National Aeronautics and Space Administration

A unified approach to data science at NASA Goddard Space Flight Center

[1]Mark Carroll, [1]Dan Duffy, [1,2]Craig Pelissier, [1,2]Jules Kouatchou,[3]Grey Nearing, [1,4]Roger Gill, [1,4]Glenn Tamkin, [1,4]Jian Li, [1,4]Thomas Maxwell, [1,4]Mary Aronne, [1]John Schnase

[1]NASA Goddard Space Flight Center, Maryland, USA

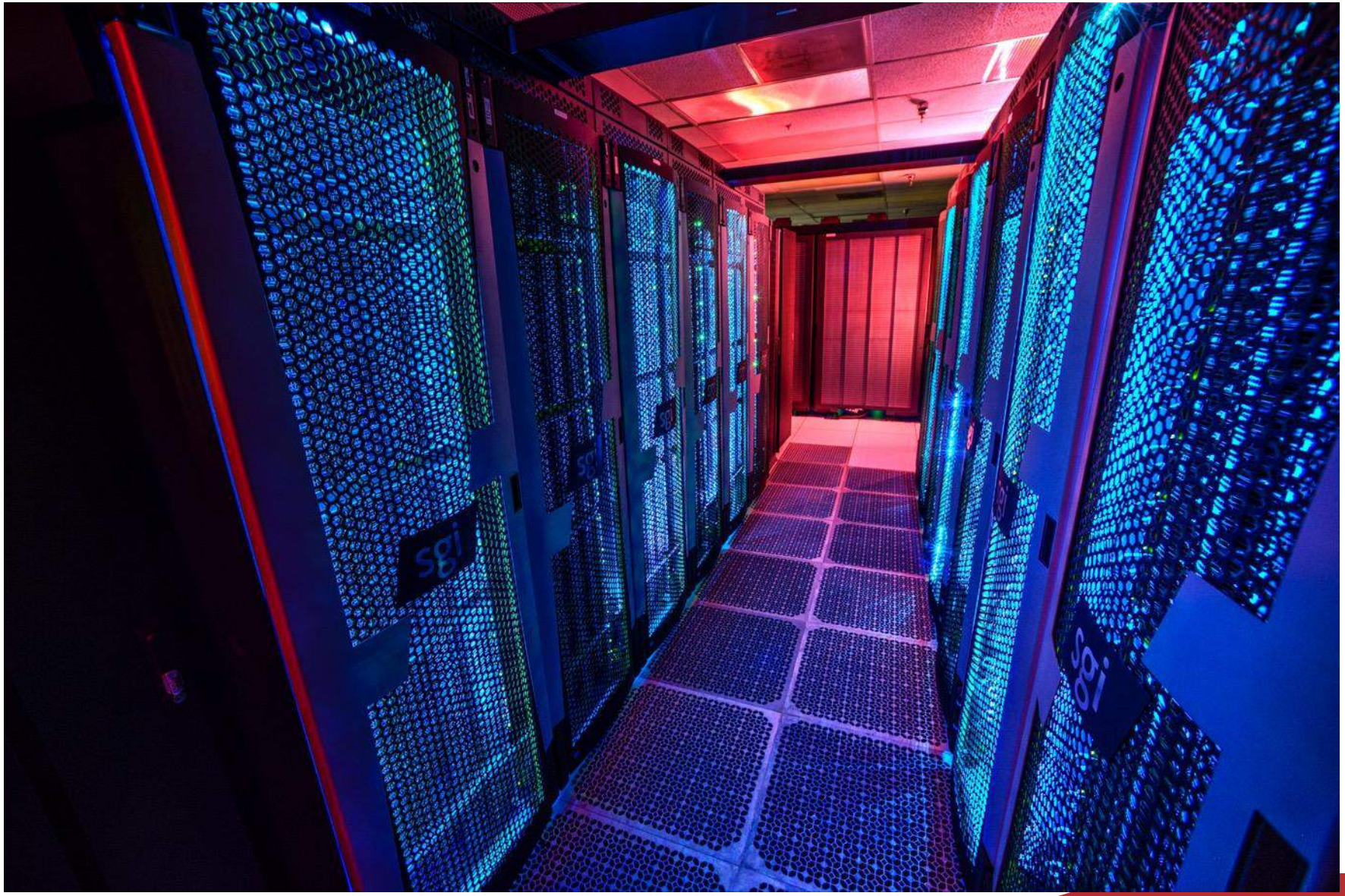
[2]Science Systems and Applications Incorporated, Maryland, USA

[3]University of Alabama, Tuscaloosa, Alabama, USA

[4]Arctic Slope Research Corporation

Presented by Liz Hoy NASA/GST, Inc.

Computational & Information Science and Technology Office (CISTO)



Concept to Accelerate Science and Engineering Discovery

Science Mission Directorate (SMD – Code 600)

Earth Science | Heliophysics | Planetary | Astrophysics

Engineering Directorate (Code 500)

Software Engineering | Electrical Engineering | Instrument Systems &Technology | Mission Engineering & System Analysis

Universities

Connections

AI Center of Excellence (GSFC)

AI Center of Excellence (Agency)

Nargess Memarsadeghi

Other NASA Centers

- JPL
- Ames
- LaRC
- JSC
- Etc.

CISTO Data Scientist

Mark Carroll

Code 500

Nargess Memarsadeghi

Digital Transformation

Jean-Guy Dubois

Purpose and Goal of the AI/ML CoE:

- The purposes of the CoE is to provide all necessary AI/ML resources needed by the SMD. The goal is to accelerate Science Discoveries through the use of AI/ML.

Scope

The of the AI/ML CoE is represented by this diagram.

- IT Resources
- Examples
- Training
- Partnership
- Support

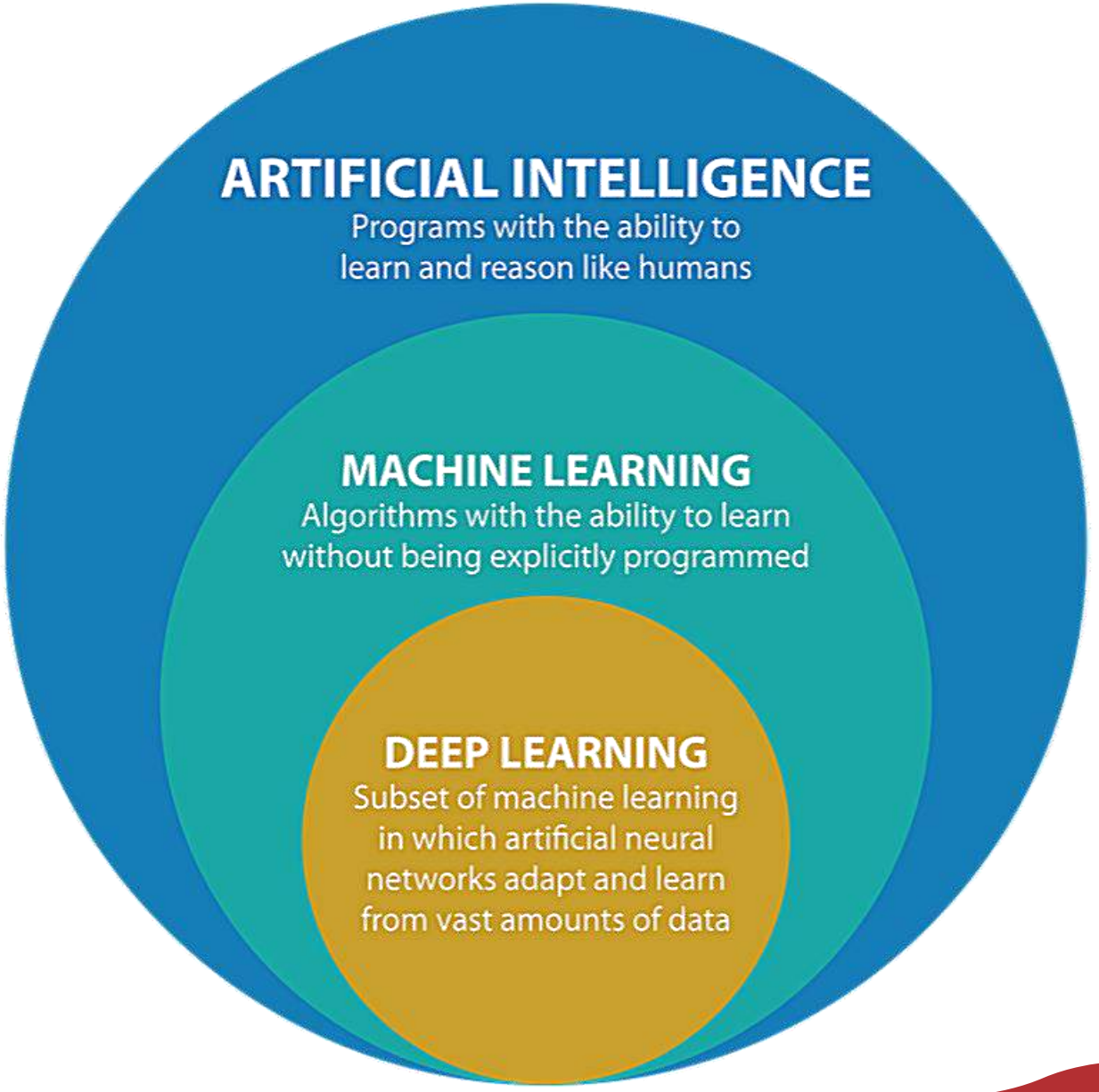
CISTO (Code 606)

Industry

- Technology
- Training
- Intel, AWS, IBM, SBIRs, etc.

Accelerating the discovery through the use of AI/ML, enhancing the competitiveness of science and engineering, and creating the next generation workforce for Goddard.

From AI to ML and DL



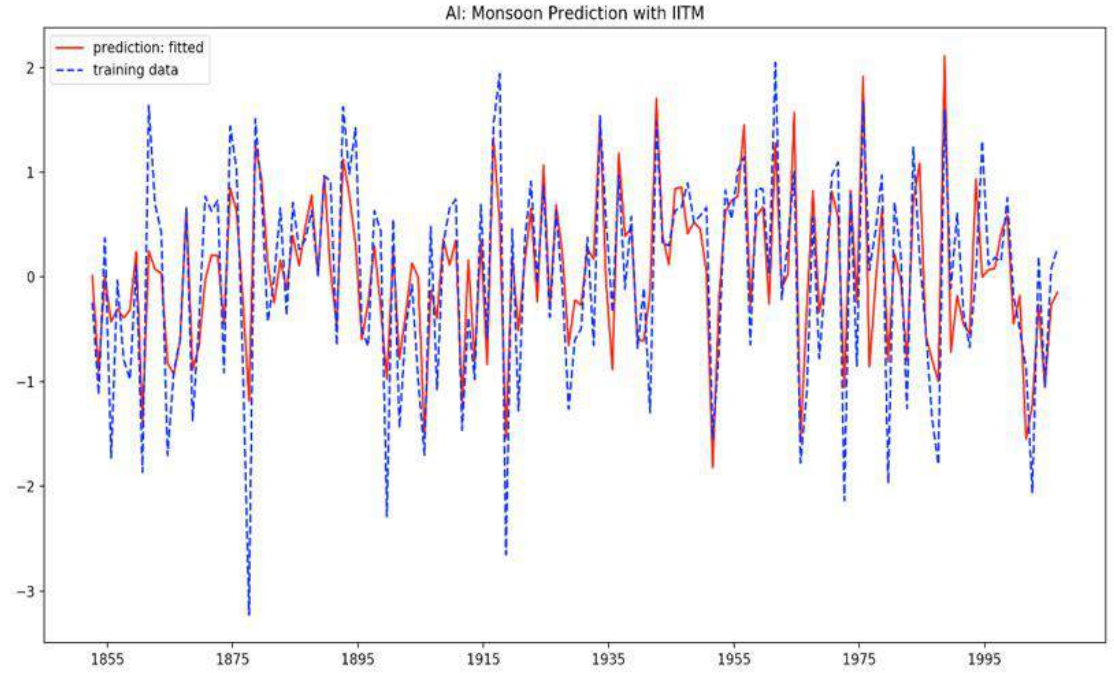
Machine Learning for Monsoon Prediction

Using Global Teleconnections

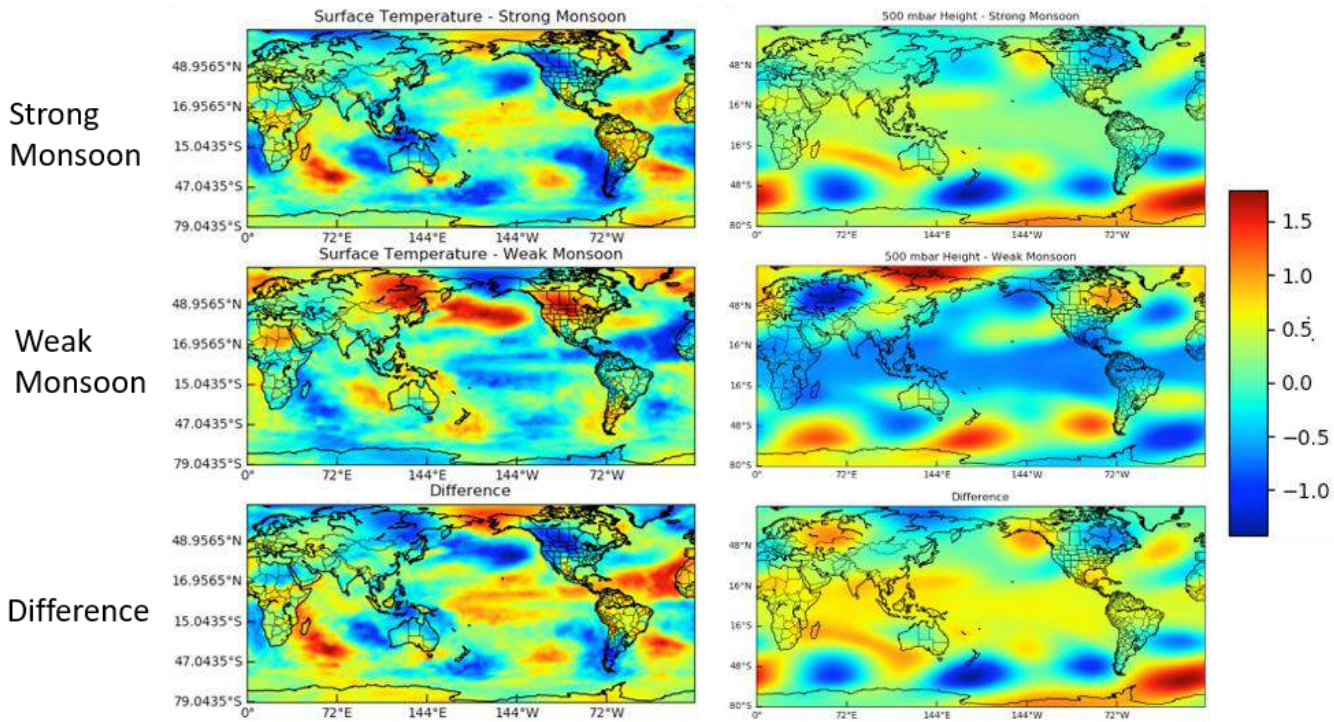
Thomas Maxwell, Jian Li, Thomas Favata

- Predict All-India Monsoon rainfall accumulation one year in advance using a two-layer neural network
- Inputs: First 16 global PCs of surface temperature & 500 mbar height (1 year lag time- August values)

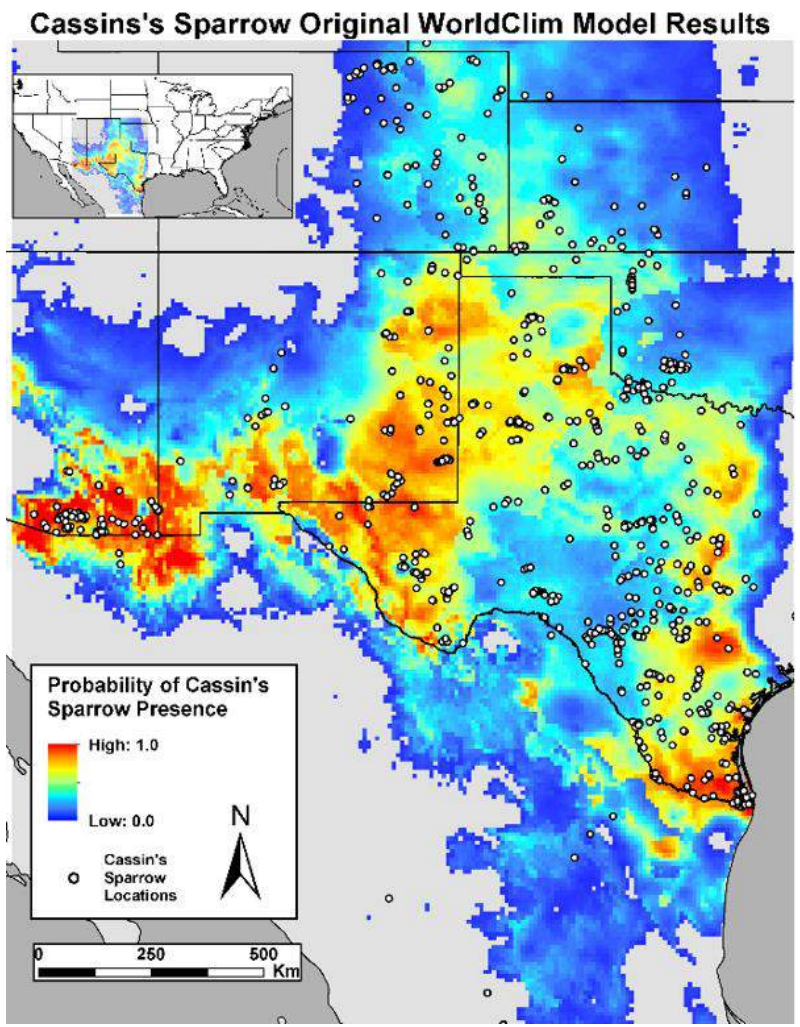
Comparison of predicted to actual monsoon precipitation

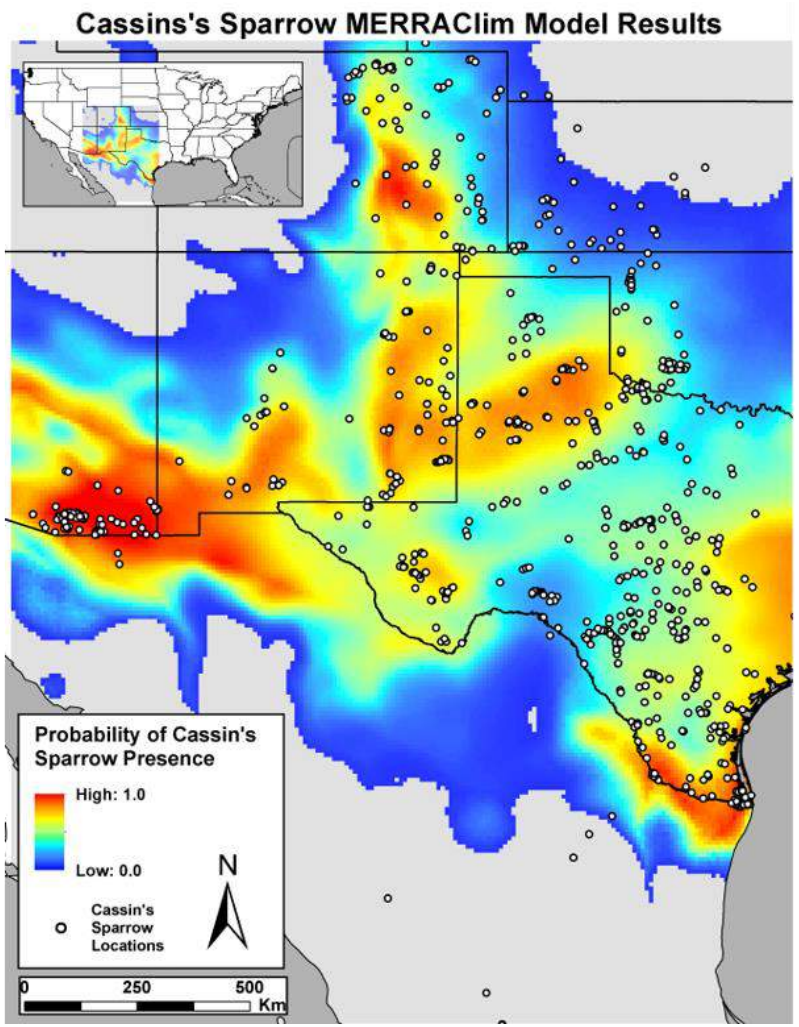


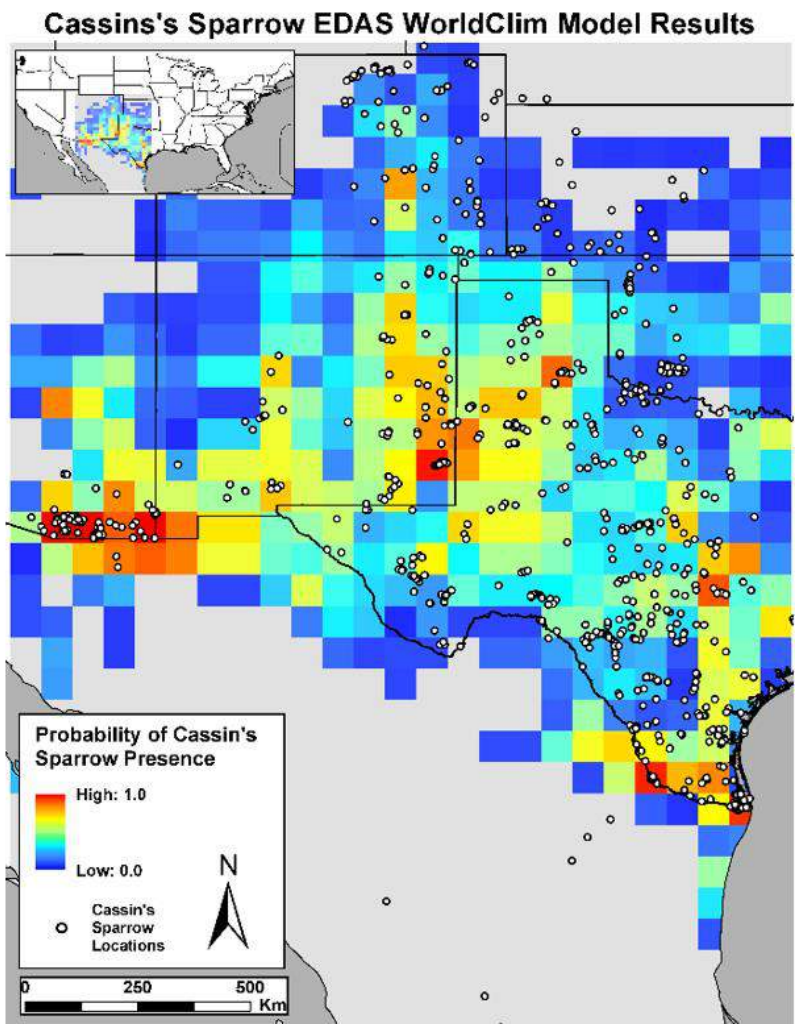
Back-projected Activation Patterns



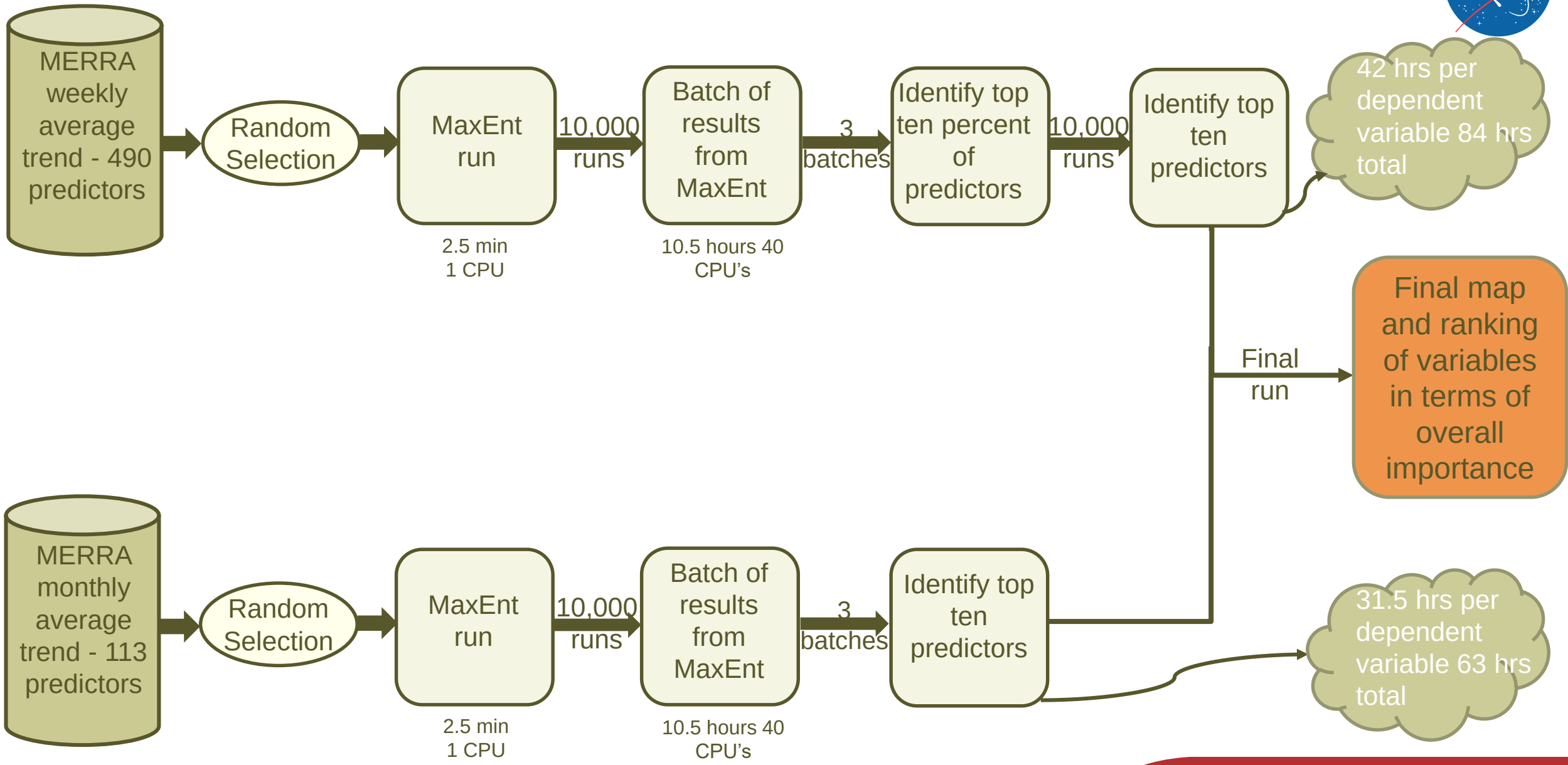
MERRAMax


Cassins's Sparrow Original WorldClim Model Results
Probability of Cassin's Sparrow Presence
High: 1.0
Low: 0.0
Cassin's Sparrow Locations
N
0 250 500 Km

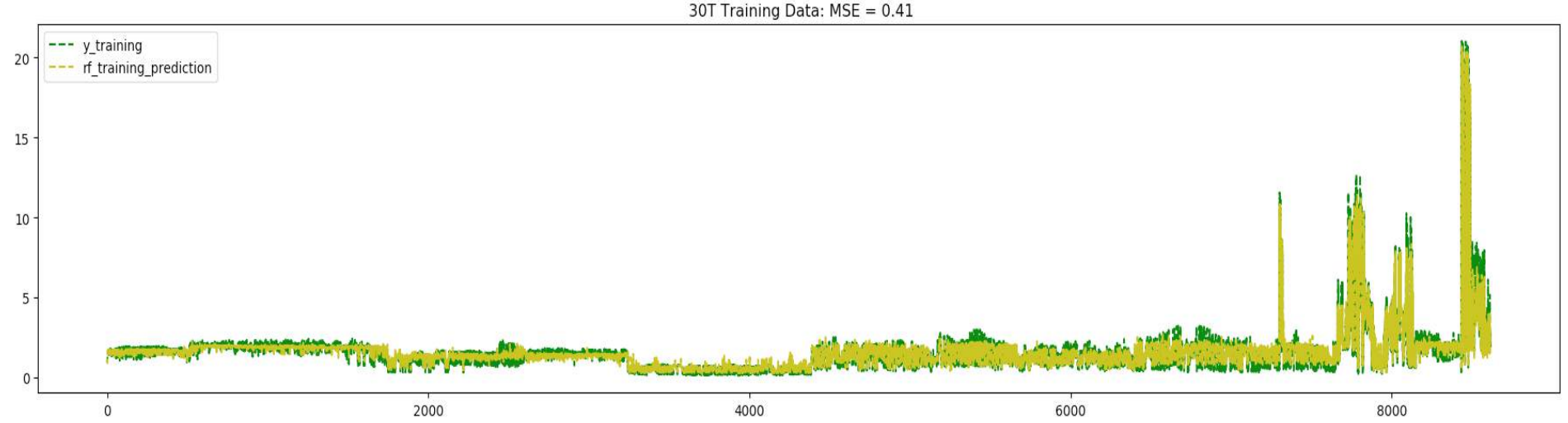

Cassins's Sparrow MERRAClim Model Results
Probability of Cassin's Sparrow Presence
High: 1.0
Low: 0.0
Cassin's Sparrow Locations
N
0 250 500 Km

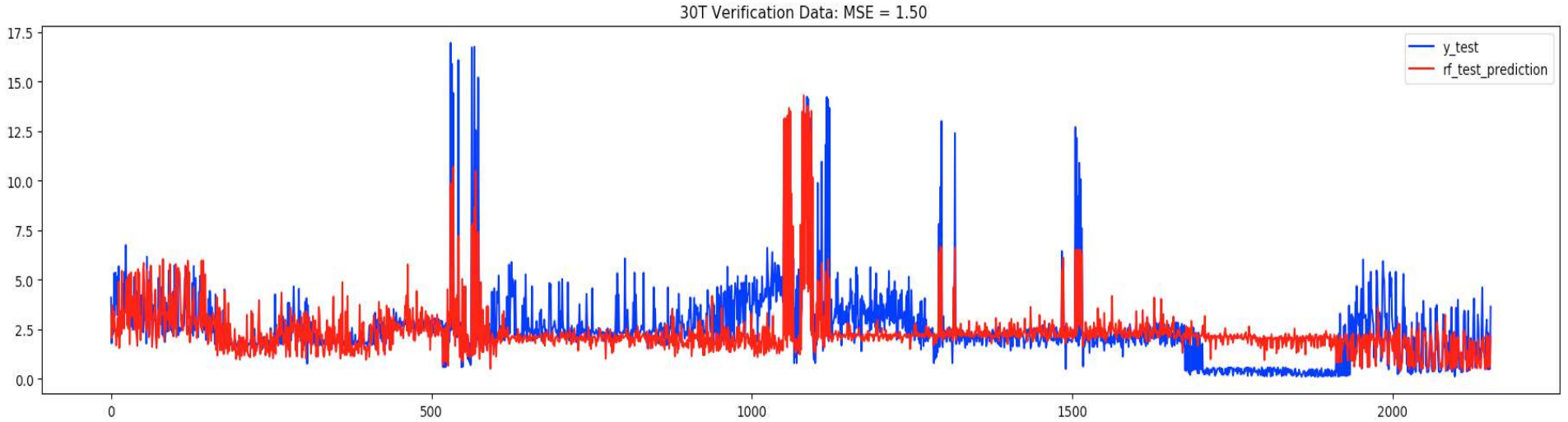

Cassins's Sparrow EDAS WorldClim Model Results
Probability of Cassin's Sparrow Presence
High: 1.0
Low: 0.0
Cassin's Sparrow Locations
N
0 250 500 Km

MERRAMax automated variable selection

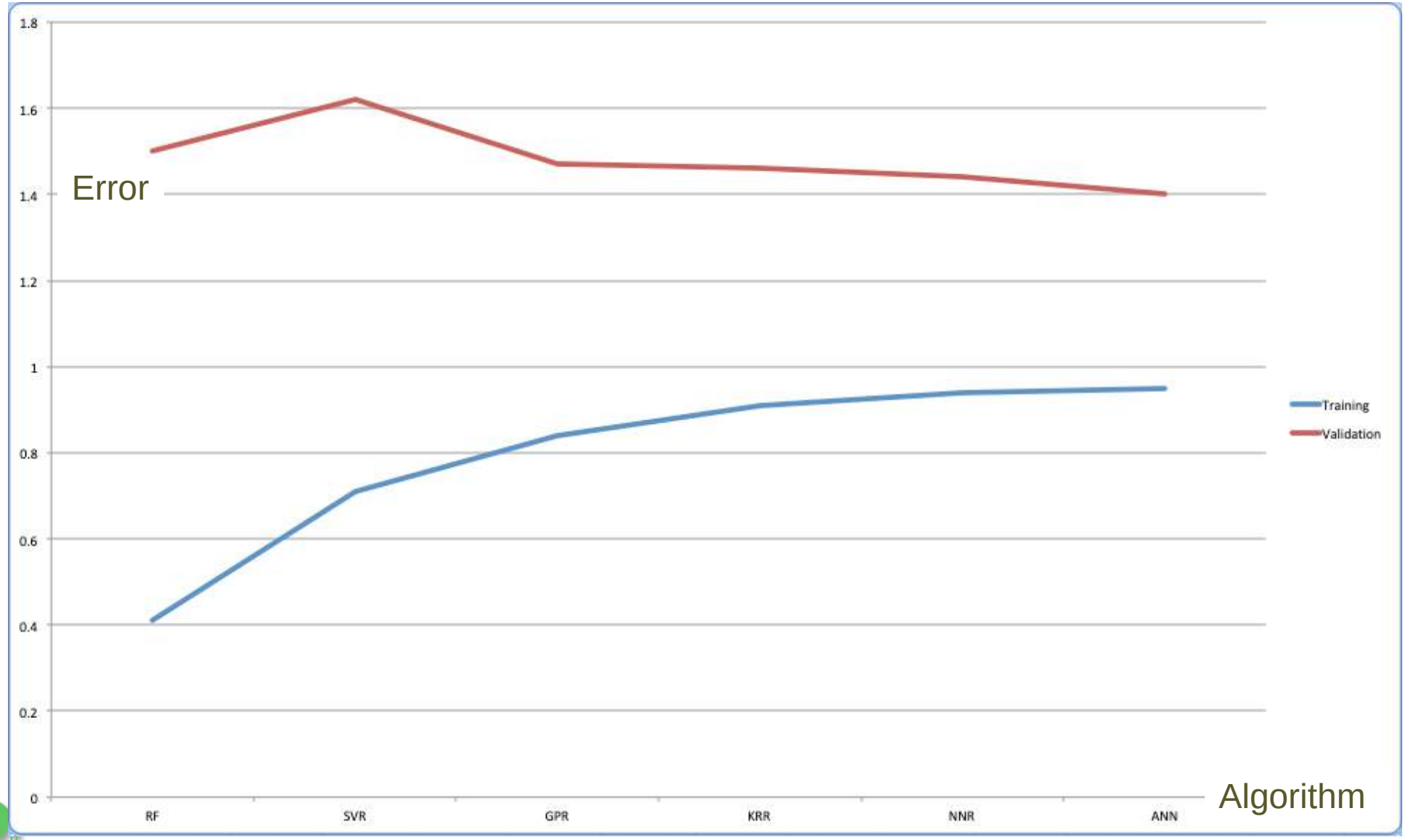


Random Forest for Lake Depth

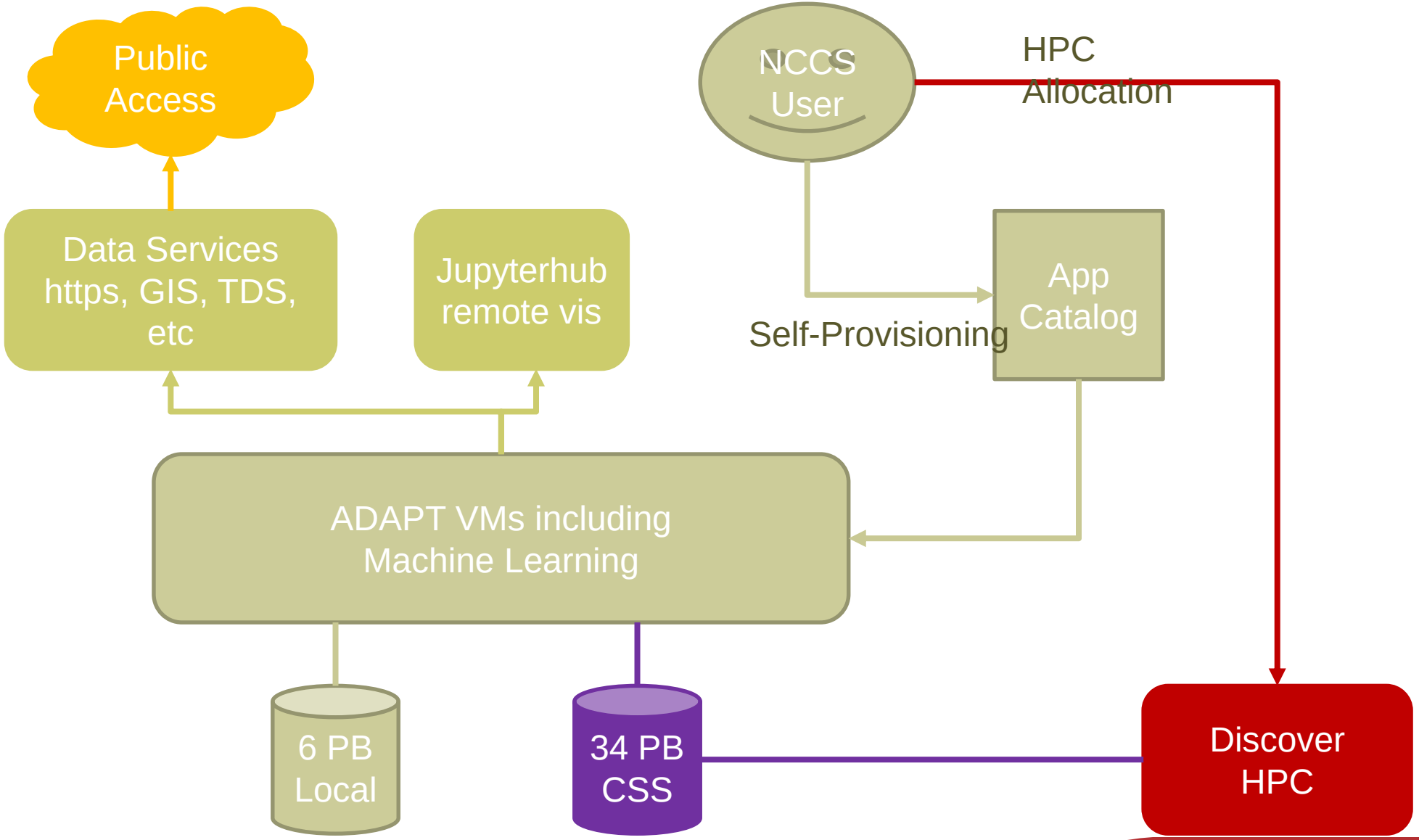




Multiple ML Comparison



ADAPT 2.0 Planned Configuration



What's next?

- **Continue to offer platform as a service with Discover and ADAPT**
  - Explore and test new hardware options to enhance the current hardware options – ADAPT 2.0
  - New GPU cluster with 20 4 GPU systems will be online by early 2020
- **Increase data holdings at NCCS to facilitate discovery through AI/ML**
  - ICESat-2 data are currently being added to the collection
  - Over 3.5 PB of Digital Globe data available
  - Data buy (Planet labs, expanded DG data) data are being stored and will be available soon
- **Develop software collaboratively with scientists to maximize utility**
- **Training and outreach**

Thanks!