

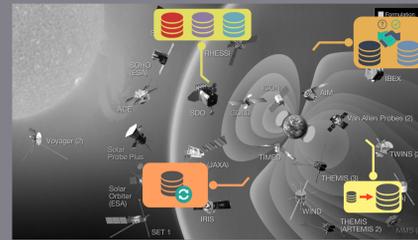
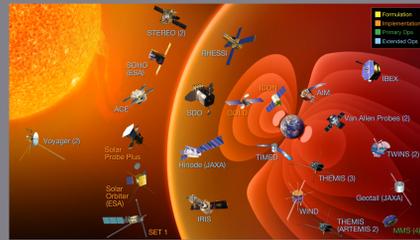
Ushering in a NEW FRONTIER in Geospace Through Data Science

Learn more, collaborate, and
be a part of the New Frontier



@AeroSciEngineer
rmcgranaghan
ryan.mcgranaghan@gmail.com

What is the current space science data landscape and why does it need data science?



'Big Data' ...not just volume

- Volume
- Variety
- Veracity
- Velocity

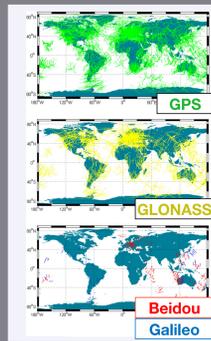
'Data Science' is...

- "Scalable architectural approaches, techniques, software and algorithms which alter the paradigm by which data are collected, managed and analyzed." – Dan Crichton, NASA JPL

Opportunity:

- Evol**ve traditional approaches
- Em**brace data science-driven discovery
- En**able interdisciplinary work

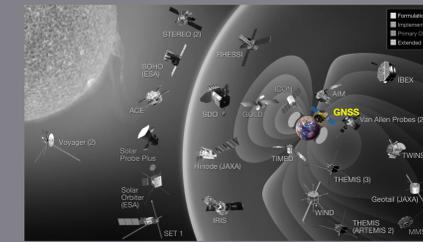
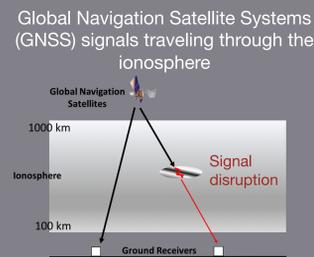
These data are illustrative of challenges and opportunities of space science data



Single day of observation locations from GNSS signals

GNSS signals:
Critical resource to understand space weather
Ideal use case for data science in space science

Total electron content (TEC) data, inferred from GNSS signal delays during passage through the ionosphere, provide critical information about the Earth's ionosphere at higher cadence and over a larger portion of the globe than any other single data set

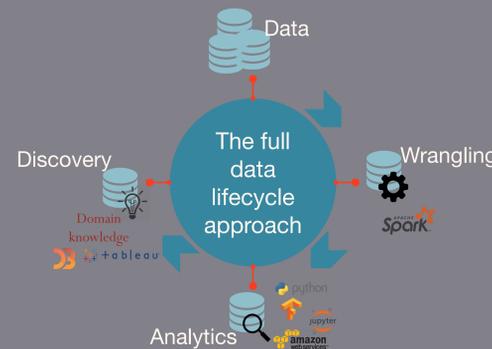


Global Navigation Satellite Systems (GNSS) are among the most important systems sensitive to space weather, but are also one of the premier tools to facilitate new space weather understanding

What if space science were an exploration, data-driven science?

What is the potential for big data technologies and machine learning to usher in a New Frontier in space science?

JPL Data Science Working Group Pilot Program



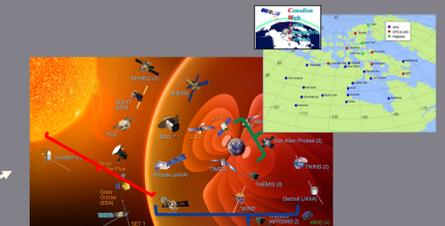
Why machine learning?

Problems well-suited to machine learning:

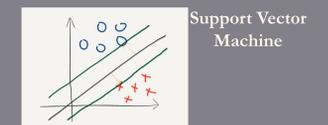
- Classification
- Event detection
- Segmentation
- Clustering
- **Prediction**
- Recommendation

Steps to successful machine learning:

1. Obtain data
2. Define predictive task
3. Choose ML algorithm
4. Understand the model

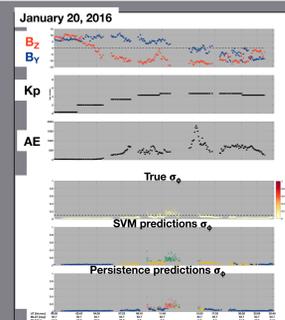
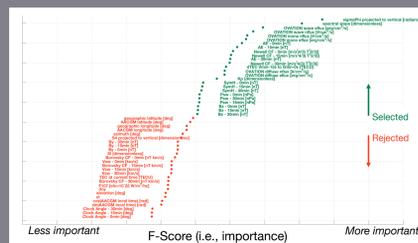


Given data available now...
...will GNSS signal be disrupted in future?



Machine Learning:
Know when and how to use
Investigate relationship to physics
Understand the model (i.e., explainability)

Investigate information in the data



Integrate model with domain knowledge

Evaluation & Explanation

- Integrate data-driven and domain knowledge
- Obtain new physical insight
- Improve the models

no scintillation	True negative	False positive	0.67	0.33
scintillation	False negative	True positive	0.09	0.91
		no scintillation	scintillation	
		Predicted label		

$$\text{TSS} = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

TSS scale from -1 (Worst Previous state of the art) to +1 (Perfect New benchmark with SVM)

Ambitious pilot projects:
Be radically interdisciplinary
Explore massive space of cutting-edge data science-driven approaches
Utilize innovative data science tools and technologies

What does this mean across the space sciences?

Key trends for the New Frontier:
Be **radically interdisciplinary**
Understand the models
Be **open by default**

Opportunities through constantly evolving data landscape

Compelling use cases

Methods for progress

Machine learning + traditional approaches

The New Frontier

