

Stormtime Plasma Sheet Recent, Ongoing, & Future efforts

Savvas Raptis¹, Viatcheslav Merkin¹, Connor O' Brien^{1,2}, Shin-ichi Ohtani¹, Anusree Devandan³, Kareem Sorathia¹, Louis Richard⁴, Amy Keese³, Matina Gkioulidou¹, Simon Wing¹

¹ APL/JHU, Laurel, MD, US

² Center for Space Physics, Boston University, Boston, MA, USA

³ University of New Hampshire, NH, USA

⁴ Swedish Institute of Space Physics, Uppsala, Sweden

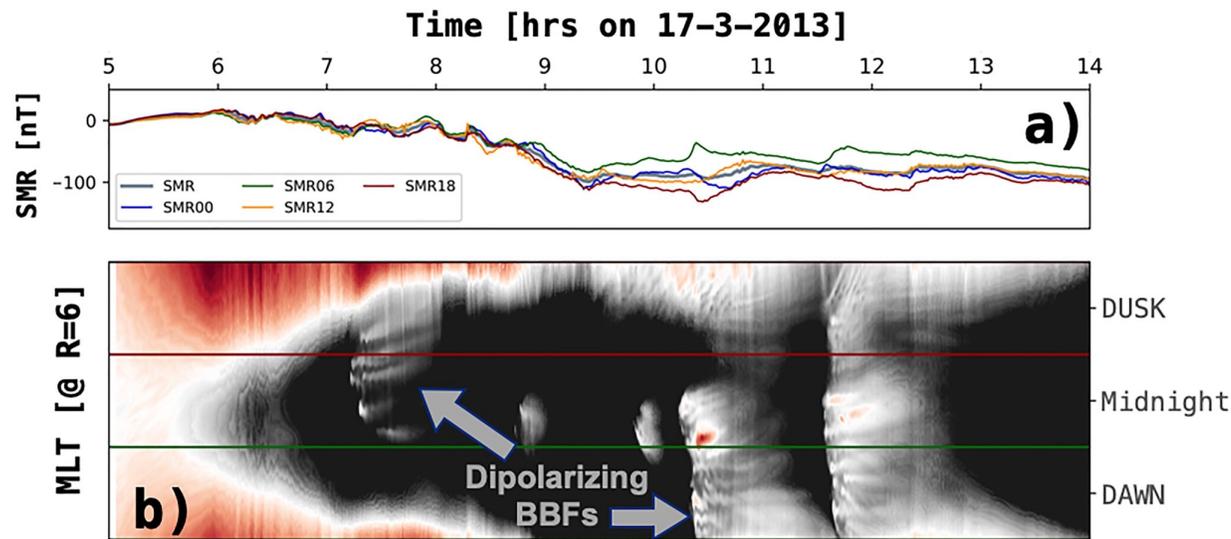


Outline

- **Global Convection Patterns of Magnetic Flux (1 published)**
 - Question: How does convection in the magnetotail change during storms and how it affects magnetic flux transport
- **Bursty Interval Contribution and Dawn/Dusk Asymmetries (1 “almost”* accepted + 1 ongoing)**
 - Question: Is the occurrence distribution of BBFs different during storms? How much they contribute to flux transport and how do their properties change during storms?
- **Machine Learning modeling of in-situ properties (1 “almost”* submitted)**
 - Question: can we use SW information and ground magnetometers to model plasmashet properties? How this changes during storms
- **Future Work Ideas**

*Very different definitions of “almost”

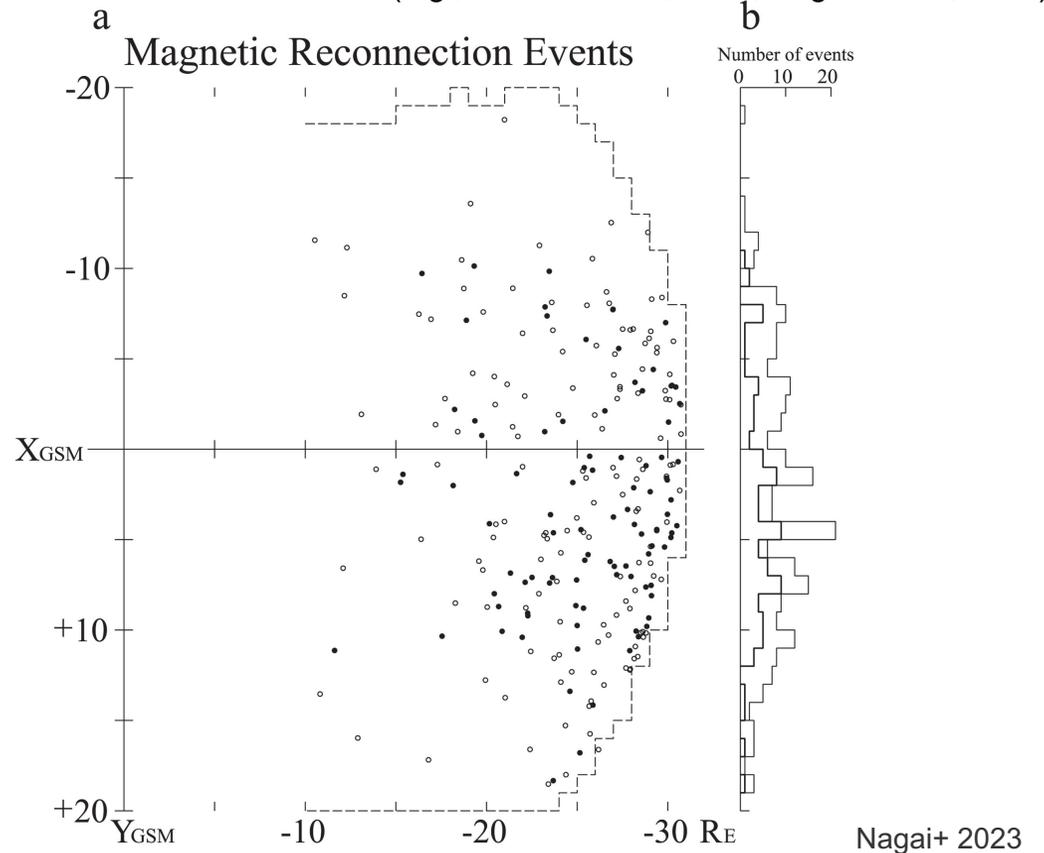
Dawnside Current Wedge & Asymmetries in Magnetopause Reconnection



Ohtani+ 2021,2024, Sorathia+ 2023

Dawnside Current Wedge is a distinct storm-time phenomenon associated with increased westward auroral electrojet (AEJ) at the dawn sector.

*Similar results with MMS (e.g., Hubbert et al., 2022; Rogers et al., 2023).

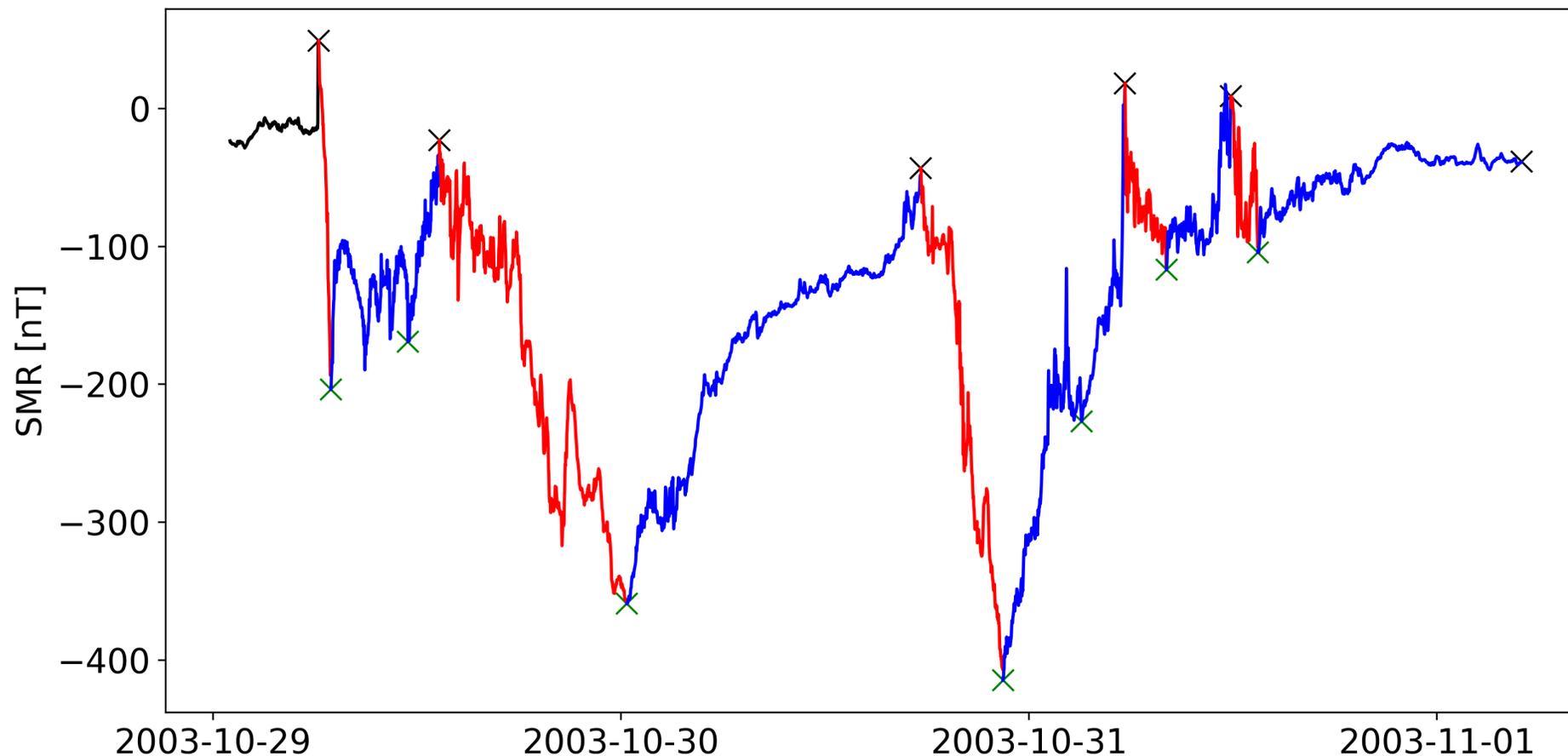


Nagai+ 2023

Magnetotail reconnection site is spatially skewed toward the dusk (pre-midnight) sector during typical activity, **but** this distribution shifts dawnward when the magnetosphere is being driven intensely (such as during major storms).

Needed Step: Storm phases classification

(semi)-automatic storm finder and classifier:



List is open, easy to use, easy to maintain

<https://zenodo.org/records/15127938>
(v3 updated to 2025, NOT manually verified)

Plasmasheet Coverage per mission

Criteria to find CSP

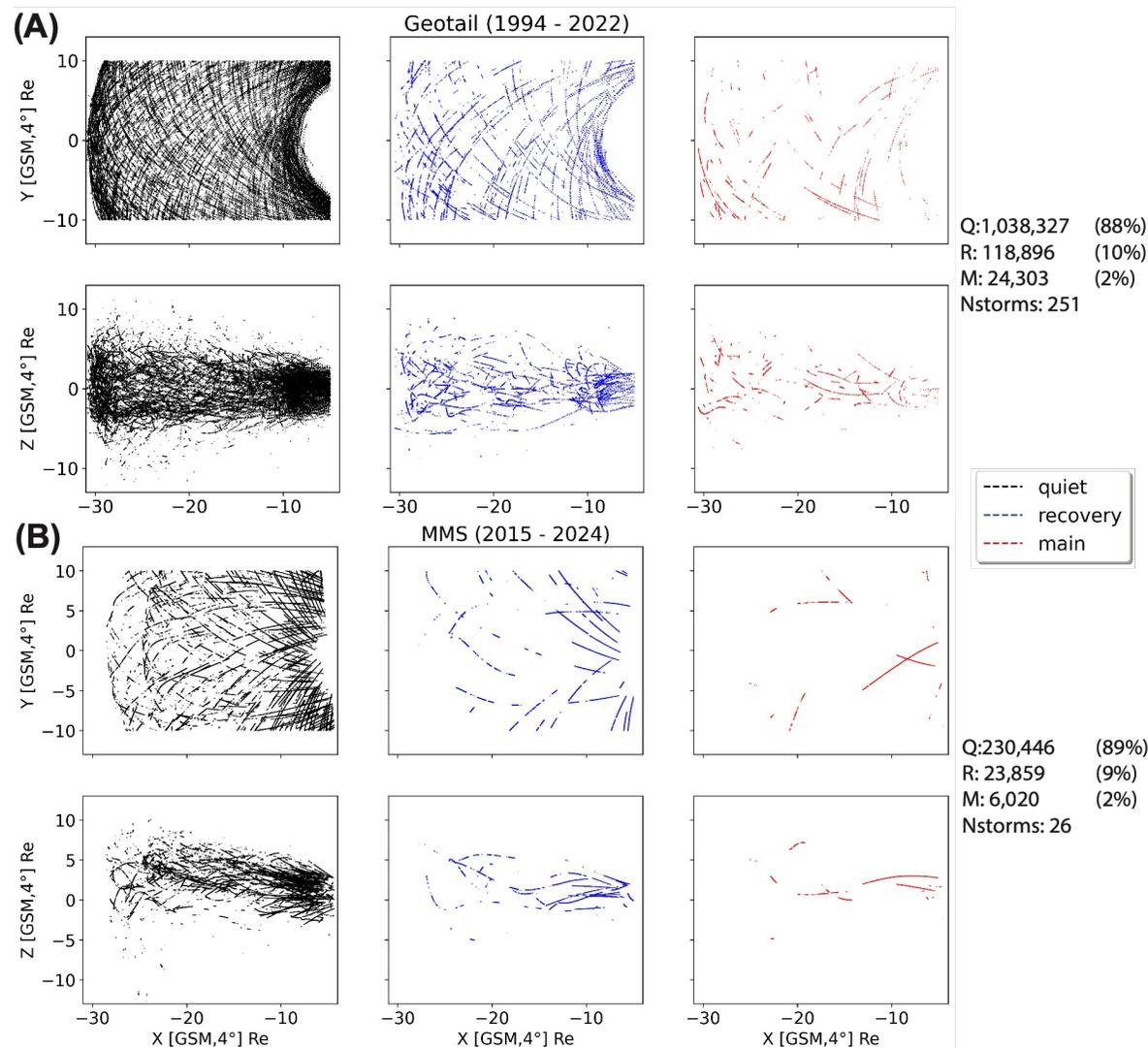
1. $|Y_{GSM,4^\circ}| < 10$
2. $-5 < X_{GSM,4^\circ} < -31$
3. $\beta = \frac{P_{the}}{P_{mag}} > 1$
4. $|Bz| > 2\sqrt{Bx^2 + By^2}$

See e.g., Ohtani+ 2008, Guild+ 2008, Roziers+ 2009, Vo+ 2023

Geotail > 1 million points ~250 storms
MMS ~ 250k points ~25 storms

Findings:

1. **MMS** have limited observations during storm times (especially main phase)
2. Main phase contains data from about 6 storms for **MMS**
3. Slightly more dawnside data during main phase for **Geotail**

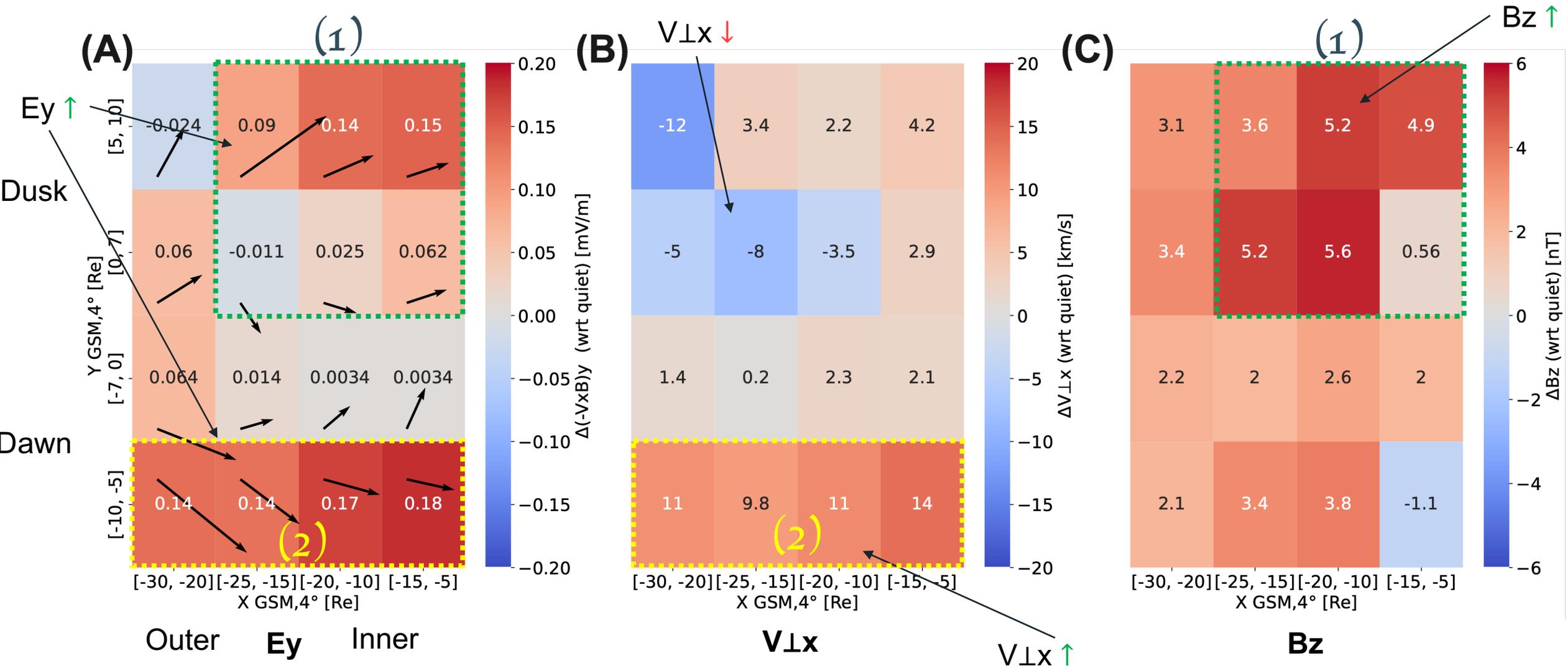


Storm - Main Phase Difference (Geotail | 1994 - 2022)



Dawn sector: storm-time magnetic flux transport linked to **faster plasma flows**

Dusk sector: storm-time magnetic flux transport linked to **stronger dipolar magnetic fields**



What do we know so far?

Plasma sheet storm time:

1. Elevated E_y associated with increased B_z , and limited enhancement of $V_{\perp x}$
2. Dusk observations showing more dipolar magnetic field ($B_z \uparrow$)
3. Dawn are associated to relatively faster flow ($V_{\perp x} \uparrow$)

Let's move to Bursty Intervals!

Bursty intervals (Jets, BBFs, BEIs, etc.) and constrains

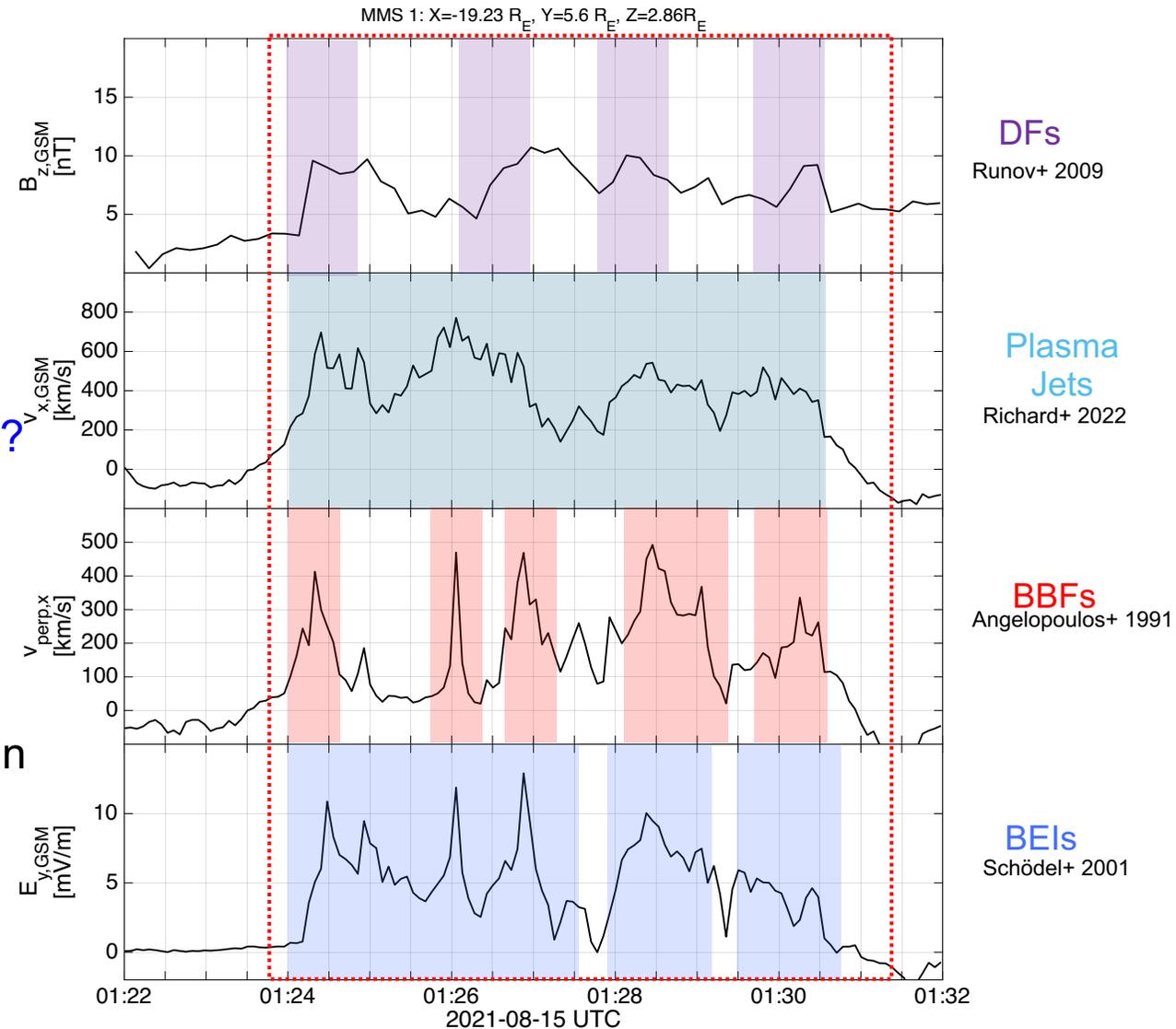
Analyzing BBFs statistically is non-trivial:

1. Arbitrary criteria and thresholds
2. BBFs are bursty, merge, and de-accelerate.
3. Orbital biases of missions
4. Are we even in the plasmashet? (can discuss more)
5. Moment calculation, composition, energy and time res
6. No knowledge of spatial scales and relative location
7. Is it a direct hit or glazing? How probable is each case?
8. General sparsity of data and driving condition distribution.
9. Fast BBFs missed by instruments (e.g., $v > 1500$ km/s)
-

So at least, (3,6,7,9) and maybe (2, 4,5) are not that bad in simulations, however number (1, 8) get complicated...

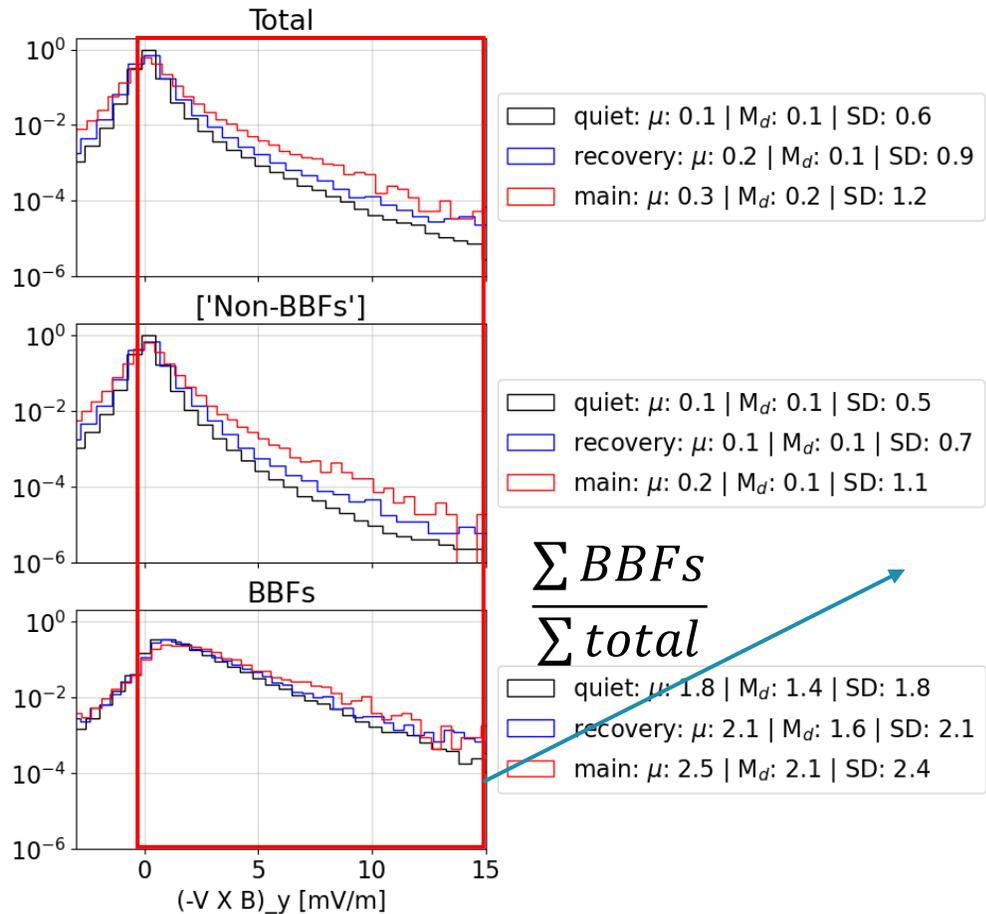
BBFs:

- Fast ion flows ($v > 400$ km/s),
- 10-100s in duration
- $\sim 4 R_E$ size

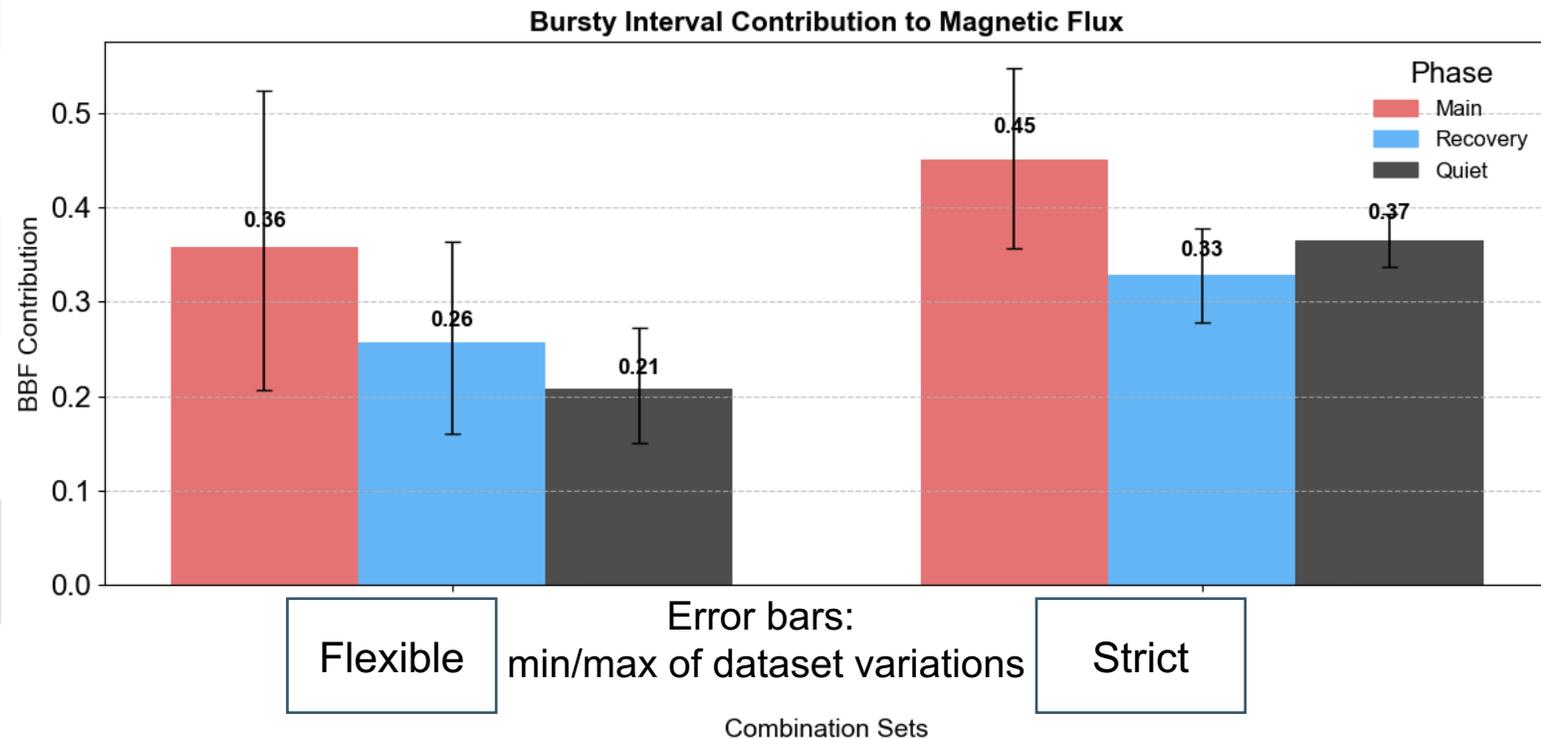


We can still do things: Earthward Bursty Magnetic Flux Transport

Histograms are normalized per phase



Strict: $|Y| < 10$ [Re] | $\beta > 1$ | $|B_z| > 2\sqrt{B_x^2 + B_y^2}$ | $n < 3$ [1/cc])
 Flexible: $|Y| < 15$ [Re] | $\beta > 0.5$ | $n < 5$ [1/cc])



After Lunch [Anusree Devanandan: Statistics of Stormtime BBFs](#)

Part #1 Summary

1. Stormtime Global Convection (**published**):

1. **Plasma sheet E_y** is elevated due to **increased B_z** , with **limited enhancement of $V_{\perp x}$**
2. **B_z enhancement** is more **prominent at Dusk**
3. **$V_{\perp x}$** is more **elevated at Dawn**

2. Plasma sheet bursty Intervals (**accepted & ongoing**):

1. **BBFs can contribute ~25% of earthward magnetic flux during quiet and ~40% during main phase.**
2. This enhancement linked to a **stronger background B_z** , while BBF velocity stays relatively constant (See *Anusree's presentation for more*).
3. **There seems to be a Dawn/Dusk Asymmetry (will discuss at the end) on BBF occurrence and stormtime.**



Interesting take by ChatGPT on data
& MAGE comparisons

Detective MAGE looks at another
simplified magnetosphere, MAGE
also looks like a fox apparently

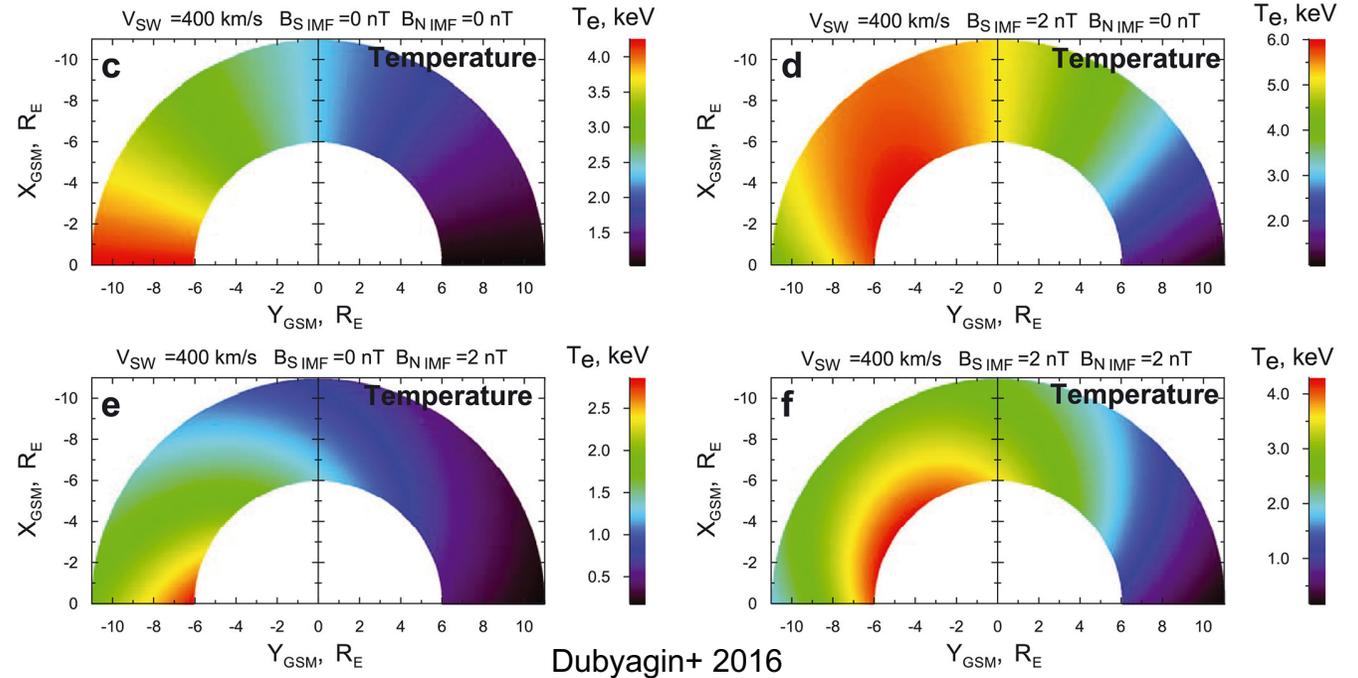
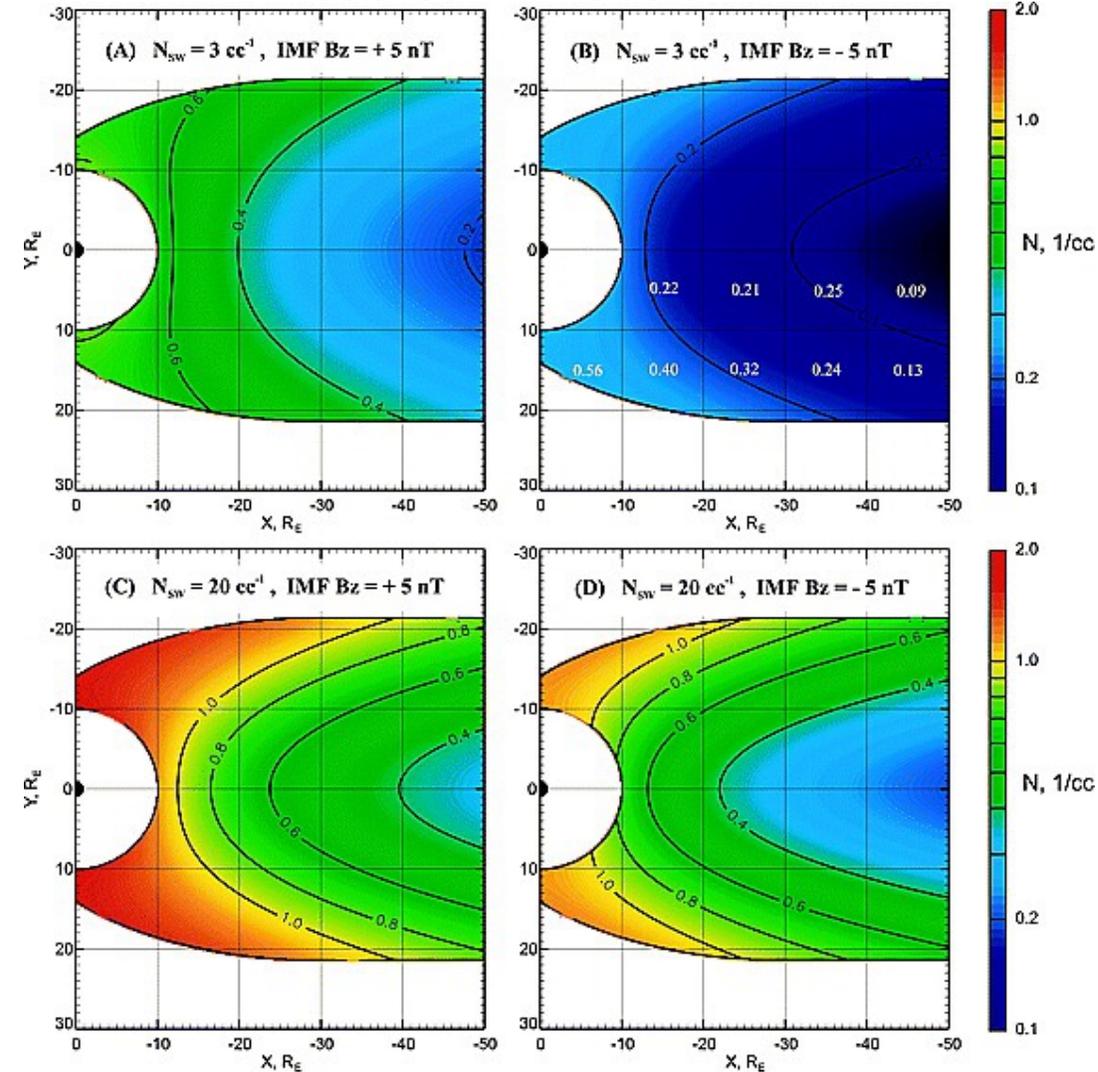
What can we do with all this data?

Machine Learning of course!

Baseline empirical models

Modelled with Geotail

Modelled with THEMIS



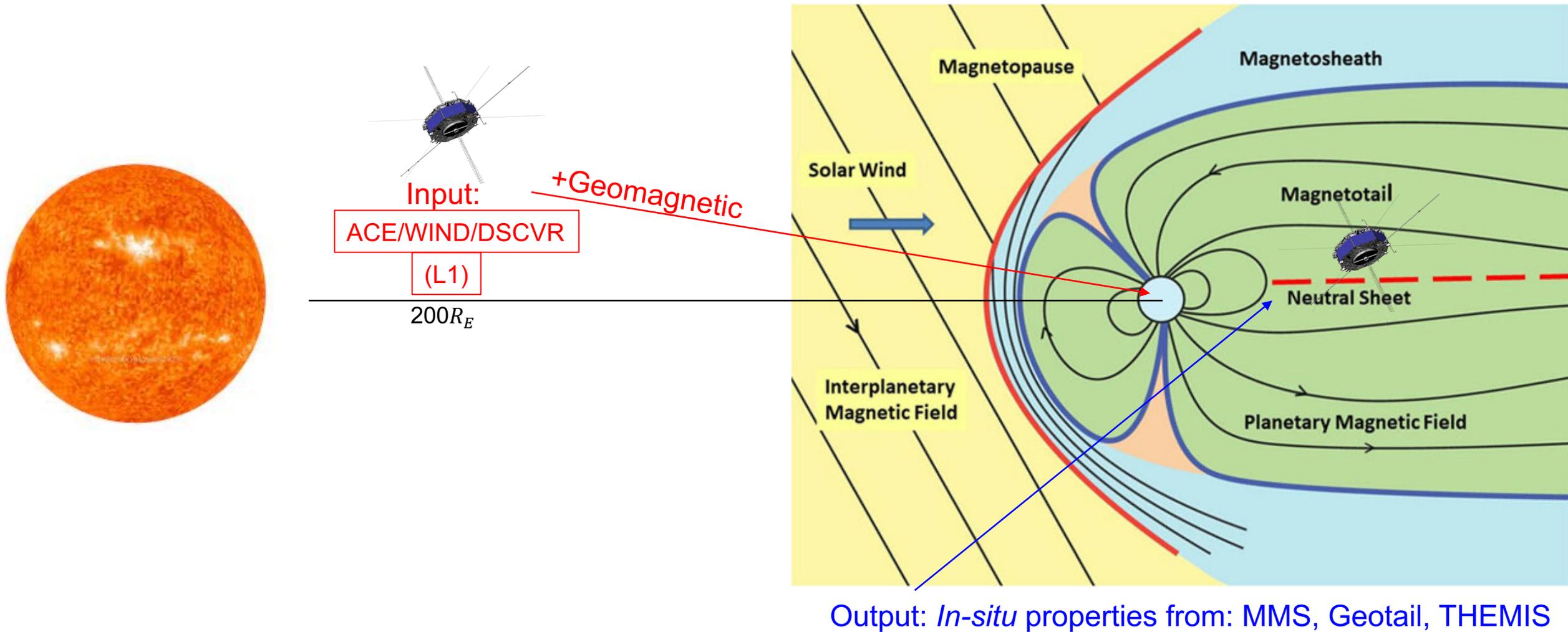
Dubyagin+ 2016

Why then work on this?

1. More data under different conditions
2. MMS was never used with its state of the art instrumentation
3. These models don't include time history
4. ML methods can reveal non-linear relationships easily

Tsyganenko & Mukai 2003

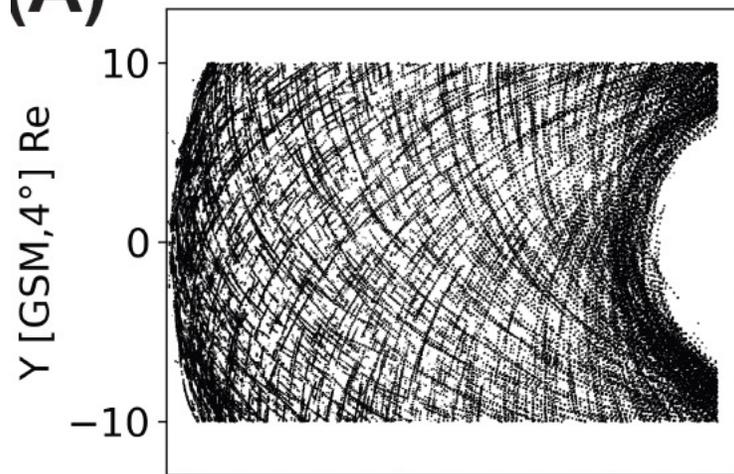
Where are we & what are we doing?



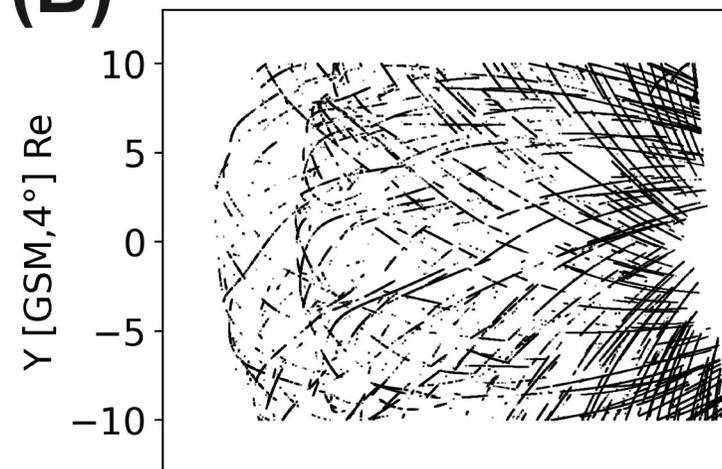
Goal: Model Plasmasheet properties based on driving (SW) and geomagnetic conditions

The dataset (output – Central Plasma Sheet)

(A)

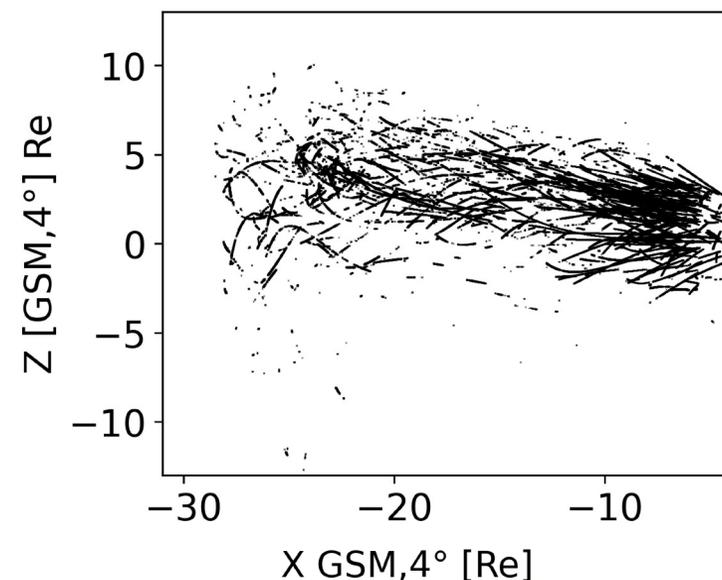
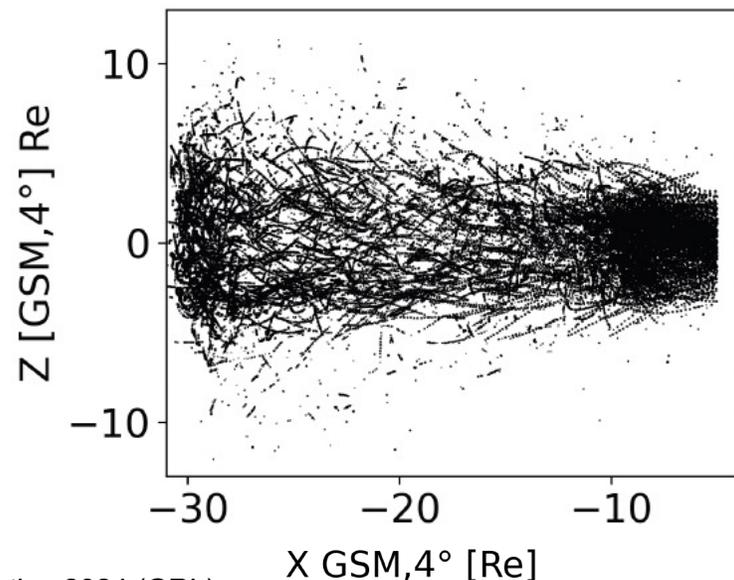


(B)

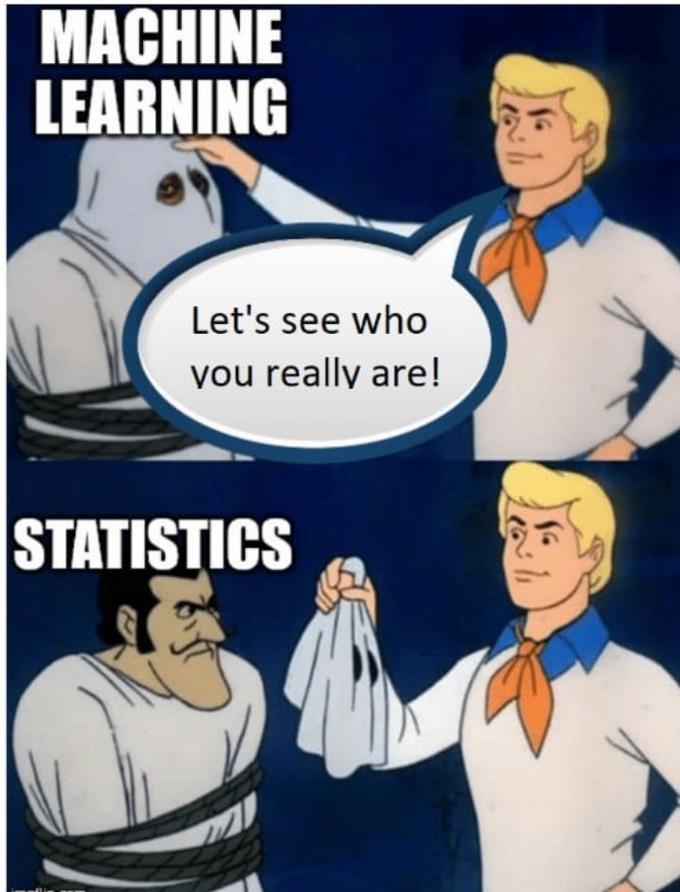


(A) Geotail (1994 - 2022)
>1 million points (~12s res)

(B) MMS (2015 – 2024)
~ 250k points (~12s res)



Output:
In principle, anything measured
(In this example plasma moments)



Metrics & Results

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Where,

\hat{y} – predicted value of y

\bar{y} – mean value of y

Quick reminder on model evaluation

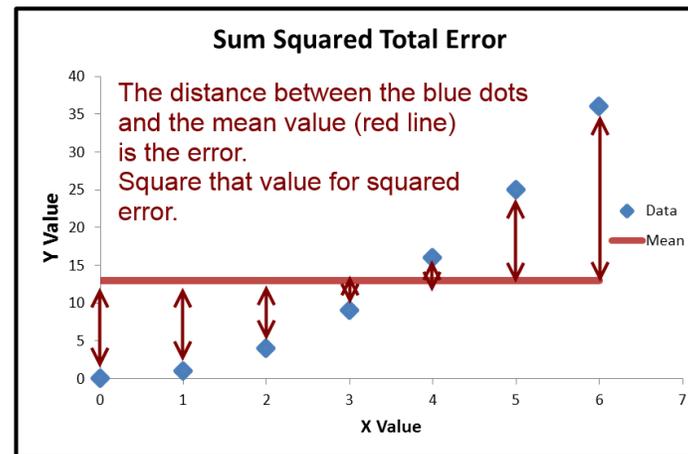
```
... 99/99 _____ 0s 808us/step
explained_variance: 0.019
median absolute error: 0.11
r2: -0.01
MAE: 0.157
MSE: 0.055
RMSE: 0.235
Cor: 0.533
```

A complex and intriguing model



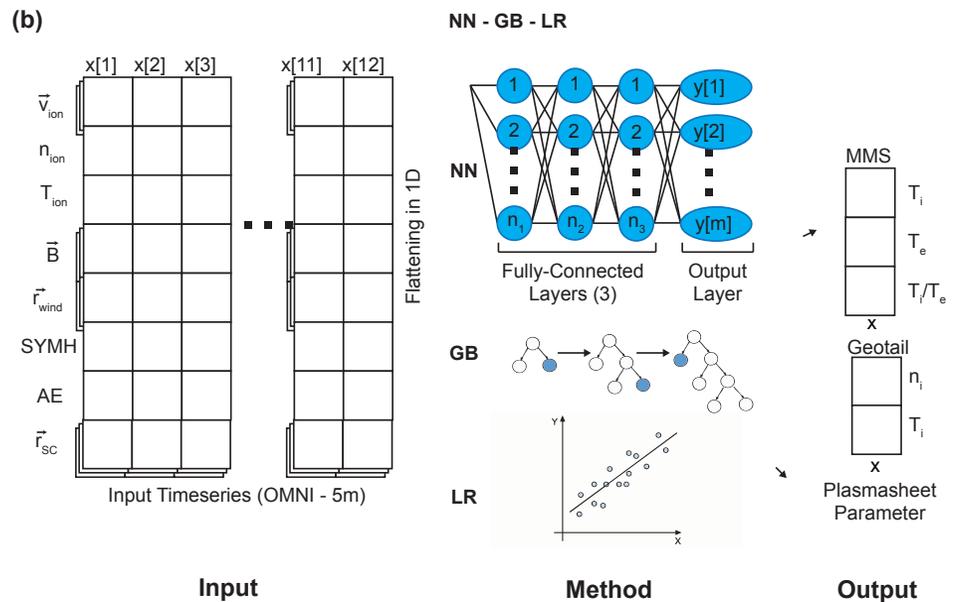
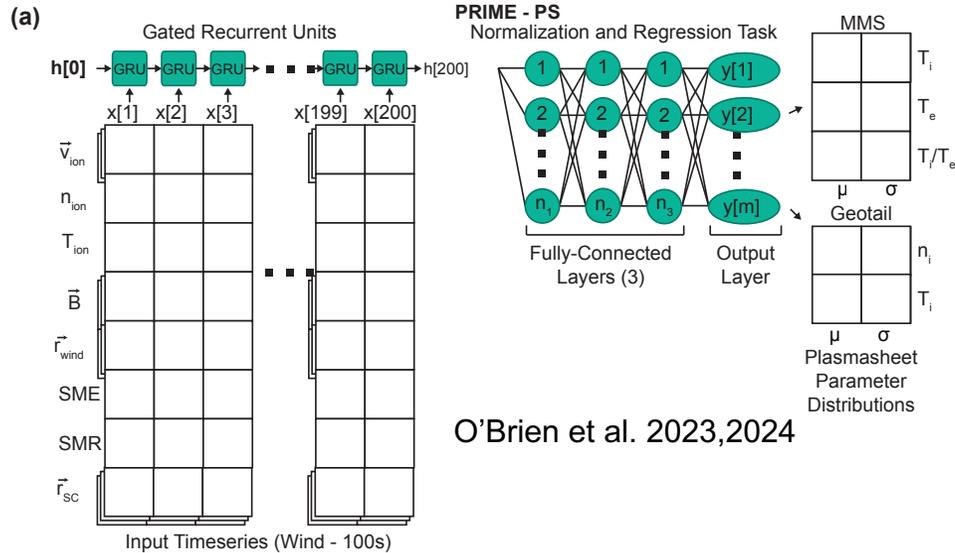
```
... 99/99 _____ 0s 778us/step
explained_variance: 0.0
median absolute error: 0.14
r2: 0.0
MAE: 0.17
MSE: 0.055
RMSE: 0.234
Cor: 0.0
```

np.mean()



When $R^2 < 0$, the horizontal line explains the data better than your model (i.e., mean of observed).

Methodologies & input space



- **PRIME Advantages:** Embedded uncertainty quantification and propagation from L1

Time History	Type of Input	Architectures
1-6h	Wind (L1)	Linear Reg
	OMNIweb	Gradient Boosting
		Neural Network
		RNN/LSTM/GRU (PRIME-PS)

Answering hypothetical questions:

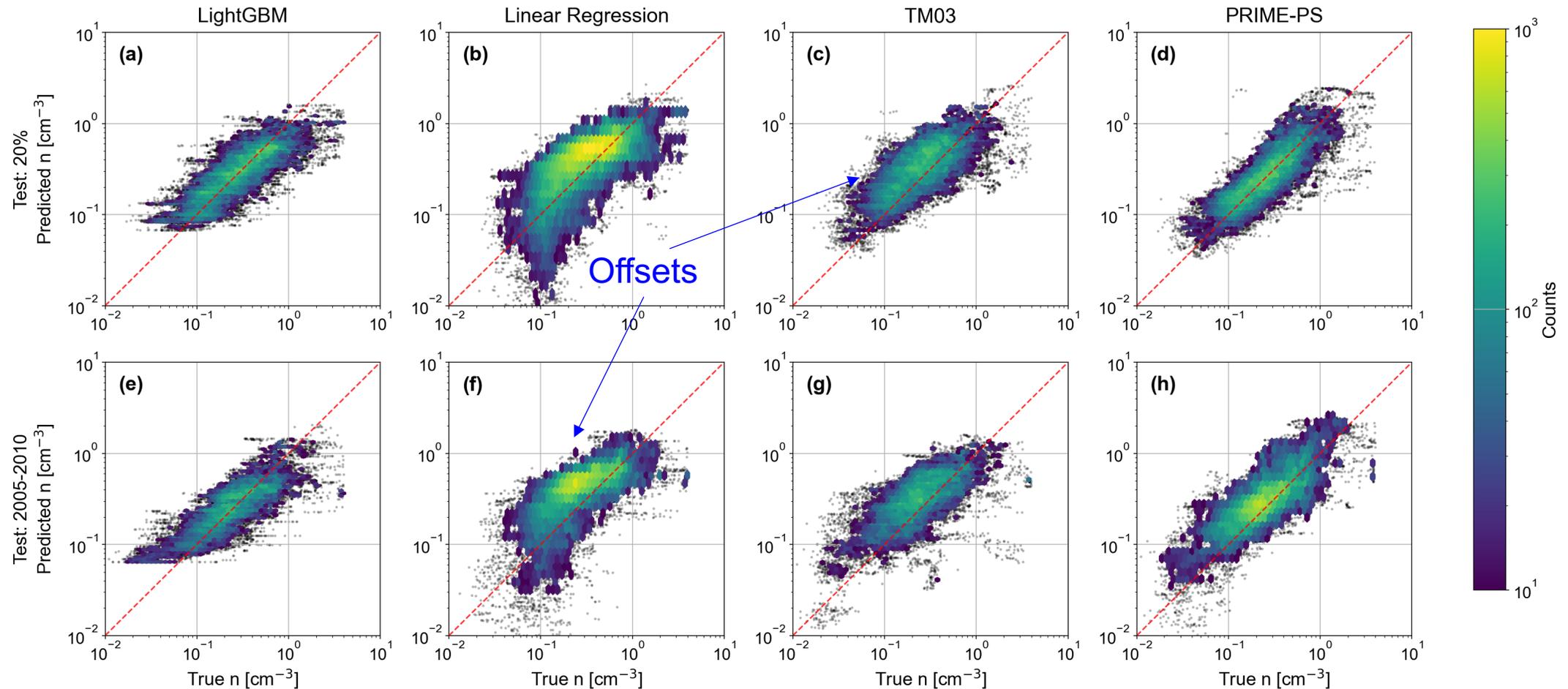
- ✓ Also tried different error functions, optimizers, hyperparameters etc.
- ✓ And different imbalanced learning techniques

Key Takeaway:
To quantify our method's impact, we tested multiple variations of the problem.

Modeling Density | Predictions vs Observations

Key Message: PRIME/GB > Baseline \approx TM03

Model Performance | Density (n)



Metrics using Test set (20% of data)

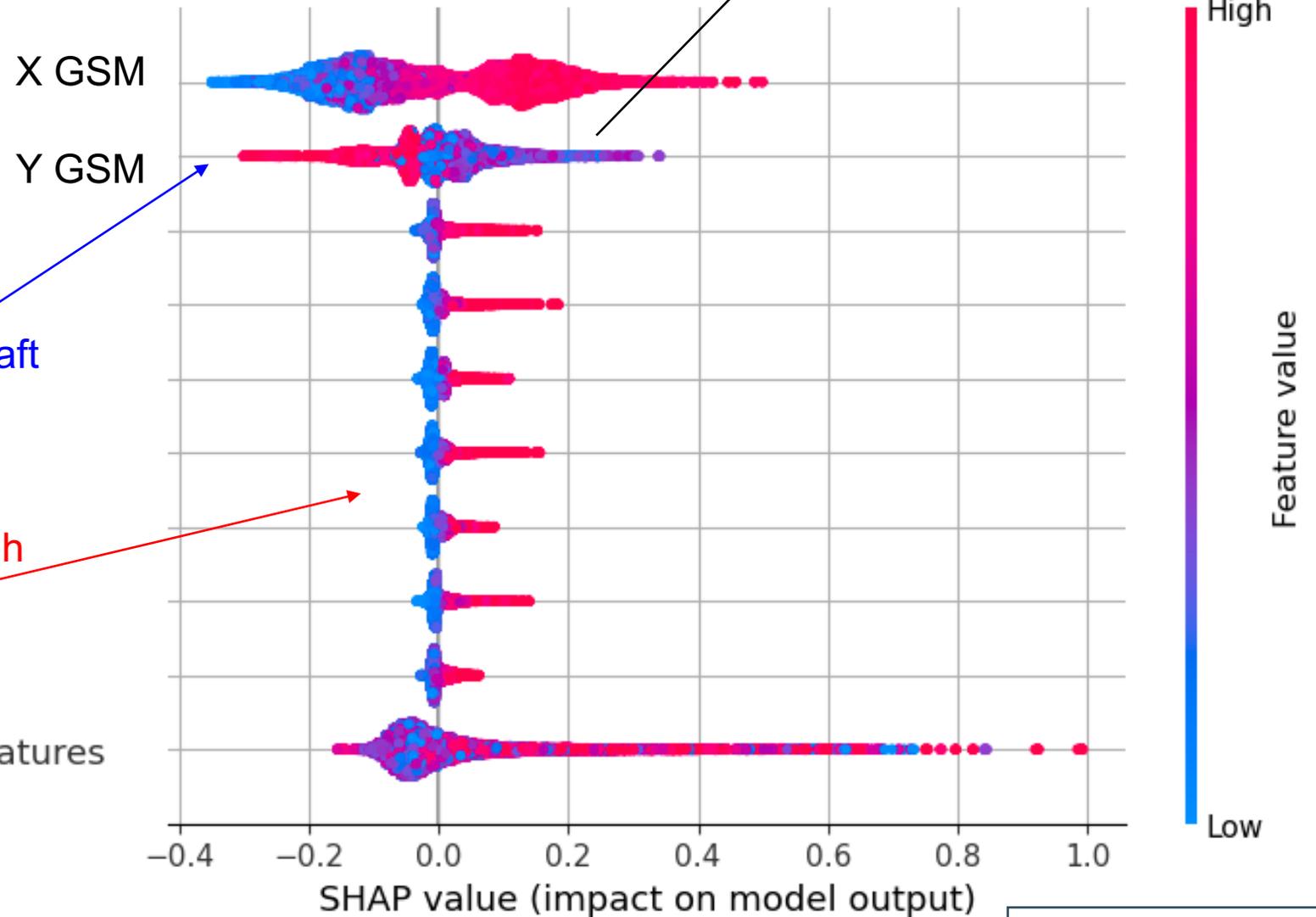
Method	MAE	R^2	r
LightGBM	0.129	0.373	0.631
Neural Net	0.152	0.325	0.603
Linear Reg	0.173	0.265	0.620
PRIME-PS	0.113	0.453	0.707
TM03	0.163	0.208	0.570

Key Results:

- PRIME-PS demonstrates a performance edge (~30% MAE from TM03 and ~15% from other ML approaches).
- This advantage can get relatively low (other train/test splits & crossvalidation).
- Different input, method, time-history, and hyperparameter tuning etc. had overall a statistically marginal effect. Why?
- Since PRIME-PS was statistically better, and Gradient boosting can't be used for modeling, we only keep this for next parts of analysis.

Feature Importance Analysis

Higher density close to earth and at dawn



Answer: In most cases (statistically):

Model is predominantly driven by spacecraft location

Solar wind input has lower effect, although cumulative history is still important

Sum of 85 other features

Blue/Red: Input value
Left/Right: Output value

SHAP Values explain why a model made a specific prediction, by showing each feature's impact.

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.

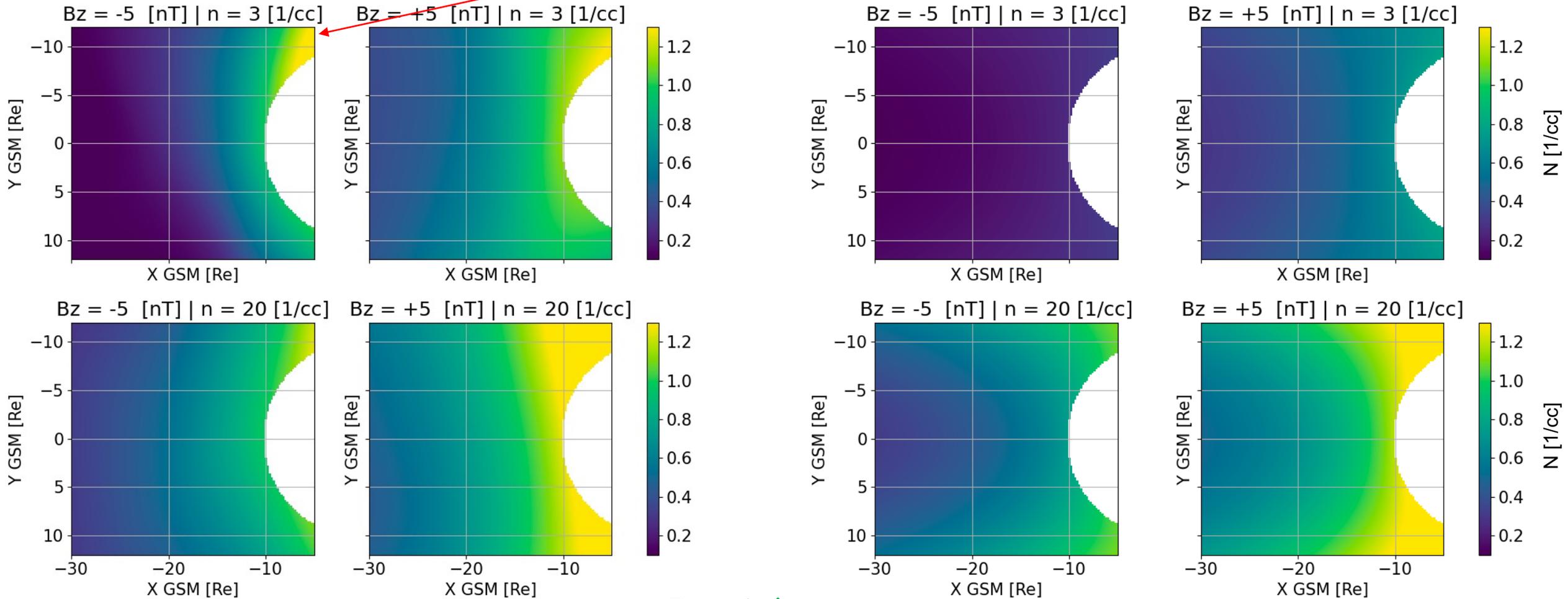


Actually pretty accurate

Modeling Efforts

Modeling Density | 2D Maps (Synthetic input)

Assymetries Introduced



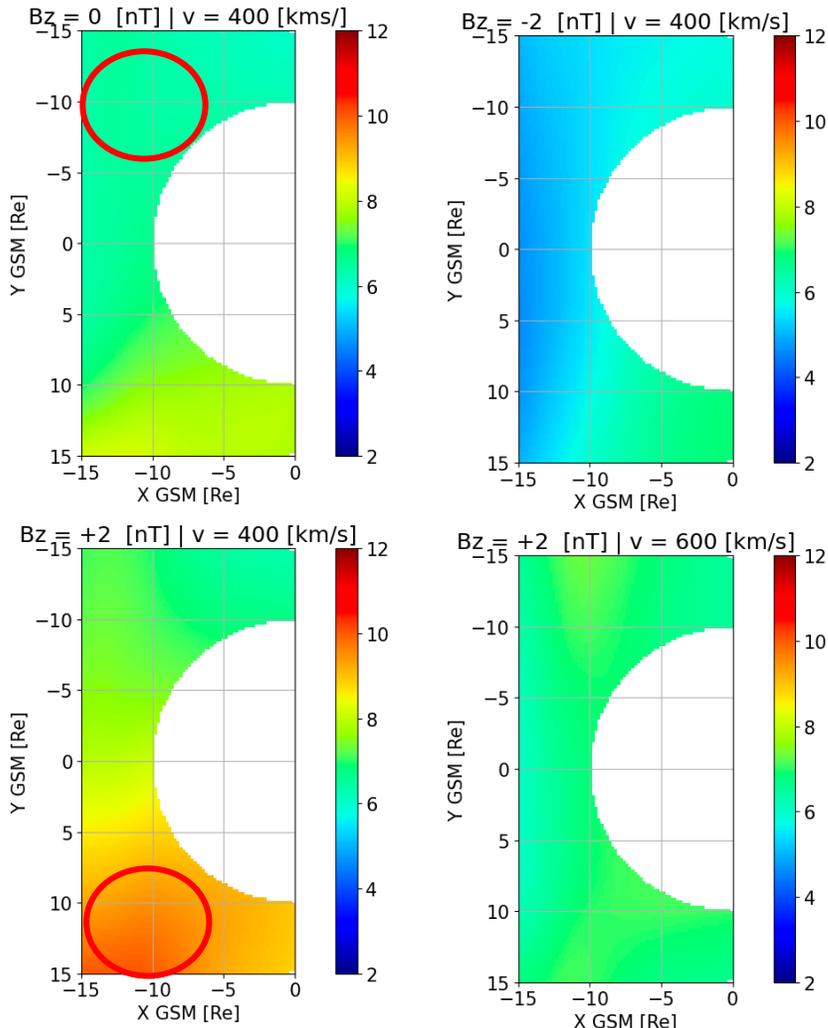
Neural Networks modeling

Empirical modeling (TM03)

$B_z > 0 \uparrow n_{ps}$
 $n_{sw} \rightarrow \uparrow n_{ps}$

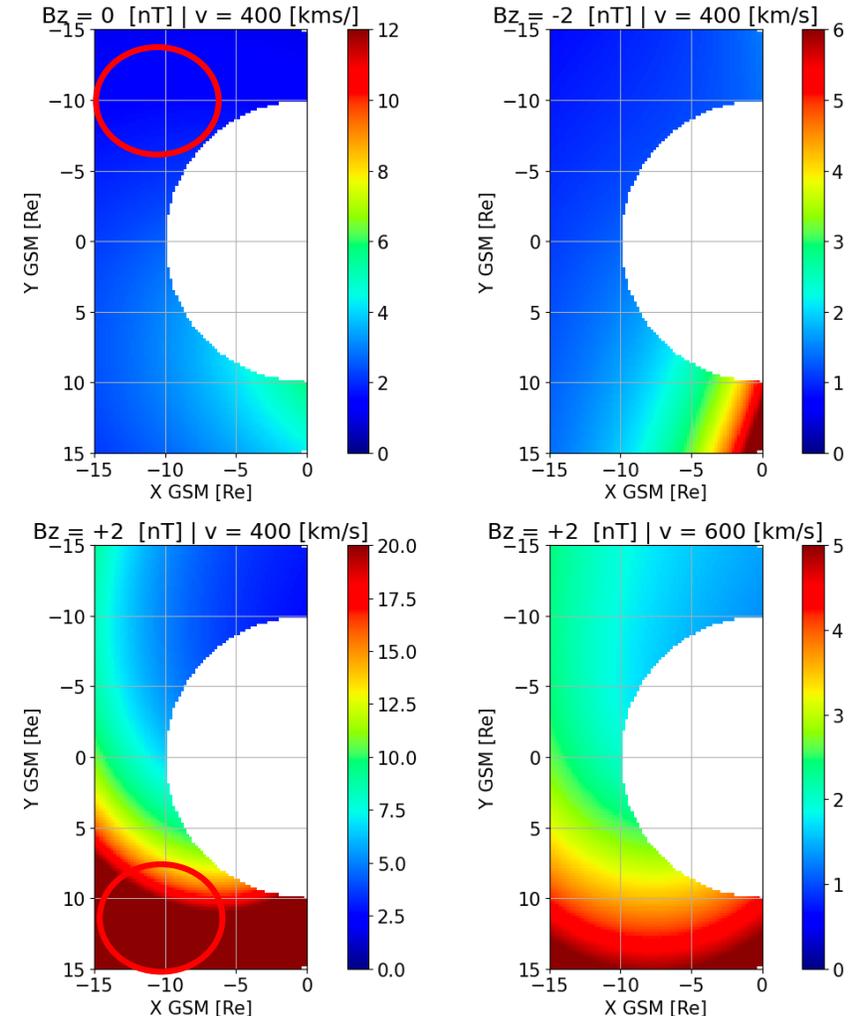
Modeling Temperature Ratios with MMS | 2D Maps

Note: Agreement Wang et al., 2010 with dusk Ti/Te higher than dawn (Using THEMIS)



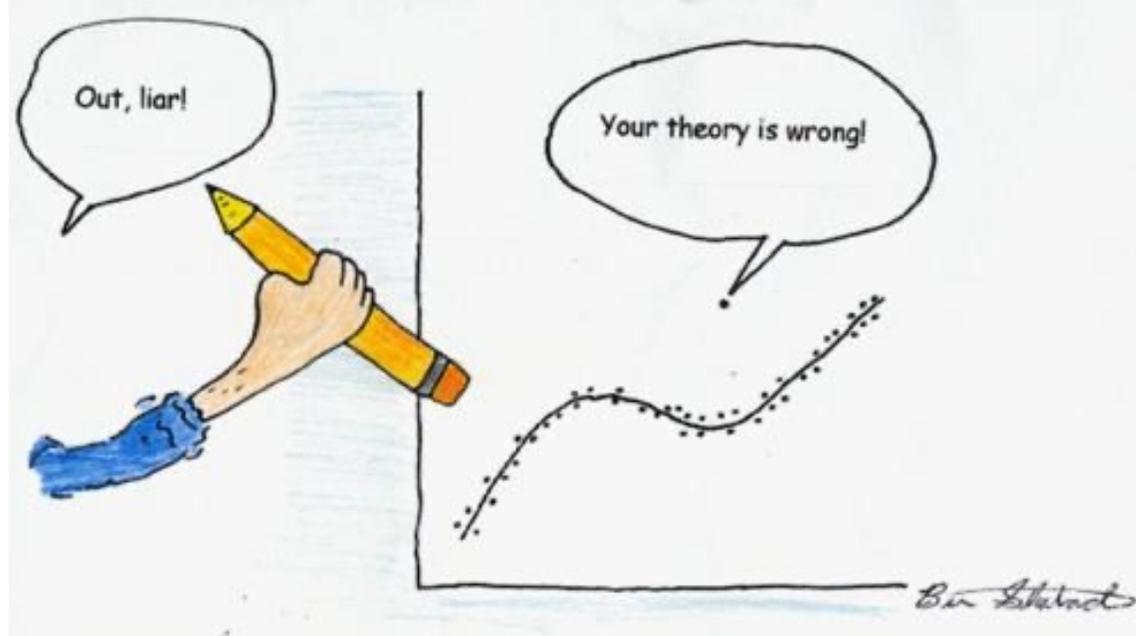
Pros:
+No extreme values
+Asymmetries shown
+ Coherent physical picture

Cons:
- Not easily available analytical form



Neural Networks modeling

Empirical modeling (TM03/DSGR16)



Storm Time Behavior and Importance of Outliers

The Problem: We use [static thresholds](#) for dynamic environments.

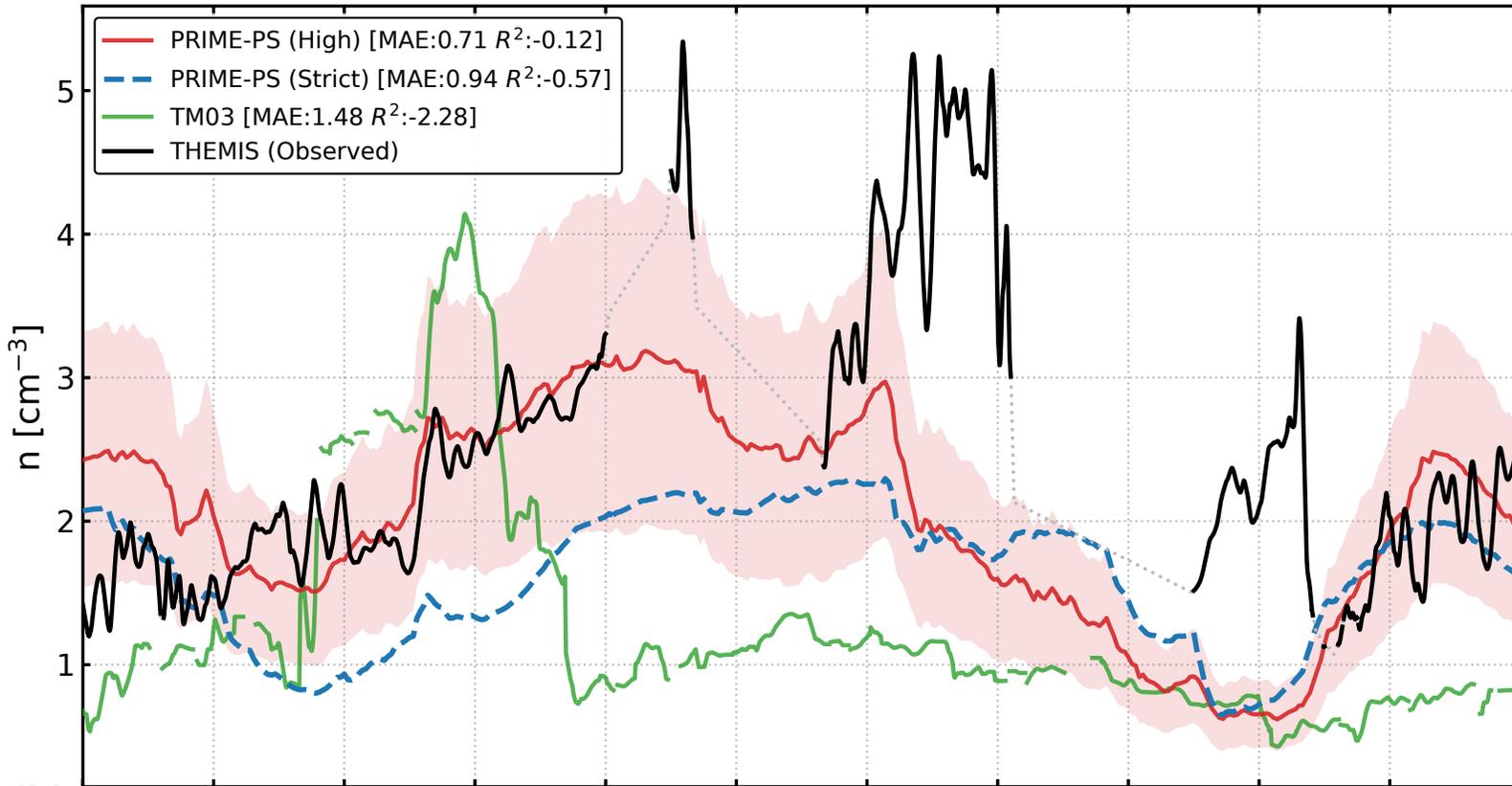
The Risk: Therefore we can [mistakenly remove the crucial "stormtime plasmashet."](#)

The "Horrible yet Practical Solution": [Manually find the missing data and add it to the dataset.](#)

Strict CPS (e.g., Ohtani et al., 2008 Raptis et al., 2024) & Flexible CPS (e.g., Richard et al., 2022)

Test case of a storm (05 Nov 2023)

(a)



Process:

High: Includes manually picked high-density intervals from Geotail

Strict: Normal threshold-based classification of plasma sheet

Test:

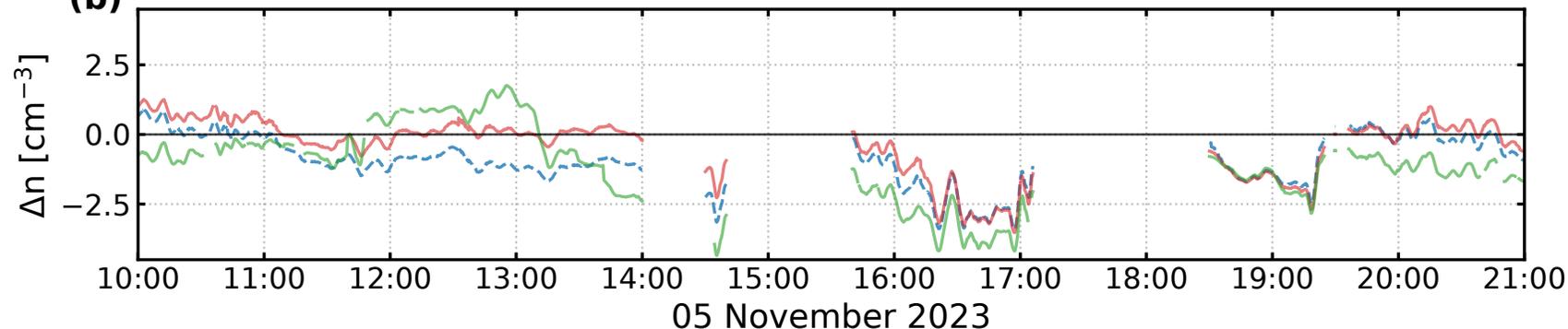
THEMIS observations

Results

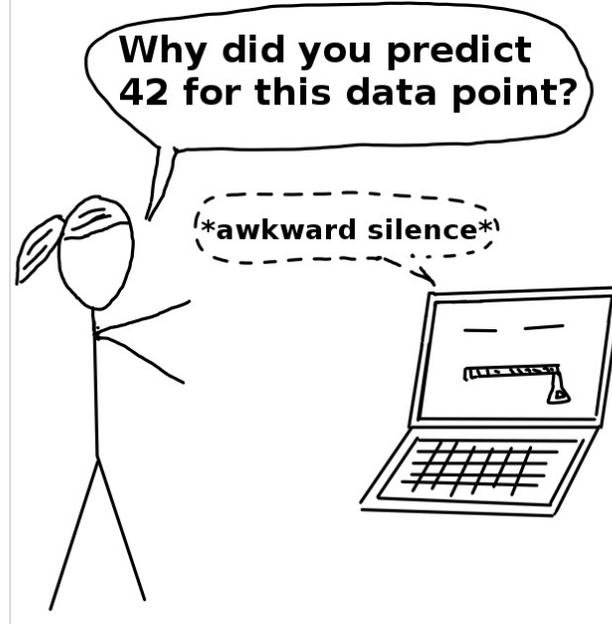
- ML model (high dens): 0.7 [1/cc]
- ML model (normal): 0.94 [1/cc]
- TM03: 1.48 [1/cc]

Key Message: >50% improvement

(b)



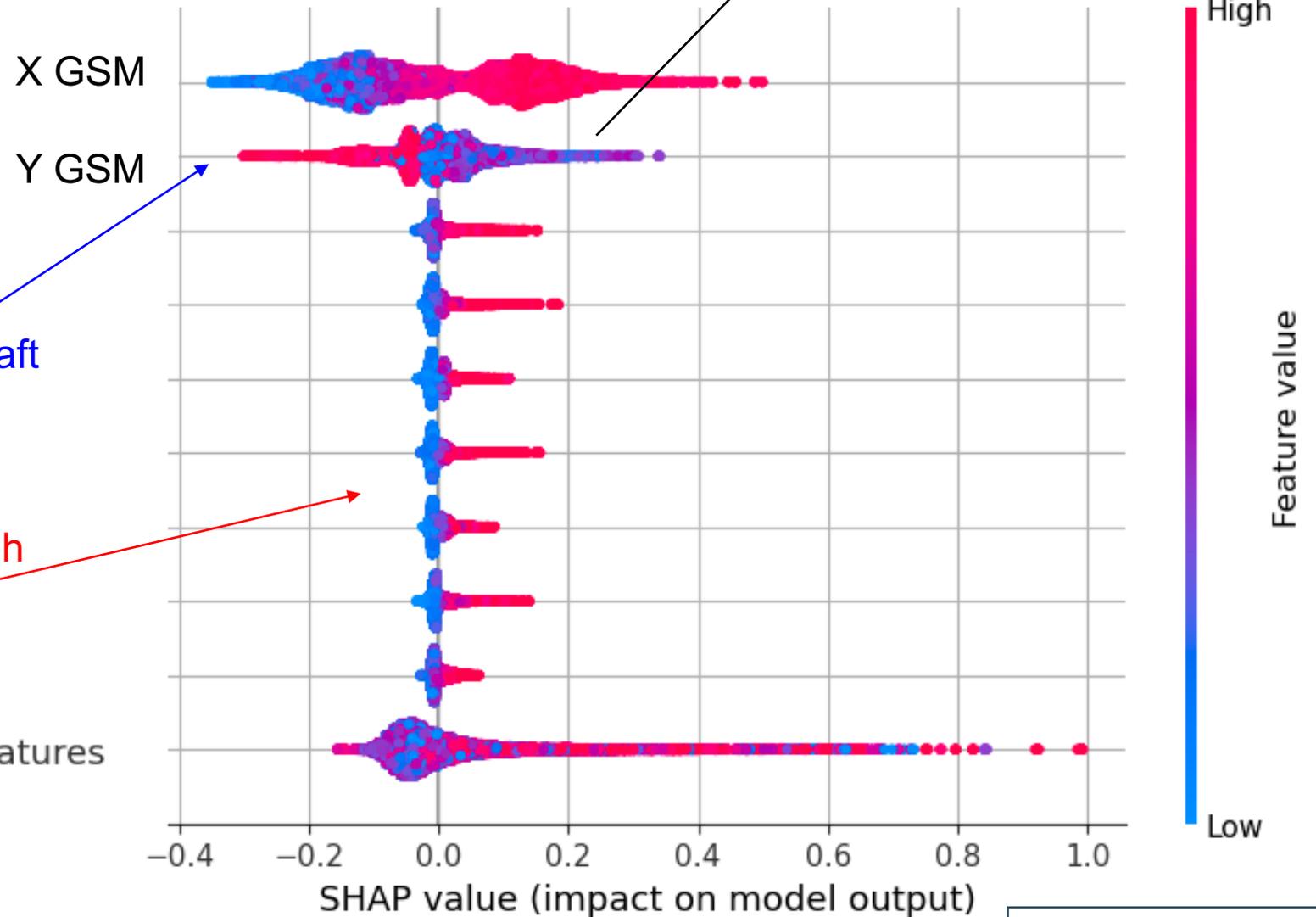
Note: values $<1 \text{ cm}^{-3}$, are boundary layer transitions, gray shaded linear interpolation



Preliminary & Simplistic Feature analysis

Feature Importance Analysis

Higher density close to earth and at dawn



Answer: In most cases (statistically):

Model is predominantly driven by spacecraft location

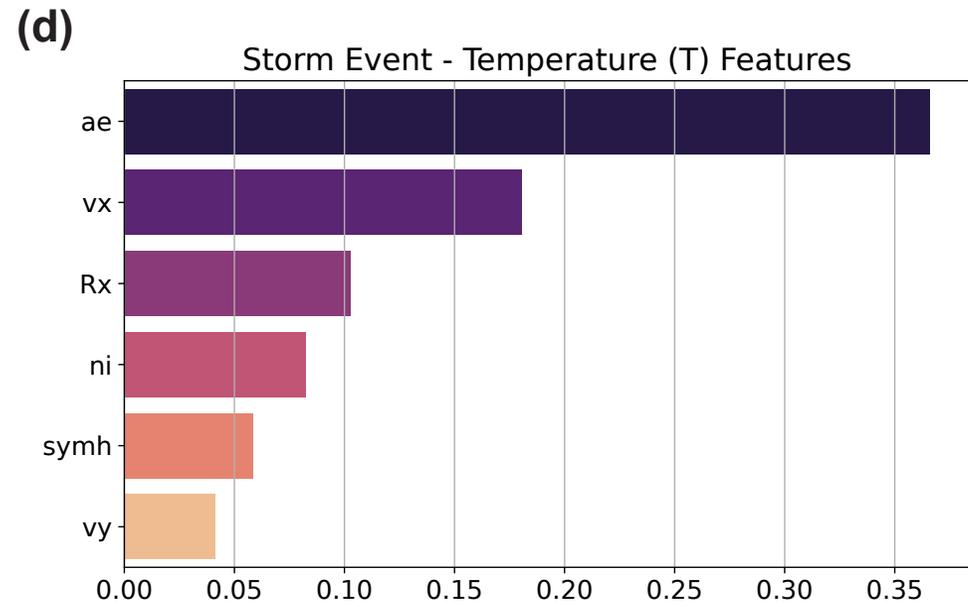
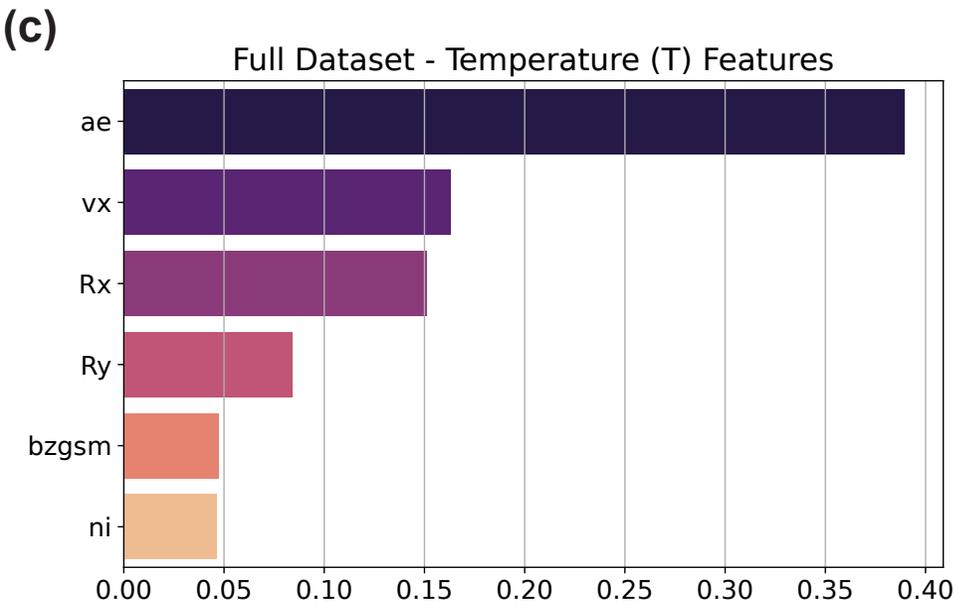
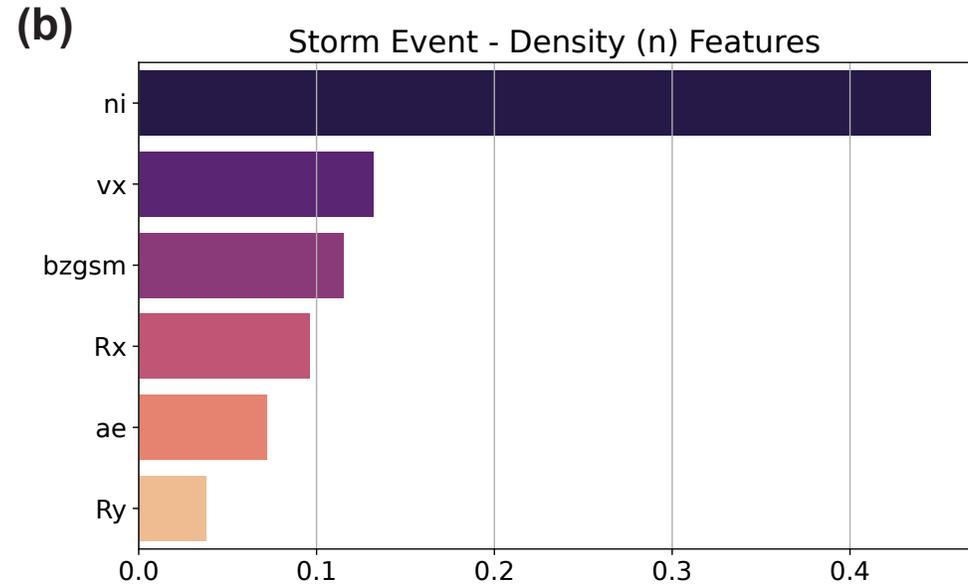
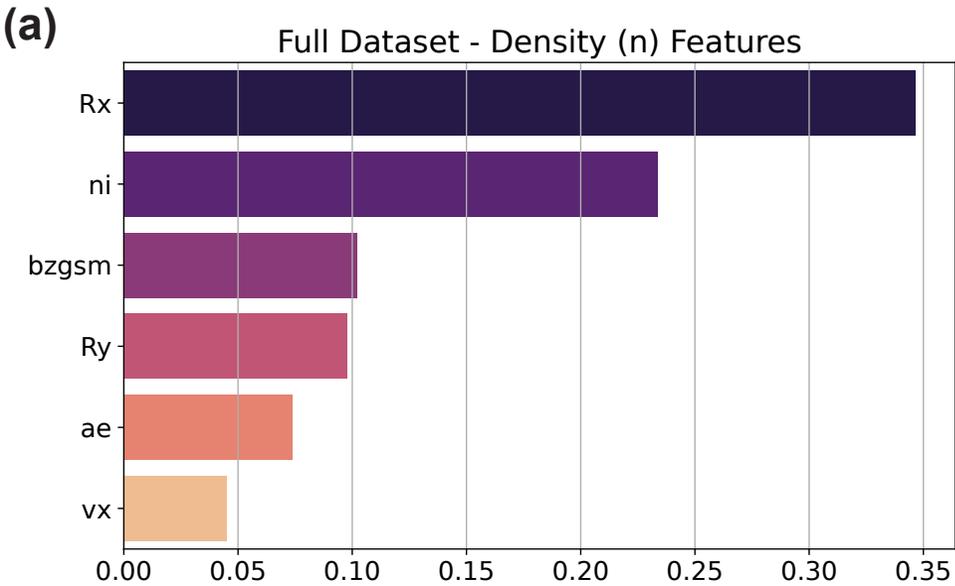
Solar wind input has lower effect, although cumulative history is still important

Sum of 85 other features

Blue/Red: Input value
Left/Right: Output value

SHAP Values explain why a model made a specific prediction, by showing each feature's impact.

Feature Importance Analysis – Merged Time history



Quiet times

- More internally driven
- T is more external, maybe
- Density is largely dictated by location

Storms:

- More externally driven
- Location matters less

Issues with this analysis:

- Intercorrelation between variables is large (ae well correlated to vx)
- Same with SW types (dense solar wind is typically slower)
- Combines all time history (1h in this particular example)

Part #2 (i.e., ML model) Summary

Results

✓ **Significant Gains:** PRIME-PS and ML models outperform analytical methods & show asymmetries.

✗ **Mediocre Storm Predictability:** We mainly capture "boring" conditions, not the rare events.

🧠 **A Core Problem:** Our training data is biased. Extreme events, are not always captured by simple threshold, and including them is methodologically challenging

Future Work

- **Understand the output:** Focus on feature importance under different conditions.
- **Hybrid Modeling:** Use simulations to generate extreme events we lack in data? (Need to be careful here).

Discussion Points/Opinions

- A model with $R^2 \sim 0$ can have a correlation of 0.7 and very low MAE depending on the problem.
- “Unique and extreme” events with parameter distribution can be more important than typical metrics.
- “Better” data can yield ~30% difference, while a better model improved metrics by ~40% (better measurements are needed for ground truth)
- Reminder GIGO applies to ML as much as physics models https://en.wikipedia.org/wiki/Garbage_in,_garbage_out

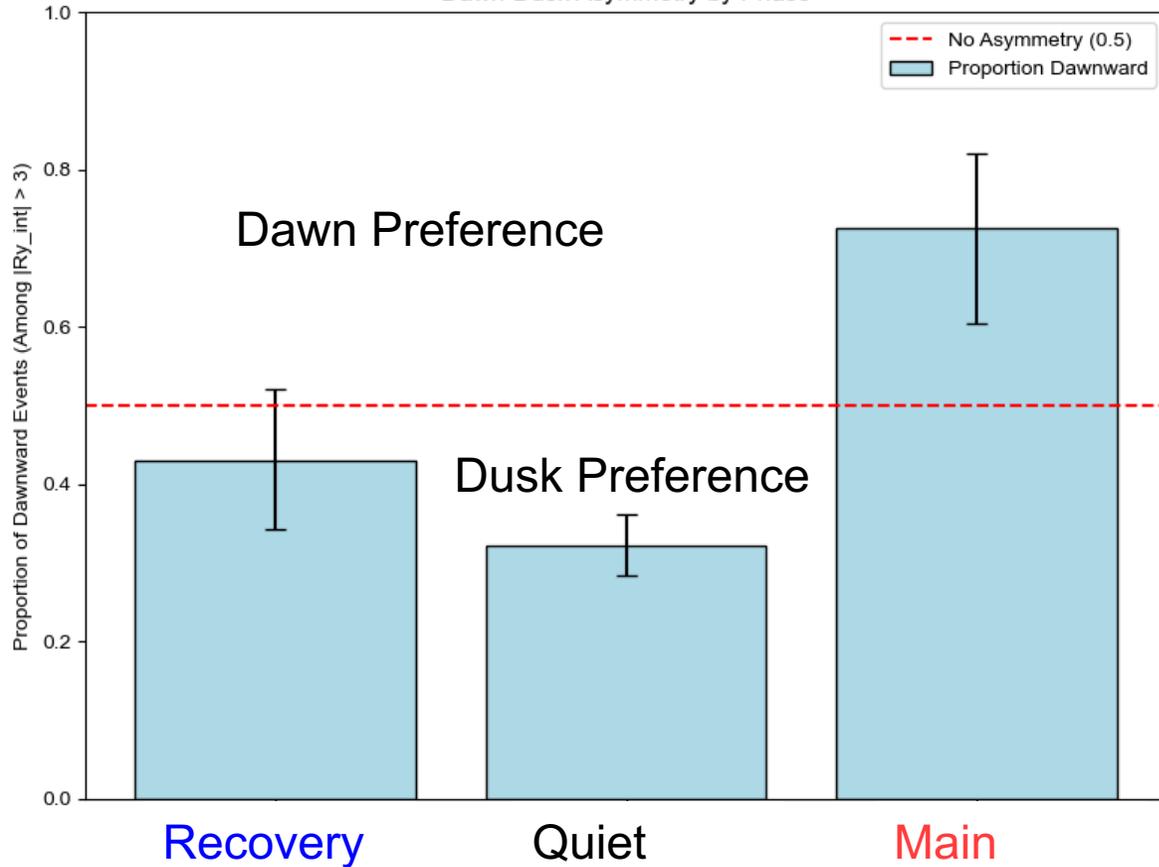


Future work Ideas Working with Modelers



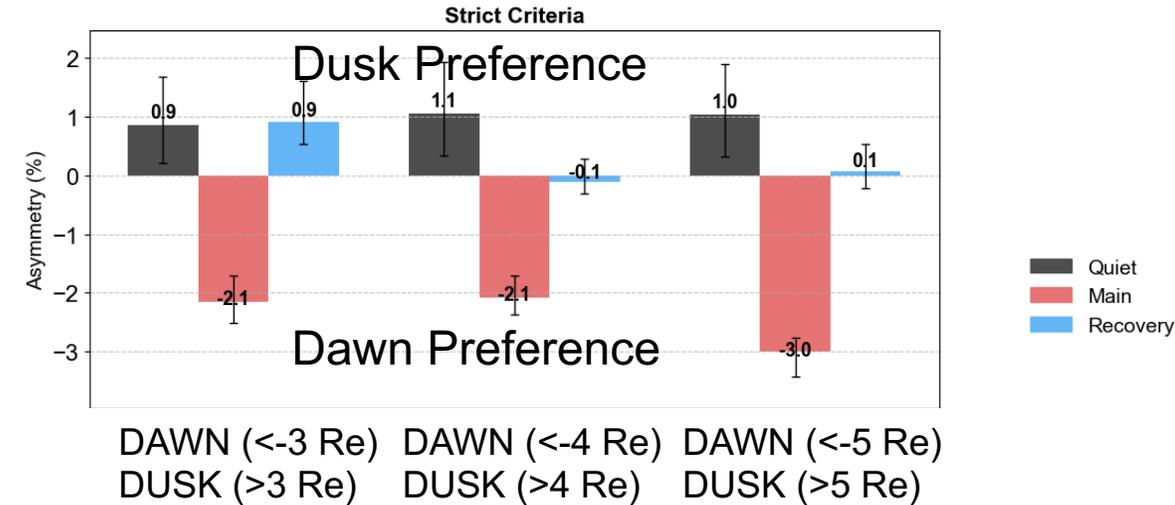
Dawn/Dusk Asymmetries - Reconnection and BBFs

Geotail dataset from Nagai+2023
Dawn-Dusk Asymmetry by Phase



Geotail dataset BBF list (Devandan+2026)

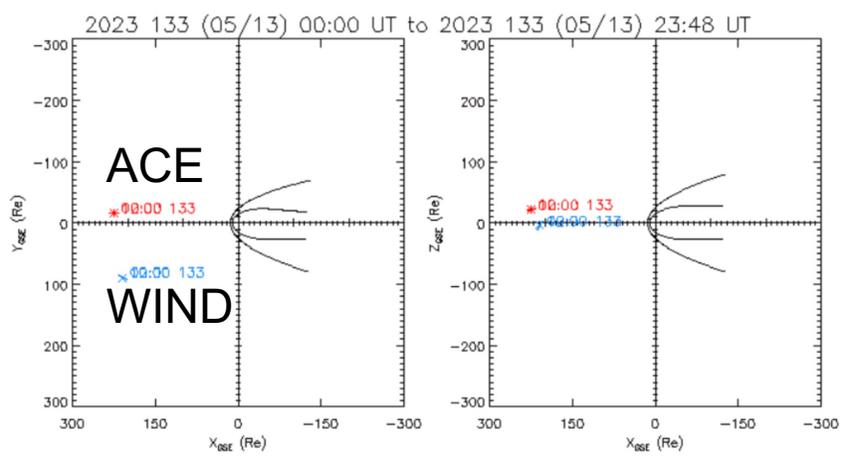
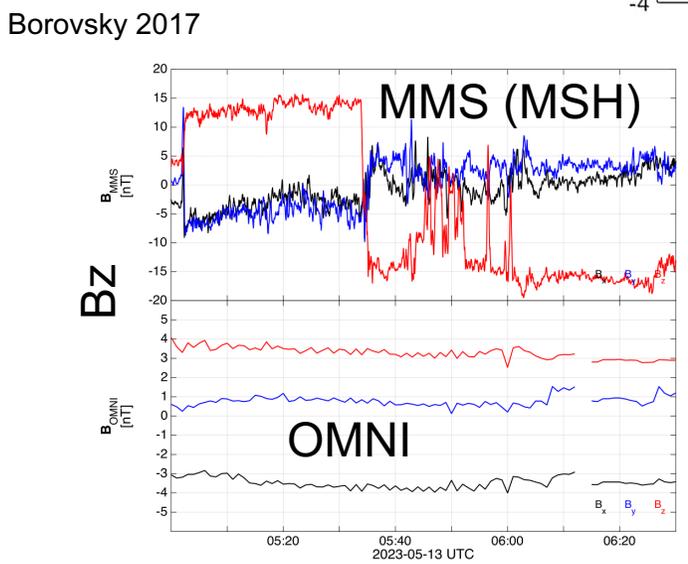
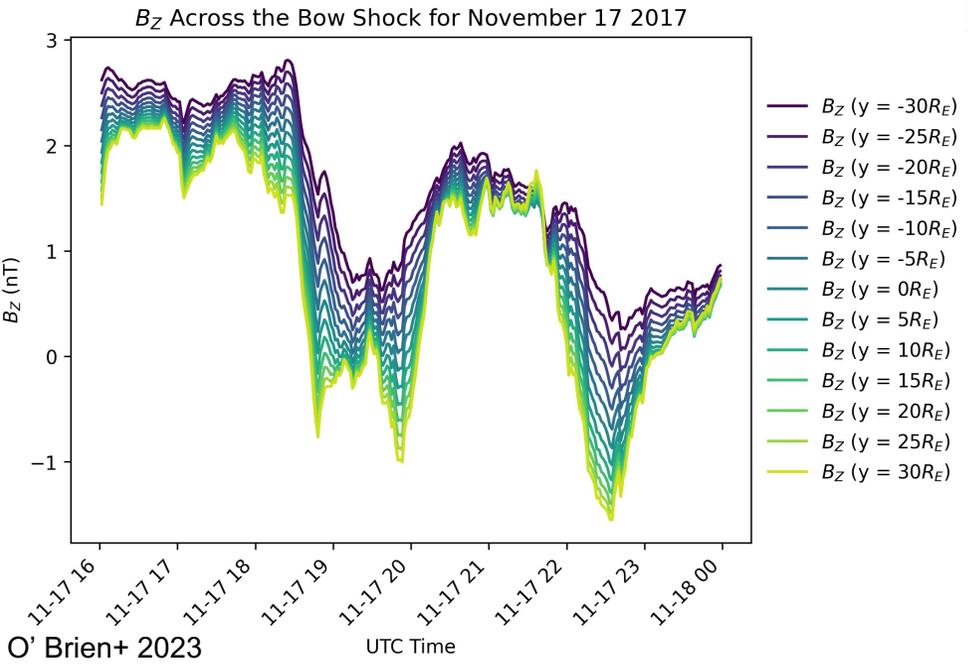
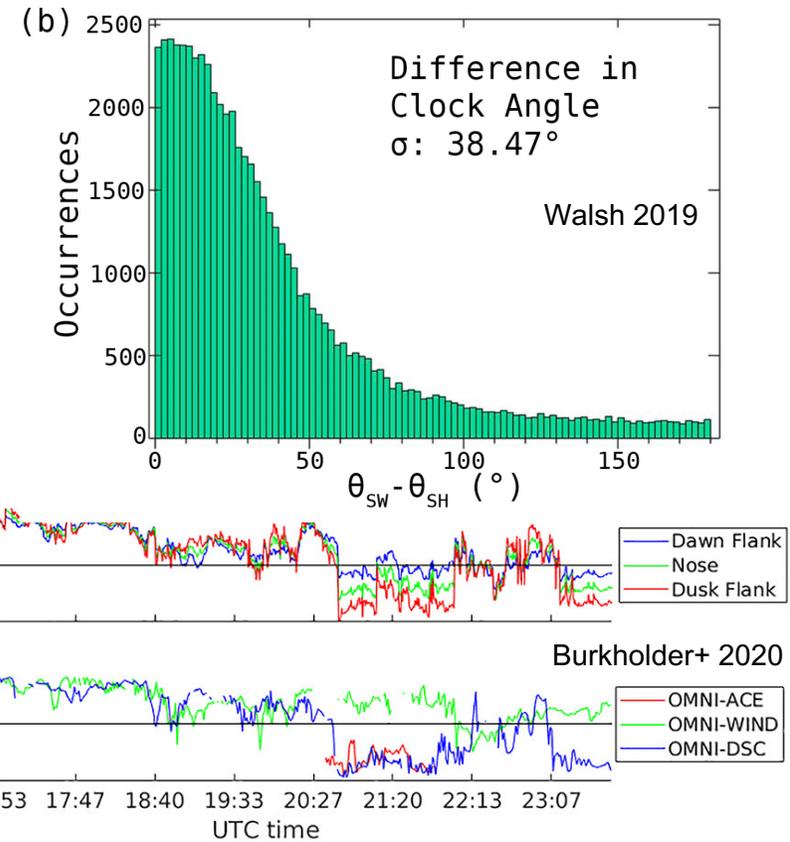
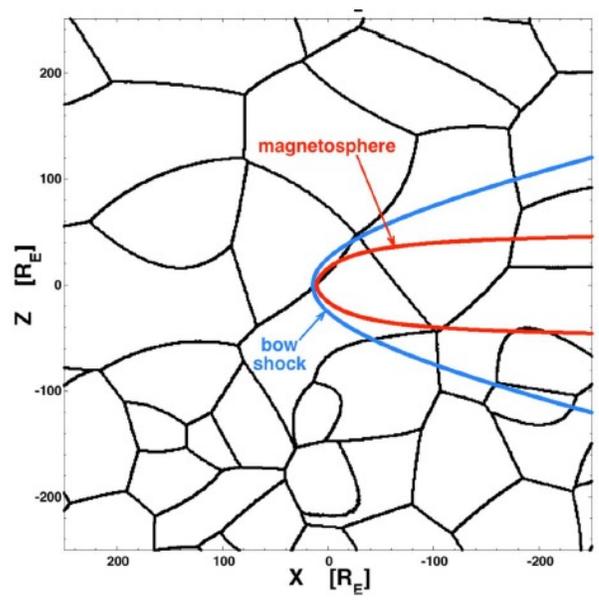
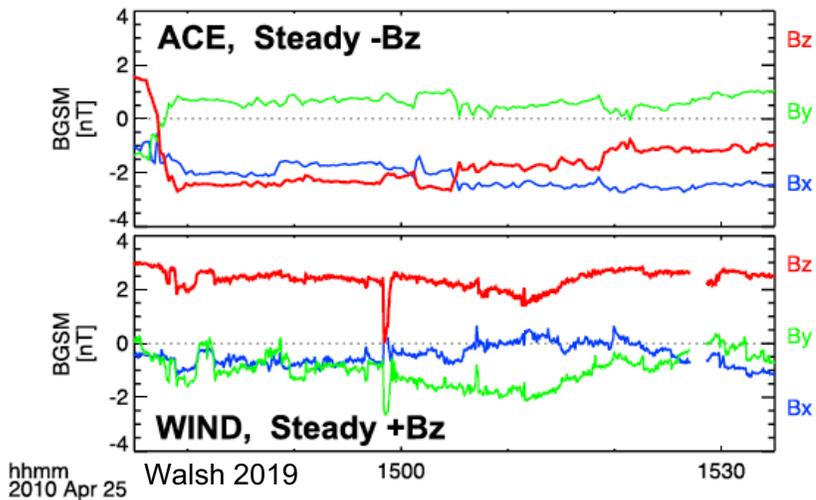
$$\text{Asymmetry} = \frac{\text{Occurrence Dusk}}{\text{Occurrence Dawn}}$$

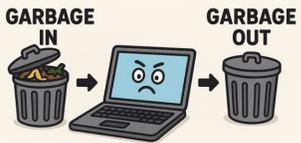


Key Point:
Asymmetries in occurrence during Main phase are shown in both BBF and Reconnection list datasets

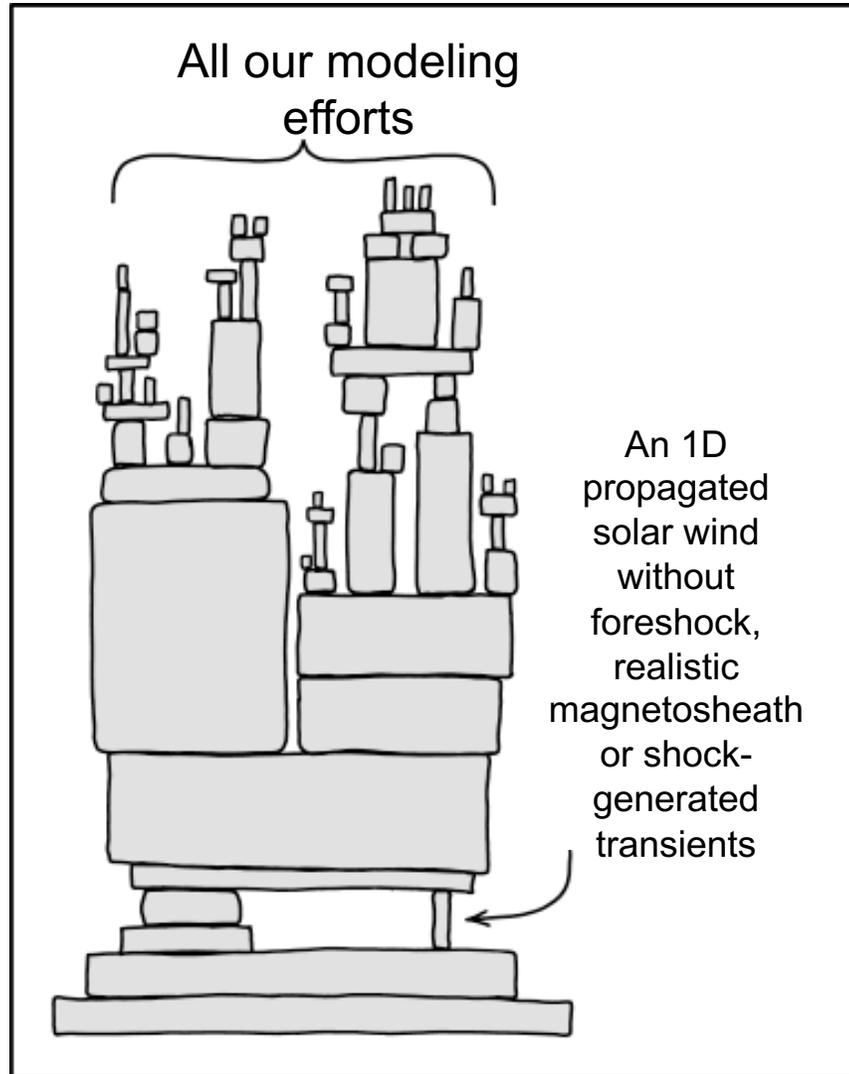
- TODO:**
- **Observational** analysis including **MMS** and **THEMIS** statistics
 - **Compare to simulation intervals** of storm and non-storm time

Spatial and Temporal Solar Wind Variability





What are we dealing with?



Two challenges:

Solar wind information limitations:

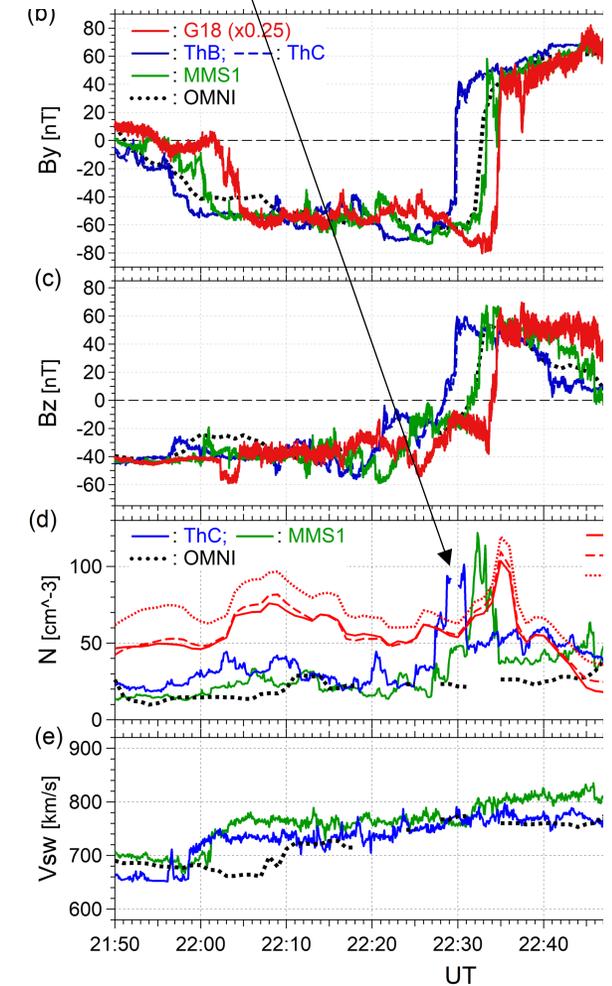
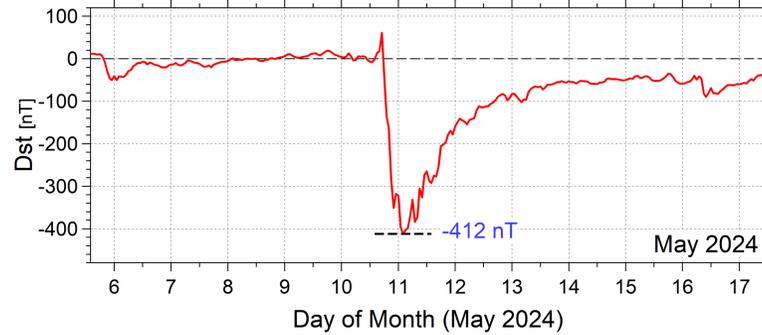
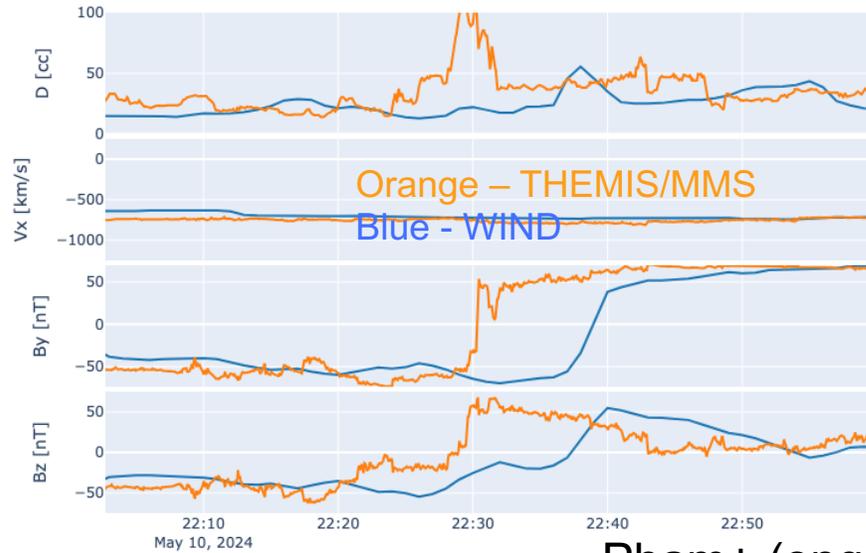
- Reality has complex 3D structure and spatial variability.
- It is easier to rely on simple 1D picture that is available than do the extra effort.

Neglected foreshock and magnetosheath transients:

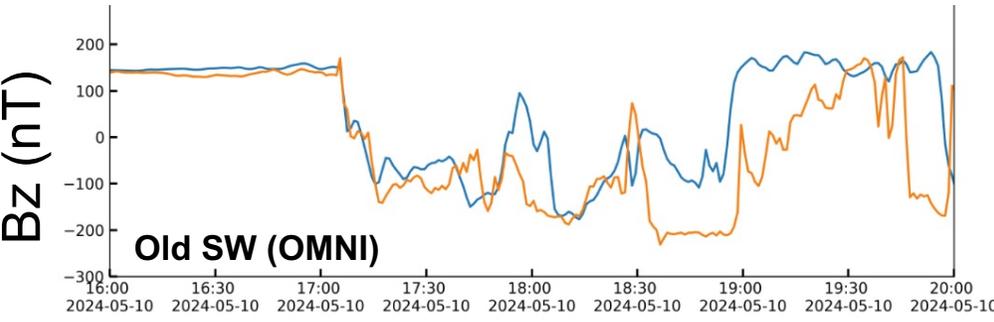
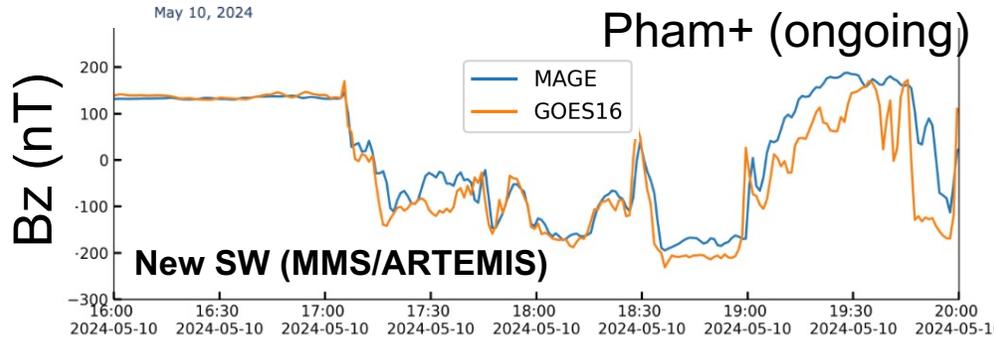
- These transients can affect magnetopause reconnection, and more.
- They are often omitted because they are difficult to include.

Examples from Gannon Storm

This density enhancement may play a crucial role



Ohtani+ (2025)

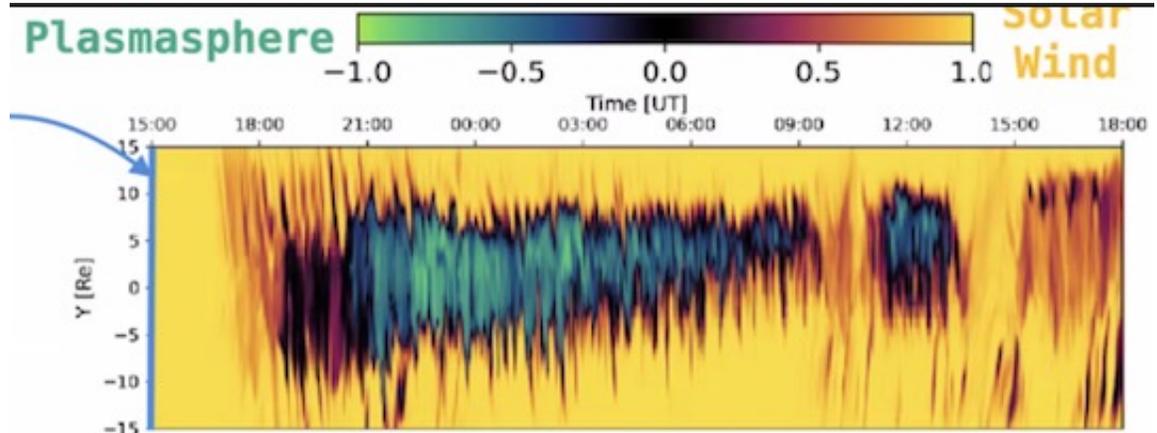
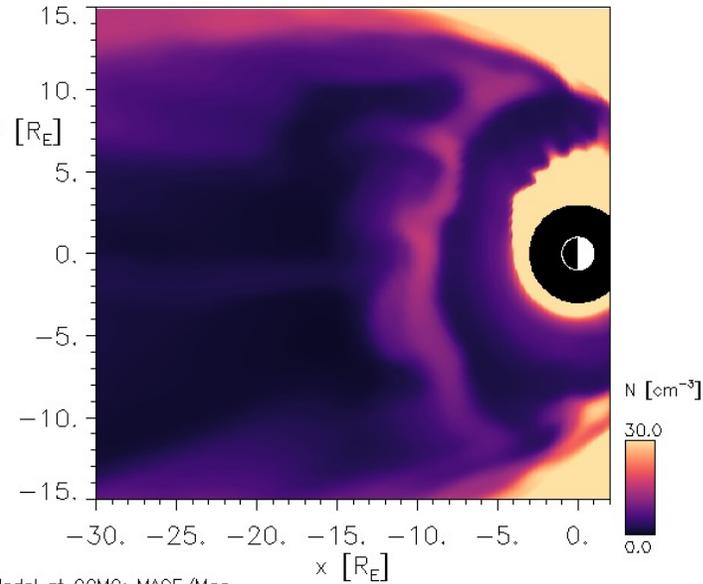


- Introduced local upstream (~10 min) transient variability from local in-situ observations
- Data-model agreement increased drastically
- Physical interpretation of ground data changed significantly

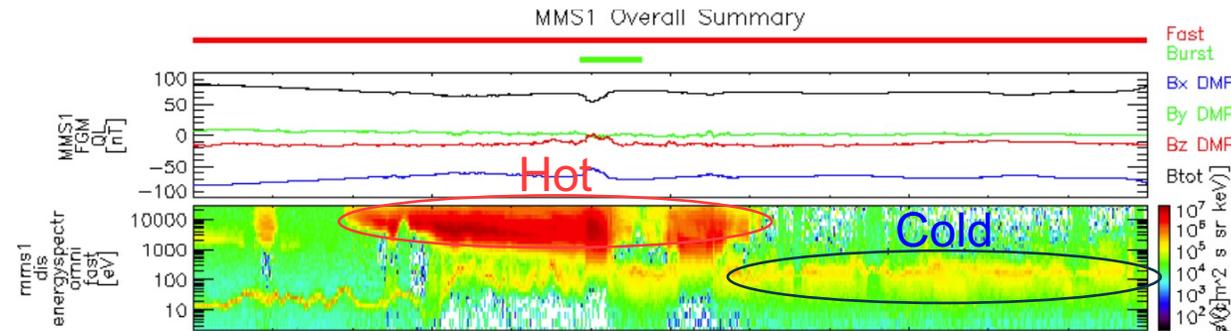
Origin of high-Density plasmashet intervals

CCMC run showing whole plasmashet with high density

11/05/2023 Time = 13:44:00 UT z= 0.000R_E

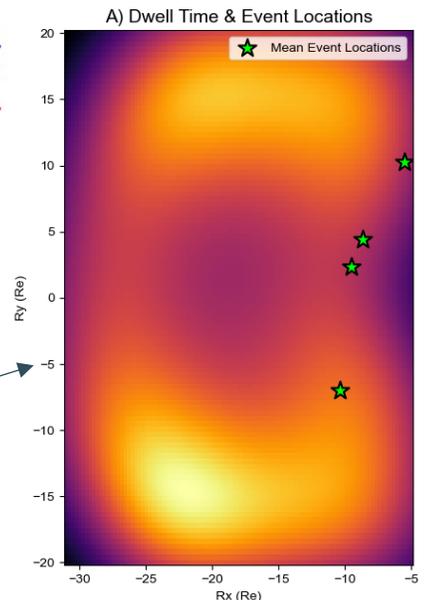


Screenshot from Kareem's presentation



MMS data during a storm, localize density peaks ($n > 4$ [1/cc]) at $x < 11R_E$

Geotail events of high density events during storm main phases



Todo: Calculate partial moments from MMS, THEMIS, Geotail during these events to evaluate presence of secondary population and compare with MAGE

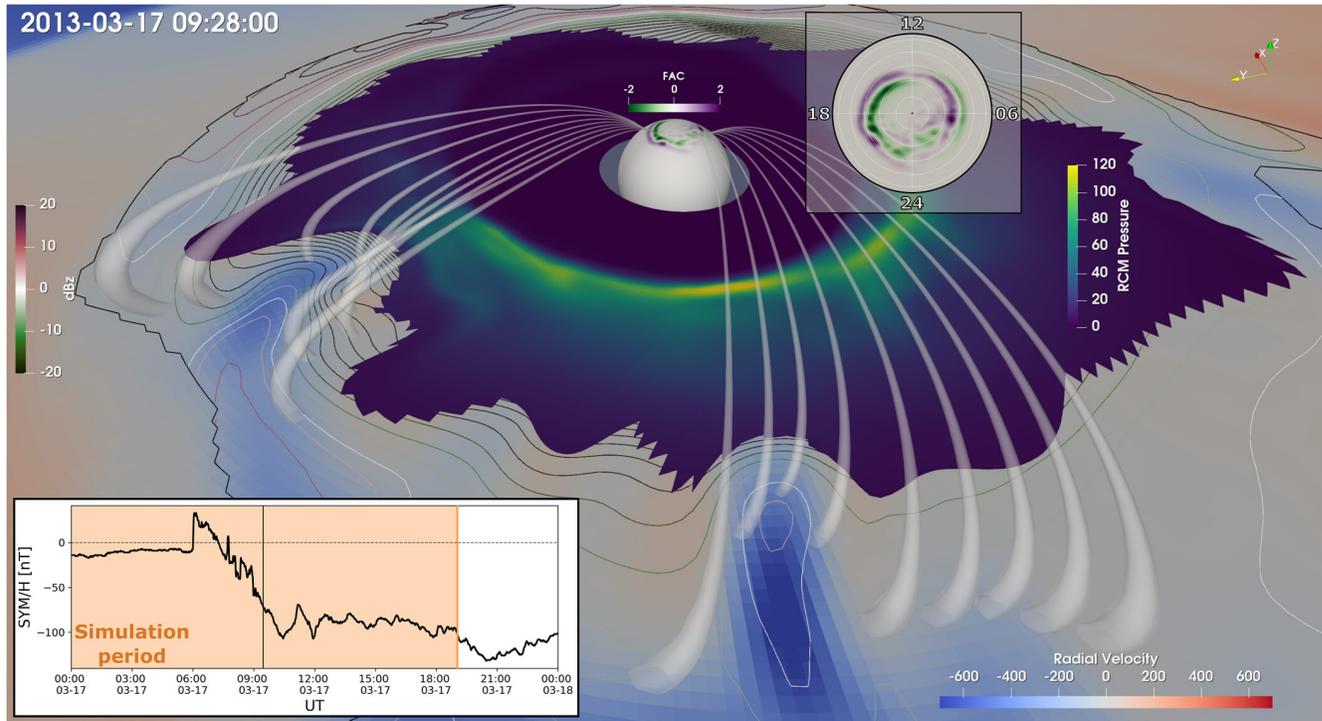
Machine learning & *Data-Model “Fusion”*

1. **Proper train/testing**, we need to make sure models generalize (magnetosphere community is not doing great here)
2. Avoid data modeling based on other in-situ quantities (*non-unique solutions*)
3. Providing incomplete physics (e.g., ideal MHD) can bias data-model fusion models towards wrong solutions rather the other way around
4. **Residual modeling**: Train ML to predict the simulation error against data (residuals) rather than the full state (i.e., the correction, can be great for storm specific modules)
5. Investigate more PINNS and data assimilation (i.e., what people mean with data-model fusion)
6. Use techniques that are easy to implement algorithmically and to interpret (e.g., Symbolic Regression)
7. Reduce in-situ observations to only these that are useful for their task

... Original dataset size: 241311
Reduced dataset size: 6978

Extras

General Context & Motivation



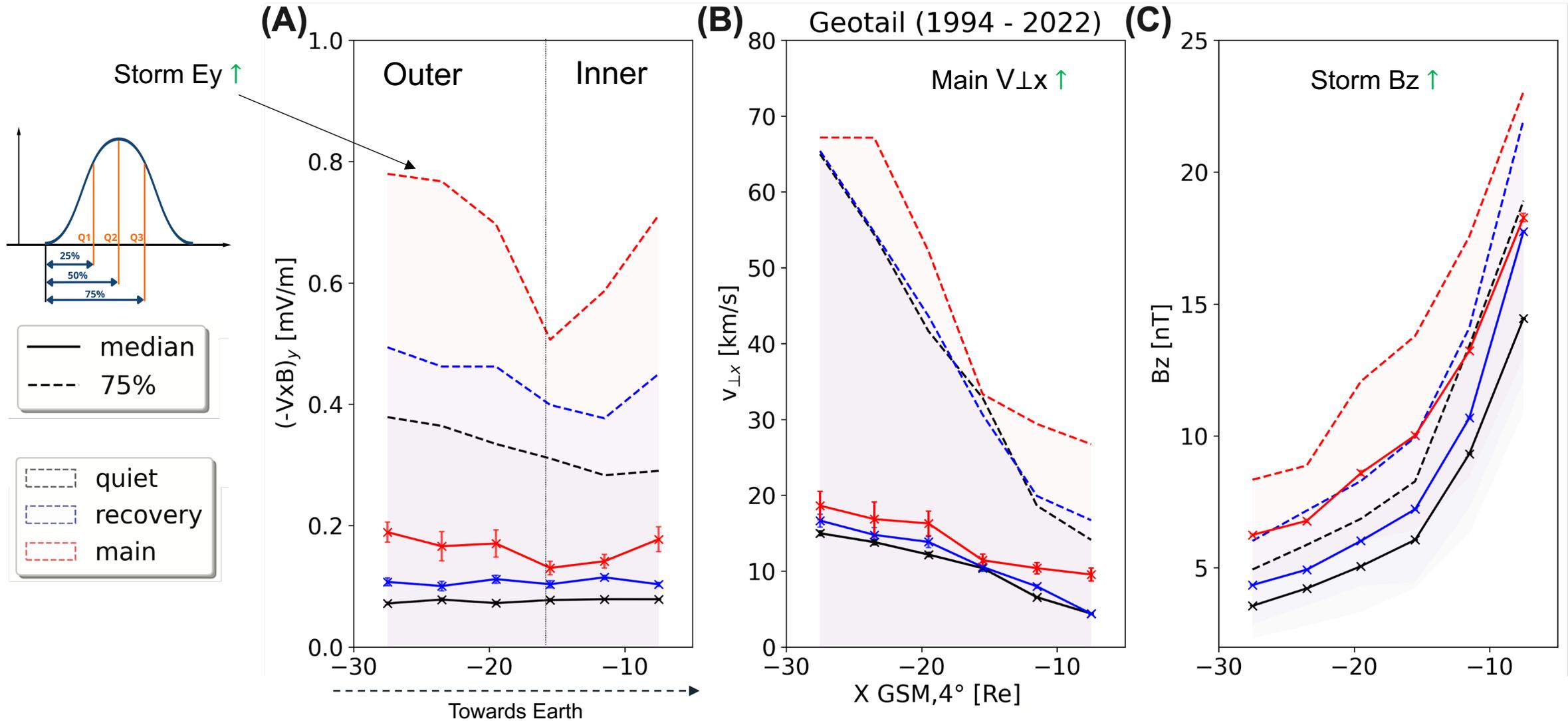
One of CGS objectives:
The role of **mesoscale plasma sheet** transport in
the **ring current** build-up

- To tackle this **we need** to establish a **clear understanding** of the overall **plasma sheet transport** during **quiet and storm times**.

Sciola+ 2023

50% of total energy flux transported into the inner magnetosphere by mesoscale structures

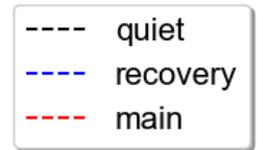
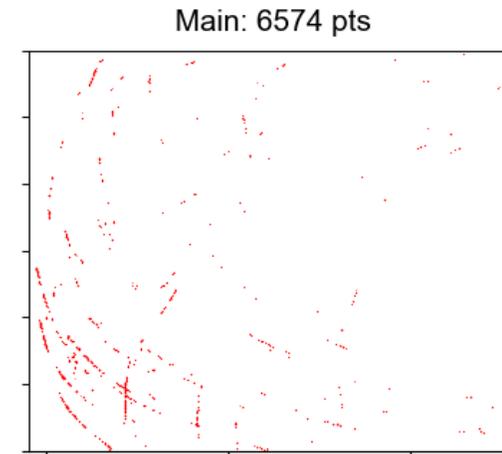
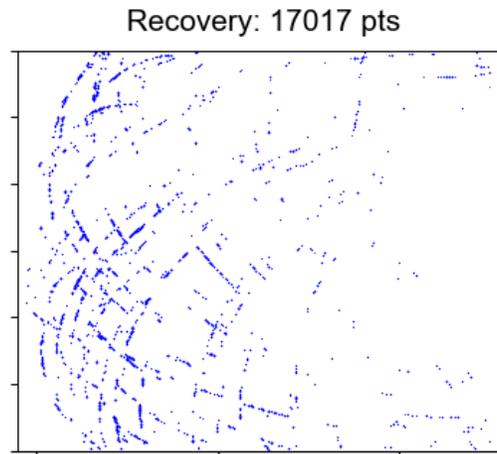
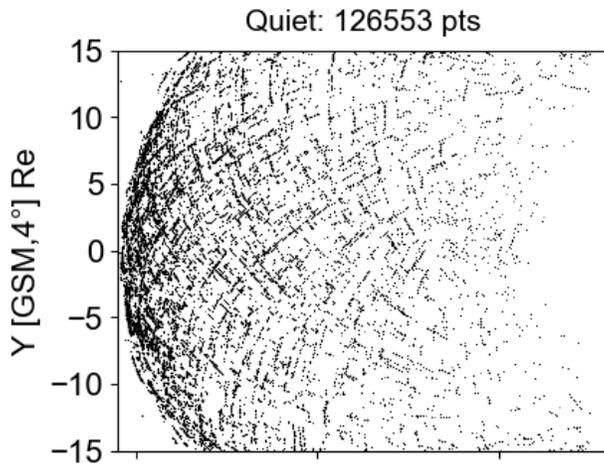
Plasma Sheet Convection – Geotail



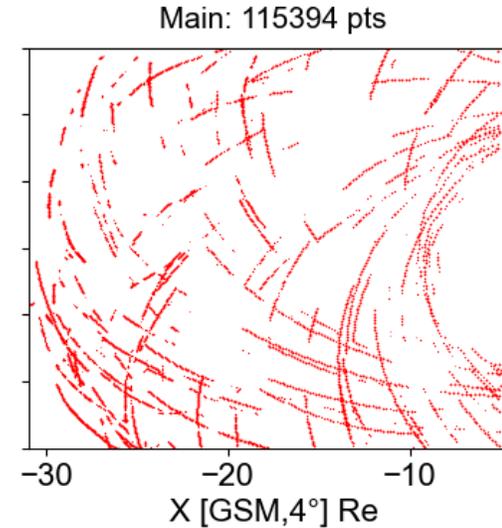
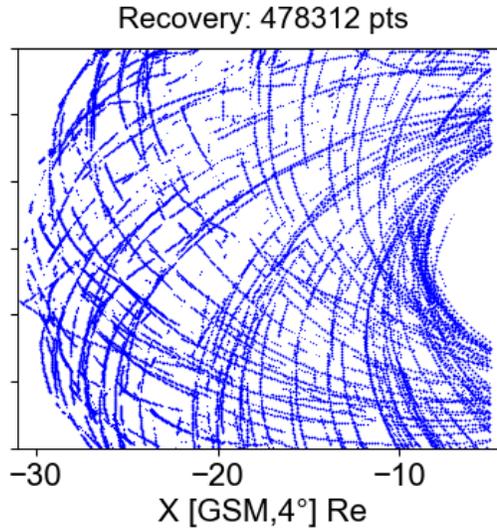
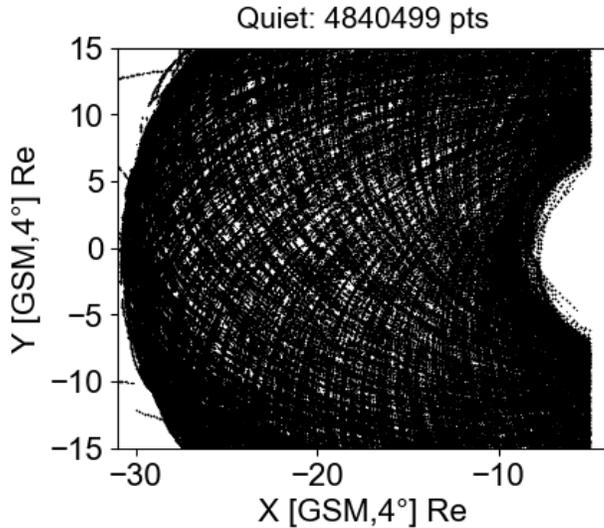
Similar for MMS, just more noisy...

Statistics Caveat for bursty flows – Geotail (1994 – 2022)

BBFs



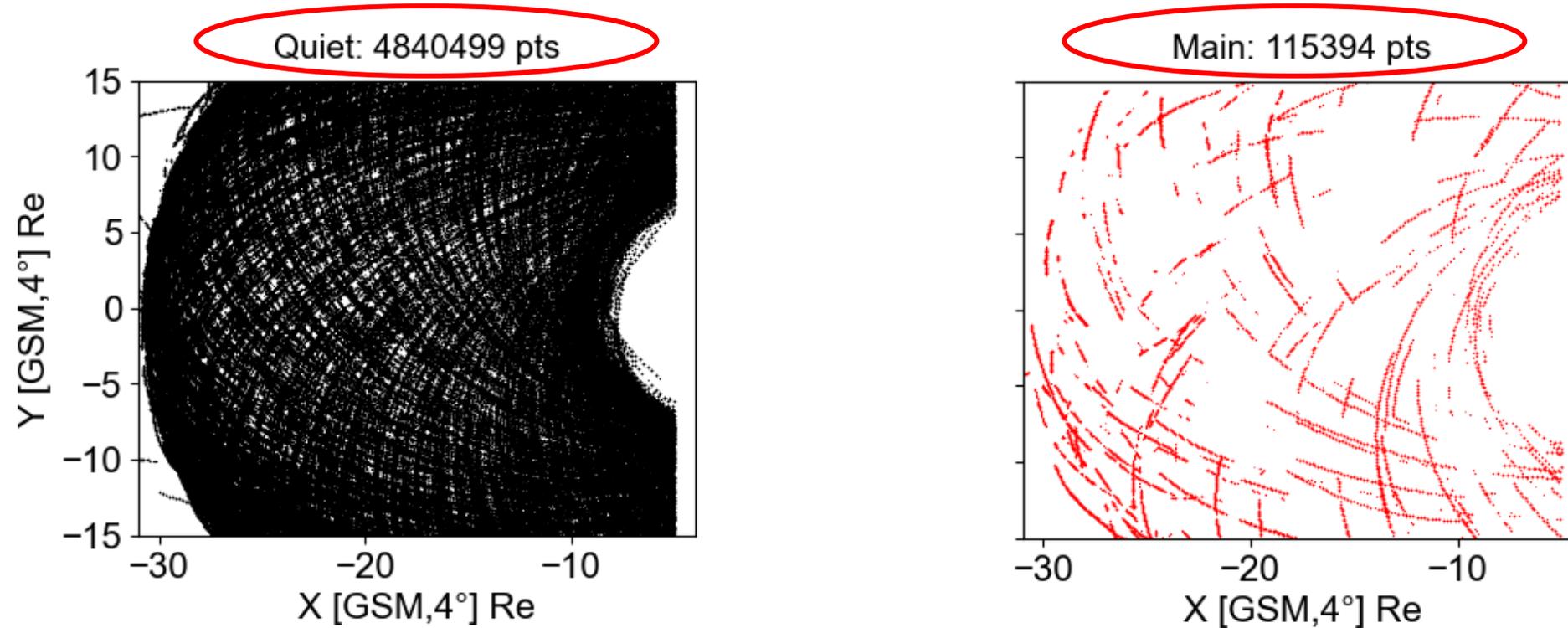
No BBFs



Keypoint: Number of data points can be misleading

Community Reminder: Data Sparsity During Extreme Events

Geotail data (1994 – 2022), time resolution: 12 seconds



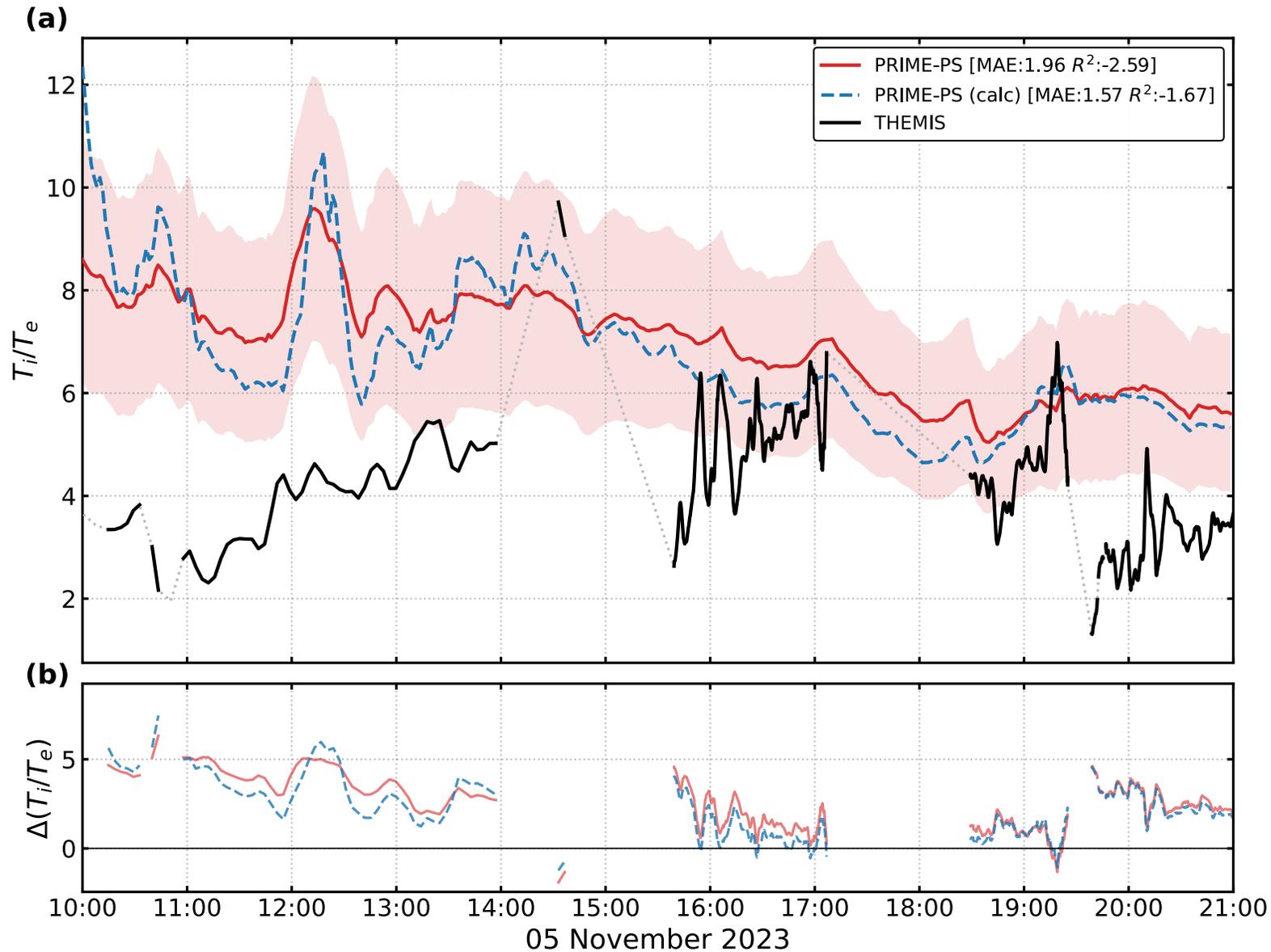
(#of unique days, # unique storms)

Geotail						
Main	7377 (36 32)	3006 (22 19)	2715 (17 17)	6240 (22 19)	6009 (21 19)	2450 (24 23)

Keypoint: Number of data points can be misleading

Note: This doesn't even include the incomplete driver (SW) distribution.

Test case of a storm (05 Nov 2023)



Process:

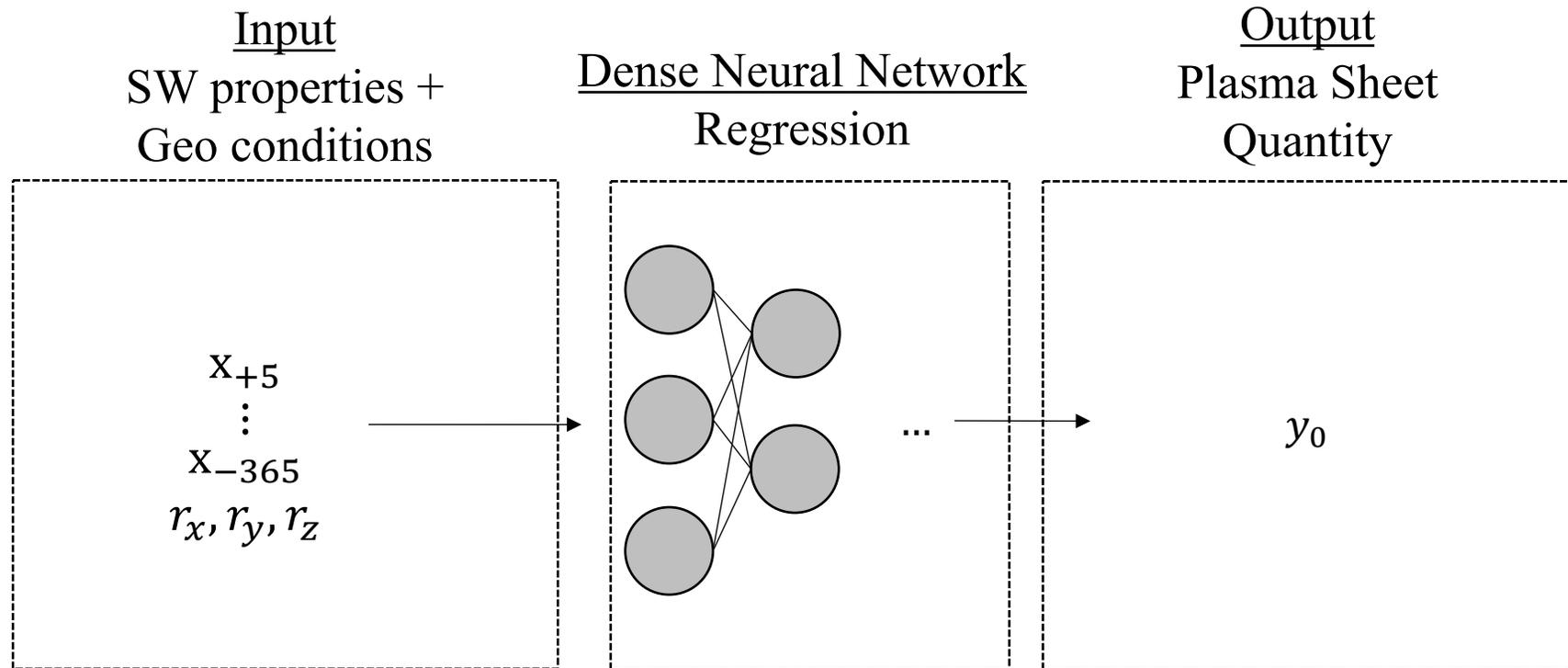
PRIME: Directly predicted

Calc: Calculated

Test:

THEMIS observations 2 years after the Geotail data stopped operating

Data Scientist POV (i.e., Input, output & regression)



Input:

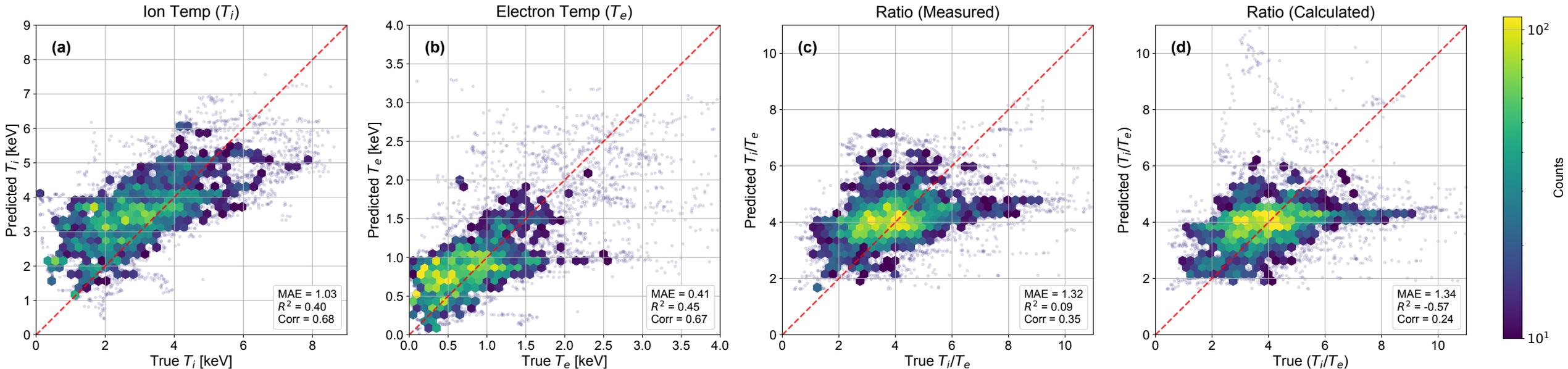
x: Different solar wind features (e.g., n, B, etc.) + geomagnetic indices including time history up to 6h
r: Location of SC measuring output

Output:

y: Different quantities at plasma sheet (e.g., n, B, T etc.)

Temperature and T_i/T_e MMS output

PRIME-PS | MMS Temperatures

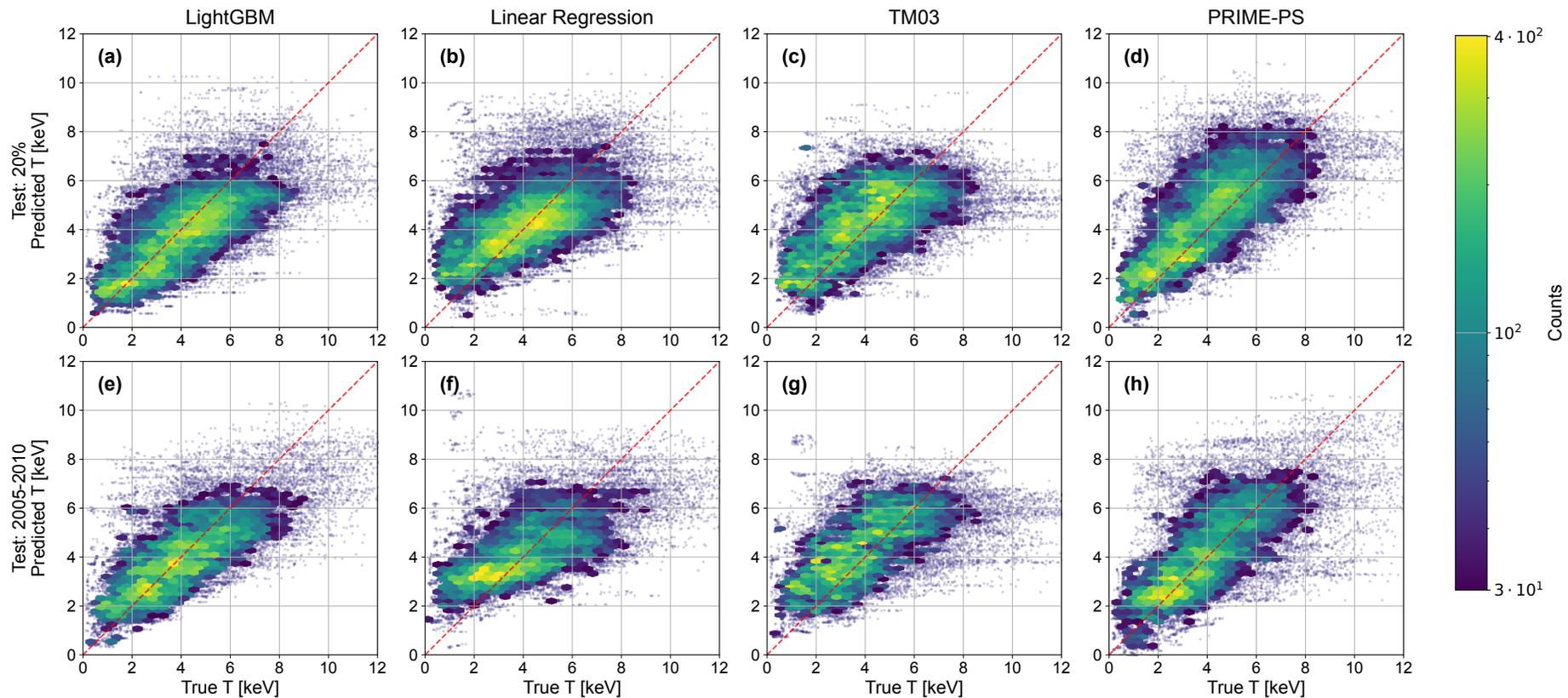


Note: Not showing analytical model, but essentially it does not agree with observations since DGSR16 and TM03 are trained on different domains, spatial distances, and from different missions

Modeling Temperature | Predictions vs Observations

Key Message: PRIME – PS/GB > Baseline \geq TM03

Model Performance | Temperature (T)



Is all this data needed? (Discussion point – Imbalanced learning)

1. Identify Outliers:

- **Find Outliers:** Detected using *Isolation Forest* (unusual feature patterns)

2. Build a Diverse Core (Farthest Point Sampling):

- Selects a *core subset* where points are maximally distant
- Ensures broad coverage and *high diversity* of the original data

3. Balance with Rare Samples (Kernel Density Estimation):

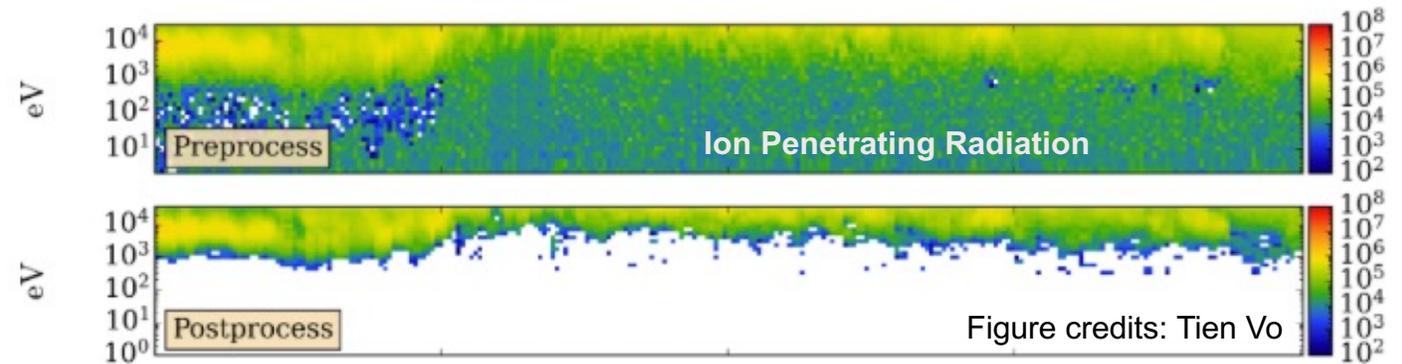
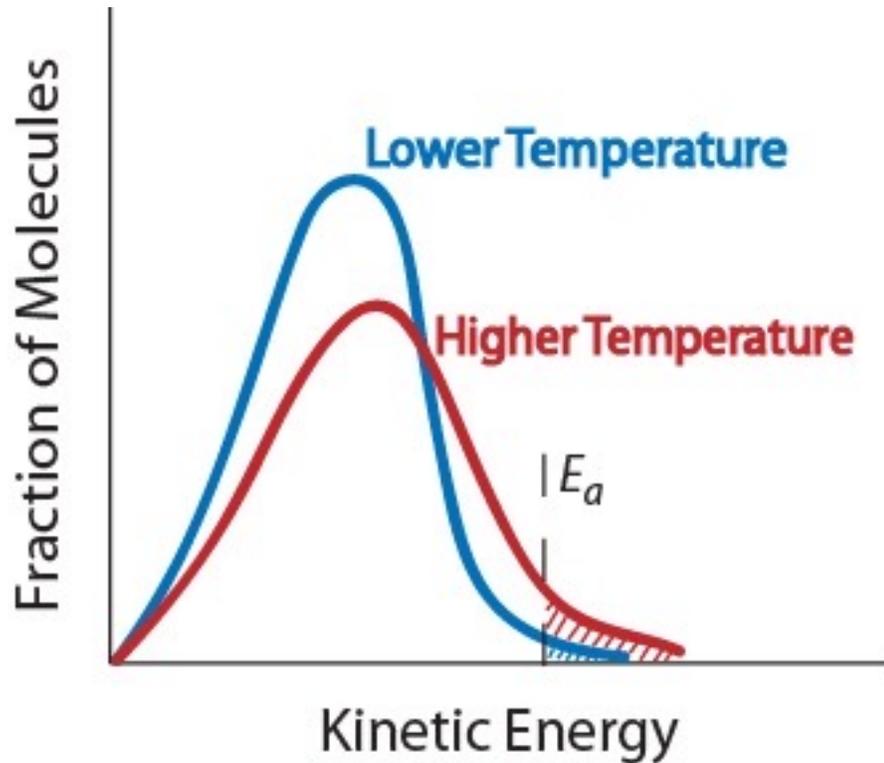
- Adds points from *under-represented regions*

Reduced by 1-2 order of magnitudes.
Almost Identical performance on test set.

```
... Original dataset size: 241311  
Reduced dataset size: 6978
```

Community Reminder on Temperature

- Temperature is the 2nd plasma moment
- **The higher the moment, the more uncertain** because you rely more on the poorly sampled tails of the distribution.
- So, 0 and 1st moment (**Density and Velocity**) are **usually ok**, but Temperature, we got to be careful



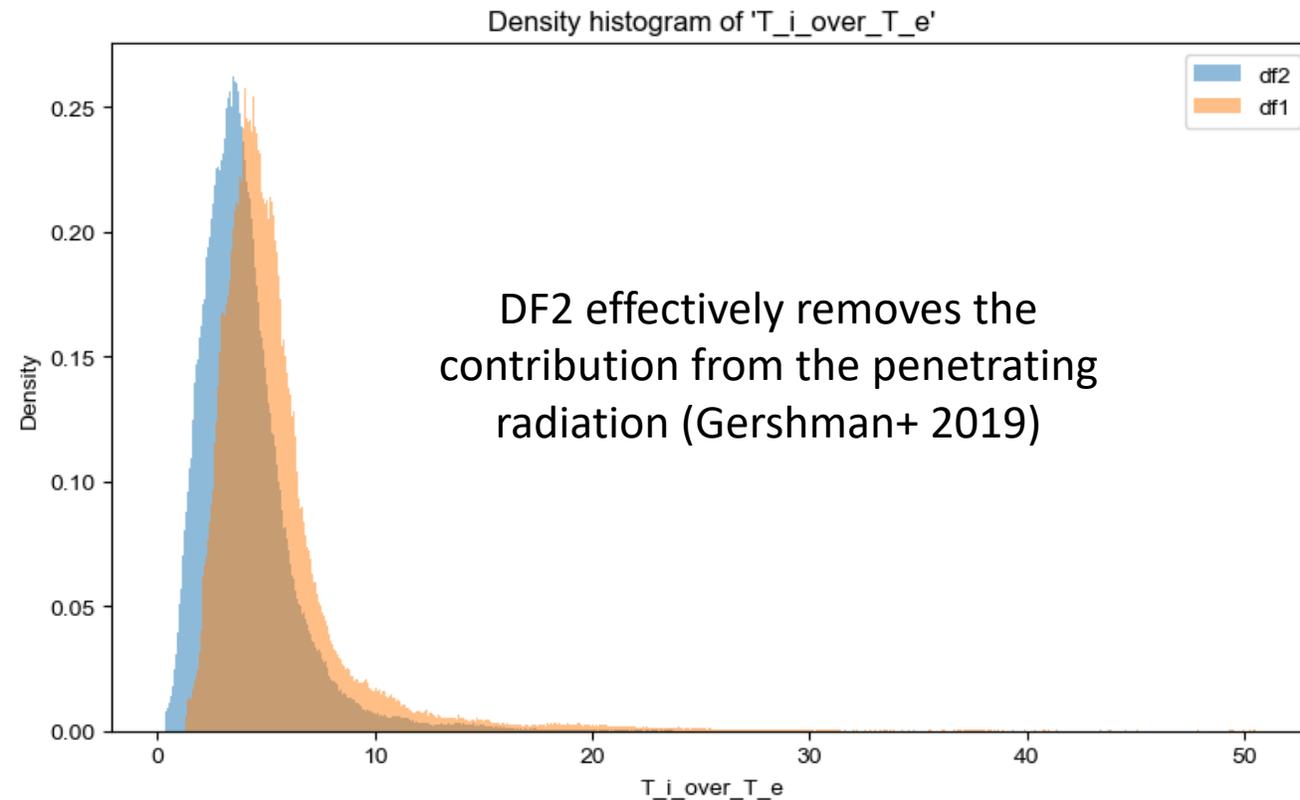
The temperature here was completely incorrect, and the velocity increased from about 200 km/s to over 1500 km/s.

$$T = \frac{m}{3k_B n} \int (\mathbf{v} - \mathbf{v}_b) \cdot (\mathbf{v} - \mathbf{v}_b) f(\mathbf{v}) d^3 v$$

MMS Ti/Te plasmashheet ratio example (Full vs Partial moments)

mean df1 (full distribution moments): 5.5513

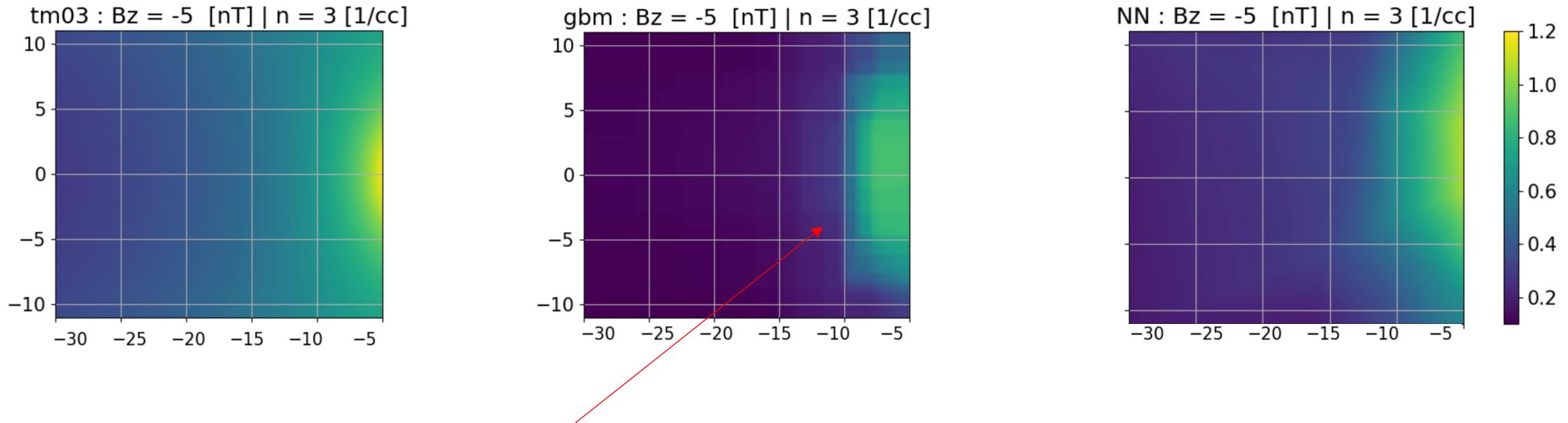
mean df2 (partial distribution moments): 4.0797



Key Message1: The mean differenced changed by 1.5 (>30%) simply by recalculating moments

Key Message2: A model with +30% is exciting, but we need to know if “ground truth” vary by the same magnitude

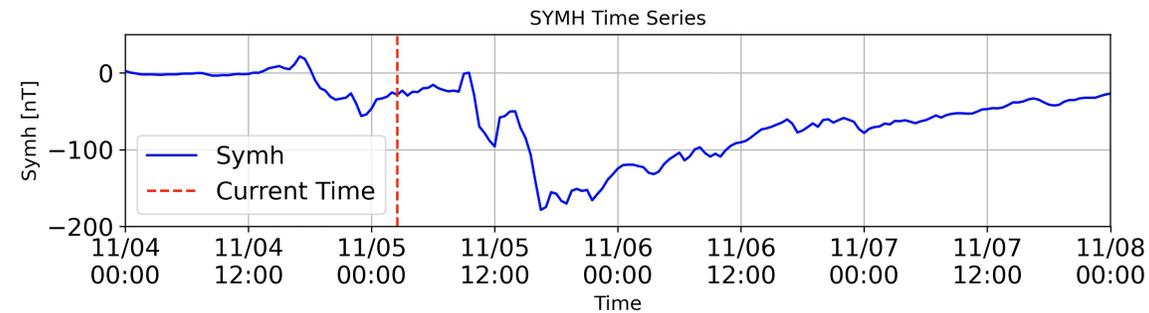
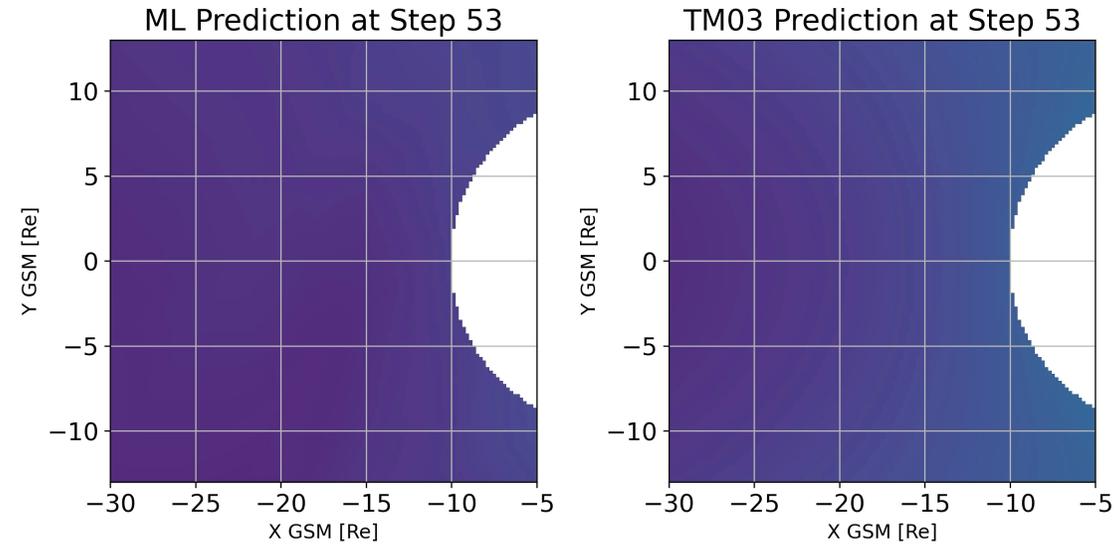
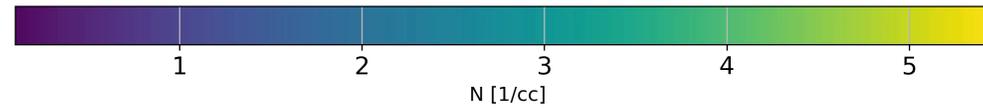
Community Reminder via Modeling Bz



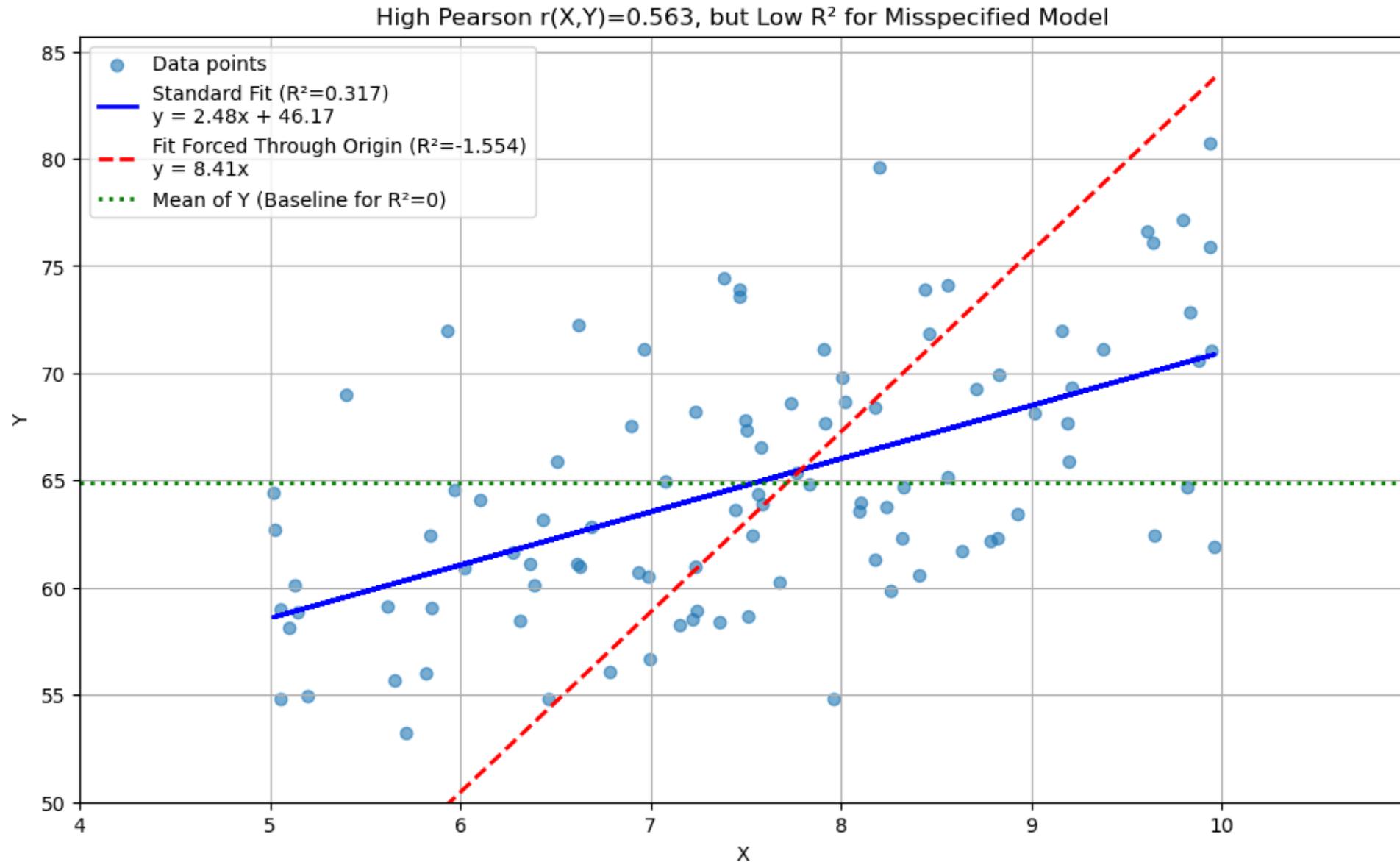
Gradient boosting is powerful, but it's a poor extrapolator.

Since it's built on decision trees, it predicts from nearby data rather than extending patterns like a neural network.

ML storm time density modeling

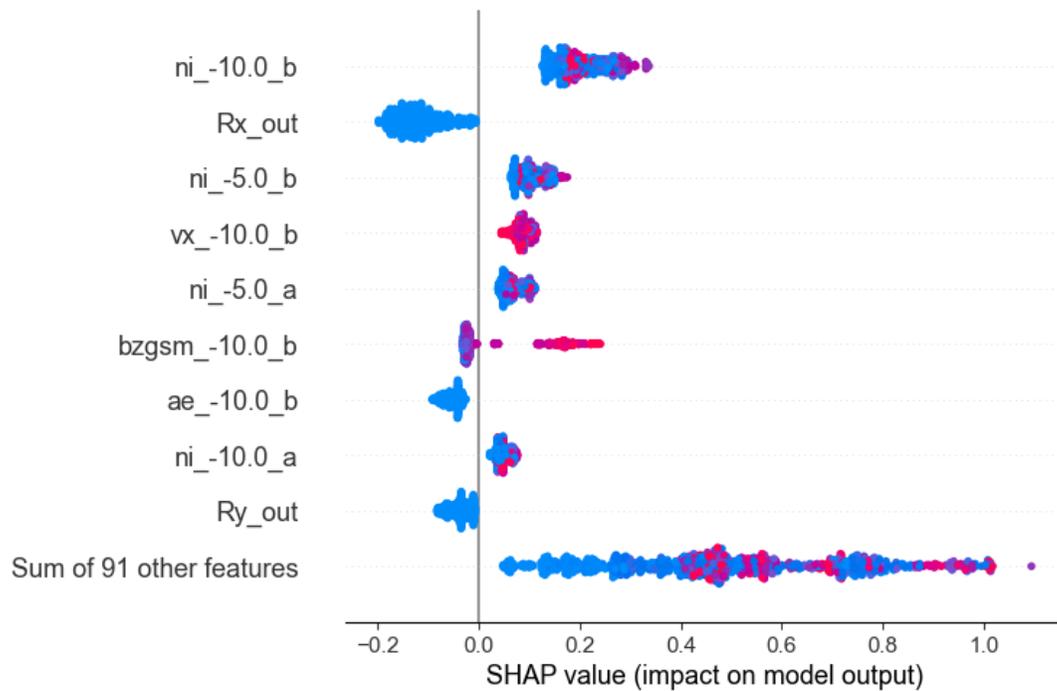


Correlation and Rsquared difference

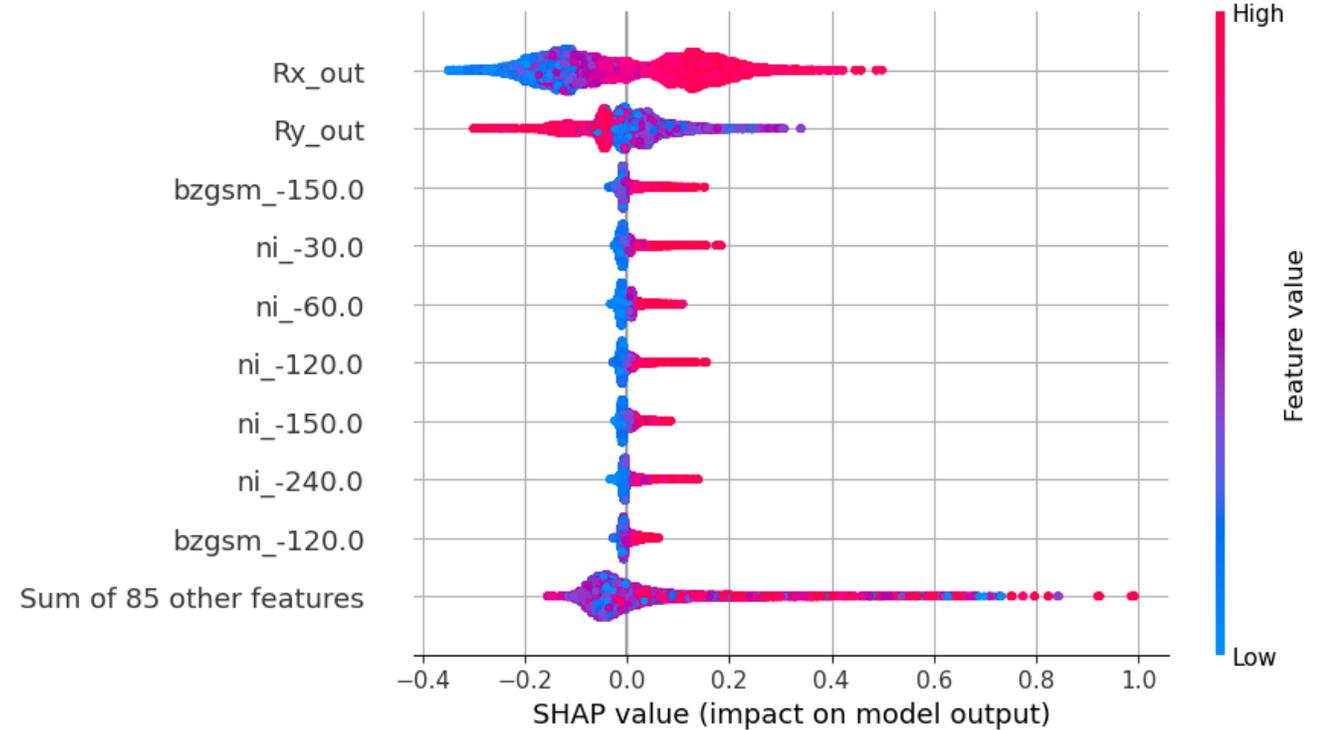


Model Feature importance storm vs quiet

In other words, the increased upstream density (-1h) had a greater impact during the storm than the SC location.

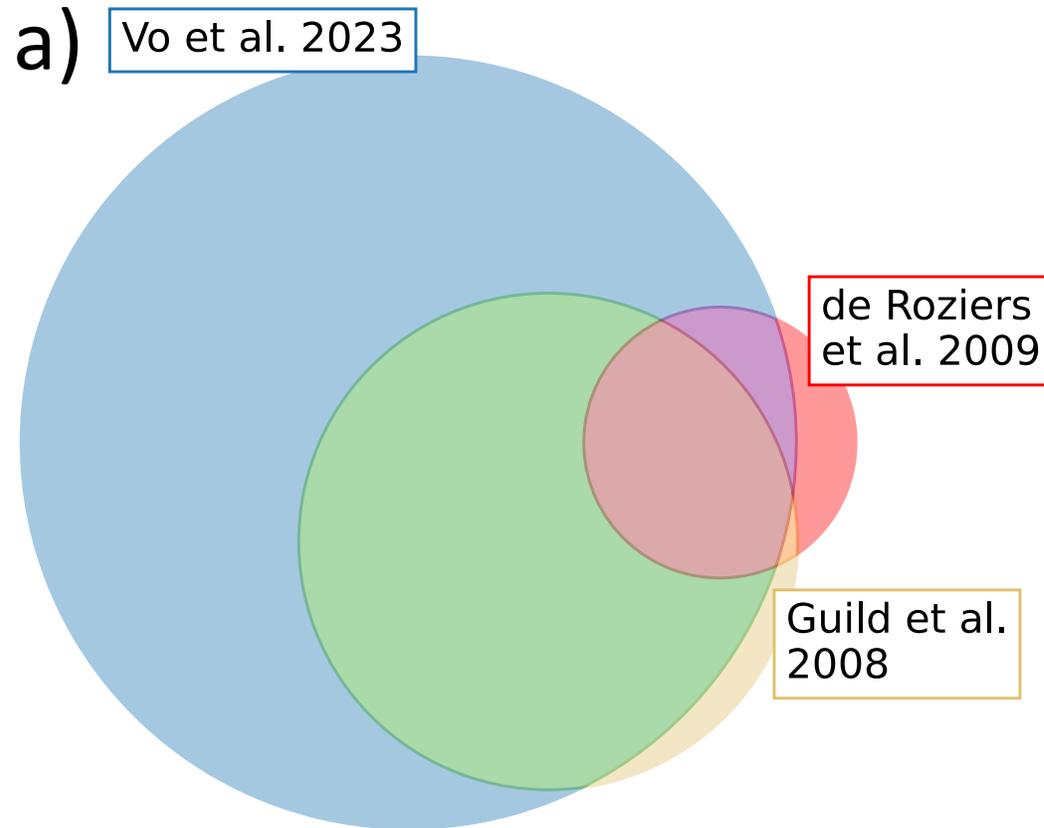


X_test_storm



X_test_total

Classifying plasma sheet is not trivial



Note Vo+2023, had a multi-step process based on interval, this is just using the point-by-point classification

Plasma Sheet Criteria			Number
Vo et al. 2023	de Roziers et al. 2009	Guild et al. 2008	
Yes	No	No	1,259,896
No	Yes	No	39,451
No	No	Yes	28,828
Yes	Yes	No	46,399
Yes	No	Yes	686,527
No	Yes	Yes	10,483
Yes	Yes	Yes	170,467

Criterion	Strict CPS	Flexible CPS	High density
$\beta > 1$	yes	—	—
$\beta > 0.5$	—	yes	—
$\sqrt{B_x^2 + B_y^2} < 2 B_z $	yes	—	—
$N < 6$	yes	—	—
$N > 6$	—	—	yes
$EA1SW0 = EA$	yes	yes	yes
$-31 < R_x < -5$	yes	yes	yes
$ R_y < 15$	yes	yes	yes
$ R_z < 10$	yes	yes	—
$V_x > -20$	—	—	yes

Table 1. Plasma sheet classification thresholds for the strict CPS, flexible CPS, and high-density subsets. *beta* is the ion plasma beta parameter, density (*N*) is in 1/cc units, *V_x* is in km/s, and all the locations (*R_{x,y,z}*) are in Earth radius. The coordinate system for all vectors is the aberrated Geocentric Solar Magnetospheric (GSM) coordinates

Storm Time Behavior and Importance of Outliers

The Problem: We use **static thresholds** for dynamic environments.

The Risk: Therefore we can **mistakenly remove the crucial "stormtime plasmashet."**

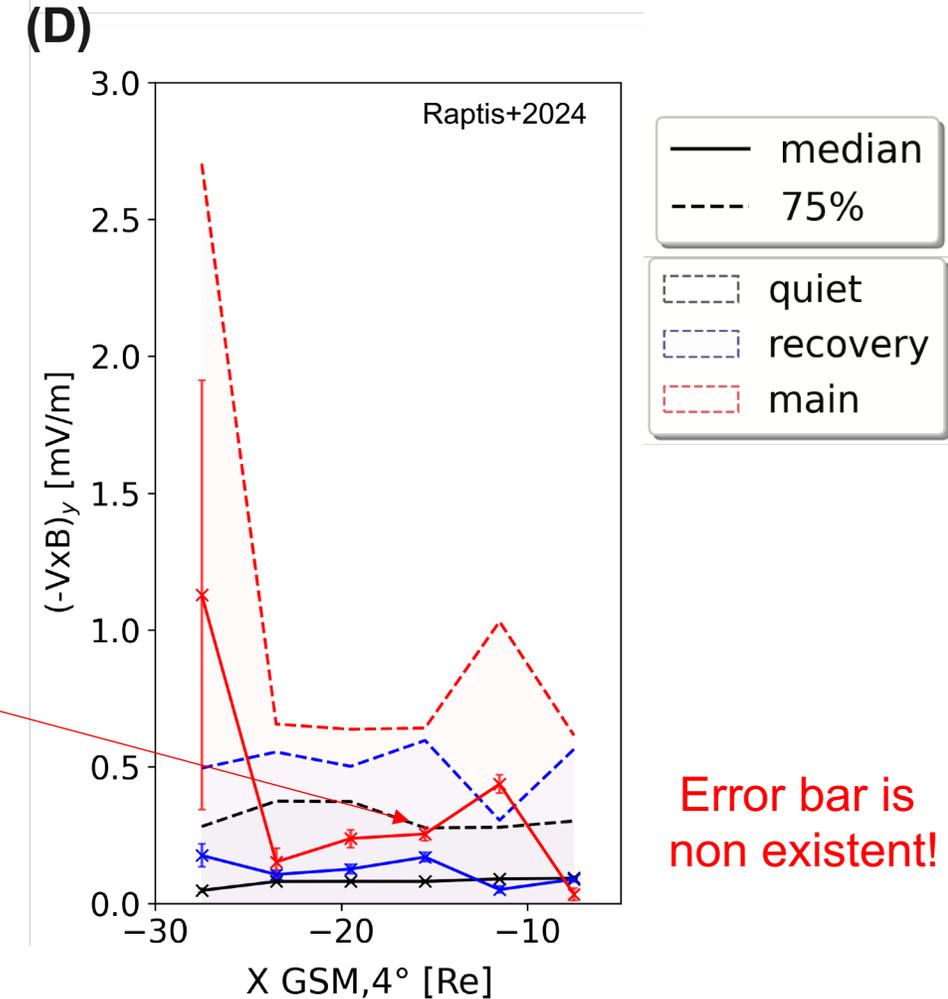
The "Solution": **Manually find the missing data and add it to the dataset.**

Strict CPS (e.g., Ohtani et al., 2008 Raptis et al., 2024) & Flexible CPS (e.g., Richard et al., 2022)

Data number & storms

Table S1. Distribution of points for Geotail and MMS, used to generate Figure 2 of the main text. Each cell indicates the number of points along with the number of unique days and unique storms (#days | #storms). The bins used for the X axis are shown in the last row.

Geotail						
Main	7377 (36 32)	3006 (22 19)	2715 (17 17)	6240 (22 19)	6009 (21 19)	2450 (24 23)
Recovery	27254 (77 66)	19855 (68 58)	17770 (60 53)	14892 (58 52)	26234 (79 68)	35488 (102 93)
Quiet	220027 (553)	122734 (425)	104663 (401)	128919 (438)	234532 (666)	382446 (833)
MMS						
Main	8 (1 1)	573 (3 3)	1476 (3 3)	1512 (3 3)	1299 (3 3)	1987 (7 5)
Recovery	781 (3 2)	4412 (8 6)	3889 (10 8)	4284 (12 10)	7907 (13 13)	7451 (18 16)
Quiet	21036 (52)	53739 (135)	40825 (126)	37538 (132)	46777 (166)	72190 (195)
Bins x	[-30, -25]	[-26, -21]	[-22, -17]	[-18, -13]	[-14, -9]	[-10, -5]



What does it mean to have 1500 data points if they originate from 3 storms/3days?

Key Message: #of unique days, # unique storms, and distribution of upstream conditions is more important than SE