2018 Workshop: Big Data Challenges

Long title
Exploring the challenges related to "Big Data" in CEDAR science
Conveners
R. Bishop
Anthea Coster
Romina Nikoukar
Kshitija Deshpande
Cheryl Huang
Description

In the near future we can expect to have ever larger volumes of many types of observational data available to the CEDAR and space science community due to the continual addition of ground-based sensors (e.g. photometers, ionosondes, MST radar, etc.) and the growing popularity of CubeSat constellation concepts. One of the most striking example of a rapidly increasing dataset is GNSS observations. The relative easy and affordable placement of ground receivers and their suitability as small satellite payloads has resulted in a high volume of data, both spatially and temporally. However, GNSS data limits the type of phenomena that can be investigated in detail. For example, dense ground-based GNSS receivers are ideal to explore Travelling Ionospheric Disturbances (TIDs) but are limited to land masses. The reverse is true for space-based GNSS observations which provides global coverage but because of the spatial ambiguity may not be the most appropriate to study phenomena like TIDs.

In order to justify future mission/facility support or the purchase of data from commercial resources, the science community must begin to quantify the optimal amount of coverage needed per data type to address specific science questions. Further, the community needs to begin to identify and address the unique challenges associated with "Big Data". The intent of this workshop is to promote discussion on various aspects of this topic such as: 1) Linking available large ground and/or space-based data sets to their ability to address specific science topics alone or in combination, 2) Present ideas for a global coverage metrics for specific types of observations, 3) Address the potential of cumulative errors resulting from ingestion of large volume data from sensor networks (e.g. Absolute TEC from ground GPS

receivers with various biases) including their impact on global models, 4) Applying "Big Data" approaches to future constellations and ground sensor networks.

Agenda

- Introduction, Rebecca Bishop
- "Ushering in a New Frontier in Geospace Through Data Science", Ryan Mcgranaghan (UCAR)
- "Challenges in accounting for multifaceted GNSS scintillation data (Low & High Rate)", Anthea Coster (MIT/Haystack)/Kay Deshpande (ERAU)
- "Data science & Data fusion: fast processing for huge GNSS, radar and imaging datasets", Michael Hirsch (BU)
- "Ham Radio for data science usage", Nathaniel Frissell (NJIT)
- "Modeling high-latitude ionospheric instabilities and scintillation: challenges and future needs", Matt Zettergren (ERAU)
- "Determining optimal LEO GPS constellation size", Rebecca Bishop (Aerospace)
- "The current state and future of the DMSP SSIES database", Marc Hairston (UTD)
- "Effectively Supporting the Data Hungry Geospace Community Now and In The Future: Experiences From the Madrigal Database", Phil Erickson (MIT/Haystack)

Justification

The availability of large datasets is allowing for a more systems approach to space weather. However, there needs to be a transition from collecting large volume data to focusing on the application of "Big Data" techniques and system science. This workshop will begin to address where that transition should occur as a function of data type and observational approach. This workshop directly supports the goal "Linking Science with Societal Needs" by supporting the application of "A robust systems approach to understanding and predicting space weather" as stated in the 2013 NSF Geospace Science Plan. Additionally, by defining the data volume needed to address specific science questions, this workshop will lay the foundation to support the 2011 Strategic Plan's Systems Perspective goal of addressing "Cross-Scale Coupling" (Section 3.2)

The topic proposed will be addressed by short presentations followed by a round-table discussion. The purpose of the workshop is to promote discussion on the topic and identify issues and potential solutions to the growth in volume of observational data. Progress from the workshop will be measured by the improved coordination

between experimenters and modelers on planning and utilizing current and future large data sets. In particular, the workshop will work to quantify the amount of data coverage required to address a subset of space weather topics.

View PDF