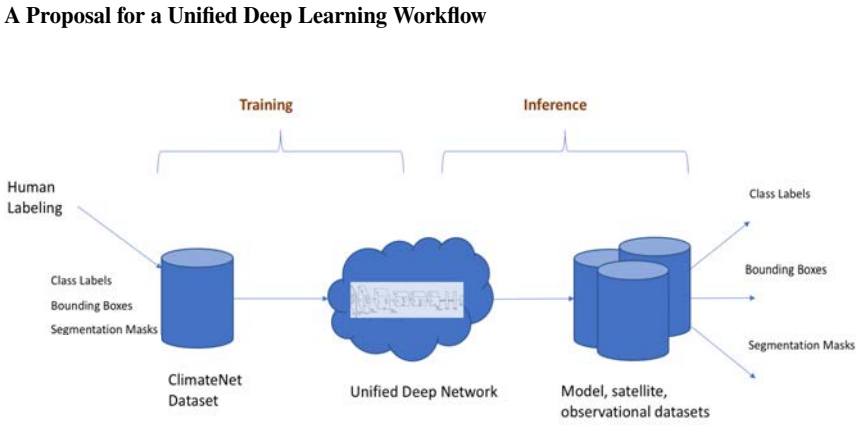
Deep Learning for the MJO

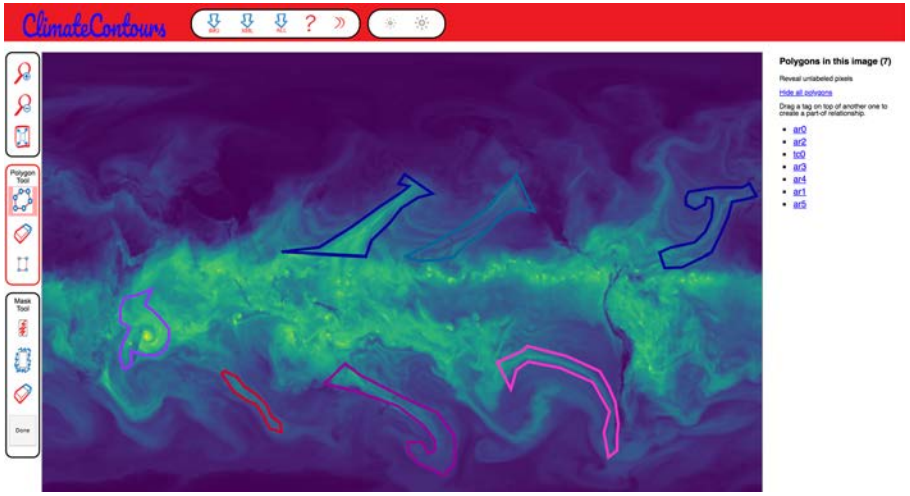
- DL can correctly classify the phase of the MJO
- DL can independently infer which variables are most important for the classification
- DL can learn the spatial structure of the MJO
- https://arxiv.org/abs/1902.04621 (*Deep Learning for Scientific Inference from Geophysical Data: The Madden-Julian Oscillation as a Test Case*)



ClimateNet: a community-sourced labeled dataset (ICML 2019)

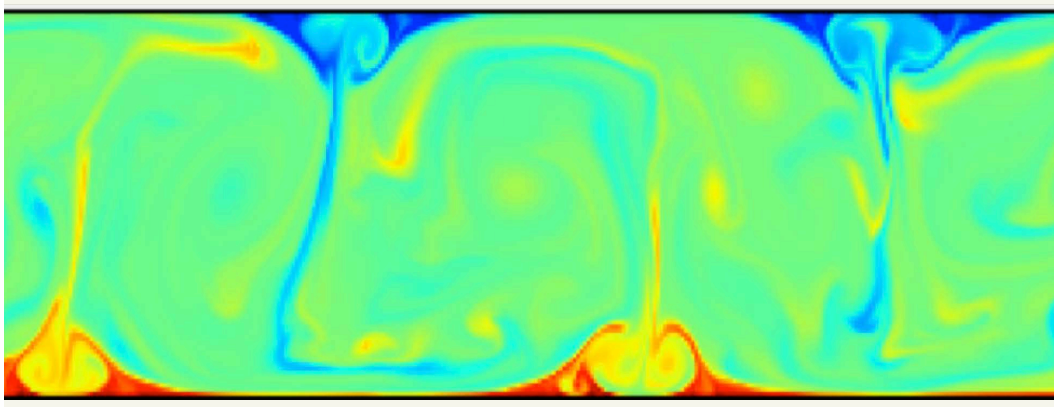
- Pattern recognition using DL works best (so far) in the supervised mode
- Reliable labeled data is sparse in climate science
- Heuristics for pattern detection in climate science fraught with discrepancies and uncertainties
- Ultimately we want to *avoid* heuristics ("feature engineering")
- https://www.nersc.gov/research-and-development/data-analytics/big-data-center/climatenet

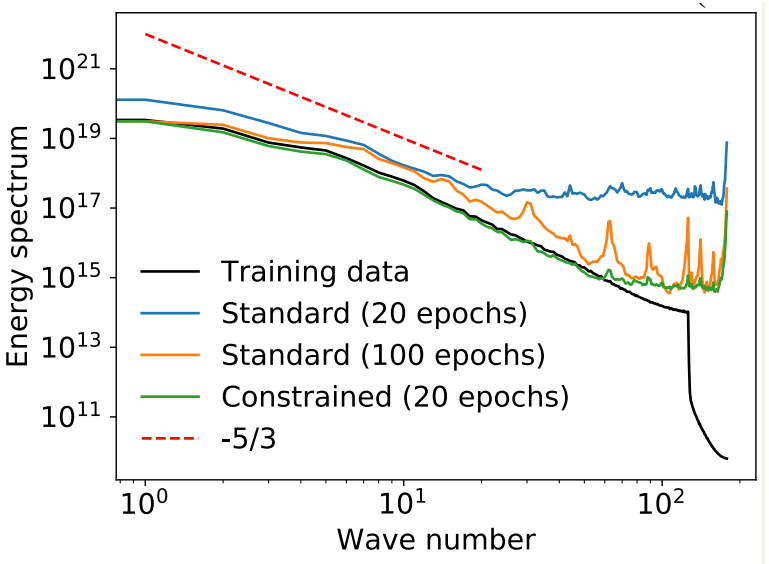




Physics-informed Deep Learning

- Reliable and robust DL models for science need to obey the laws of nature
- A systematic framework for physics-informed approaches in ML and DL is needed
- We've made a first step in this direction with PI-GANs for atmospheric flows





[1] J.-L. Wu, K. Kashinath, A. Albert, D. Chirila, Prabhat and H. Xiao, "Enforcing Statistical Constraints in Generative Adversarial Networks for Modeling Chaotic Dynamical Systems", Submitted, available at arXiv:1905.06841, 2019.

[2] Y. Zeng, J.-L. Wu, K. Kashinath, A. Albert, Prabhat and H. Xiao, "Physics-informed Generative Learning to Emulate PDE-Governed Systems by Incorporating Conservation Laws", in preparation, 2019.

Short-Term Challenges

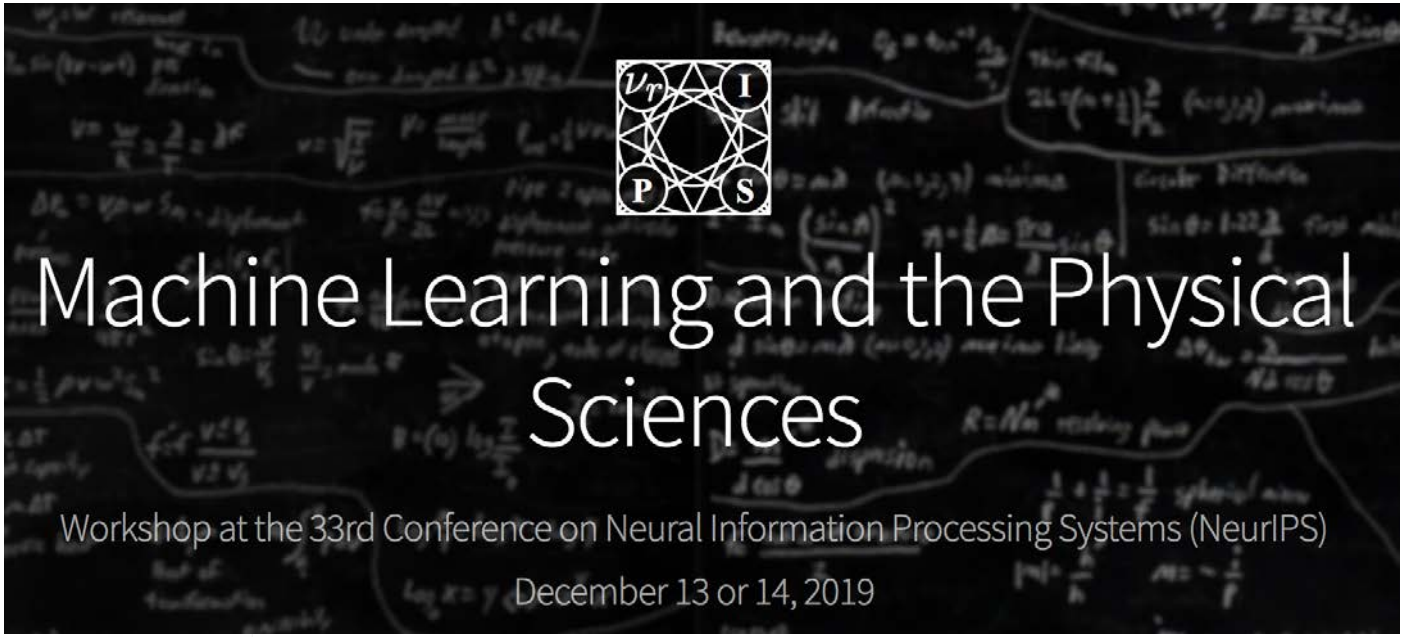
- **Complex Data**
  - 2D/3D/4D, multi-channel, dense/sparse
  - Multi-scale, Spatial and Temporal coherence
- **Hyper-Parameter Optimization**
  - Tuning #layers, #filters, learning rates is a black art
- **Performance and Scaling**
  - Current networks take days to train on O(10) GB datasets, we have O(100) TB datasets on hand
- **Scarcity of reliable, high-quality Labeled Data**
  - Communities need to self-organize and run labeling campaigns

Long-Term Challenges

- **Theory**
  - Limits of supervised, unsupervised, semi-supervised learning
  - GANs are promising, but unstable and hard to train
- **Interpretability and Explainability**
  - Need to visualize representations
  - Incorporate domain science principles (physical constraints, conservation laws, etc.)
- **Uncertainty Quantification**
- **Formal protocol for applying Deep Learning**
  - Applied Math has developed methodology for solving PDEs over decades, no analog in DL

IN033: Incorporating physics and domain knowledge to improve interpretability, explainability, reliability and generalization on Machine Learning Models

AGU Tutorial on Machine Learning and Deep Learning for the environmental and geosciences



Questions?

kkashinath@lbl.gov