2017 Workshop: Next generation systems science

Long title
Next generation systems science: Embracing data fusion and data science methods to understand geospace complexities
Conveners
Ryan McGranaghan
Tomoko Matsuo
Alex Chartier
Marcin Pilinski
Description

The session will be coordinated by the following objectives and activities.

Objectives:

1. Evaluate the tools available and those still necessary to address the rapidly evolving heliophysics research landscape
2. Inspire the community to embrace advanced methods in space sciences and plan for future workshops and conferences
   a. Fall AGU 2017: Prepare for formal discussions and to involve a larger community
   b. New Geospace Environment Modeling Focus Group (GEM FG): Identify strong links between GEM and CEDAR communities to address geospace complexity

This workshop will serve as a critical component in the long-term goal to promote stronger collaboration within (low atmosphere, upper atmosphere, magnetosphere) and outside (space physics, computer science, applied math) of the space sciences discipline. As such, we suggest that this workshop could be a multi-year effort, depending on interest within the community.

Activities:

- 3 invited scene-setting presentations (15 minutes each)
- Short contributed presentations (~5-10 minutes)
- Culminating discussion of new methods and future plans to enable geospace system science

- Topics will include, but will not be limited to:
- Data fusion and assimilation: comprehensive approaches to combining observations from various geospace observational systems (leveraging the efforts and outcomes of two 2016 CEDAR-GEM workshops: (1) Data assimilation for space weather; and (2) Making sense of high-latitude geospace observations)
- Machine learning: data-mining, classification, and regression
- Big data technologies: software tools for the analysis of large and heterogeneous data sets

Agenda

10:00 - 10:10 -- Opening Remarks (Ryan McGranaghan)

10:10 - 10:25 -- Seebany Datta-Barua (innovative methods for ionosphere-thermosphere discovery)

10:30 - 10:45 -- Chunming Wang (data mining approaches to improve medium-range ionosphere-thermosphere forecasts)

10:50 - 11:05 -- Josh Semeter (scientific insight through computer-aided discovery)

11:10 - 11:20 -- Roger Varney contributed talk (forward-looking in terms of research to operations)

11:20 - 11:30 -- Sierra Flynn contributed talk (Nitric Oxide flux EOFs)

11:30 - 11:40 -- Romina Nikoukar contributed talk (novel data assimilation and extension under new NASA grant)

11:40 - 11:50 -- Asti Bhatt contributed talk (Integrated Geoscience Observatory EarthCube project and coordinated science campaigns)

11:50 - 12:00 -- Open discussion

12:00 - 1:00 -- Extension of discussion over lunch and beyond

Justification

The NRC Decadal Survey emphasizes that the magnetosphere-ionosphere-thermosphere (MIT) system is complex and highly coupled. Understanding and modeling of observed MIT phenomena requires a systems-level (global) approach and a wide range of expertise. Issues of deciphering complexity through systematic data utilization also confront fields such as economics, social sciences, and medicine, and their solution has overwhelmingly been to embrace advanced mathematical tools and computational technologies to advance understanding. Given the emphasis on complexity and system science in the CEDAR Strategic Plan the CEDAR community has a great opportunity to take advantage of the latest tools and technologies. The CEDAR summer workshop is an ideal forum to coordinate such an effort.

CEDAR has already begun to focus on necessary new methods and technologies. Previous workshop sessions have focused on data fusion/assimilation (Making sense of high-latitude geospace observations: Modeling, data fusion and assimilation; Data assimilation for space weather), data processing technologies (Snakes on a spaceship: Python in space science), and system science (Grand Challenge: The high-latitude geospace system). This session will leverage these previous efforts and discuss how to create new progress.

Our aim is to bring together multi-disciplinary approaches (space physics, statistical analysis, and computer and data sciences) to:

1. Identify promising methodologies that support the systems science approach to understand geospace;
2. Outline the paths from methodology to new fundamental understanding; and
3. Determine how to integrate these new tools into CEDAR research.

To accomplish these goals this workshop will focus on innovation to address the complexities of system science research, using the MIT system as a use case. We most directly respond to the Key Science Goal 2 of the NRC Decadal Survey: Determine the dynamics and coupling of Earth’s magnetosphere, ionosphere, and atmosphere and their response to solar and terrestrial inputs. Additionally, this session supports progress towards several of the CEDAR Strategic Plan Thrusts:

1. Strategic Thrust #1: Encourage and Undertake a Systems Perspective to Geospace
    - We will coordinate methods capable of providing new understanding about complex processes in the MIT system, including multi-scale specification, cross-scale feedback, and nonlinearity.
2. Strategic Thrust #5: Fuse the Knowledge Base across Disciplines
    - We will start a conversation to identify synergies between the field of

space science and the quickly emerging fields of data fusion, data science and machine learning.

• We hope to create a multi-disciplinary community dedicated to the objectives outlined above to promote sessions at future workshops and conferences.

3. Strategic Thrust #6: Manage, Mine, and Manipulate Geoscience Data and Models

• Innovative methods will contribute to more effective utilization of the geospace observational system.

Summary

Central points:

- Given the expanse of the geospace system, we must rely on distributed, diverse observations to provide observational support for new understanding
- Innovation is required to gain the most utility from our diverse observational system
- Data-driven efforts deliver on understanding through innovation. Three categories were highlighted in our session:
  1. Data fusion
  2. Machine learning
  3. Improved feature identification (through both statistical and computer-aided methods)

We heard from an excellent set of speakers that covered a wide range of analysis techniques aimed to investigate and understand the complex and highly coupled geospace system. Ryan McGranaghan opened the session with a scene-setting presentation. McGranaghan Defining the Next Generation of Geospace Research (pdf)

First, Seebany Datta-Barua (sdattaba@iit.edu) spoke about coherence in ionosphere-thermosphere flows and introduced Lagrangian Coherent Structures as a promising method to characterize the coherence. She also discussed broader themes pertaining to innovation in geospace research, specifically highlighting the importance of uncertainty quantification and placing emphasis on producing actionable forecasts. Datta-Barua Innovative Methods for Ionosphere-Thermosphere Discovery (pdf)

Chunming Wang (cwang@dornsife.usc.edu) next presented results from the Space Weather Forecasting Testbed (SWFT), a machine learning tool created through the ‘Medium Range Ionosphere-Thermosphere Forecasts’ Living With a Star (LWS) project. Chunming stressed that effective forecasting systems must be a combination of model and data-driven approaches, and demonstrated the potential of machine learning techniques to improve space weather forecasting with existing models and data. Wang Machine-Learning Approach for Medium Range Forecast of Ionosphere Anomalies (pdf)

Josh Semeter (jls@bu.edu) gave the final invited talk of the session. He discussed that improving technologies (processing power of an iPhone, for instance) has made heliophysics a data-intensive field. The Mahali Space Weather Monitoring Project was highlighted as a prototype to a revolutionary new architecture that will harness mobile devices to form a global space weather monitoring network. He concluded by echoing a theme that developed from Chunming’s presentation, and continued throughout the session, that scalable machine assistance provides great potential to help humans in the scientific discovery process. Semeter Computer-Aided Discovery in Geospace (pdf)

Roger Varney (roger.varney@sri.com) spoke about the future of geospace observational facilities, raising two important questions: 1) How do we change the way we use our existing facilities to better meet community needs? and 2) How should we plan future facilities? He detailed the requirements for an operational system and highlighted the success of the Poker Flat Incoherent Scatter Radar in meeting these requirements. He concluded by outlining the near, medium, and long range directions for geospace observational systems to bring about innovation in CEDAR science. Varney Incoherent Scatter Radars for System Science and Operational Applications (pdf)

Sierra Flynn (sierra.flynn@colorado.edu) presented results from an Empirical Orthogonal Function (EOF) analysis of nitric oxide flux, and extrapolated the broader implications of such advanced statistical approaches. These broader implications included better understanding complex data sets in compact and comprehensible ways and gaining new scientific insight through data mining and advanced analytics. Flynn Empirical Orthogonal Function (EOF) Analysis and Thermospheric Nitric Oxide Flux (pdf)

Romina Nikoukar (romina.nikoukar@jhuapl.edu) presented the results of a novel data assimilation method for the plasmasphere, which was also given as a science highlight on the final day of the workshop. Romina demonstrated that assimilation, statistically fusing observations with models, is capable of drastically improving geospace specification, extending this paradigm from the ionosphere-thermosphere (where assimilation is generally used in our field) into the inner magnetosphere. Nikoukar A Novel Data Assimilation Model for the Plasmasphere (pdf)

Finally, Asti Bhatt (asti.bhatt@sri.com) spoke about how next generation system science can be approached through networks of homogeneous instruments providing consistent data products using NSF CEDAR program funded Mid-latitude All-sky-imager Network for Geospace Observations (MANGO) as an example. She also provided a software-centric perspective for innovation in geospace by detailing the Integrated Geoscience Observatory (InGeO). InGeO, funded by the National Science Foundation EarthCube program, is a new community resource based on JupyterHub that allows users to discover and collect data from different ionospheric data sources. Networks of homogeneous instruments and Integrated Geoscience Observatory (pdf)

View PDF