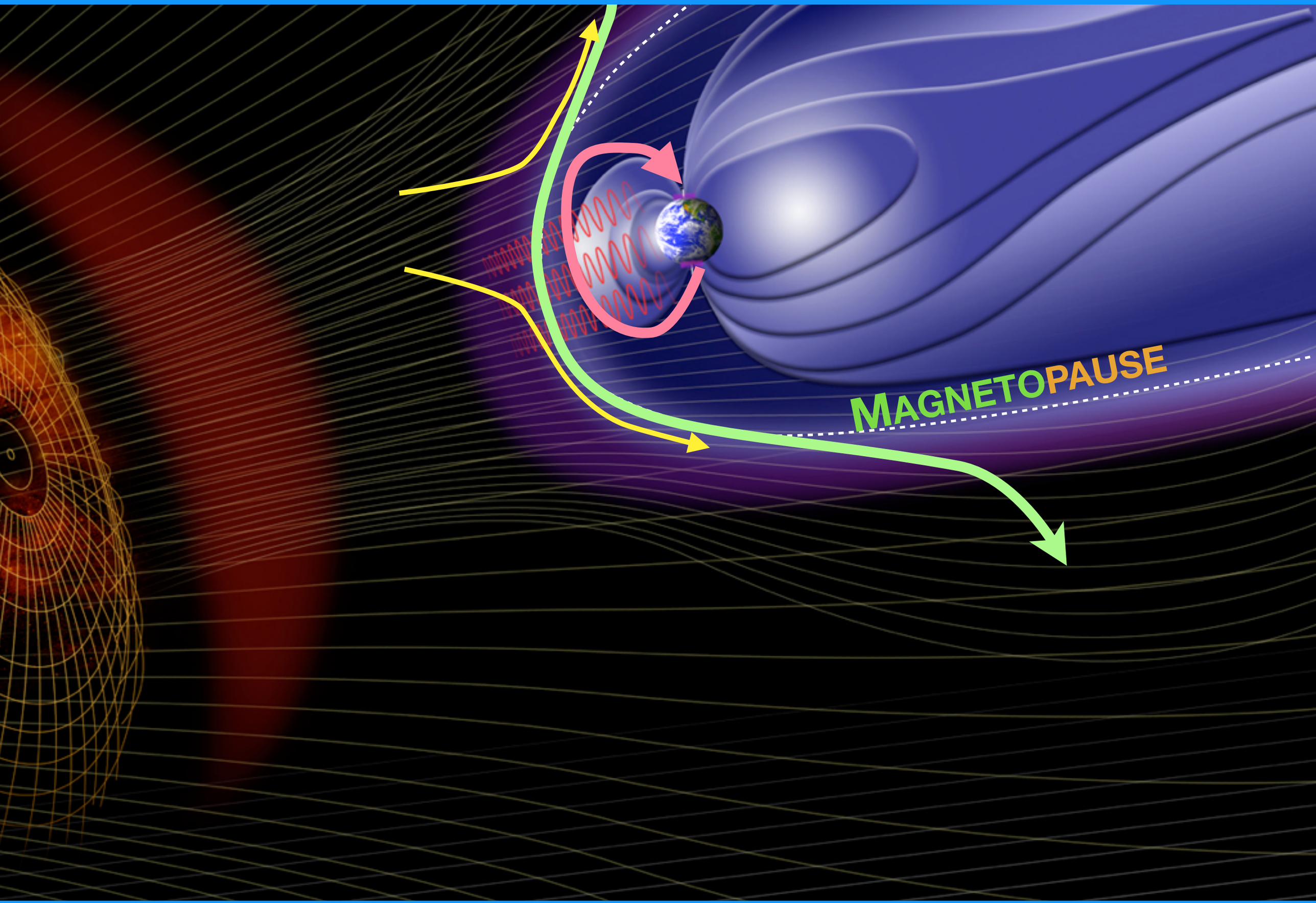
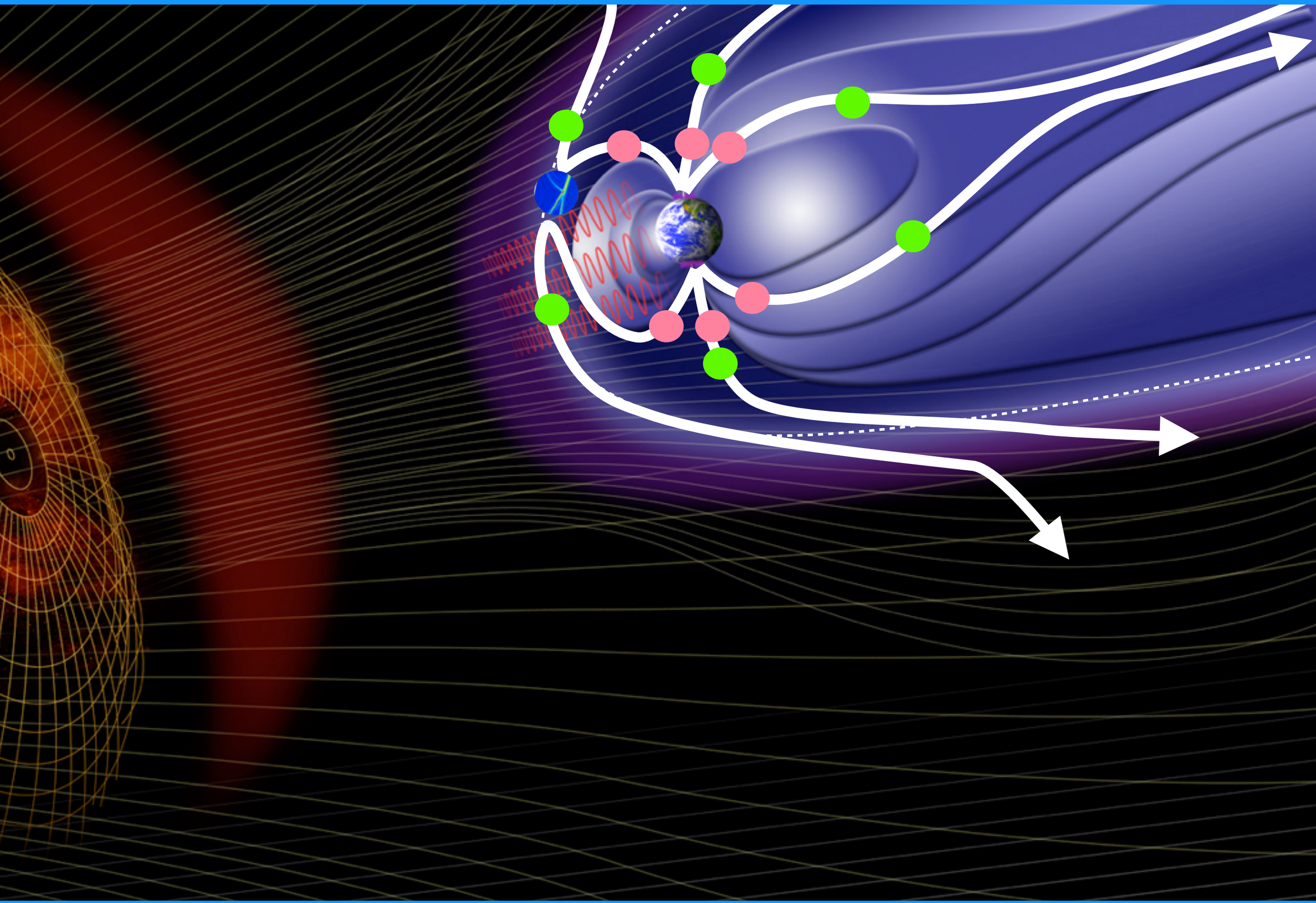
SciQLOP

Alexis Jeandet, Nicolas Aunai, Laurent Mirioni

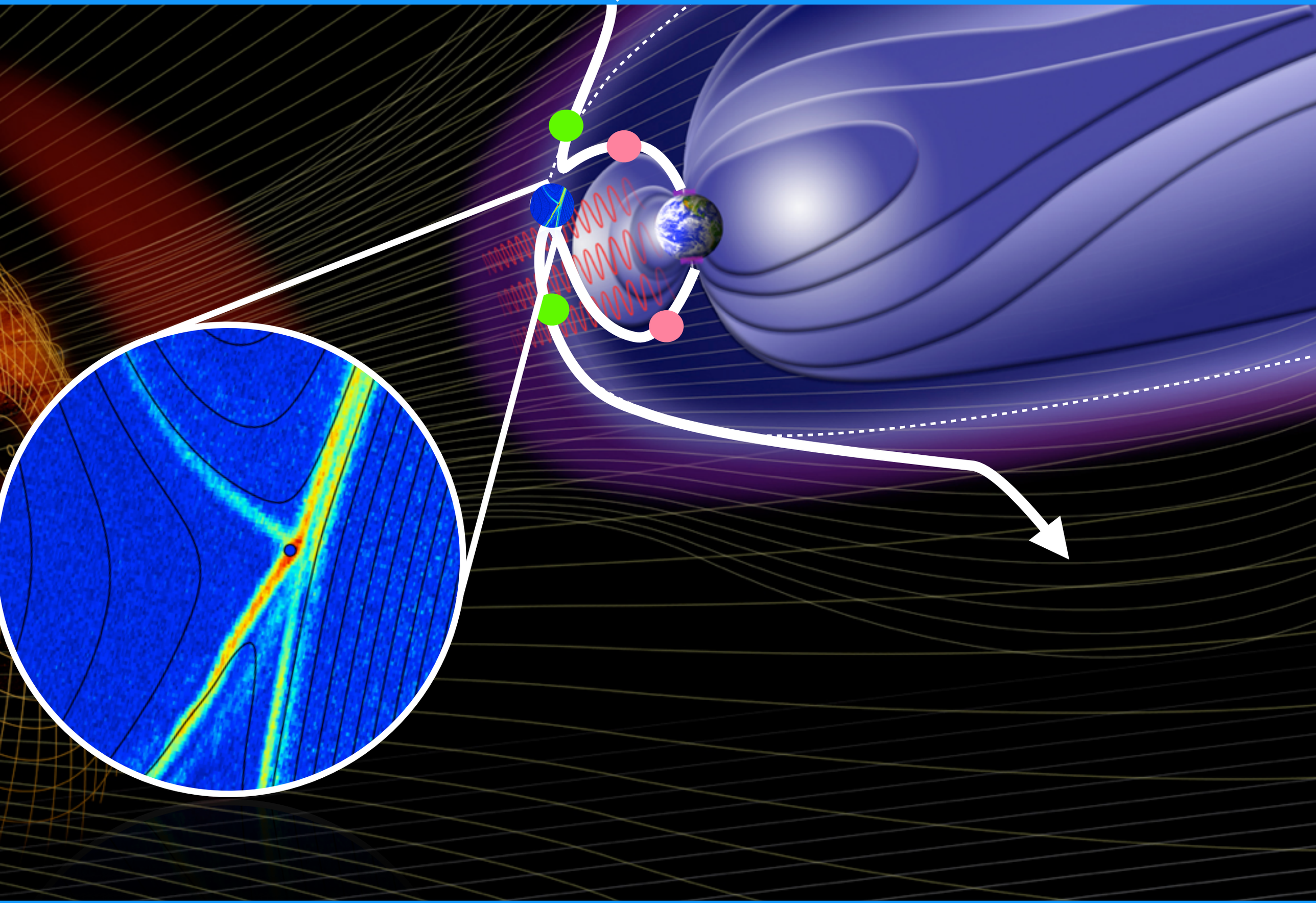
SOLAR WIND - MAGNETOSPHERE COUPLING



SOLAR WIND - MAGNETOSPHERE COUPLING



Big Questions



Big Questions

Where does reconnection occurs at the magnetopause?

How does solar wind and sheath plasma/B affect it?

What are the 3D shape and properties of the sheath?

When/where/how does KHI occur at the magnetopause?

What are the properties of ICMEs, shocks, sheaths, etc.?

What’s common to all these question?

Big Questions

Where does reconnection occurs at the magnetopause?

How does solar wind and sheath plasma/B affect it?

What are the 3D shape and properties of the sheath?

When/where/how does KHI occur at the magnetopause?

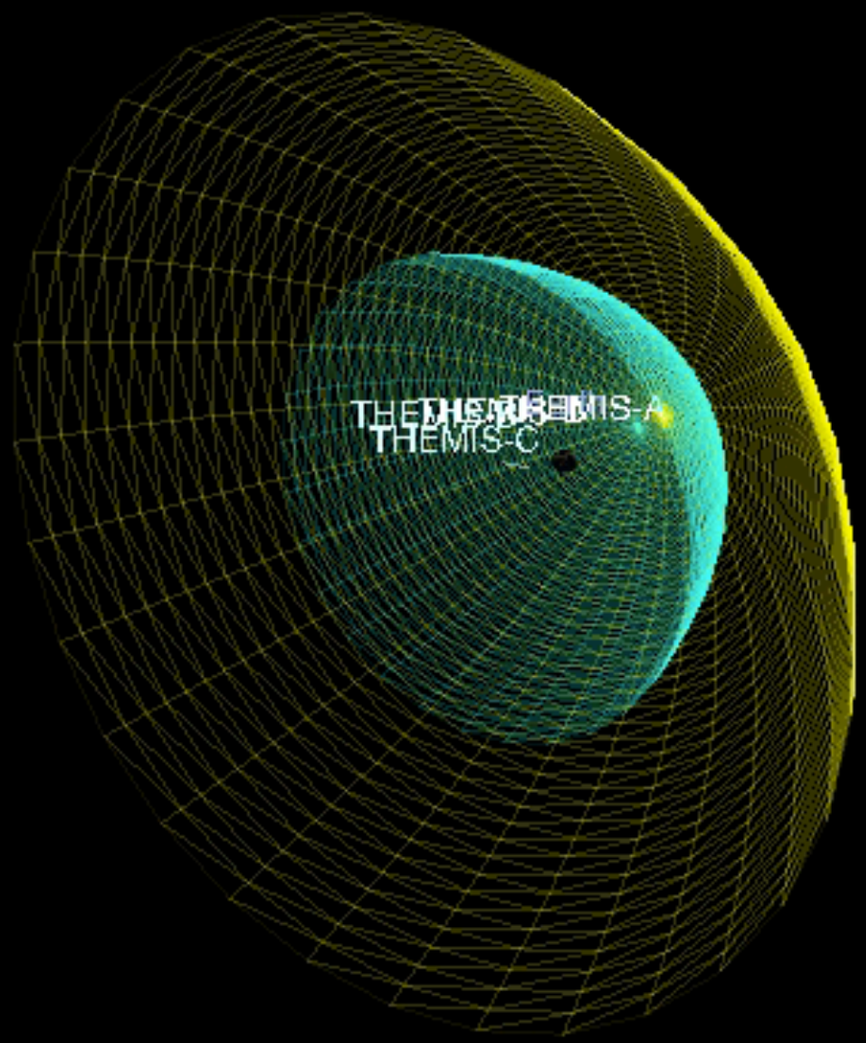
What are the properties of ICMEs, shocks, sheaths, etc.?

What's common to all these question?

They're all defined at the **global scale** but can only be answered by looking at **small scales.**

can**not** be answered with case studies

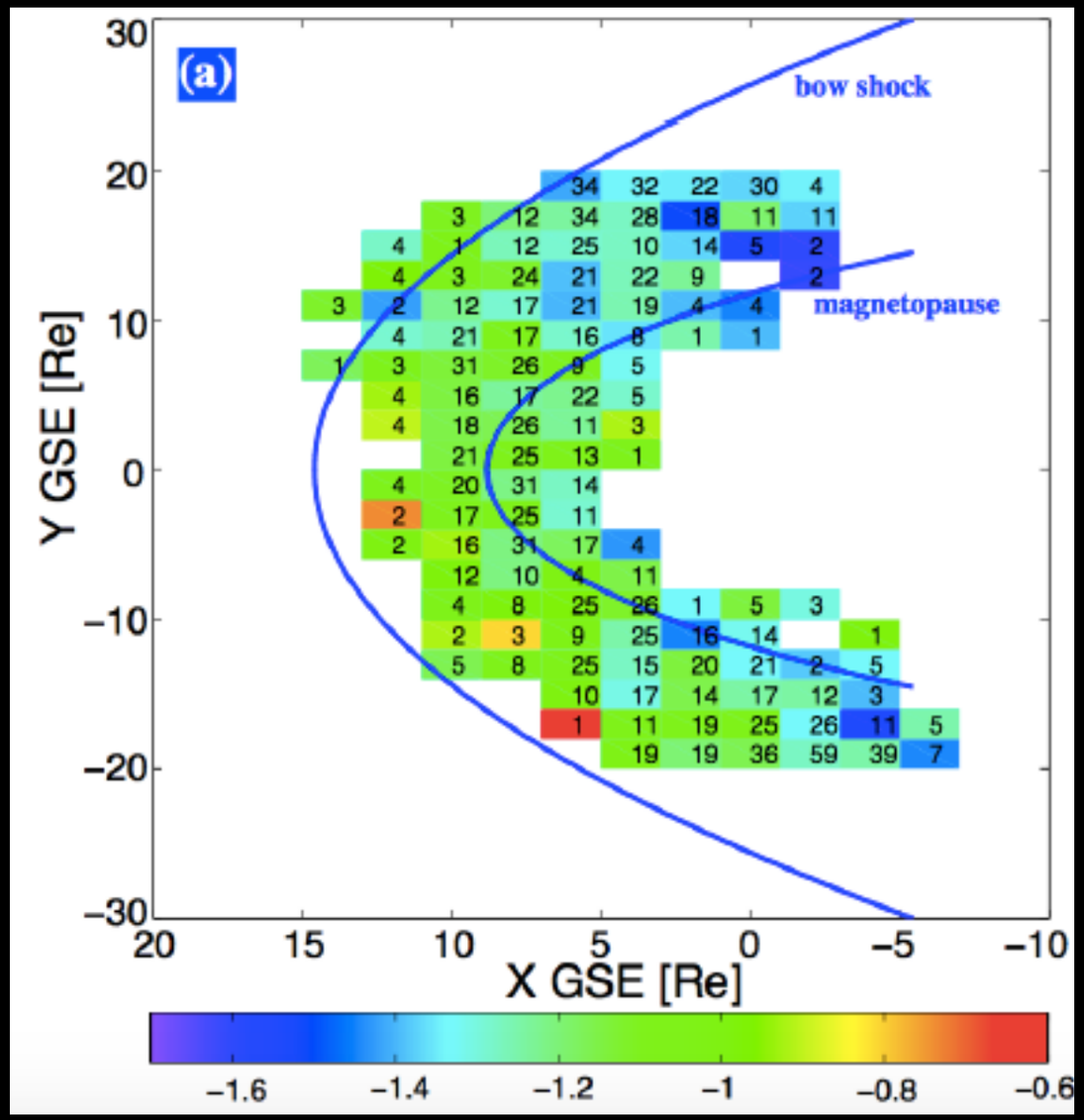
Massive amount of data available



3DView

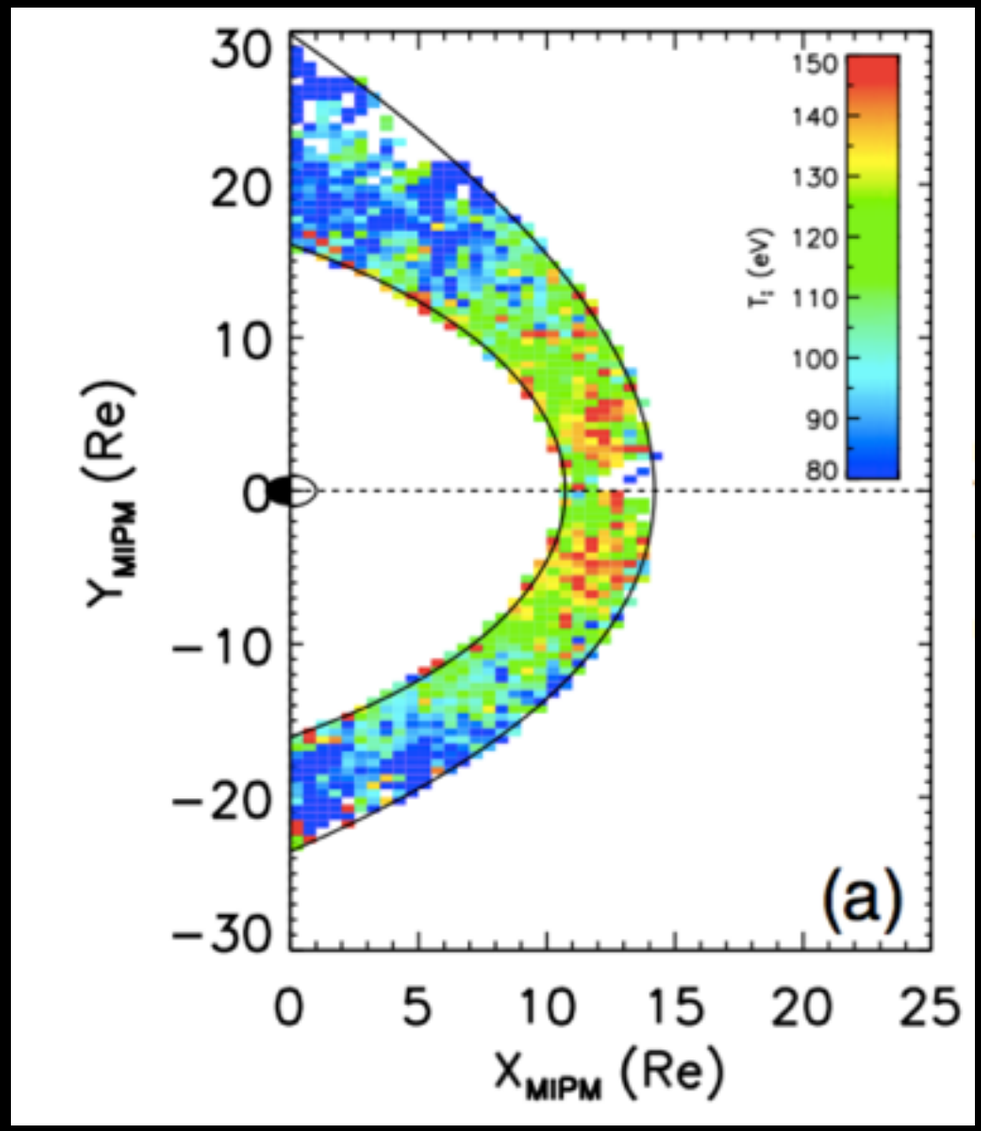
MAGNETOSHEATH : BOUNDARY CONDITION FOR RECONNECTION

3YRS CLUSTER DATA
1600 1.6HRS INTERVALS



Huang et al. 2016

7YRS THEMIS
BOUNDARY MODELS+ SAFE DISTANCE



Dimmock et al. 2015

VALIDITY OF BOUNDARY MODELS? (SHUE ETC.)

GET DATA CLOSE TO BOUNDARIES?

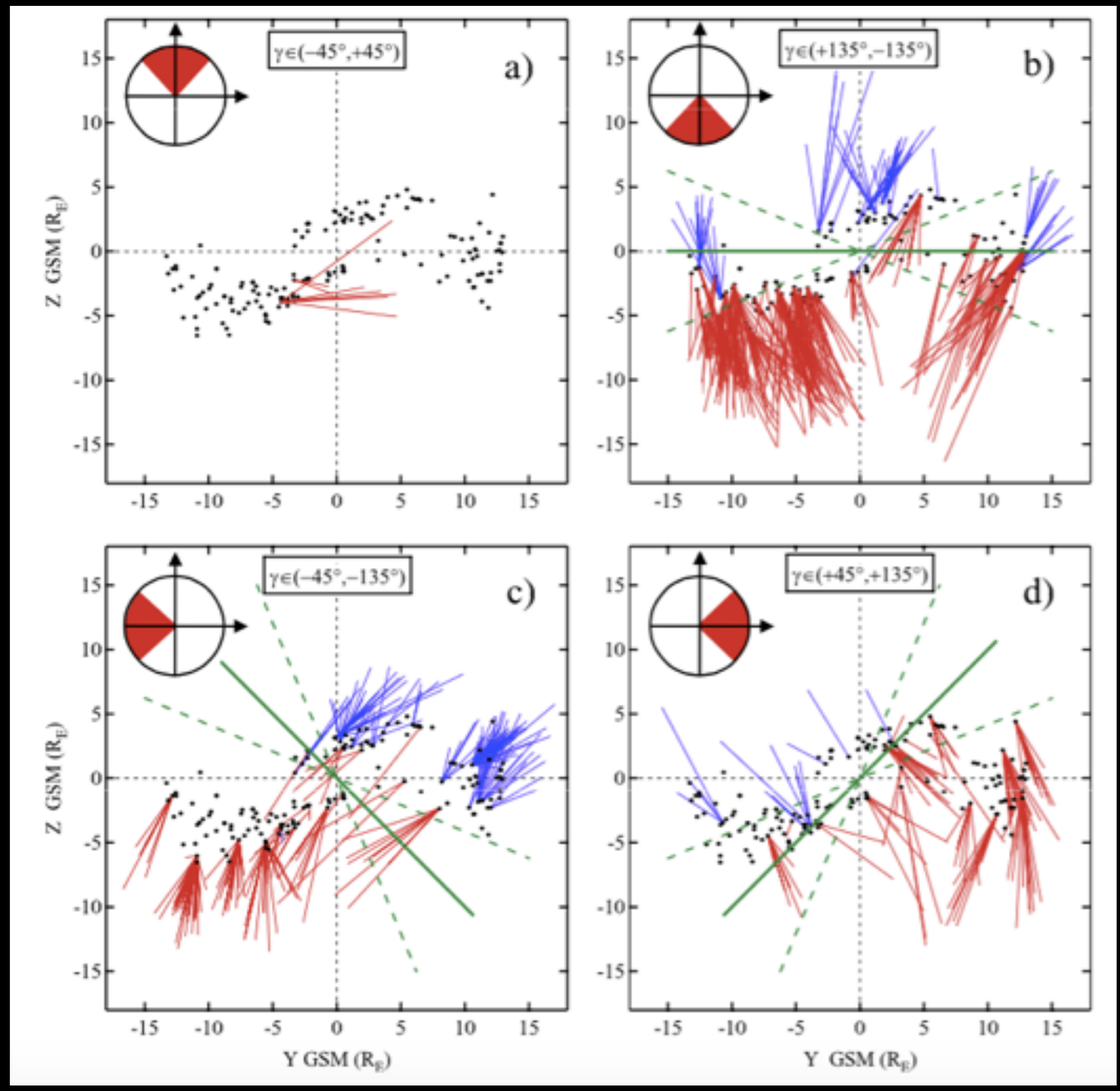
MAGNETOPAUSE MODEL? SHOCK MODEL? MAGNETOSHEATH FLOW, FIELD, ETC. MODELS?

WHY NOT MULTI-MISSION?

MAGNETOPAUSE RECONNECTION: WHERE? HOW?

[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°) 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-

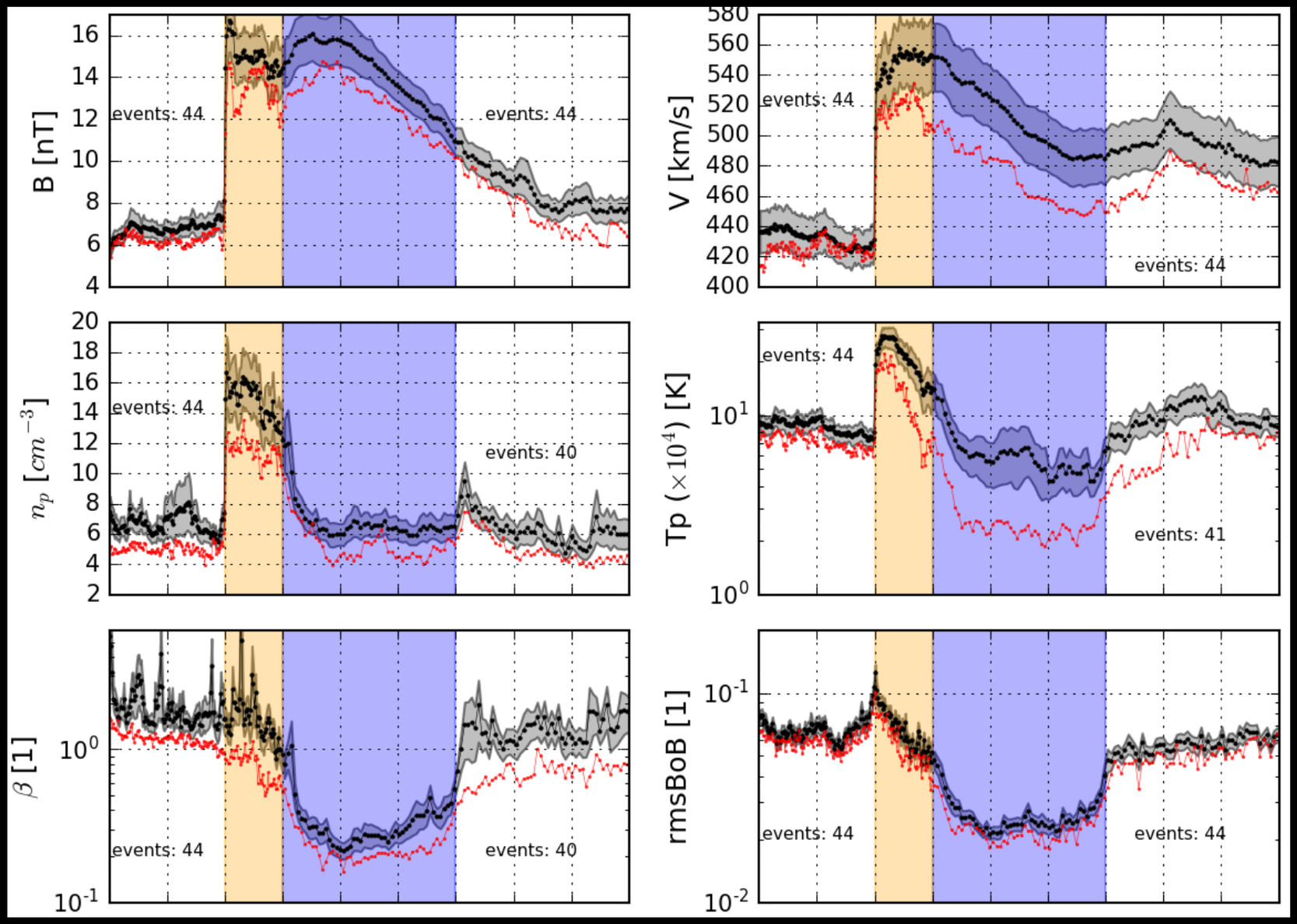
[4] Here, we use Double Star TC1 plasma and magnetic field data to study reconnection occurrence at the dayside low-latitude magnetopause. The orbit of the TC1 satellite, which is almost equatorial with an apogee of 12.4 R_E, is particularly suitable for this study. We analyzed 239 MP crossings, 143 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



Trenchi et al. 2008

patterns exist but statistics are hard because event selection is hard

ICME PROPERTIES, CLOUDS, SHEATH, FORESHOCKS?

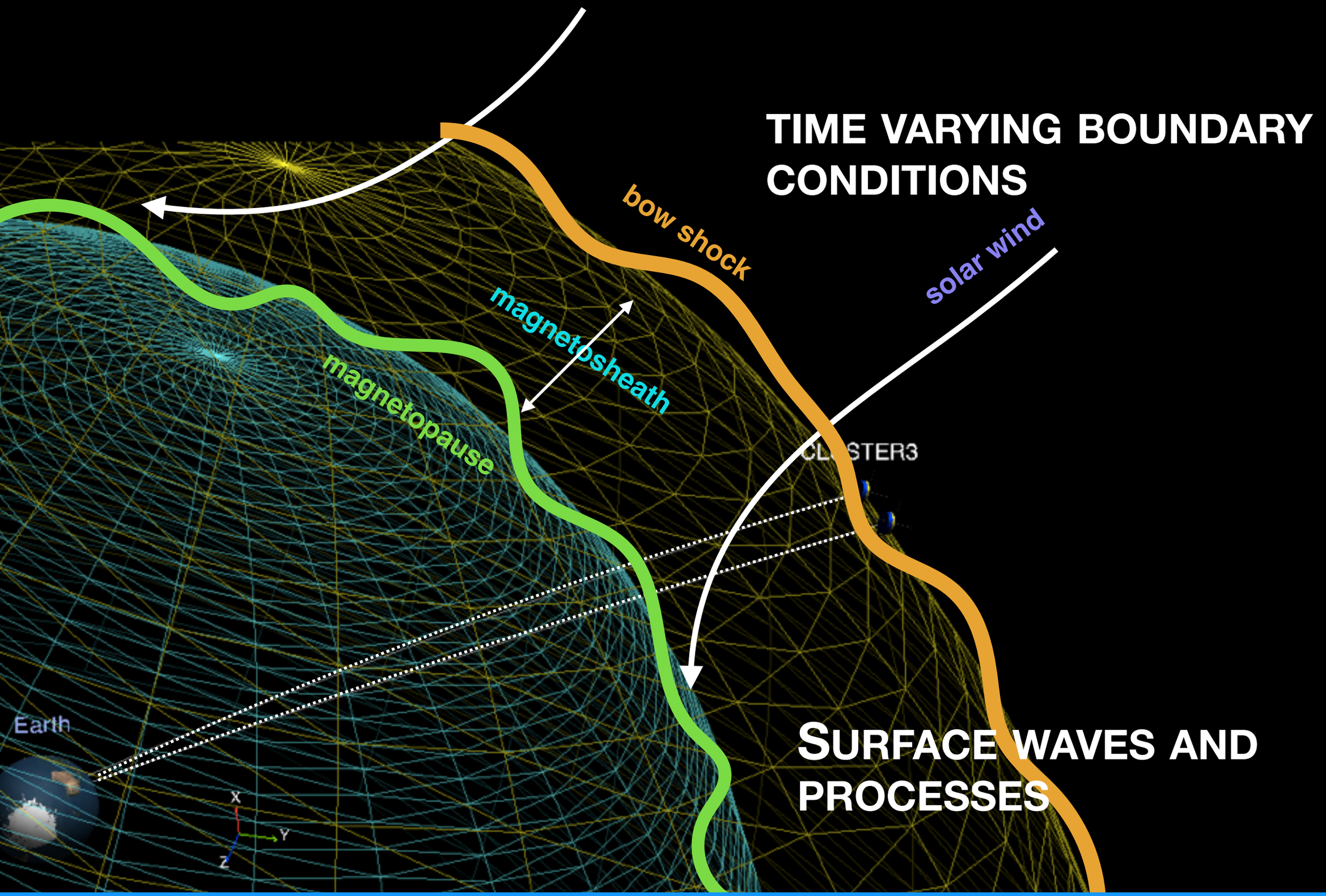


Masías-Meza, J.J. 2016

Reproducible way to identify key regions?

BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detecting thousands of « events » is hard

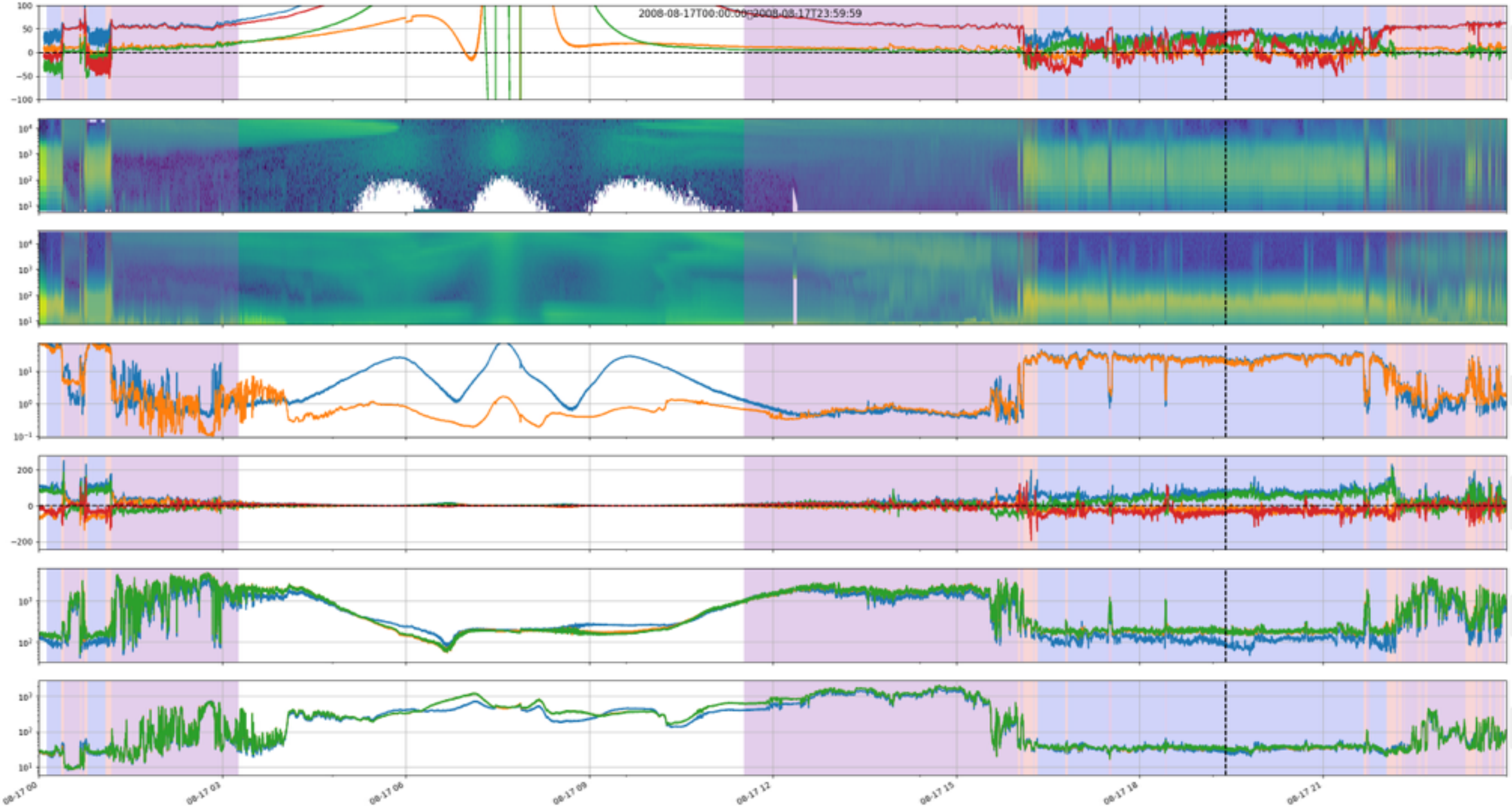


TIME VARYING BOUNDARY CONDITIONS

SURFACE WAVES AND PROCESSES

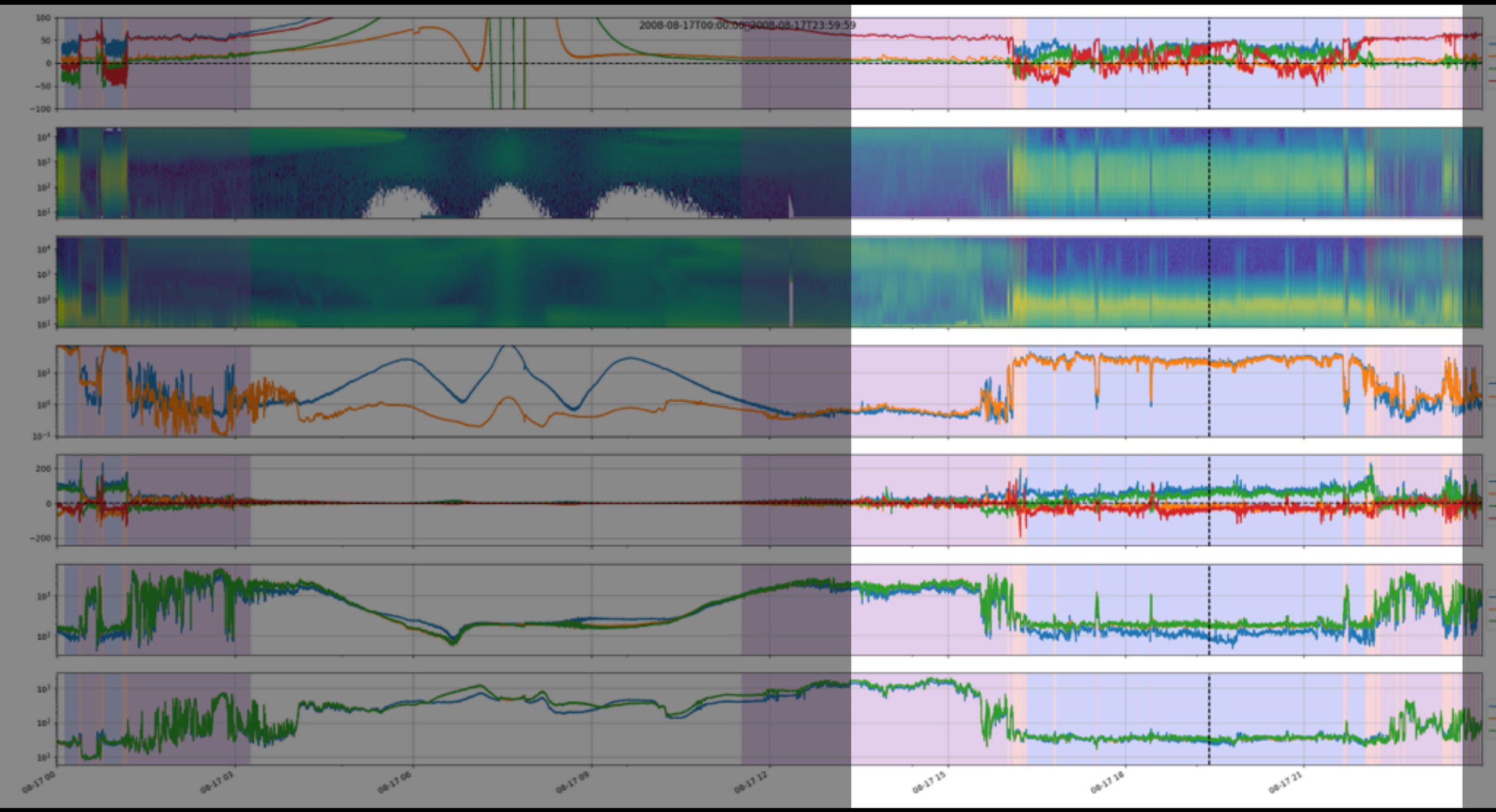
BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detecting thousands of « events » is hard



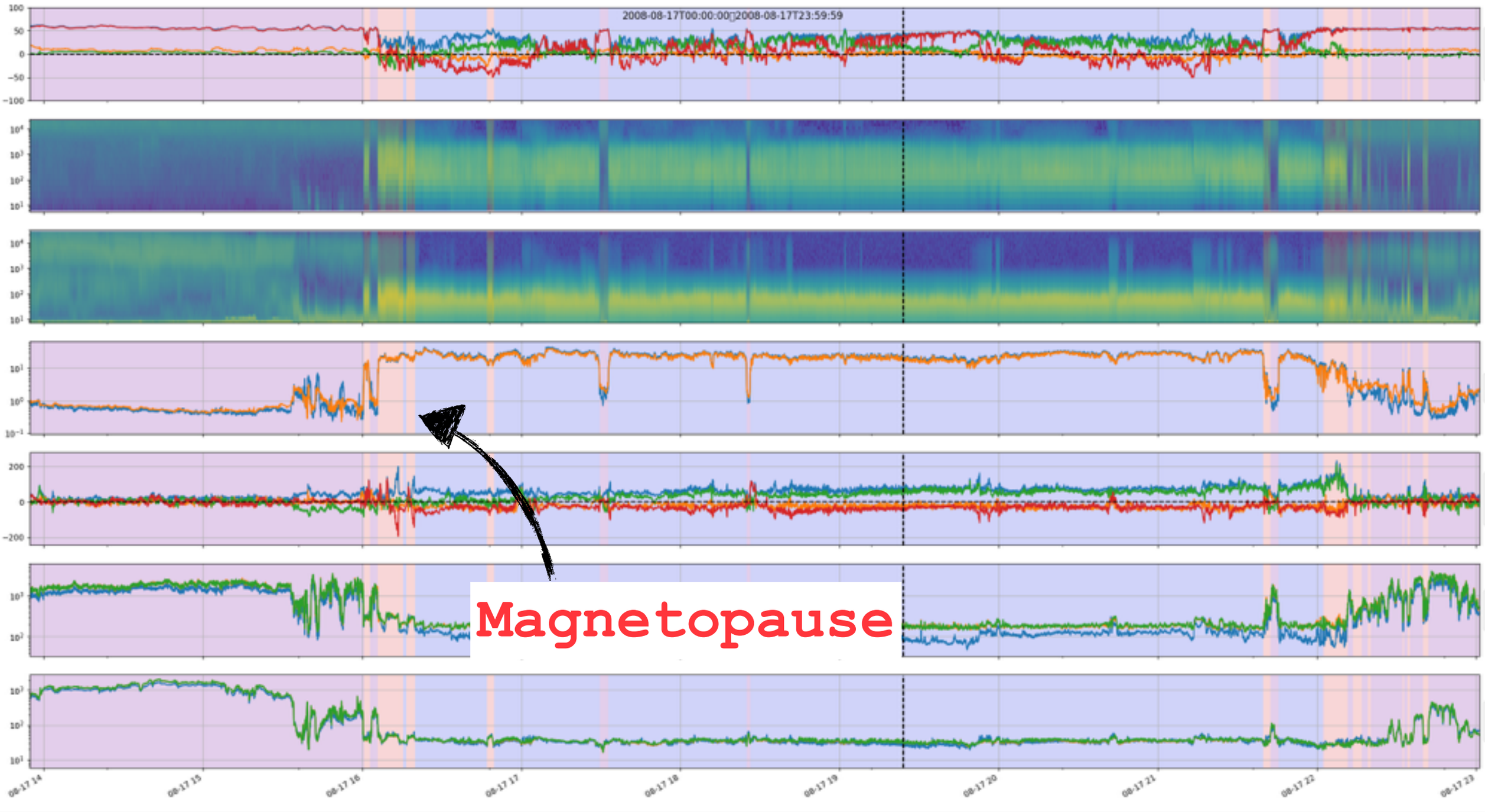
BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

They're all defined at the **global scale** but can only be answered by looking at **small scales**.



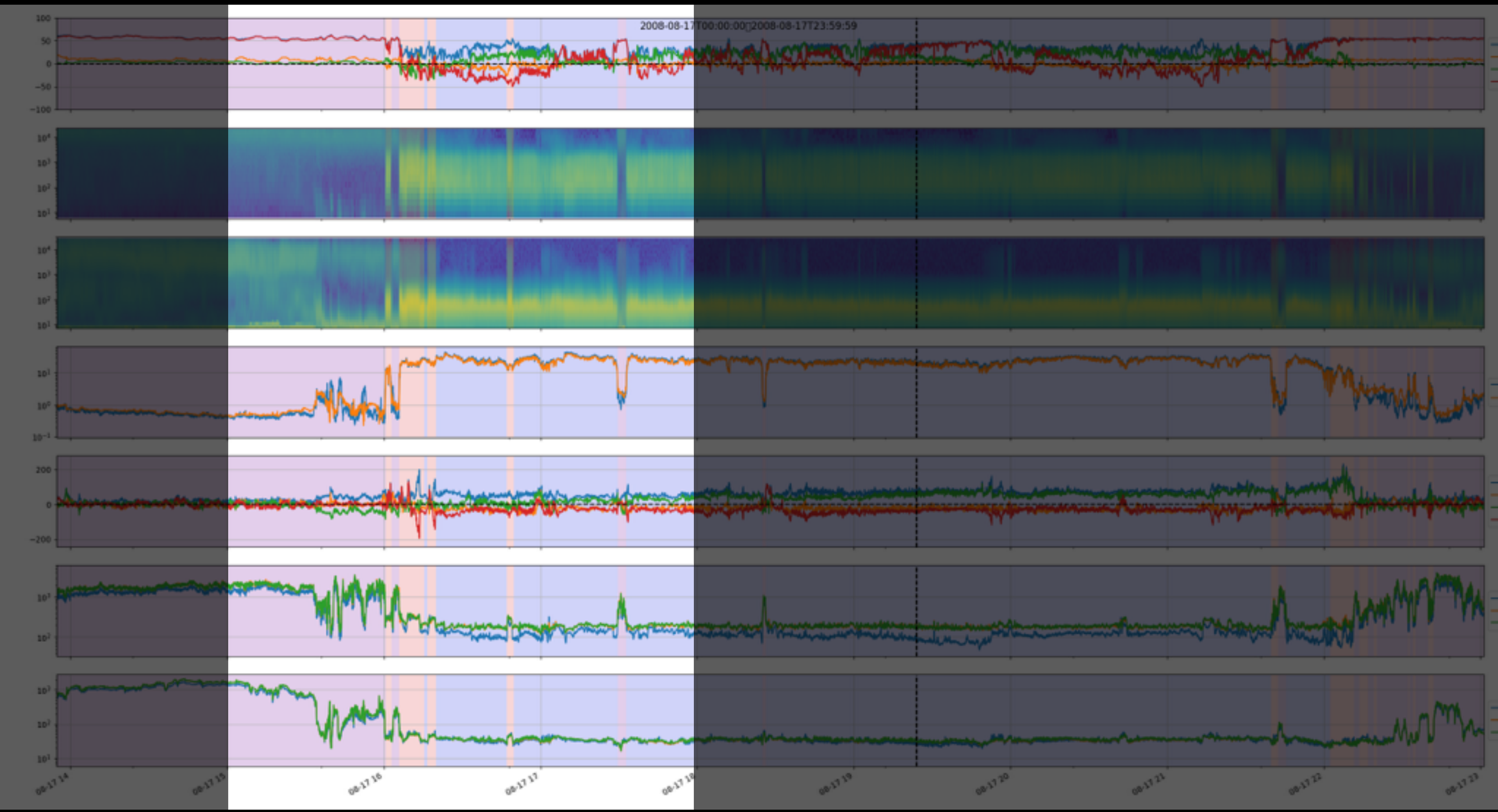
BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detect thousands of « events » is stupidly hard



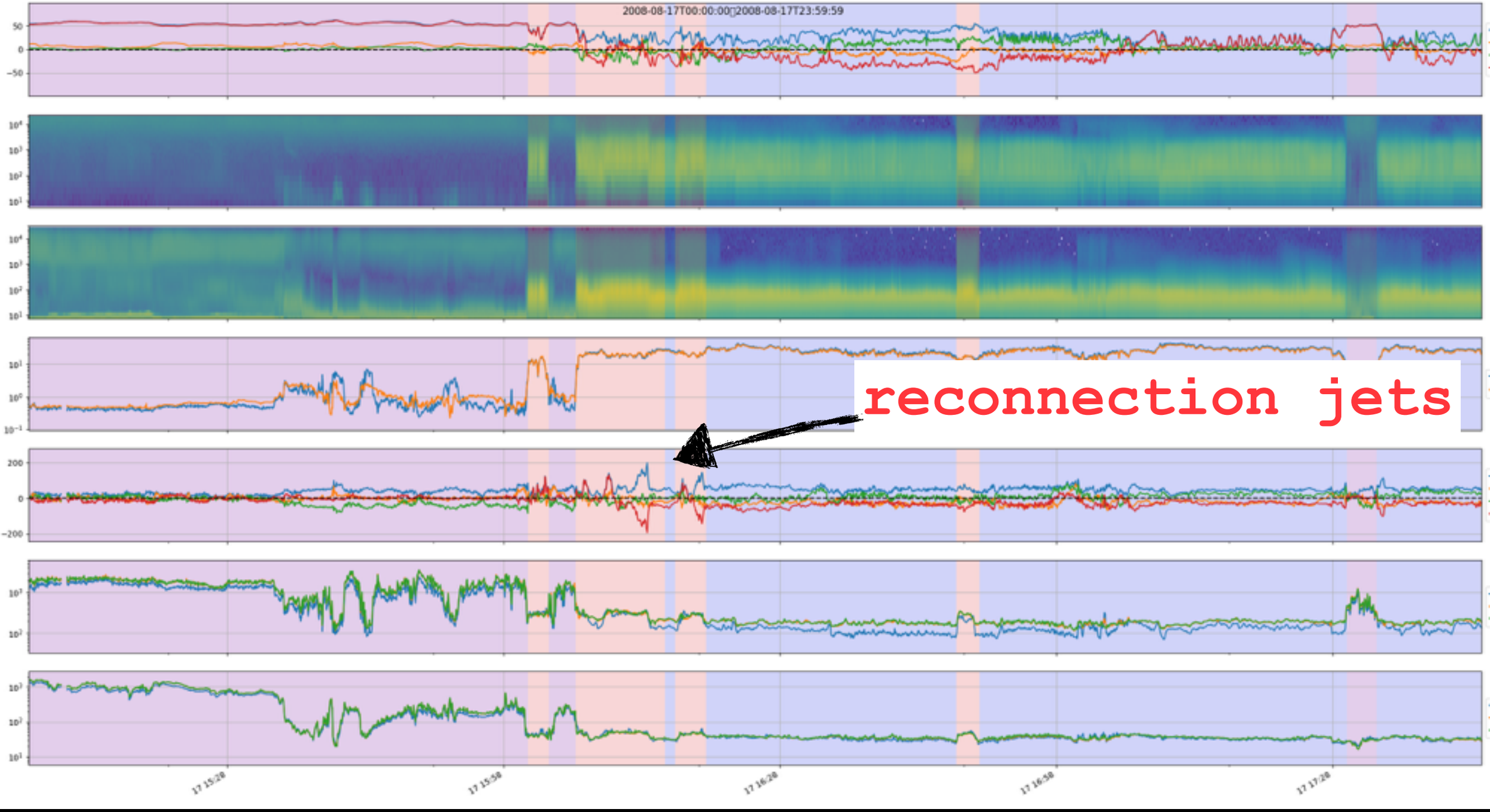
BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detect thousands of « events » is stupidly hard



BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detect thousands of « events » is stupidly hard



BIG QUESTIONS MAY BE HARD FOR WRONG REASONS

detect thousands of « events » is stupidly hard

VARIABILITY IS HARD

- Detecting signatures is difficult
- Strong *apparent* variability of signatures
- Strong variability of the environment

VISUAL DETECTION DOESN'T SCALE

- **Massive** statistical studies of processes hard mainly because data selection is hard
- Months of tedious data selection
- Selection criteria shift in time, depend on people
- Event lists are rarely shared, not associated with selection criteria… poor reproducibility

ANALYTICAL MODELS ARE OFTEN TOO SIMPLE

- Shue, etc. only for operational usage (no physics) but often too simplistic (e.g. axisymmetric etc.)

Automatic feature detection and data driven models

It is possible to have 10 to 30 years of multi-mission catalogs of:

- Magnetopause/shock crossings correlated with solar wind
- Magnetopause reconnection events detected in situ
- Kelvin Helmholtz vortices at the magnetopause
- magnetotail reconnection events
- magnetosheath intervals
- shock/sheaths/cloud intervals
- mirror, firehose, etc. instabilities signatures
- specific waves signatures
- etc.

automatically detected with reproducible criteria and sharable

thousands of intervals to give us a statistically relevant viewpoint of space plasma processes

change the way we analyse in situ data

MACHINE LEARNING

- **Supervised: learned from labeled data**
- Unsupervised: discover features in unlabeled data

Useful to make predictions P based on data D for which the function P(D) is not known.

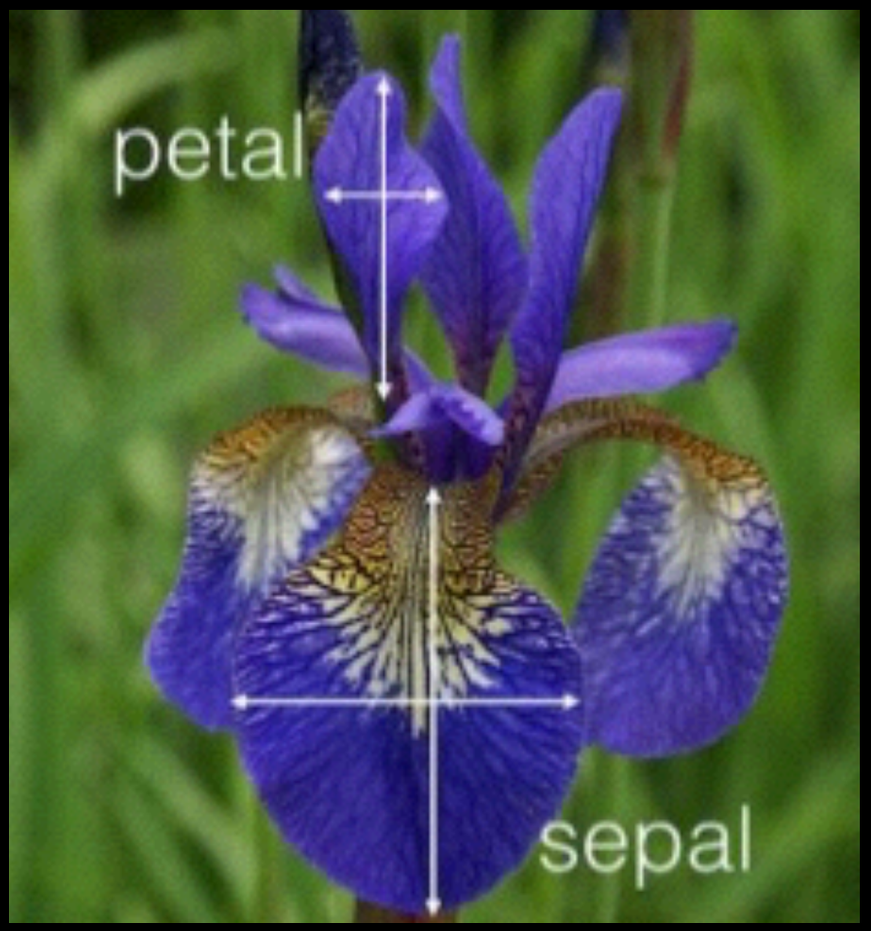
*« A computer program is said to **learn from experience** E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E »* [12]

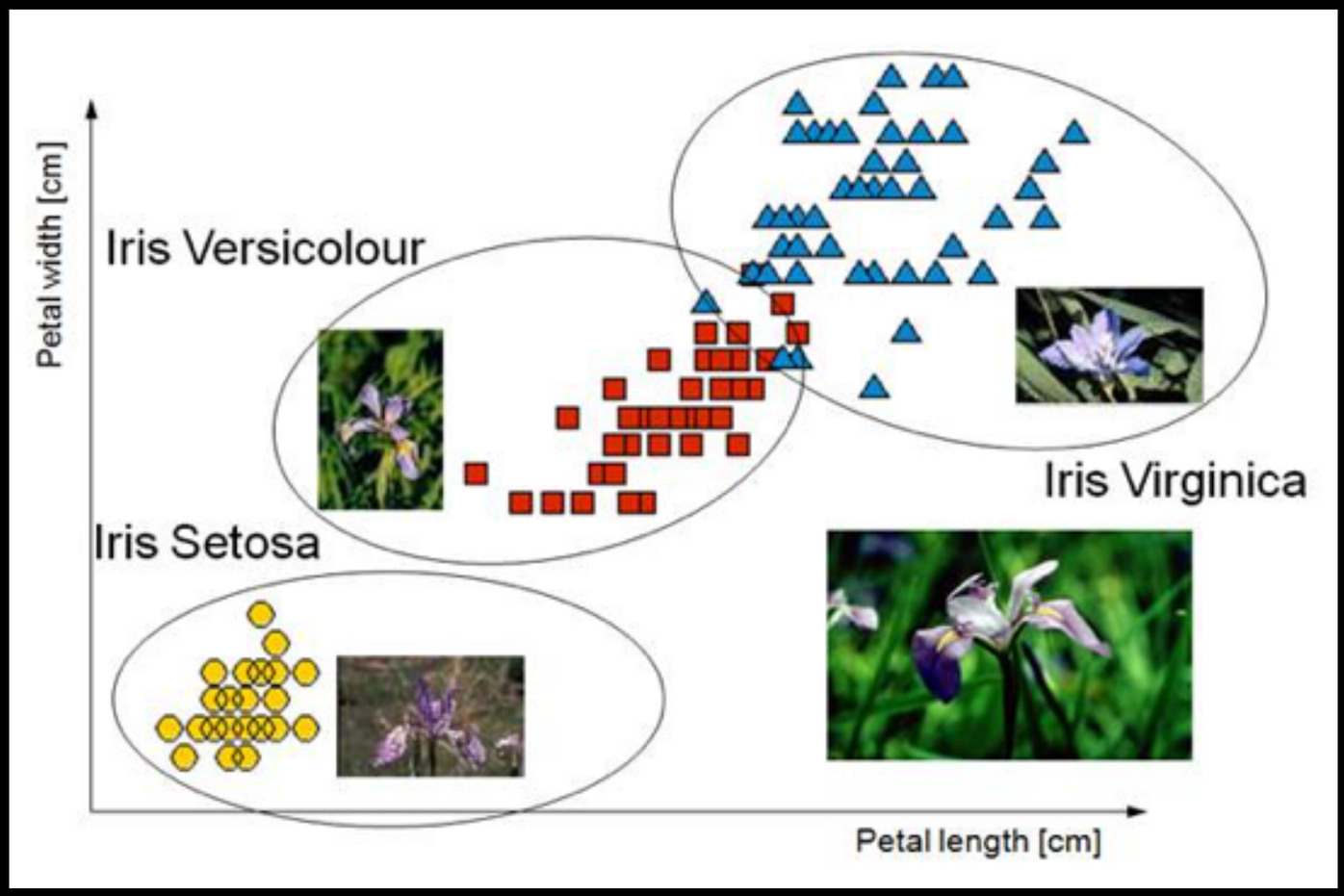
T. M. Mitchell

MACHINE LEARNING

- **Supervised: learned from labeled data**
- Unsupervised: discover features in unlabeled data

- **Classification problem: learn discrete classes**



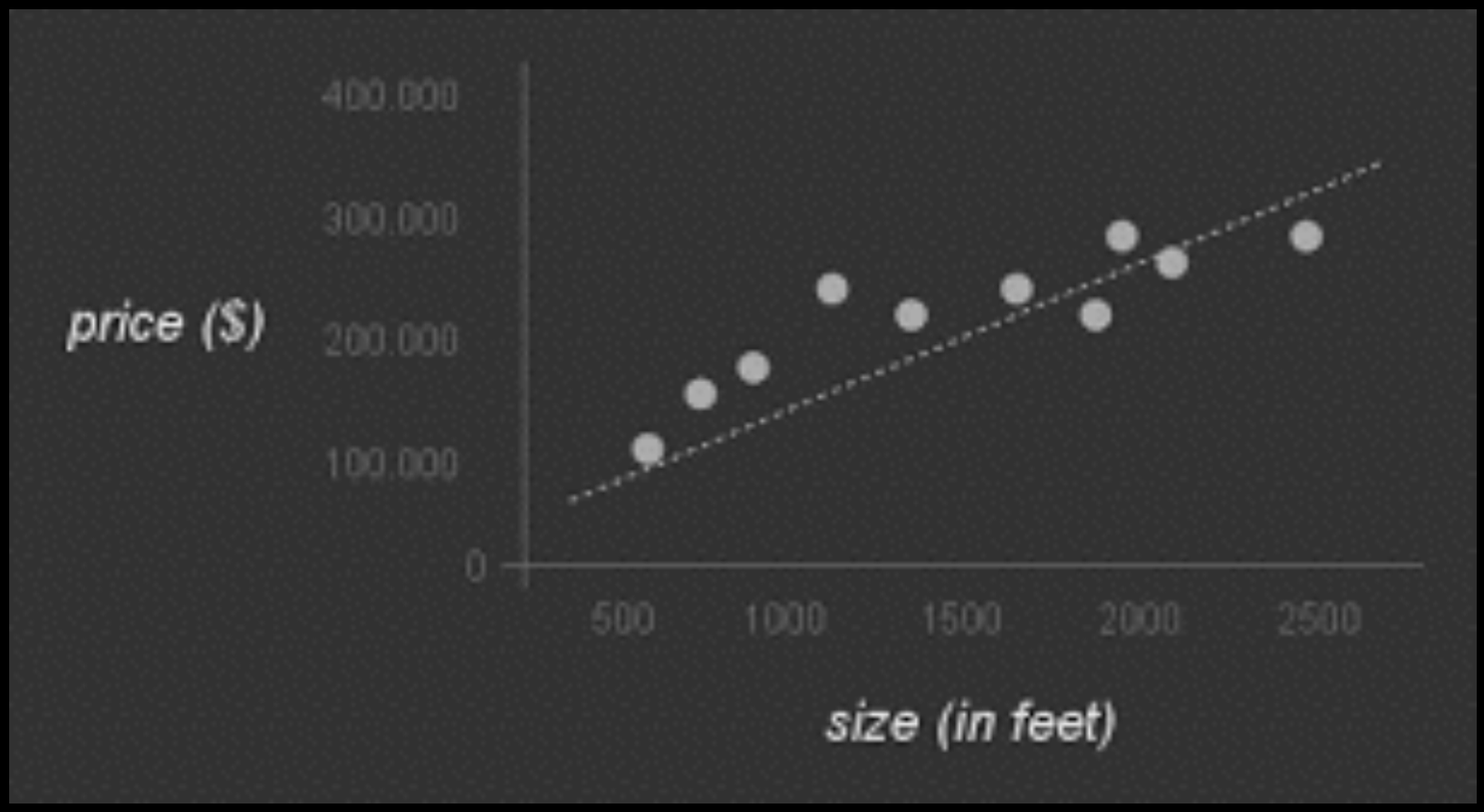


used to predict the class of unknown objects

MACHINE LEARNING

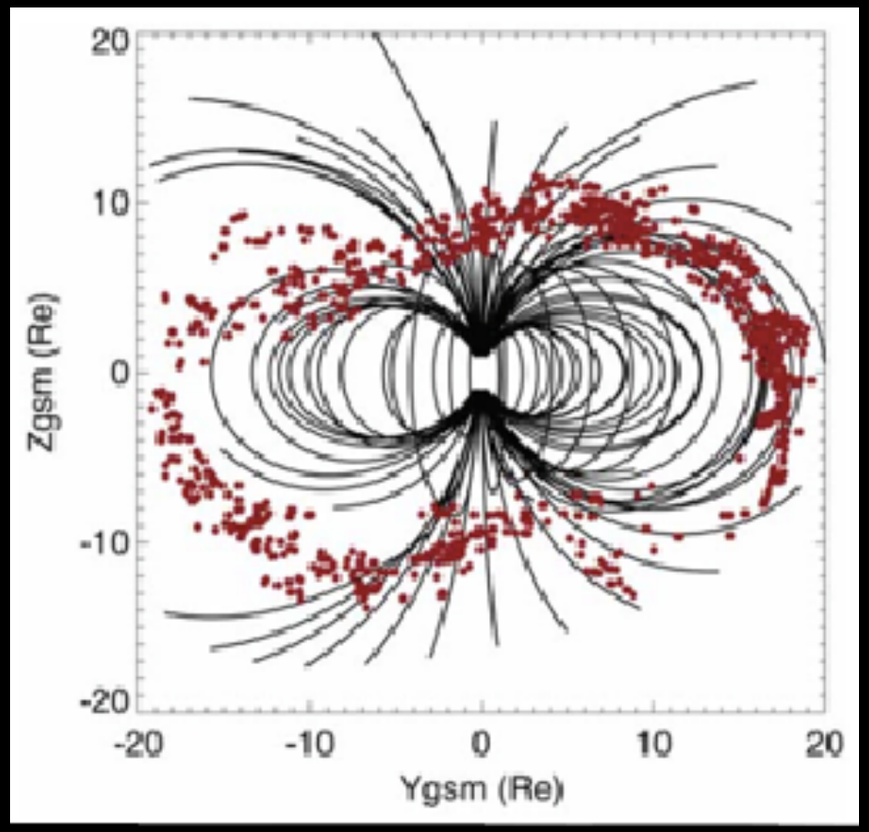
- **Supervised: learned from labeled data**
- Unsupervised: discover features in unlabeled data

  - **regression problem: learn *continuous* variables**

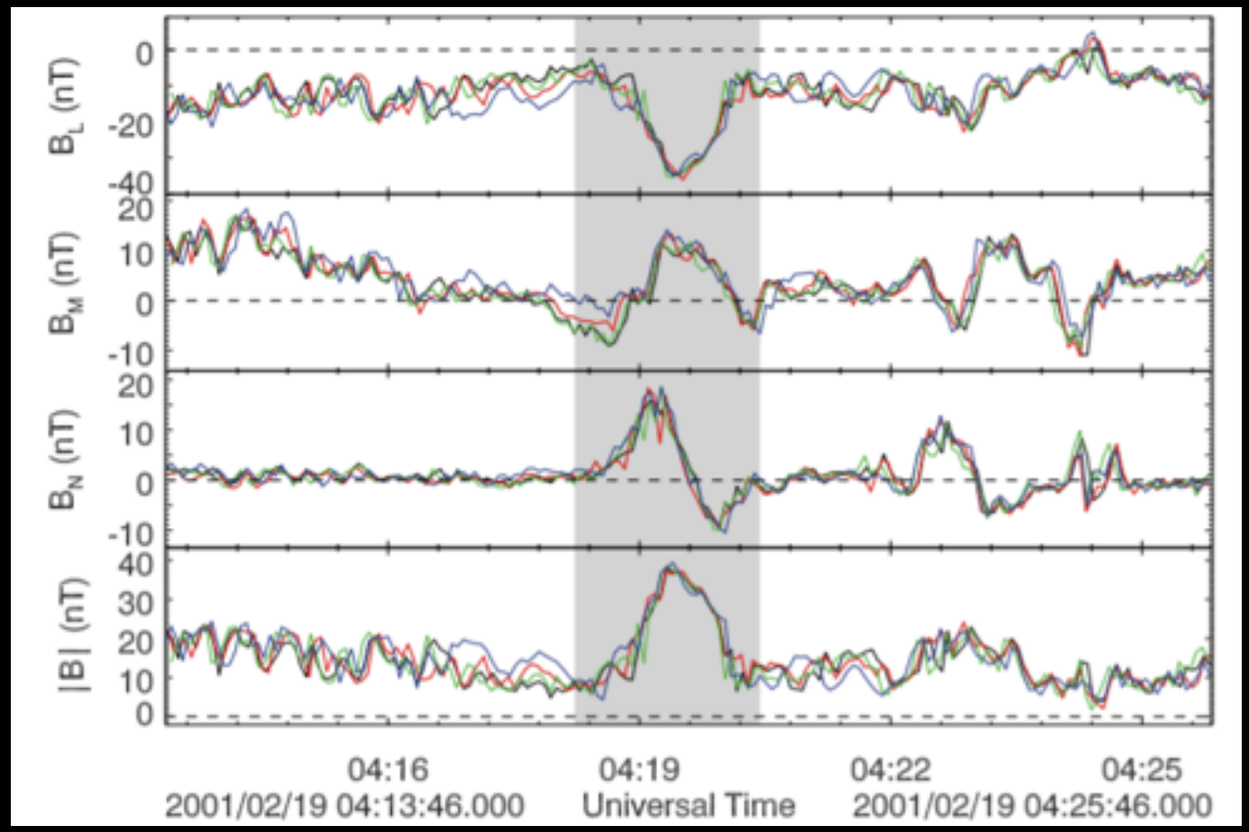


used to predict value of unknown/complex functions

SIGNATURE DETECTION



Karimabadi et al. 2009



- >1000 FTEs
- Discovery of FTEs
- Classification MSH/MSP FTEs
- **[95-99]% success rate**

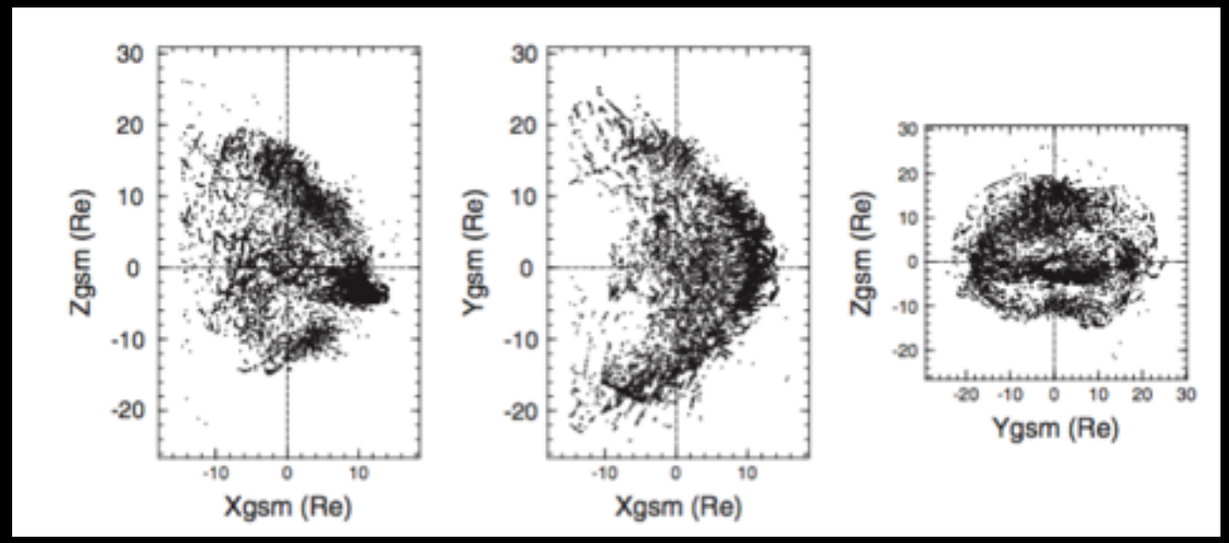
- Magnetopause, shock, ICMEs, sheath, etc.

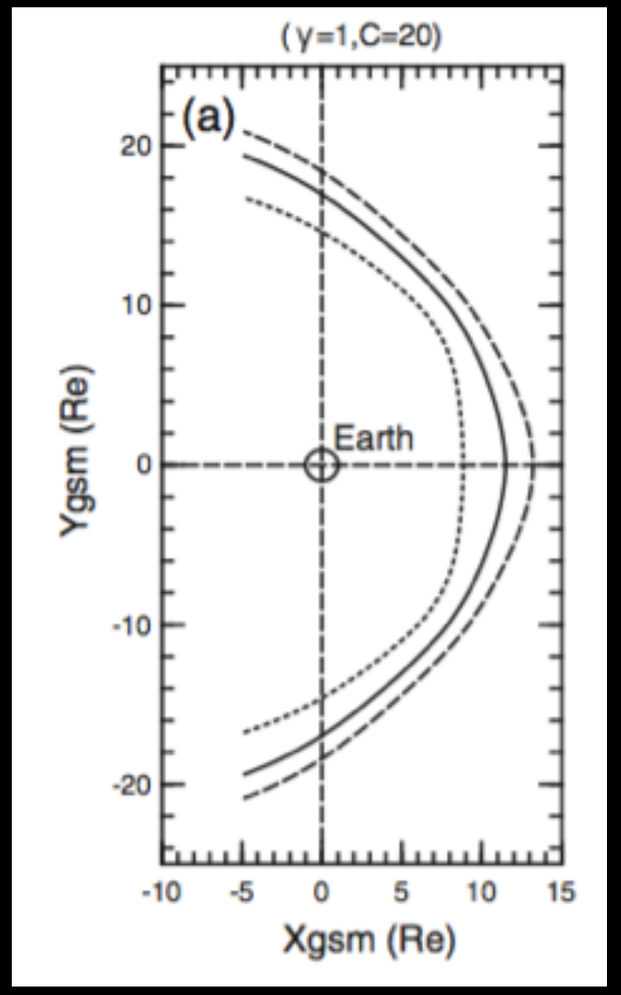
DATA DRIVEN MODELS

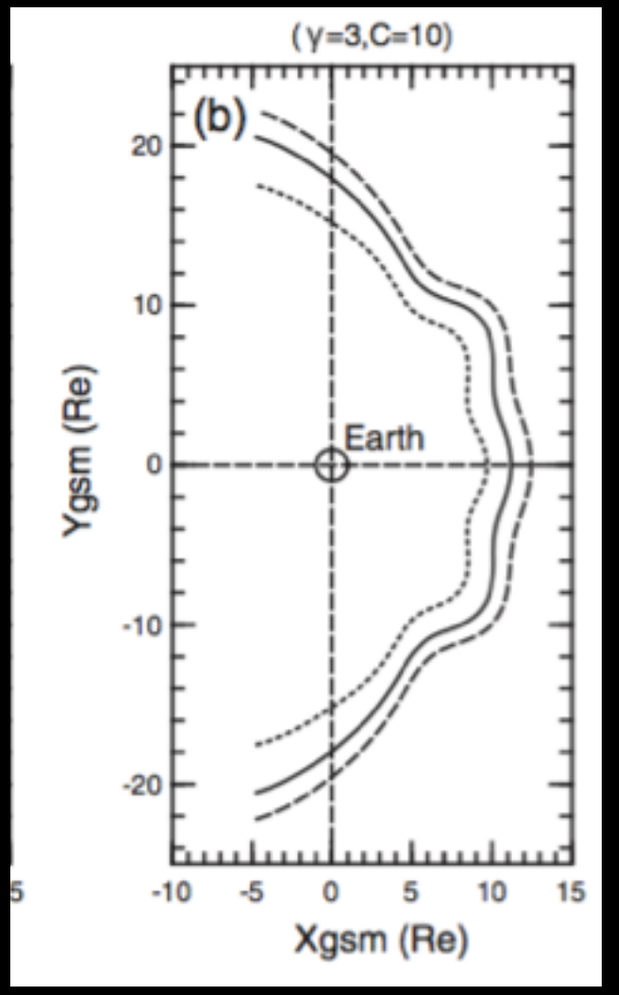
- ~15000 magnetopause crossings
- Solar wind monitoring
- Non-parametric support vector regression

- **Use data to infer the function :**
  - **Magnetopause(x,y,z ; SW)**

- Use more data
- Shock model?
- Magnetosheath model?
- Tail model, etc. etc.







Wang et al. 2013

SCIENCE ROADMAP

Now

- Machine Learning for in situ data
- ICME detection
- Magnetopause detection
- 3D data driven shock, sheath, magnetopause models
- Magnetopause reconnection and flow patterns models
- …

FORESEEABLE FUTURE

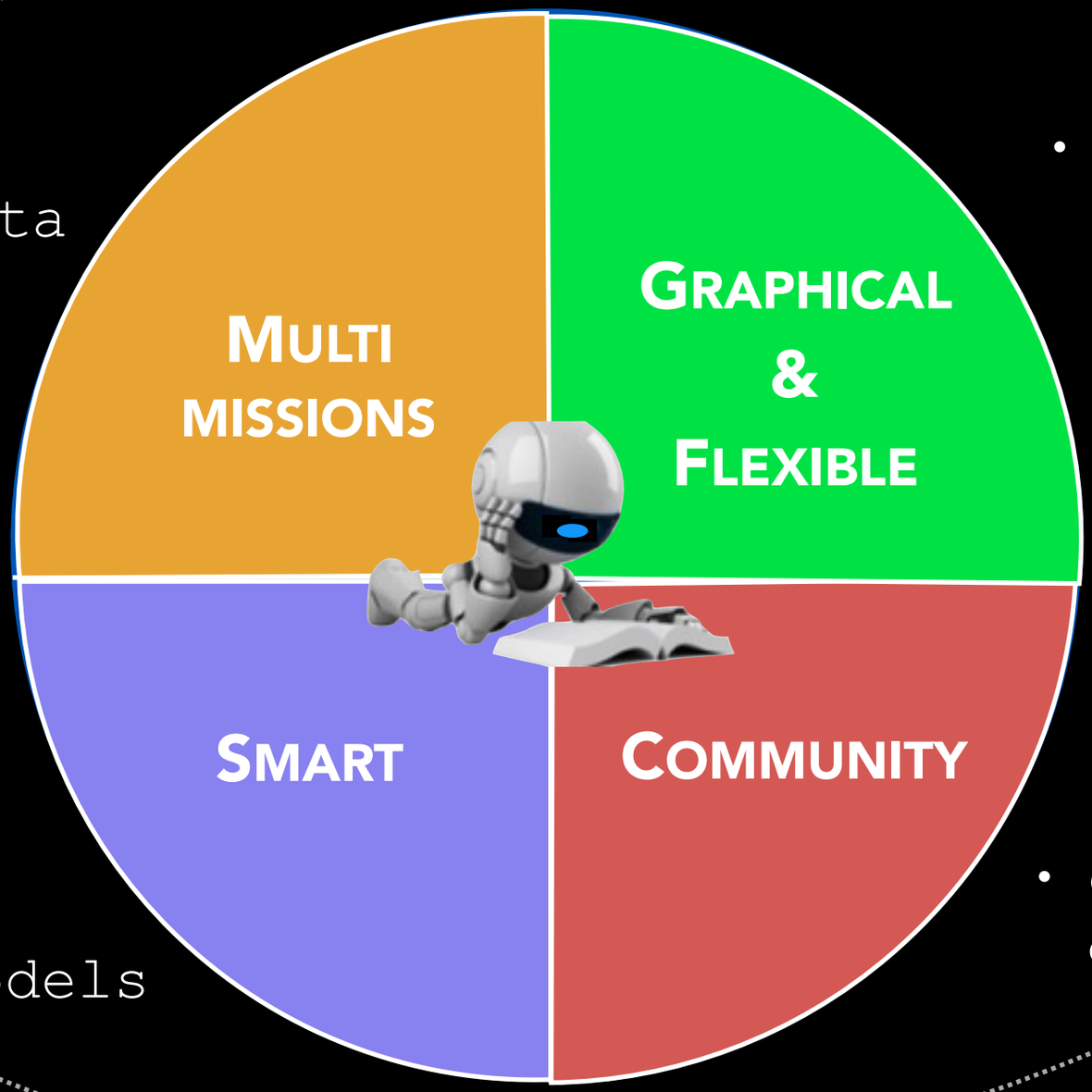
- Machine learning feature discoveries
- thousands of public collaborative catalogs
- …
- Not a single minute of data not labelled

SciQLOP: 4 main features

reduce the amount of engineering code, increase performance & interactivity

- Transparent download of data

- Efficient data browsing



- Smart detection
- data driven models

- collaborative event catalogs

Reduce detection bottlenecks and improve reproducibility

SciQLOP: HISTORY & FUNDING

2015

- In house prototype developed on free-time

2016 - 2017

- GUI : 50k€ - 1yr CDD
- ML : intern students

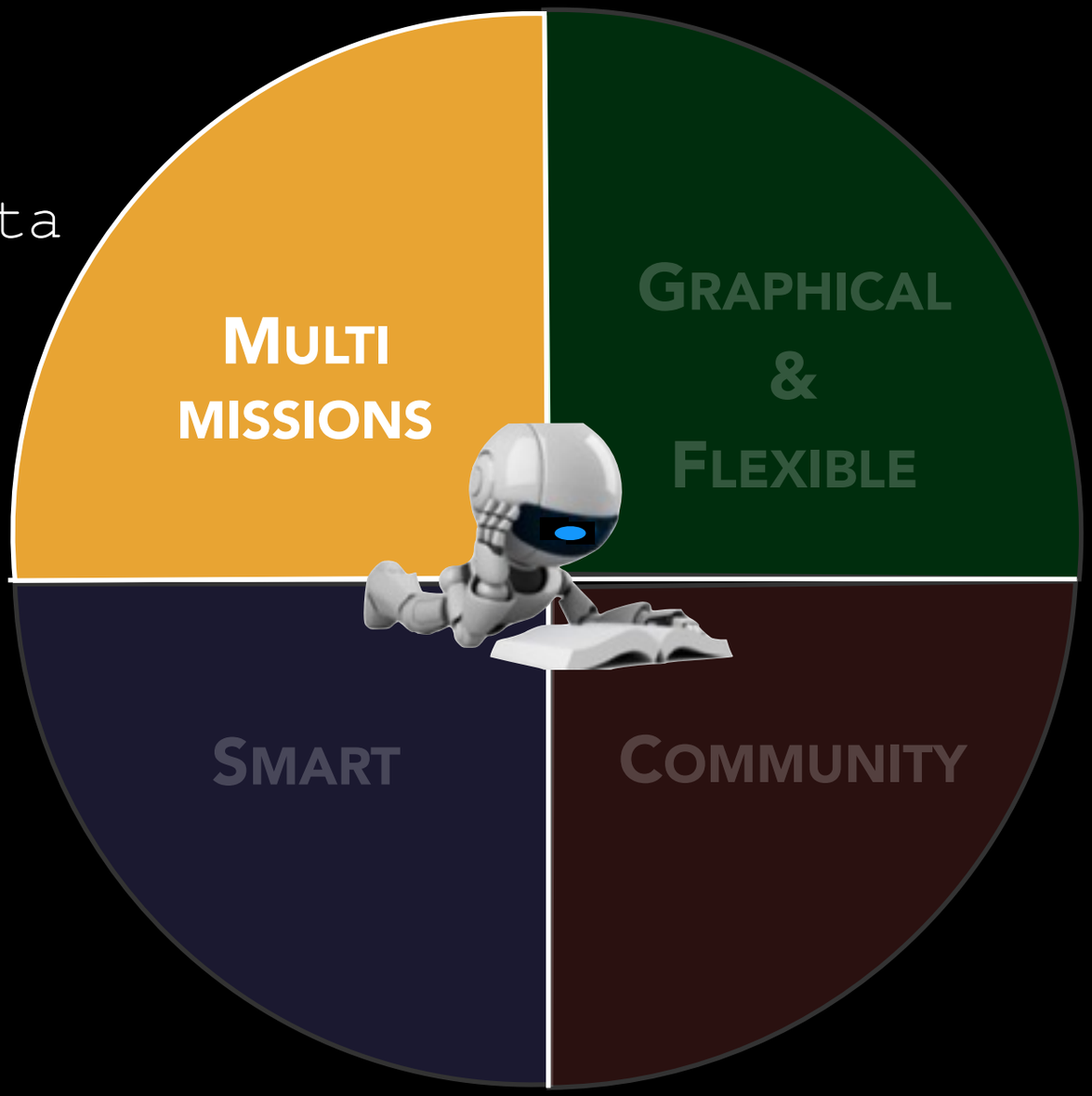
- 150k€ from WP1 SPACEOBS
- GUI: IPSIS
- ML : thesis + CDS

2018?

- **Need funding**.....

SciQLOP: 4 main features

- Transparent download of data



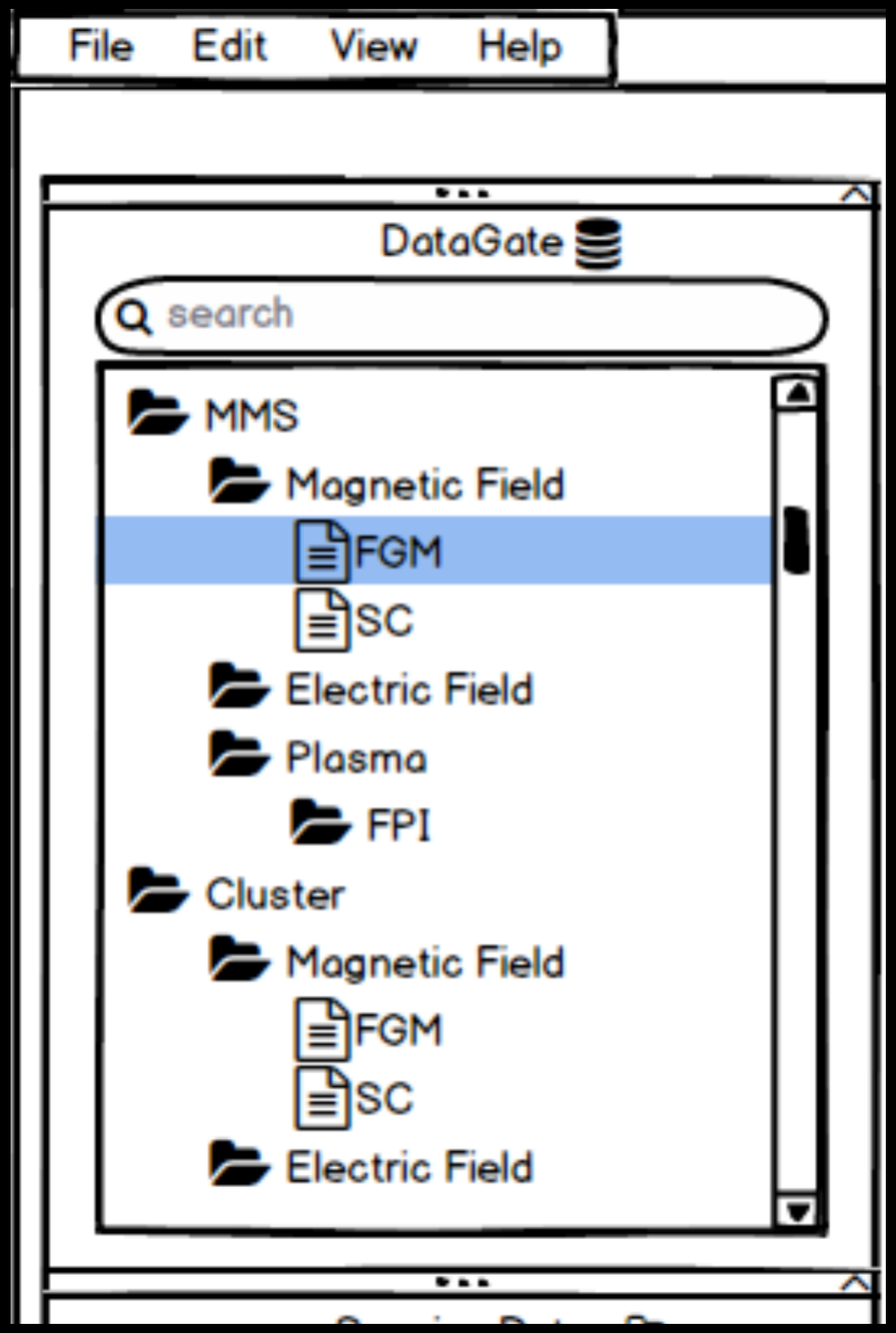
why limiting studies to only ONE mission??

SciQLOP: a Multi-mission tool

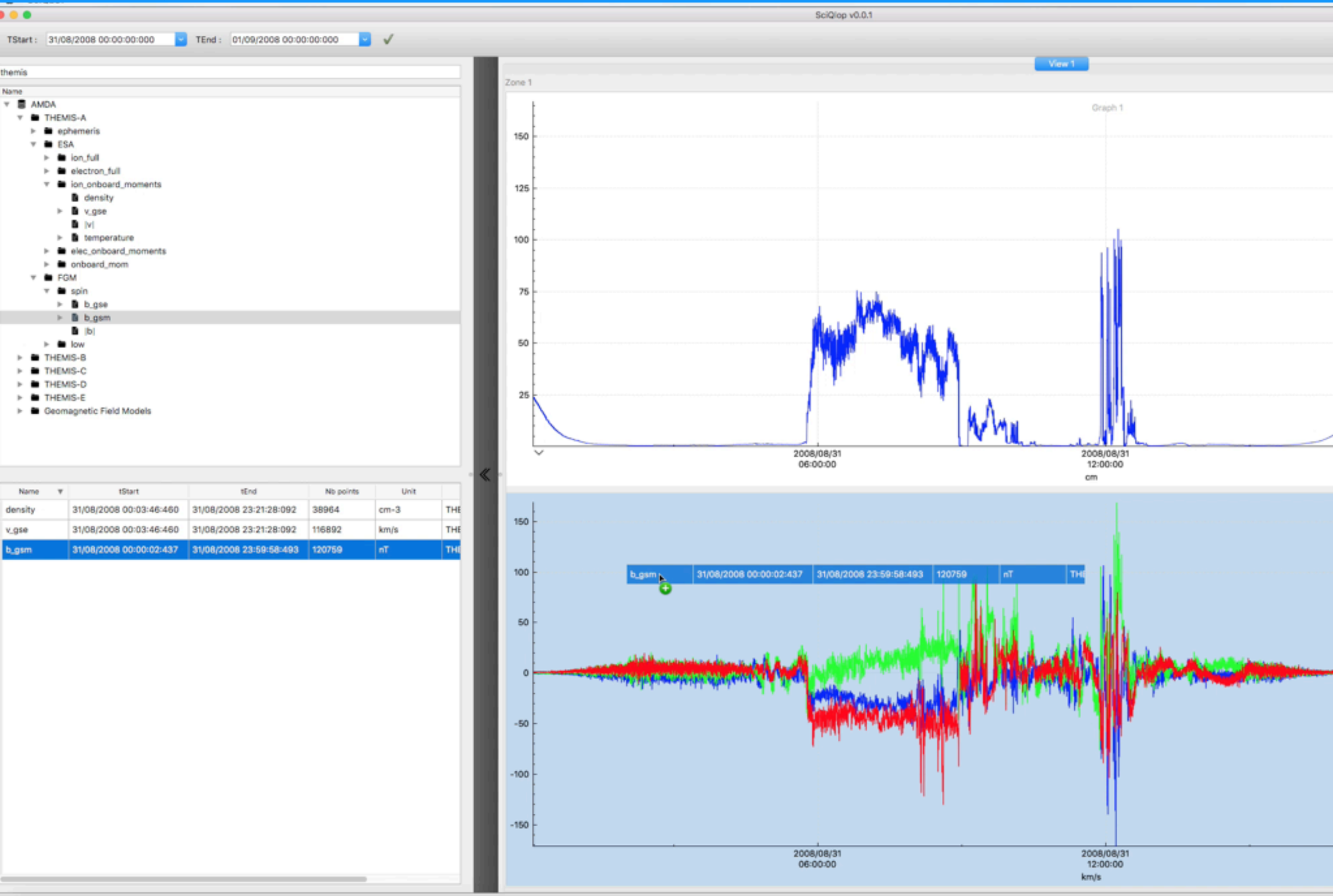
Amda

NASA

- REST, SOAP protocols

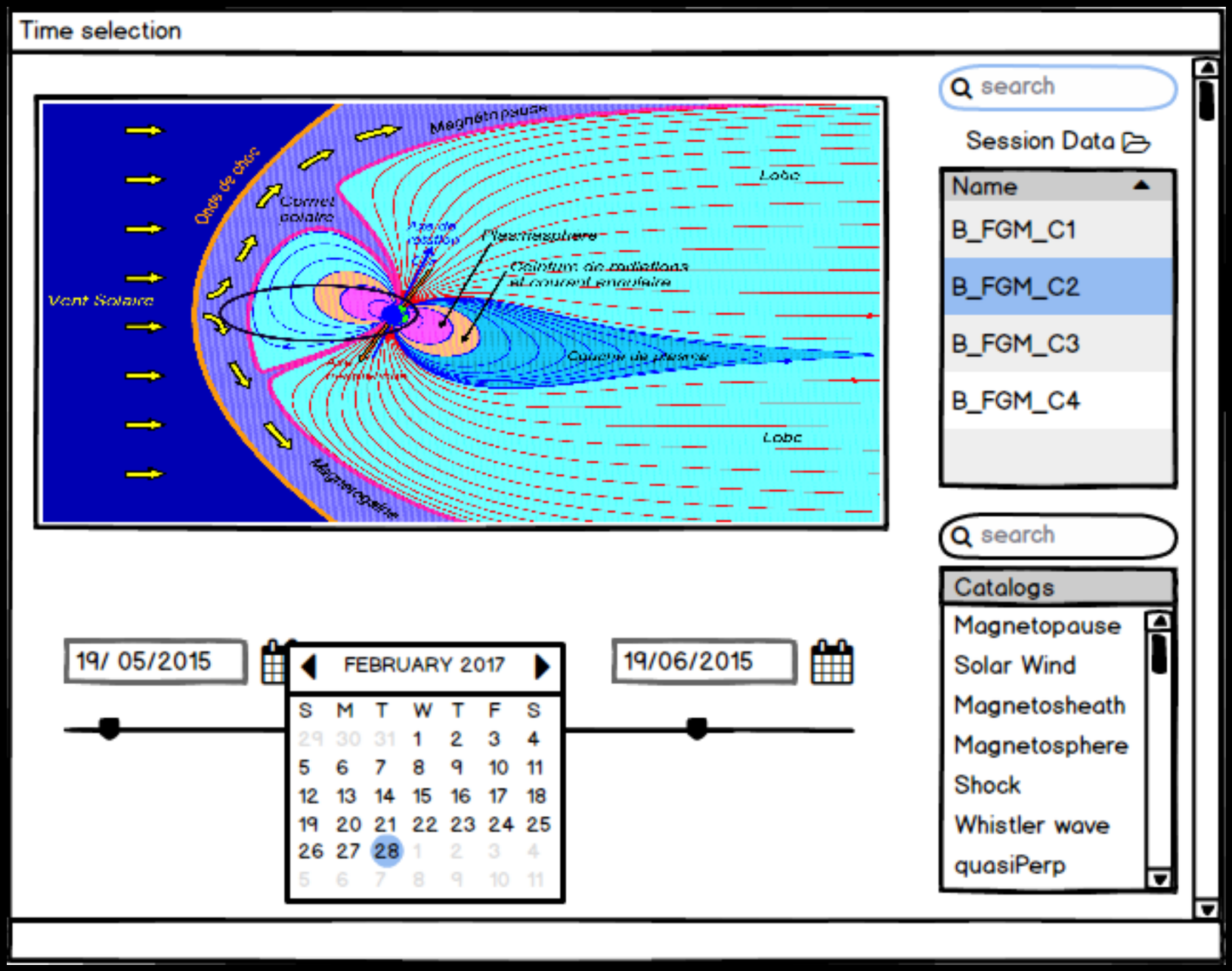


SciQLOP: LOADING DATA

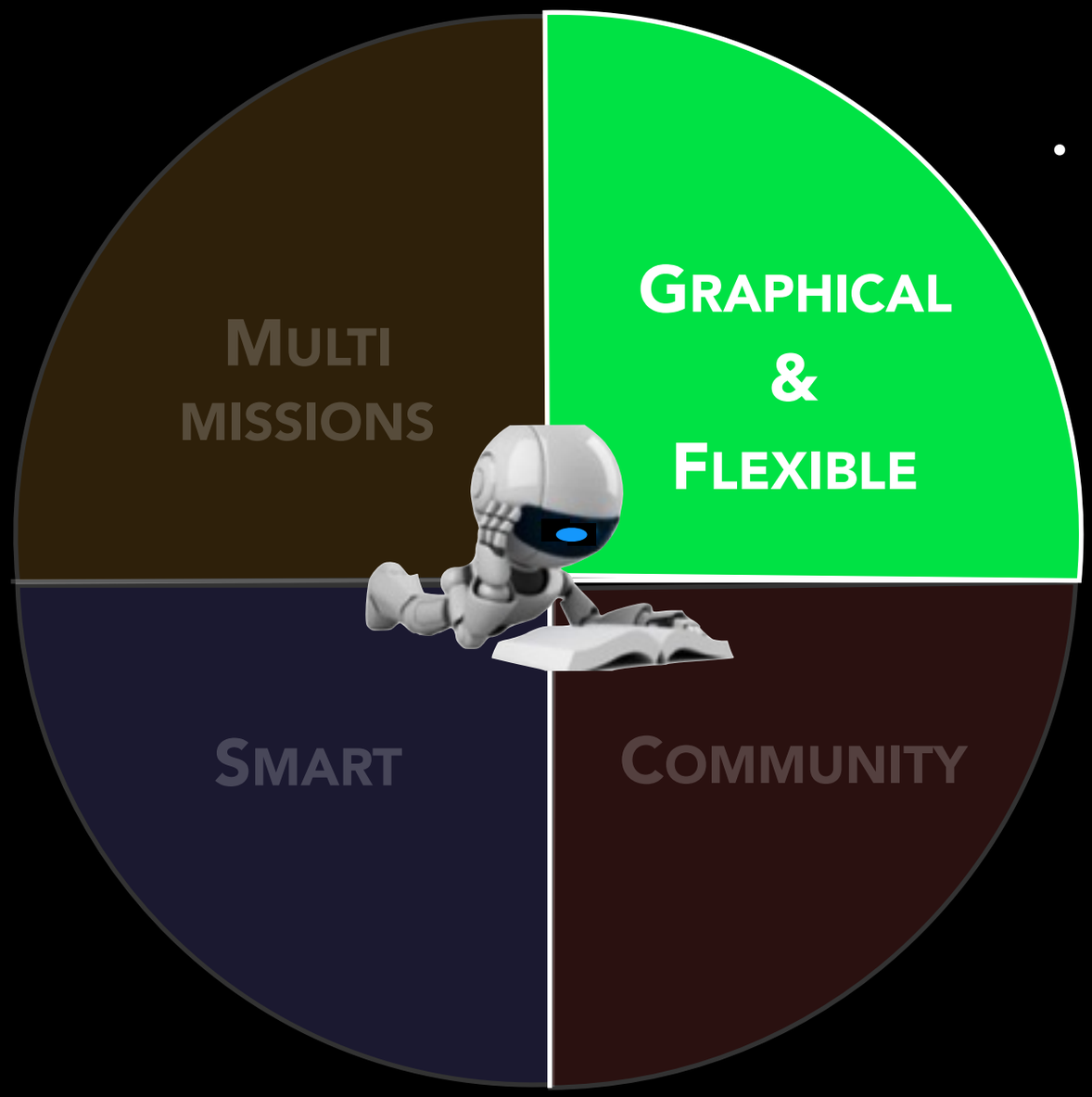


SciQLOP: TIME SELECTION

Several ways to adjust the time interval



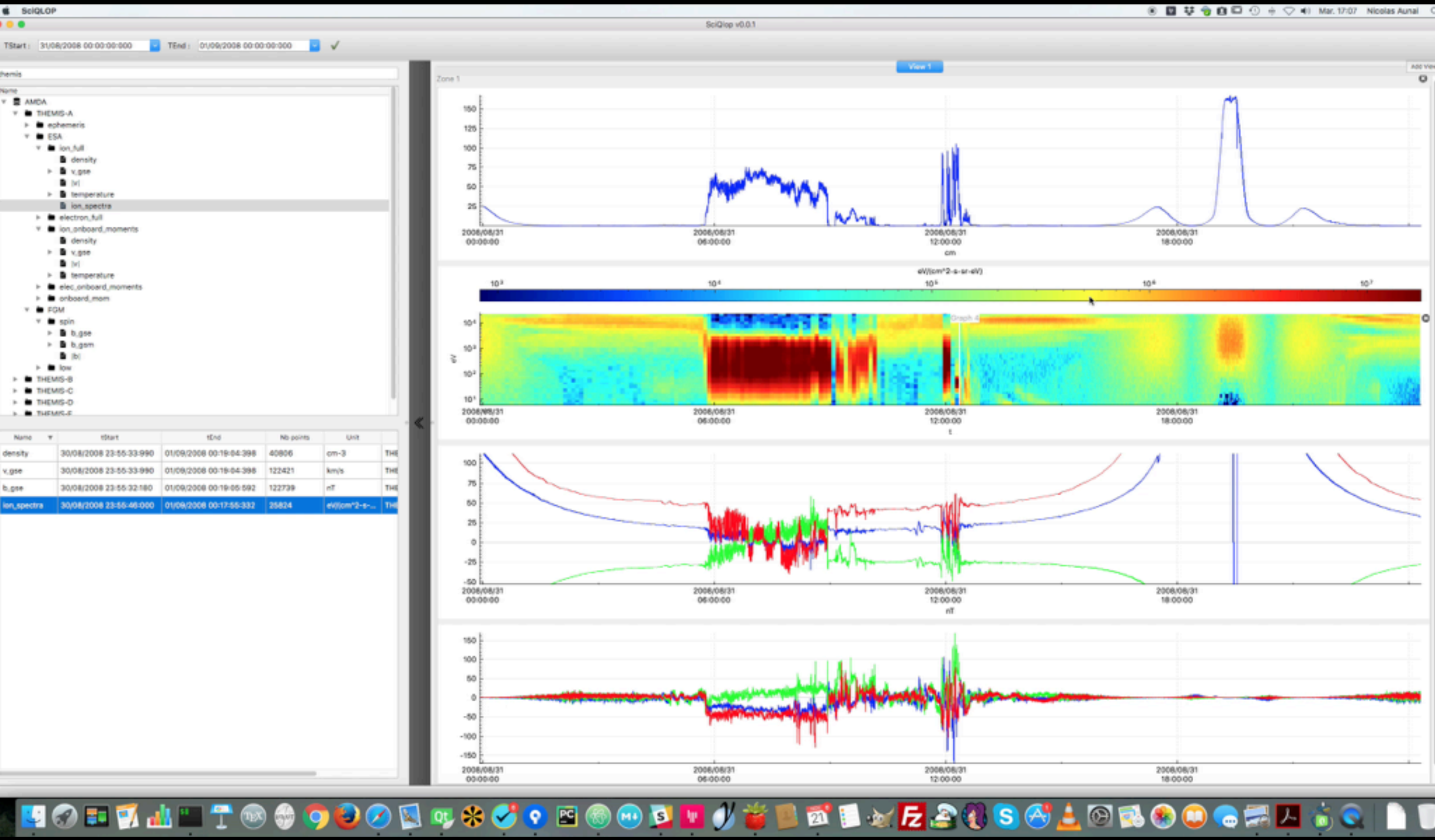
SciQLOP: 4 main features



- Efficient data browsing

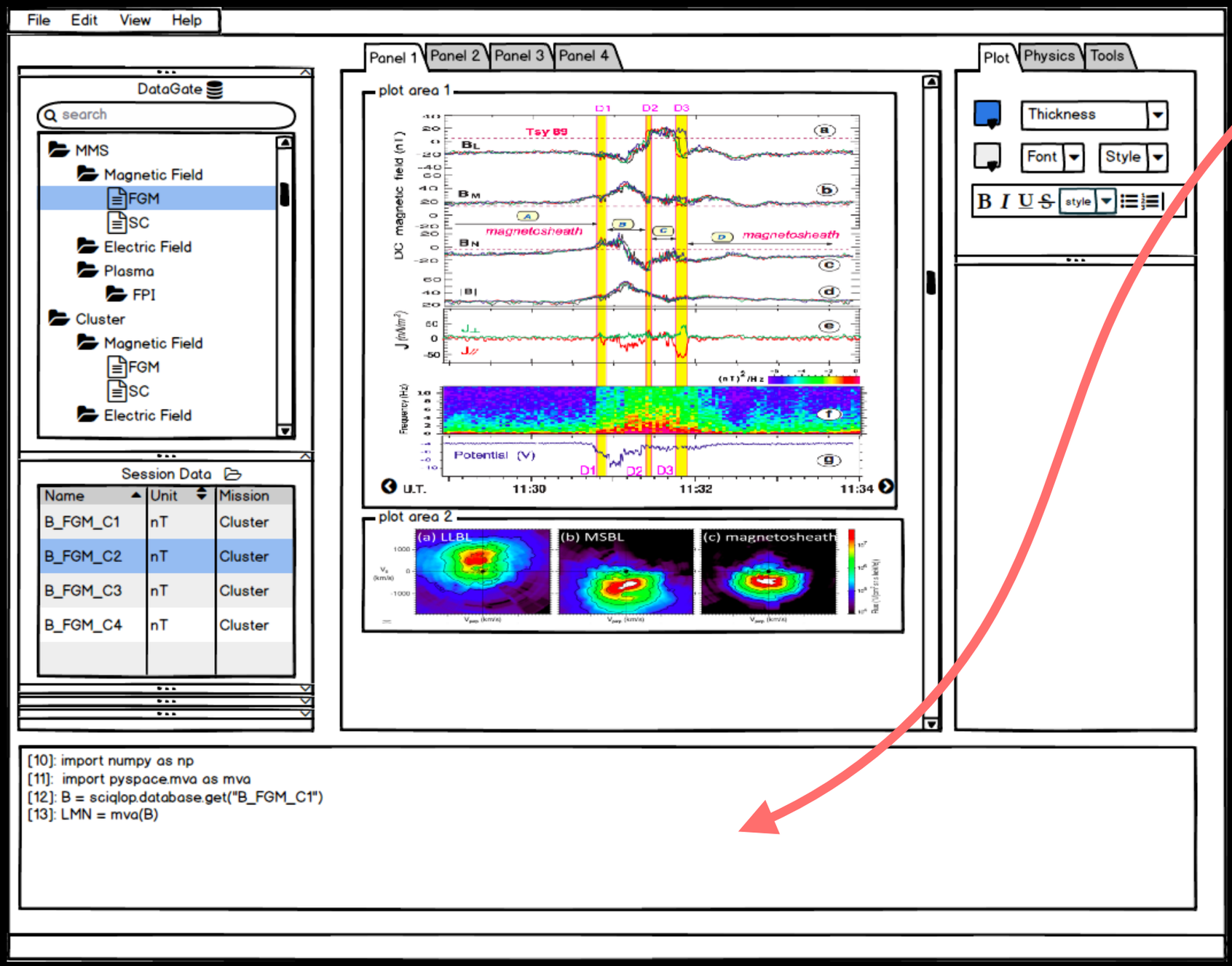
performance is crucial to browse lots of data

SciQLOP: Fluid visualization with > 1M points



SciQLOP: Graph synchronization

SciQLOP: FUTURE PYTHON TERMINAL

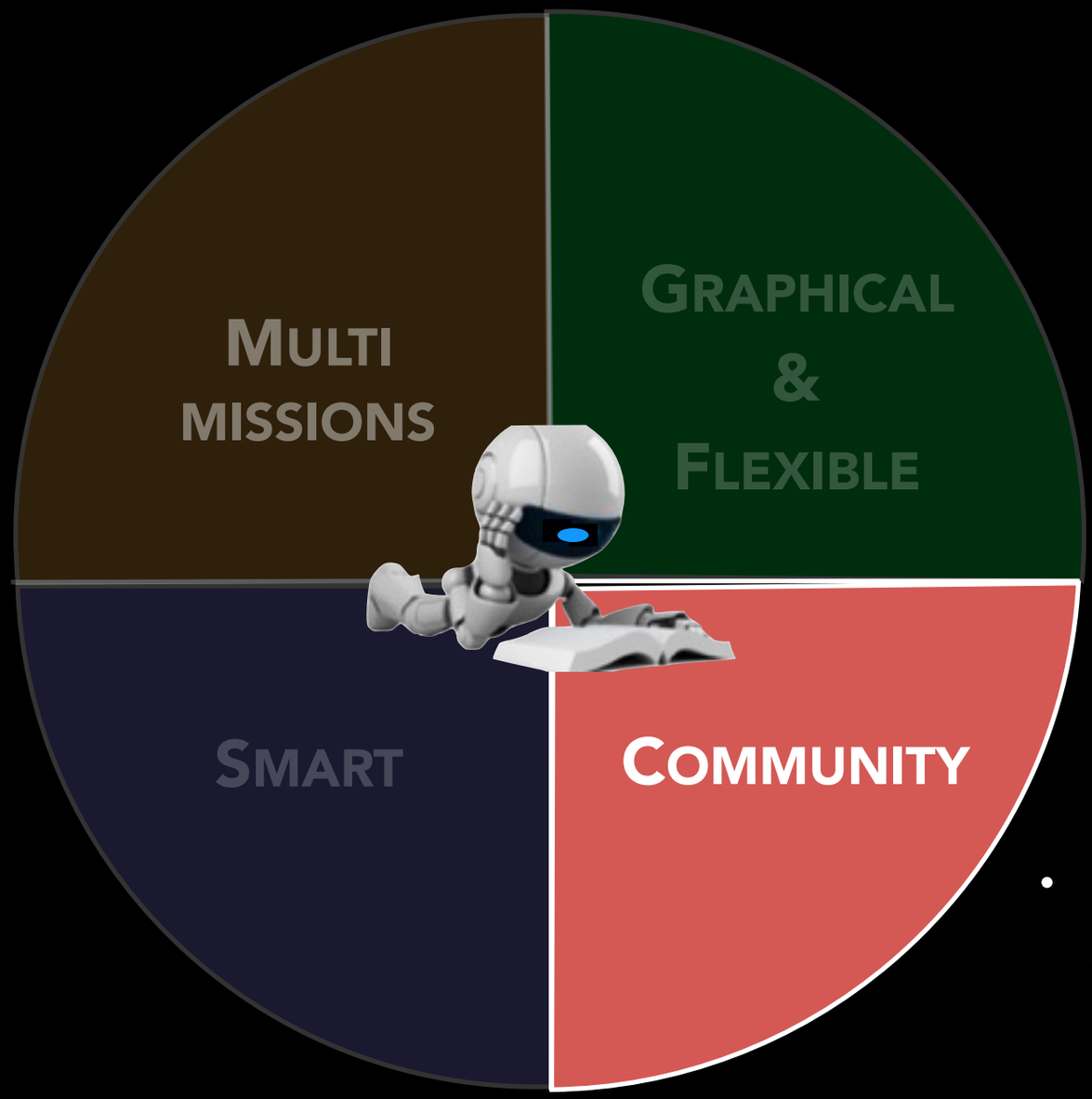


Use loaded data, define function, « virtual products ». Load any community/ custom packages: limitless analysis options

Deploy python scripts close to (remote)data in dedicated Docker containers

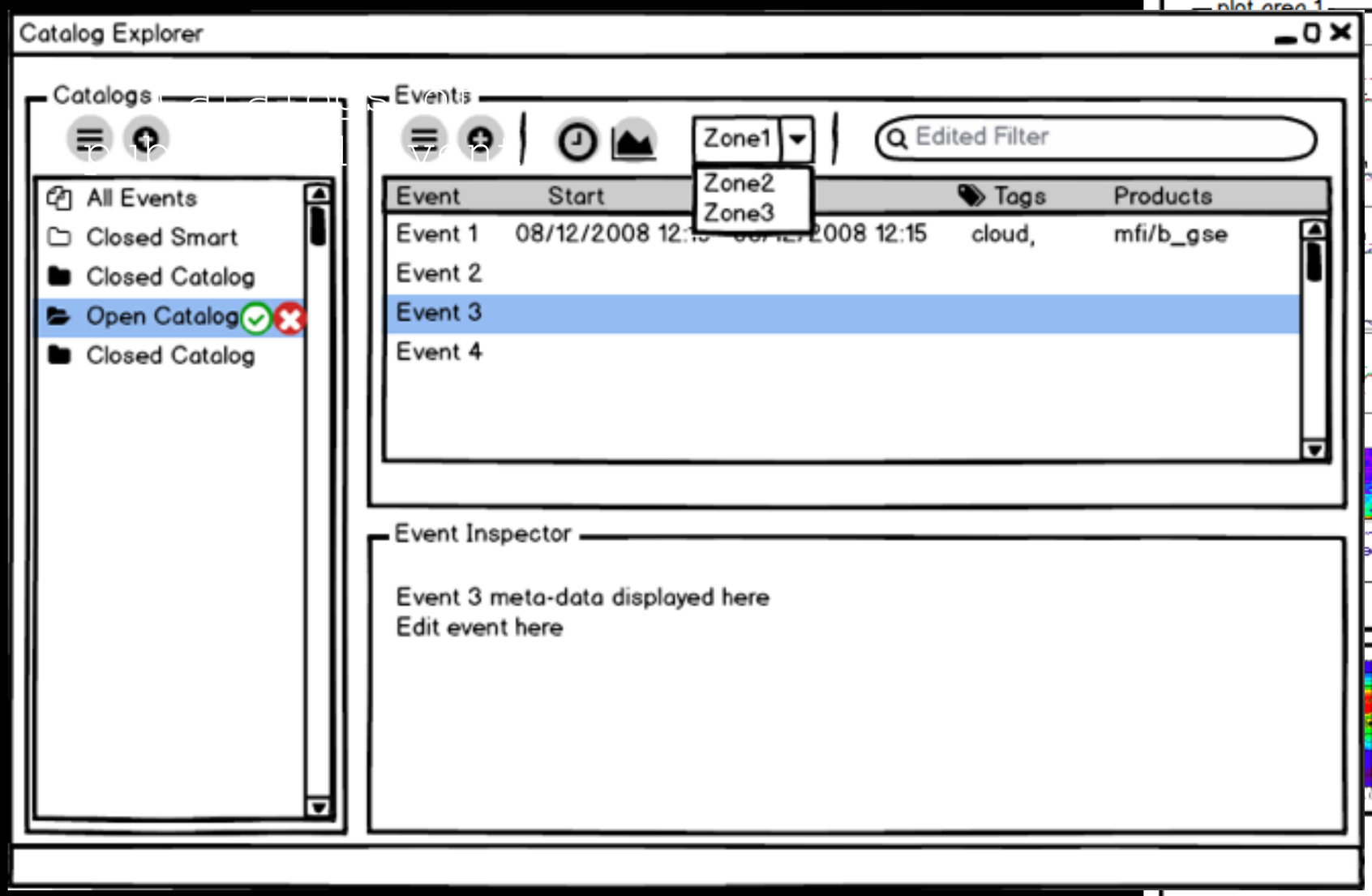
SciQLOP: 4 main features

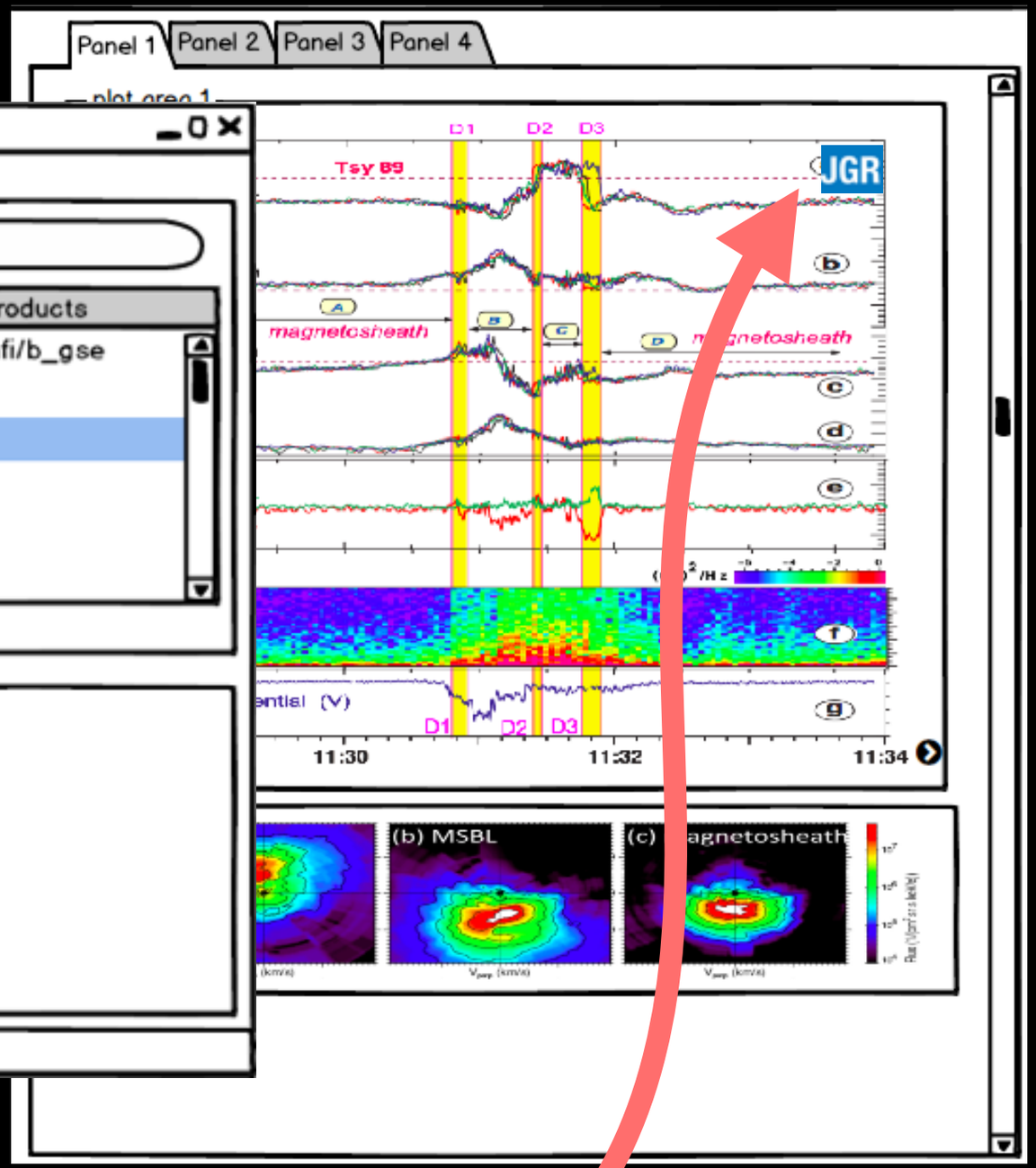
labeling events is a tedious process. We need to **build flexible catalogs together**



- collaborative event catalogs

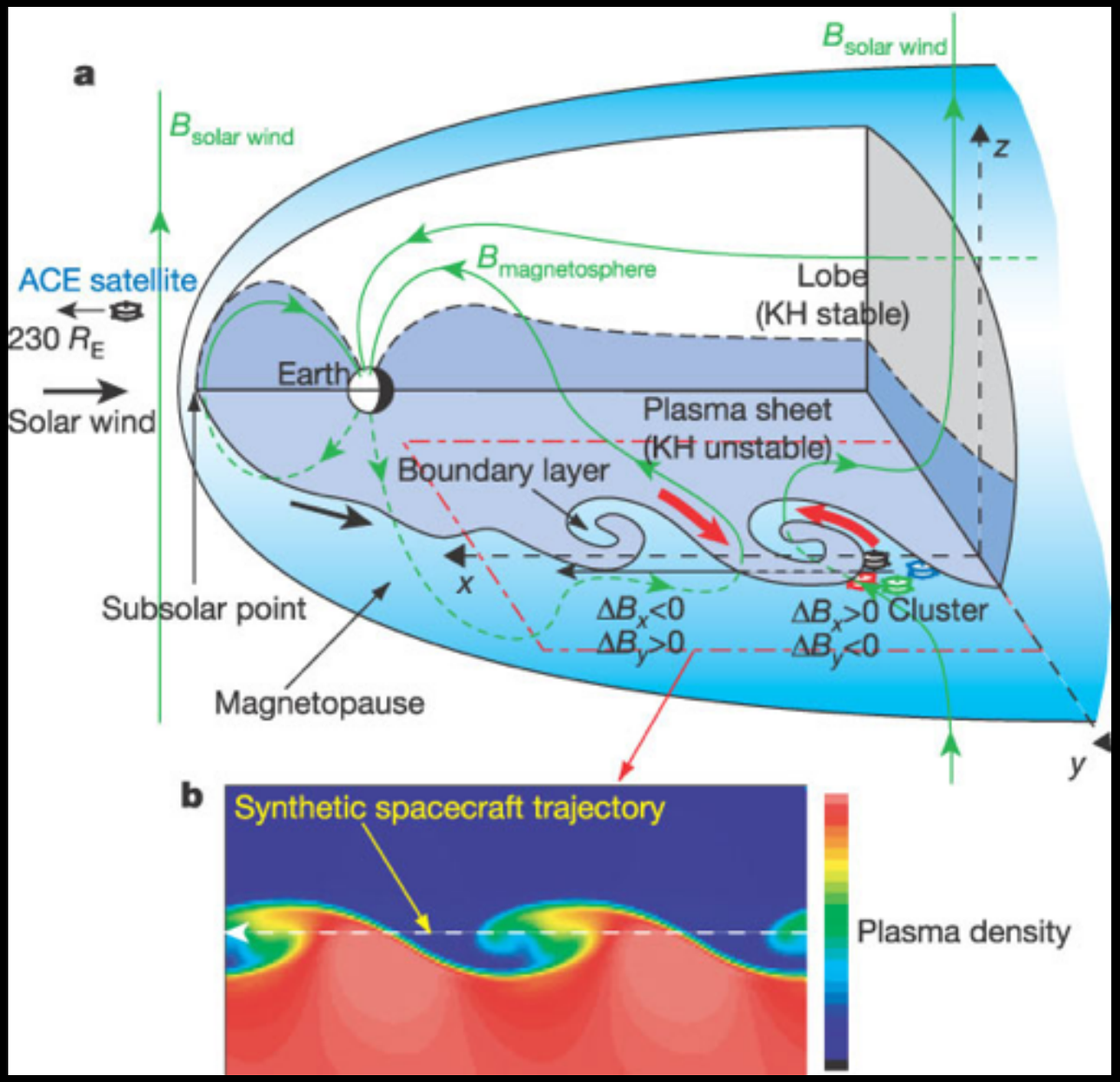
CATALOGS

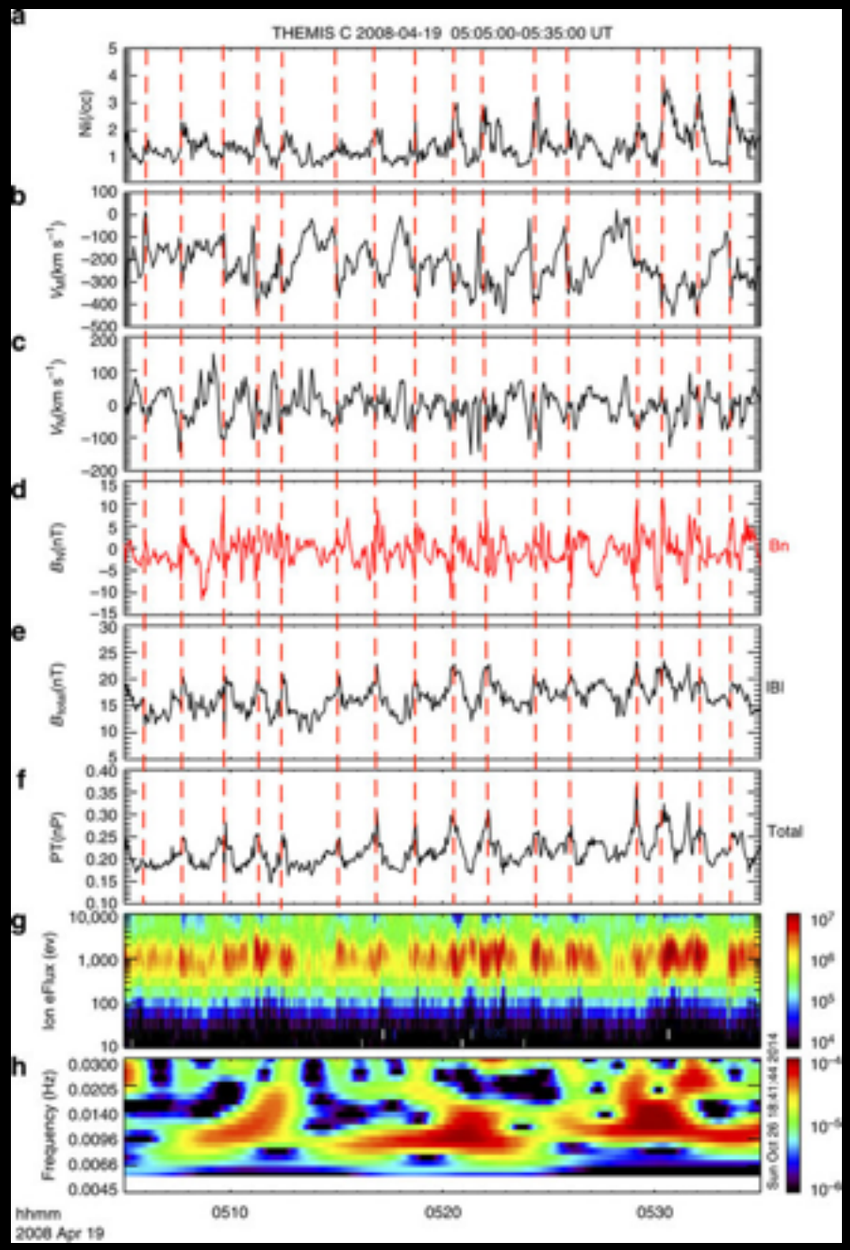




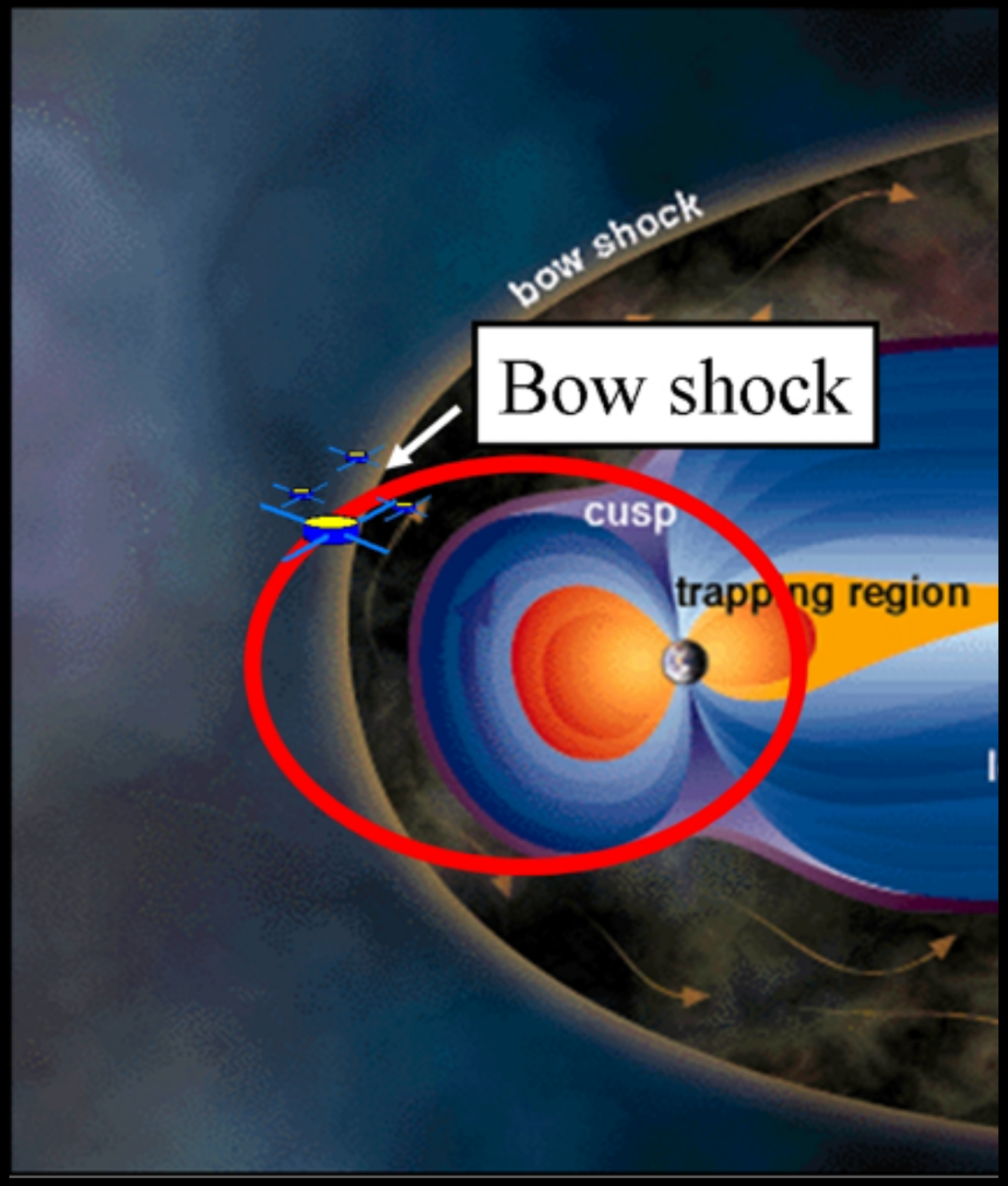
- manual & smart catalogs
- collaborative edition / shared on github
- automatic completion from Machine learning
- ex: catalog of published events

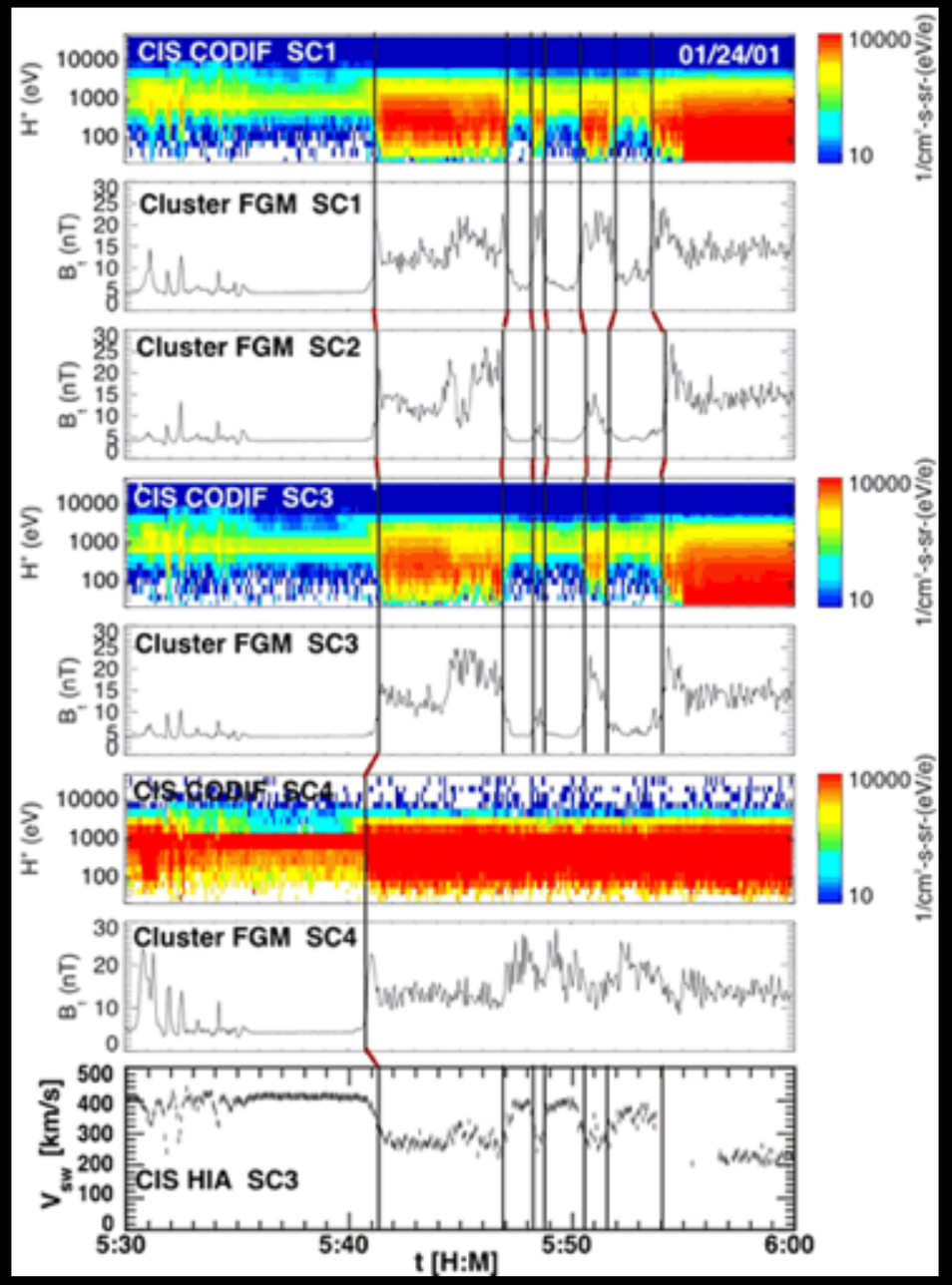
EXEMPLE : KHI CATALOG





EXEMPLE: SHOCK CROSSING CATALOGS





SEE AND EDIT CATALOGS EVENTS ON PLOTS

M'sheath

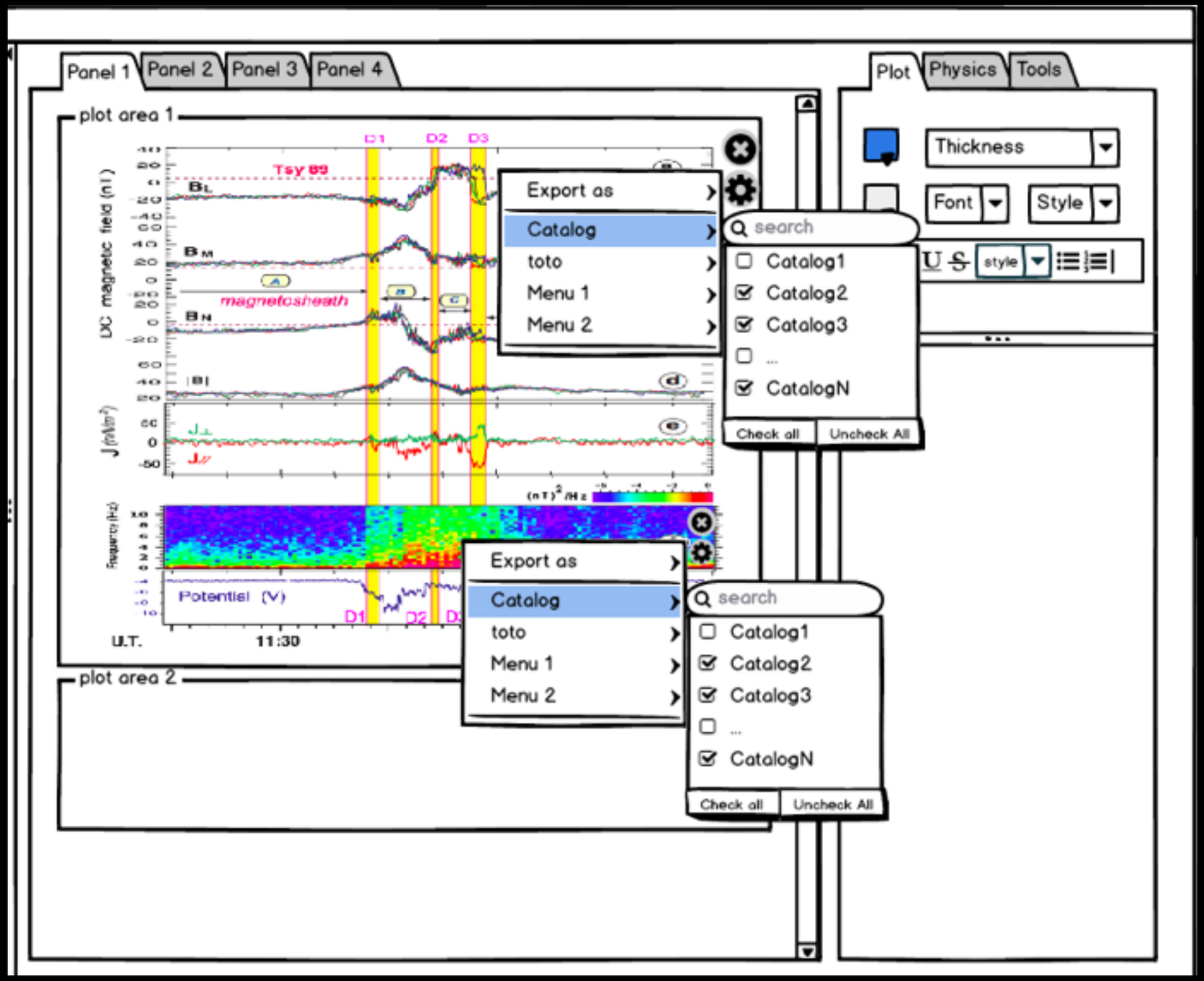
M'pause

M'sphere

augmented data visualization

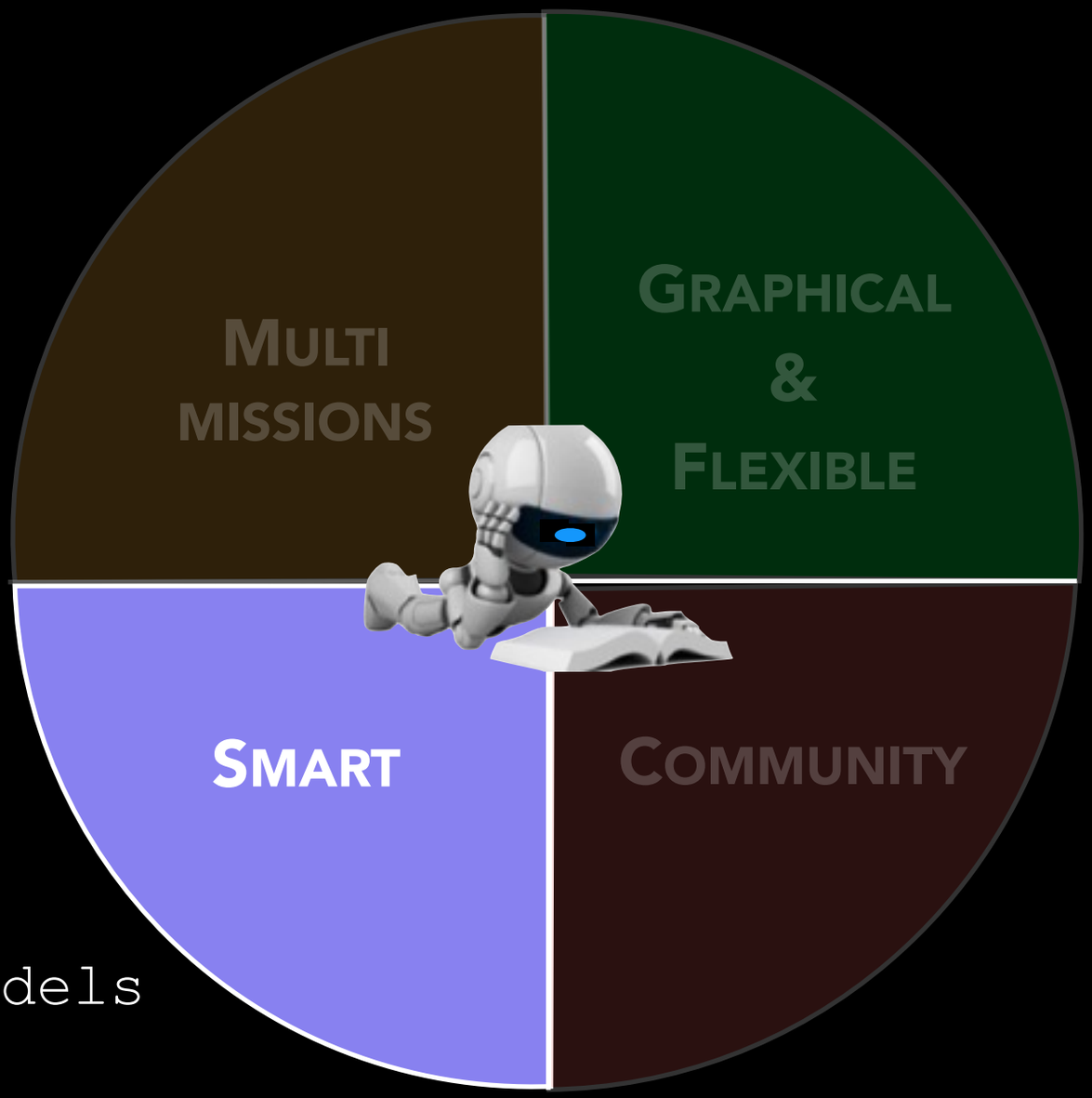
ease data analysis

great tool for students



SciQLOP: 4 main features

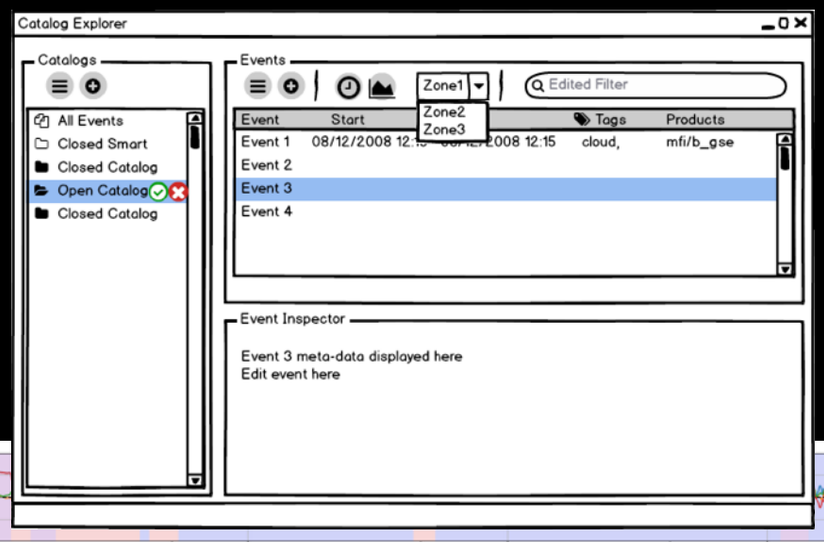
Use catalogs to *learn*
make *predictions*

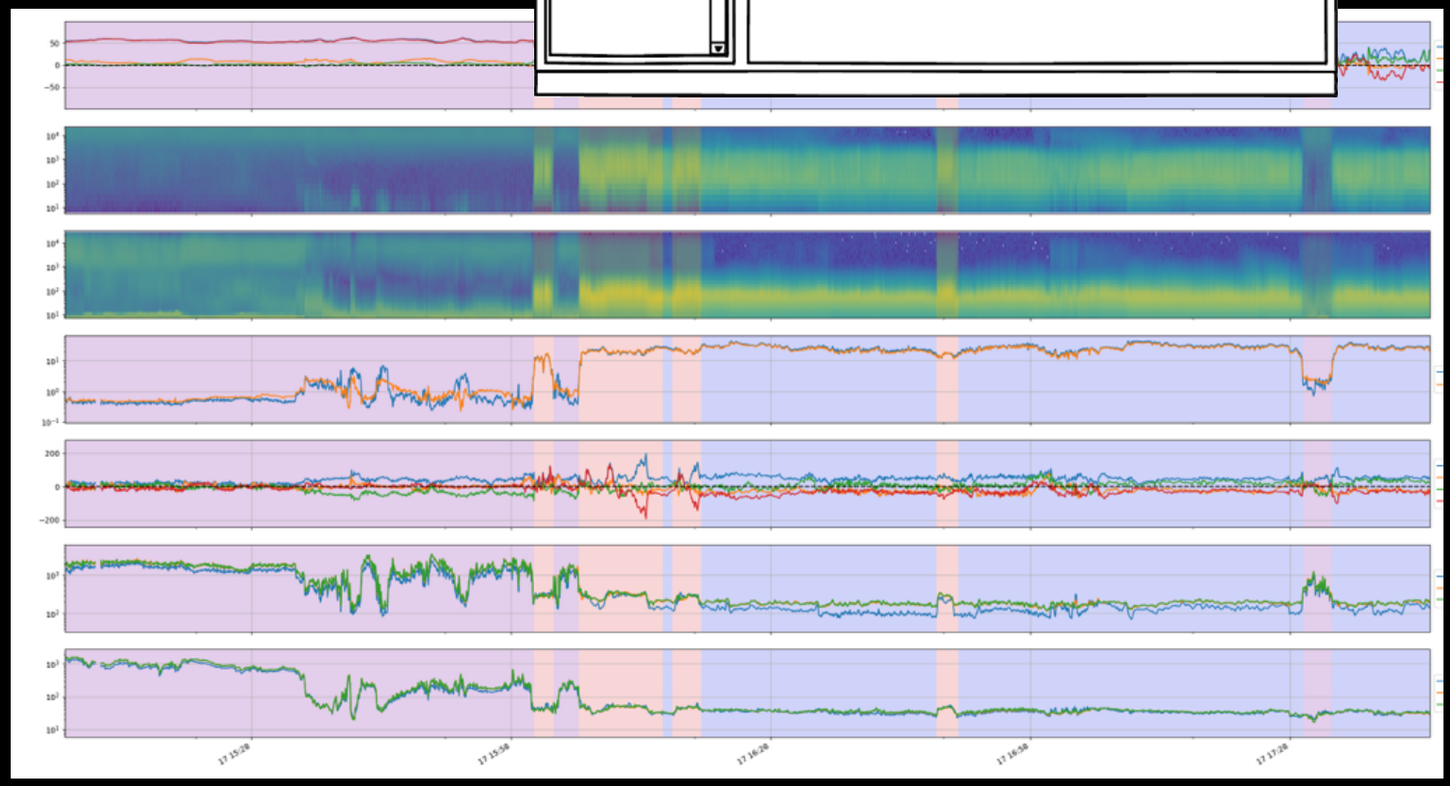


- Smart detection
  - data driven models

LABEL SUGGESTIONS / DATA DISCOVERY

MACHINE LEARNING





M'sheath

M'pause

M'sphere

suggestions / discovery will get better as catalogs grow

« models » (learning) will be sharable and deployed in docker containers to run close to (remote) data.

R&D: LEARN FROM ICME CATALOG

AUTOMATIC DETECTION OF ICMES

PRL X : 1/2DAY /WEEK - 4MHTS

- Brice Bakkali
- Bruno Regaldo
- Louis Vignoli

« WIND FULL LIST »

~400 ICMES DETECTED AT L1

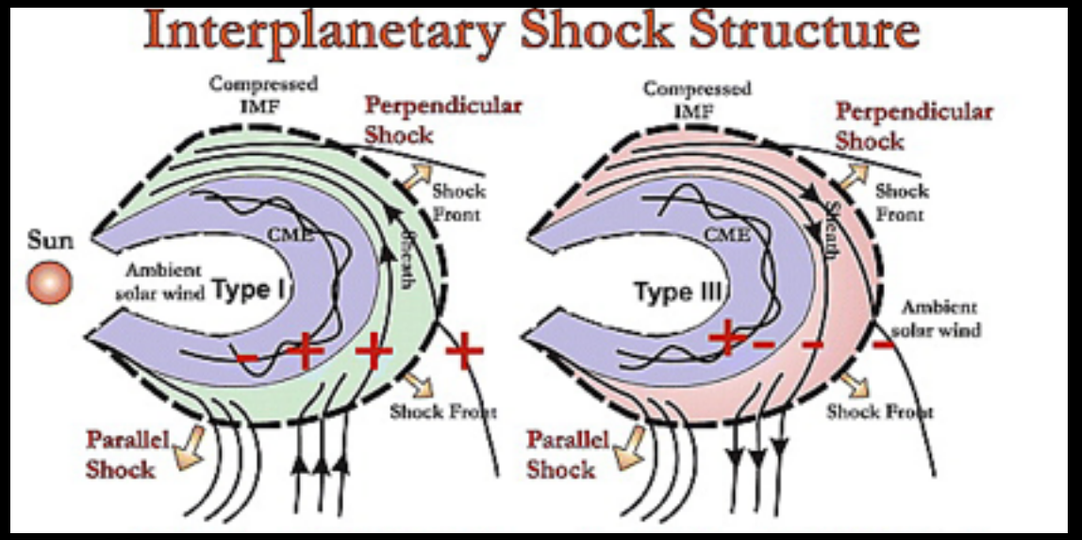
few events, worst case scenario

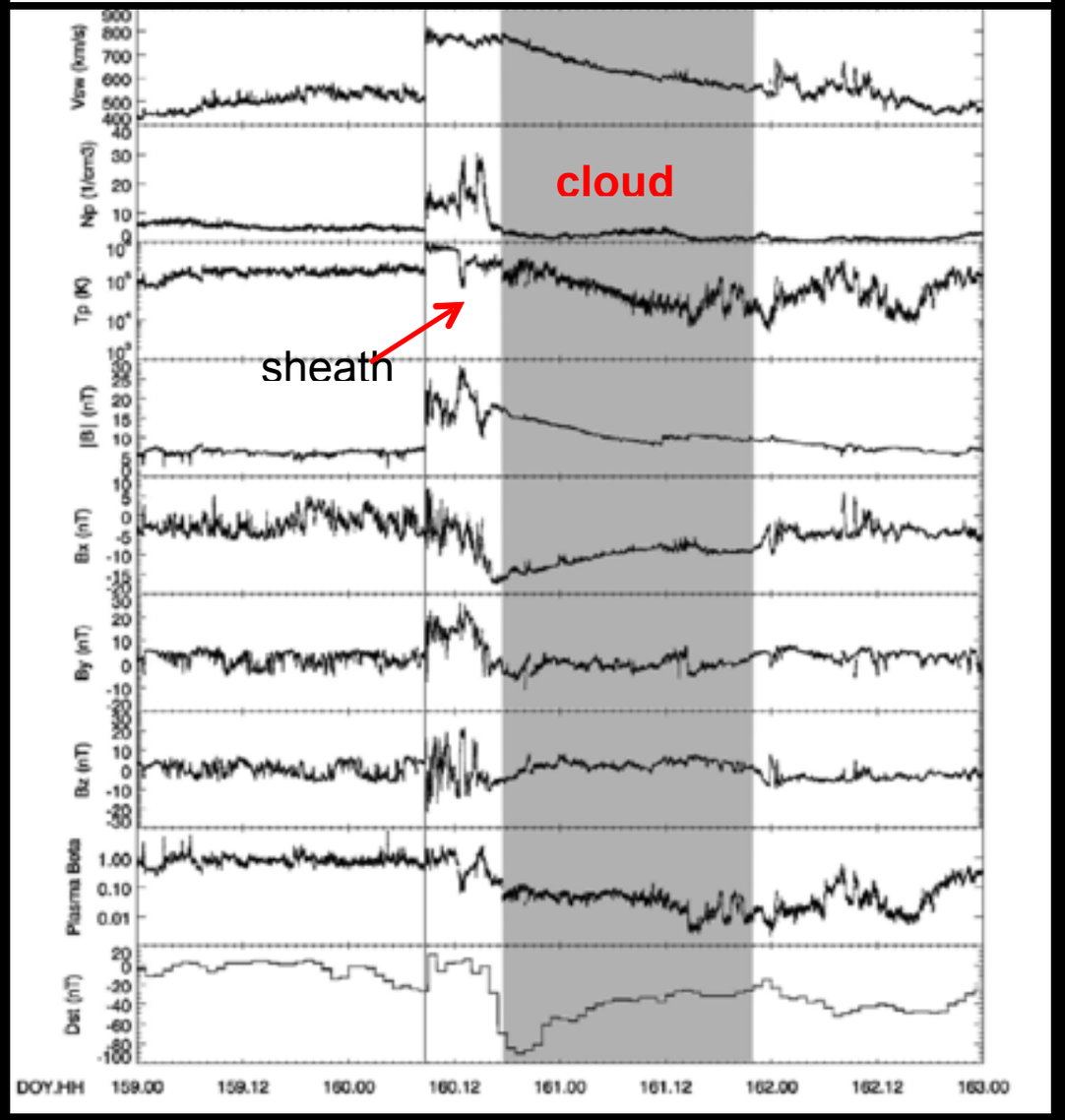
4 DATES:

- START DATE
- START CLOUD DATE
- (END CLOUD DATE)
- END DATE

20 YEARS OF DATA

CAN WE LEARN FROM THIS CATALOG?





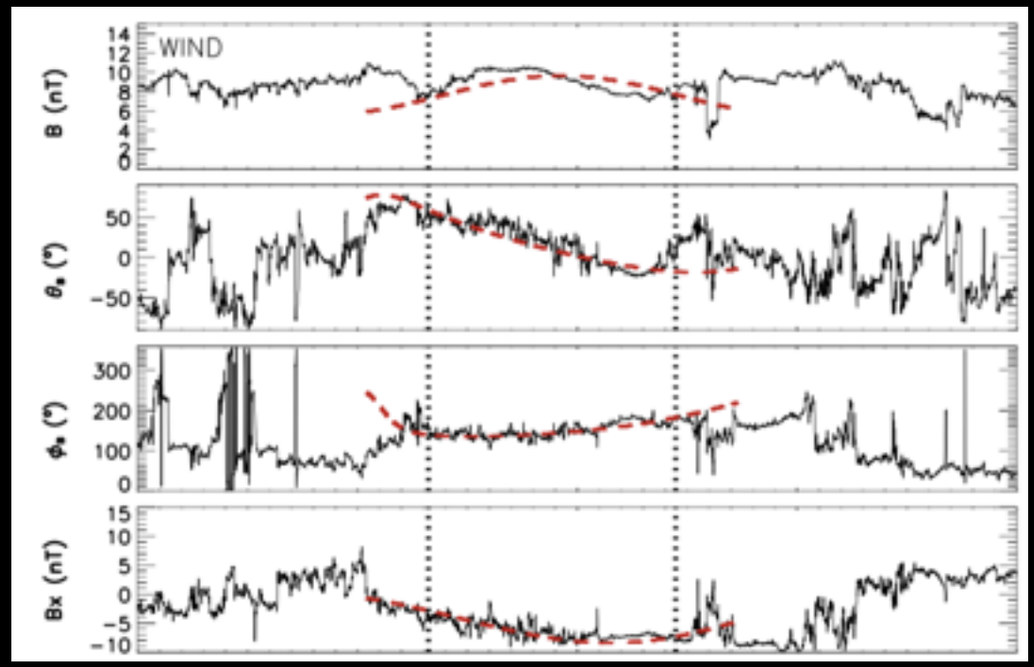
STATE OF THE ART

Lepping et al. 2005

local scale thresholds

Parameter	Limit value	
	Strict Set	Loose Set
Δt(min)	30	25
$<\beta_p>_L$	0.3	0.3
χ^2	450	500
ΔT_L (h)	8	8
$<\|B\|>$(nT)	8	7
$<V_{Th}>$(km/s)	30	30
% of black interval for $<V_{Th}>$	100%	Central 33%
$\Delta\theta_{B,L}$ (deg)	45	35

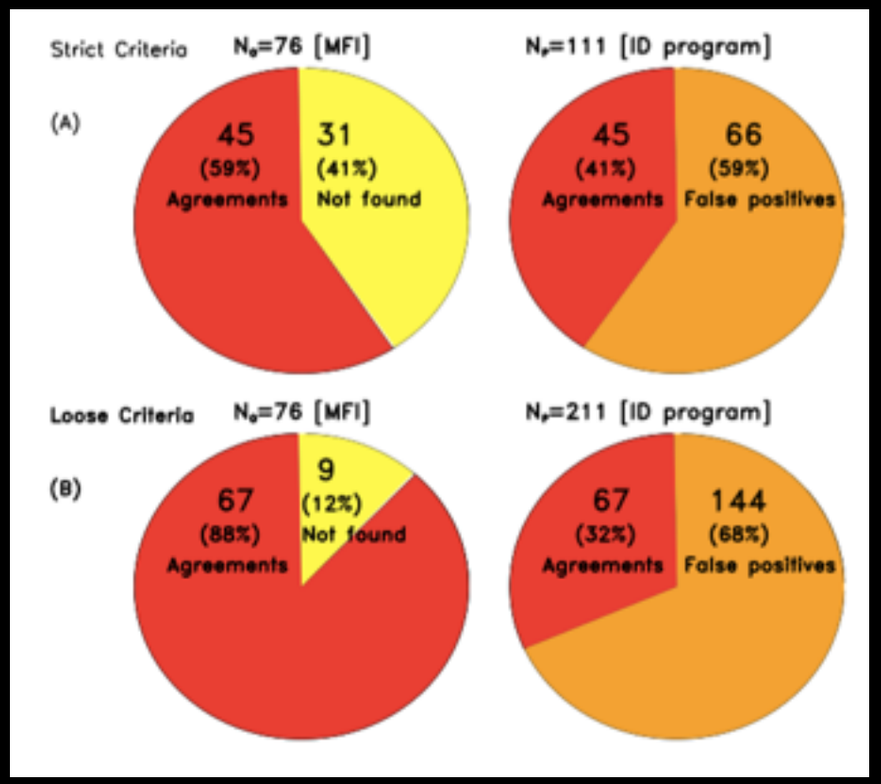
global fit



lots and lots of empirical knowledge on ICMEs required

redo everything for e.g. magnetopause detection

does not work well



EXEMPLE: LEARN FROM ICME CATALOG

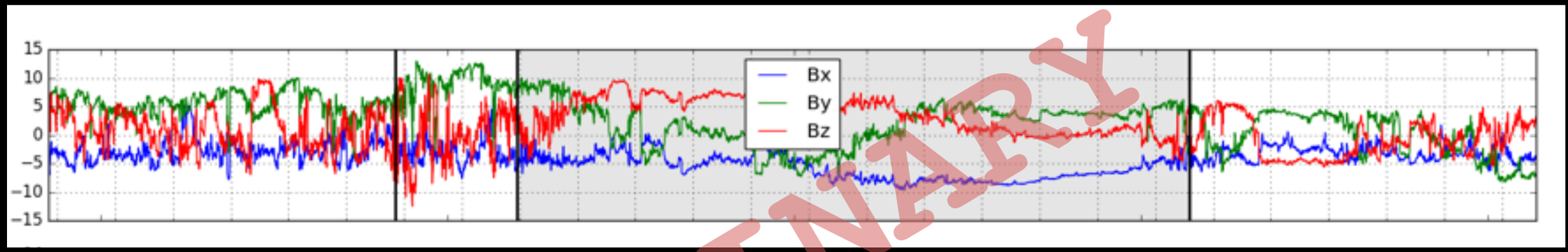
- cut in dt=1h intervals
- 100Go—>100Mo
- use as many measurements as possible (B, density, temperature, beta, spectrograms)
- for each « point », attach the last 24 features to introduce « context »

- train Artificial Neural Networks (ANN)

EXEMPLE: LEARN FROM ICME CATALOG

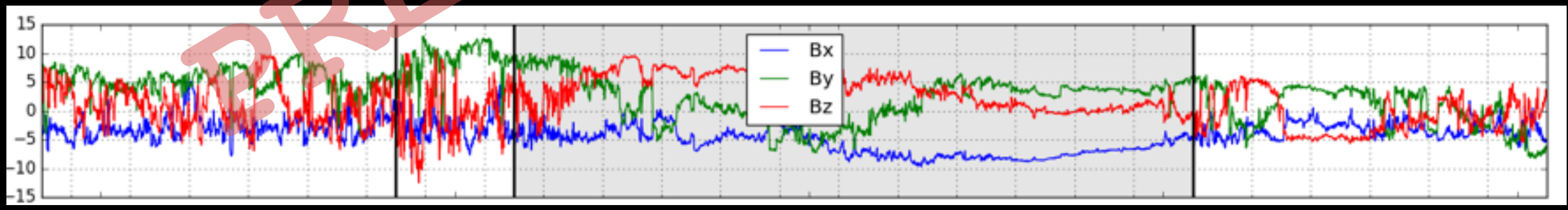
Labeled event

sheath cloud



PRELIMINARY

prediction ! (good case ;-)



May be the best algorithm by far.

EXEMPLE: LEARN FROM ICME CATALOG

Train period: 18 years ~400 IMCEs
Test period: 2 years : 34 ICMEs labeled

Detected clouds: 75 (34 labeled)
- 48 (64%) false pos. 27 (36%) true pos.
- 27 (79%) detected
- 6 (17%)not detected

preliminary results equivalent than state of the art with NO knowledge of ICME required

Sheaths: 30 official
- 81 detected:
  - 14 (17%) true pos. and 67 (82%) false pos.
  - 16 (53%) not detected

only algo. we know that detects sheaths

PRELIMINARY

ROADMAP

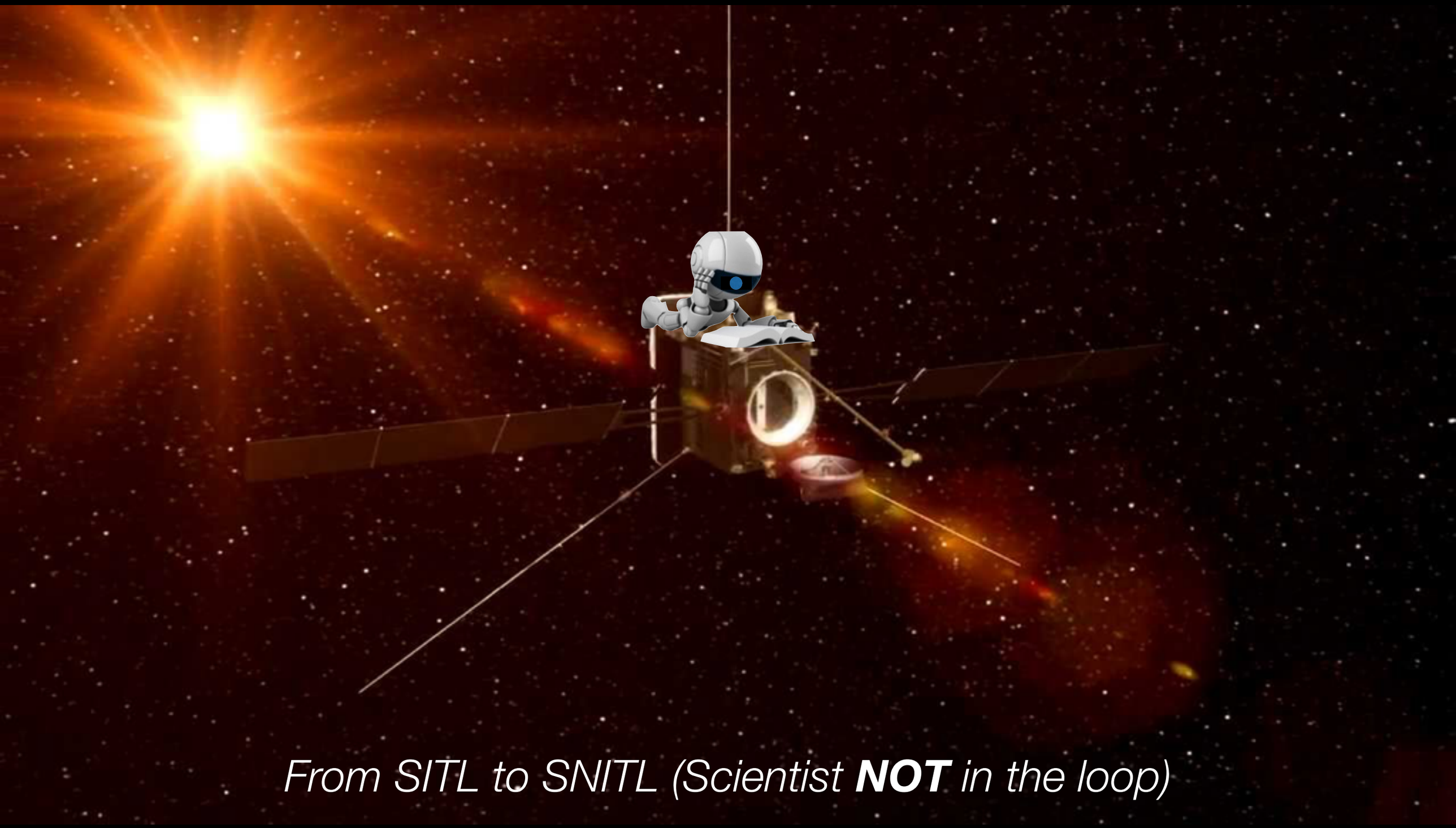
- **Data**
  - Better AMDA integration
  - more products
  - improve interactions/download from WSs
  - Include more servers (ex. ESA/Cluster, CDAWeb…)

- **Catalogs**
  - Share catalogs (import/export VO-tables) (2017)
  - catalogs on github + merge/diff/commit history (2018/9?)

- **GUI**
  - Improve ergonomy / performance (2018)
  - Python terminal / virtual products (2018)

- **Machine Learning**
  - R&D ICMEs and magnetopause : these G. Nguyen (2017-2020)
  - Machine learning integration ?

FUTURE: BUILT-IN SCIQLOP ENGINE?



*From SITL to SNITL (Scientist **NOT** in the loop)*