Ushering in a new frontier in geospace through Data Science

Ryan McGranaghan[1], Anthony Mannucci[1], Chris Mattmann[1], and Brian Wilson[1]

[1]NASA Jet Propulsion Laboratory

November 23, 2022

Abstract

We are at a unique time in the study of our place in space. On one hand, we operate in the same paradigm that has guided the study of space science for the past couple of decades, and on the other a rising dependence of our economic and social well-being on space demands a shift. Everywhere in our society 'big data' (defined by four V's: volume, variety, veracity, and velocity) and the advent of sophisticated and efficient methods to explore these data (i.e., data science) present new opportunities for discovery, and the time is ripe for these methods to shift how we study the physics of space. We will first discuss the meaning of data science in the context of space science, and then demonstrate the potential for new discovery through a power use case: leveraging Global Navigation Satellite Systems (GNSS) signals for space weather prediction. In this use case, we take advantage of a large volume of data from GNSS signals, data science-driven technologies, and a machine learning algorithm known as the Support Vector Machine (SVM) to develop a novel predictive model for high-latitude ionospheric phase scintillation. This talk will conclude with a perspective on opportunities in space science through 'big data' and creating new scientific discovery at the intersection of traditional approaches and data science-driven innovation.

Ryan McGranaghan[1,2], Anthony J. Mannucci[2], Brian D. Wilson[2], Chris A. Mattmann[2]

[1] CPAESS

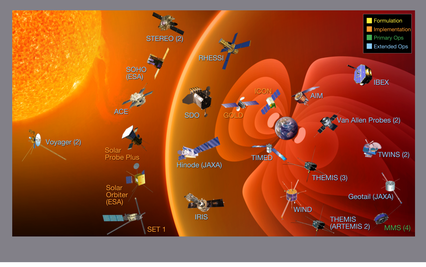
[2] Jet Propulsion Laboratory, California Institute of Technology

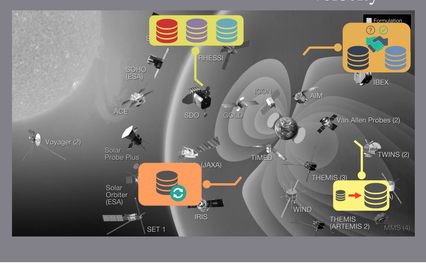
Ushering in a NEW FRONTIER in Geospace Through Data Science

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?





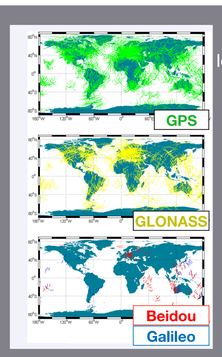
'Big Data'…not just volume
- Volume
- Variety
- Veracity
- Velocity

'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:
Evolve traditional approaches
Embrace data science-driven discovery
Enable interdisciplinary work

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

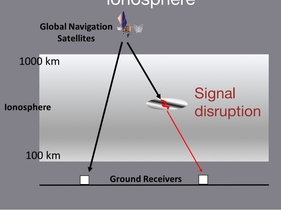
GNSS signals:
Critical resource to understand space weather
Ideal use case for data science in space science

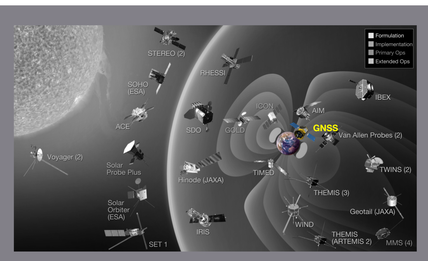


Single day of observation locations from GNSS signals

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) signals traveling through the ionosphere



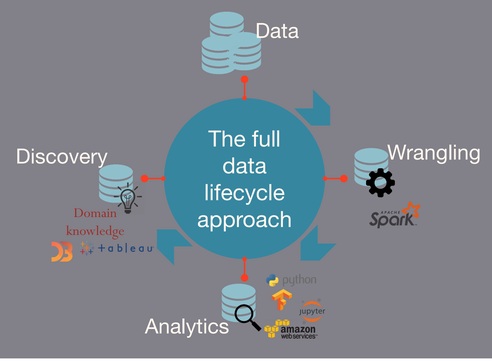


Global Navigation Satellite Systems (GNSS) are among the most important systems sensitive to space weather, but are also one of the premier tools to facilitate new space weather understanding

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?

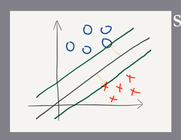
JPL Data Science Working Group Pilot Program



Why machine learning?

Problems well-suited to machine learning:
- Classification
- Event detection
- Segmentation
- Clustering
- **Prediction**
- Recommendation

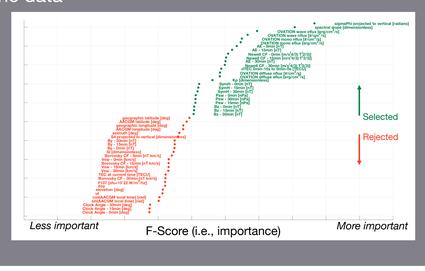
Steps to successful machine learning:
1. Obtain data
2. Define predictive task
3. Choose ML algorithm
4. Understand the model

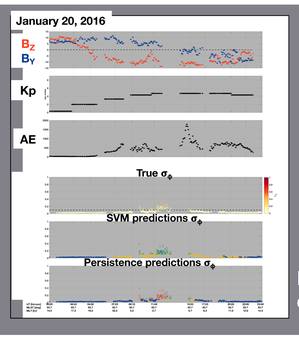


Support Vector Machine

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

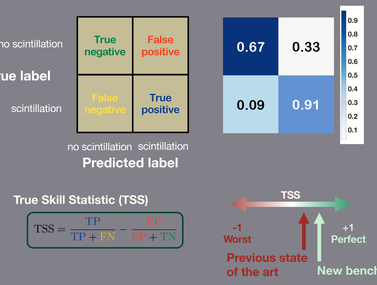
Investigate information in the data





Integrate model with domain knowledge

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



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