2017 Workshop: Digital Geospace

Long title
Geospace science in the digital age: New tools and methods
Conveners
Asti Bhatt
Ethan Miller
Todd Valentic
Nathaniel Frissell
Tomoko Matsuo
Elizabeth MacDonald
Description

The geospace community relies on data from disparate sets of instruments and various modeling frameworks involving different boundary conditions. Facilities like CEDAR and later Madrigal database and CCMC have come into existence to store and curate the data and models produced by the community. However, we are still lagging significantly in making effective use of all the data that is generated from a multitude of instruments with high cadence. We are making some progress towards creating data integration tools that would allow models like AMIE to imbibe the data with ease. But the individual researchers still face challenges in navigating the data/software landscape to both use and share with ease. While there are some who have embraced the open source ways of documenting and sharing their software, many tools used in our communities reside on researchers' own computers in proprietary languages. How do we navigate these diverse sets of resources and make them available for effective collaboration?

The data and software are going to only increase as we go forward. Both NSF and NASA have put significant resources towards data management that has also brought in computer scientists to collaborate effectively with geoscientists. For example, the NSF has put their energy behind 'EarthCube', a program that explicitly brings together geoscientists and computer scientists to solve the modern problems in data and software management in geosciences. Similarly, many new digital tools and methods from open source sharing to crowdsourcing have been commonplace that the geospace community is only catching up to now. We want to continue the

conversation on effective practices of data and software sharing and learn about new digital initiatives that the field of geospace sciences can employ.

This session will follow the 'Geoscience paper of the future (GPF)' plenary tutorial by Dr. Yolanda Gil from University of Southern California, who is a computer scientist with a focus on geosciences. The GPF is an initiative to encourage geoscientists to publish papers together with the associated digital products of their research. She will be at the session for further discussion generated from the plenary talk.

Tutorial description for Geoscience Paper of the Future (pdf)

We invite short 2-5 slide presentations or demos on existing or planned digital resources, tools, observation platforms, as well as needs and challenges related to digital practices in geospace science. We would like to have plenty of time for discussions on how to make these tools widely available and create best practices for our community.

Agenda

1:30-2:00 Q & A on Geoscience paper of the future (Yolanda Gil) - Following up on the tutorial on Monday morning, Yolanda will answer more in-depth questions about best practices for data and software stewardship in CEDAR community

2:00-2:10 Embracing data science in heliophysics (Ryan McGranahan) - Approaches to accomplish system-level geospace science

2:10-2:25 HamSCI: The Ham Radio Science Citizen Investigation (pdf) (Nathaniel Frissell) - What is HamSCI? Current techniques for gathering and analyzing data from the amateur radio community and How we are applying this to the 2017 Total Solar Eclipse.

2:25-2:35 Red Pitaya as an Ionospheric Sounder (Michael Hirsch) - Open source hardware and software for ionospheric applications

Second hour will focus on community software resources followed by open discussion.

2:35-2:45 DavitPy notebooks (pdf) (Ashton Reimer) - SuperDARN's data access interface and Jupyter notebook way of writing tutorials and other documentation practices

2:45-3:00 AMIEpy (Liam Kilcommons, Tomoko Matsuo) - Openly accessible version of Assimilative Mapping of Ionospheric Electrodynamics model

3:00-3:10 Integrated Geoscience Observatory (pdf)(Todd Valentic, Asti Bhatt) - A new community resource based on JupyterHub that allows users to discover and collect data from different ionospheric data sources (e.g. SuperDARN, All-sky imagers) and analyze/visualize it along with running AMIE

3:10-3:30 Open discussion on vision for improving geospace science in digital age

Justification

Characterization of geospace system relies on observations from various instruments, models with a variety of frameworks and inputs, effective analysis of data and assimilation of these data into the models to improve predictive capability of the models. The 2012 Decadal Survey for Solar and Space Physics recognized the need to create an effective data environment to enhance the space physics research. This includes coordinated development of data systems infrastructure, community based data mining and assimilation tools, exploitation of emergeing technologies and community oversight of emerging, integrated data systems. The decadal survey also recognized that much of the data infrastructure and analysis tools development has happened in an uncoordinated fashion. With the exception of Madrigal database and CCMC, this is largely true for the geospace communities, CEDAR and GEM included.

The 2011 CEDAR strategic plan thrust #6 calls out the need to Manage, Mine and Manipulate Geoscience Data and Models. The rationale for this approach is to discover correlations, understanding causalities, and contribute to determining the evolution of geospace using data spanning space and time. This includes developing and implementing standardized data formats, and developing data assimilation schemes to integrate data with models among other things.

NASA and NSF have put significant resources in the cyberinfrastructure and data science development in recent years. One of the latest NSF initiatives is called EarthCube that aims to bring geoscientists and computer scientists together to solve scientific challenges in geoscience. Geospace sciences is woefully underrepresented in the EarthCube community even though the concepts being developed through it are equally applicable to geospace sciences. Many other geoscience communities have developed data and software standards and frameworks that aid the scientists

in curating, managing and mining both the data and software tools. Additionally, recent advances in technology and networking offer new data sources that are potentially of great value to the geospace community, but are not yet fully understood or exploited. This includes data sources such as smartphone magentometer readings and observations by amateur radio observers.

We would like to continue the conversation on the topic of data and software standardization, mining, assimilation and curation, and emerging data sources in the CEDAR and GEM fields. This workshop is a follow-up of the workshop conducted during joint CEDAR-GEM meeting in 2016 and we plan for it to be a recurring workshop in subsequent CEDAR and CEDAR-GEM joint meetings.

View PDF