JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

Storm-Time Plasma Sheet: Global Convection Patterns, Bursty Bulk Flows, and Data-Driven Modeling

Savvas Raptis

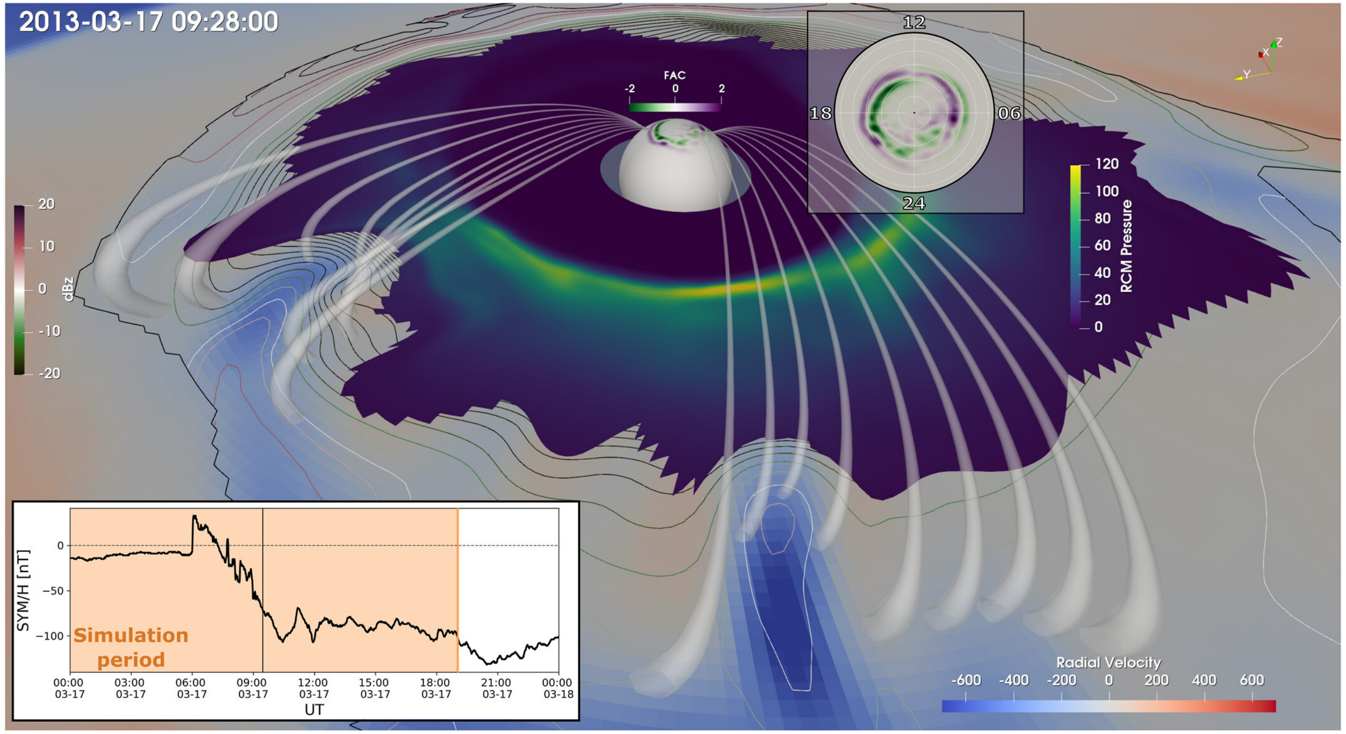
APL/JHU, Laurel, MD, USA

Acknowledgments:

- CGS NASA Drive Center

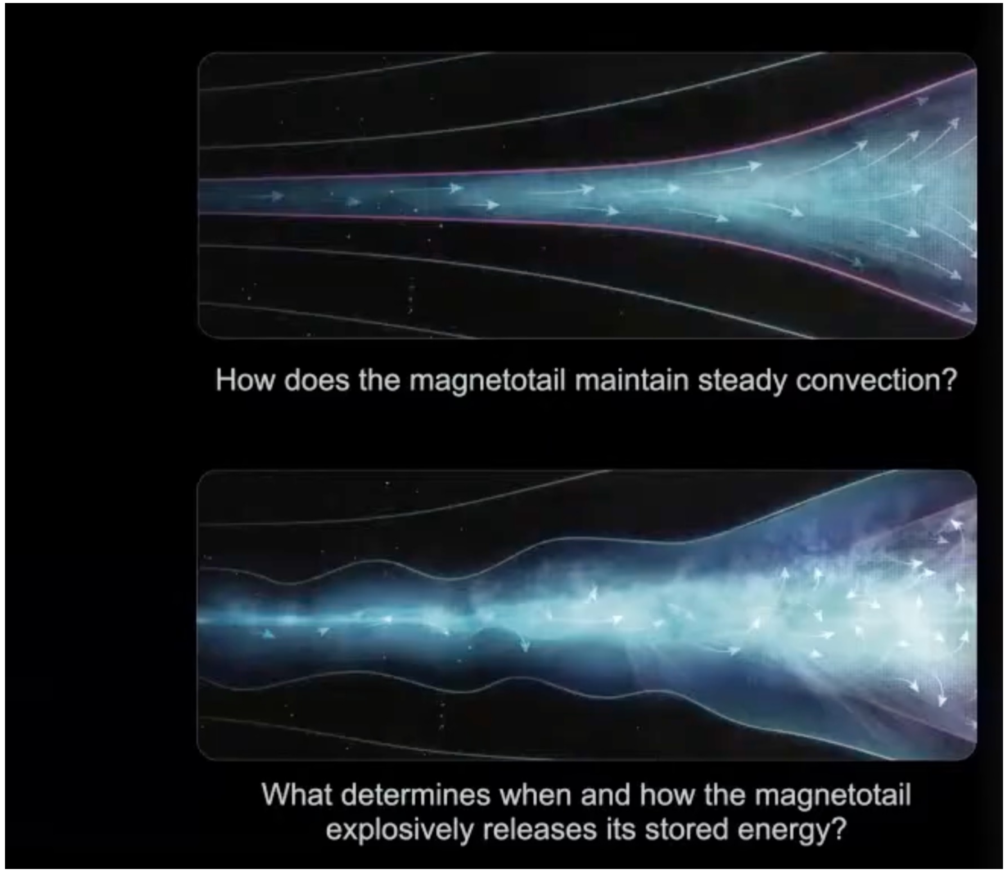
savvas.raptis@jhuapl.edu / https://savvasraptis.github.io

General Context & Synergy



Sciola+ 2023

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



Cropped from Robyn Millan Presentation 26 Feb 2026

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 **plasmasheet transport during quiet and storm times**.

Overaching Goal of CINEMA

Understand the role of plasma sheet structure and evolution in Earth's multiscale magnetospheric convection cycle

Outline

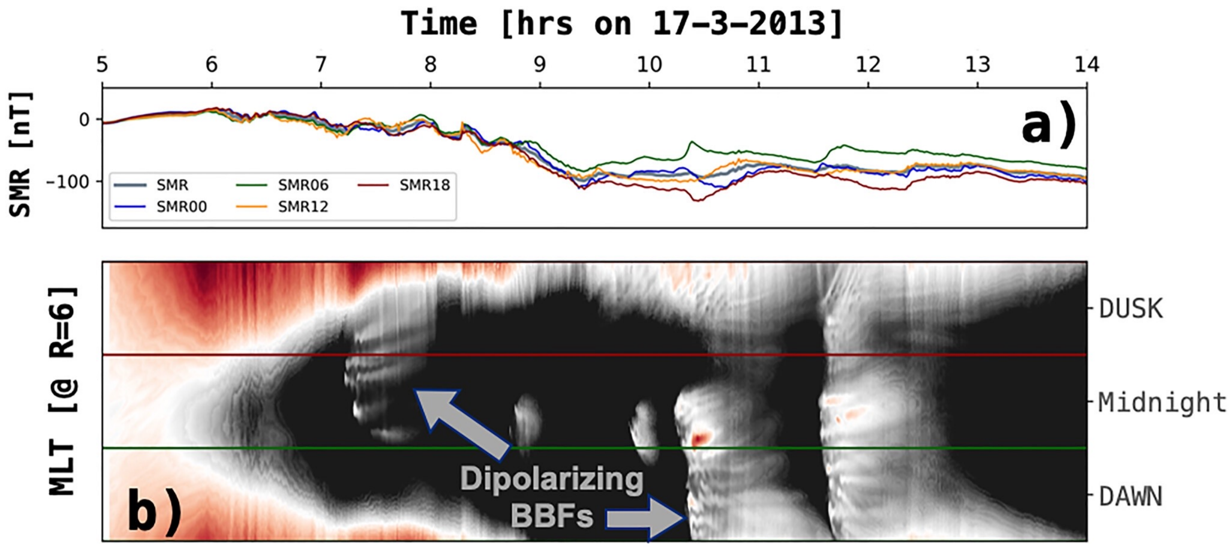
- Global Convection Patterns of Magnetic Flux
  - Question: How does convection in the magnetotail change during storms and how it affects magnetic flux transport
- Bursty Interval Contribution and Dawn/Dusk Asymmetries
  - Question: Is the occurrence distribution of BBFs different during storms? How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

Geotail

Basically data analysis of Geotail (1994 – 2022), and MMS (2015 – 2024) magnetotail plasmasheet observations

*Similar definition to the one used in fusion research

Dawnside Current Wedge & Asymmetries in Magnetotail Reconnection

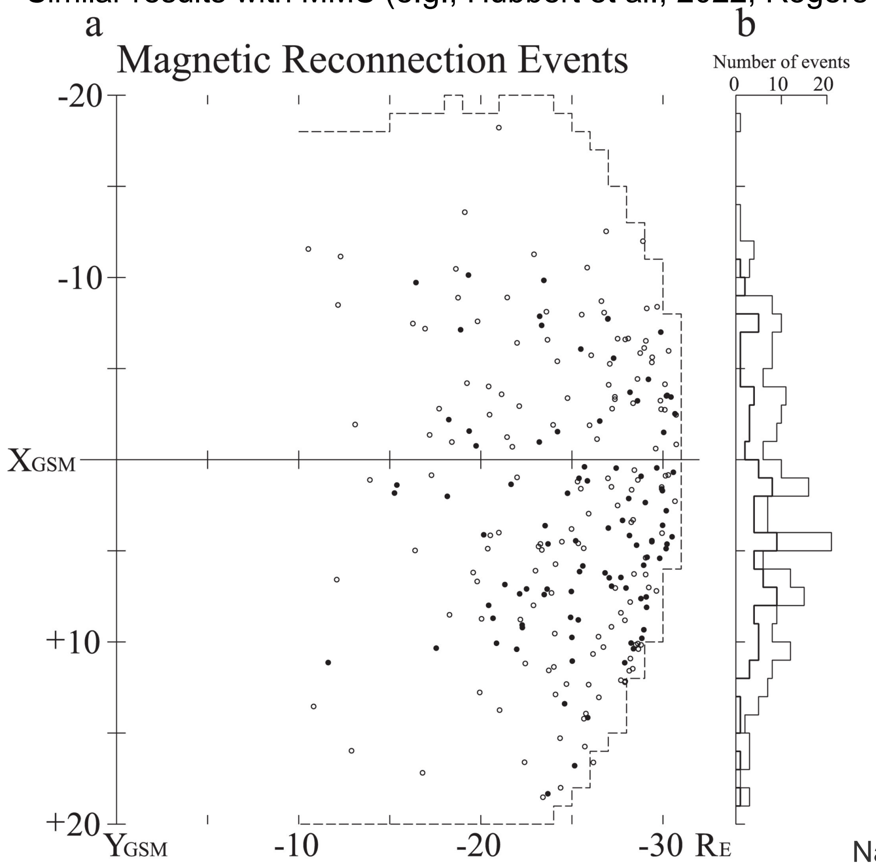


Ohtani+ 2021, 2024, Sorathia+ 2023

Dawnside Current Wedge

DCW 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).

How to do data analysis for storms

- STEP 1: Gather data
- STEP 2: Associate these data to storms
- STEP 3: ??
- STEP 4: Profit (i.e., publish papers?)

How to do data analysis for storms

- STEP 1: Gather data

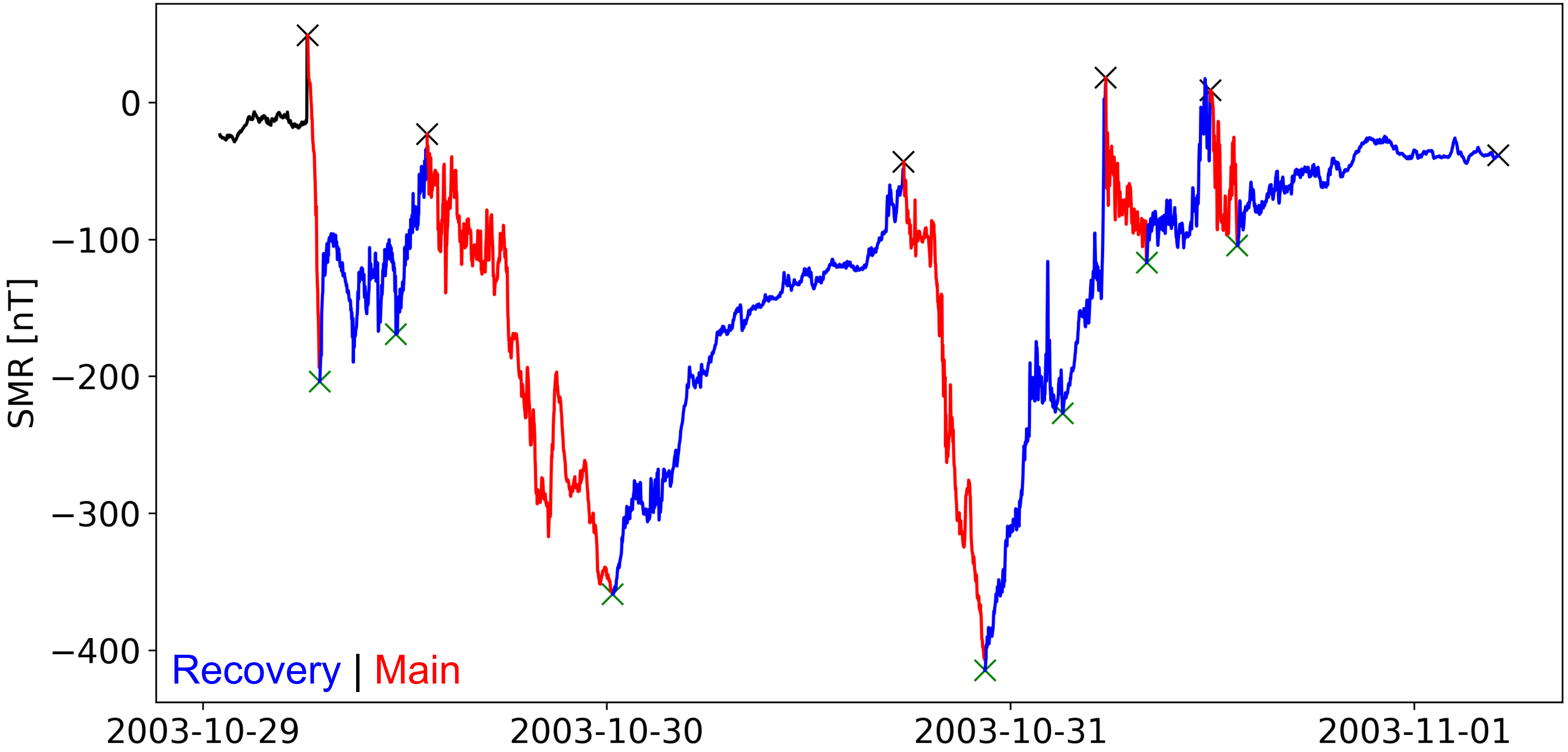
- STEP 2: Associate these data to storms

- STEP 3: ??

- STEP 4: Profit (i.e., publish papers?)

STEP 2: Storm phases classification

(semi)-automatic storm finder and classifier:



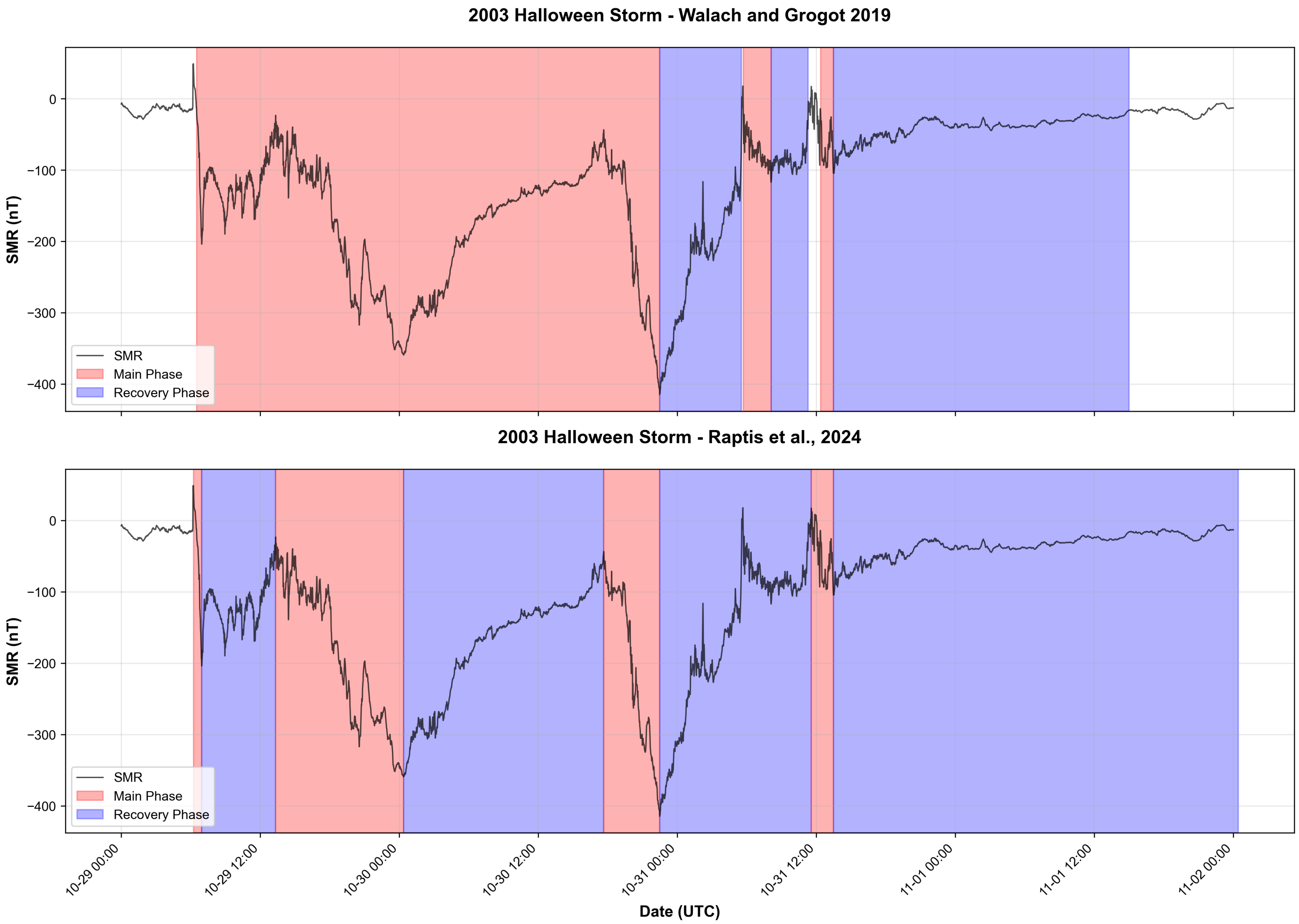
List is open, easy to use, easy to maintain

https://zenodo.org/records/15127938
(v3 updated to 2025, not yet manually verified)

Recent Update Regarding Storm Lists

<u>Our list is semi-automatic:</u>

1. Simple Peak finder using *scipy find_peaks function*.
2. Hyperparameters tuned based on Halloween storm & simple cases
3. ~2% manual changes visually



There are more lists out there. See Paper Published this Monday:

"How Does the Definition of a Geomagnetic Storm Affect the Contents of the Resulting Storm List? "

Patrick et al., 2026
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025SW004857

STEP3: Plasmasheet Coverage per mission

<u>Criteria to find CSP</u>

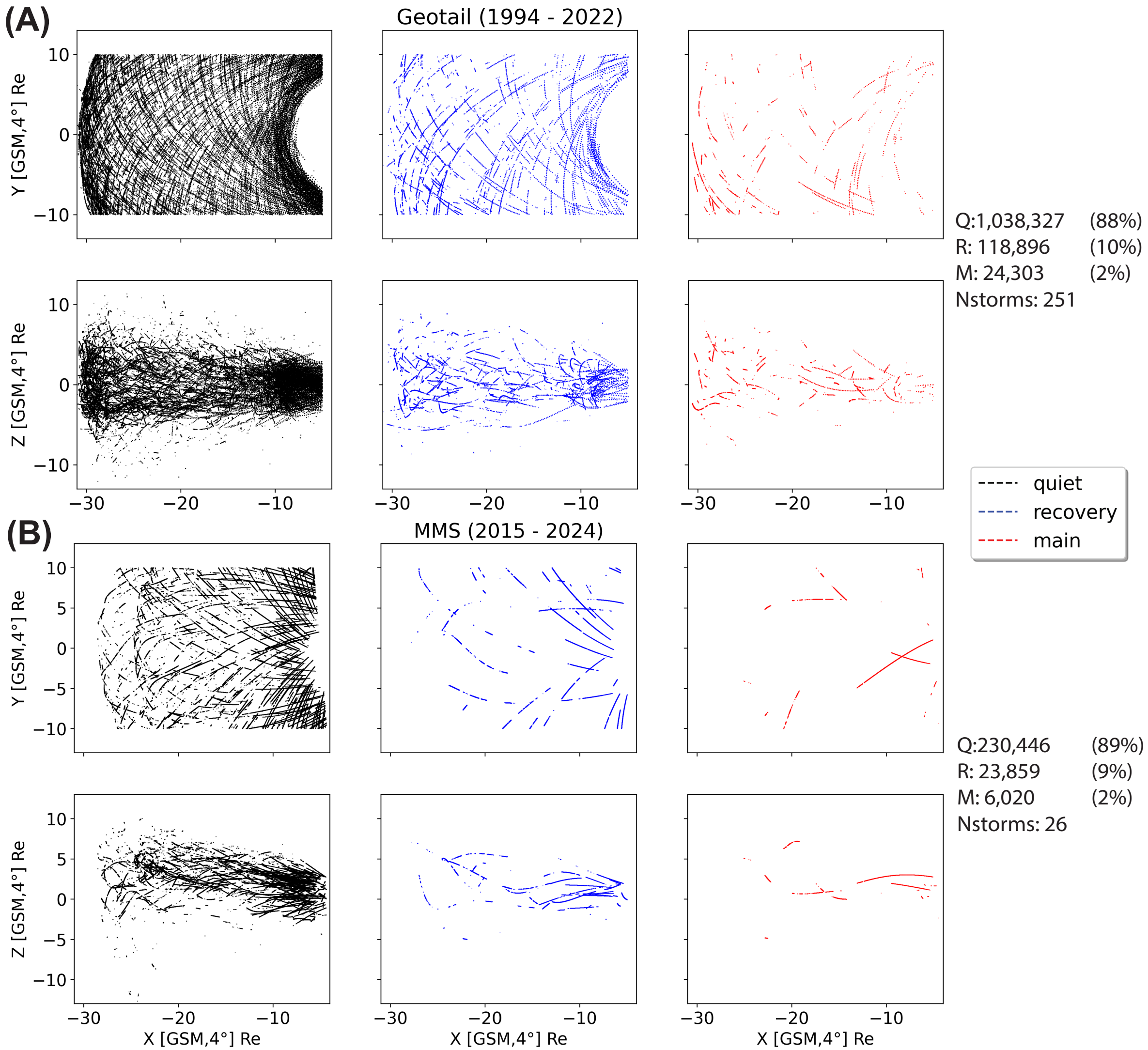
1. $|Y_{GSM,4°}| < 10$ Re
2. $-5 < X_{GSM,4°} < -31$
3. $\beta = \frac{P_{the}}{P_{mag}} > 1$
4. $|Bz| > 2\sqrt{Bx^2 + By^2}$

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

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

<u>Findings:</u>

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**



Raptis+ 2024 | GRL

Classifying plasma sheet is not trivial

a)

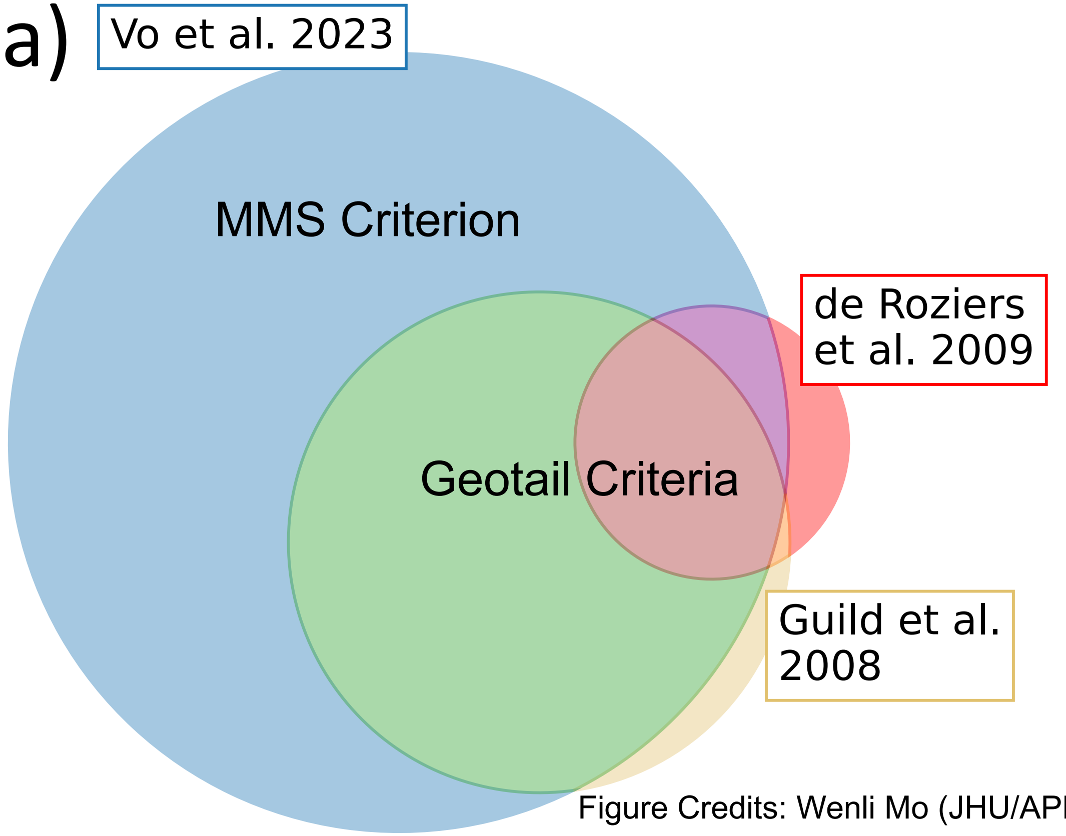


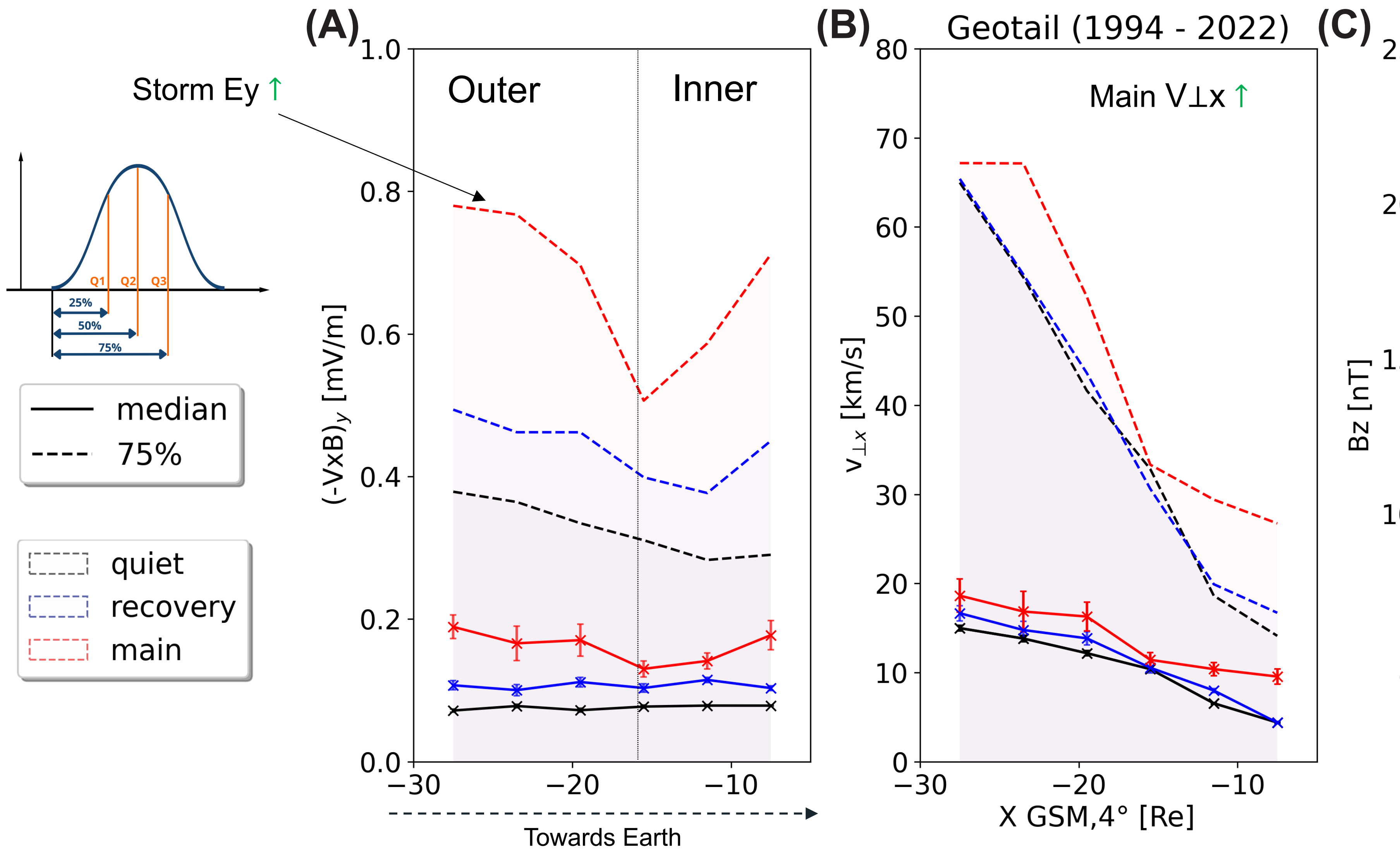
Figure Credits: Wenli Mo (JHU/APL)

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

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

Results

Plasma Sheet Convection – Geotail

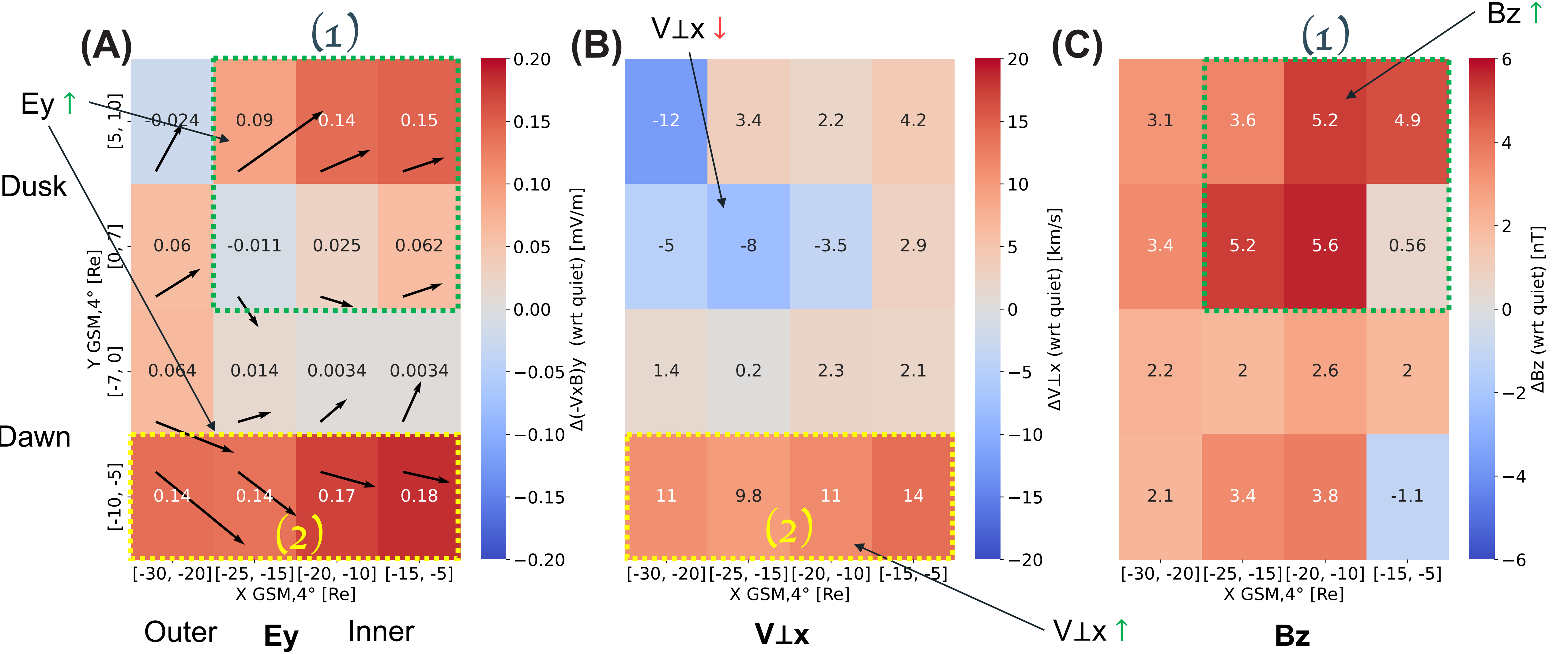


Similar for MMS, just noisier…

Raptis et al., 2023

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**



Plasmasheet changes in an asymmetric way during storms

Raptis et al., 2023

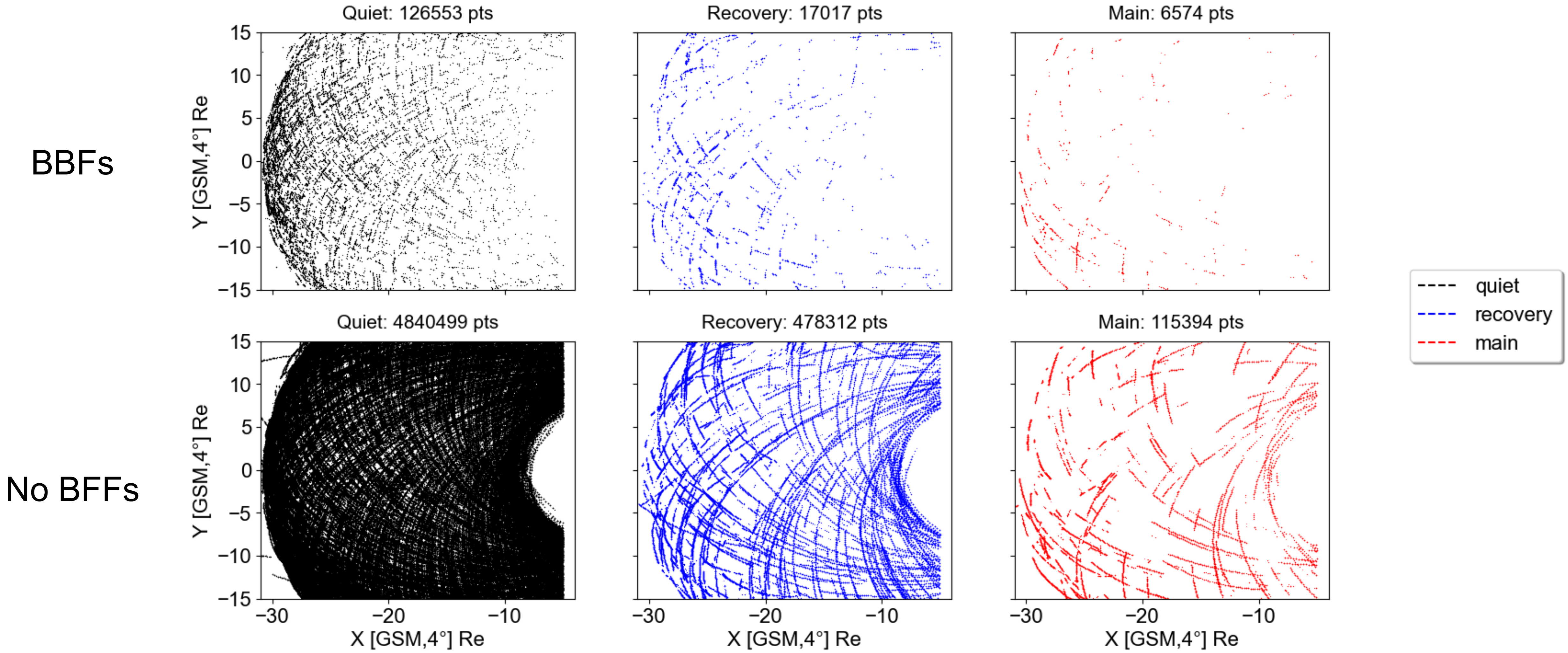
What do we know so far?

Plasma sheet storm time:

1. **Elevated Ey** associated with **increased Bz**, and **limited enhancement of** V⊥x
2. **Dusk observations showing more dipolar magnetic field (Bz** ↑)
3. **Dawn** are associated to **relatively faster flow (V⊥x** ↑)

Let’s move to Bursty Mesoscale Intervals!

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



Takeway: Number of data points can be misleading

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

<u>Analyzing BBFs statistically is non-trivial:</u>

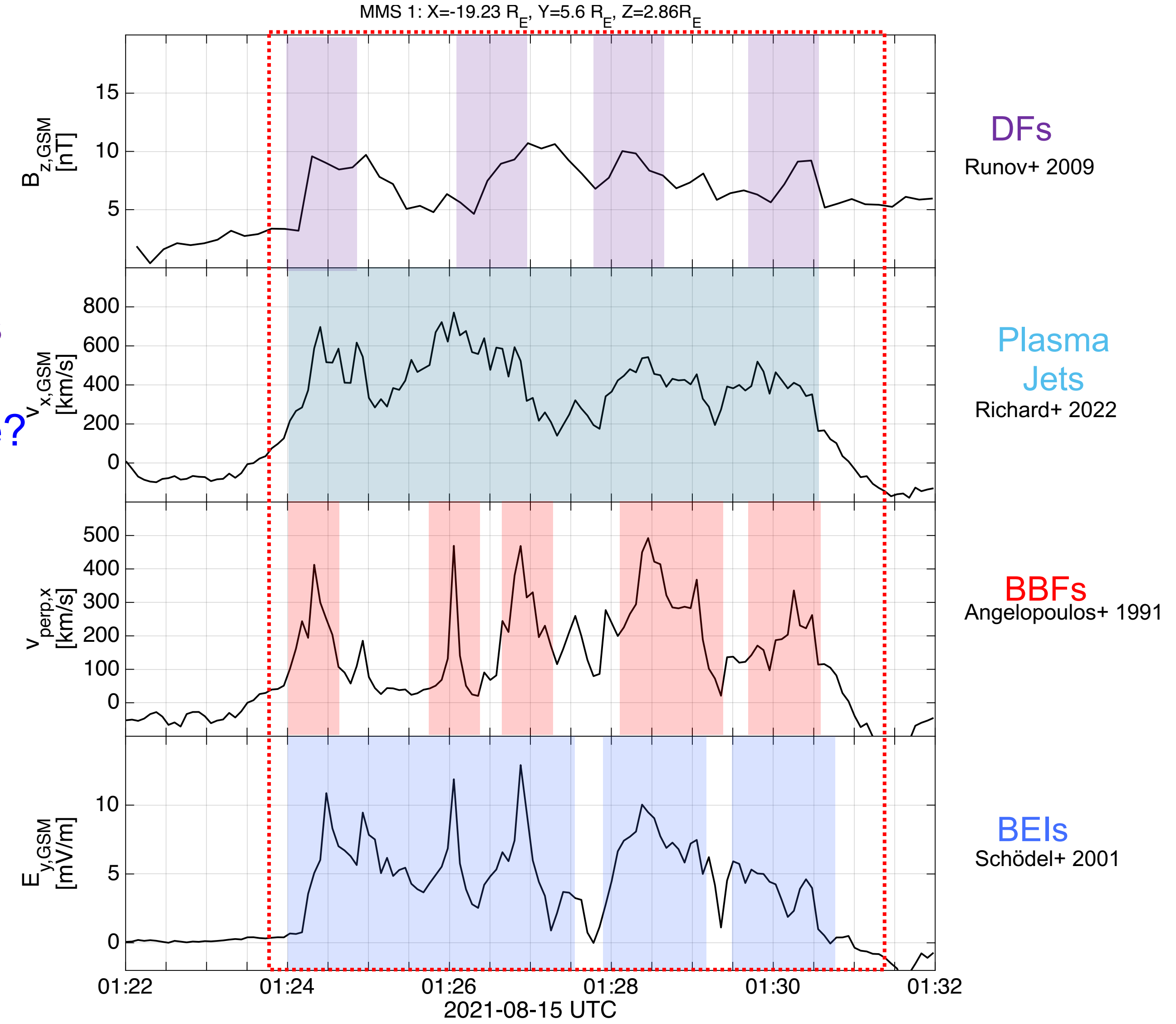
1. Arbitrary criteria and thresholds
2. BBFs are bursty, they merge, and de-accelerate.
3. Orbital biases of each mission
4. Are we even in the plasmasheet? (see previous slide)
5. Moment calculation, composition, energy and time res
6. No knowledge of spatial scales and relative location
7. Is it a direct hit or glazing? How probable is each case?
8. 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 in principle approachable by data/model comparisons, however number (1, 8) are very complicated…

<u>BBFs:</u>

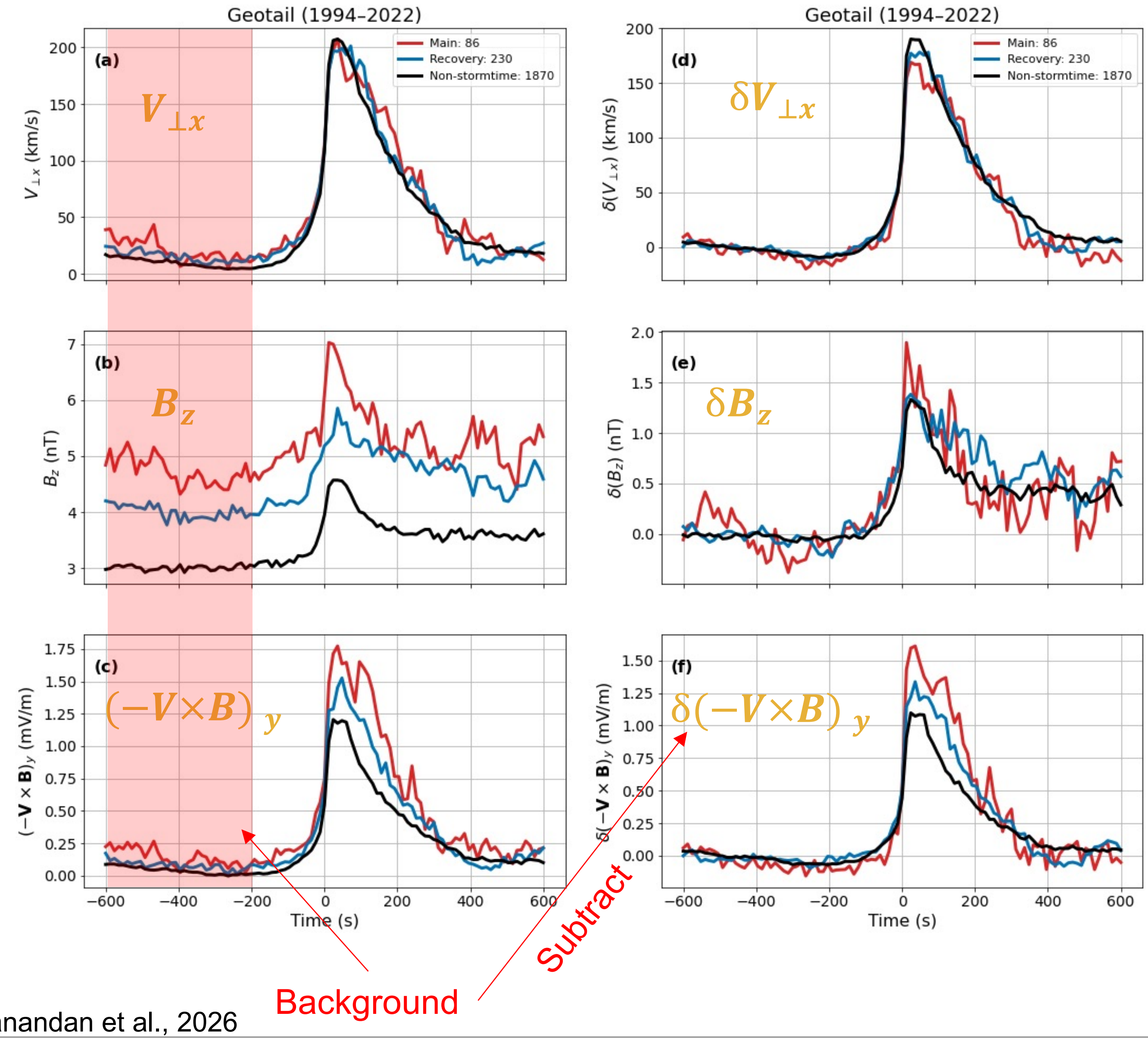
- Fast ion flows (v >400 km/s),
- 10-100s in duration
- ~ 4 Re size



But we can still do things!

We just have to be careful

Superposed Epoch Analysis (SEA) of BBFs during storms



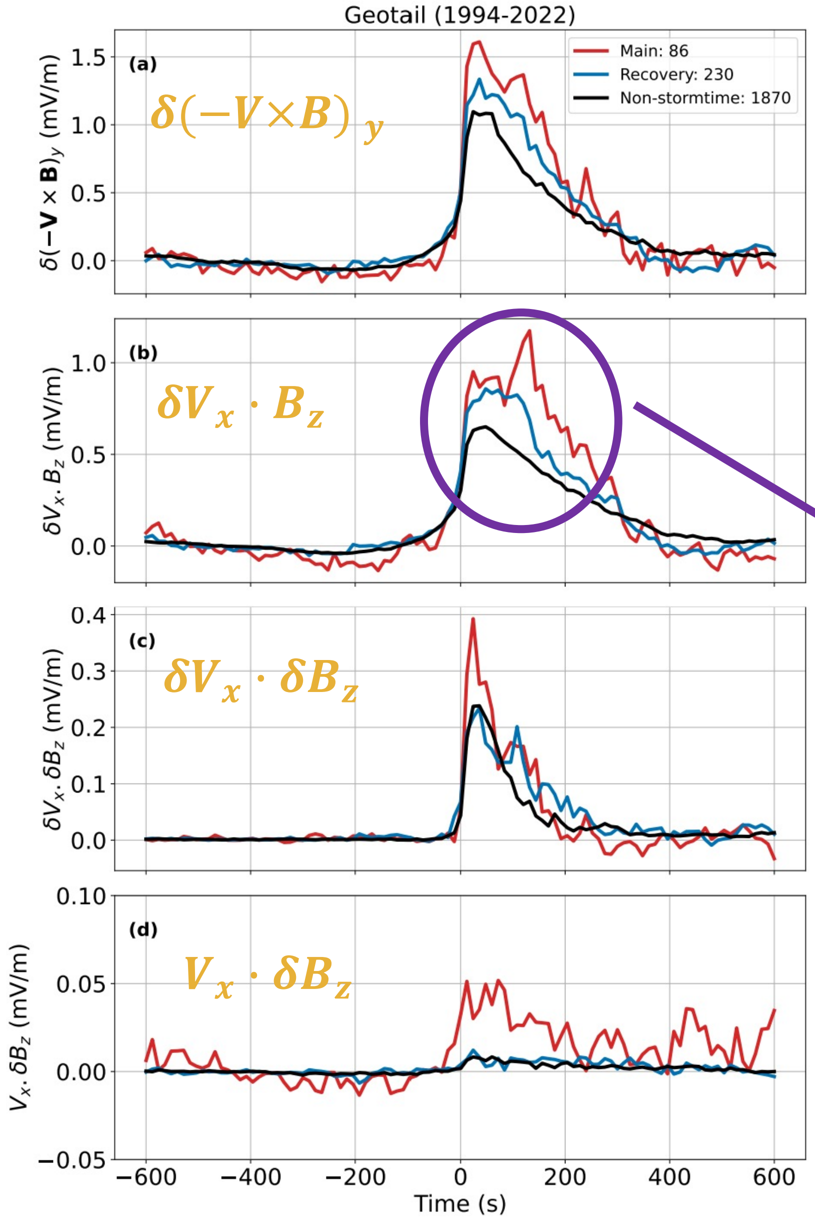
Devanandan et al., 2026

<u>Key Results:</u>

1. Storm-time BBFs transport **more magnetic flux** than non-storm BBFs.

2. This is linked to a **stronger background Bz**, while BBF velocity stays about the same.

Decomposition of Magnetic Flux Enhancement



Devanandan et al., 2026

$$\delta(-V\times B)_y = \delta V_x \cdot B_z + V_x \cdot \delta B_z + \delta V_x \cdot \delta B_z - (\delta V_z \cdot B_x + V_z \cdot \delta B_x + \delta V_z \cdot \delta B_x)$$

Negligible

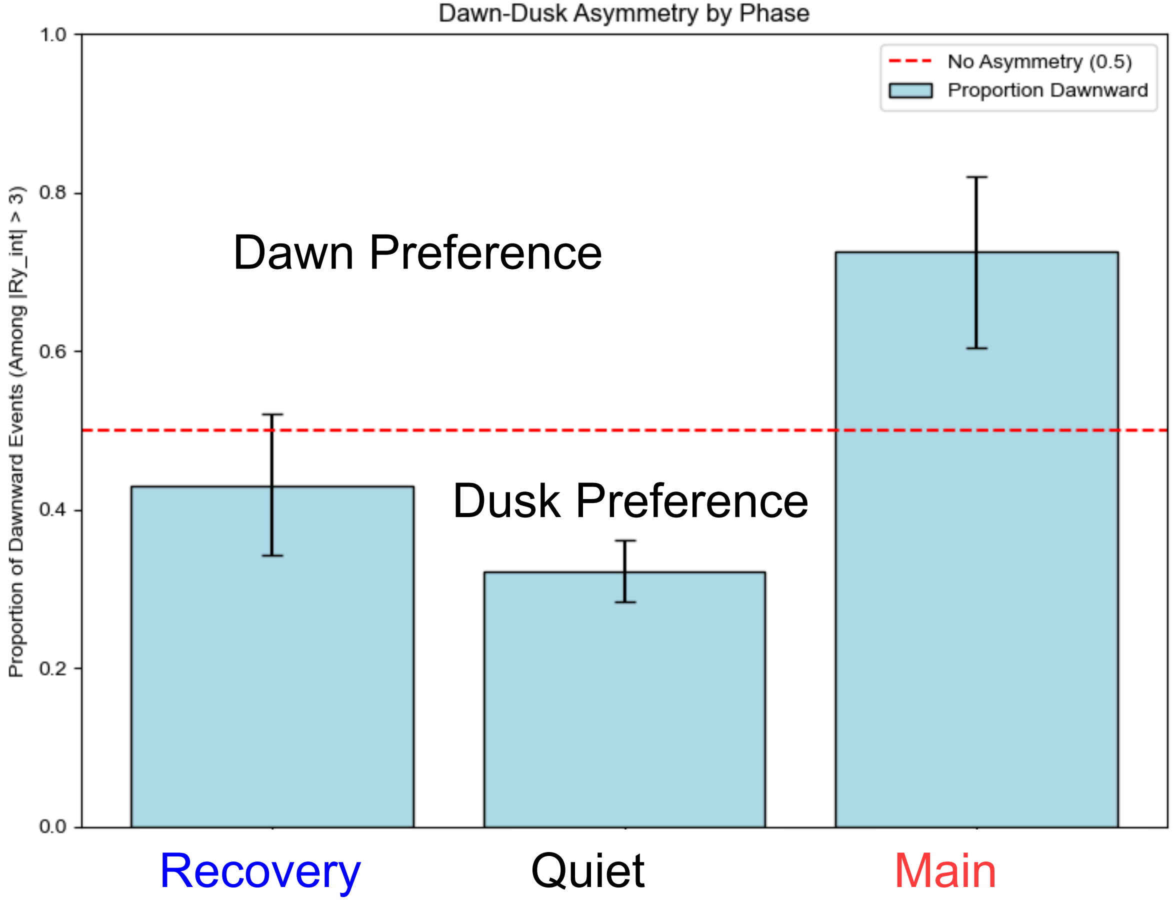
$$\delta(-V\times B)_y = \delta V_x \cdot B_z + V_x \cdot \delta B_z + \delta V_x \cdot \delta B_z$$

The contribution from $\delta V_x \cdot B_z$ is significantly larger compared to other terms.

In both storm and non-storm cases, the **flux enhancement comes mainly from the elevated BBF velocity.**

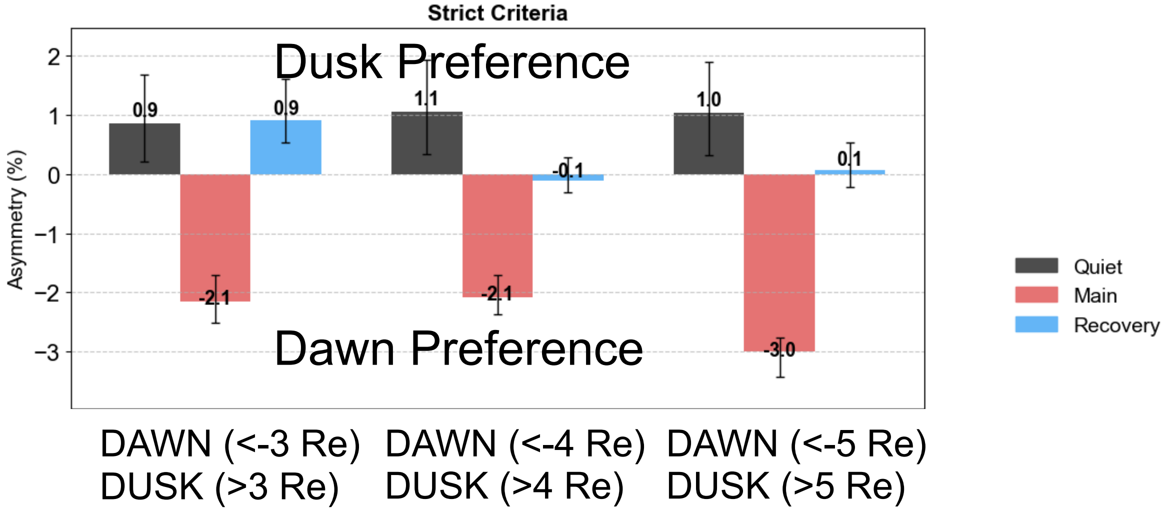
Dawn/Dusk Asymmetries - Reconnection and BBFs

Geotail dataset from Nagai+2023, Raptis+2026



Geotail dataset BBF list (Devandan+2026)

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



<u>Key Point:</u>

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

<u>TODO:</u>

- **Combine MMS and THEMIS** statistics to get a general picture

Ongoing effort

Part #1 Summary

1. Stormtime Global Convection:
    1. **Plasma sheet Ey** is elevated due to **increased Bz**, with **limited enhancement of** $\mathbf{V}_{\perp x}$
    2. **Bz enhancement** is more **prominent** at **Dusk**
    3. $\mathbf{V}_{\perp x}$ is more **elevated at Dawn**
2. Plasma sheet bursty Intervals:
    1. **Storm-time BBFs transport more magnetic flux than non-storm BBFs.**
    2. This enhancement linked to a **stronger background Bz**, while BBF velocity stays relatively constant (Devanandan+2026, GRL)
    3. **There seems to be a Dawn/Dusk Asymmetry on BBF occurrence and stormtime.** Statistics are still low, **Ongoing effort**: Combining THEMIS, Geotail, and MMS dataset to understand this better (Ongoing**)**

Raptis+2024

Devanandan+2026

What else can we do with all this data?

Machine Learning* of course!

Statistical Modeling ~ Supervised (labelled) Machine Learning

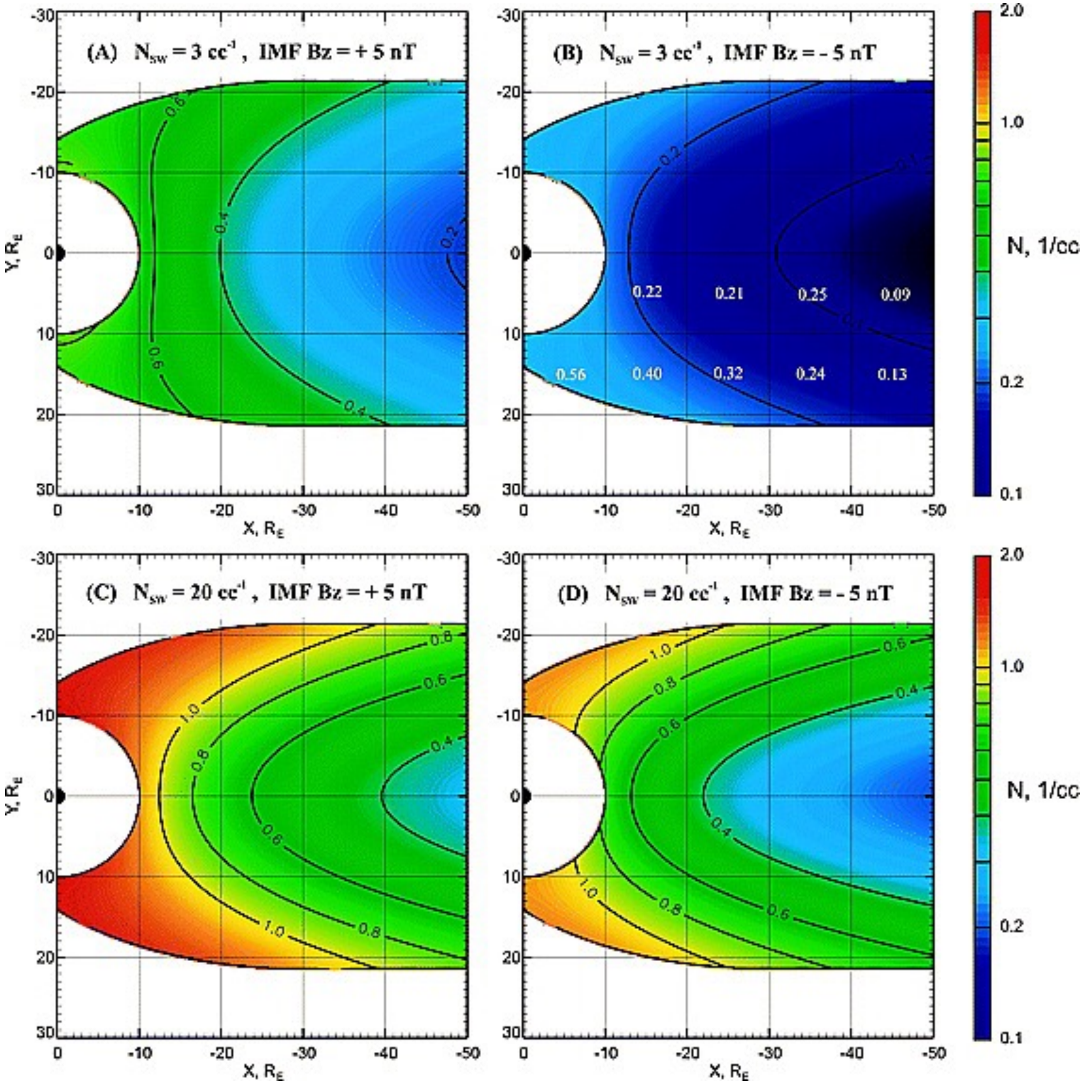
Outline

- Global Convection Patterns of Magnetic Flux
  - Question: How does convection in the magnetotail change during storms and how it affects magnetic flux transport
- Bursty Interval Contribution and Dawn/Dusk Asymmetries
  - 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 (“almost”* submitted)
  - Question: **can we use SW information and ground magnetometers to model plasmasheet properties**? How this changes during storms

*Similar definition to the one used in fusion research

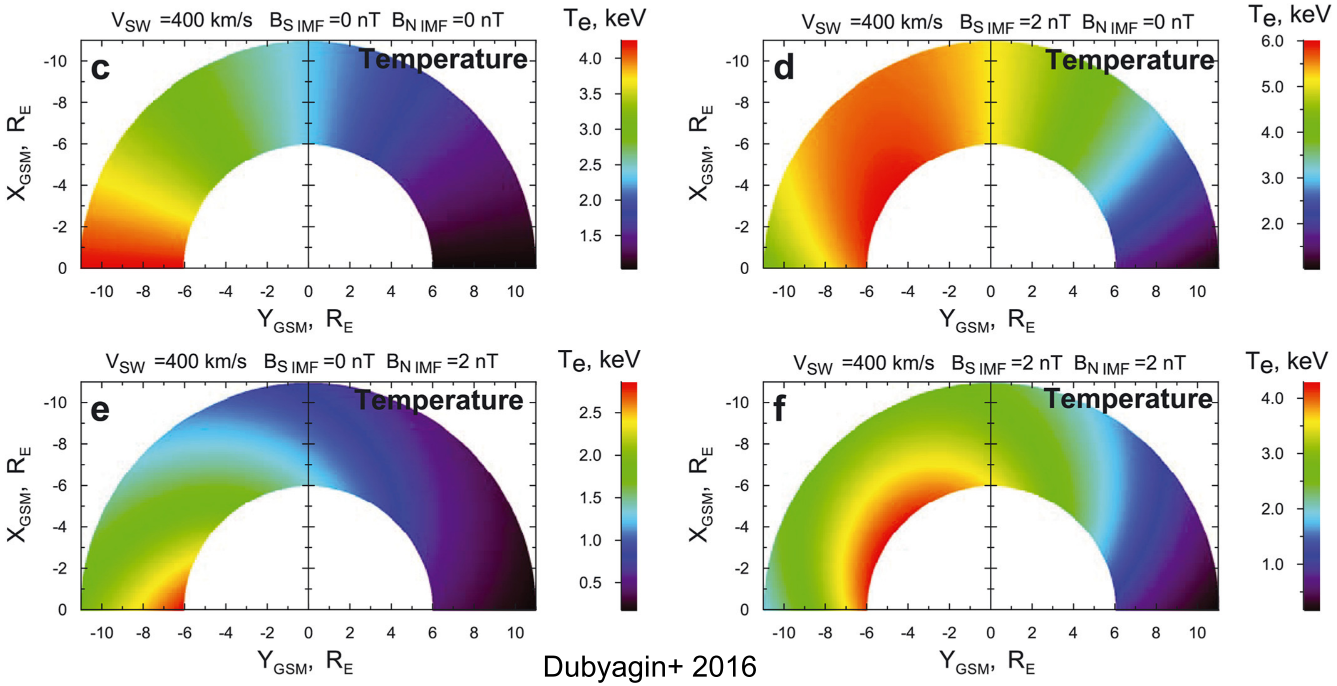
Baseline empirical models

Modelled with Geotail



Tsyganenko & Mukai 2003

Modelled with THEMIS



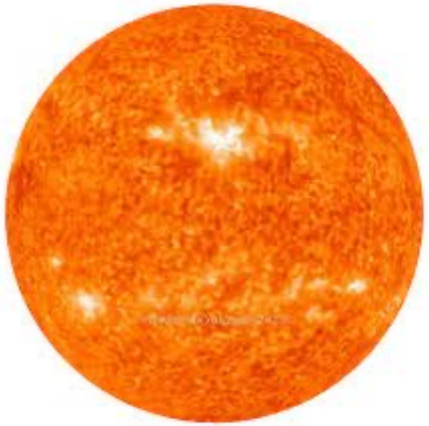
Dubyagin+ 2016

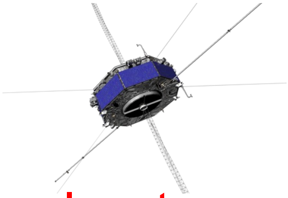
<u>Why then work on this?</u>

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 take advantage of non-linear relationships
5. Never validated with statistical/ML practices (i.e., train/test splits)

This presentation: Modeling of Ti/Te and ni

Where are we & what are we doing?





Input:

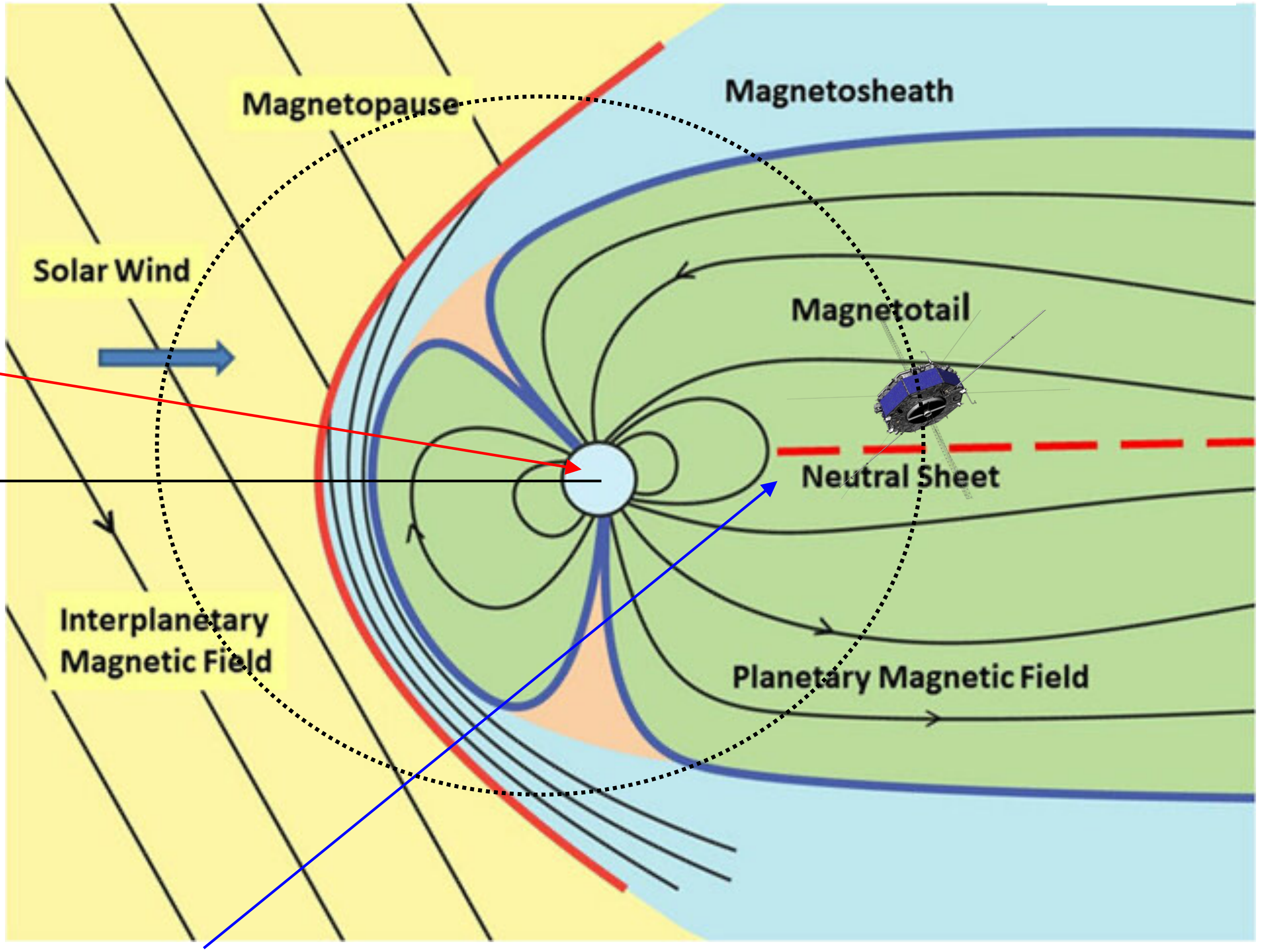
ACE/WIND/DSCVR

(L1)

+Geomagnetic

$200R_E$

Credits: NASA

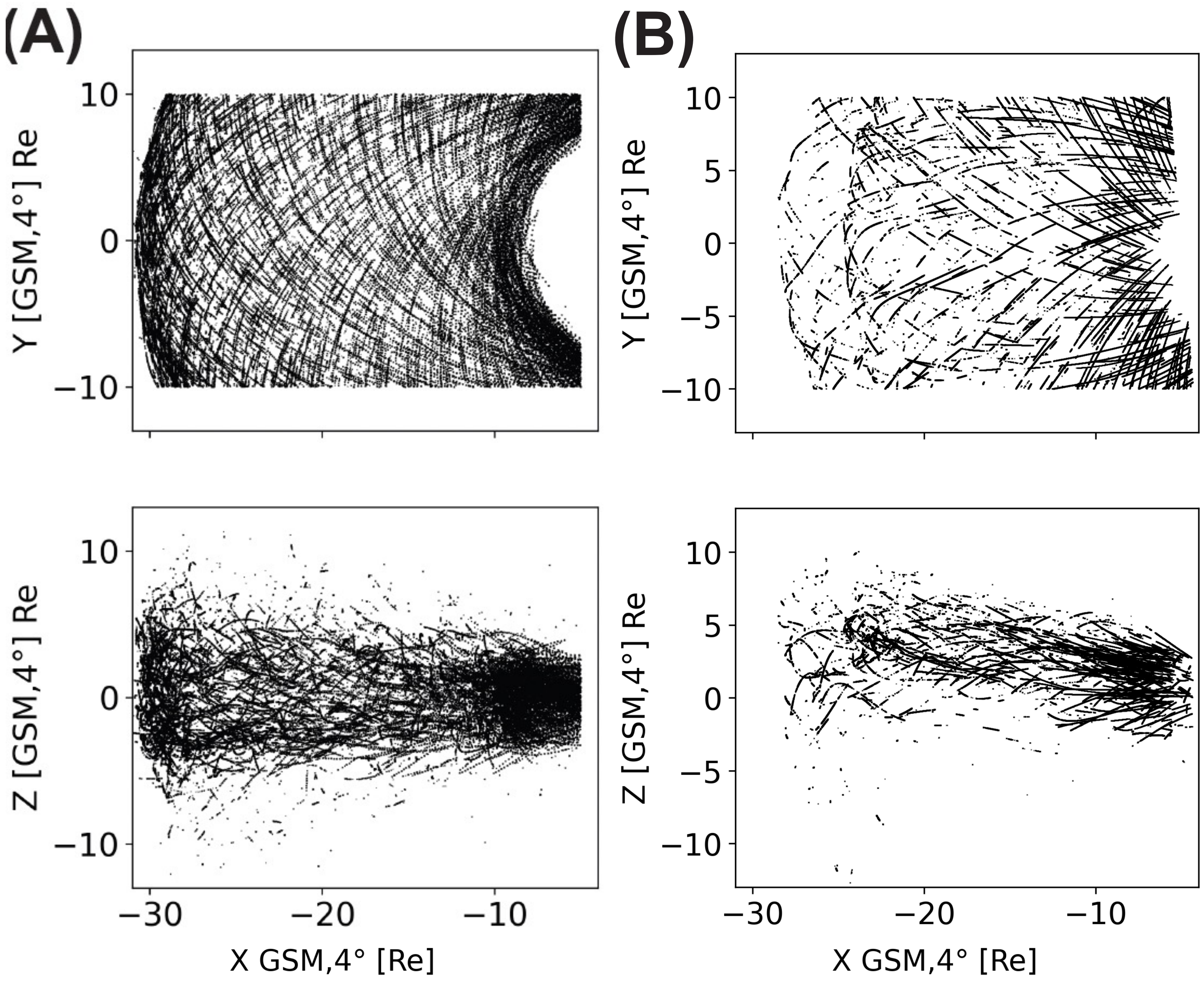


Output: *In-situ* properties from: MMS, Geotail, THEMIS

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

$R_E = 6371.2\ [km]$

The dataset (output – Central Plasma Sheet)



(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

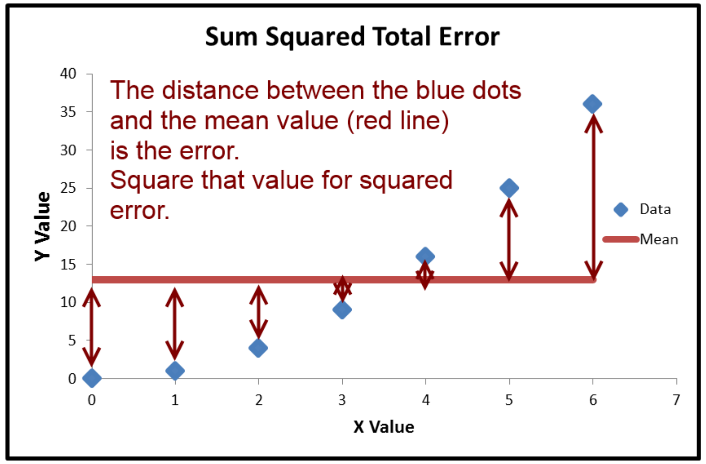
```
99/99 ━━━━━━━━━━━━━━━━━━━━ 0s 808us/step
explained_variance:  0.019
median absolute error:  0.11
r2:  -0.01
MAE:  0.157
MSE:  0.055
RMSE:  0.235
Cor:  0.533
```

A complex and intriguing model

VS

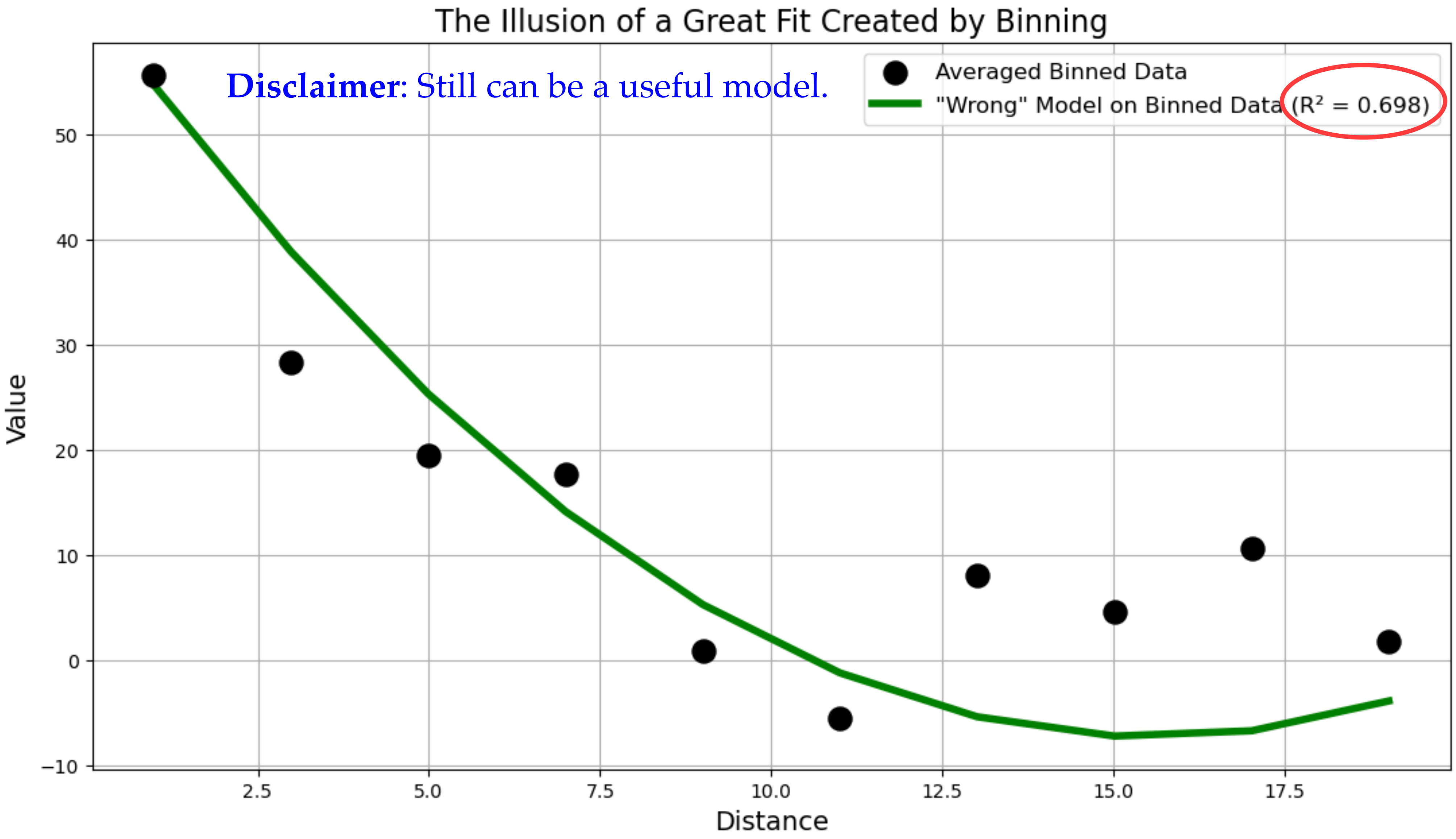
```
99/99 ━━━━━━━━━━━━━━━━━━━━ 0s 778us/step
explained_variance:  0.0
median absolute error:  0.14
r2:  0.0
MAE:  0.17
MSE:  0.055
RMSE:  0.234
Cor:  0.0
```

np.mean()

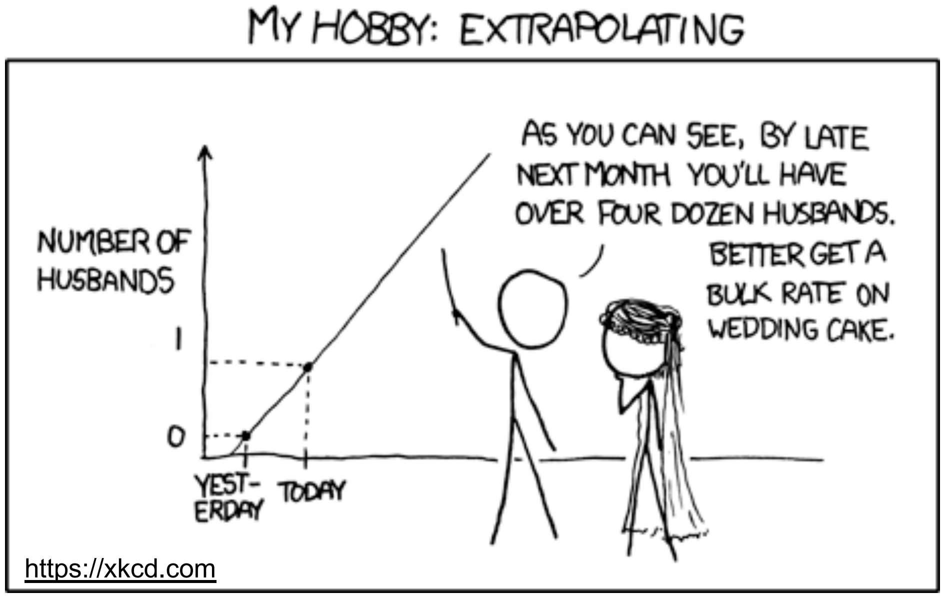


When $R2$<0, the horizontal line explains the data better than your model (i.e., mean of observed).

The illusion success through averaging our data



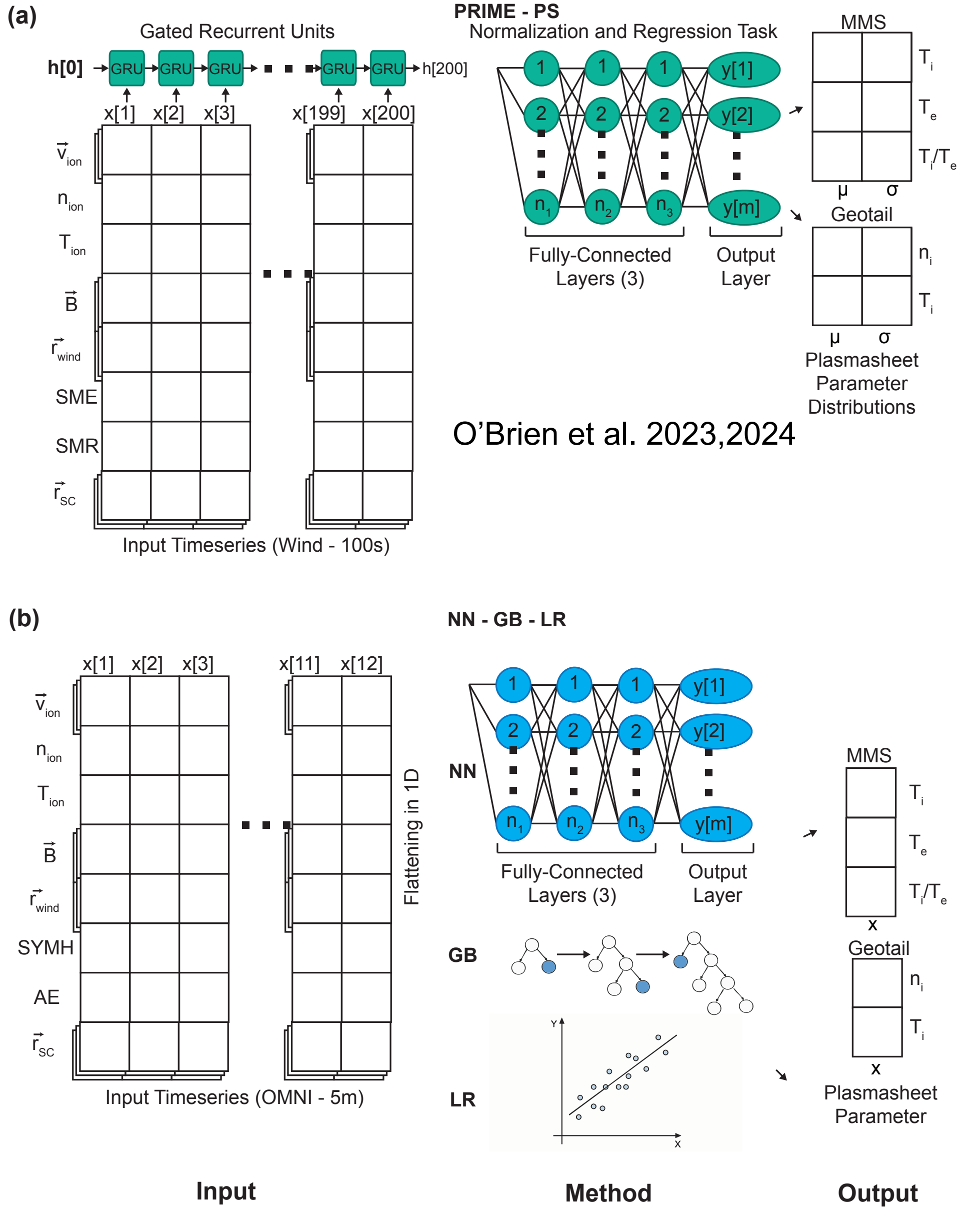
Underlying process is structured, but the measurements and their local variations are hidden.



https://xkcd.com

Results

Methodologies & input space



Raptis, O'Brien et al., 2026 (under prep.)

- **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)

<u>Answering hypothetical questions:</u>

✓ *Also tried different error functions, optimizers, hyperparameters etc.*

✓ *And different imbalanced learning techniques*

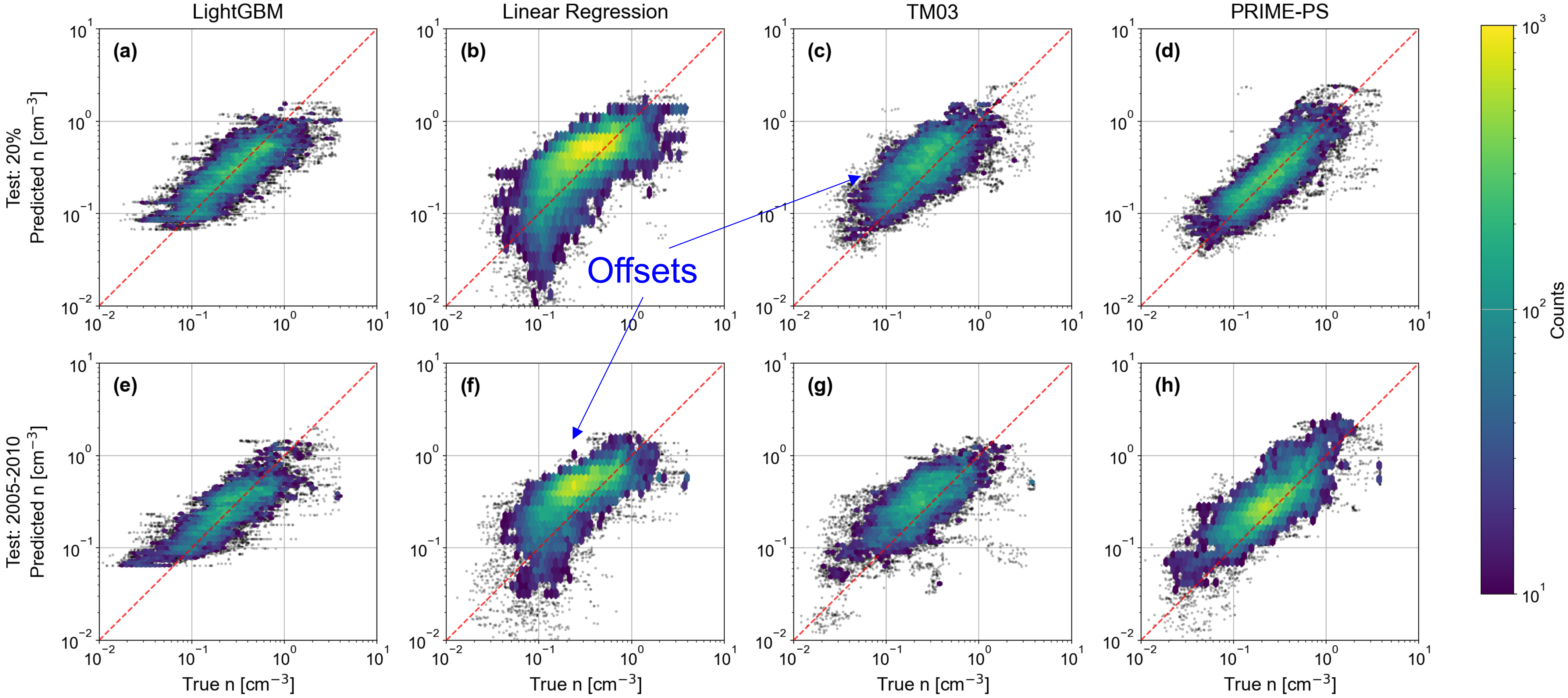
<u>Key Takeaway:</u>

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)



Raptis, O'Brien et al., 2026 (under prep.)

Reminder: Close to x=y → good

Metrics using Test set (20% of data)

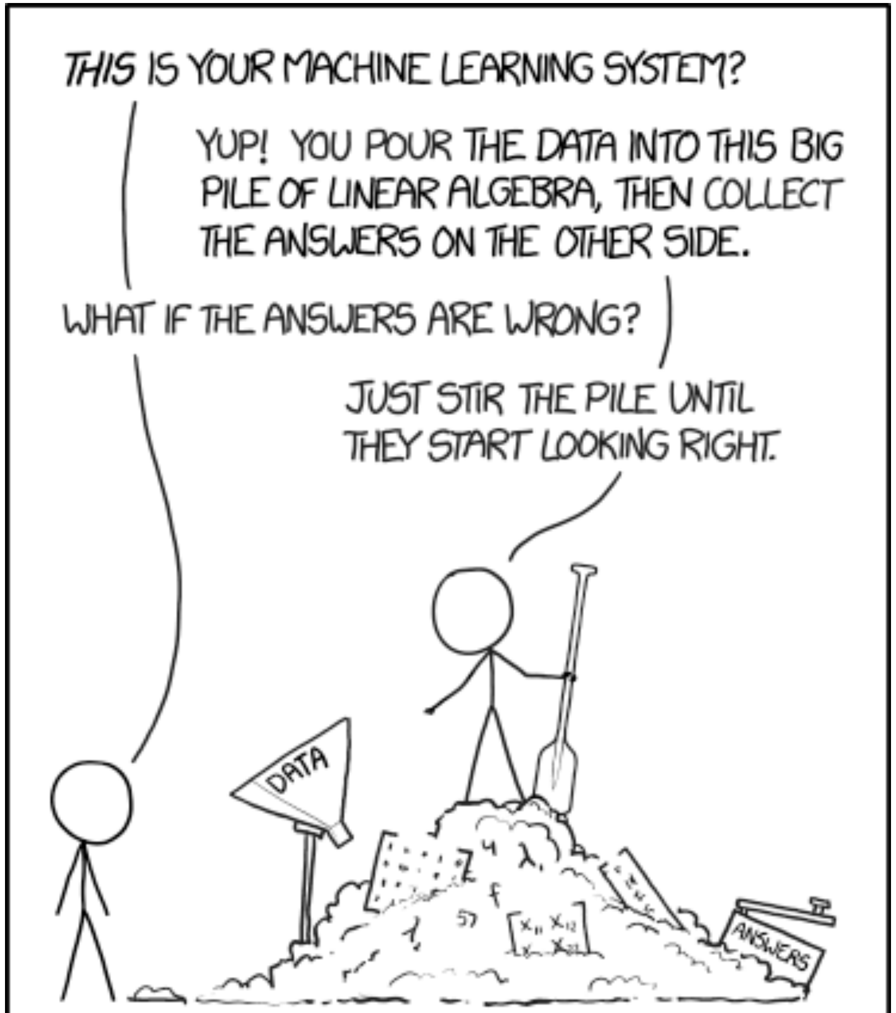
Method	MAE	R^2	r
LightGBM	0.129	0.373	0.631
Neural Net	0.152	0.325	0.603
Linear Reg	0.173	0.265	0.620
PRIME-PS	**0.113**	**0.453**	**0.707**
TM03	0.163	0.208	0.570

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 (discuss later)?

➢ Since PRIME-PS was statistically better, and Gradient boosting can't be used for modeling, we only keep this for next parts of analysis.

Raptis, O'Brien et al., 2026 (under prep.)

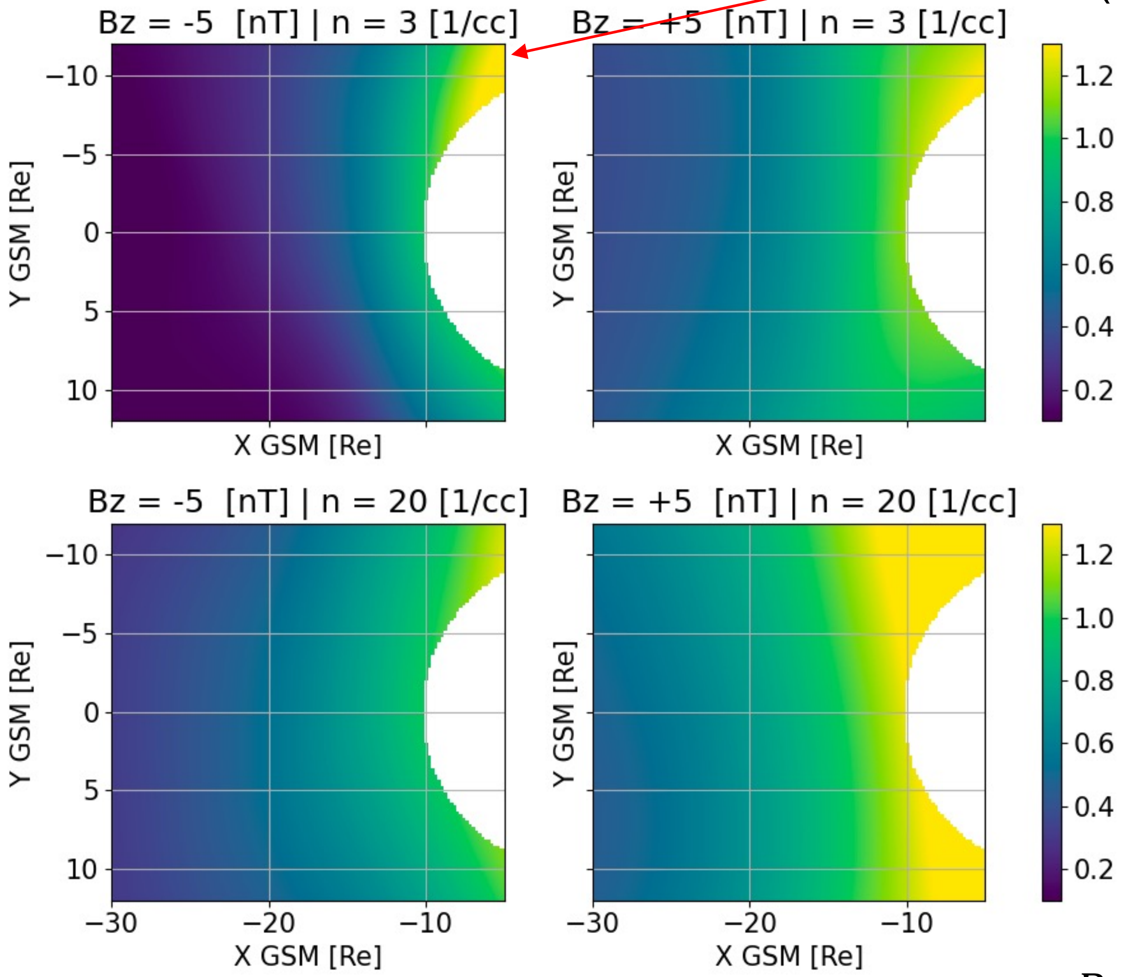


https://xkcd.com

Modeling Efforts

Modeling Density with Geotail | 2D Synthetic Maps

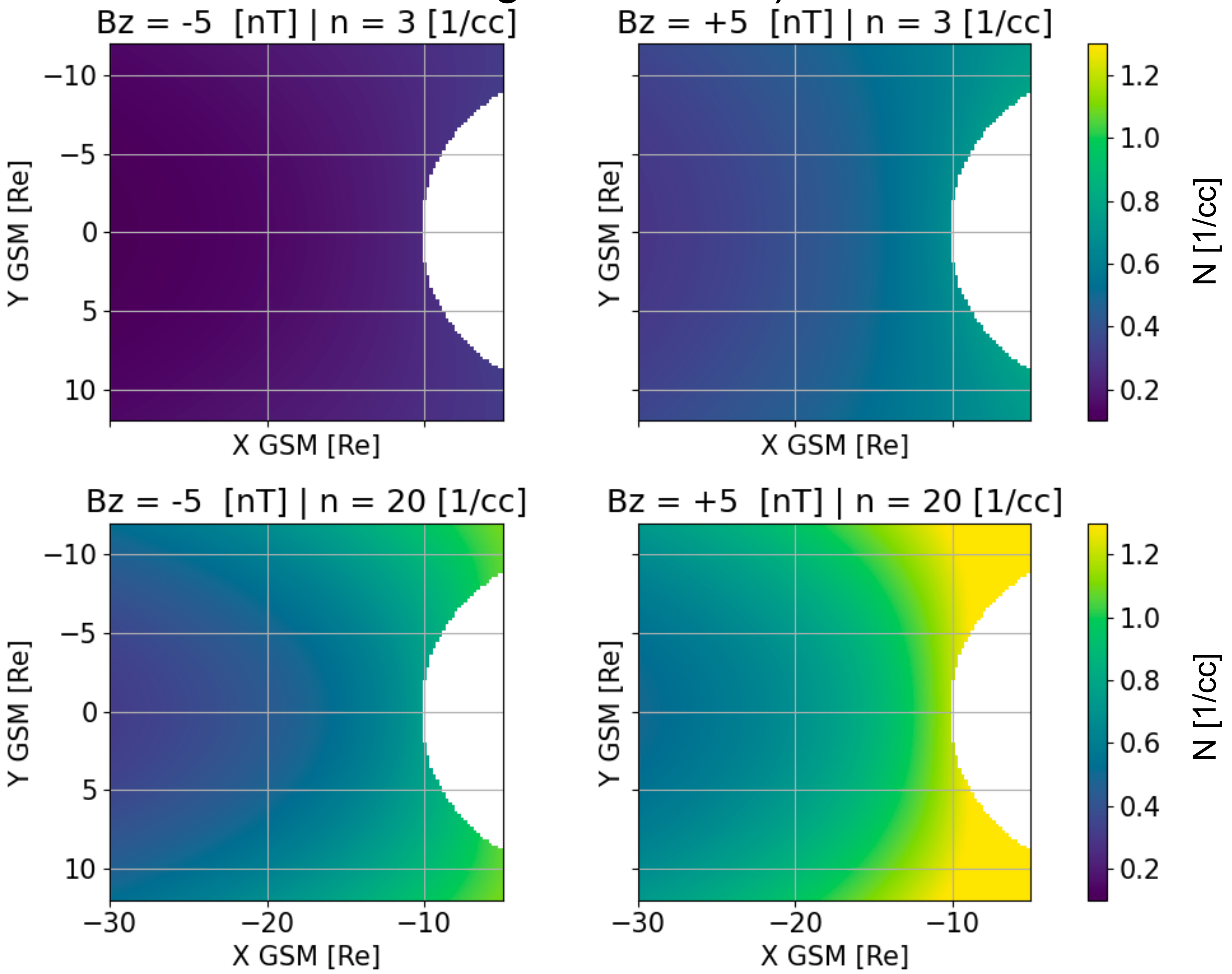
Asymmetries Introduced In agreement with previous studies
(Wing & Newell, 1998; C.-P. Wang et al., 2007)



Neural Networks modeling

$Bz > 0$ ↑ n_{ps}

n_{sw} → ↑ n_{ps}

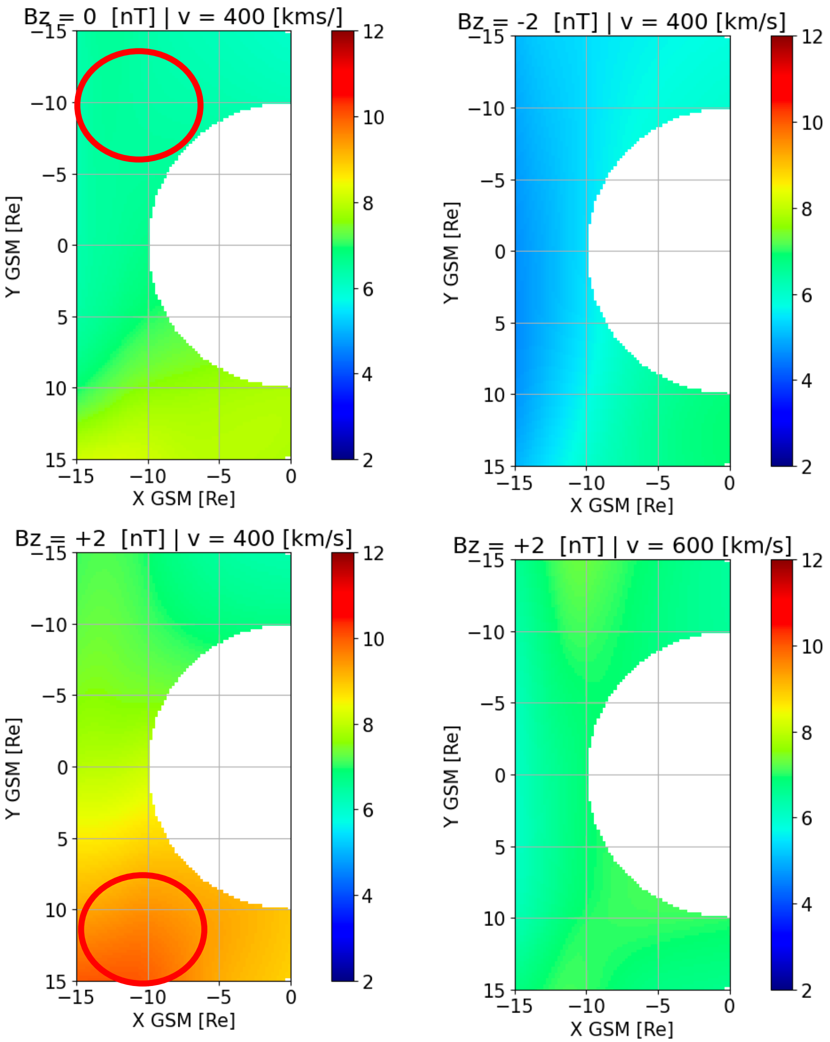


Empirical modeling (TM03)

Raptis, O'Brien et al., 2026 (under prep.)

Modeling Temperature Ratios with MMS | 2D Synthetic Maps

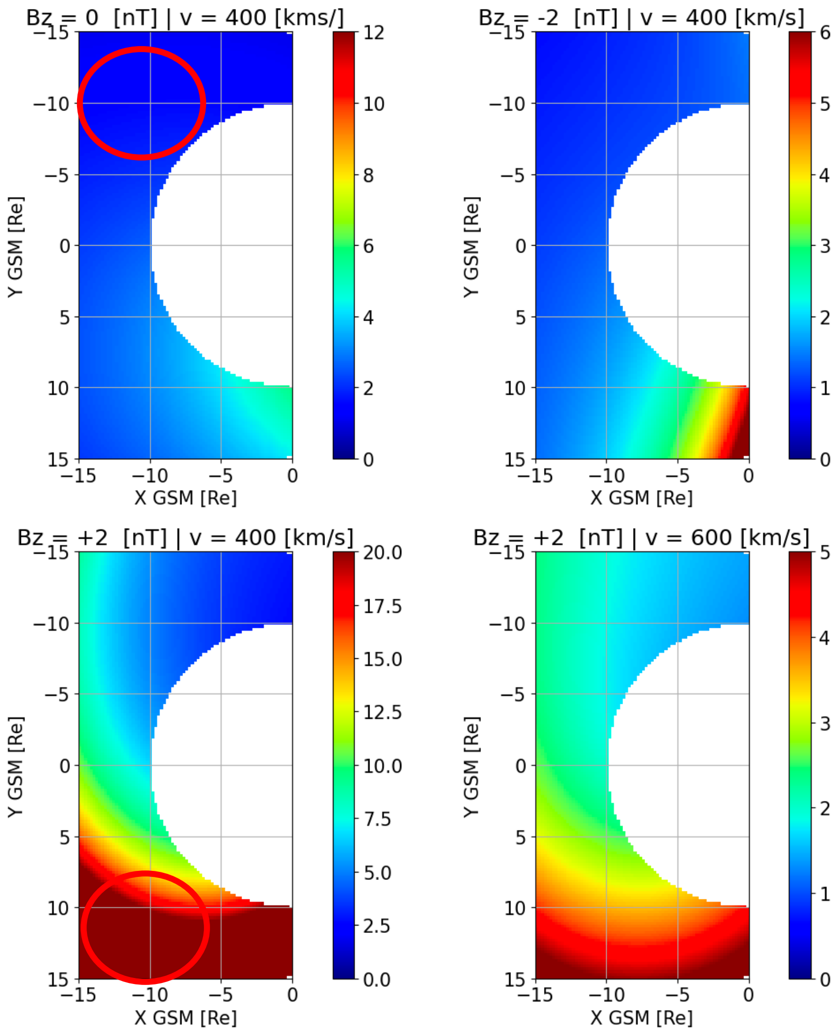
Note: Agreement Wang et al., 2010 with dusk Ti/Te higher than dawn (Using THEMIS)



Neural Networks modeling

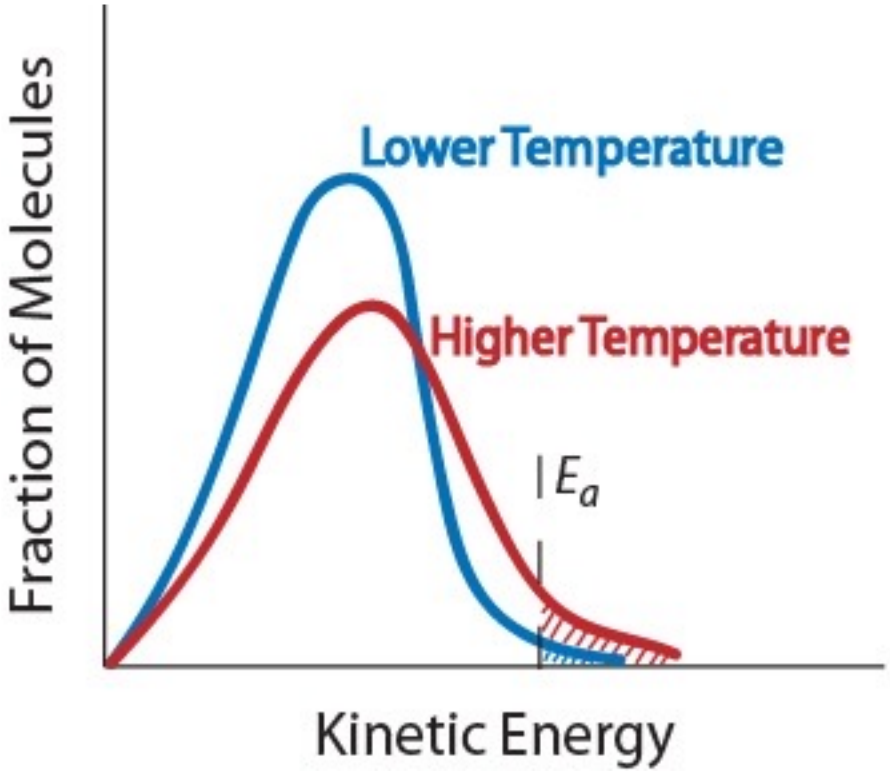
<u>Pros:</u>
+No extreme values
+Asymmetries shown
+ Coherent physical picture

<u>Cons:</u>
- Not easily available analytical form



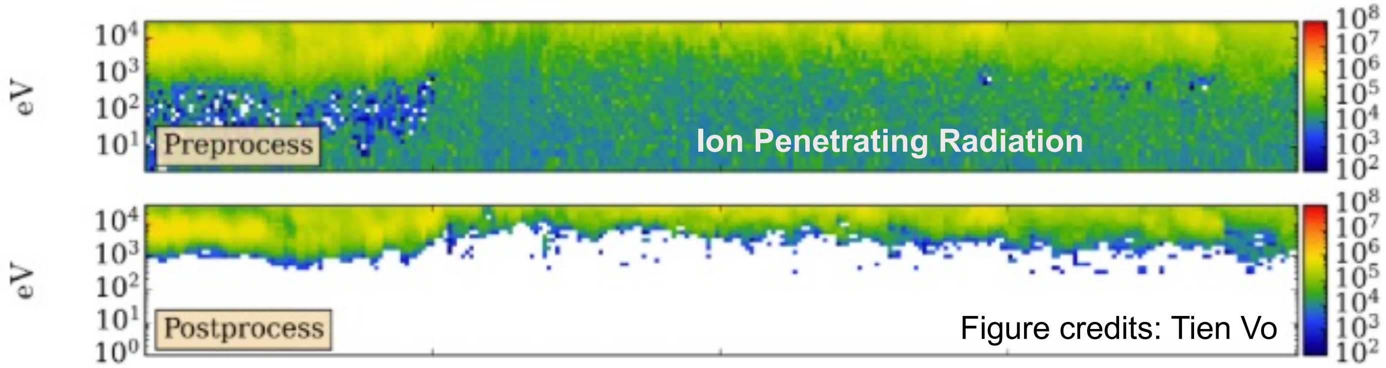
Empirical modeling (TM03/DSGR16)

Community Reminder on Temperature



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

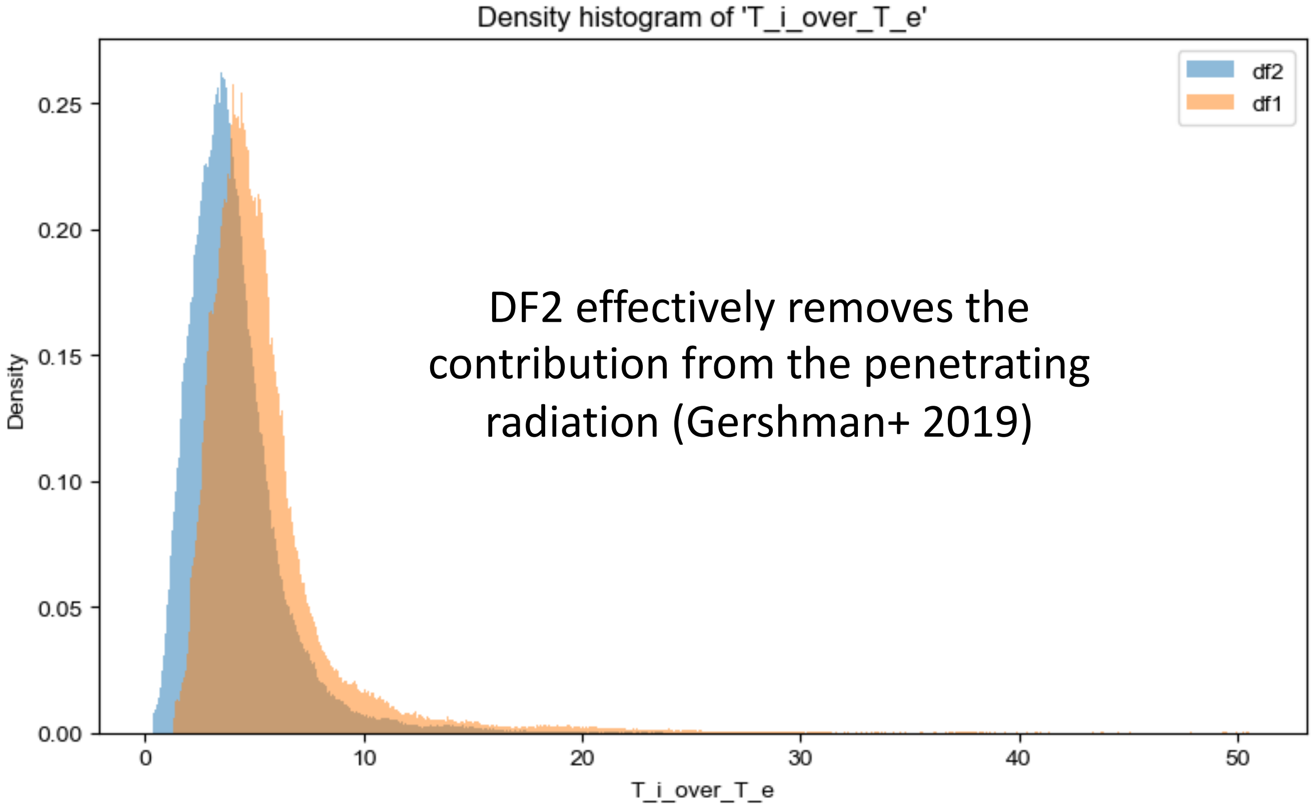
- Temperature is the 2nd plasma moment
- The higher the moment, the more uncertain because you rely more on the poorly sampled tails of the distribution.
- So, 0 and 1st moment (**Density and Velocity**) are usually ok, but Temperature, we got to be careful



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

MMS Ti/Te plasmasheet 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

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

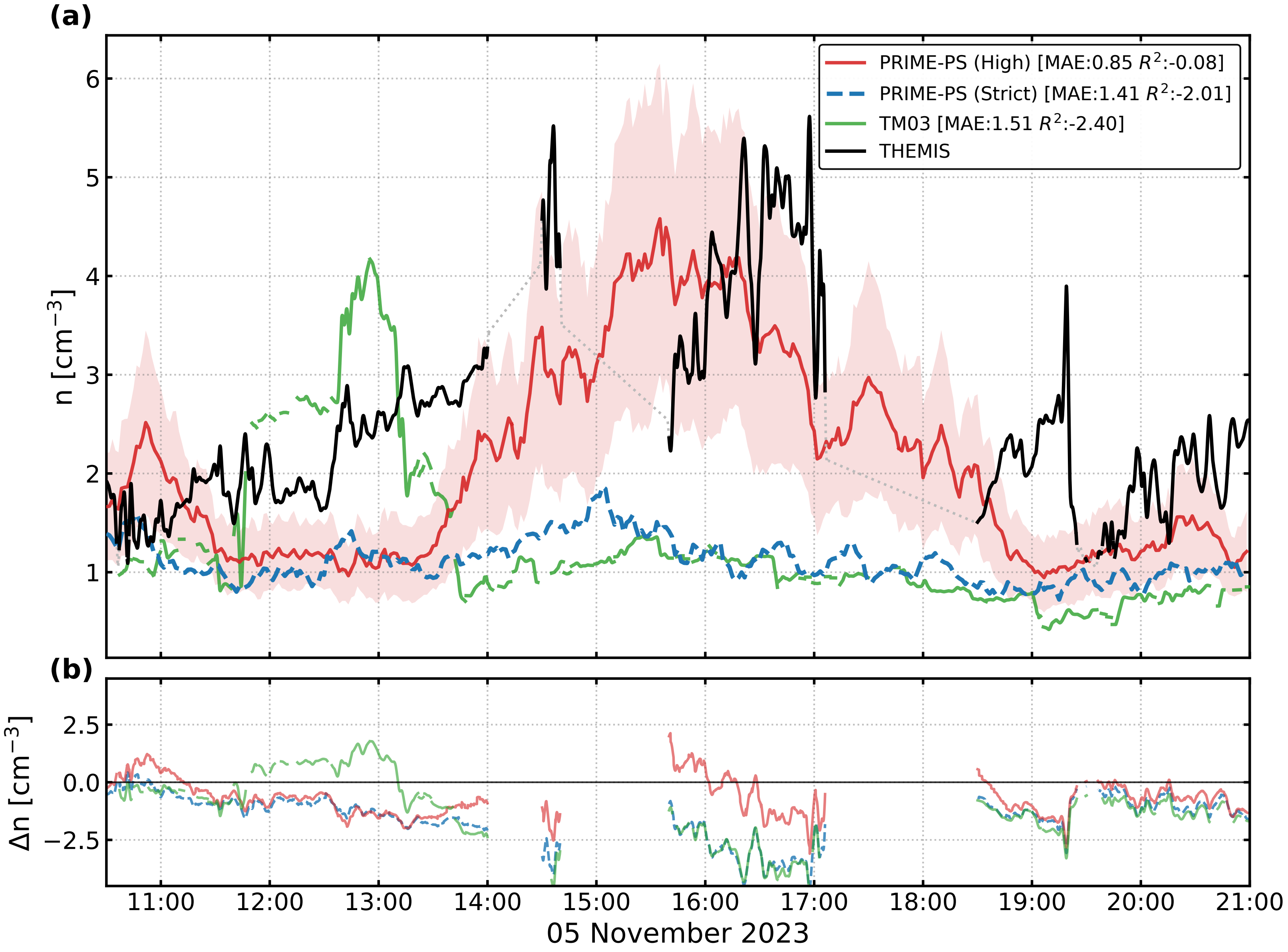
The Problem: We use static thresholds for dynamic environments.

The Risk: Therefore we can mistakenly remove crucial "stormtime plasmasheet" datapoints.

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)

Test case of a storm (05 Nov 2023) – Density (Geotail Model)



Process:

High: Includes manually picked high-density intervals from Geotail

Strict: Normal threshold-based classification of plasma sheet

Test:

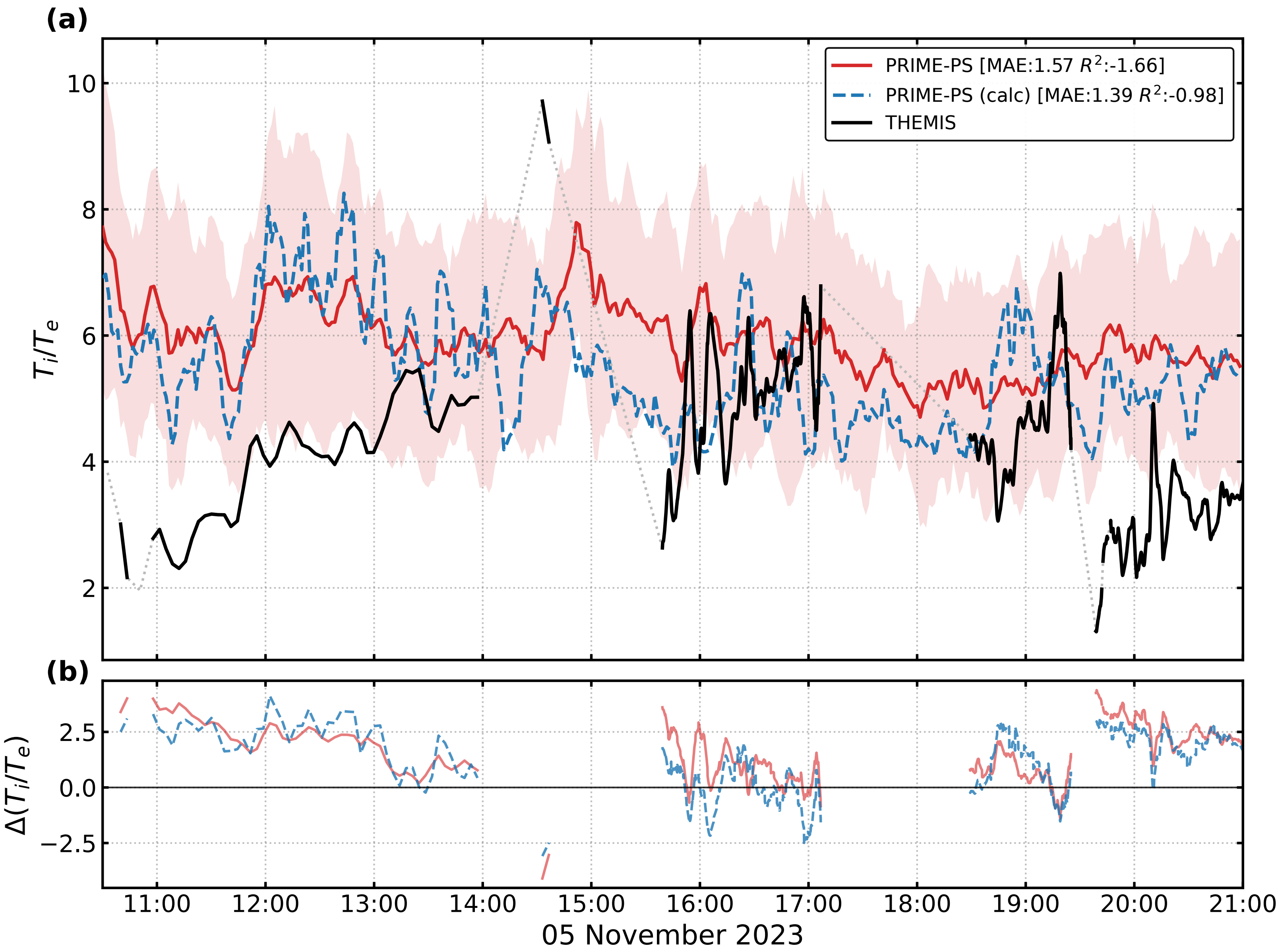
THEMIS observations

Results

- ML model (high dens): 0.85 [1/cc]
- ML model (normal): 1.41 [1/cc]
- TM03: 1.51 [1/cc]

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

Test case of a storm (05 Nov 2023) – Ti/Te (MMS Model)



Process:

- Directly Predict Ti/Te as trained from MMS data
- Calculate Ti/Te using the predicted Ti and Te from MMS data

Test:
THEMIS observations

Question to ask ourselves:
How bad would Ti/Te~4 would be?

Answer we reply to ourselves:
Fair, but that's boring.

Note: values <1 cm⁻³, are boundary layer transitions, gray shaded linear interpolation

Preliminary Feature analysis

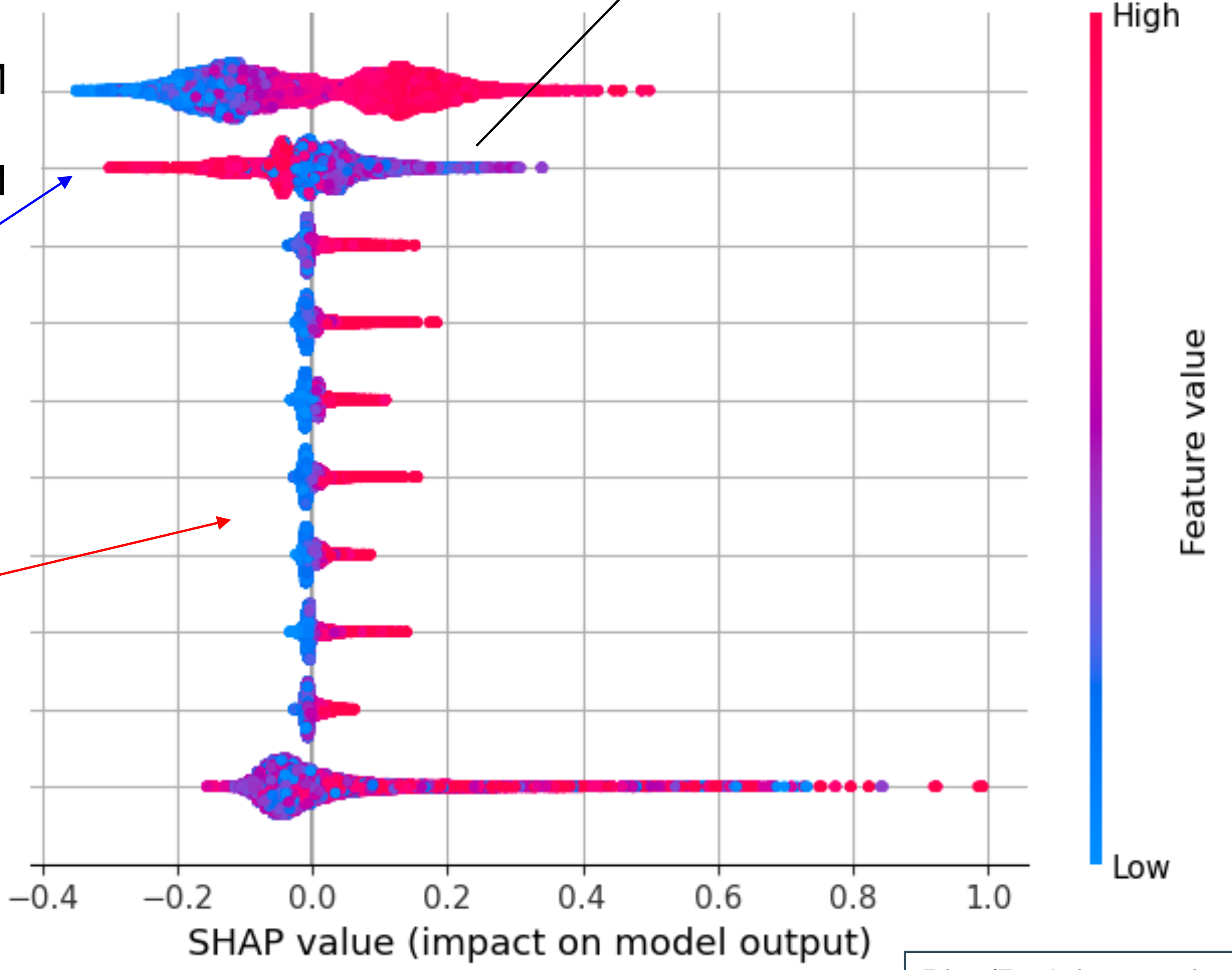
Feature Importance Analysis

*Higher density **close to earth and at dawn***

Answer: In most cases (statistically):

Model is predominantly driven by spacecraft location

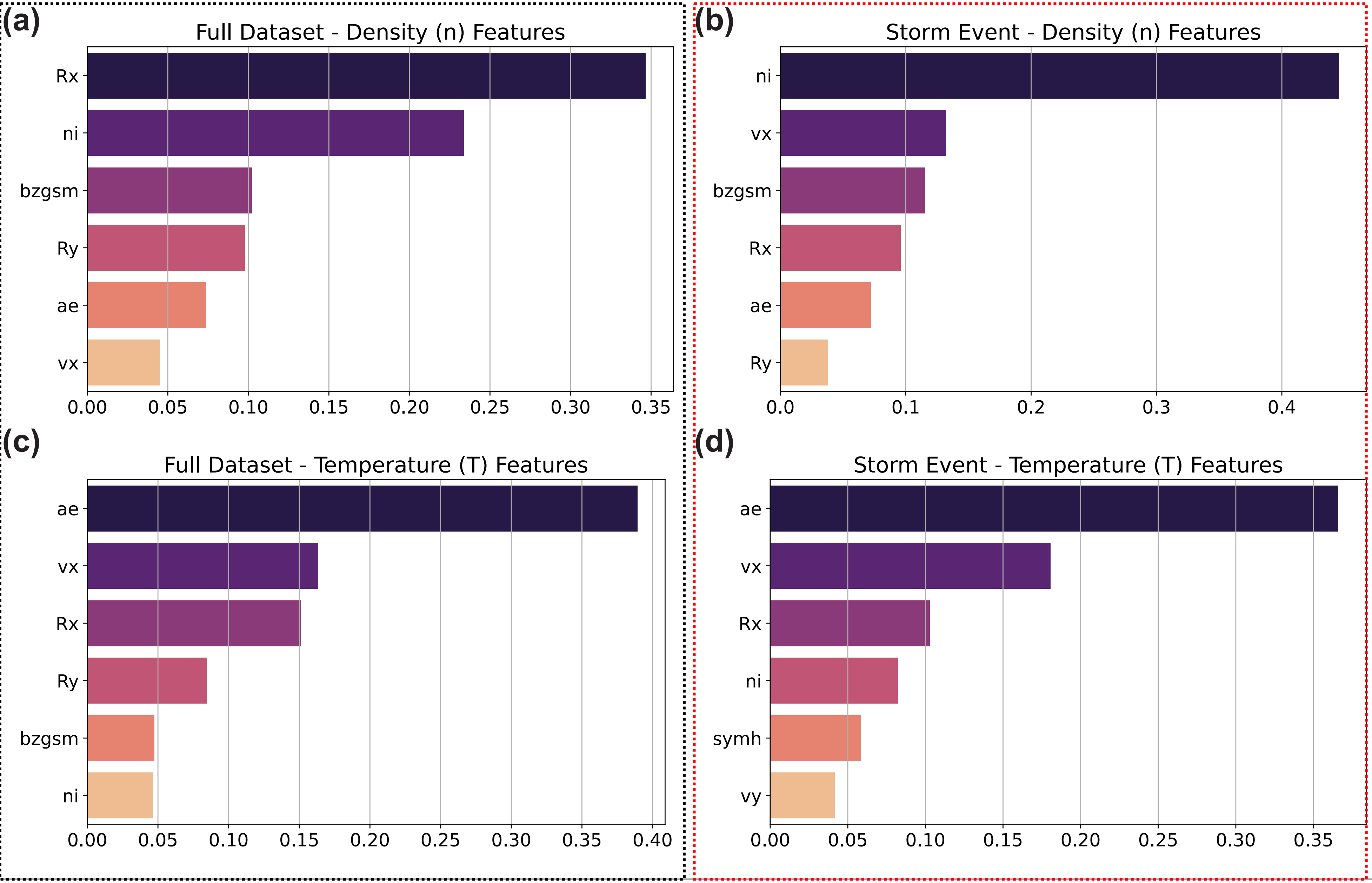
Solar wind input has lower effect, although cumulative history is still important



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

- Generally less SW driven
- T more external
- Density dictated by location

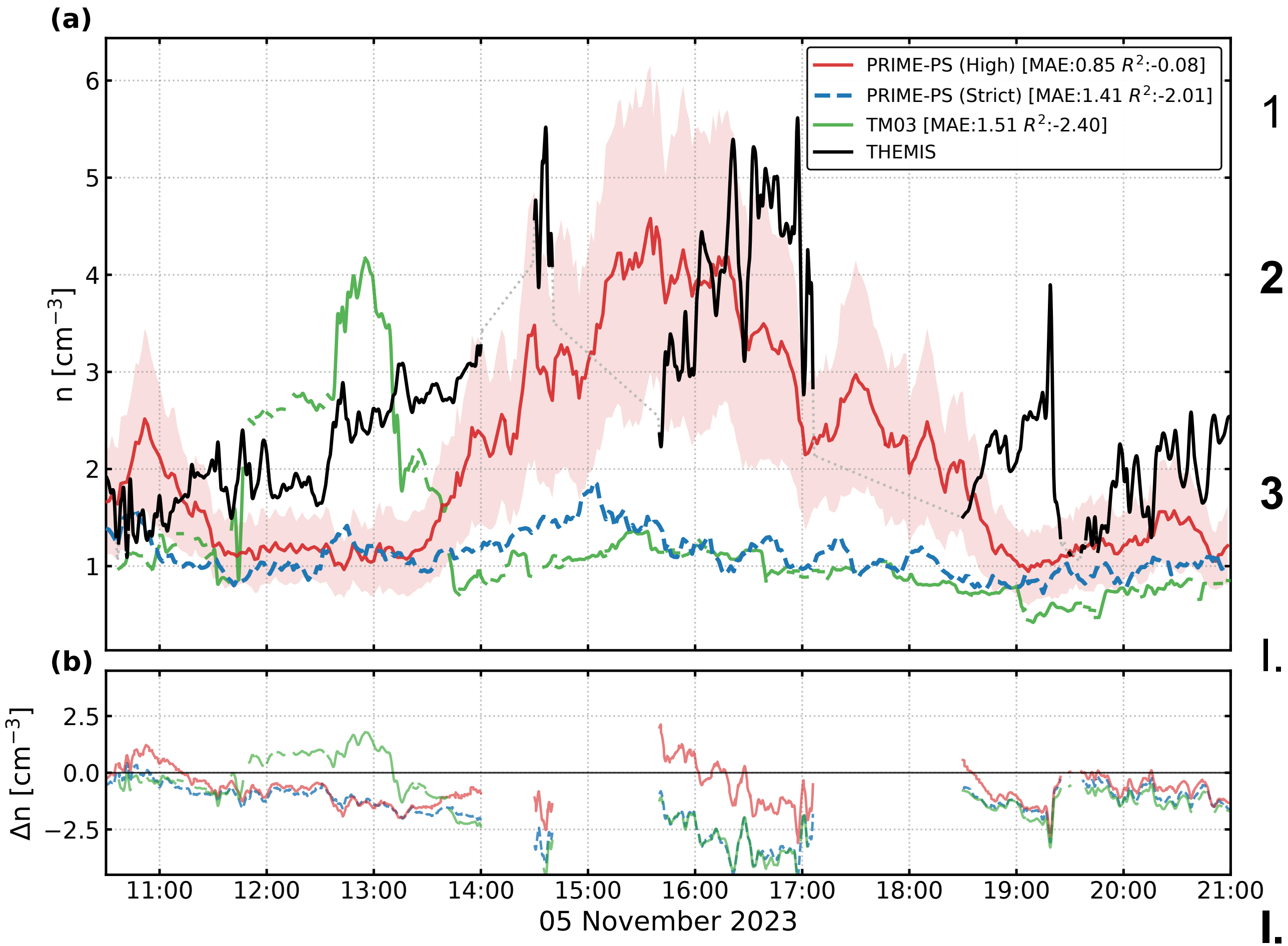
Storms:

- More externally driven
- Location less important

Caveats

- Intercorrelation between variables (e.g., ae correlated to vx)
- Hidden correlations of SW type (dense solar wind is typically slower)
- Combines all time history in one variable

High Level Takeways



Main Points

1. Global **convection** is different **during storms**, with **duskward Bz and dawnward Vx enhancements**.

2. **Storm-time BBFs transport more magnetic flux** than quiet-time BBFs, **affected mostly by the global increase in Bz** across the plasma sheet.

3. **BBF** observations appears to **occur more at dusk during quiet times and shift dawnward during storms**.

I. A **machine-learning plasma sheet mode**l, driven by solar wind input and geomagnetic state, **outperforms earlier empirical models, generating a physically accurate picture**.

I. **Sparse data during rare conditions still degrades the model performance** during extreme events such as storms.

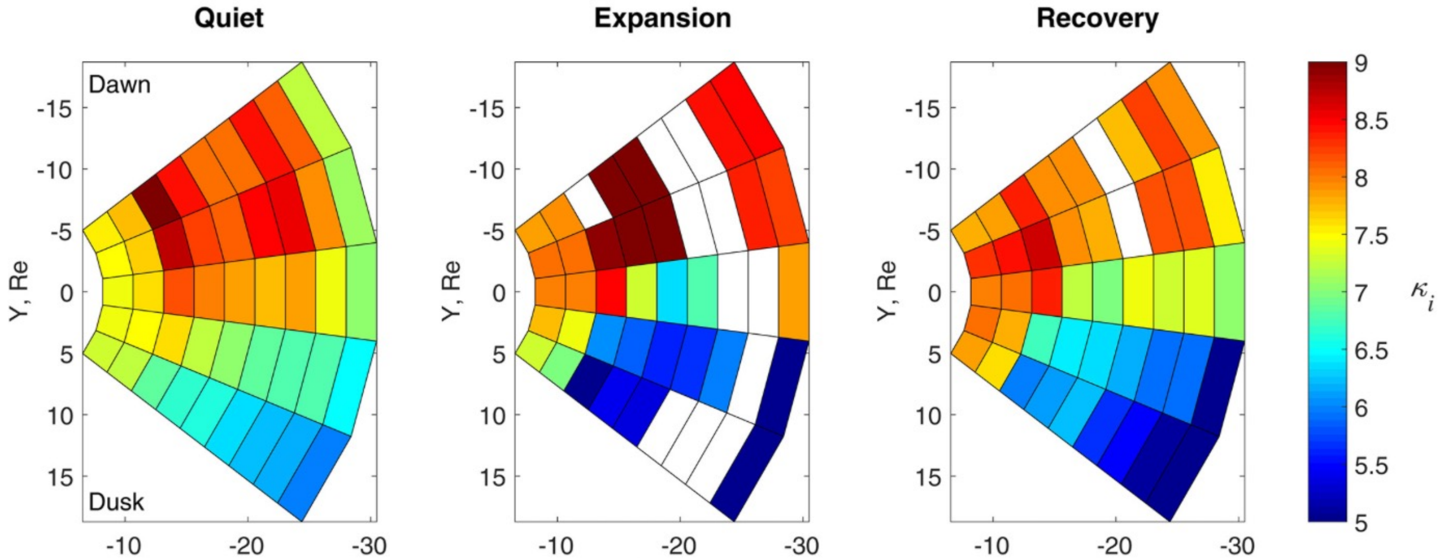
Future Steps & Discussion

Future Steps

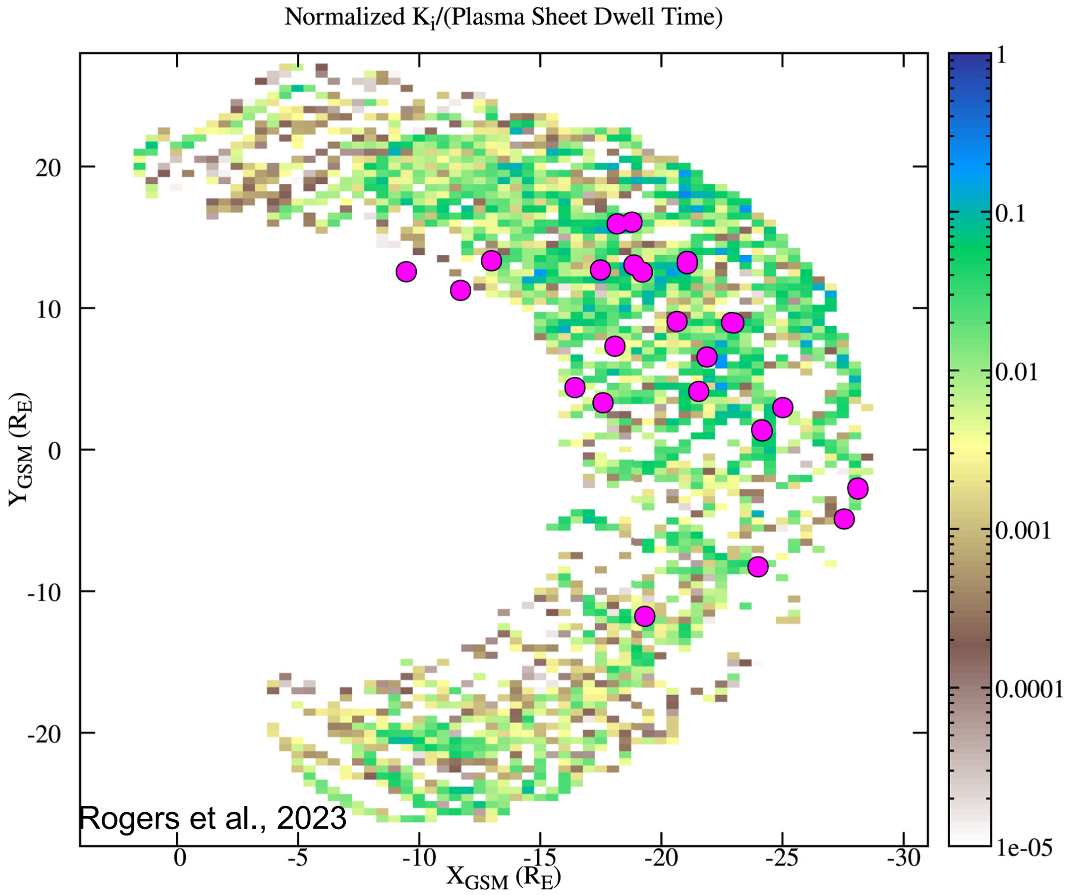
- **Extend MMS coverage** from 2015-2022 to 2015-2026 and **add THEMIS intervals** to enlarge the statistical sample. This can help us **understand dawn/dusk asymmetries and improve model performance.**

- Extend the analysis to **substorm activity** and test model performance.

- Evaluate more complex physical quantities, such as **non-Maxwellianity** of particle distributions and **magnetic curvature**.



Espinoza et al., 2018



Rogers et al., 2023