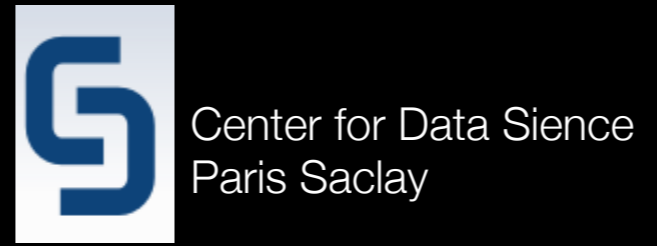


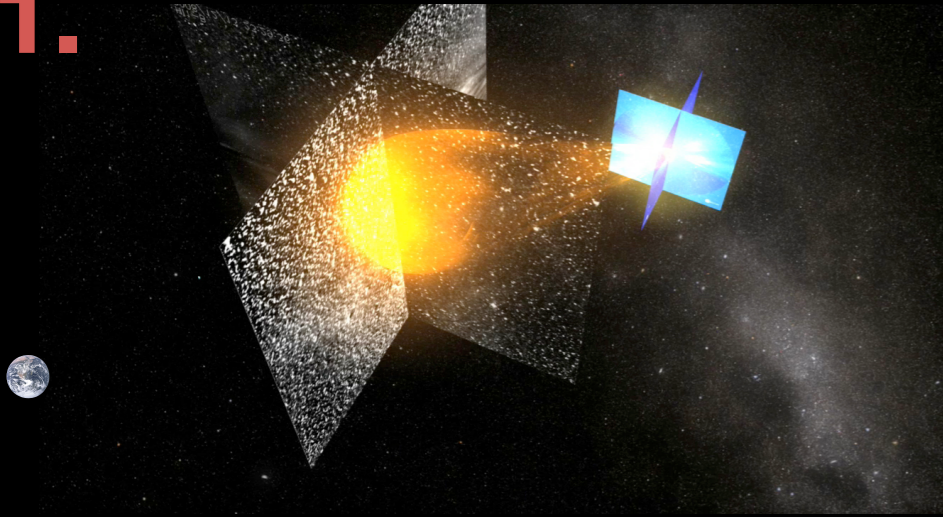
SciQLoP

Scientific Qt application for **Learning from Space Data** from observations of Plasmas



IN THIS PRESENTATION

1.



2.

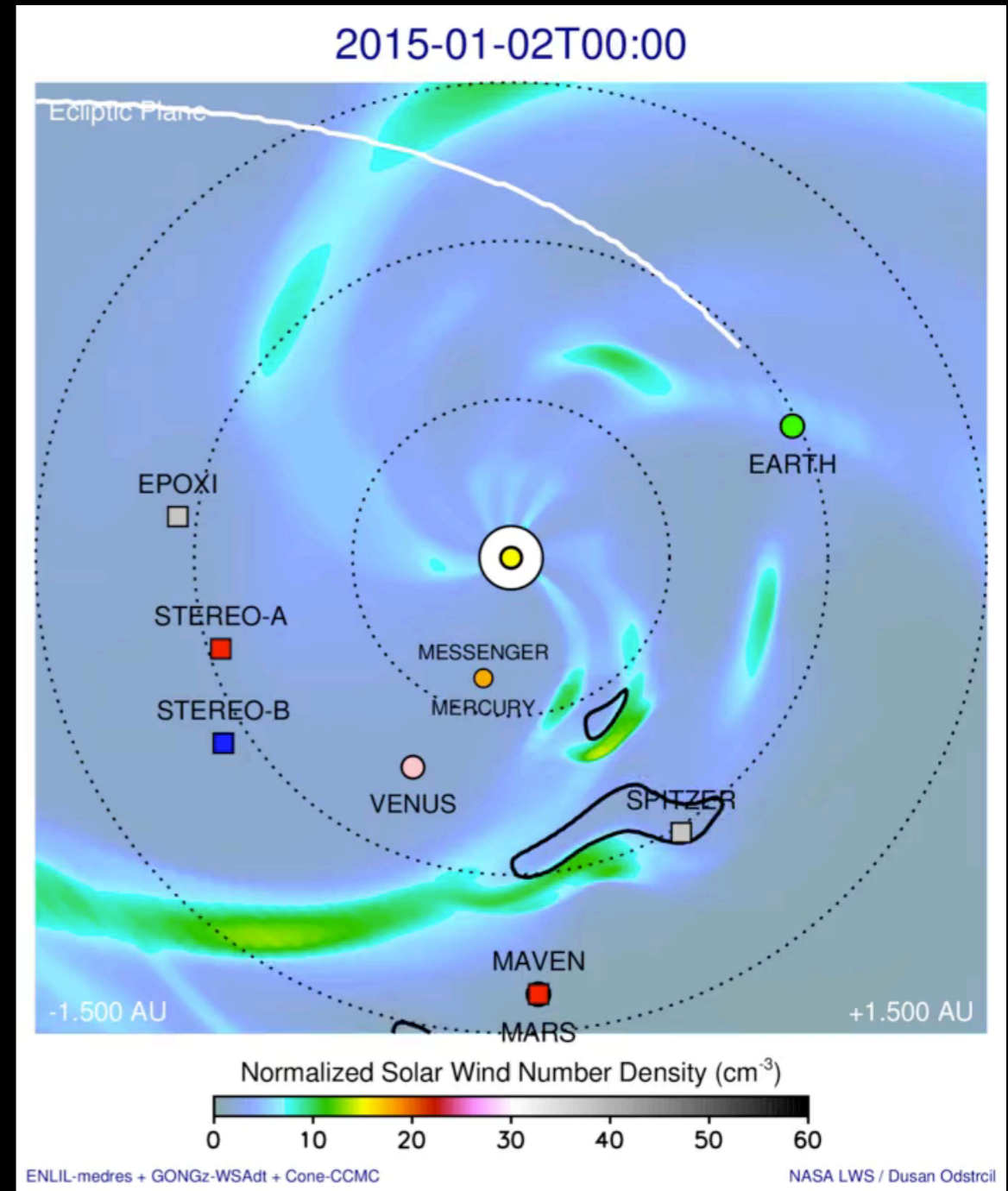
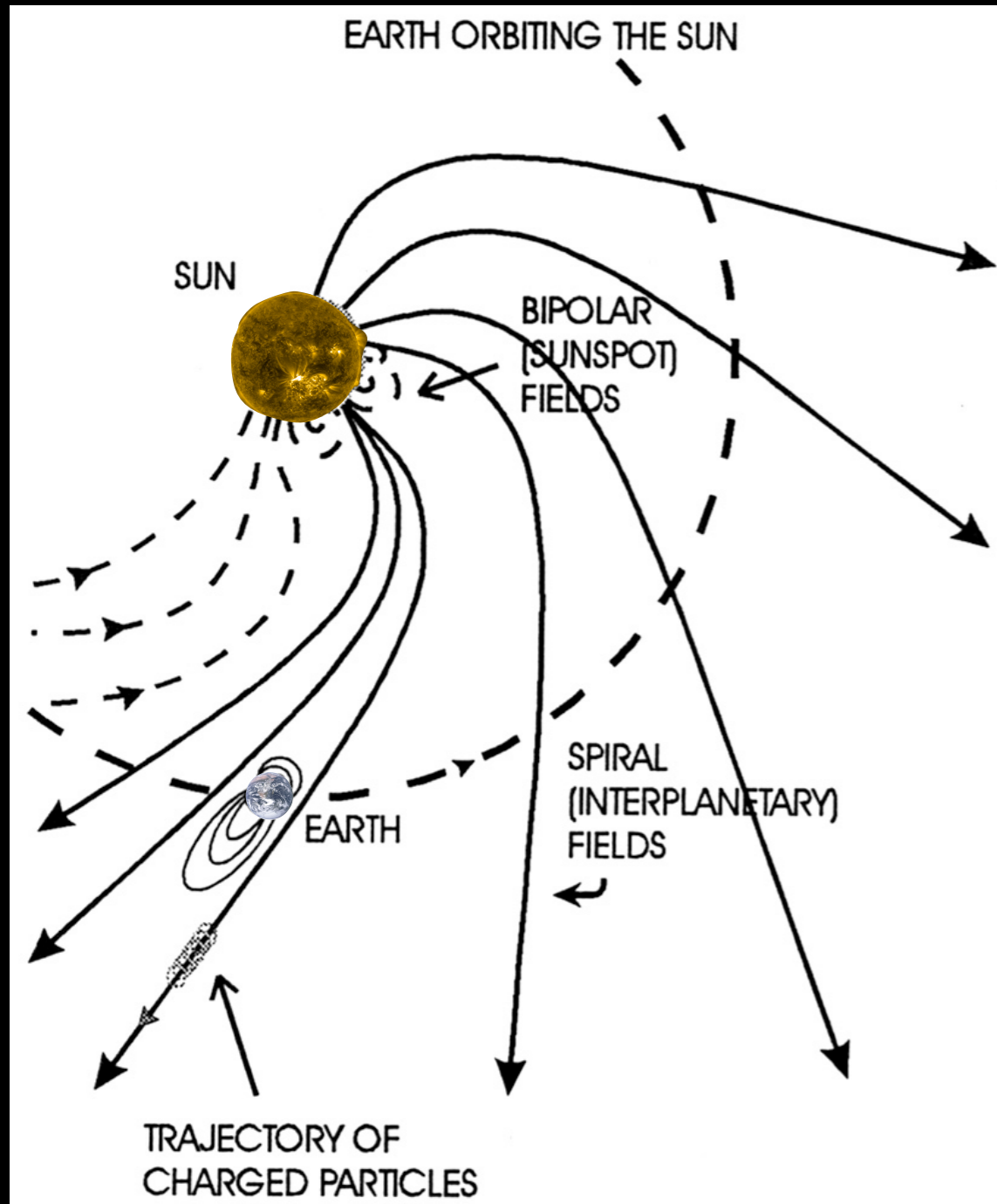


3.

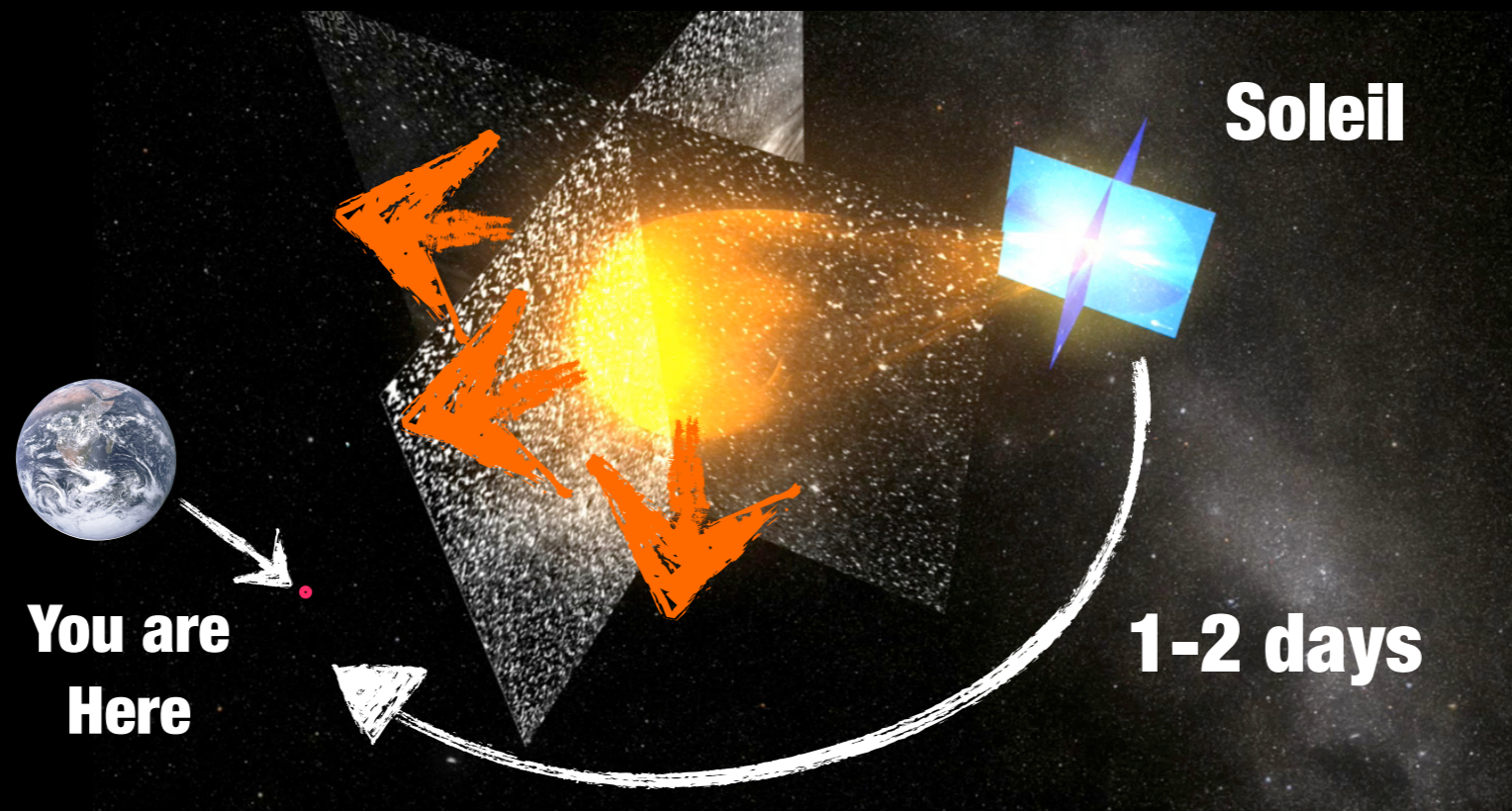
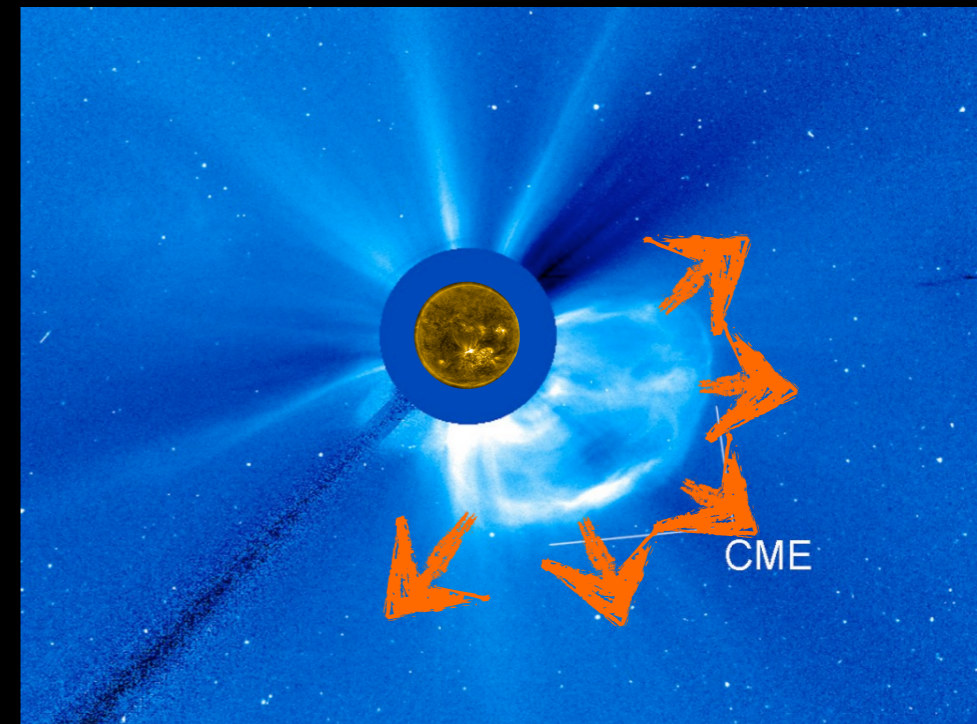
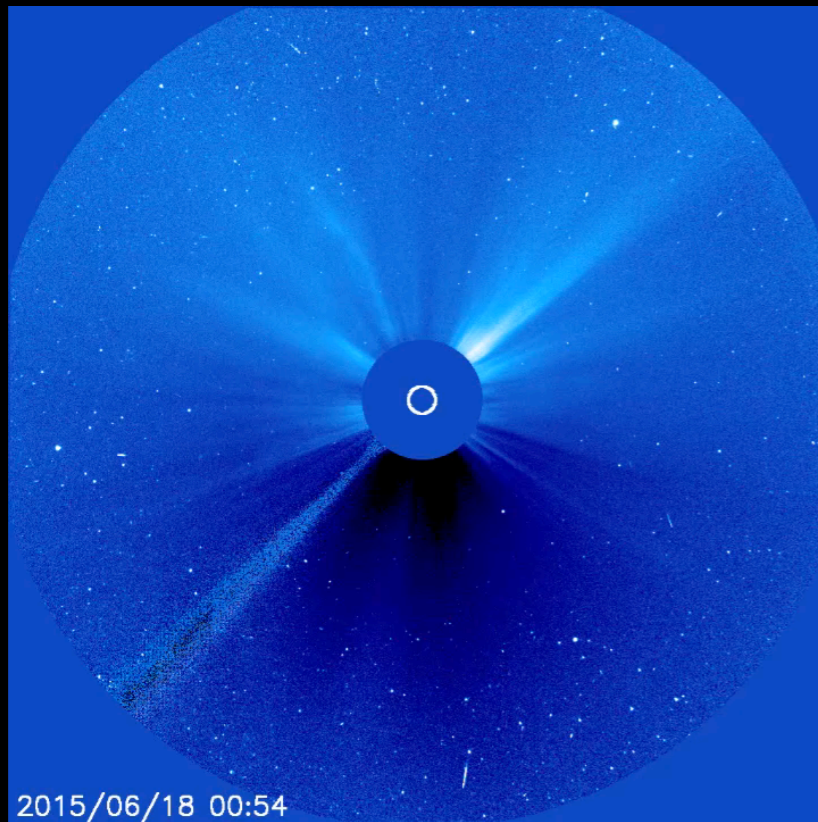
université
PARIS-SACLAY

Learning
from
Space
Data

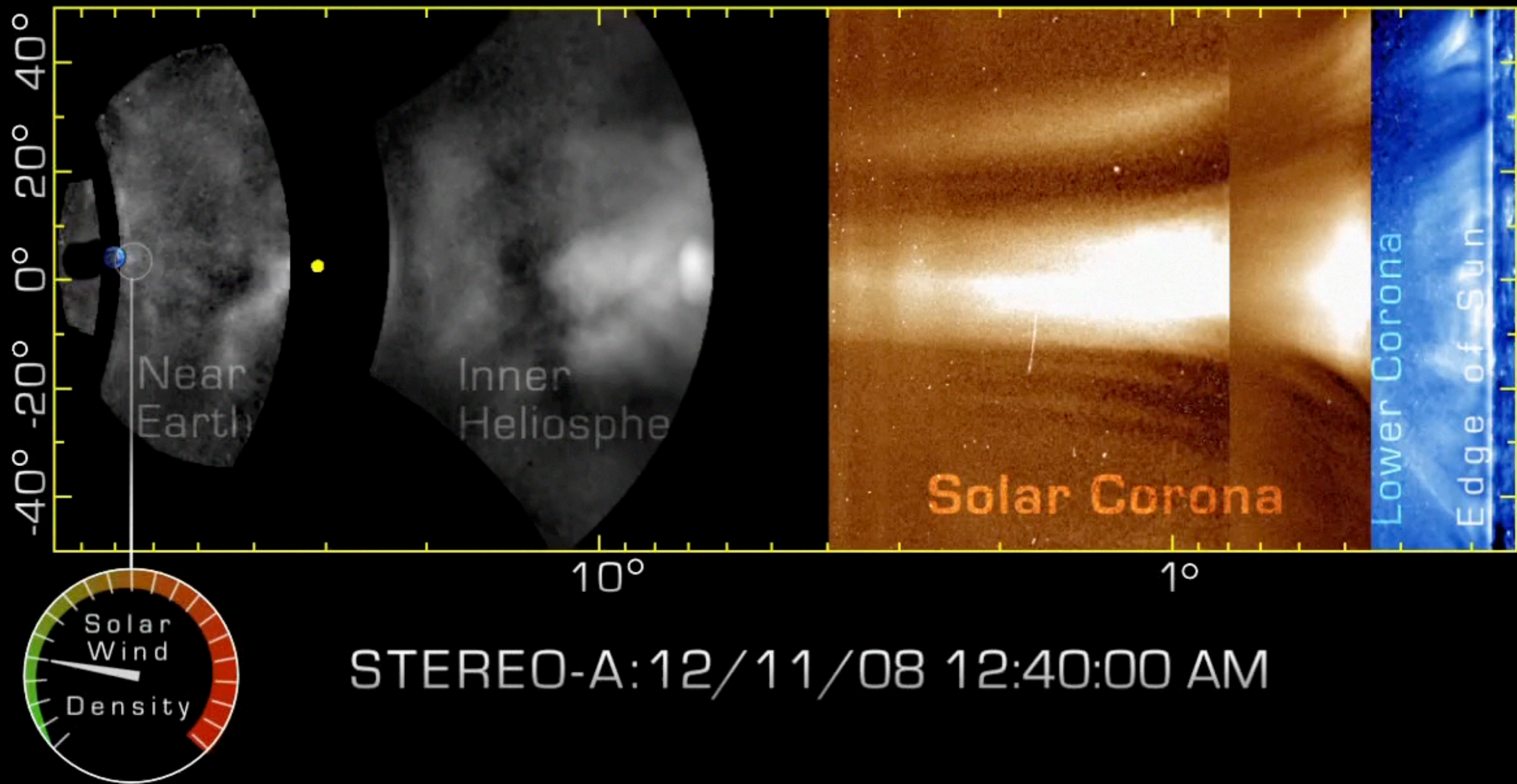
SUN-EARTH SYSTEM



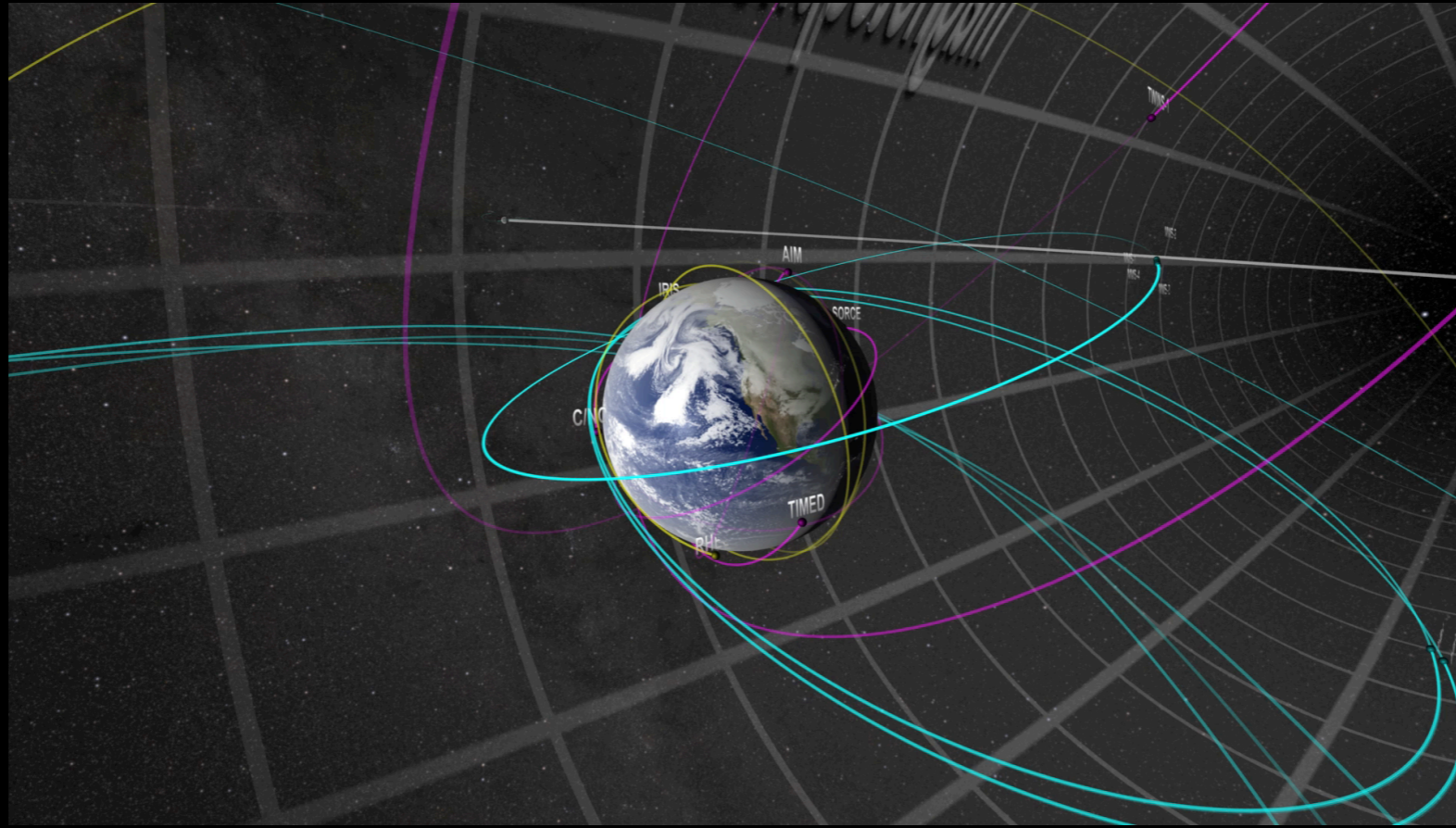
CORONAL MASS EJECTIONS



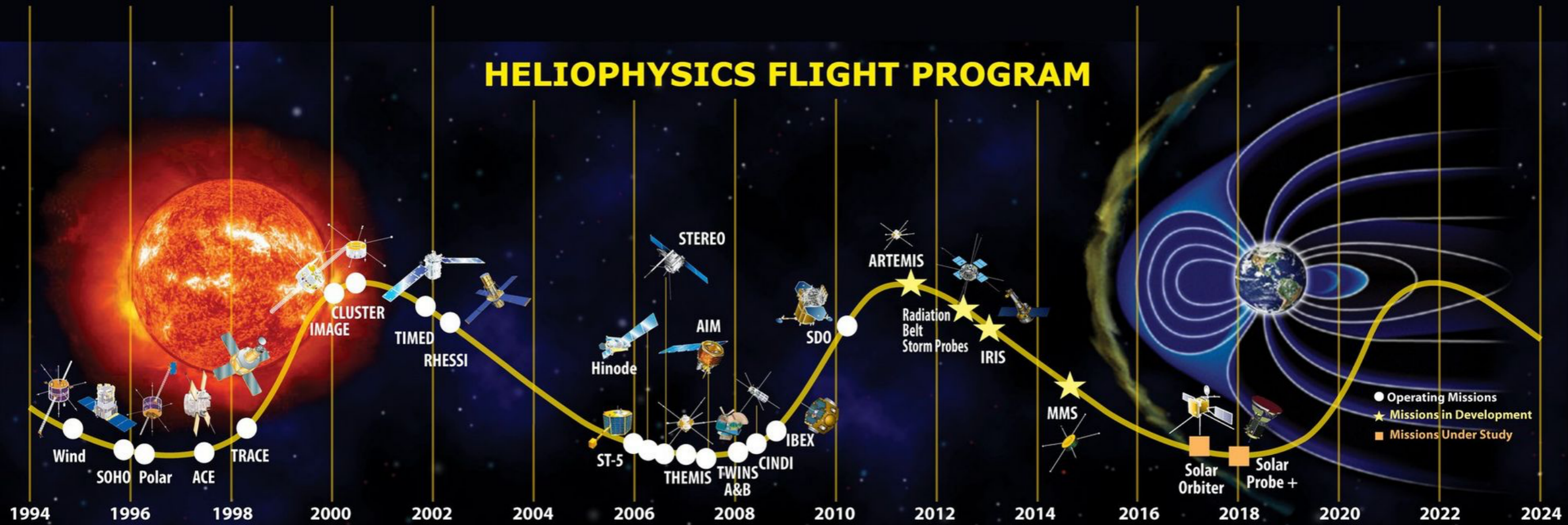
SOLAR WIND

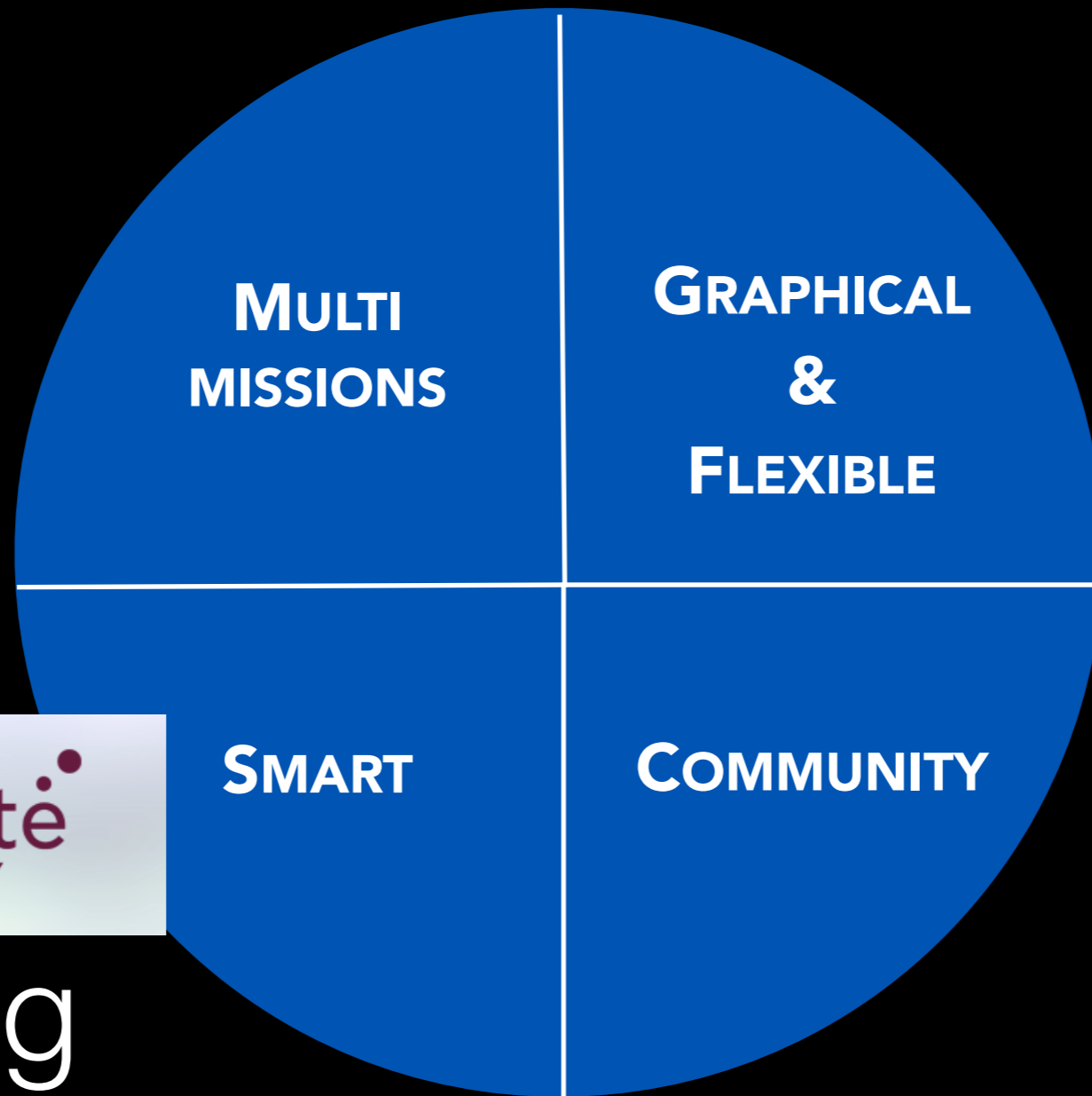


LOTS OF MISSIONS



HELIOPHYSICS FLIGHT PROGRAM





Learning
from
Space
Data

Automatic event
detection with Machine
Learning

TEAM



Nico A.

*« typical user »
interface with observers @ LPP*

Alexis Jeandet

*Main code designer
expert C++/GUI*

HUGO WINTER -

CDD - 12 MOIS

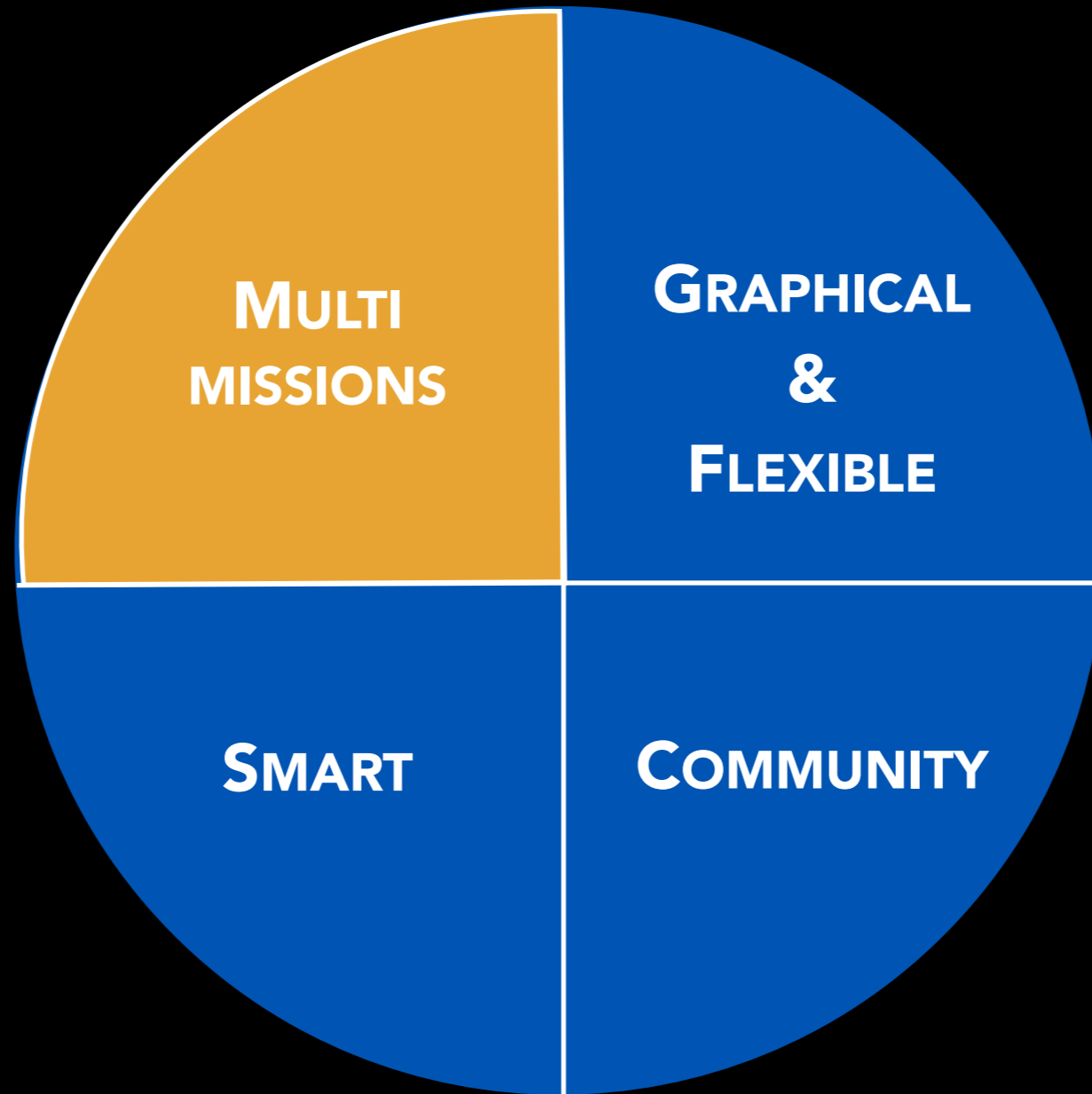
*Main developer
GUI Qt
Signal*

Erwan Le Pennec

*Expert/consulting
Machine Learning*

Rodrigue Piberne

Space Data products / scientific visualization



WHY MULTI-MISSIONS ?

SAME DATA

PLASMA

DENSITY, TEMPERATURE

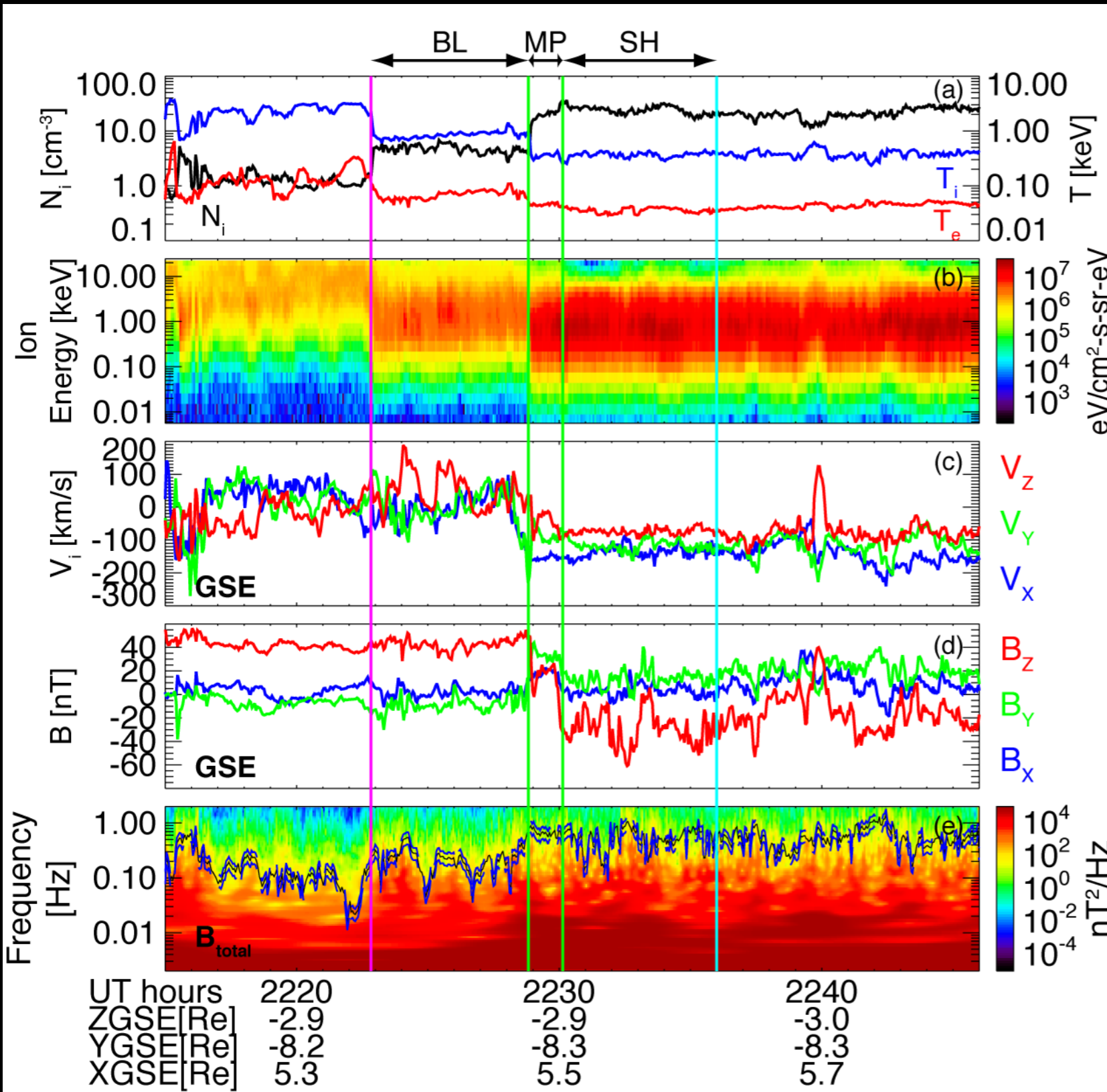
FLOWS

PRESSURES

DISTRIBUTIONS

ELECTROMAG.

E, B, POTENTIAL...



WHY MULTI-MISSIONS ?

Multiple URL Shortener
GODDARD SPACE FLIGHT CENTER
Space Physics Data Facility
+ Goddard Home
+ Visit NASA.gov

+ SPDF HOME + DATA & ORBITS + MODELS at CCMC + SCIENCE ENABLED + AND MORE

CDAWeb
+ SPDF HOME
+ FEEDBACK
+ ABOUT CDAWEB

CDAWeb Mirror Site
+ RAL/UK

Guides and Tutorials
+ CDAWeb help
+ Internet browser help

Additional Services
+ CDAWeb Inside IDL
+ HTTP and Anonymous FTP access to public CDAWeb database
+ Overview of Alternative Data Access Methods
+ Autoplot.org (non-NASA) interface to public CDAWeb database

CDAWeb Data Views

- Public data from current (1992 -> present) space physics missions (including ACE, Cluster, C/NOFS, FAST, Geotail, GOES 5-12, IMAGE, LANL 1989-2002, NOAA 10-14, OMNI, Polar, STEREO, THEMIS, TIMED, Ulysses, Van Allen Probes, Voyager, Wind and others).
- Public data from older missions (including Alouette, CRRES, DE, Hawkeye, IMP-8, ISIS, NOAA 5-10, OMNI and others).
- Public data from all current and past space physics missions

CDAWeb Data Views

- Public data from current (1992 -> present) space physics missions (including ACE, Cluster, C/NOFS, FAST, Geotail, GOES 5-12, IMAGE, LANL 1989-2002, NOAA 10-14, OMNI, Polar, STEREO, THEMIS, TIMED, Ulysses, Van Allen Probes, Voyager, Wind and others).
- Public data from older missions (including Alouette, CRRES, DE, Hawkeye, IMP-8, ISIS, NOAA 5-10, OMNI and others).
- Public data from all current and past space physics missions

Centre de Données de la Physique des Plasmas
Plasma Physics Data Centre

CDPP
Home
About the CDPP
Publications
Presentations
Rules of the road
DATA
Overview
CNES database
AMDA
Mirror Themis database
Event List
SERVICES

Home

Welcome to CDPP

The CDPP (Centre de Données de la Physique des Plasmas) was created in 1998 jointly by CNES and INSU. The CDPP is the French national data centre for natural plasmas of the solar system. The CDPP assures the long term preservation of data obtained primarily from instruments built using French resources, and renders them readily accessible and exploitable by the international community. The CDPP also provides services to enable on-line data analysis (AMDA), 3D data visualization in context (3DView), and a propagation tool which bridges solar perturbations to in-situ measurements. The CDPP is involved in the development of interoperability, participates in several Virtual Observatory projects, and supports data distribution for scientific missions (Solar Orbiter, JUICE).

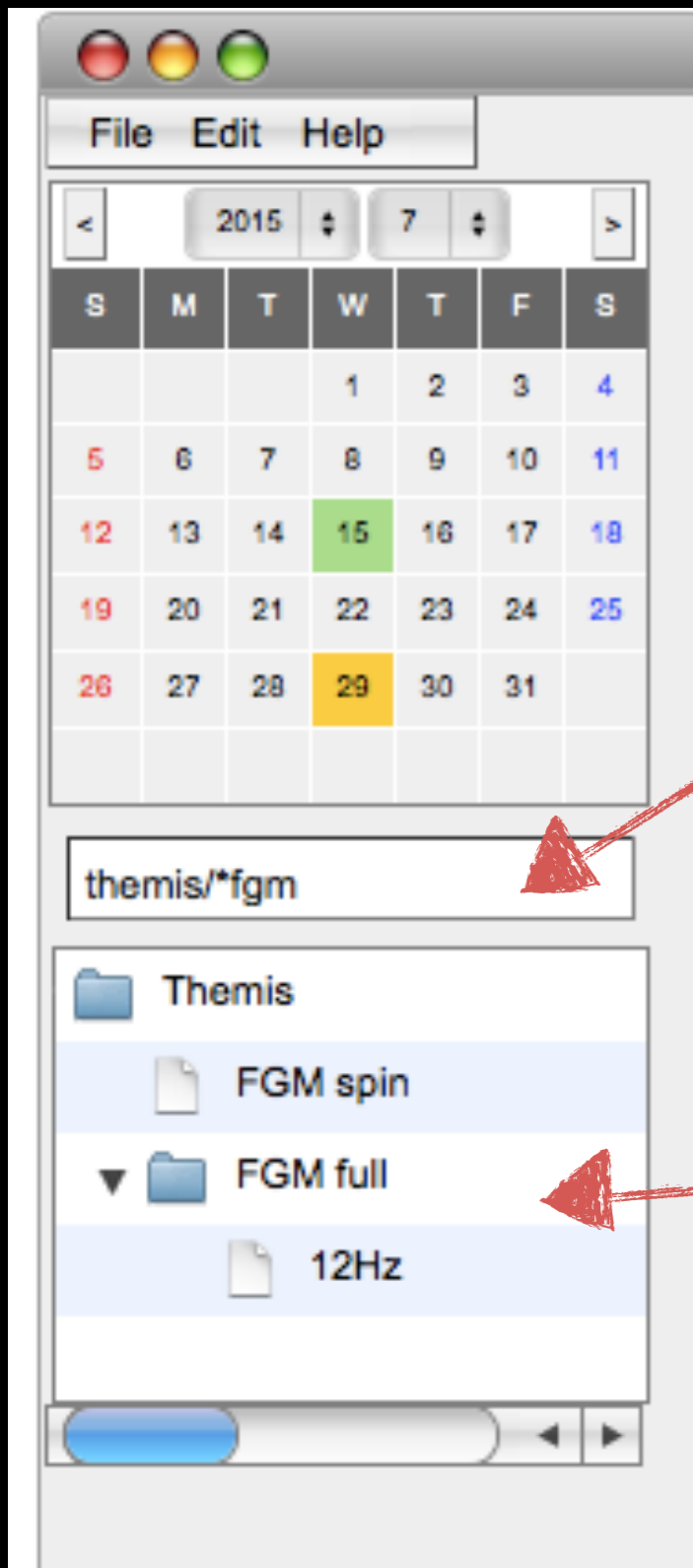


SAME DATA FORMAT : CDF

e.g. Mission ESA/Cluster, 130TB since 2001
mission NASA/MMS, launched 2015 > 10TB/year



- **Multi-mission, intuitive GUI**
- **EASILY BROWSE DATA PRODUCTS**
- **INTEROPERABILITY WITH CDP, NASA, ETC.**
- **LOAD ASCII/CDF FILES...**
- **SIMPLY DRAG PRODUCTS TO PLOT AREA**



Search data

any text will be searched in product meta-data

All known products

Dynamically filtered data products

local, distant...



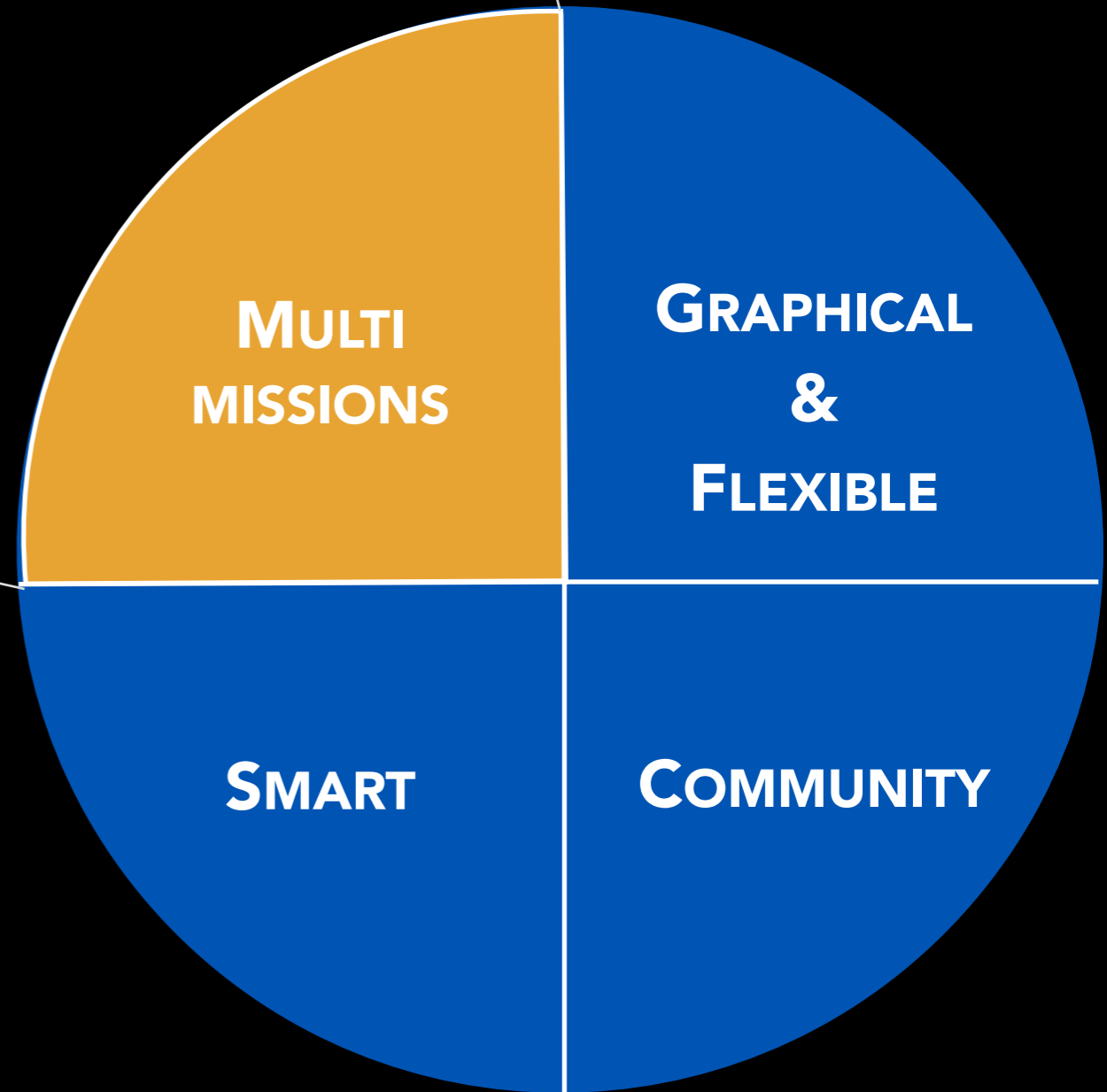
• **READ CDF, ASCII...**

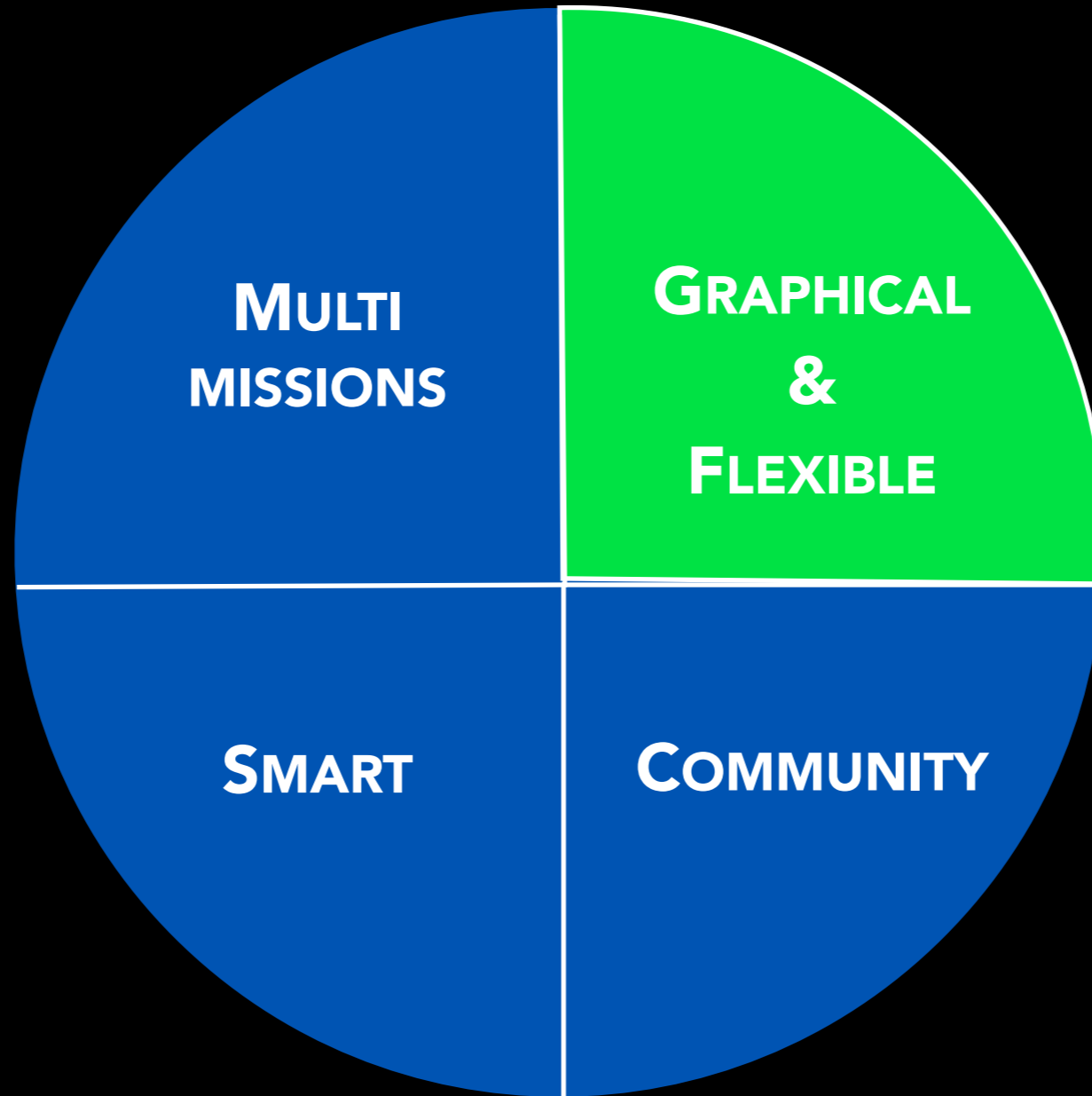
• **MISSIONS PLUGINS**



• **AMDA/NASA INTEROPERABILITY**

Just get data



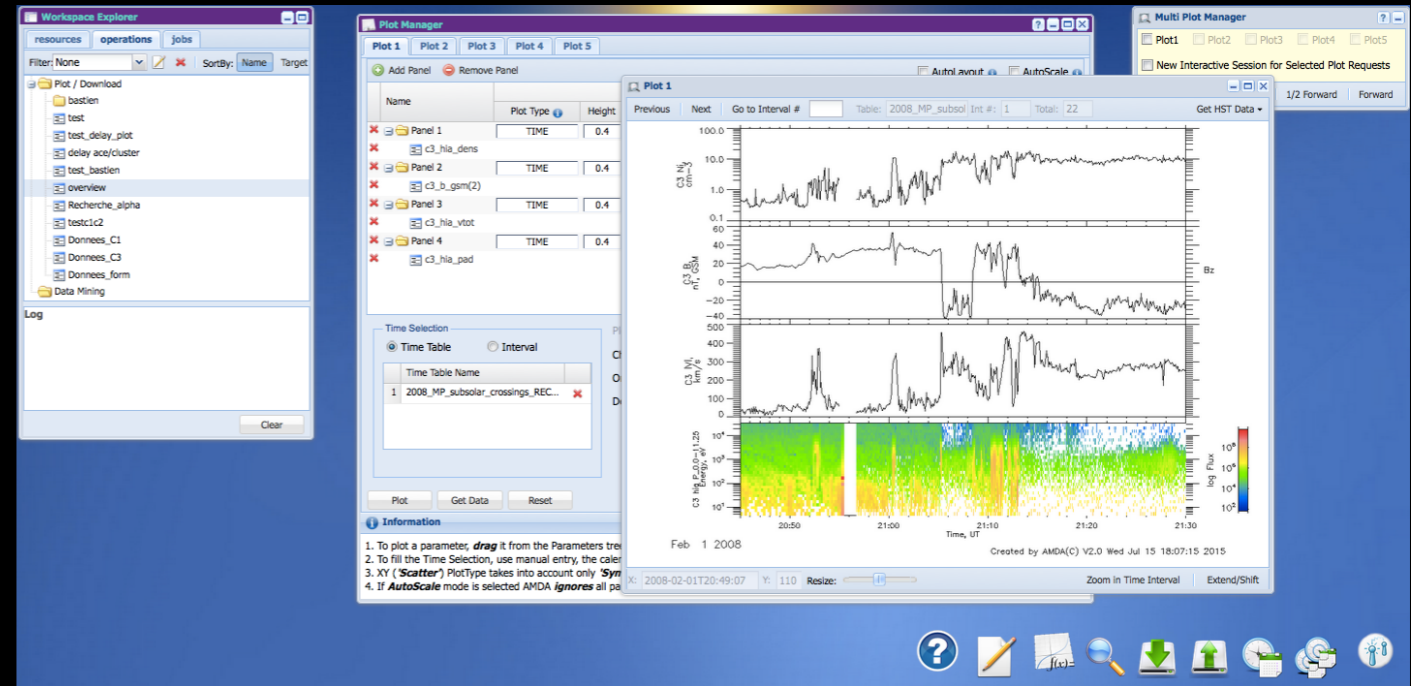


EXISTING TOOLS?

SCRIPTING



GUIs



NO:

VERY BAD FOR JUST DATA BROWSING
REINVENTING ALL WHEELS
LOTS OF CRAPPY CODE IN NATURE...

NO:

NOT VERY FLEXIBLE
TENDENCY FOR « USINES À GAZ »

YES:

BRINGS STRONG FLEXIBILITY
REQUIRED BY RESEARCH
SHARING CODE
BATCH ANALYSIS

YES:

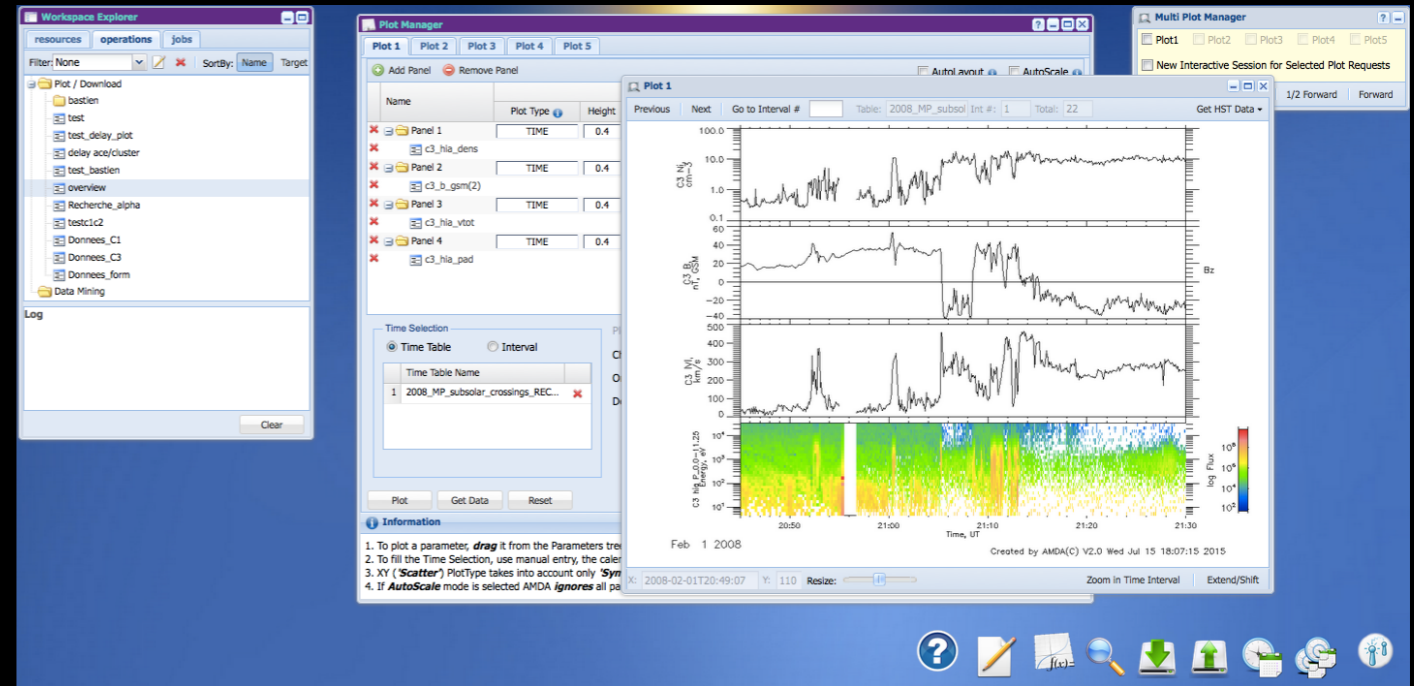
EASY DATA BROWSING
EASY ROUTINE TREATMENTS
EASY FOR STUDENTS

EXISTING TOOLS?

SCRIPTING



GUIs



NO:

VERY BAD FOR JUST DATA BROWSING
REINVENTING ALL WHEELS
LOTS OF CRAPPY CODE IN NATURE...

NO:

NOT VERY FLEXIBLE
TENDENCY FOR « USINES À GAZ »

YES:

BRINGS STRONG FLEXIBILITY
REQUIRED BY RESEARCH
SHARING CODE
BATCH ANALYSIS



YES:

EASY DATA BROWSING
EASY ROUTINE TREATMENTS
EASY FOR STUDENTS

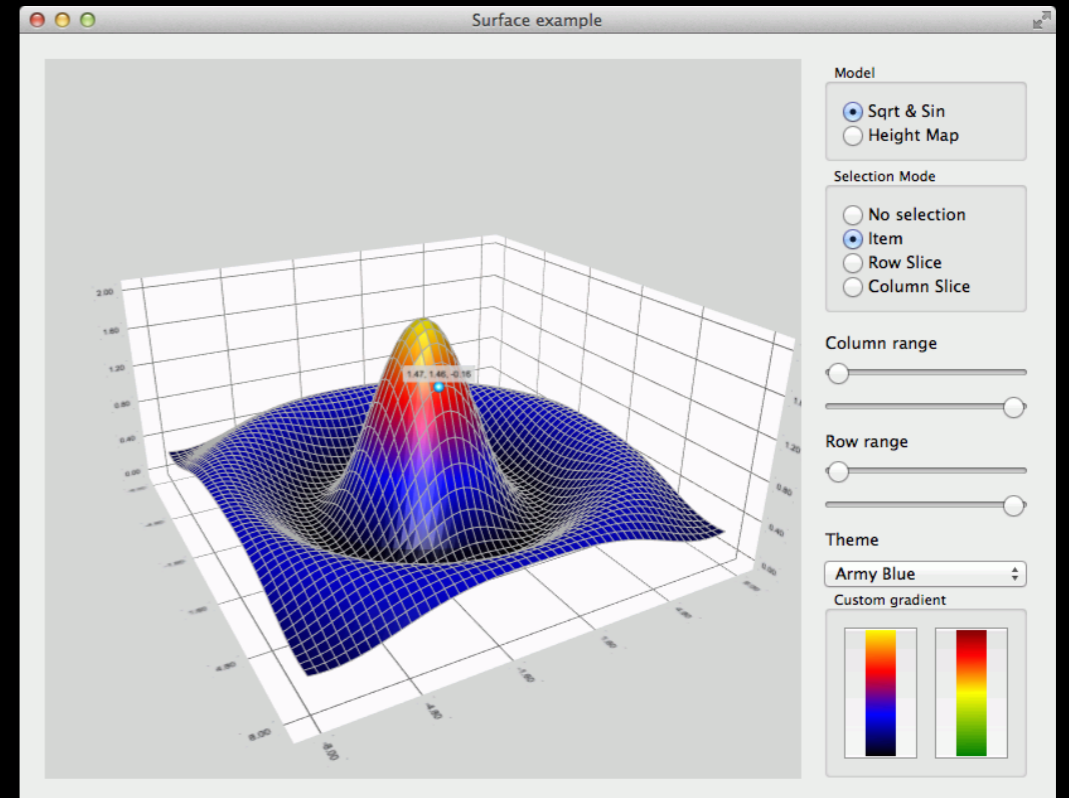


• Technology choices



C++ QT

SIMPLE CODE,
PORTABLE,
HUGE COMMUNITY
OPEN SOURCE



PERFORMANCE, GOOD
COMMUNITY



• Multi-mission, intuitive GUI

Interactive high perf panels

Scroll and transparently download data

easy browsing of data products based on keywords

The screenshot displays the SciQLOP GUI with several key components:

- Calendar:** A calendar for the year 2015, with the 7th of the month highlighted.
- File Browser:** A search bar containing "themis/*fgm" and a list of files including "Themis", "FGM spin", "FGM full", and "12Hz".
- Main Plot Area:** A multi-panel plot showing four time-series data series: "C3 M, cm⁻³", "C3 B, nT, GCM", "C3 M, km/s", and "C3 Hg P, 0.0-11.26 Energy, eV". A red arrow labeled "real time update" points to the plot area.
- Distribution Plot:** A smaller window titled "distribution" showing a contour plot of "Vperp (km/s)" vs "Vpara (km/s)" for the time interval "2002-05-04/14:14:51->14:15:03". A red arrow labeled "specific toolboxes" points to a list of options: XZ, XY, YZ, V1V2, VparaV1, and VparaV2.
- Command Line:** A text area at the bottom left containing the command "In [1]: Import spacepy as sp".

specific toolboxes



- **Embedded iPython : power of custom toolkits (homemade or not)**

- Easy access to user libraries
- terminal \longleftrightarrow plots
- enable very specific data manipulation (not GUI)

The screenshot displays the SciQLOP application interface. On the left, there is a file browser showing a directory structure for 'themis/*fgm' with subfolders 'FGM spin', 'FGM full', and '12Hz'. Above the file browser is a calendar for the year 2015, with the date 29 highlighted. Below the calendar is a terminal window containing the Python code: `In [1]: Import spacepy as sp`. The main area of the application is occupied by several data plots. The top plot shows a time-series of $C3 N_y$ in cm^{-3} on a logarithmic scale. The second plot shows $C3 B_{nt, CSM}$ in nmT . The third plot shows $C3 V_L$ in km/e . The bottom plot is a heatmap showing $C3 N_{E_{p,0.0-1.26}}$ in $Energy, eV$. A red arrow points from the terminal window to the top plot. In the bottom right corner, there is a 'distribution' window showing a contour plot of V_{perp} (km/s) for the time interval 2002-05-04/14:14:51 to 14:15:03. The distribution plot has axes ranging from -1000 to 1000 km/s. To the right of the distribution plot is a control panel with radio buttons for selecting different variables: XZ (selected), XY, YZ, V1V2, VparaV1, and VparaV2.



- **ELEGANT AND ERGONOMIC**

- **PERFORMANCE AND REAL TIME PLOTTING**

- **INTERACT WITH DATA AND PLOTS**

- **POWER AND FLEXIBILITY OF PYTHON SCRIPTS**

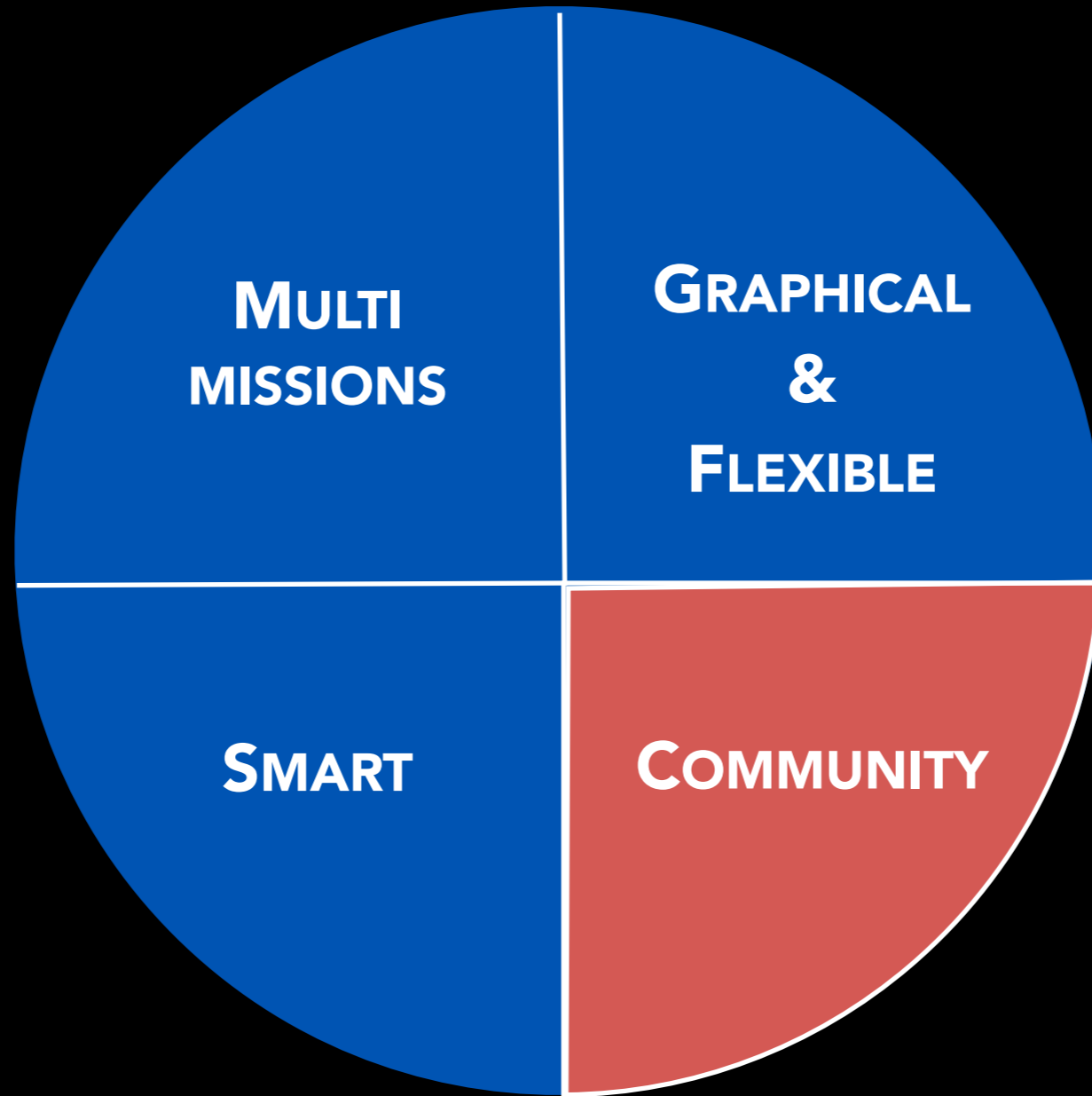
**MULTI
MISSIONS**

**GRAPHICAL
&
FLEXIBLE**

SMART

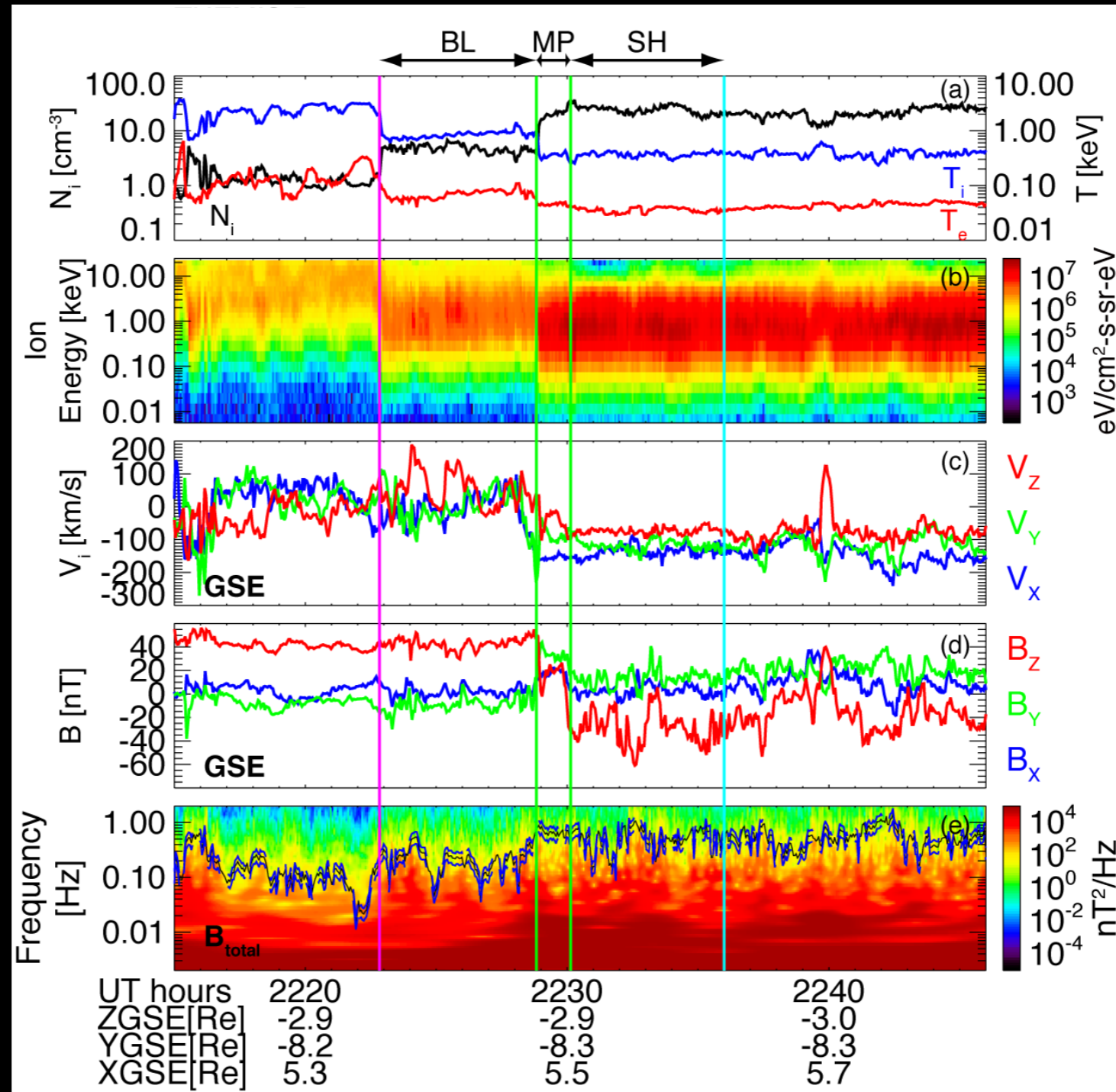
COMMUNITY

*Visualize complex data
interact with data*



« EVENT »

Time interval where measures show signatures associated with a physical phenomenon of interest. Usually group them to do statistical studies

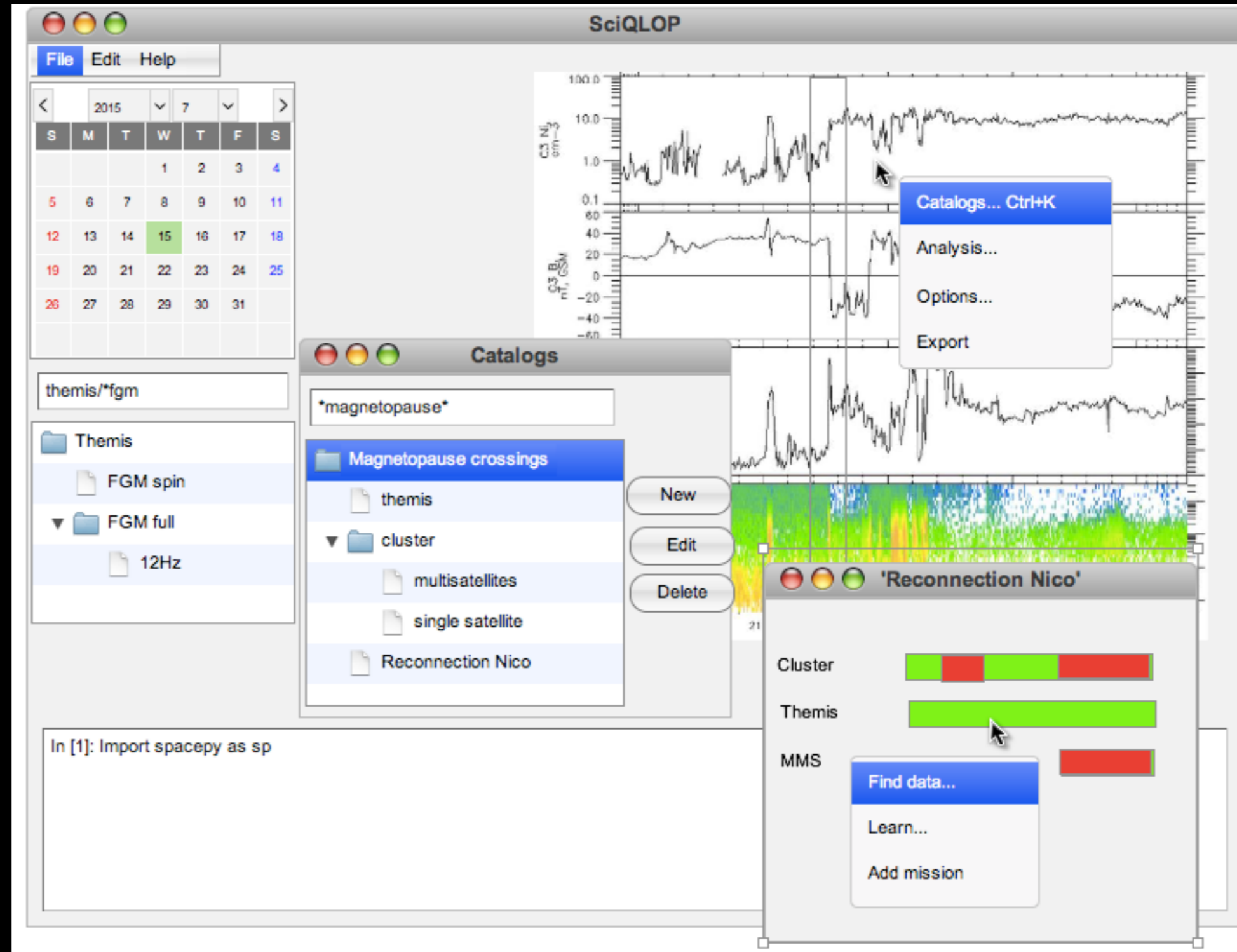




• Catalogs of data

Gather data for statistics

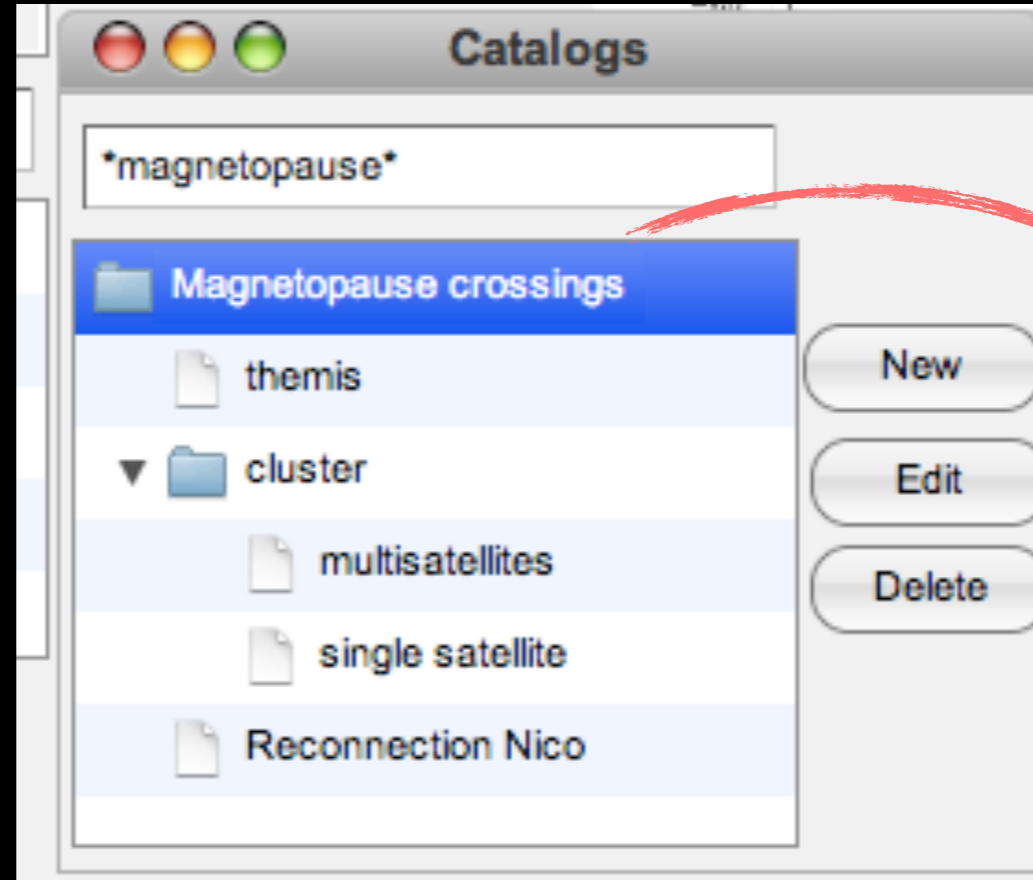
- Catalog = group of data intervals
- Data can belong to multiple catalogs
- « add to catalog » directly from plot panels
- clone/extend features



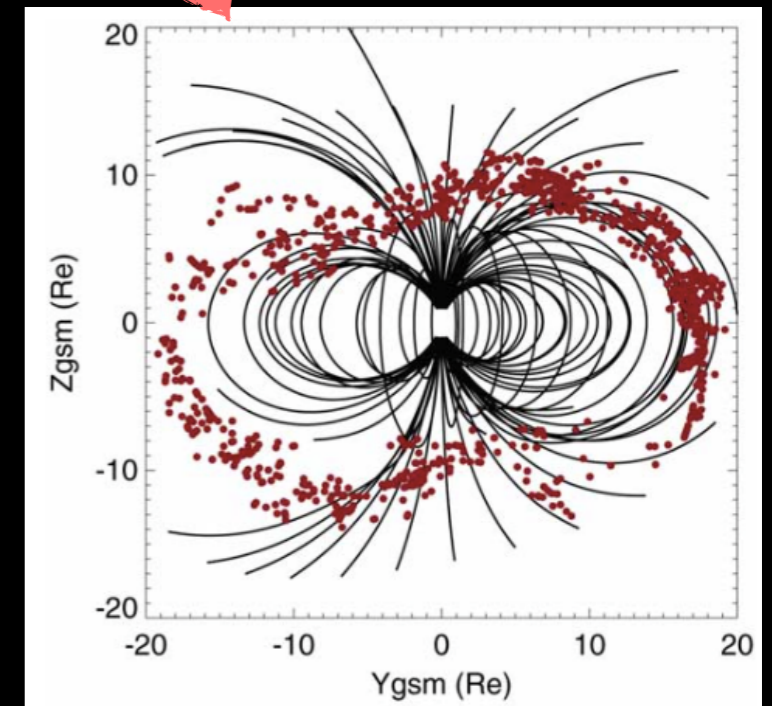


• **Visualizing catalogs** *Extract and visualize metadata*

- Rich *automatic* metadata (user, spacecraft, data products etc.) not just start/stop date and optional description
- Easily extract and visualize information from your catalog



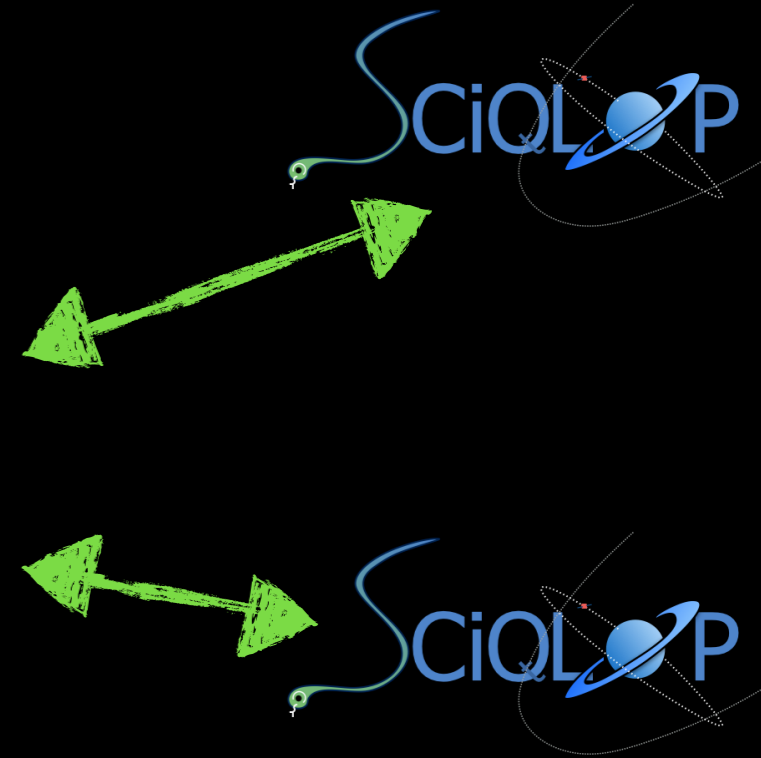
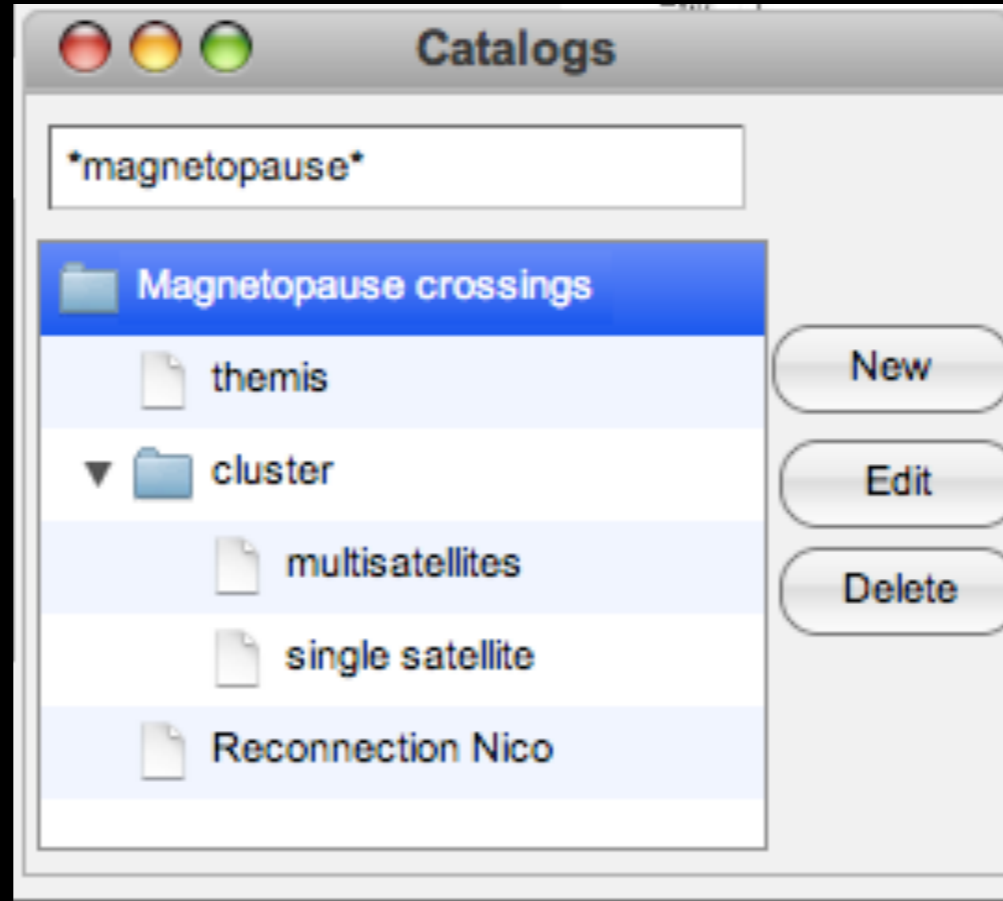
ex : where are all my intervals located?





- **Online community based catalog** (> *SciQLOP v.1*)
Improve reproducibility - ANTI-reinventing-the-wheel-tool

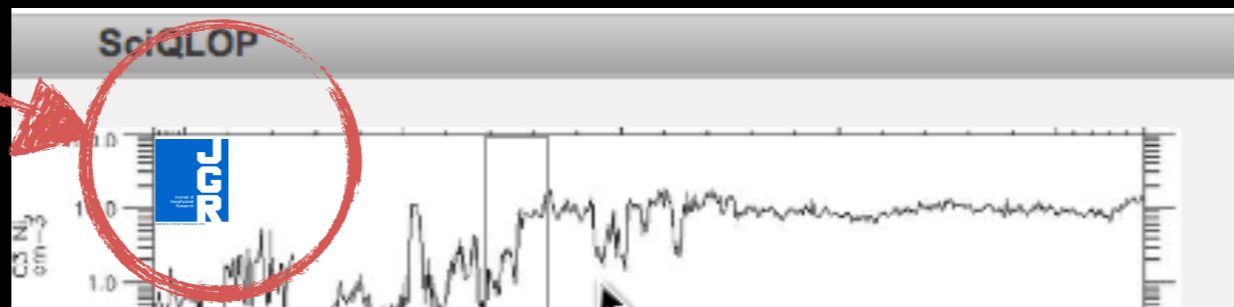
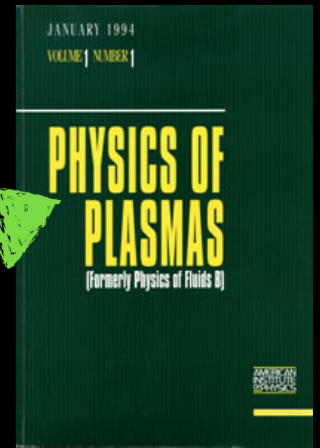
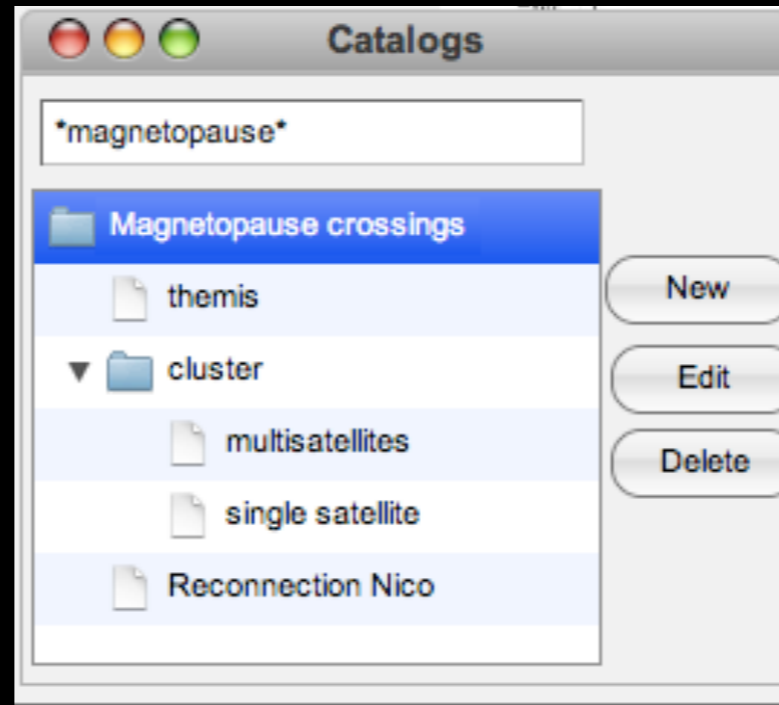
- Public and group catalogs
- Online sharing between all SciQLOP instances
- Build catalogs with colleagues





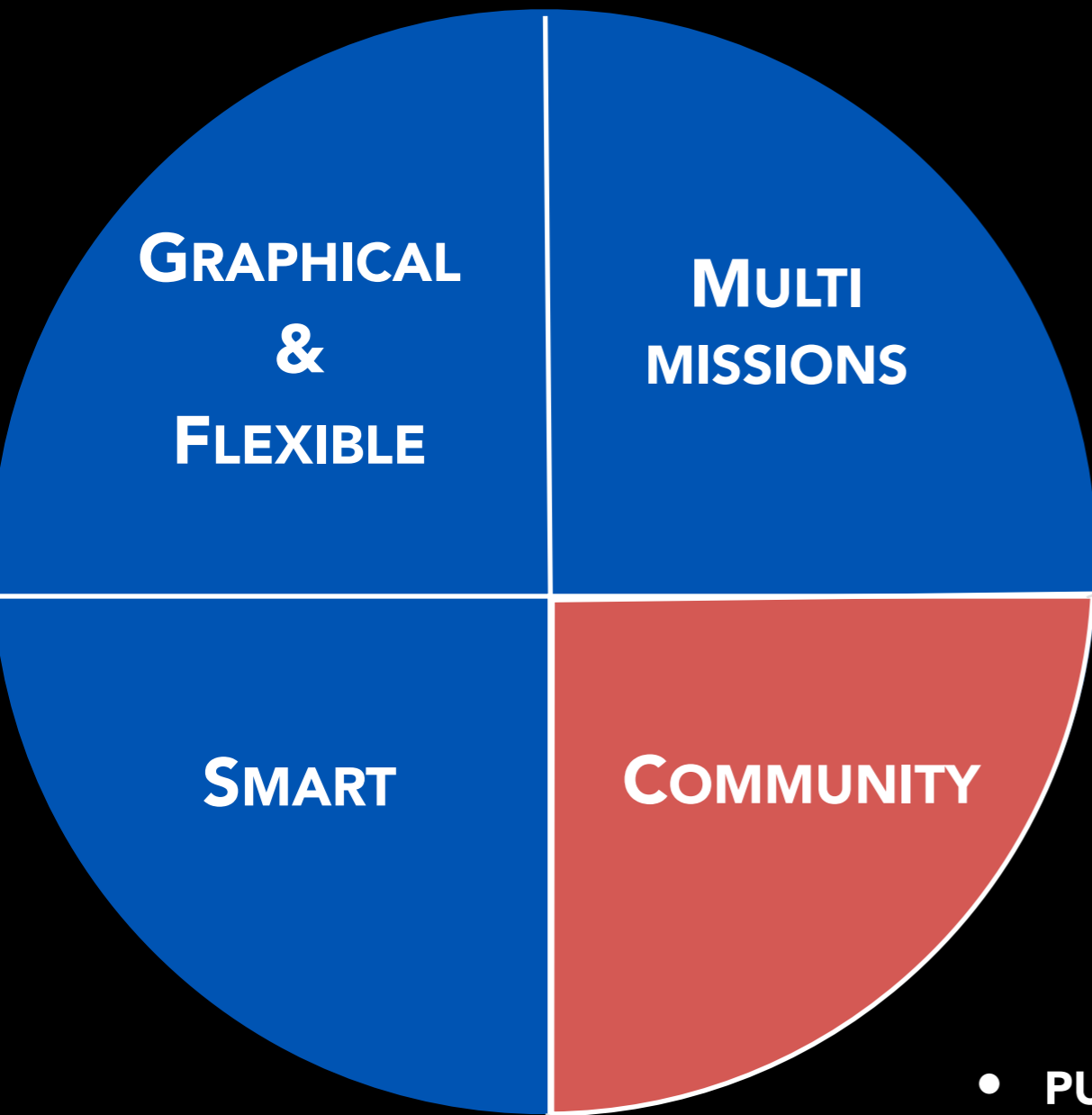
- **Catalogs and published studies** (> *SciQLOP v.1*)
Improve reproducibility - ANTI-reinventing-the-wheel-tool

- Export to publishable additional material catalogs with custom fields
- Catalog type = « published event »
 - Register an event as « published » and add DOI/ paper meta data
 - SciQLOP will let you know visually that the event you're looking at has been published and let you easily grab the paper





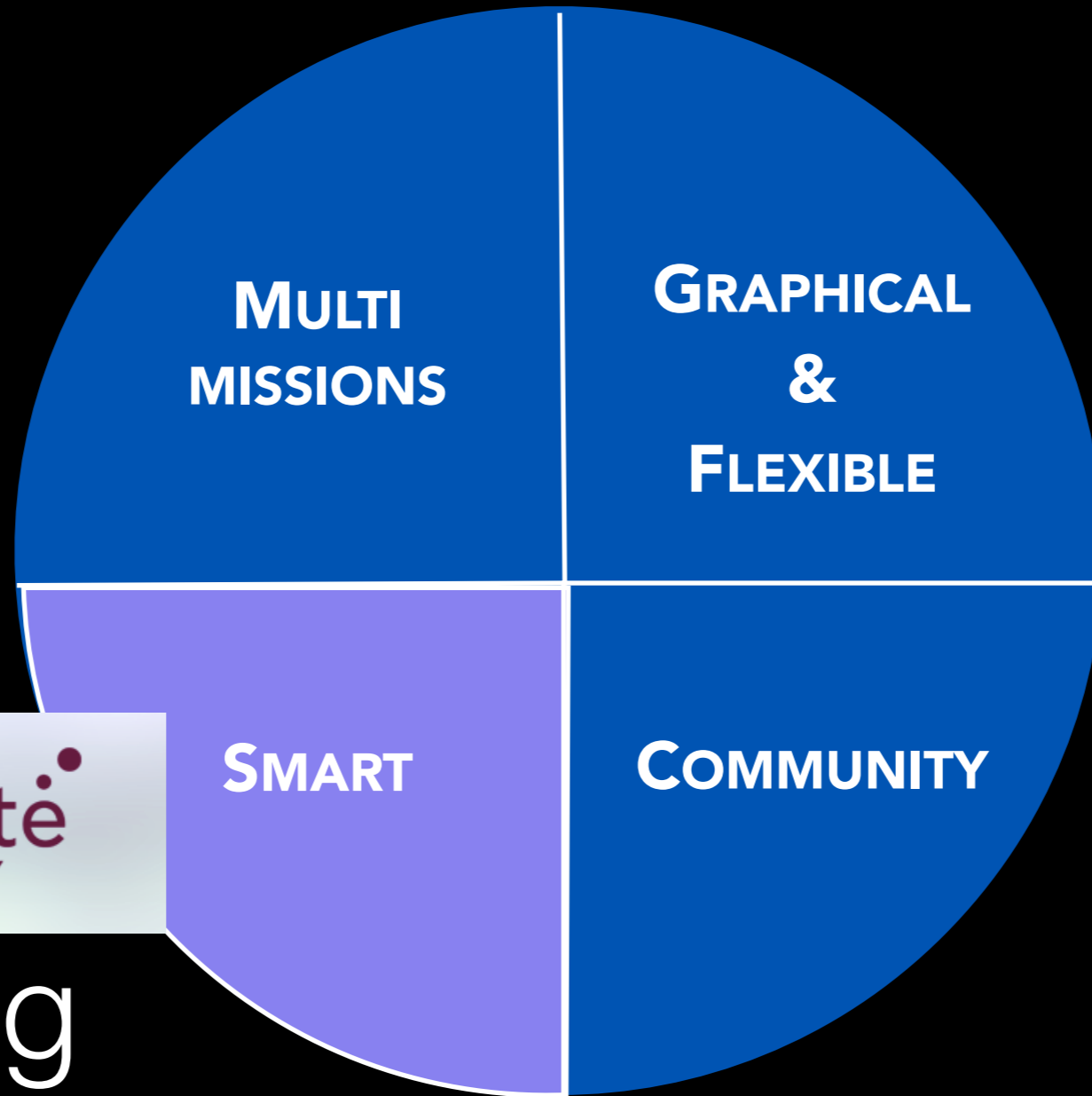
Share science



- **ORGANIZE DATA INTO CATALOGS**

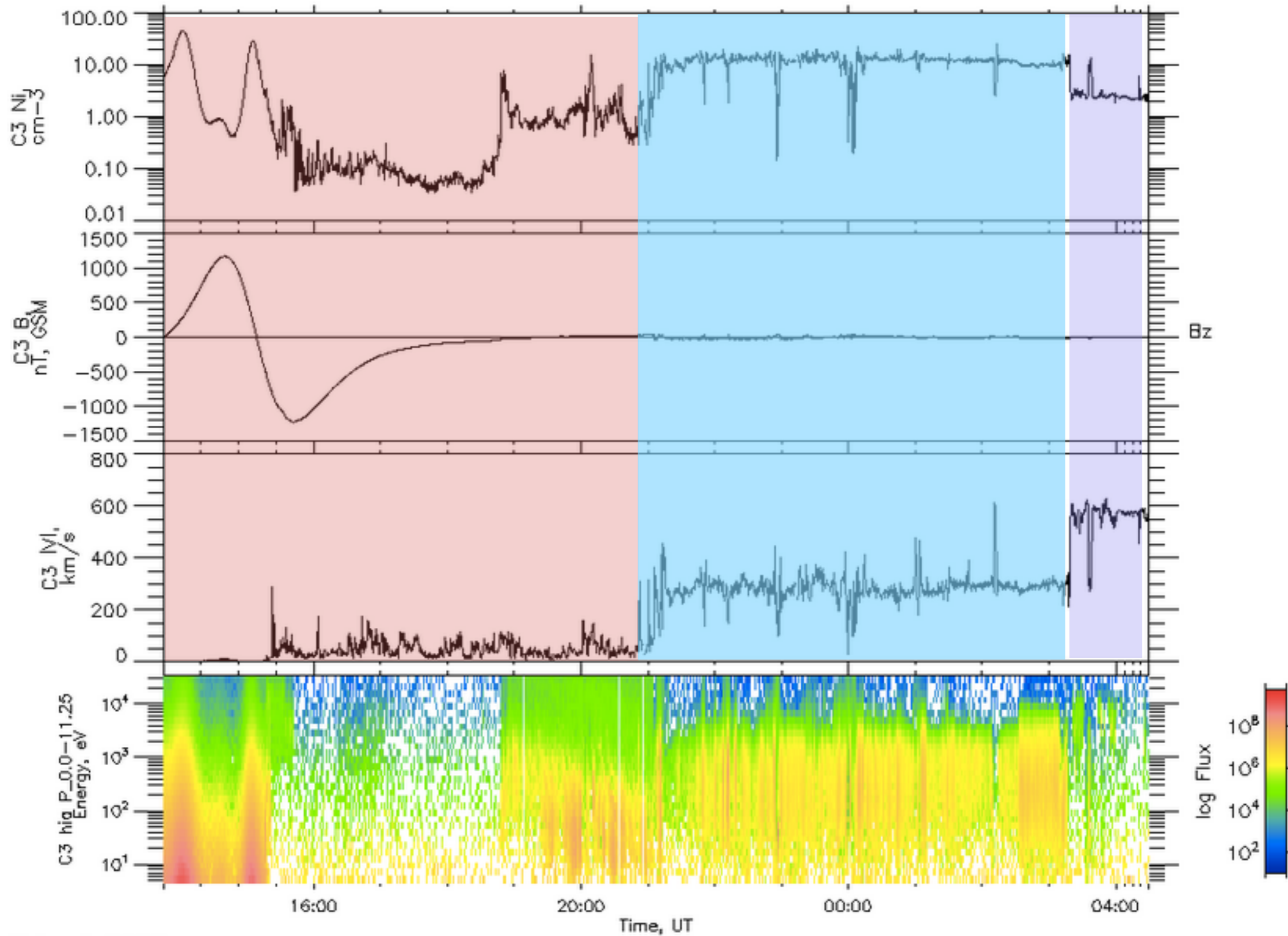
- **COLLABORATIVE CATALOGS**

- **PUSH AND PULL PUBLISHED DATA**

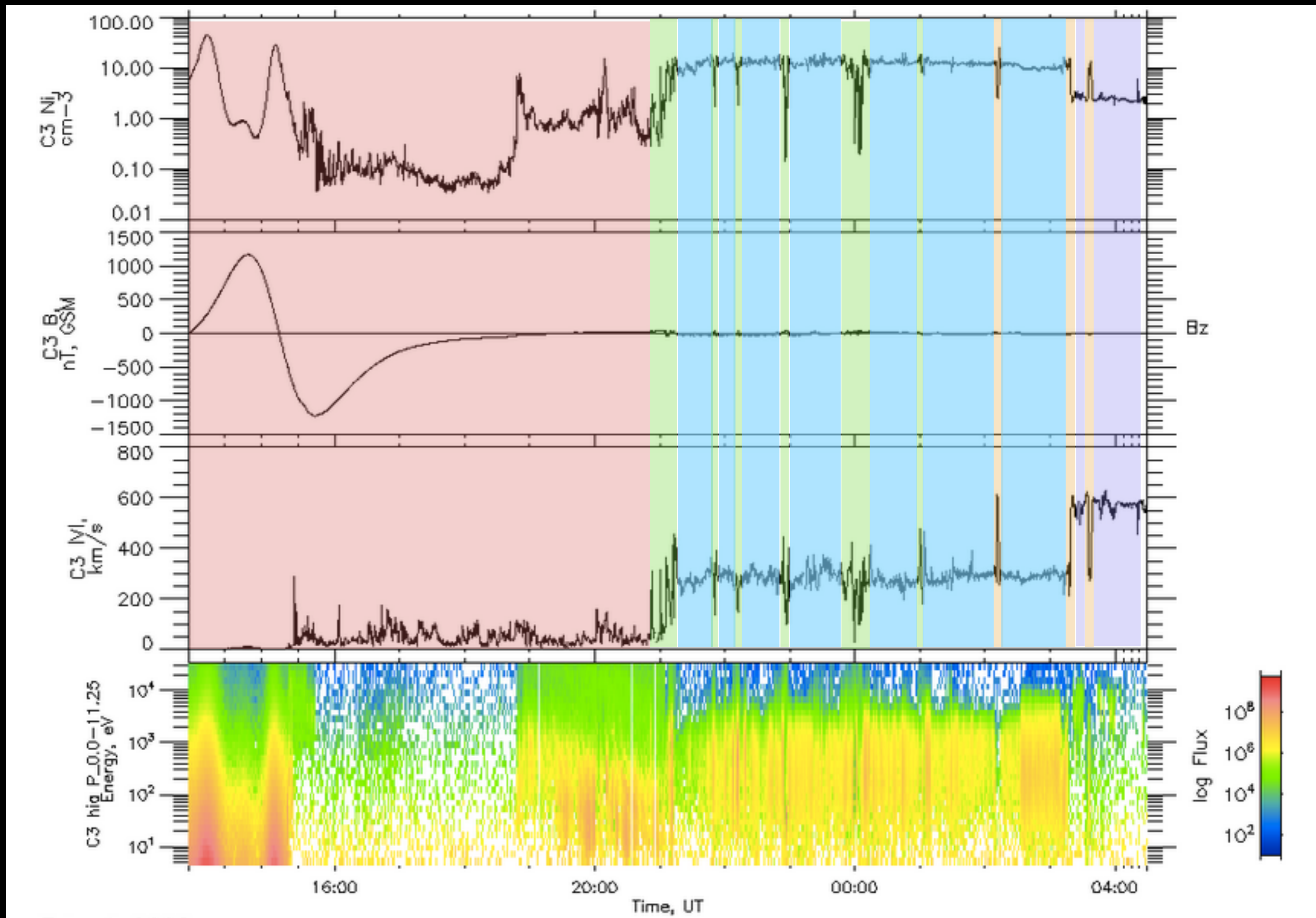


Learning
from
Space
Data

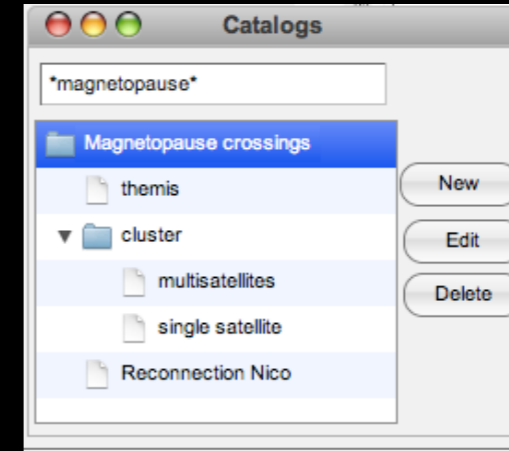
Automatic event
detection with Machine
Learning



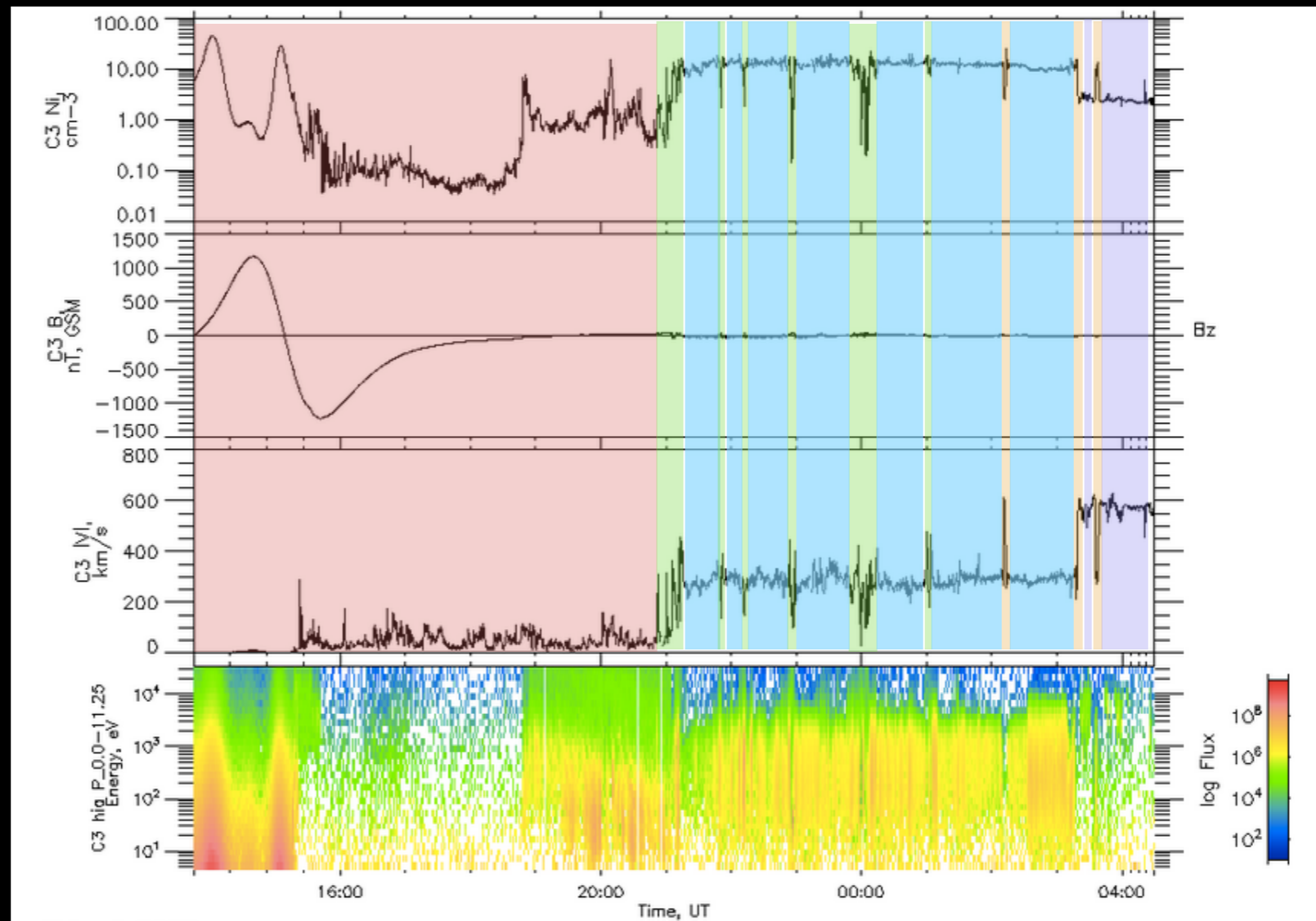
SPACE TIME AMBIGUITY



MACHINE LEARNING

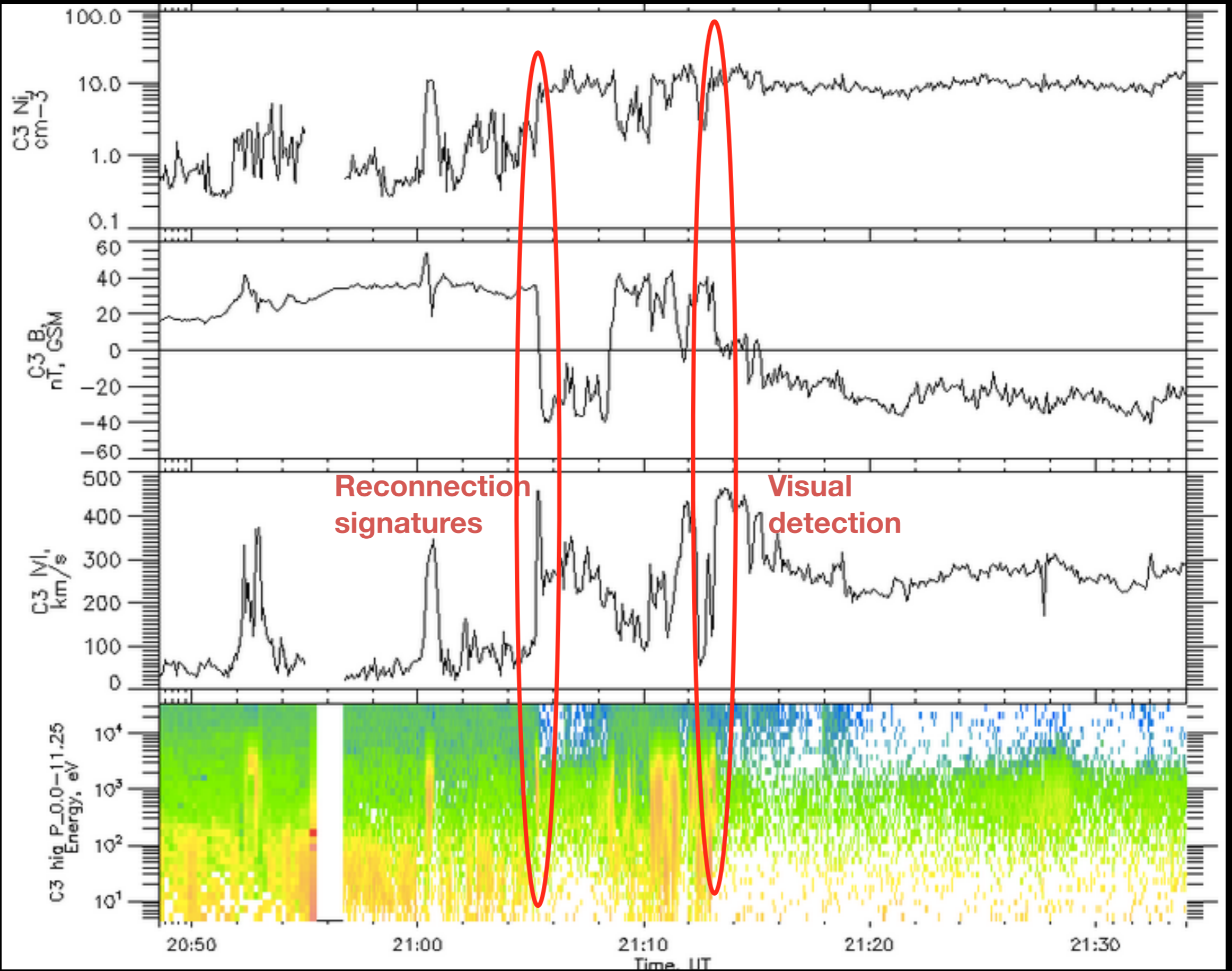


-  *M'sphere*
-  *M'pause*
-  *M'sheath*
-  *Shock*
-  *Solar wind*



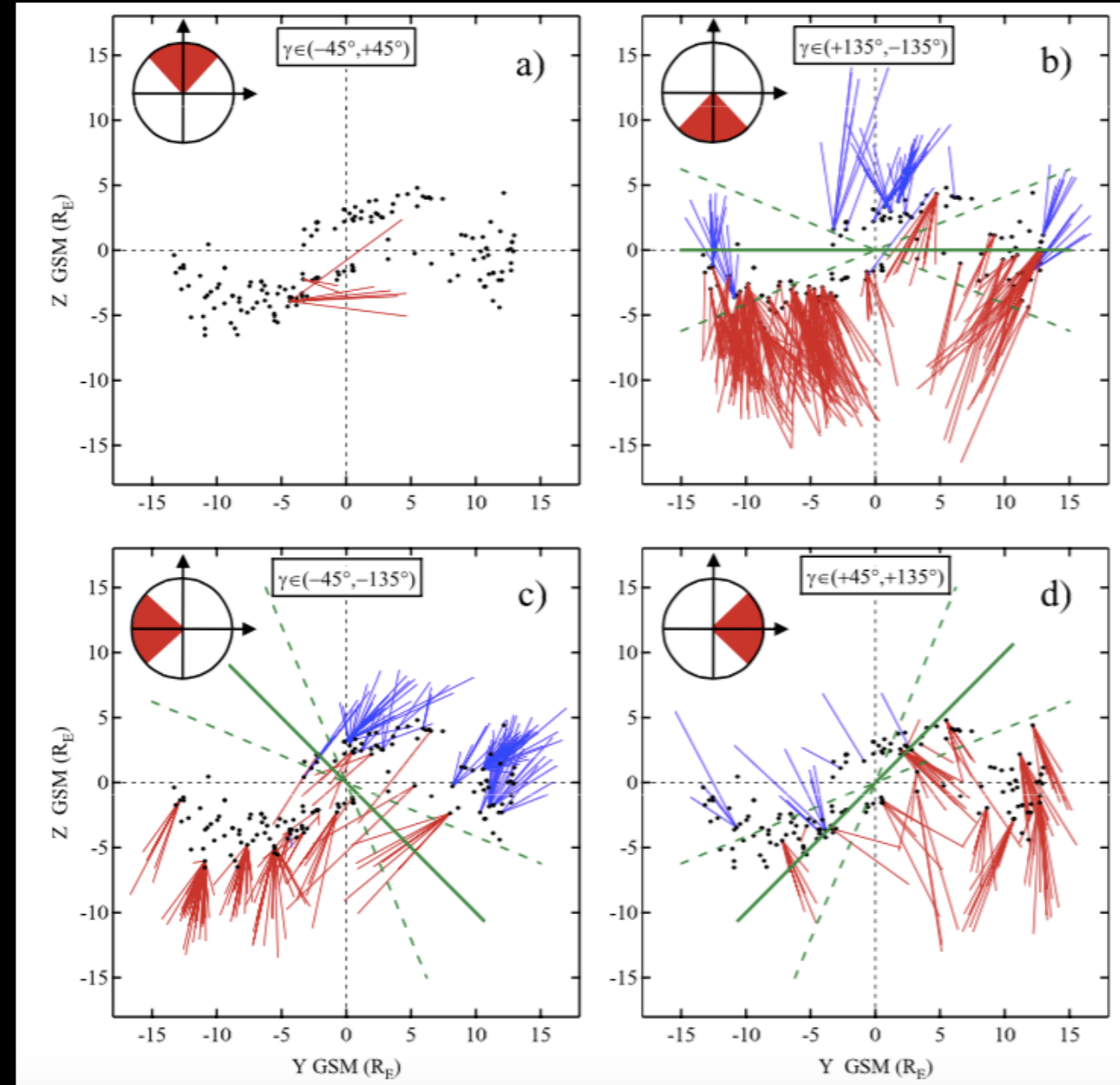
*Auto
select
regions*



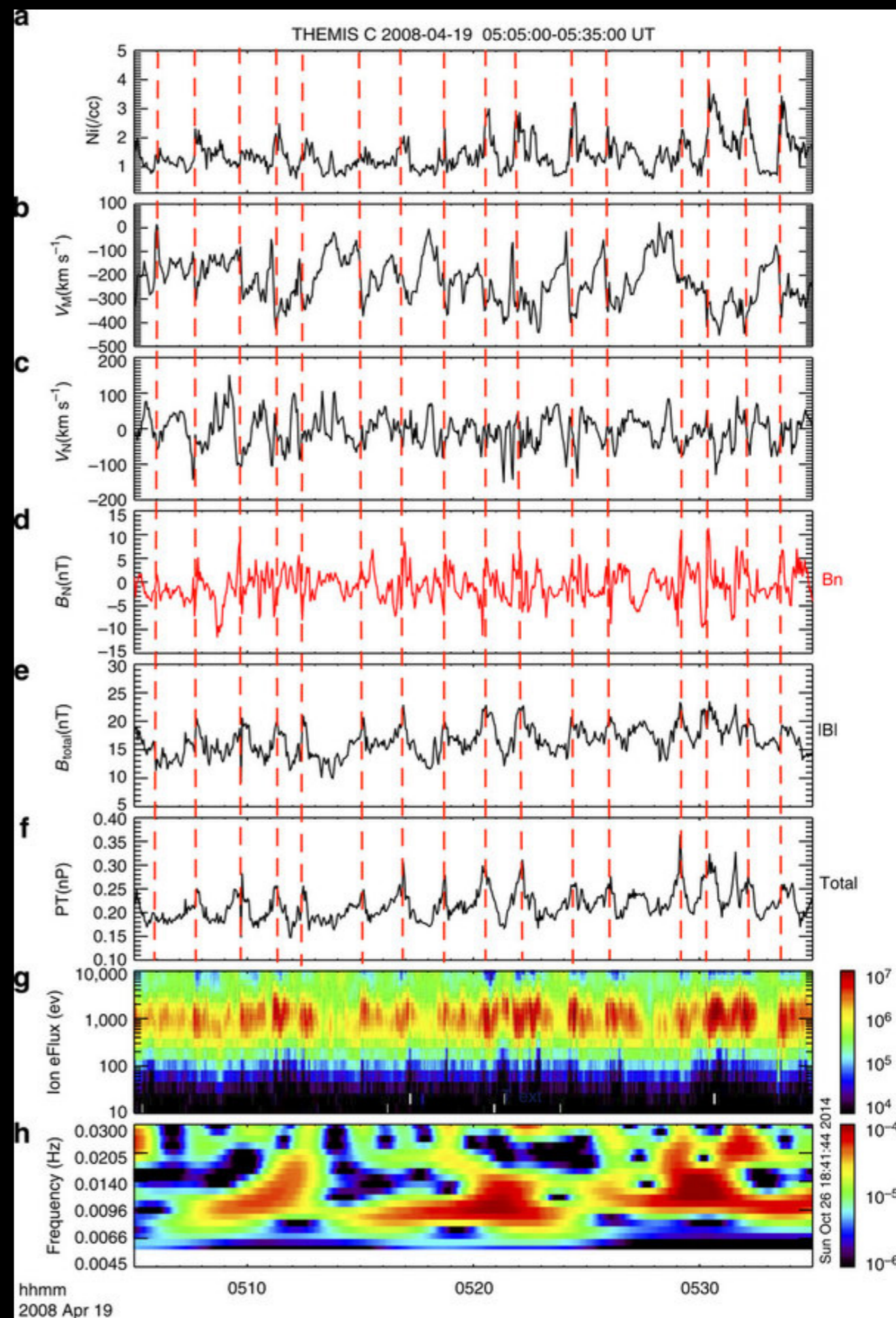
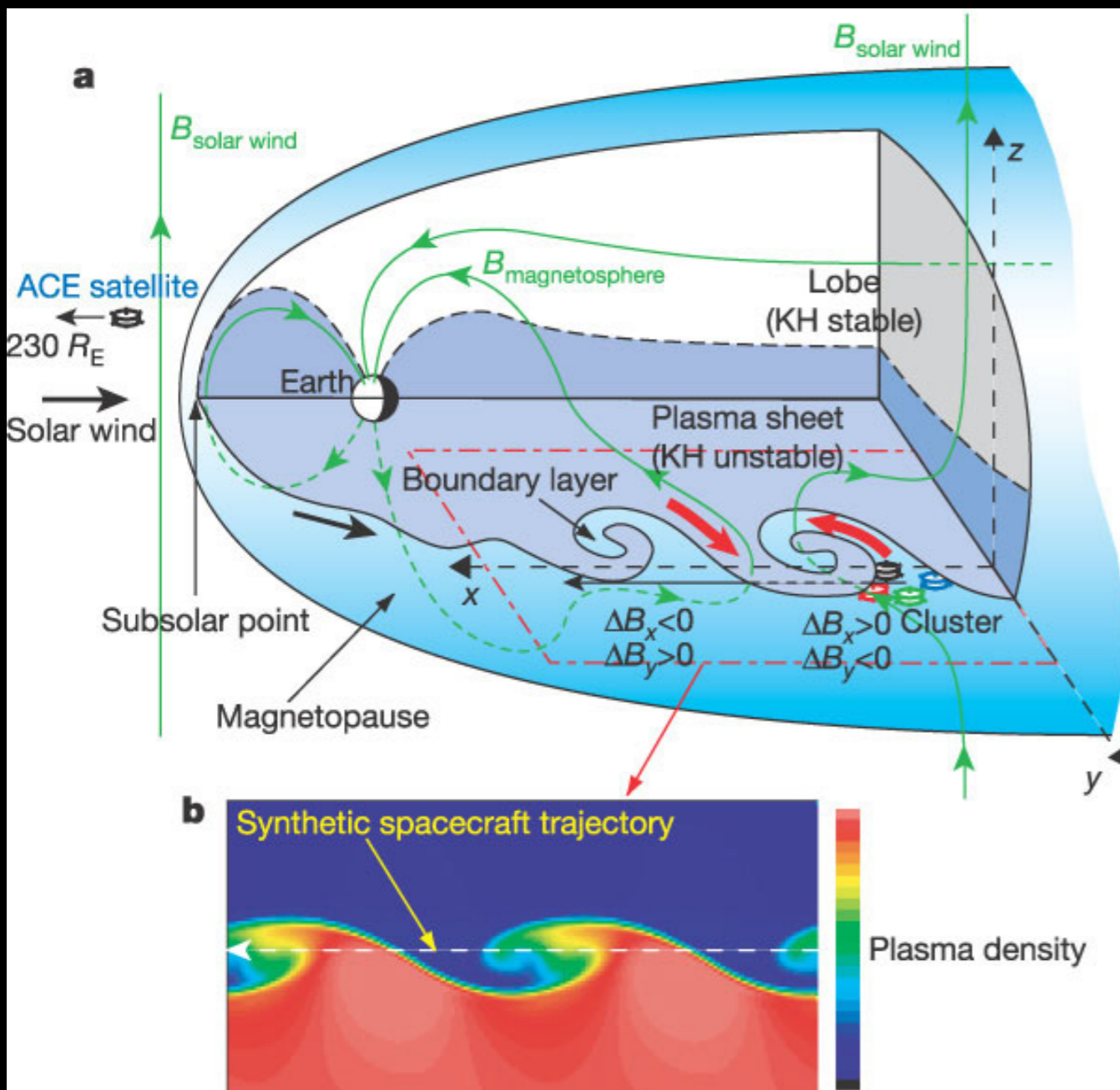


[3] Several statistical studies have been performed in the past to address these questions. *Gosling et al.* [1990] presented ISEE 2 data relative to 17 events for which reconnection flows at the dayside low latitude boundary layer had velocities opposite-directed with respect to the magnetosheath flow in order to study the influence of the IMF B_y component. *Scurry et al.* [1994] instead selected and studied 58 events at the low-latitude dayside MP for which the reconnection flows had a higher flow speed than the adjacent magnetosheath. *Phan et al.* [1996] studied 69 MP crossings with high magnetic shear ($>45^\circ$) at the low-latitude dayside MP and quantitatively analyzed the agreement of the flow change across the MP with that predicted for reconnection. More recently, *Paschmann et al.* [2005] discussed the results of a study of 60 MP crossings, among which 19 show reconnection signatures, observed by Clus-

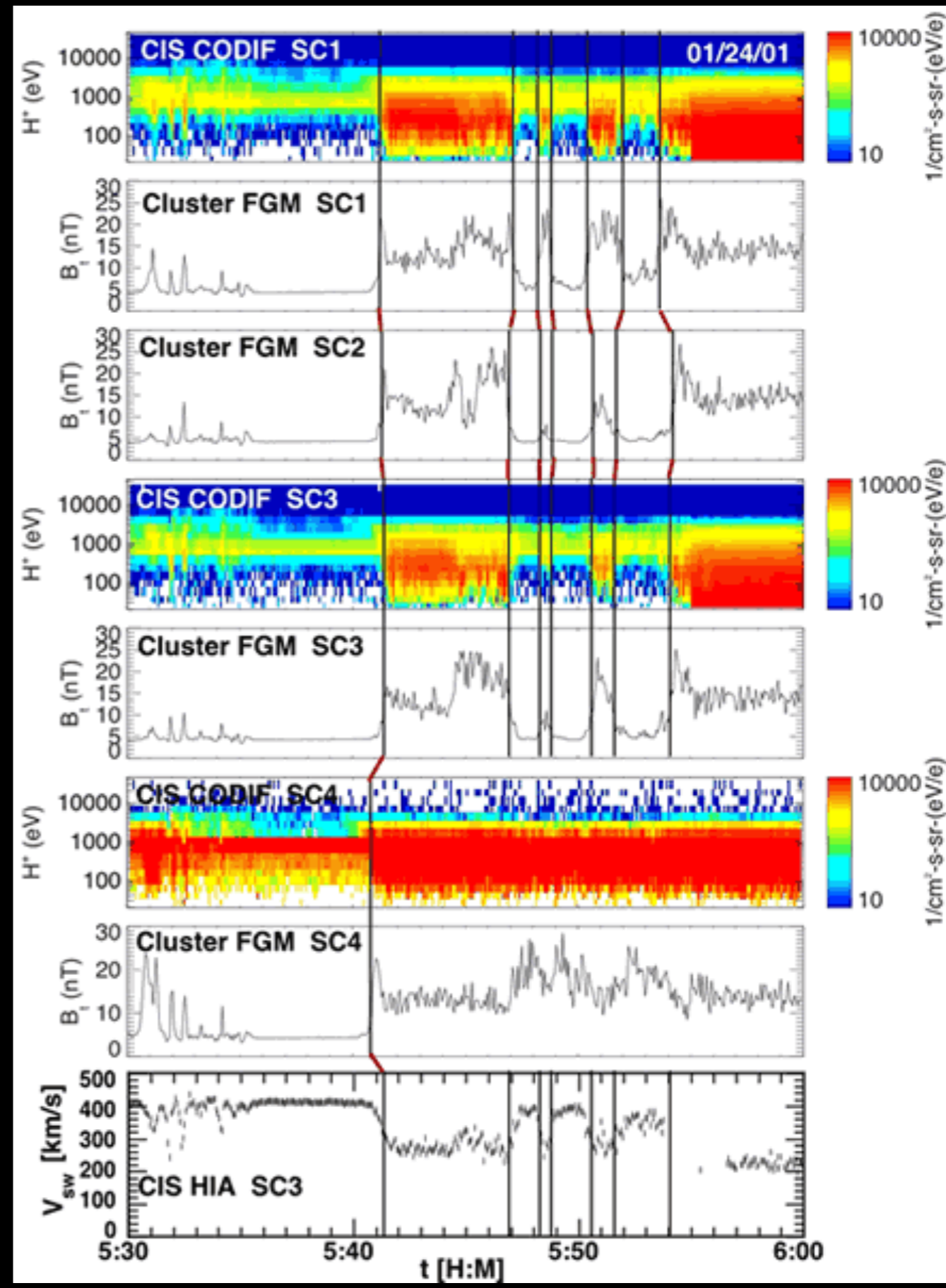
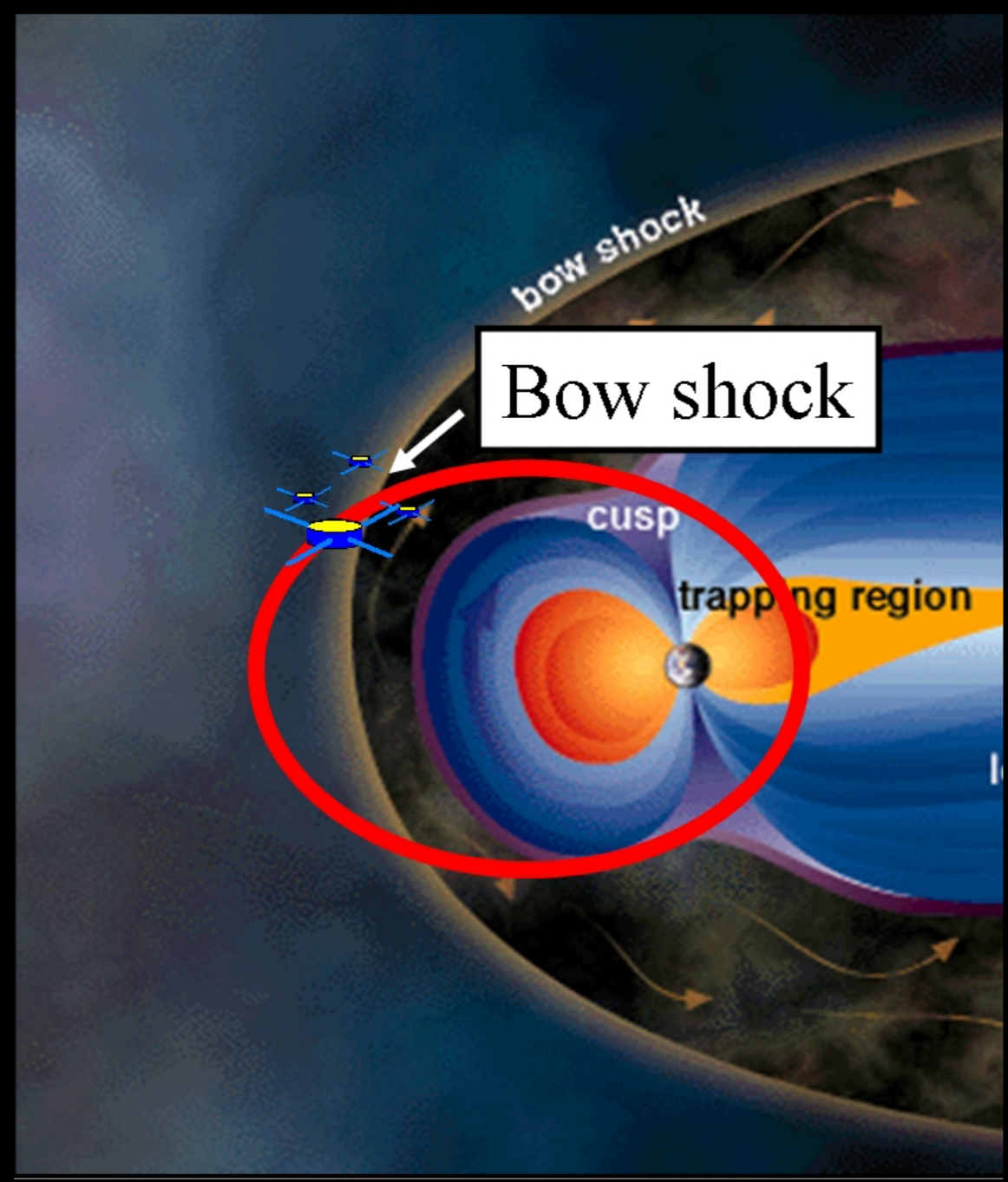
[4] Here, we use Double Star TC1 plasma and magnetic field data to study reconnection occurrence at the dayside low-latitude magnetopause. The orbit of the TC1 satellite, which is almost equatorial with an apogee of $12.4 R_E$, is particularly suitable for this study. We analyzed 239 MP crossings, 143 of which show reconnection signatures. Twenty three out of the 143 MP crossings show flow reversals, which indicate that the crossings occur near the reconnection X point. We present the occurrence of the



KELVIN HELMHOLTZ



SHOCK CROSSING



COLLECTING (AUTOMATICALLY) DATA IS HARD

- DATA IS COMPLEX, NOT REPRODUCIBLE
- NAÏVE DETECTION ALGO. BASED ON FIXED RULES GIVE $> 70\%$ FALSE DETECTIONS

EASIEST THING IS STILL THE EYE

- EVERYONE KNOWS THE « TEXTBOOK » EXAMPLE OF OUR FAVORITE PHENOMENA REPRESENTS LESS THAN 1% OF EVENTS
- PREVENTS STATISTICAL STUDIES OF PHENOMENA
- HOW DO WE USE YEARS OF ARCHIVED DATA??
- LISTS ARE COMPILED HERE AND THERE ... BAD REPRODUCIBILITY
- WHAT DO WE DO WHEN WE RUN OUT OF INTERNS TO SELECT INTERVALS?

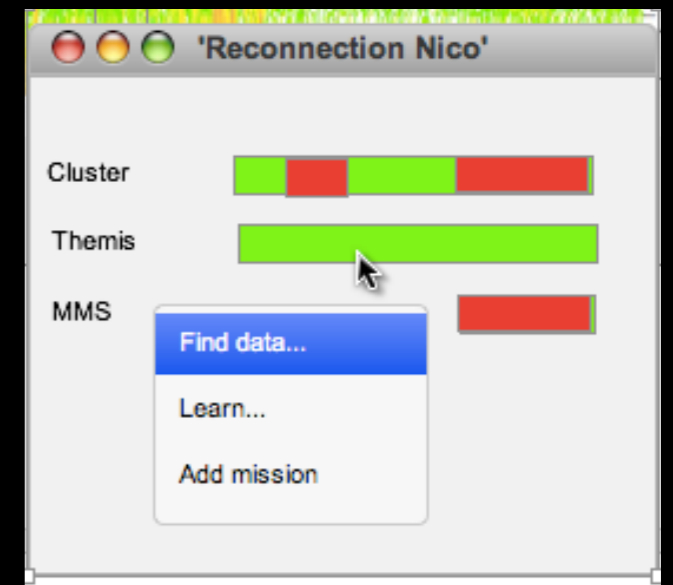
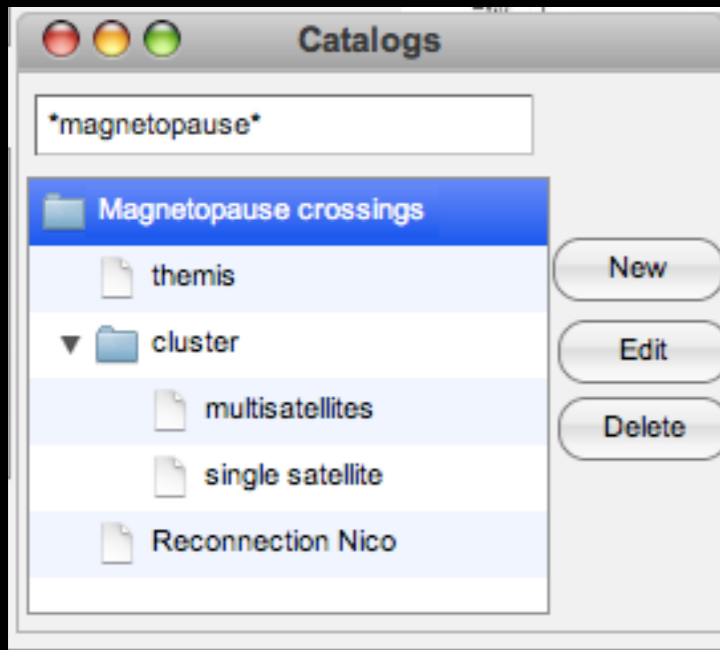


- **ML from and for catalogs**

- Learn from catalogs

- suggest new events

- scan databases



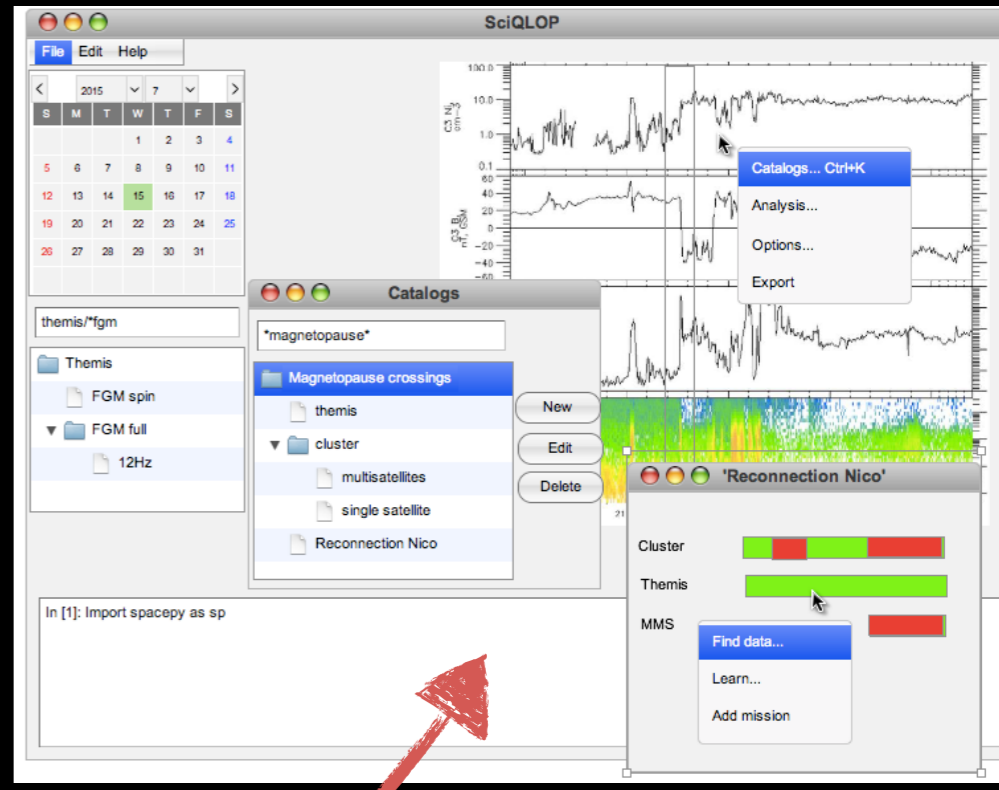
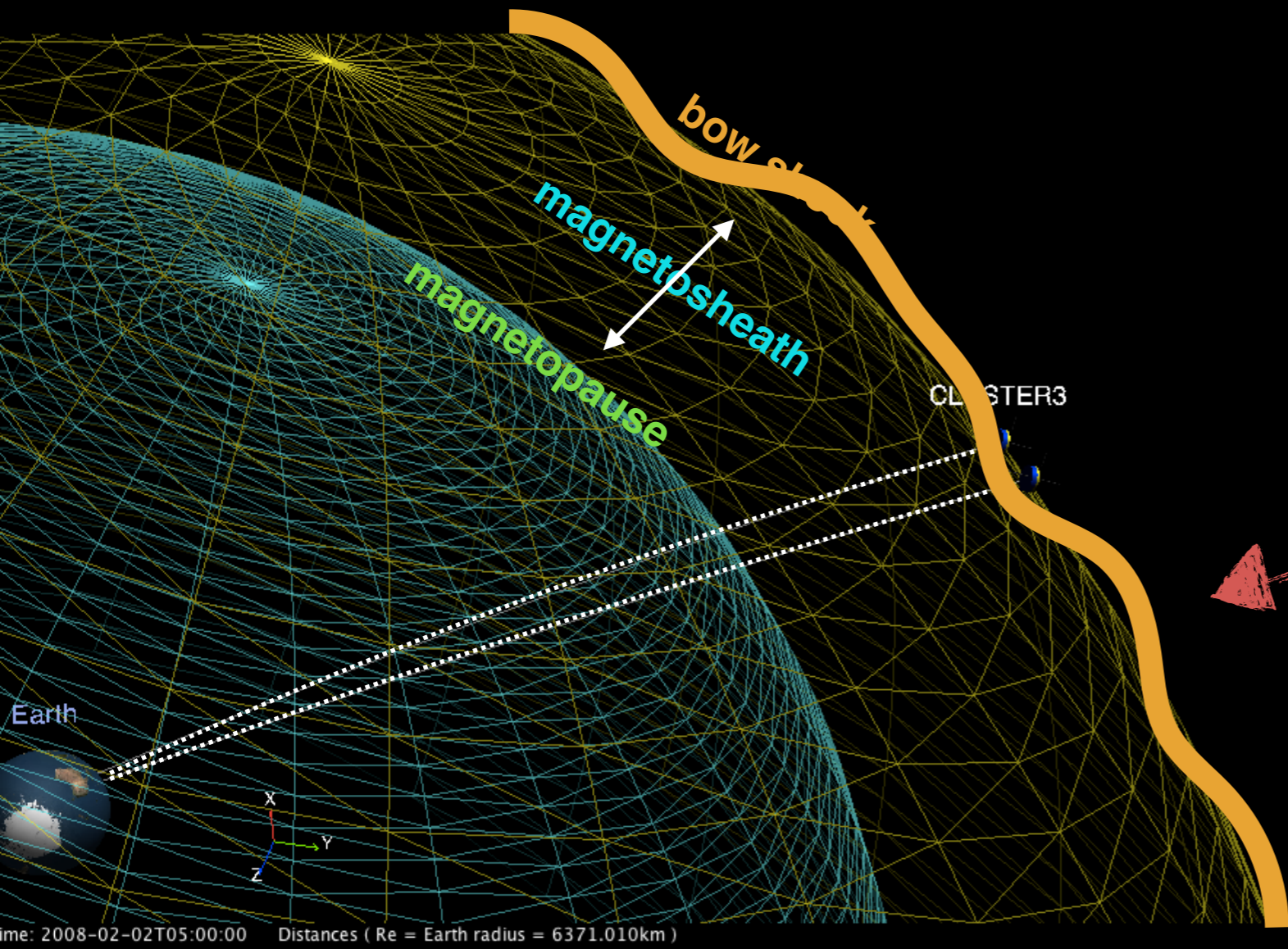
- Extend catalogs

- Test performance



• Using catalogs to do science. E.g. shock model as a function of IMF and Sw Mach nber.

• What is the 3D shape/position of the shock as a function of solar wind control parameters ?

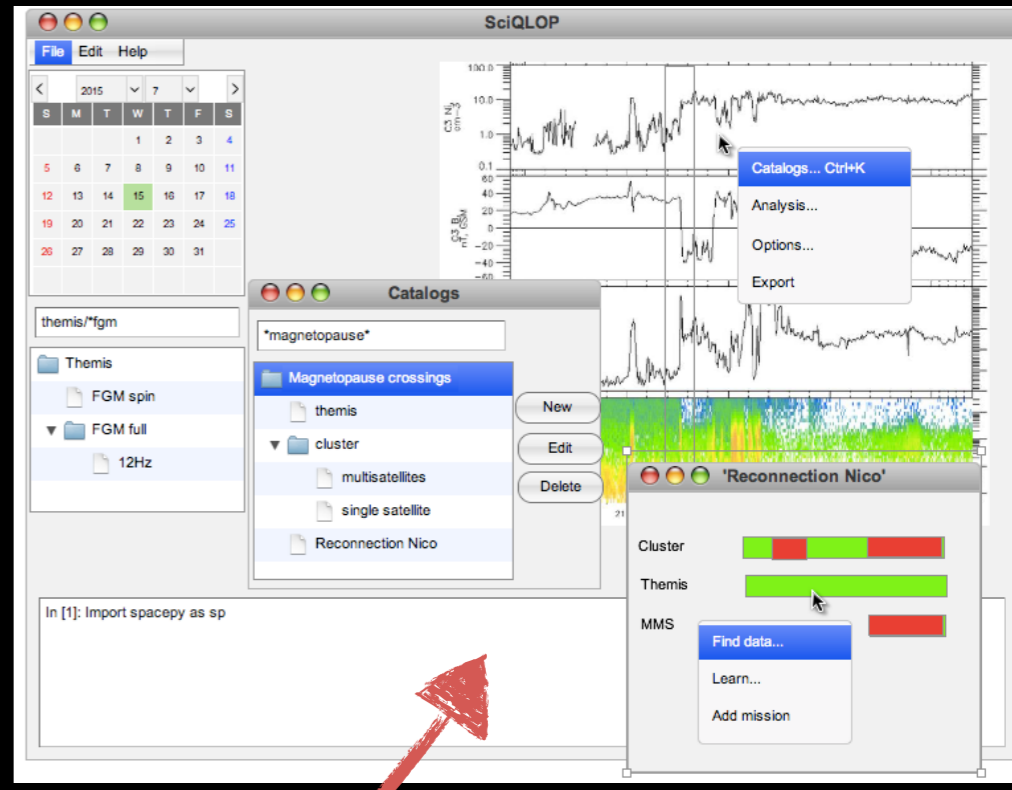
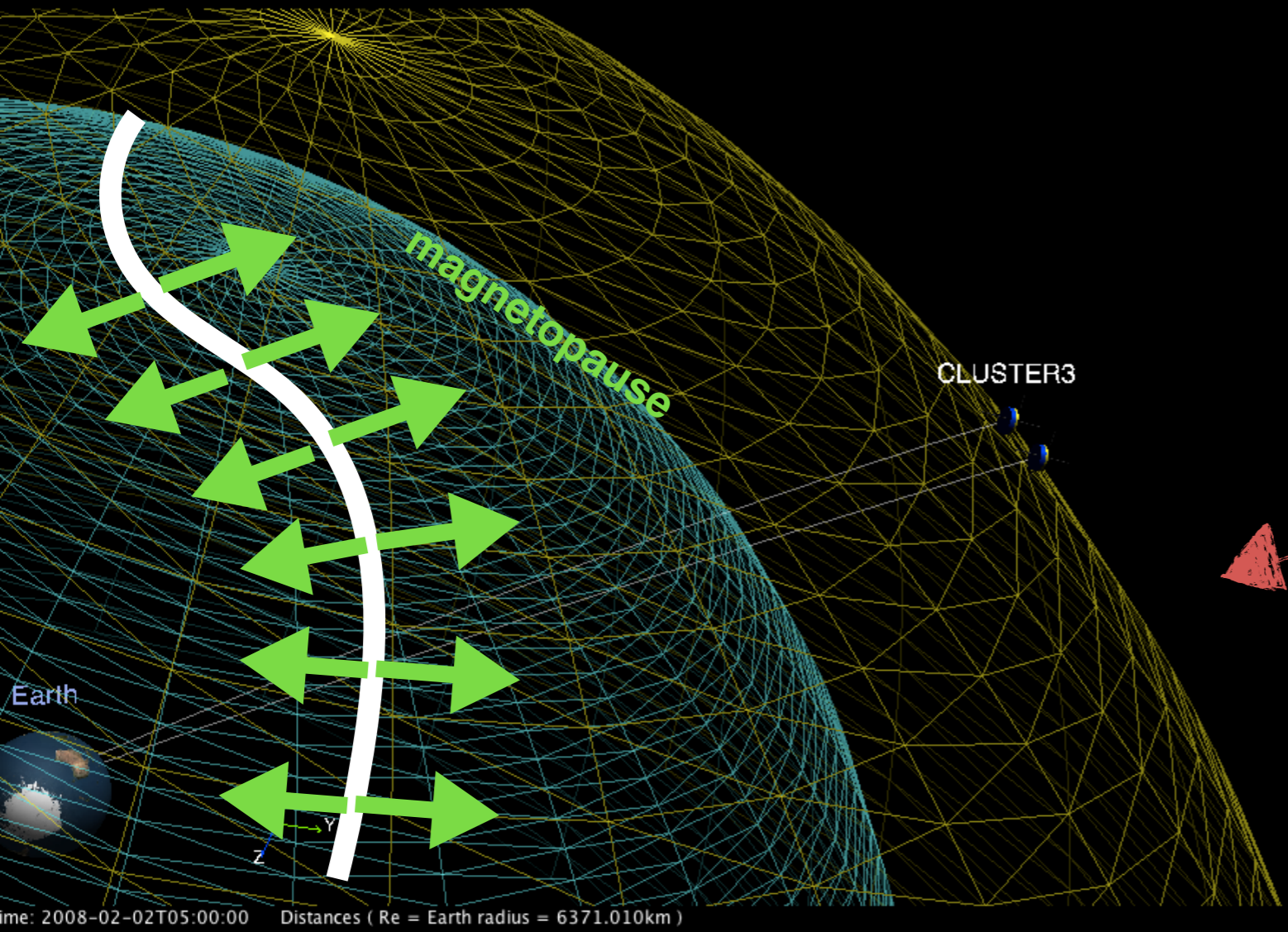


- build and share models based on catalogued data
- Export model to 3DView (collaboration with CDPP)



- Using catalogs to do science. e.g. reconnection at the magnetopause

- What is the position of the X line on the magnetopause as a function of solar wind control parameters ?



- build and share models based on catalogued data
- Export model to 3DView (collaboration with CDPP)

AUTOMATIC DETECTION OF ICMEs

Magnetic clouds:

Very geoefficient structure

Huge structure lasting typically 1 day

Start with a discontinuity : jumps in B , V , n , T than in preceding solar wind

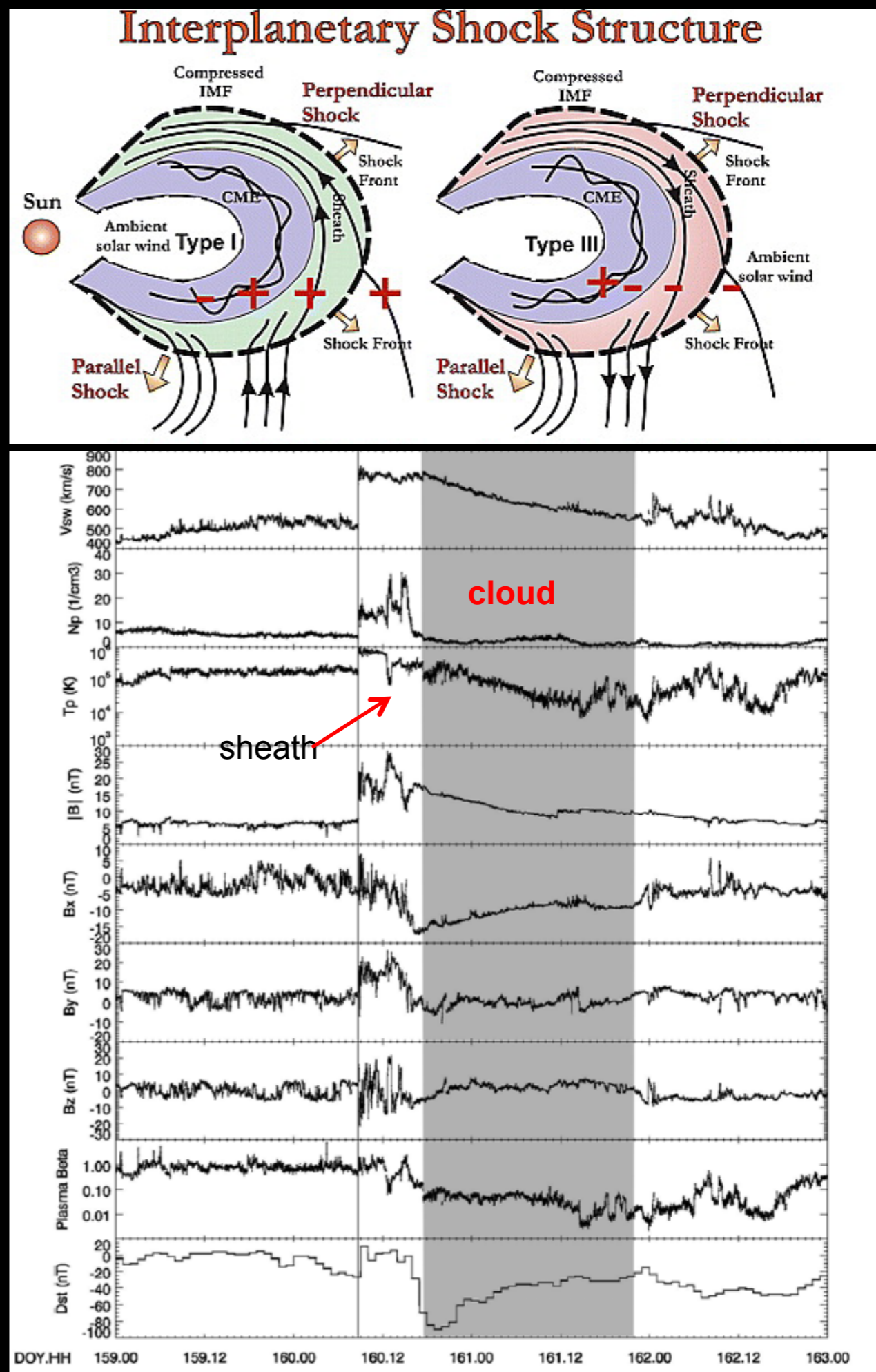
Then 2 parts:

(1) sheath:

large fluctuations

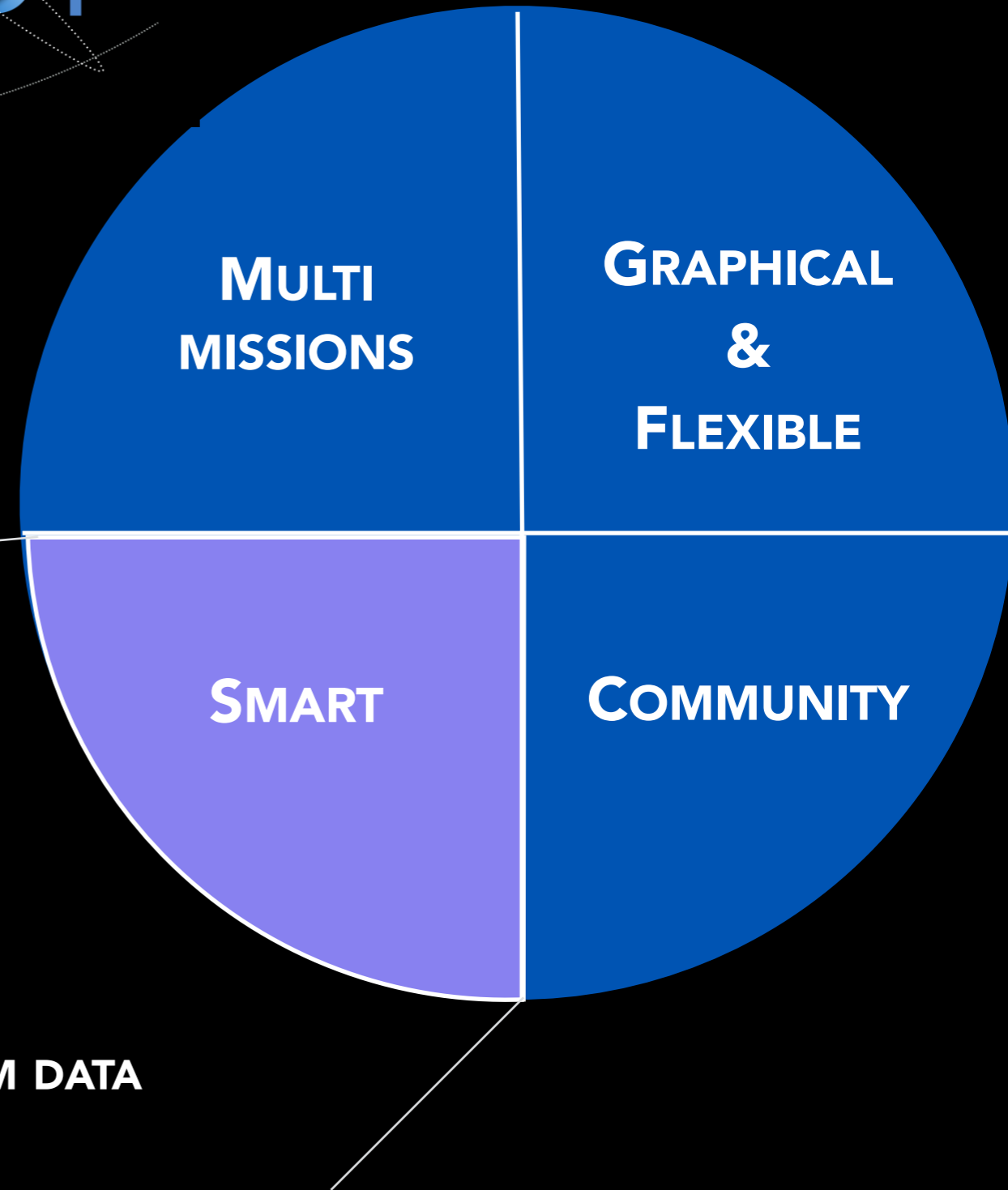
(2) Magnetic cloud itself:

- smooth variations
- Smooth B rotation





Learn data from/for users



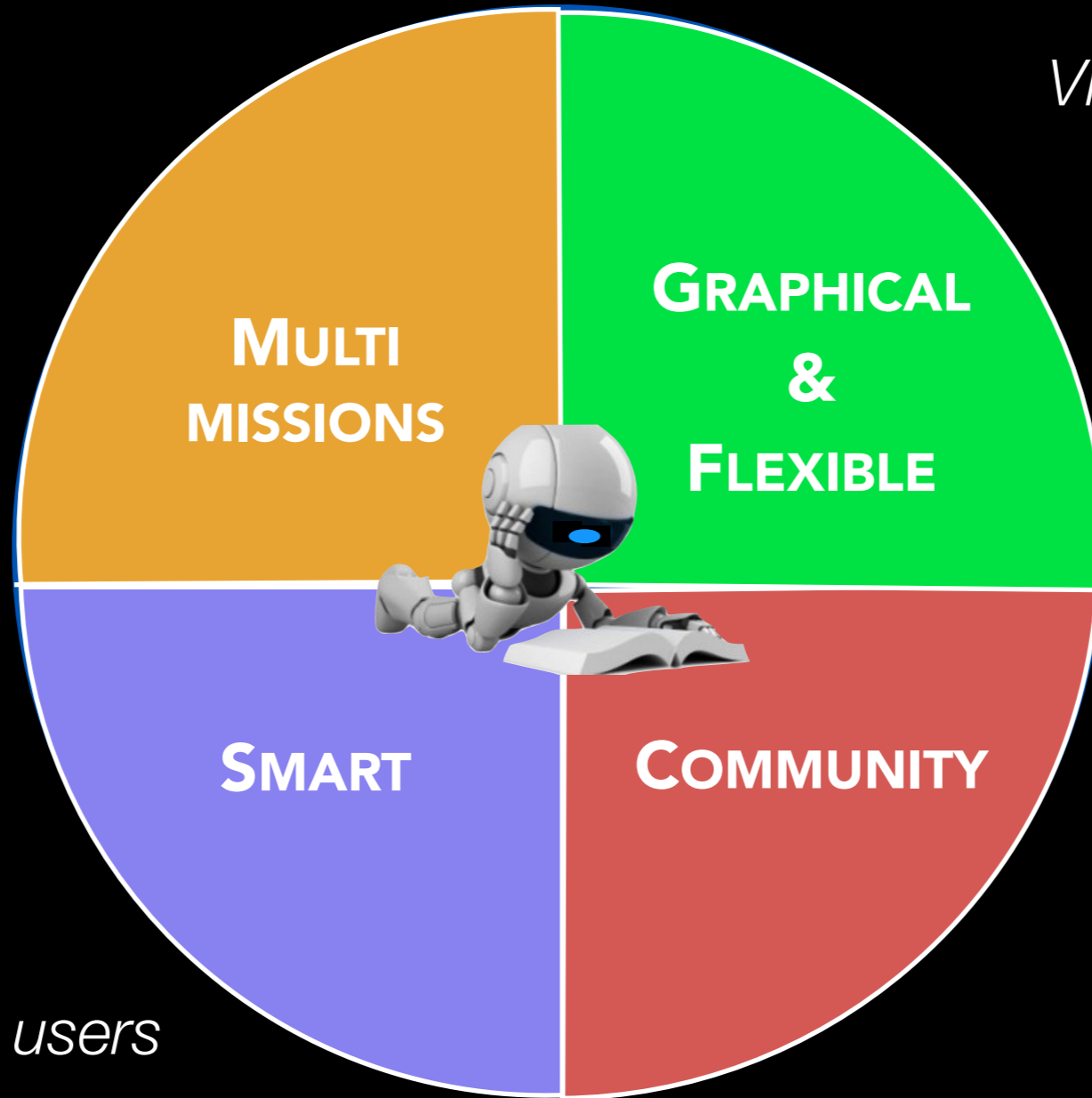
- **LEARN FROM CATALOGS**

- **SUGGEST DATA AND EXTEND CATALOGS**

- **BUILD COMPLEX MODELS FROM DATA**



Just get data



*Visualize complex data
interact with data*

Learn data from/for users

Share science

Learning from Space Data

- **DEFINE STRATEGIES TO DETECT :**
 - **REGIONS/BOUNDARIES**
 - **TAIL / M'PAUSE / SHOCK / ETC.**
 - **SOLAR WIND**
 - **EVENTS**
 - **SOLAR WIND SHOCKS**
 - **M'PAUSE RECONNECTION**
 - **M'PAUSE KH**
- **INTEGRATION IN SciQLOP**
 - **LEARN FROM CATALOGS**
 - **SCAN DATABASES**
 - **SUGGEST EVENTS**



FUTURE : Built-in SciQLoP engine?

