

Modeling Earth's Plasma Sheet using Machine Learning

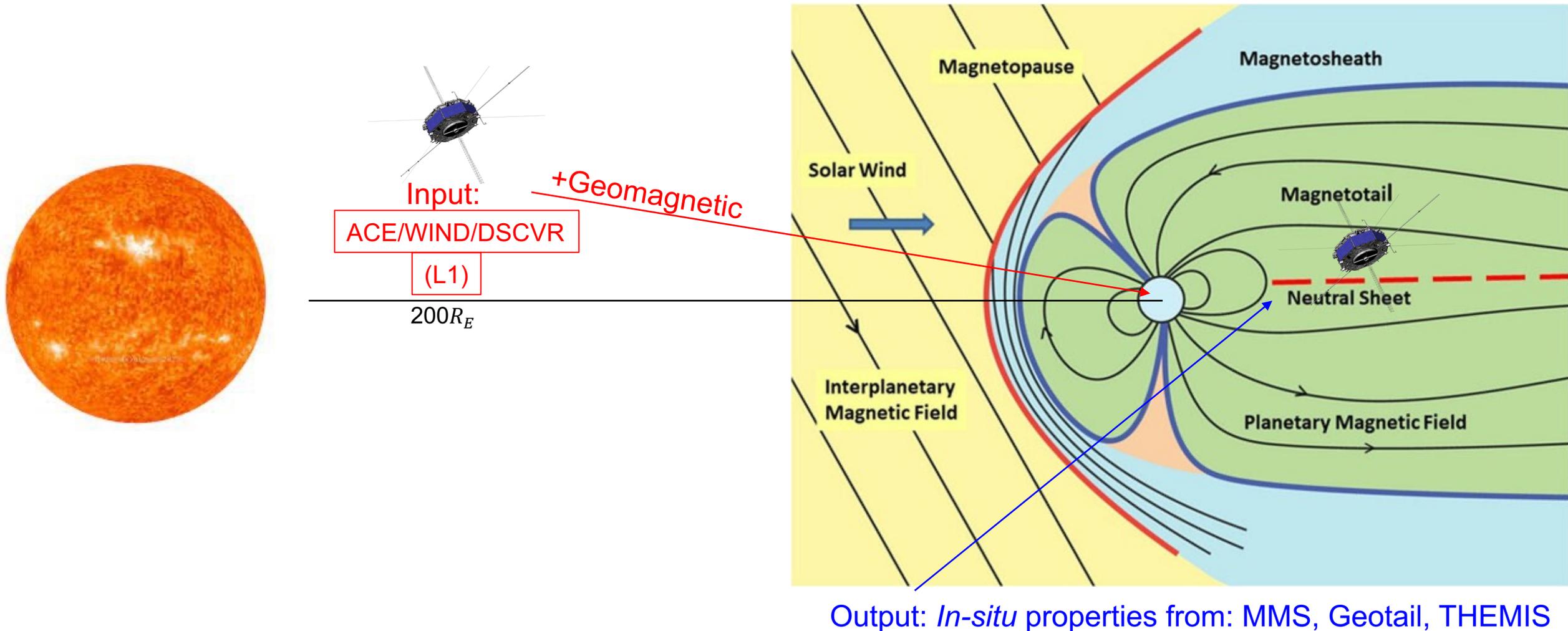
Savvas Raptis¹, Connor O' Brien^{1,2}, Kareem Sorathia¹, Viatcheslav Merkin¹, Shinichi Ohtani¹, Louis Richard³, Simon Wing¹

¹ APL/JHU, Laurel, MD, US

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

³ Swedish Institute of Space Physics, Uppsala, Sweden

Where are we & what are we doing?

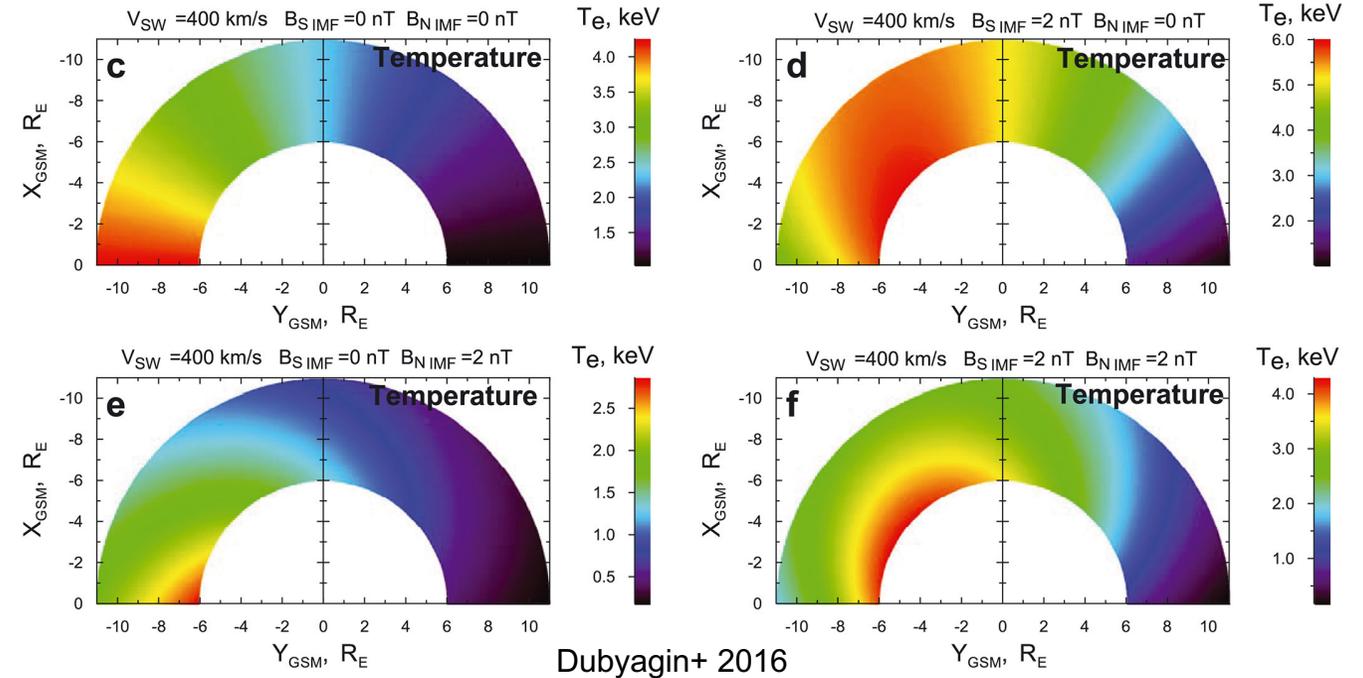
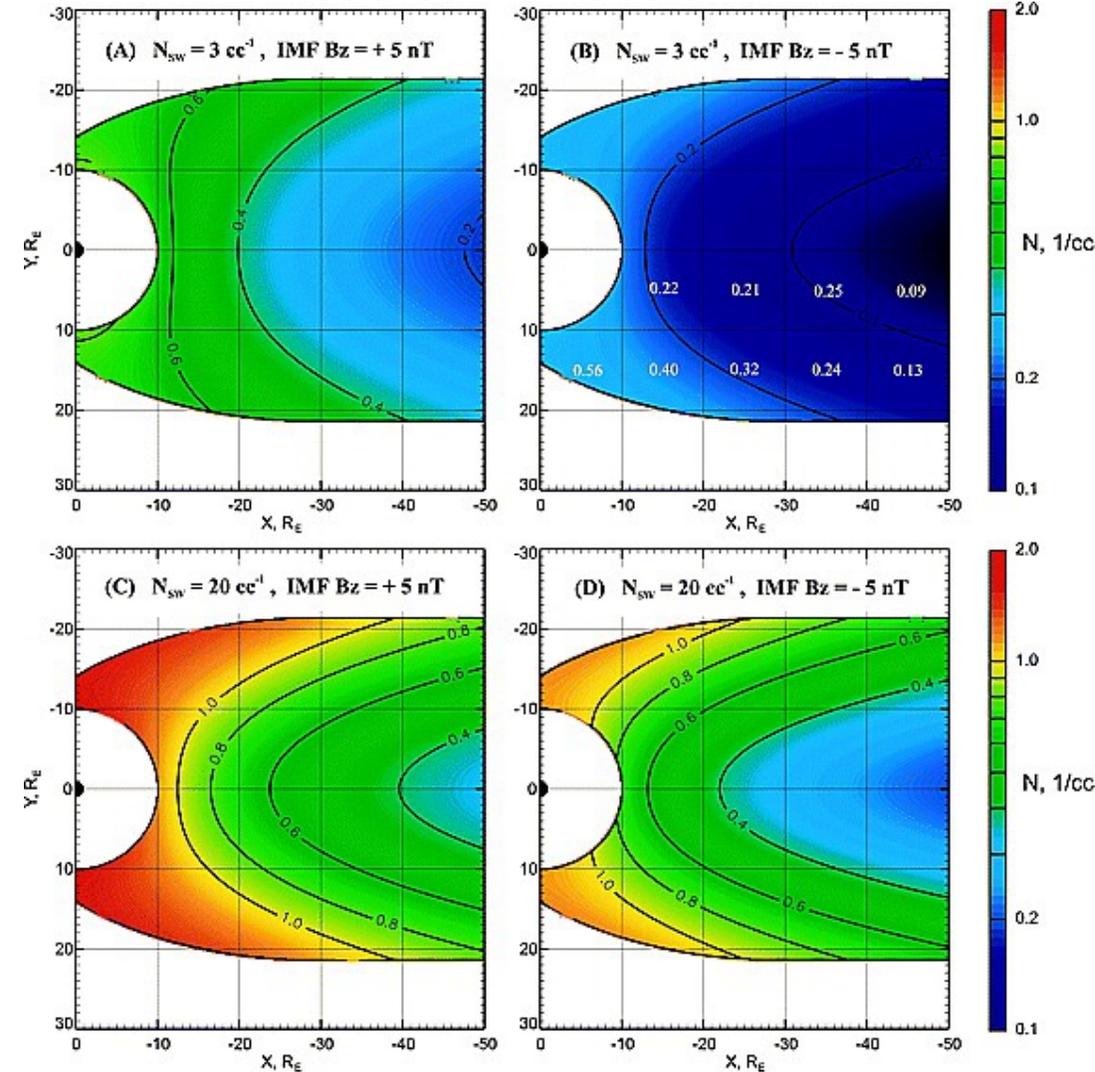


Goal: Model Plasmasheet properties based on driving (SW) and geomagnetic conditions

Baseline empirical models

Modelled with Geotail

Modelled with THEMIS



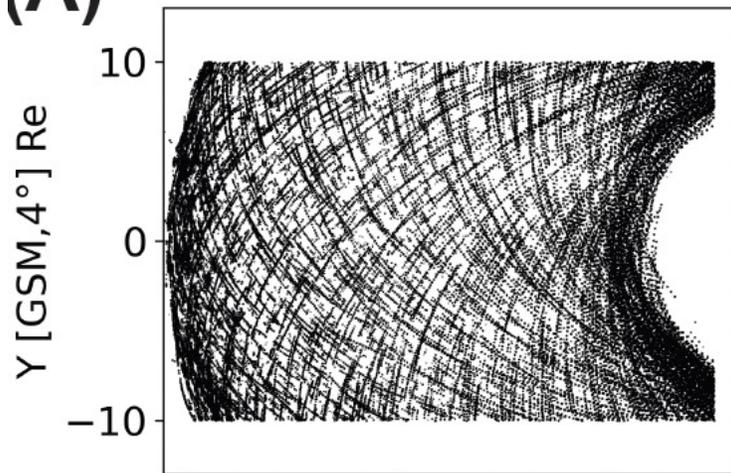
Why keep working on this?

1. More Data
2. No MMS Model
3. These models don't include time history
4. New methods can utilize non-linear relationship of input space
5. We can learn more now based on feature evaluation

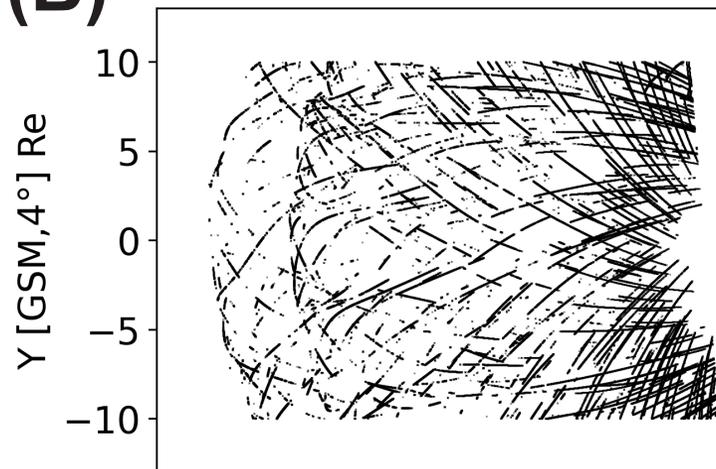
Tsyganenko & Mukai 2003

The dataset (output – Central Plasma Sheet)

(A)

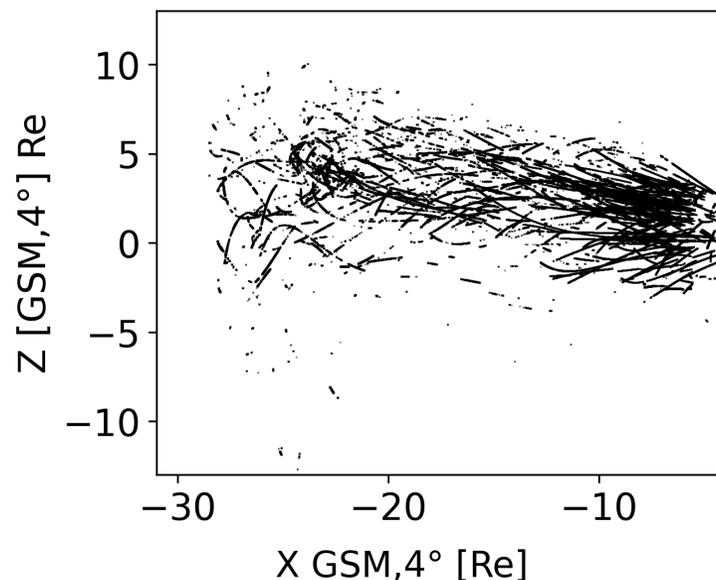
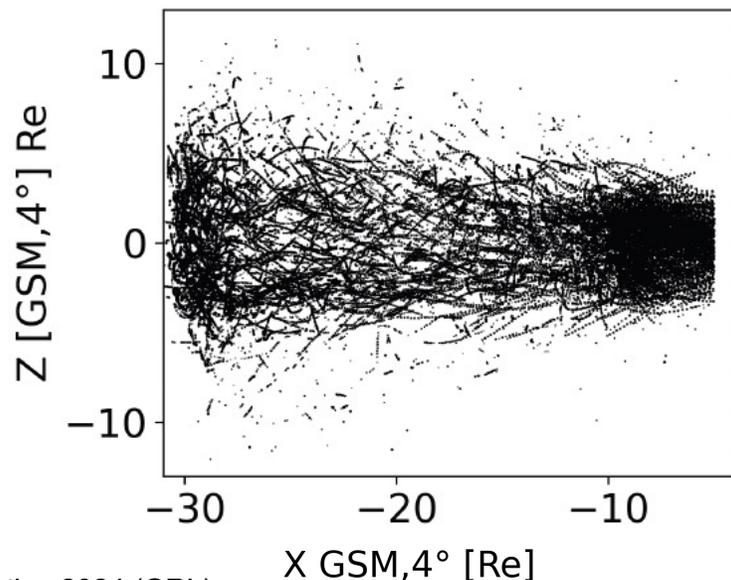


(B)

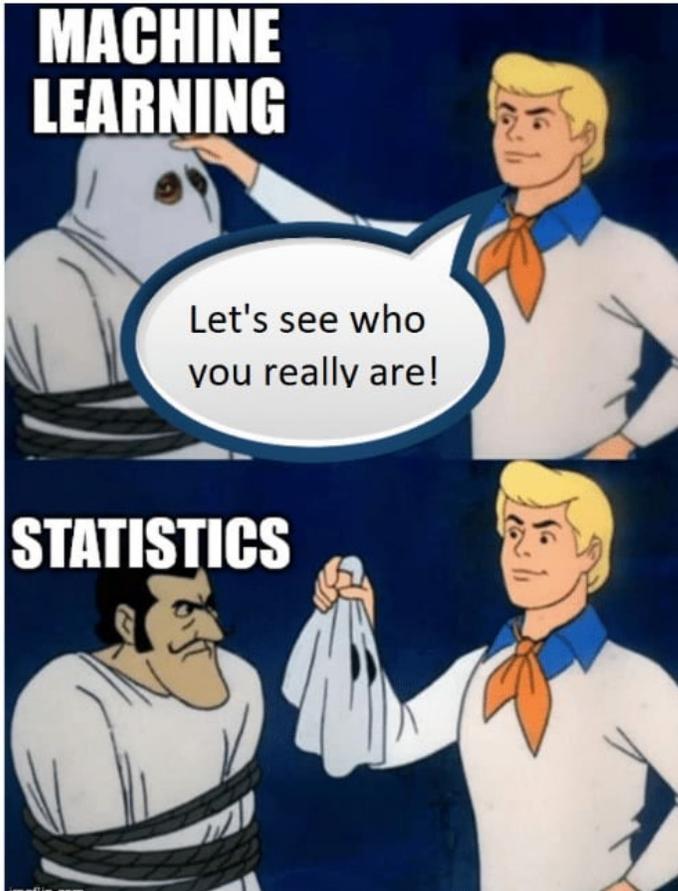


(A) Geotail (1994 - 2022)
>1 million points (~12s res)

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

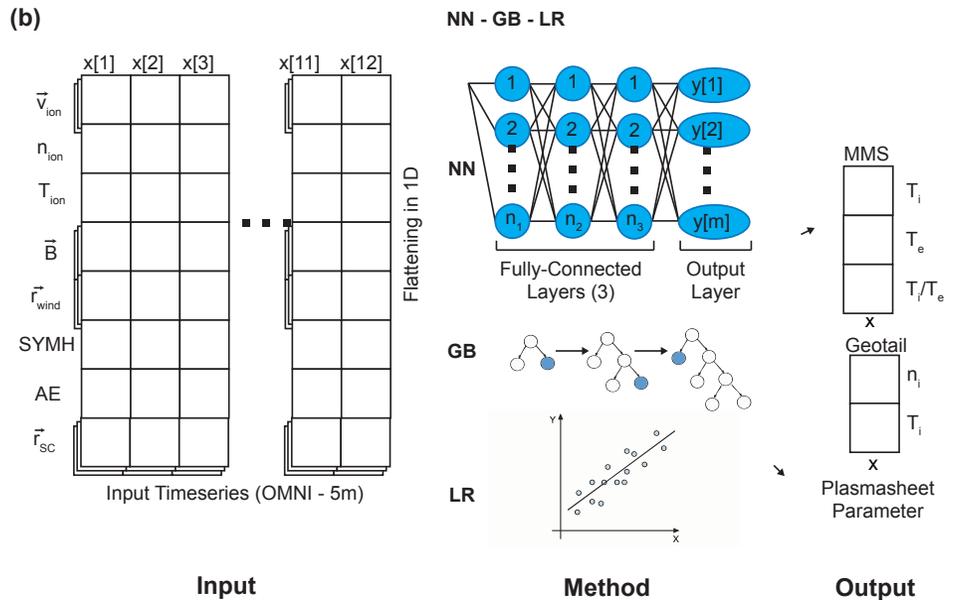
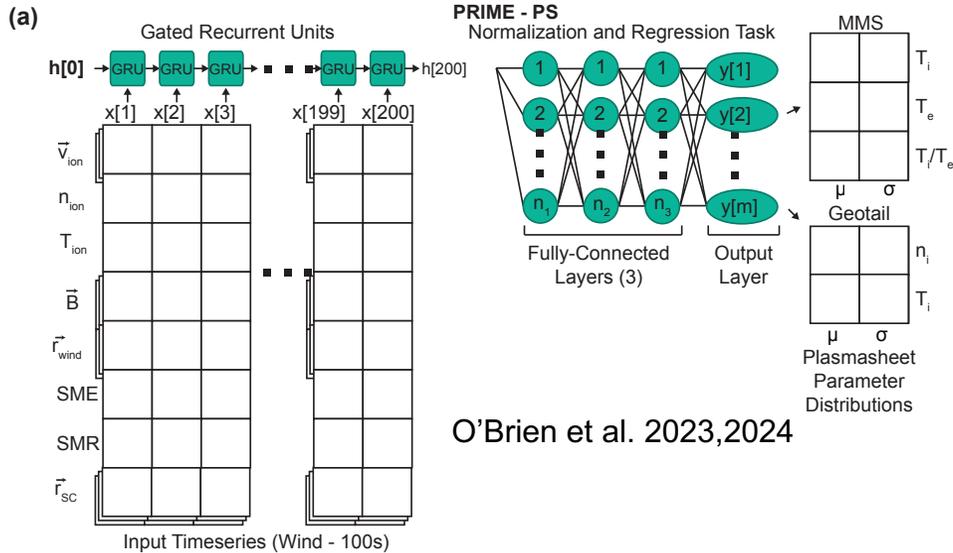


Output:
Anything locally measured
(In this example plasma moments)



Metrics & Results

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:

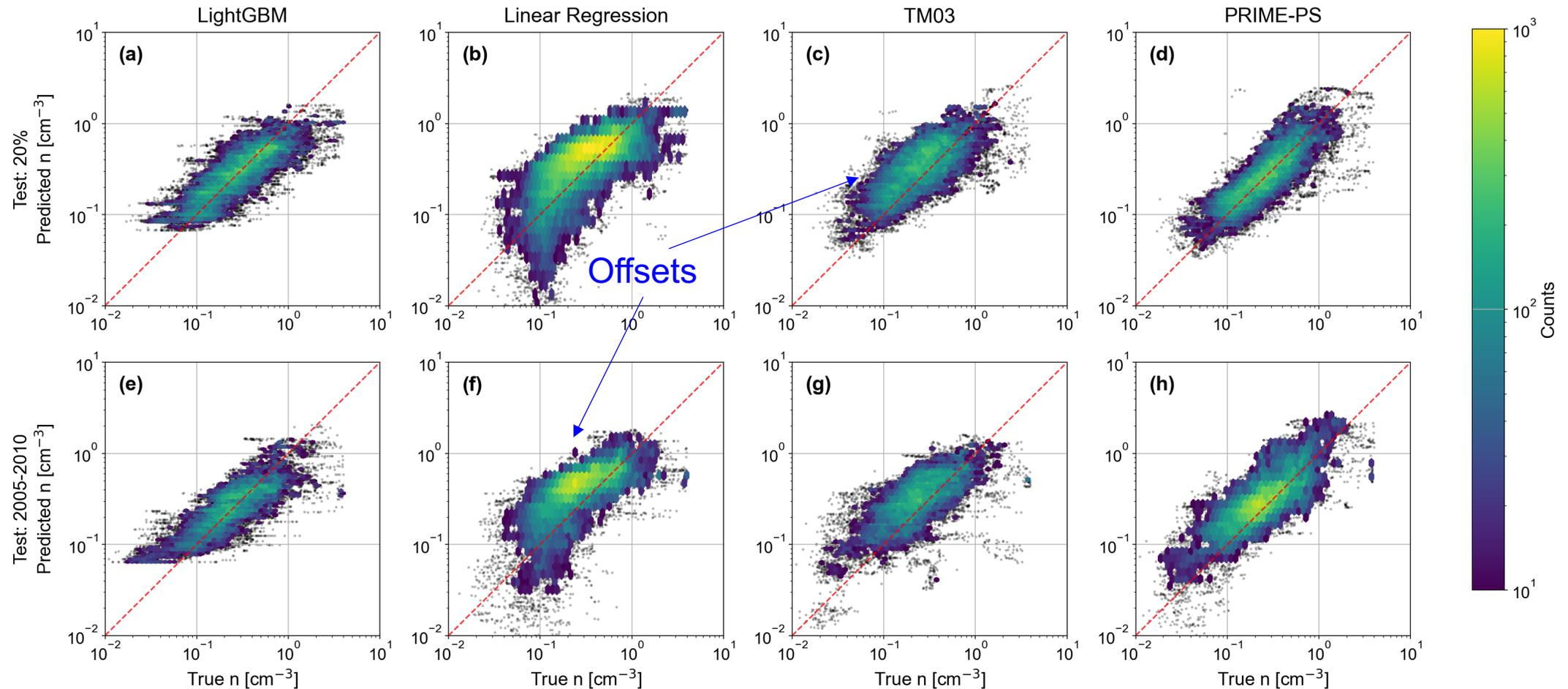
- ✓ Also tried different error functions, optimizers, hyperparameters etc.
- ✓ And different imbalanced learning techniques

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)

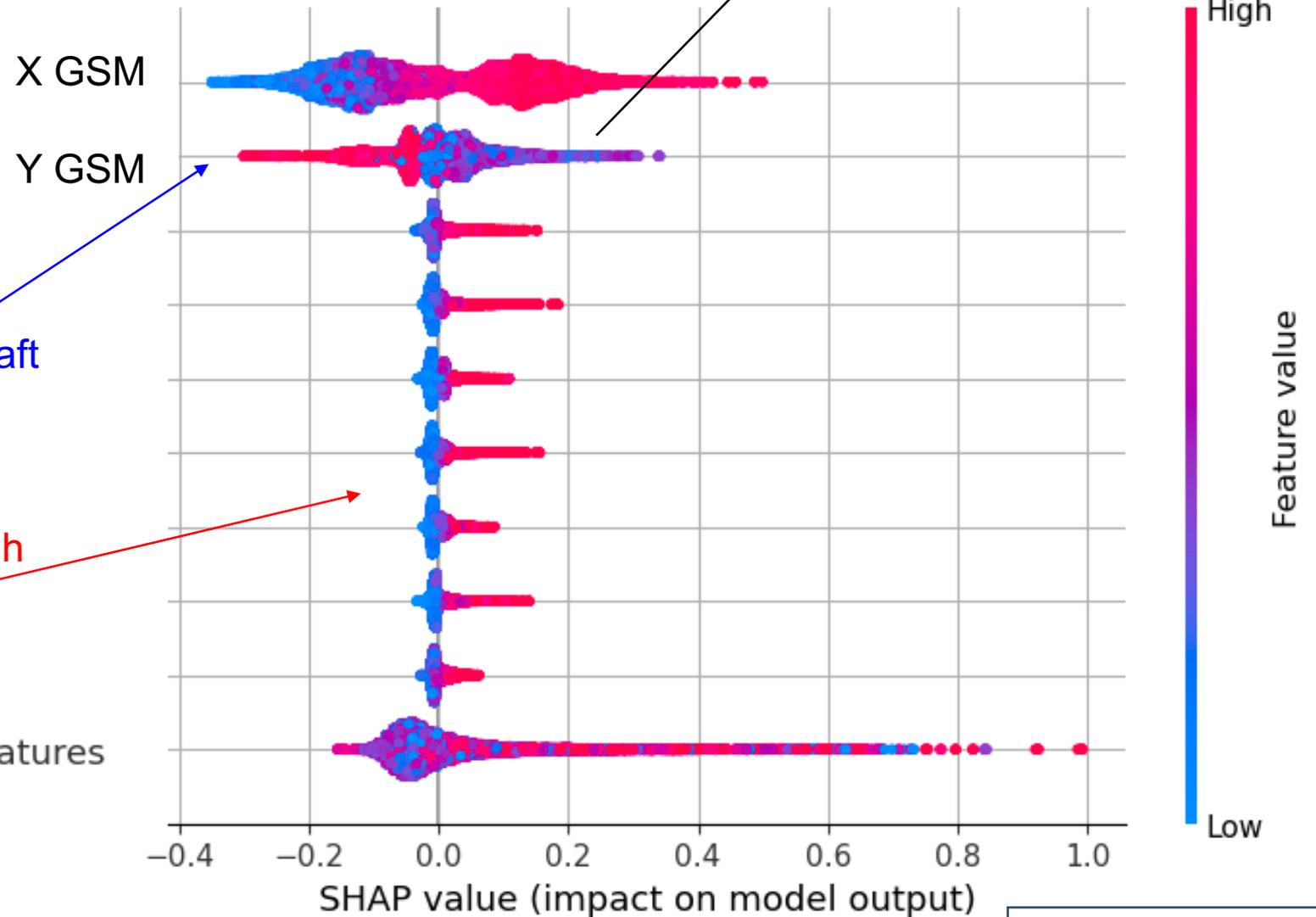
Method	MAE	R^2	r
LightGBM	0.129	0.373	0.631
Neural Net	0.152	0.325	0.603
Linear Reg	0.173	0.265	0.620
PRIME-PS	0.113	0.453	0.707
TM03	0.163	0.208	0.570

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.

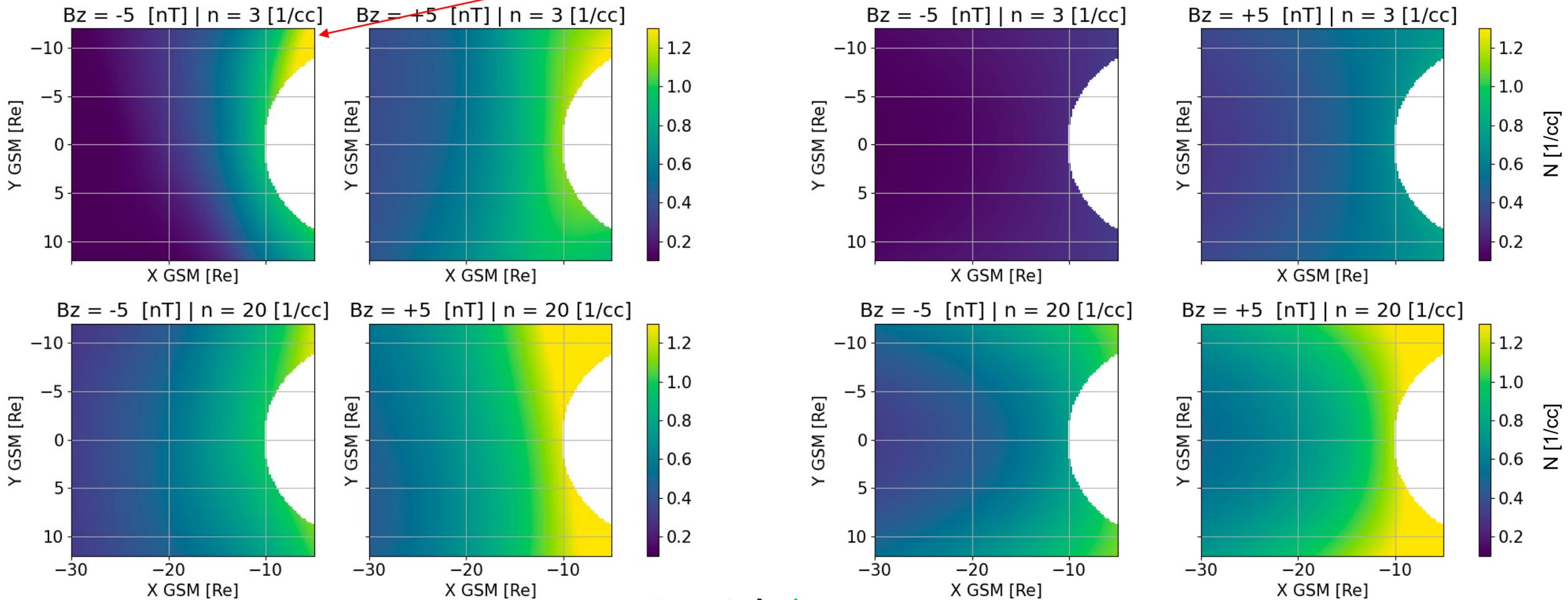


Seriously, we did exactly that.

Modeling Efforts

Modeling Density | 2D Maps (Synthetic input)

Assymetries Introduced



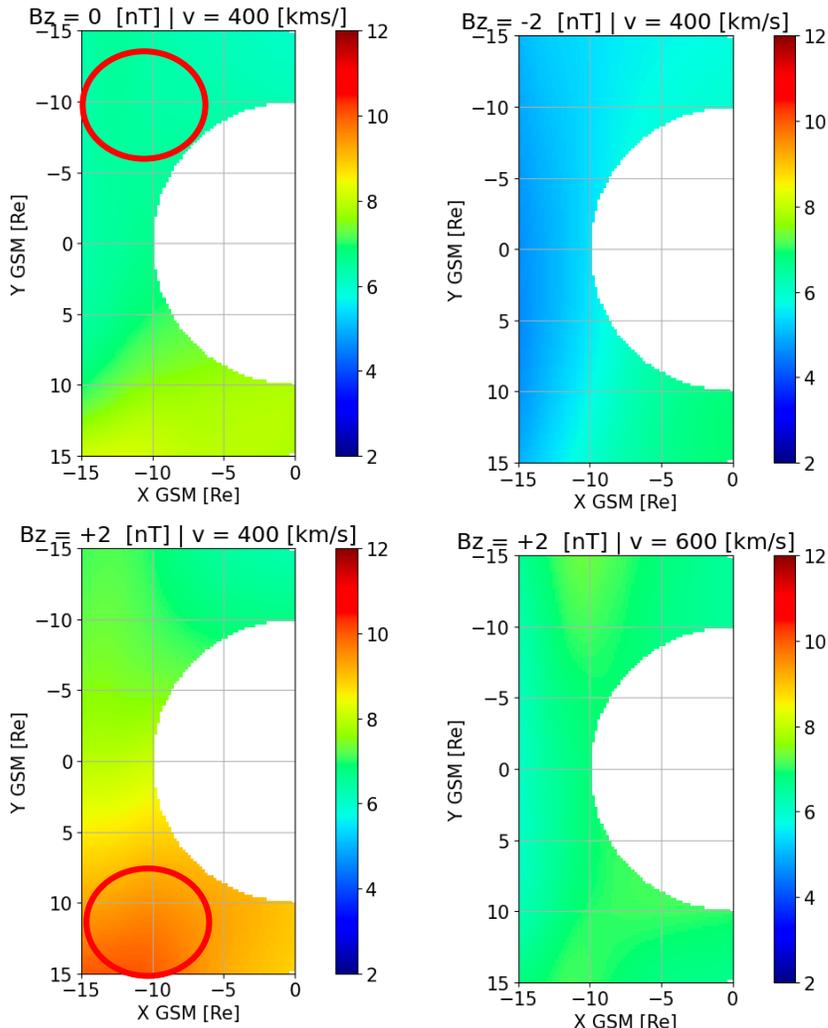
Neural Networks modeling

Empirical modeling (TM03)

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

Modeling Temperature Ratios with MMS | 2D Maps

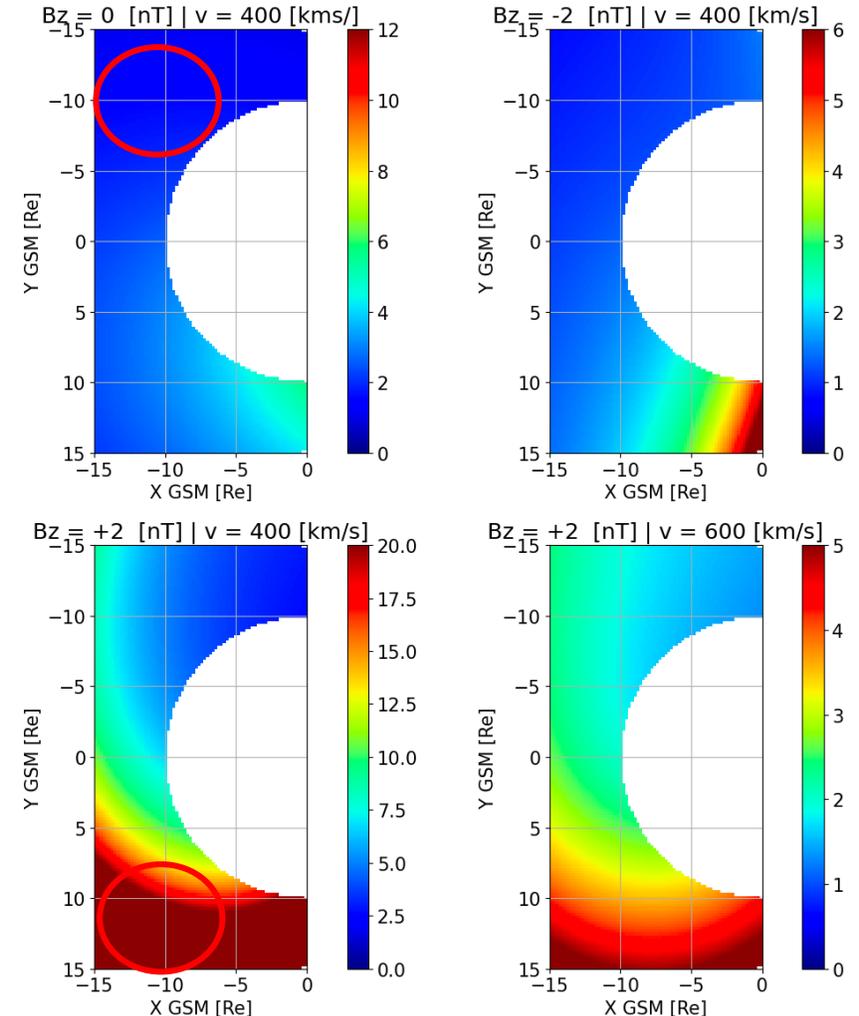
Note: Agreement Wang et al., 2010 with dusk Ti/Te higher than dawn (Using THEMIS)



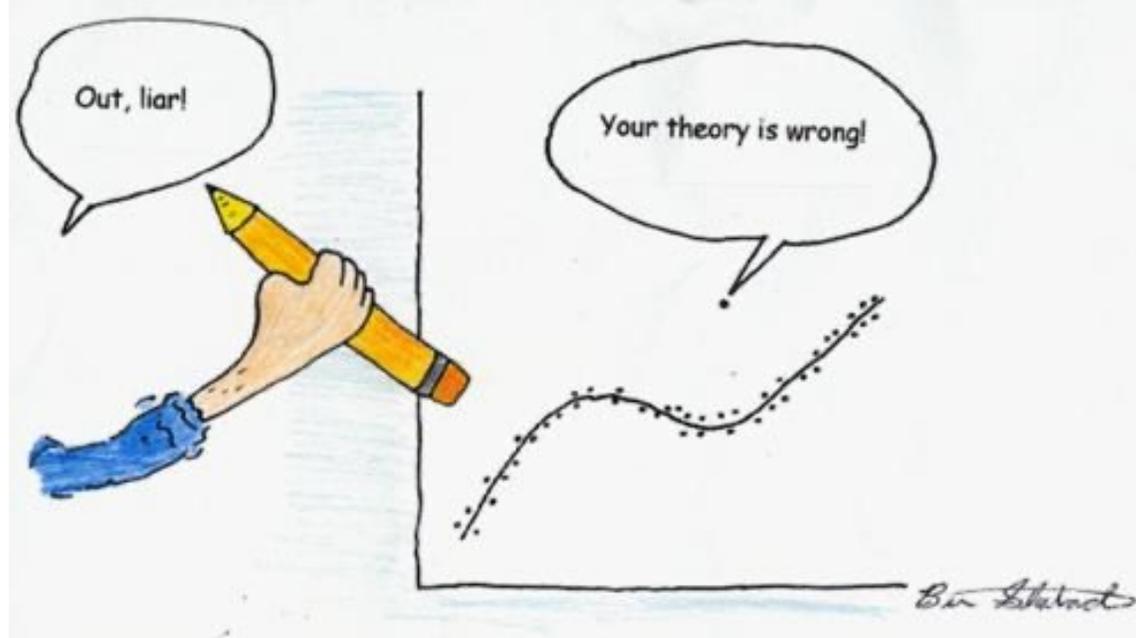
Neural Networks modeling

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

Cons:
 - No analytical form



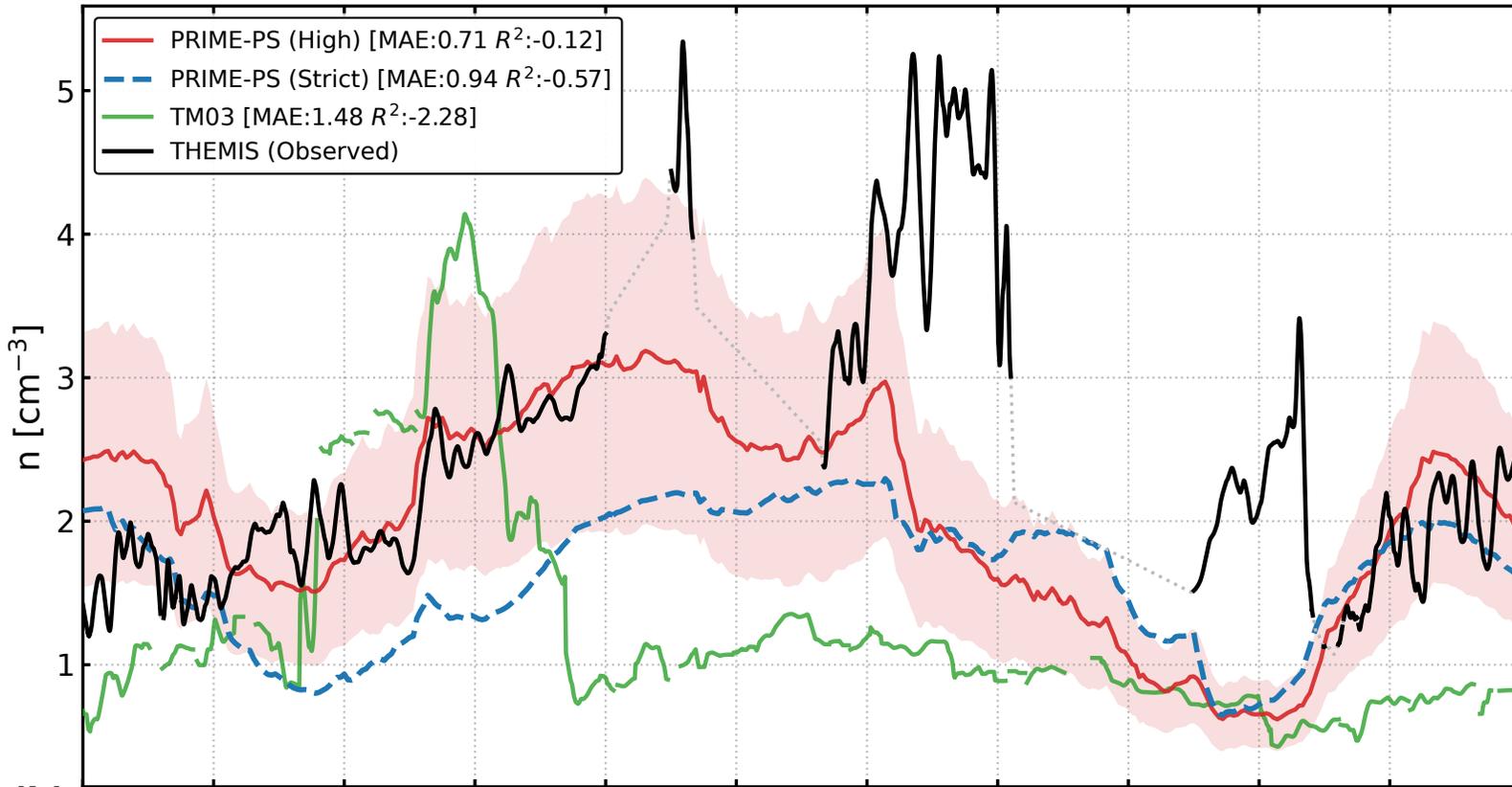
Empirical modeling (TM03/DSGR16)



Storm Time Behavior and Importance of Outliers

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

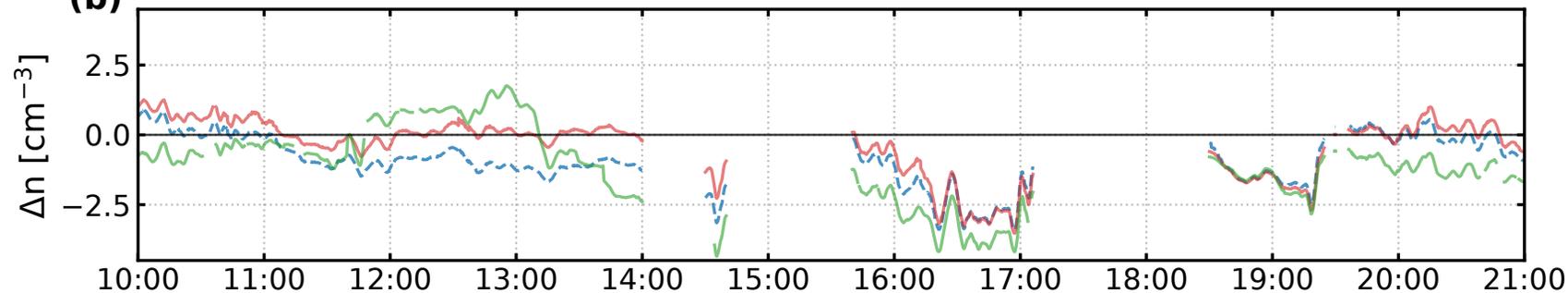
ML model (high dens): 0.7 [1/cc]

• ML model (normal): 0.94 [1/cc]

• TM03: 1.48 [1/cc]

Key Message: >50% improvement

(b)



05 November 2023

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

Summary & Discussion

Results

- ✅ **Significant Gains:** PRIME-PS and ML models outperform analytical methods & show asymmetries.
- ❌ **Mediocre Storm Predictability:** We mainly capture "boring" conditions, not the critical rare events.
- 🧠 **A Core Problem:** Our training data is biased. Extreme events, are not always captured by simple thresholds, are very rare, 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).

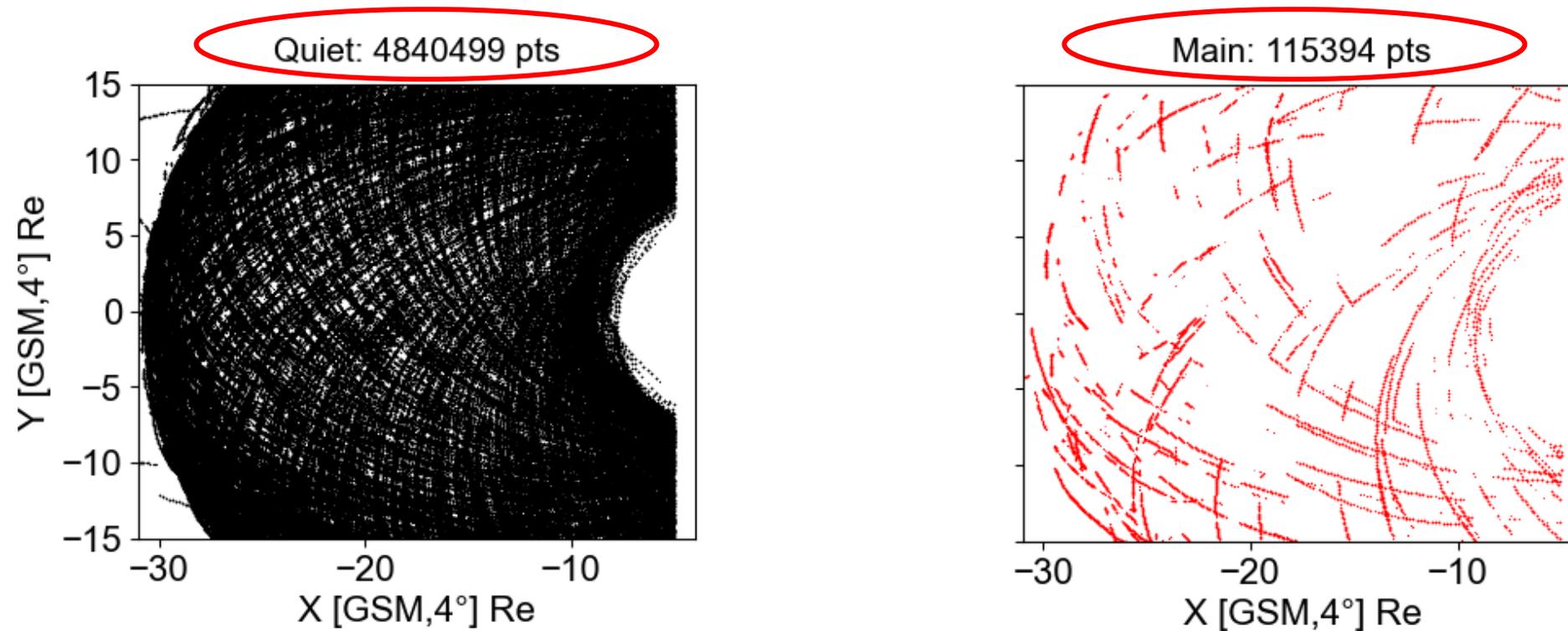
Discussion Points/Opinions

- 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.

Extras

Community Reminder: Data Sparsity During Extreme Events

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

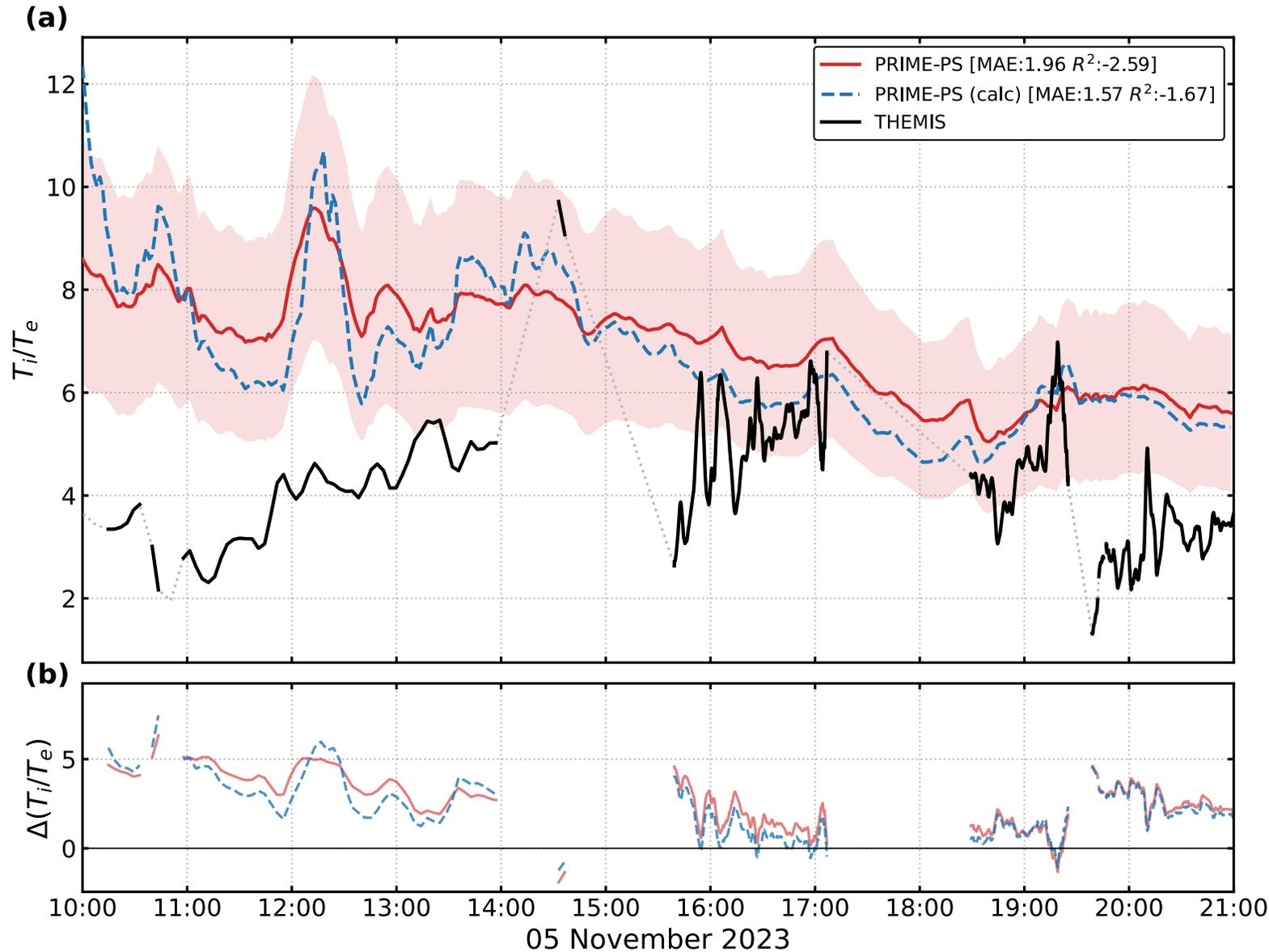


(#of unique days, # unique storms)

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

Test case of a storm (05 Nov 2023)



Process:

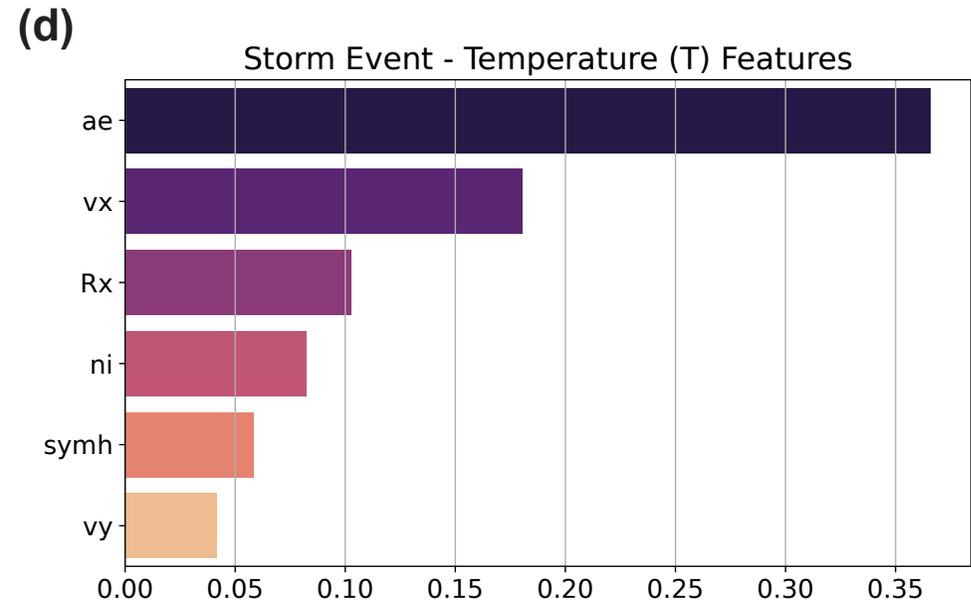
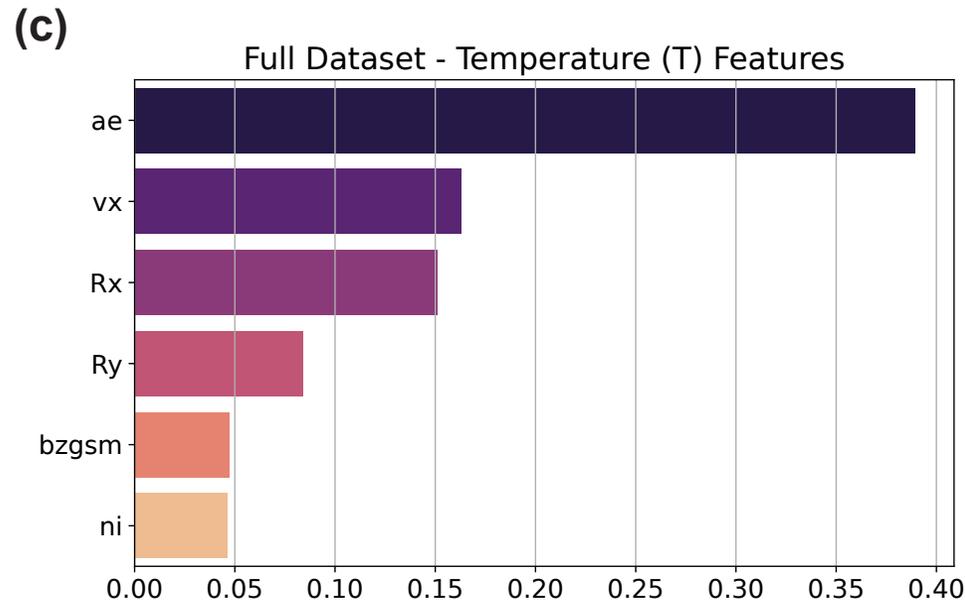
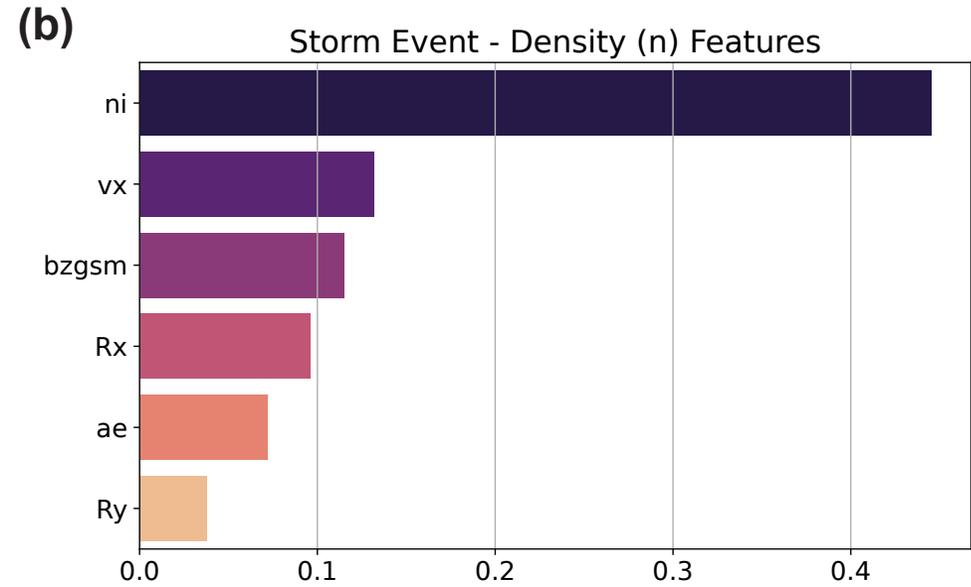
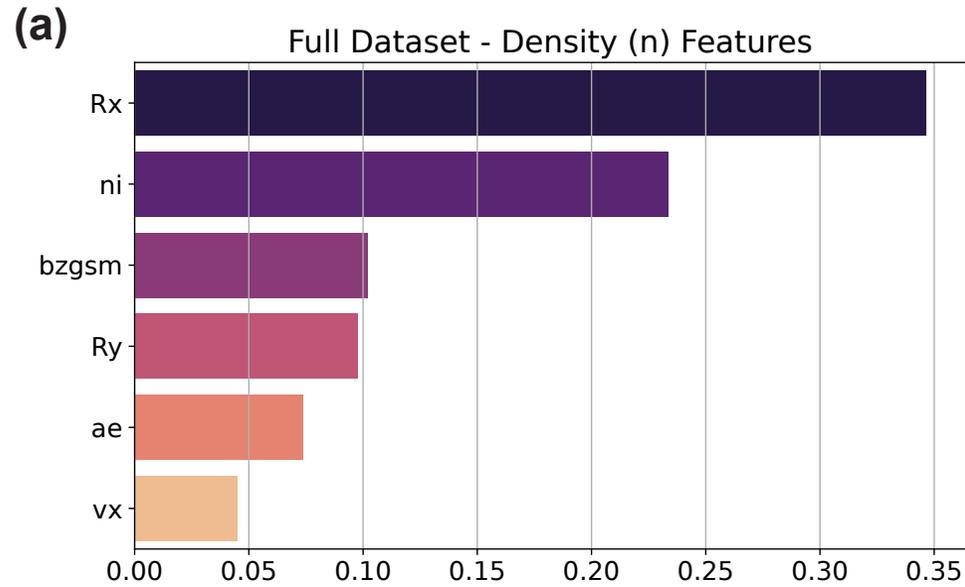
PRIME: Directly predicted

Calc: Calculated

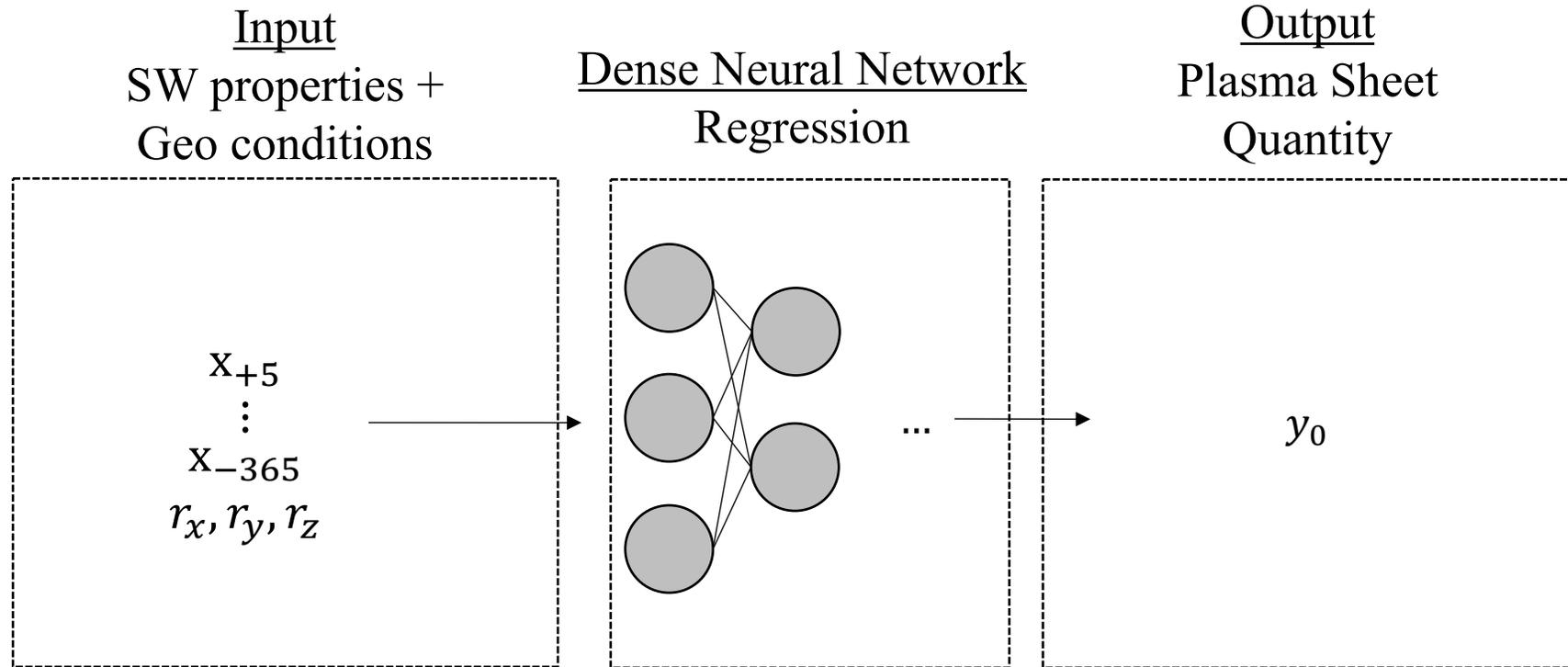
Test:

THEMIS observations 2 years after the Geotail data stopped operating

Feature Importance Analysis



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



Input:

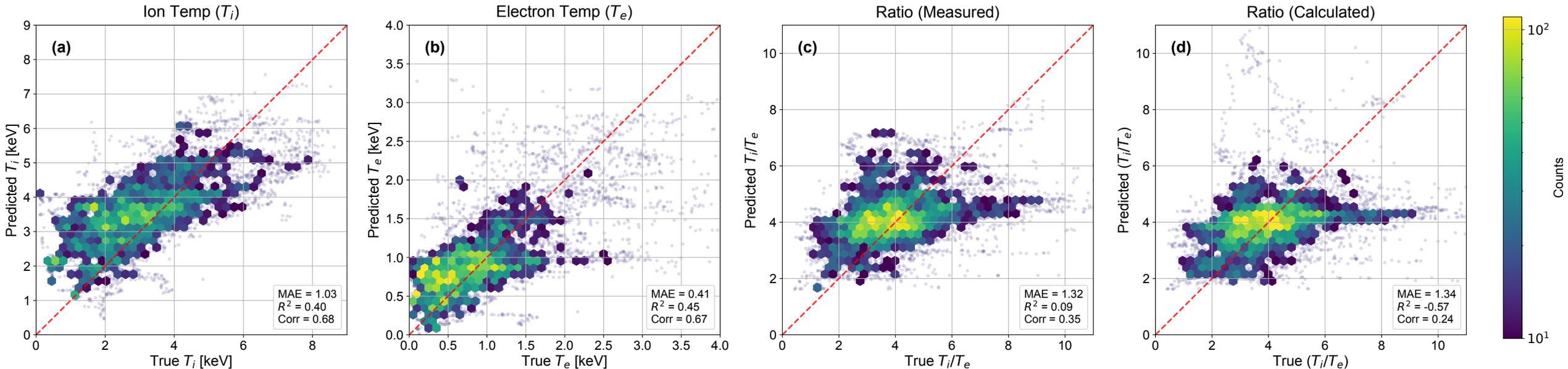
x: Different solar wind features (e.g., n, B, etc.) + geomagnetic indices including time history up to 6h
r: Location of SC measuring output

Output:

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

Temperature and Ti/Te 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

What are the metrics we need to use

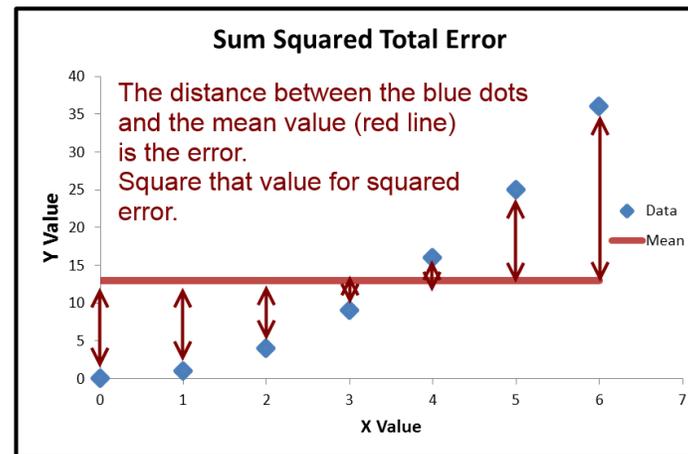
```
... 99/99 _____ 0s 808us/step
explained_variance: 0.019
median absolute error: 0.11
r2: -0.01
MAE: 0.157
MSE: 0.055
RMSE: 0.235
Cor: 0.533
```



```
... 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
```

A complex and intriguing model

np.mean()

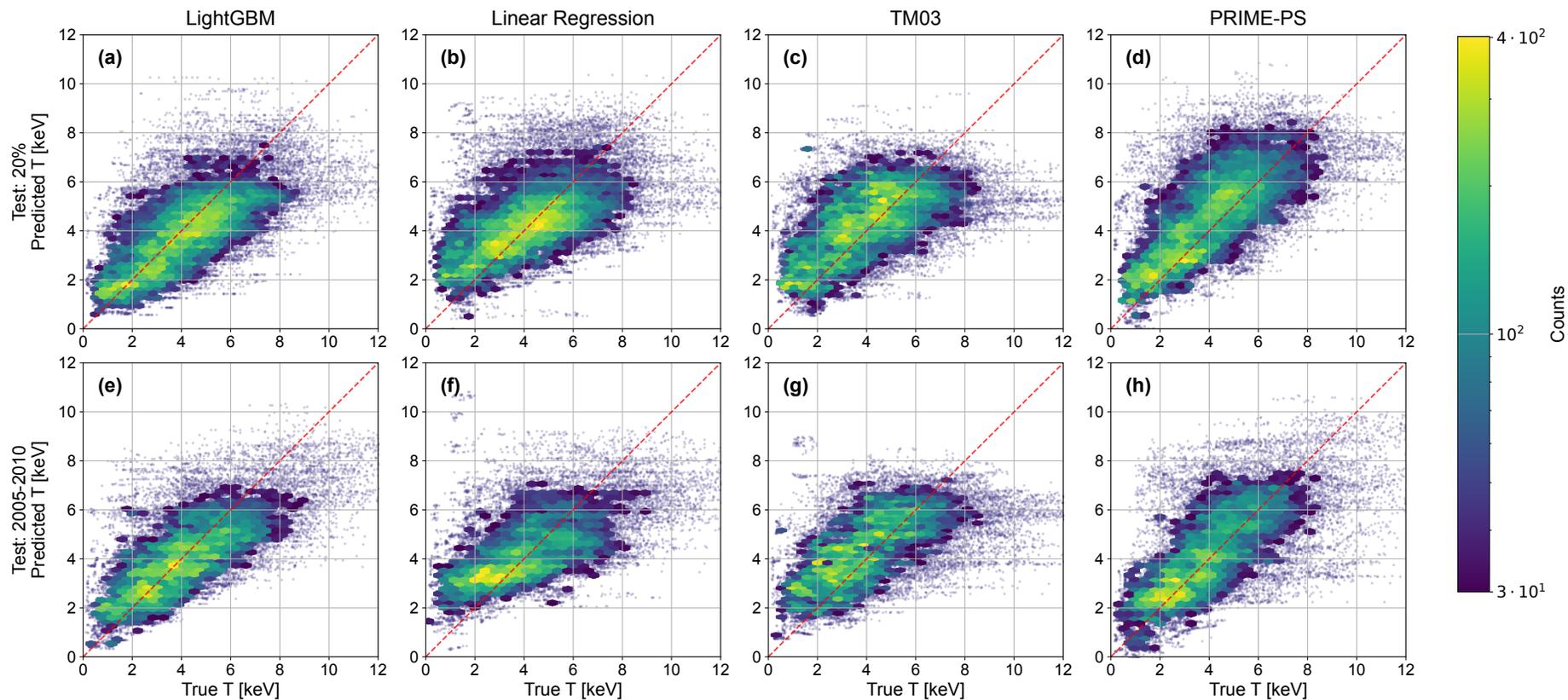


When $R^2 < 0$, the horizontal line explains the data better than your model (i.e., mean of observed).

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:

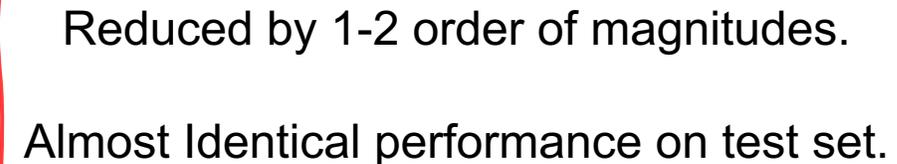
- **Find Outliers:** Detected using *Isolation Forest* (unusual feature patterns)

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

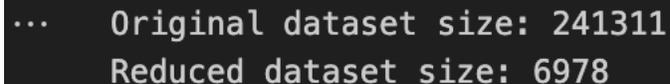
- Selects a *core subset* where points are maximally distant
- Ensures broad coverage and *high diversity* of the original data

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

- Adds points from *under-represented regions*



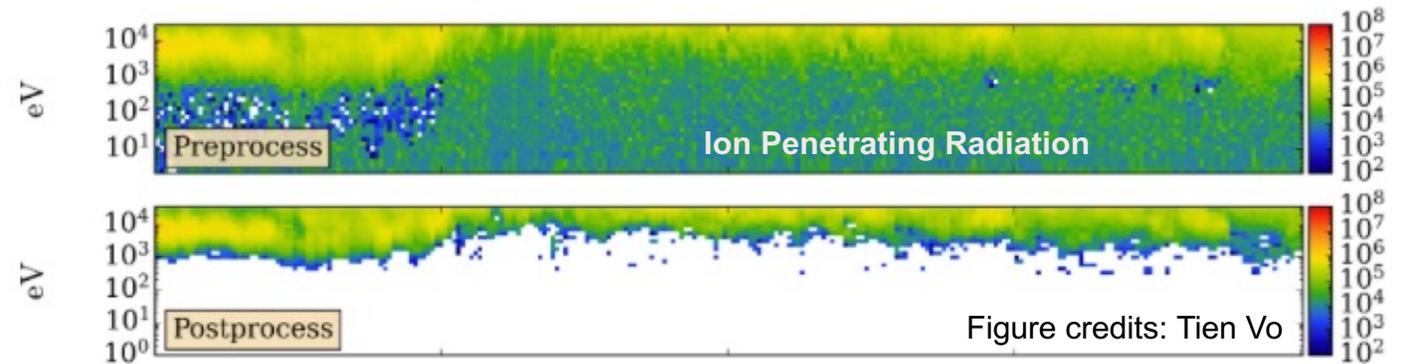
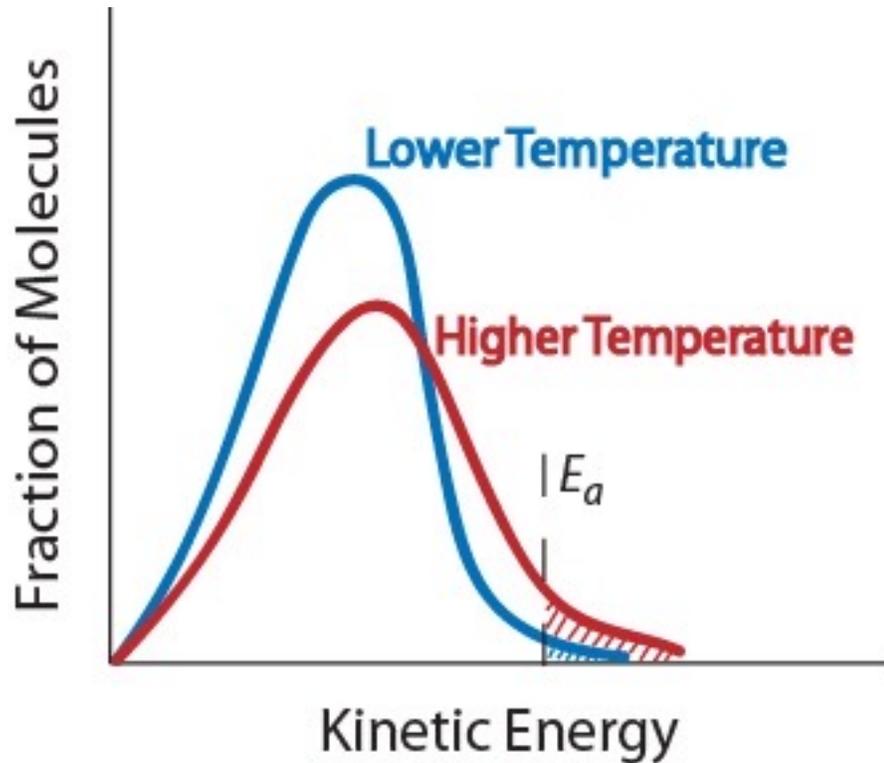
Reduced by 1-2 order of magnitudes.
Almost Identical performance on test set.



```
... 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



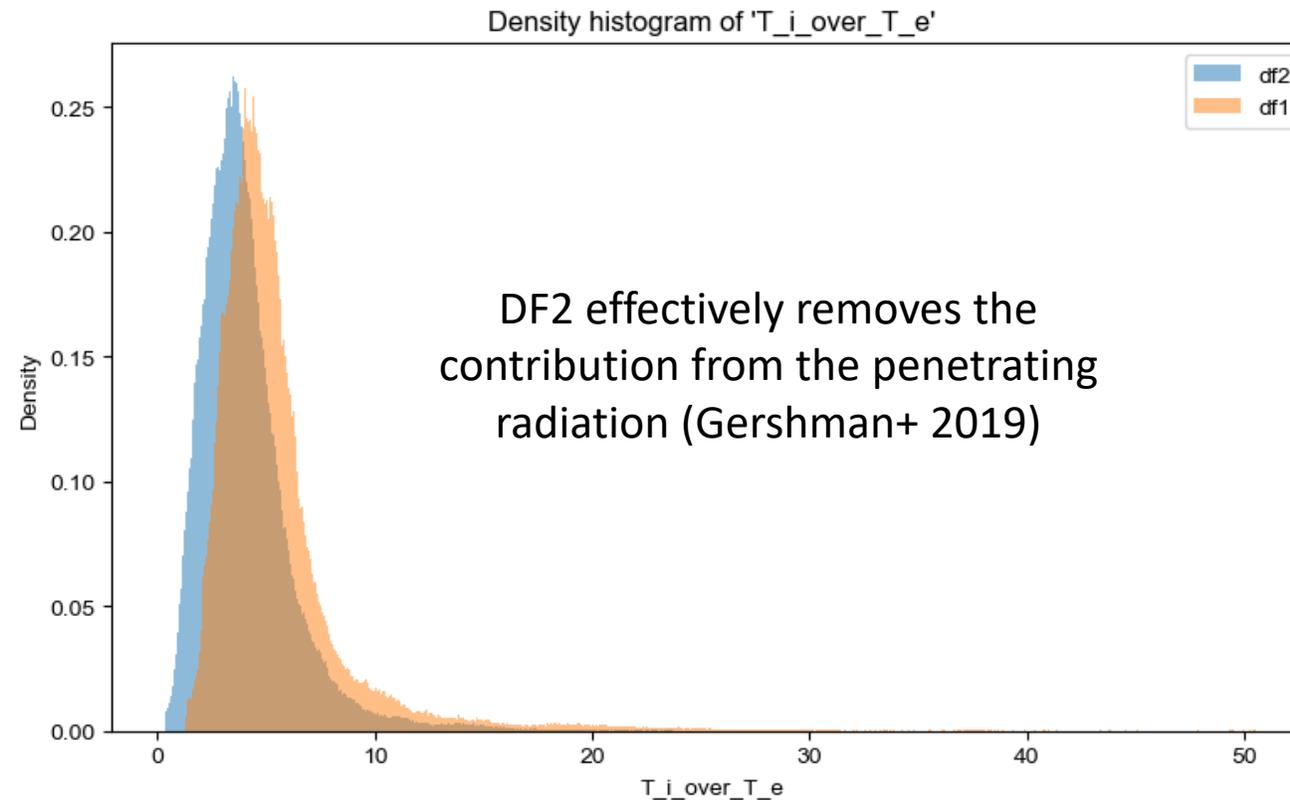
The temperature here was completely incorrect, and the velocity increased from about 200 km/s to over 1500 km/s.

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

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

mean df1 (full distribution moments): 5.5513

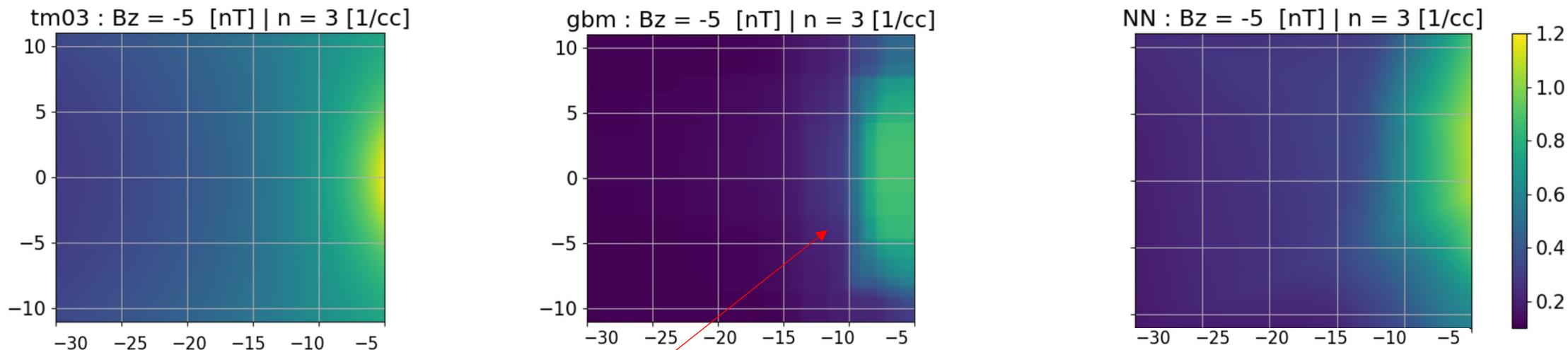
mean df2 (partial distribution moments): 4.0797



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

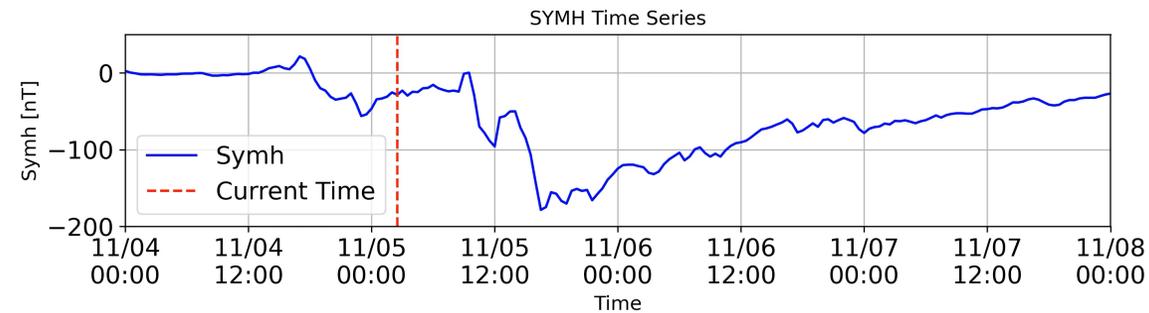
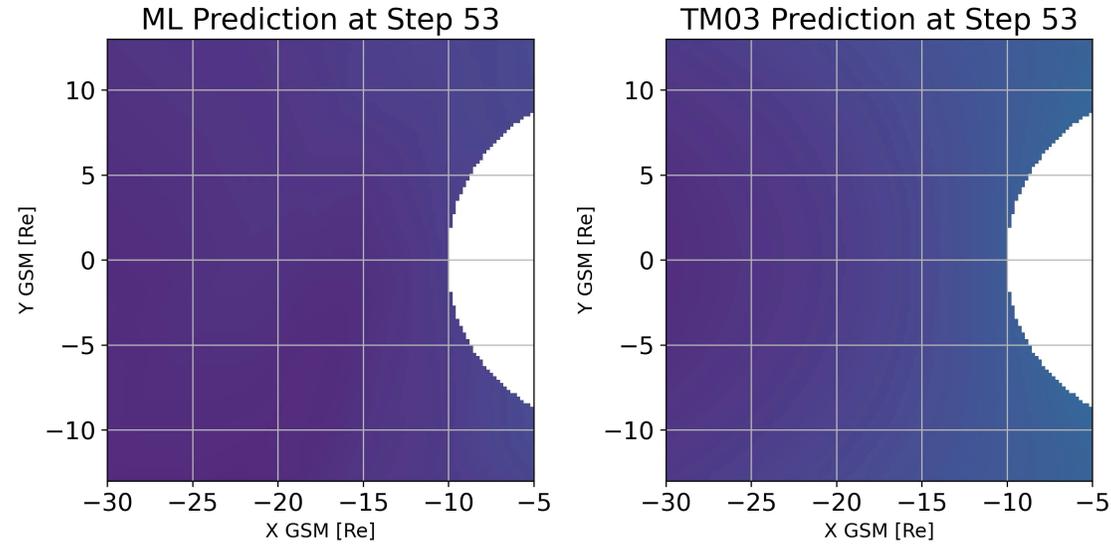
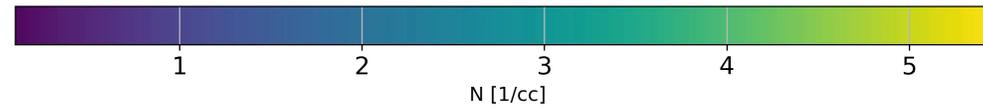
Community Reminder via Modeling Bz



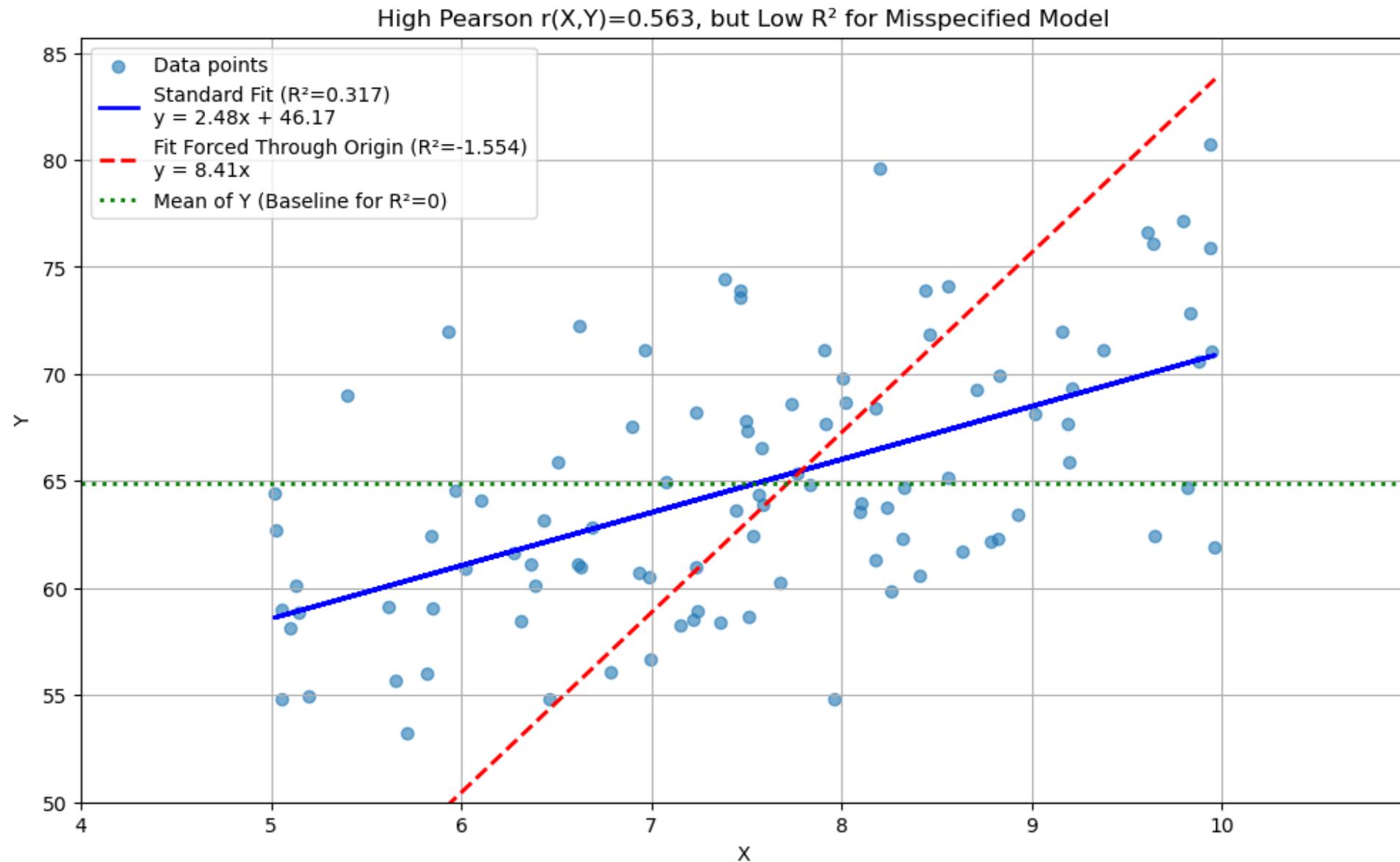
Gradient boosting is powerful, but it's a poor extrapolator.

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



Goal of Presentation

Spark a discussion on problems we have with data-driven modeling.

◆ Evaluating Models

What do our typical metrics really tell us about predictive power?

◆ Outliers & Sampling

How should we treat rare events? Can we expand or restrict our dataset meaningfully?

◆ Ground Truth in Space

How do we use in-situ data, and what counts as “truth” relative to model performance?

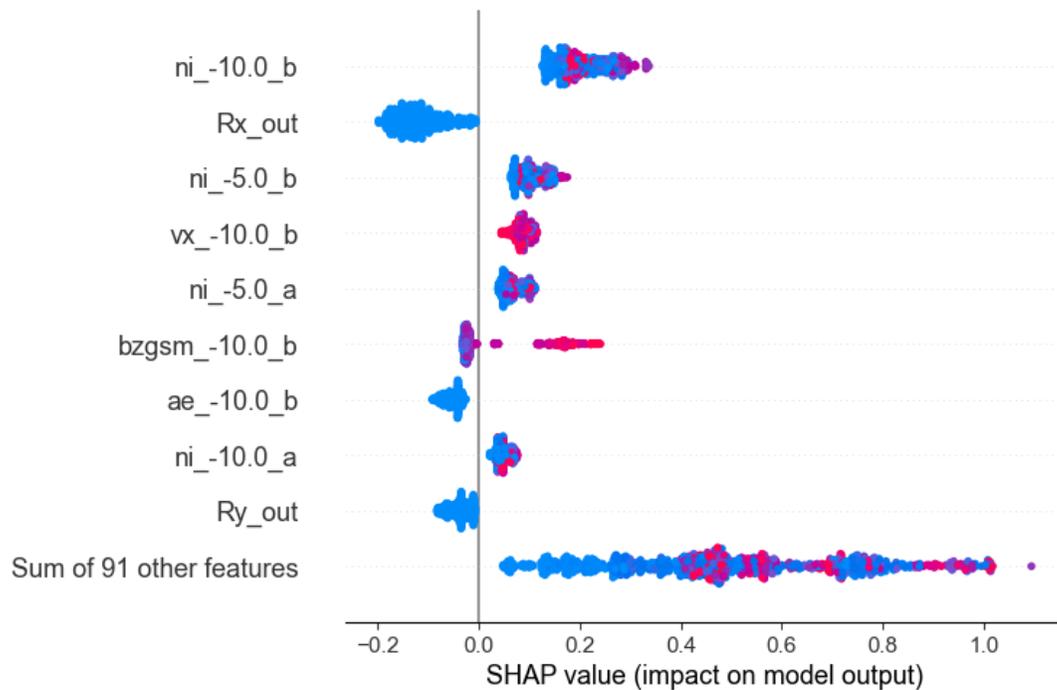
◆ Model Architectures

Do architectures always drive performance? or do data and physics matter more?

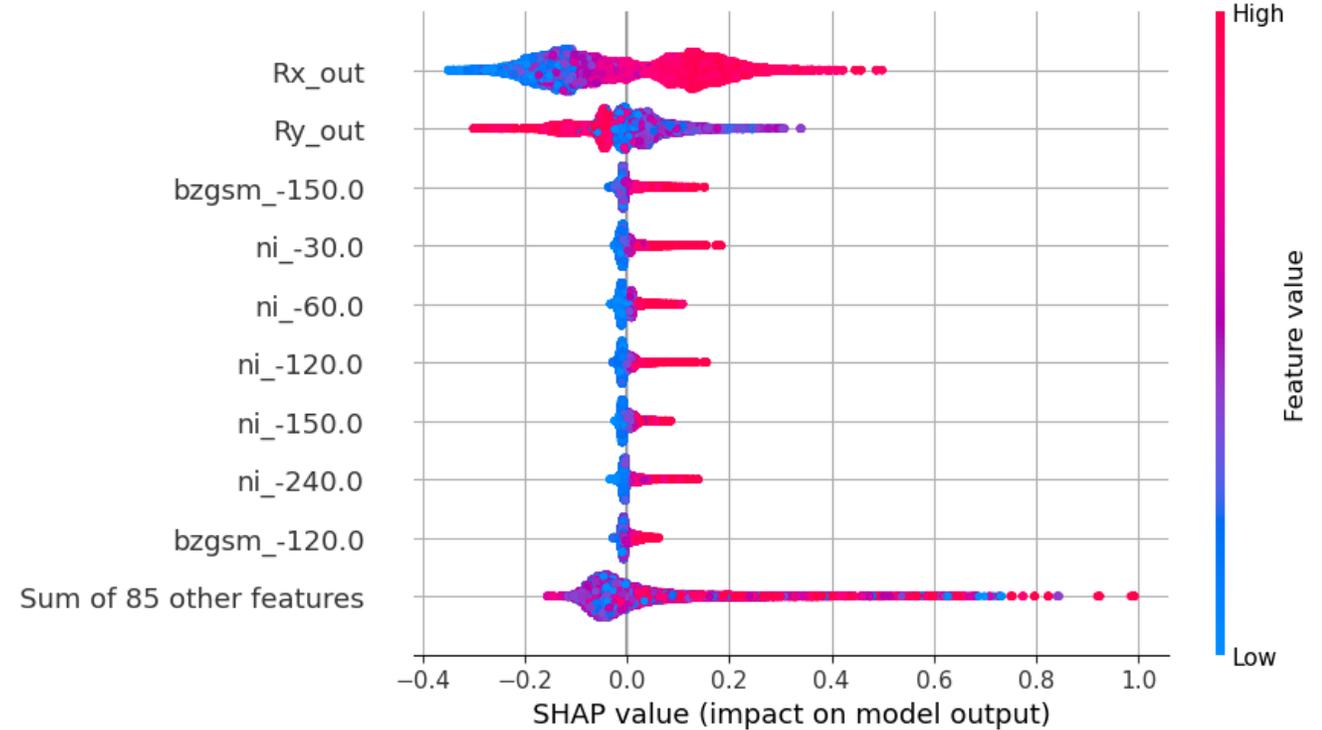
All these questions will be discussed through the challenge of predicting Earth’s plasma sheet properties from solar wind and geomagnetic conditions

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.

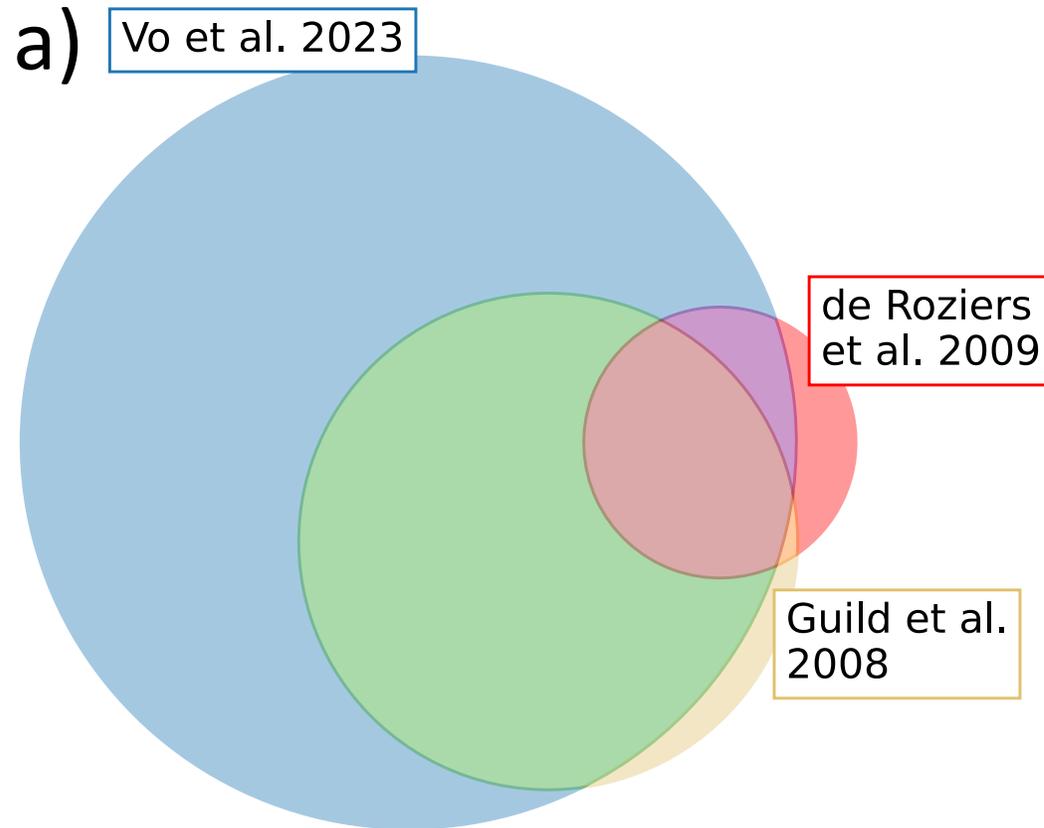


X_test_storm



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

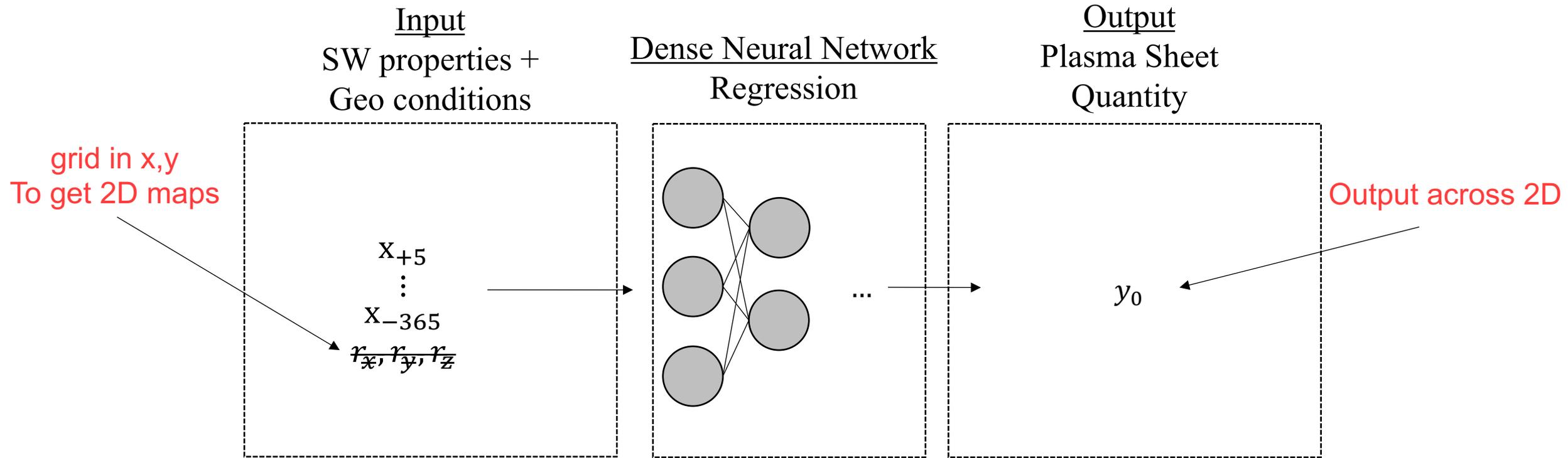
The Problem: We use **static thresholds** for dynamic environments.

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)

Next step: 2D modeling



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:

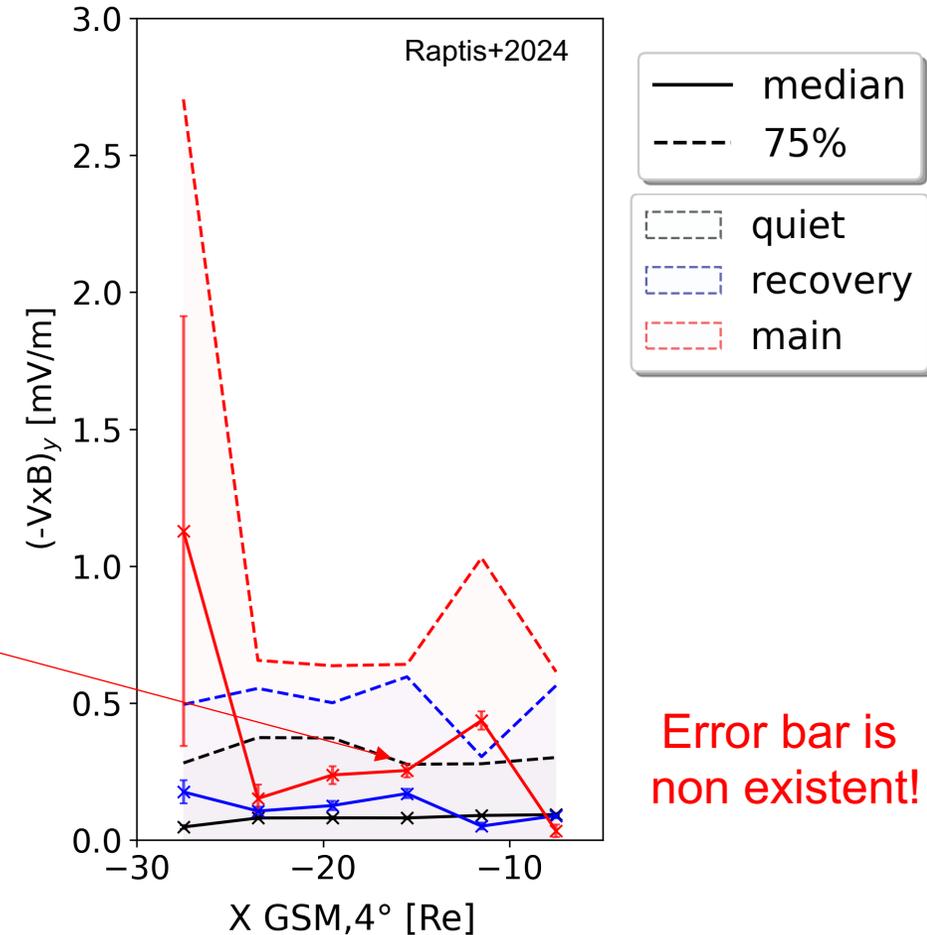
y: Different quantities at plasma sheet (e.g., n, B, T etc.)

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]

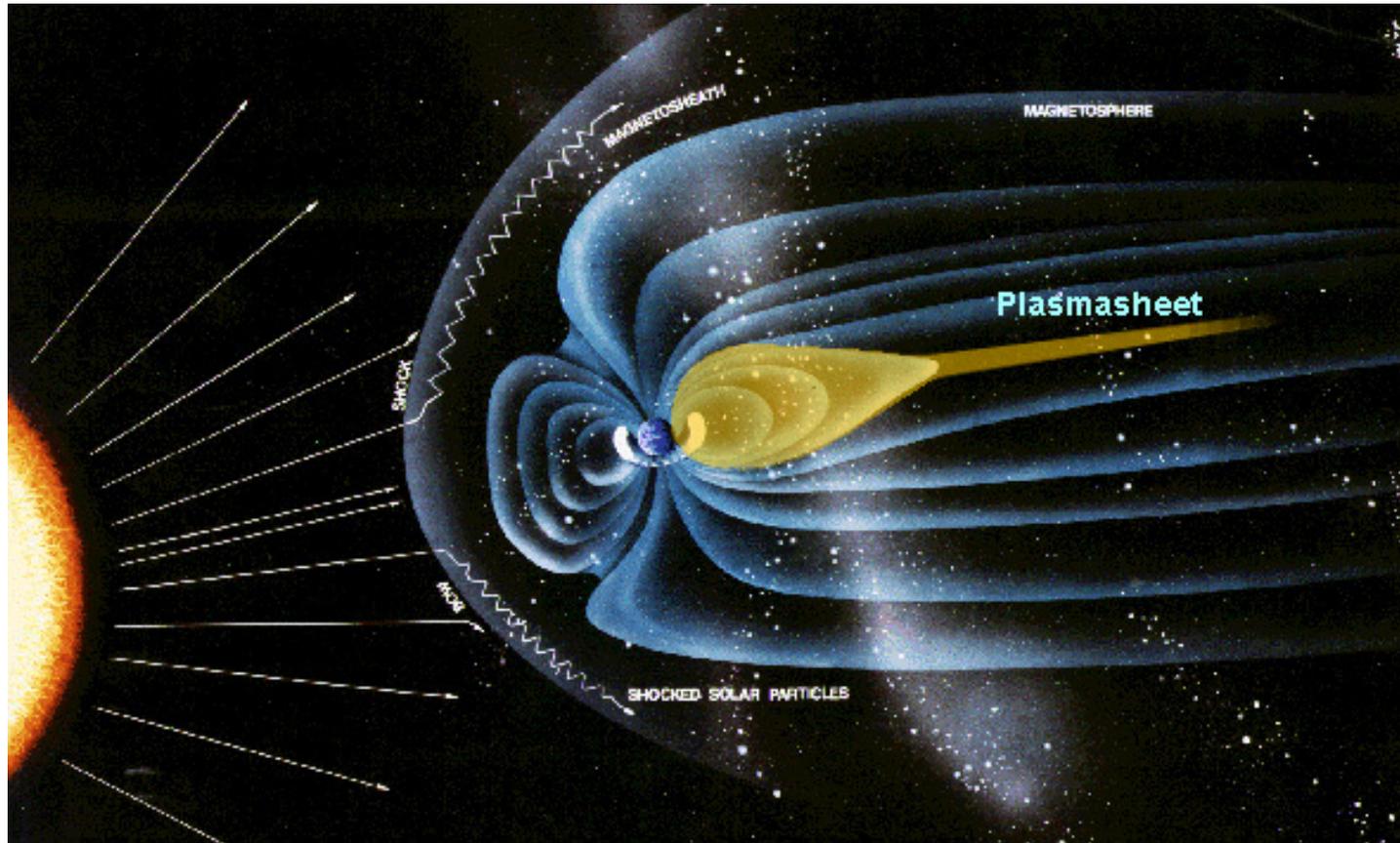
(D)



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

Earth's plasma sheet



Credits: NASA

- Magnetotail reconnection
- Bursty Bulk Flows (BBFs)
- Global Convection Patterns
- Ring current

Modeling PS is useful for:

- (a) Understanding storm/substorm dynamics
- (b) Explain ring current configuration
- (c) Facilitate space weather modeling
- (d) Understand inner magnetosphere
- (e) Source for radiation belts