

Data Mining for Science of the Sun-Earth Connection as a Single System

Primary Convenor: Nathalia Alzate (NASA GSFC/ADNET)

Co-Convenors: Simone Di Matteo (NASA GSFC/CUA)

Craig DeForest (SwRI)

Matt West (SwRI)

Key Questions

- a) What is needed to overcome the **challenges of past/current missions** when trying to establish a Sun-Earth connection?
- b) How can current data, tools and models be exploited to create the synergy required for a **system treatment of the heliosphere?**

Limitations and Challenges

Data and Computational Limitations:

- **Datasets and catalogs** – different formats, different analysis tools required
- **Different datasets** – different region coverage
- **Observations limited** by spacecraft and ground-based instrument configuration
- **Numerous analysis approaches** not fully optimized
- **Various programming languages**

Challenges of Past/Current Missions:

- We lack an **uninterrupted field of view** of the solar atmosphere
- Change of **plasma regimes**
- **Comparison with in-situ** observations
- **Windows of opportunities** to effectively exploit current remote sensing observations and in-situ measurements limited by spacecraft configurations
- **Lack of information from the poles**

Discussion Highlights

The discussion naturally evolved into three key points:

1. Data mining, **don't forget previous datasets**, they are available and can still provide a lot of science!
 2. **Proper documentation** for instrument caveats, but users should read them and contact instrument team to avoid misinterpretations
 3. **Best practice** for tools developments
- Executable paper or Jupiter notebook with data handling examples
 - We should think about the longevity of software (e.g. some data analysis tool written in IDL are not compatible with newer IDL version; SSW has lasted ≈30 years)
 - Standardization of data format will increase the compatibility of science tools
 - Unique data repository mentioned but difficult for practical purpose (management and maintenance)

Discussion Highlights

The discussion naturally evolved into three key points:

1. Data mining, **don't forget previous datasets**, they are available and can still provide a lot of science!
 2. **Proper documentation** for instrument caveats, but users should read them and contact instrument people to avoid misinterpretations
 3. **Best practice** for tools developments
- **PLEASE READ THE INSTRUMENT DOCUMENTATION** or contact instruments POC
 - Would be helpful to have special issue review or updates on lesson-learned/caveats about mission/instruments after a certain amount of time
 - For proper documentation teach new generation of scientists the best practice and provide workshop for mid-career to catch up (e.g., Center for HelioAnalytics, US-RSE)
 - Time spent on pretty codes don't pay off... recognize the effort of developers (e.g., include name in papers)
 - Funding opportunities tailored for data preparation and tool development (NASA, NSF)

Moving Forward

- How can we work together to innovate and share ideas as a community?
- What should our plan forward be? (Living Review, ISSI Team, Workshops, etc.)
 - **White Paper**
 - **SHIPWork (previously SIPWork)**

Key Message

Stay informed and share information!

Send Comments & Suggestions!

Naty Alzate:

nathalia.alzate@nasa.gov