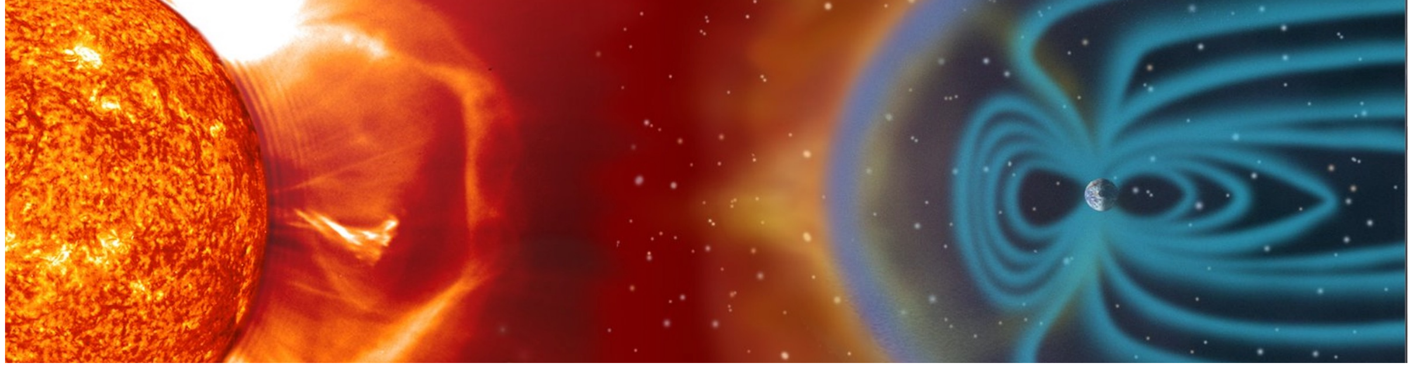


Plasma-Sheet Bubble Identification **Using Multivariate Time Series Classification**

Xuedong Feng, Jian Yang

12131165@mail.sustech.edu.cn & xuedongfeng1@gmail.com

[1]Department of Earth and Space Sciences, Southern University of Science and Technology, Shenzhen, China

2023.04.25

Outline

① Background

② Bubble Criteria and Dataset

③ Multivariate Time Series Classification (MTSC) Models

④ Results

⑤ Summary

1. Background

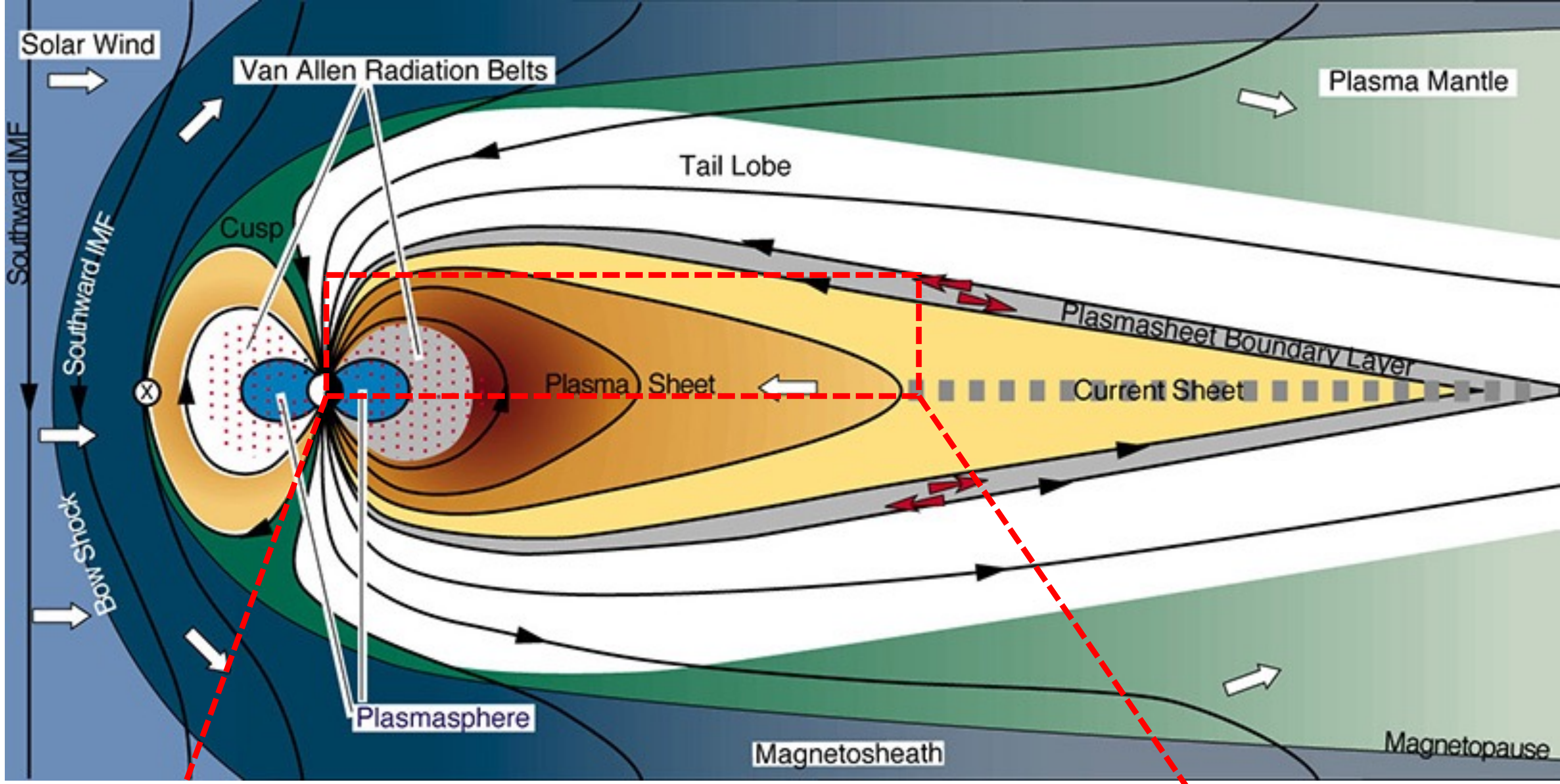
Bubble?

What is a **plasma-sheet Bubble?**

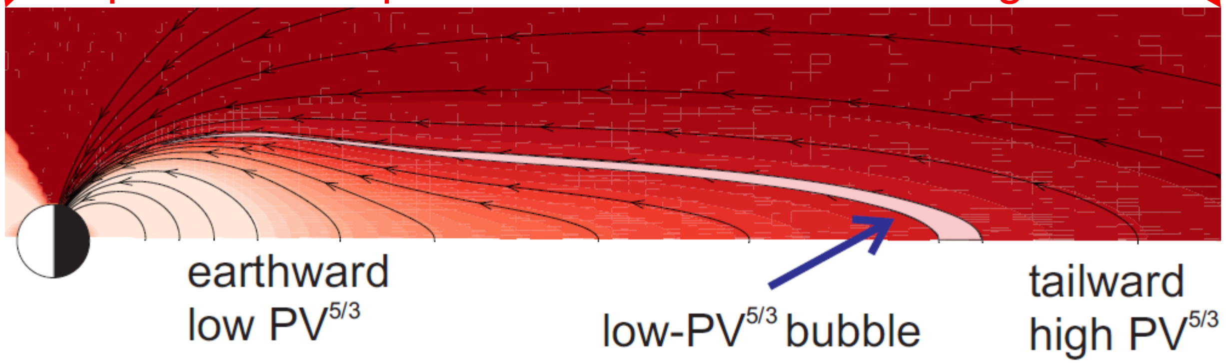
(1) A **plasma-sheet bubble** is a flux tube in the **nightside** magnetotail with its entropy **lower** than its **neighbors** [Pontius and Wolf, 1990].

(2) Bubbles are also referred to as bursty bulk flows (**BBFs**) in **observation** [Angelopoulos et al., 1992; Wolf et al., 2009].

(3) Bubbles/BBFs are the **primary carrier** for **substorm-time particle injection** from the plasma sheet to the inner magnetosphere [Yang et al., 2011].



plasma transport in the closed-field-line region



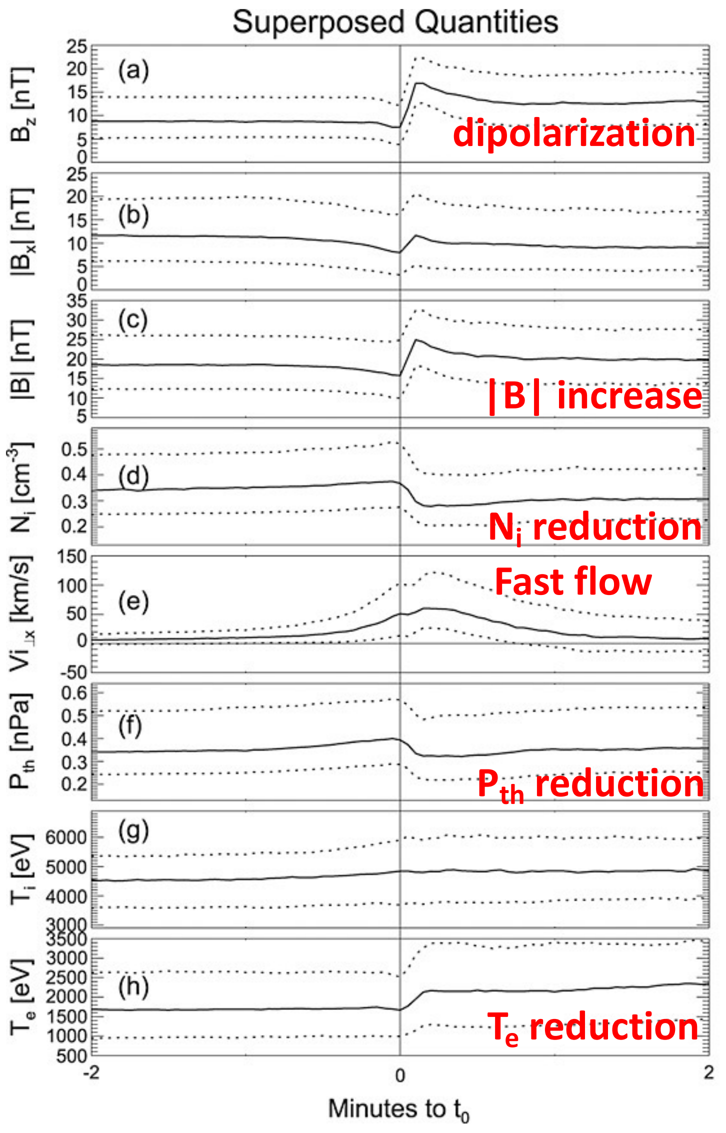
1. Background

Bubble's observation:

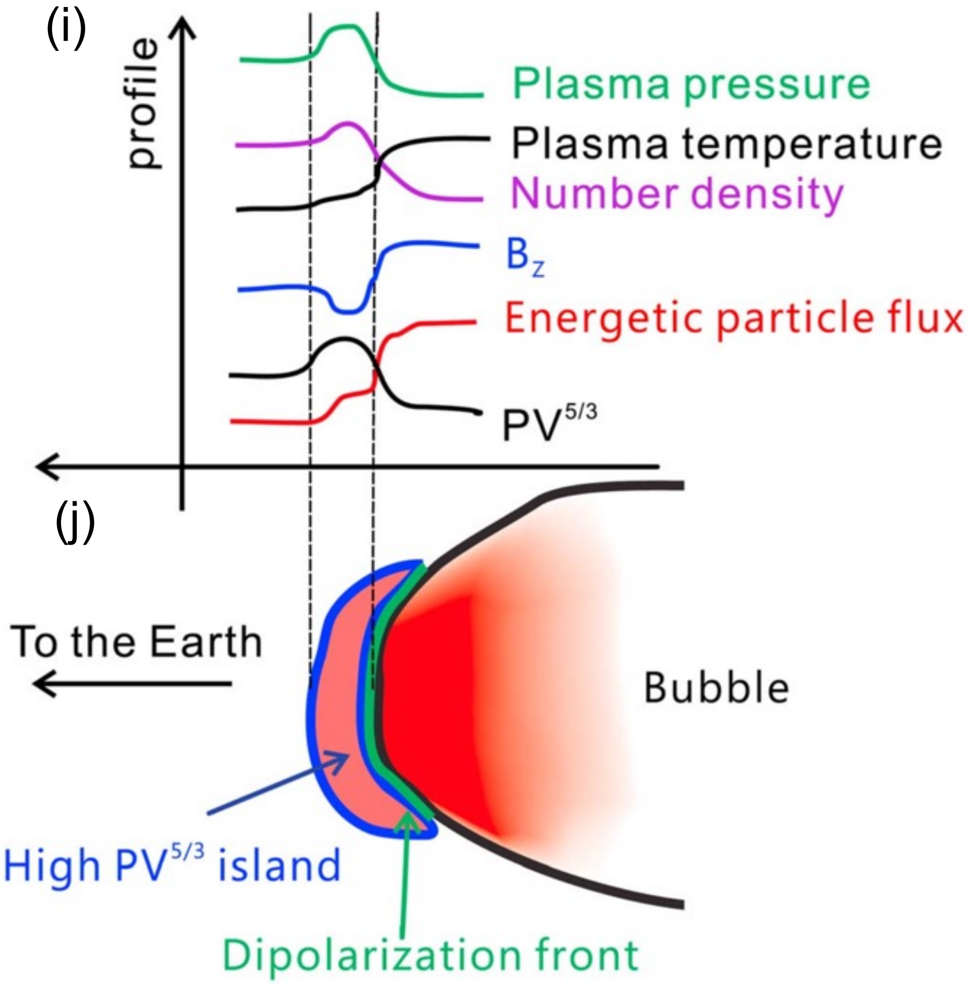
The most remarkable features inside a bubble normally include：
(a) magnetic field dipolarization
(c) magnetic strength increase
(d) density reduction
(e) fast earthward flow
(f) plasma pressure reduction
(h) electron temperature increase

Bubble's simulation:

Using a fluid model to simulate a bubble injection with significantly reduced entropy, showed a similar pattern (i and j).



[Liu et al., 2013]



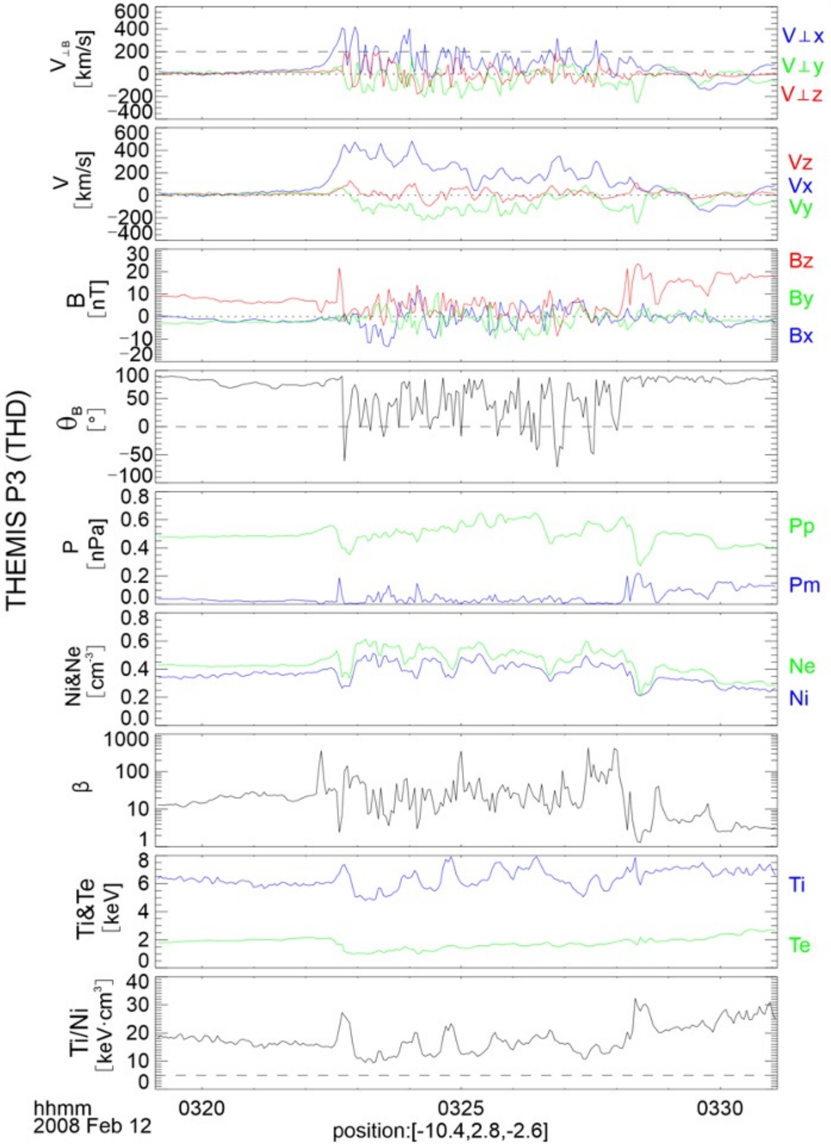
[Yang et al., 2011]

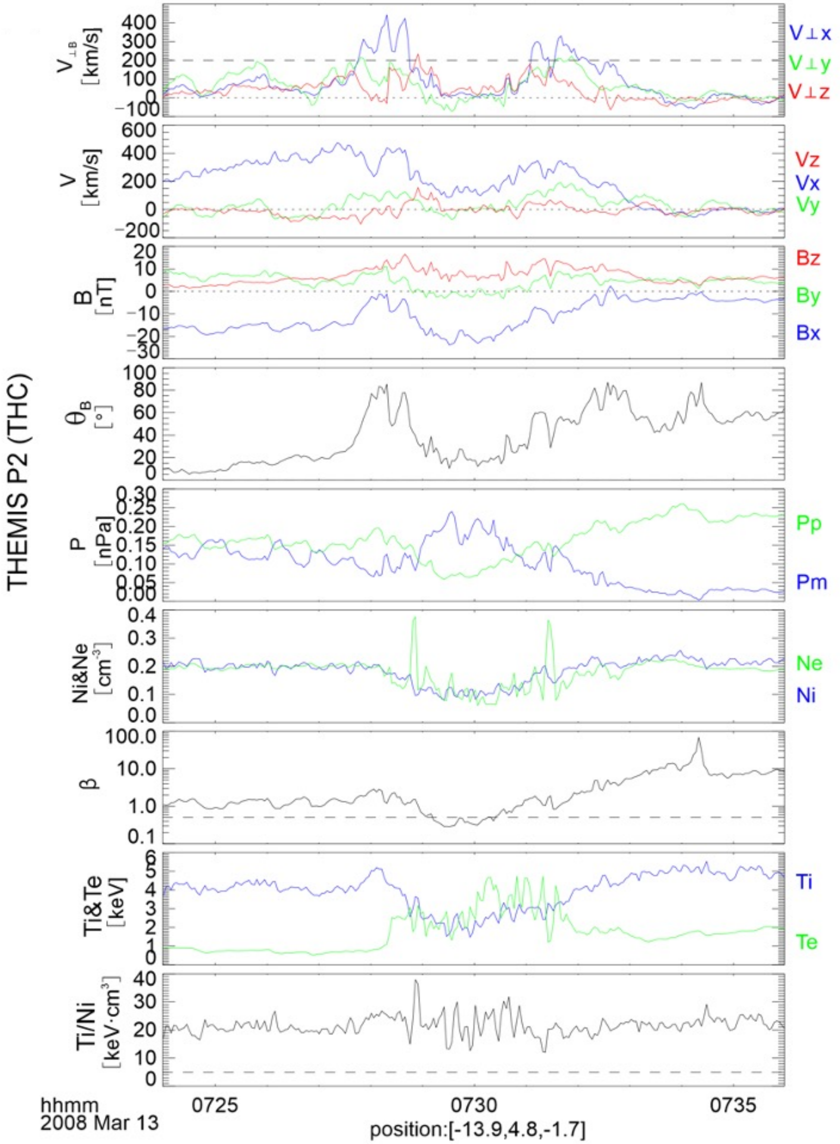
1. Background

SUSTech Southern University of Science and Technology

Why using machine learning?

a) the rapid **development** of machine learning **techniques**
b) Abundant data
c) Alleviate the manual inspection **workload** on scientists
d) Traditional recognition can easily **confuse** plasma-sheet bubbles with other categories of events
e) Provide a list of bubbles obtained by machine learning methods



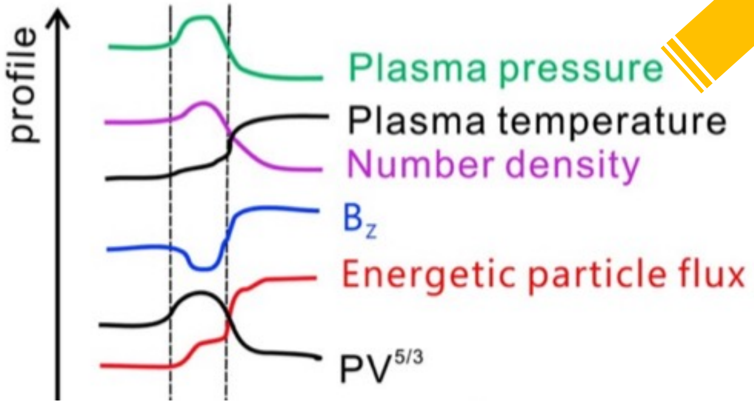


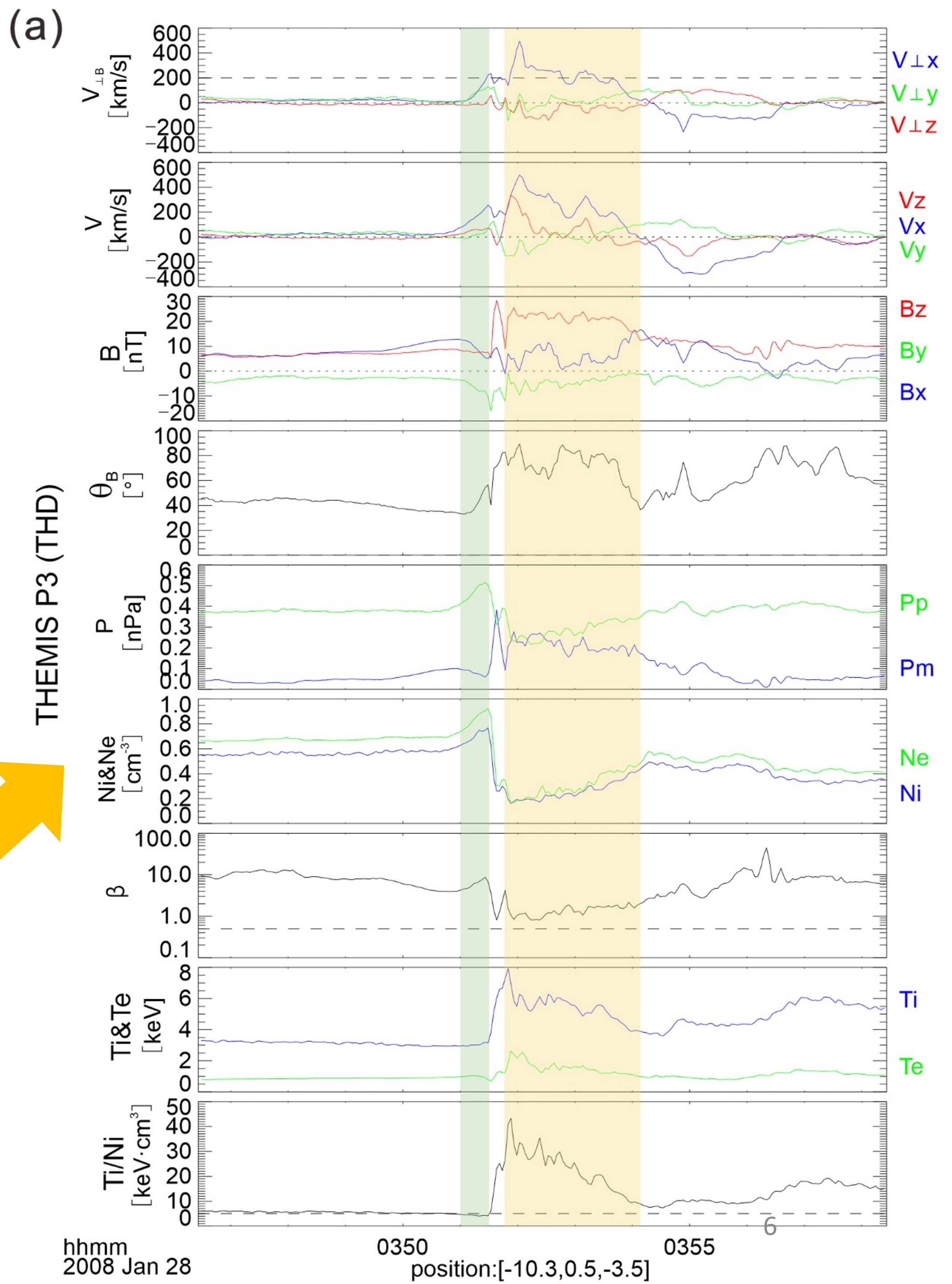
2. Bubble Criteria and Dataset

- **Data**: THEMIS measurements in the magnetosphere from **FGM**, **ESA**, and **SST** instruments.
- **Time Duration**: from year 2007 to 2021;
- **Resolution: 3 seconds**;
- **12 minutes** data is taken as one sample;
- **18 variables**: B_x,B_y, B_z, $\boldsymbol{\theta_B}(arctan\frac{B_z}{|B_x|})$, N_i, N_e, $\boldsymbol{plasma\ \beta}$, $\boldsymbol{P_m}$, $\boldsymbol{P_p}$, T_i , T_e , V_x, V_y, V_z, $(\boldsymbol{V}\perp\boldsymbol{B})_x$, $(\boldsymbol{V}\perp\boldsymbol{B})_y$, $(\boldsymbol{V}\perp\boldsymbol{B})_z$, $T_i/\ T_i$

Traditional criteria:

$-20R_E \leq X \leq -6R_E$
$\lvert Y\rvert \leq 10R_E$
$\lvert V_{x\_}pred\_B\rvert \geq 200km/s$
$plasma\ \beta\ \geq 0.5$
$B_z > 0\ nT$





2. Bubble Criteria and Dataset

Dataset: Positive negative samples ratio is **1:40** (1:1, 1:3, 1:10,1:20 had been tested)

- **Positive samples**: **2668** bubbles (identify bubbles between 2007 and 2020 using ***traditional criteria and manual inspection***)
- **Negative samples**: **106,720** non-bubbles(consists of non-bubbles that are manually excluded and non-bubble that are randomly selected at other times).

Train-validation-test dataset split: **6:2:2**

Normalization: maximum-minimum normalization

Prediction dataset: 82,152 12-minute intervals data from 2021

3. MTSC Models

3.1 MINIROCKET

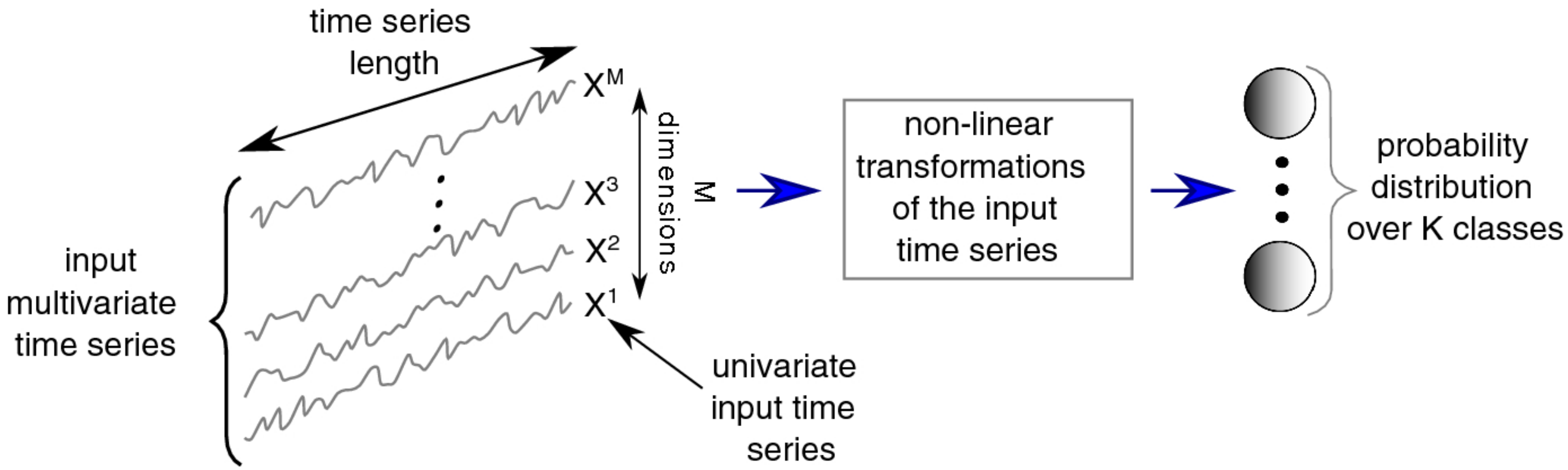
• MINIROCKET achieves state-of-the-art accuracy for time series classification by transforming input time series using **random convolutional kernels**, and using the **transformed features** to train a **linear classifier** [Angus Dempster, 2021].

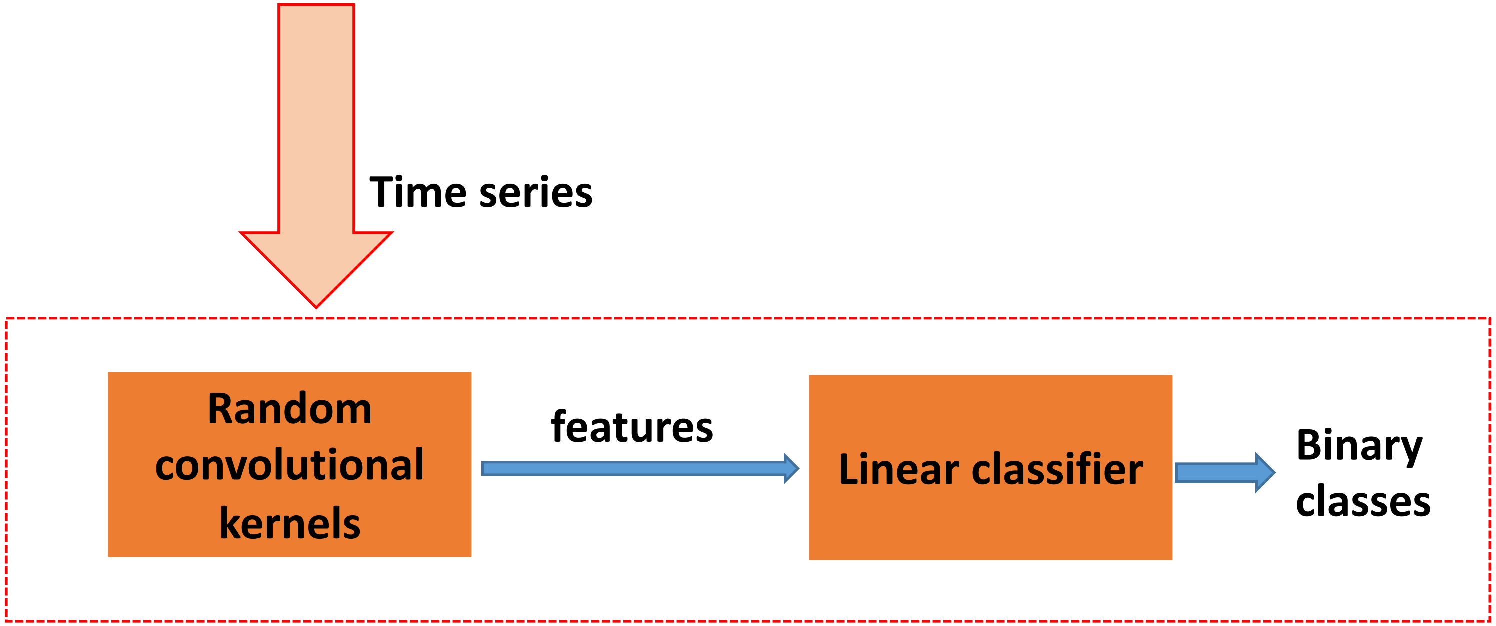
•Transformed features:

•Each input time series is convolved with 10000 random convolutional kernels. Kernels with random length, weights, bias, dilation, and padding.

Linear Classifier:

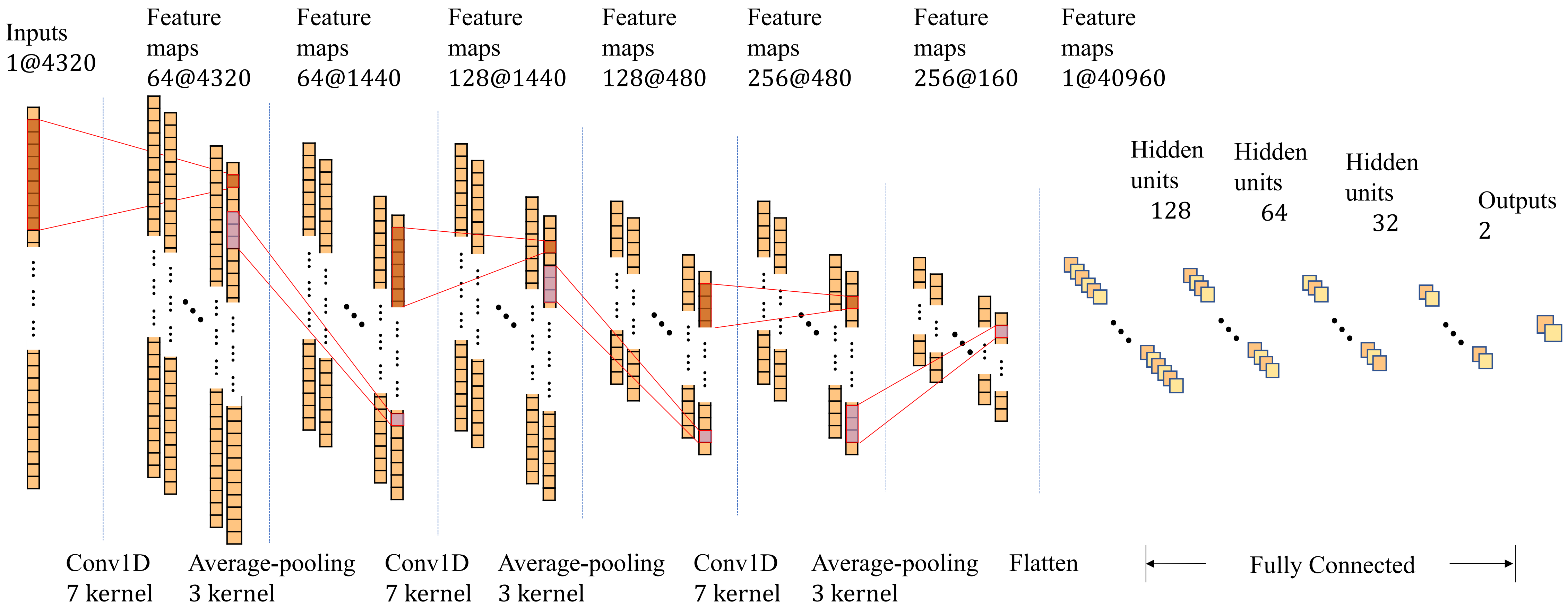
•The model uses logistic regression as classifier.





3.2 1D-CNN

- The convolutional layer **slides a filter across the time series** of the plasma bubble, which is known as the ‘convolutional kernel’.
- Each sample is **flattened** into a single column to serve as input for the model.

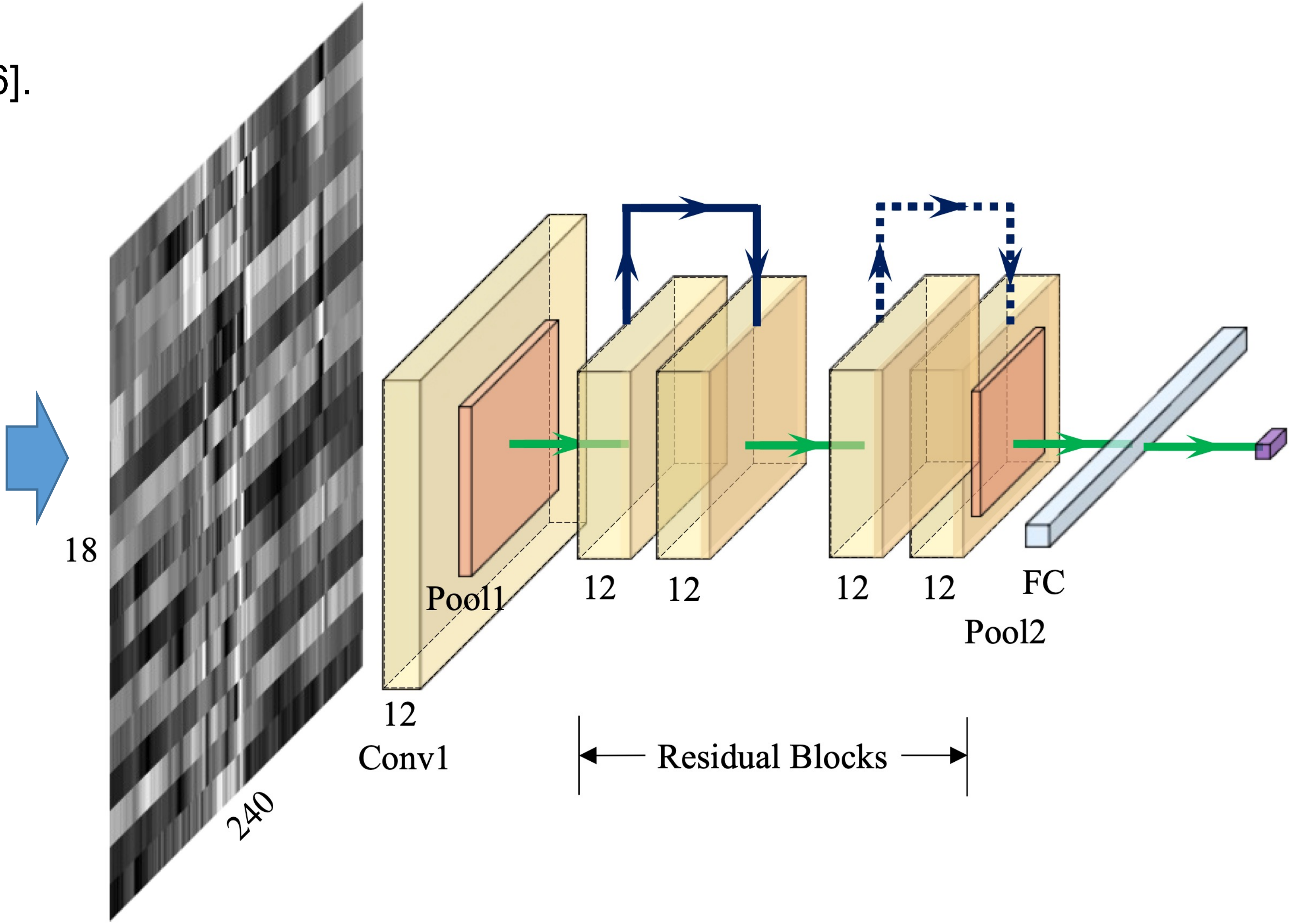


3.3 The residual network

Residual Network (ResNet), which has shown success in **image recognition** [*He et al.*, 2016].

The input time series of the bubble can be converted to **grayscale image** data.

Layer name	Output size	6-layer
conv1	120×9	$7\times7, 12, stride\ 2$
Conv2	120×9	$3\times3\ \max pool, stride\ 2$
	60×5	$\begin{bmatrix}3\times3, 12\\3\times3, 12\end{bmatrix}\times2$
Pool2	59×4	Average pooling
FC1-Softmax	2×1	Fully-Connected, Softmax



4.1 Results – model training

Southern University of Science and Technology

Confusion matrices	*Pred. Null*	*Pred. Event*
Obs. Null	tn	fp
Obs. Event	fn	tp

MINIROCKET	Train		Validation		Test	
	Pred. Null	*Pred. Event*	*Pred. Null*	*Pred. Event*	*Pred. Null*	*Pred. Event*
Obs. Null	63947	116	21251	66	21253	87
Obs. Event	126	1443	81	480	106	432

1D-CNN	Train		Validation		Test	
	Pred. Null	*Pred. Event*	*Pred. Null*	*Pred. Event*	*Pred. Null*	*Pred. Event*
Obs. Null	63865	198	21246	71	21245	95
Obs. Event	110	1459	64	497	78	460

Resnet2D	Train		Validation		Test	
	Pred. Null	*Pred. Event*	*Pred. Null*	*Pred. Event*	*Pred. Null*	*Pred. Event*
Obs. Null	63762	301	21211	106	21229	111
Obs. Event	178	1391	76	485	89	449

4.1 Results – model training

- **All three models** exhibit **over 99% accuracy**, with **precision** and **recall** rates **exceeding 80%.** Moreover, their $\boldsymbol{F_2}$ **scores** are **above 80%.**
- The $\boldsymbol{F_2}$ **score** for MINIROCKET, 1D-CNN, and ResNet in **test sets** are 81%, 85%, and 83%, respectively.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN},$$
$$Precision = \frac{TP}{TP + FP},$$
$$Recall = \frac{TP}{TP + FN},$$
$$\text{and } F_\beta = (1 + \beta^2) \cdot \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}.$$

F_2 score	Train	Validation	Test
MINIROCKET	**0.9209**	0.8602	0.8087
1D-CNN	0.9196	**0.8837**	**0.8496**
2D-ResNet	0.8729	0.8554	0.8278

4.2 Results - prediction in 2021

SUSTech Southern University of Science and Technology

We utilized observed data from P3(THD), P4(THE), and P5(THA) in the year 2021.

Only traditional Crateria	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81782	181
Obs. Event	0	189

MINIROCKET	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81880	83
Obs. Event	93	96

CNN1D	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81848	115
Obs. Event	101	88

ResNet	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81741	222
Obs. Event	89	100

To improve the recall rate (detectable rate) of bubbles identified by the three models, we combined their respective prediction results using intersection and union.

Intersection	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81928	35
Obs. Event	123	66

Union	**2021**	
	Pred. Null	*Pred. Event*
Obs. Null	81682	281
Obs. Event	63	126

4.2 Results - prediction in 2021

Due to the **extremely unbalanced** sample in 2021, which has a ratio of 435 non-bubbles to bubbles, the model can detect about half of the actual bubbles.

Methods	Only Traditional Criteria	MINIROCKET	CNN-1D	ResNet	Intersection	Union
Precision	0.5108	0.5363	0.4335	0.3106	0.6535	**0.3096**
Recall	**1.0000**	**0.5079**	**0.4656**	**0.5291**	**0.3492**	**0.6667**
F_2 score	0.8393	0.5134	0.4588	0.4638	0.3851	**0.5417**

- F_2 score indicates that the highest score is obtained when the union set is used, suggesting that it provides the best identification of bubbles.
- Based on the recall rate of the union results, we can identify two-thirds (126 out of 189) of the bubbles in 2021.

Where the term "intersection" refers to a situation where all three models predict a bubble, resulting in the final combination also being classified as a bubble. The term "union" means that the event is classified as a bubble if at least one of the three models predicts it as such.

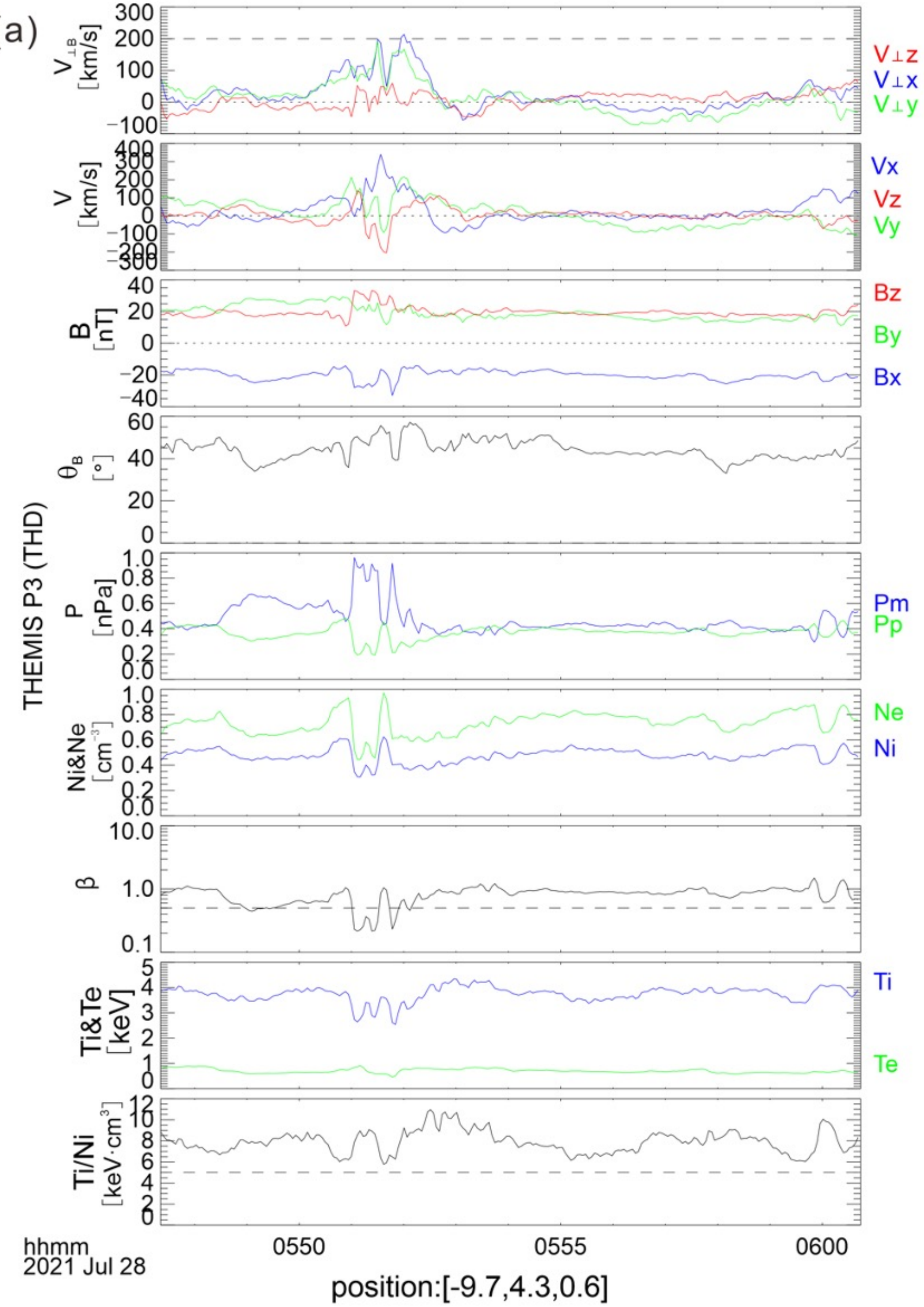
5. Summary

- 1. This study utilized a plasma-sheet **bubble dataset** that was created through a combination of traditional criteria and manual inspection with an **imbalanced** ratio of 1:40.
- 2. **Bubble identification**: (defined as multivariate time series classification). The models included **MINIROCKET**, a traditional machine learning method, **1D-CNN**, a deep learning technique, and **ResNet**, a two-dimensional deep learning technique typically used for image recognition.
- 3. **All three models were effective** in recognizing bubbles, with precision, recall, and F_2 score reaching 80% **on the training-validation-test set**.
- 4. When predicting in 2021, combining the results of three models to create a **union set improved** the accuracy of **predicted results**, **increasing** the **recall rate and F_2 score**.

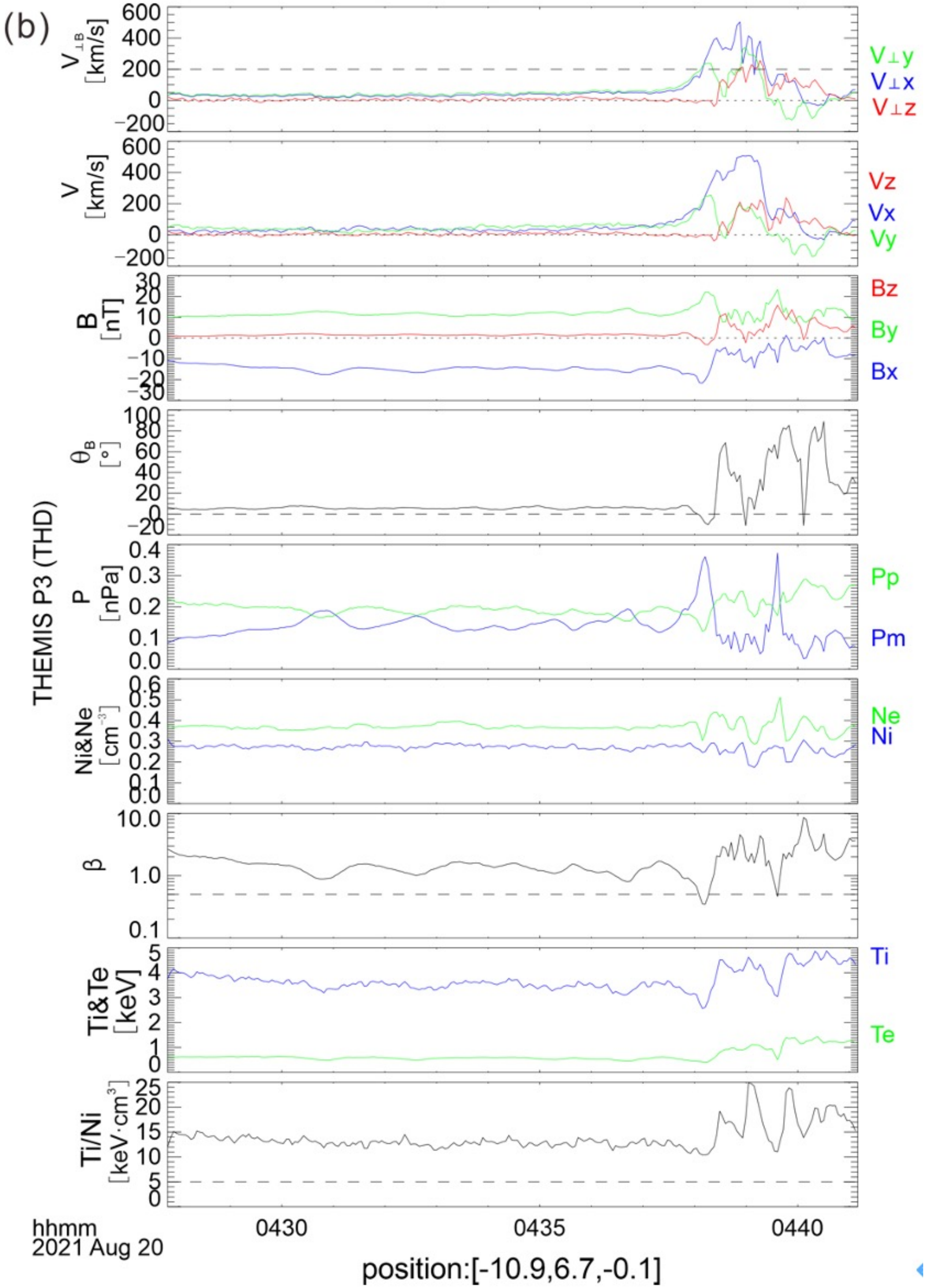
The data and model files are available at https://github.com/pinecypressfxd/Bubble_Identification_MTS

- Thanks for your attention!

FN cases

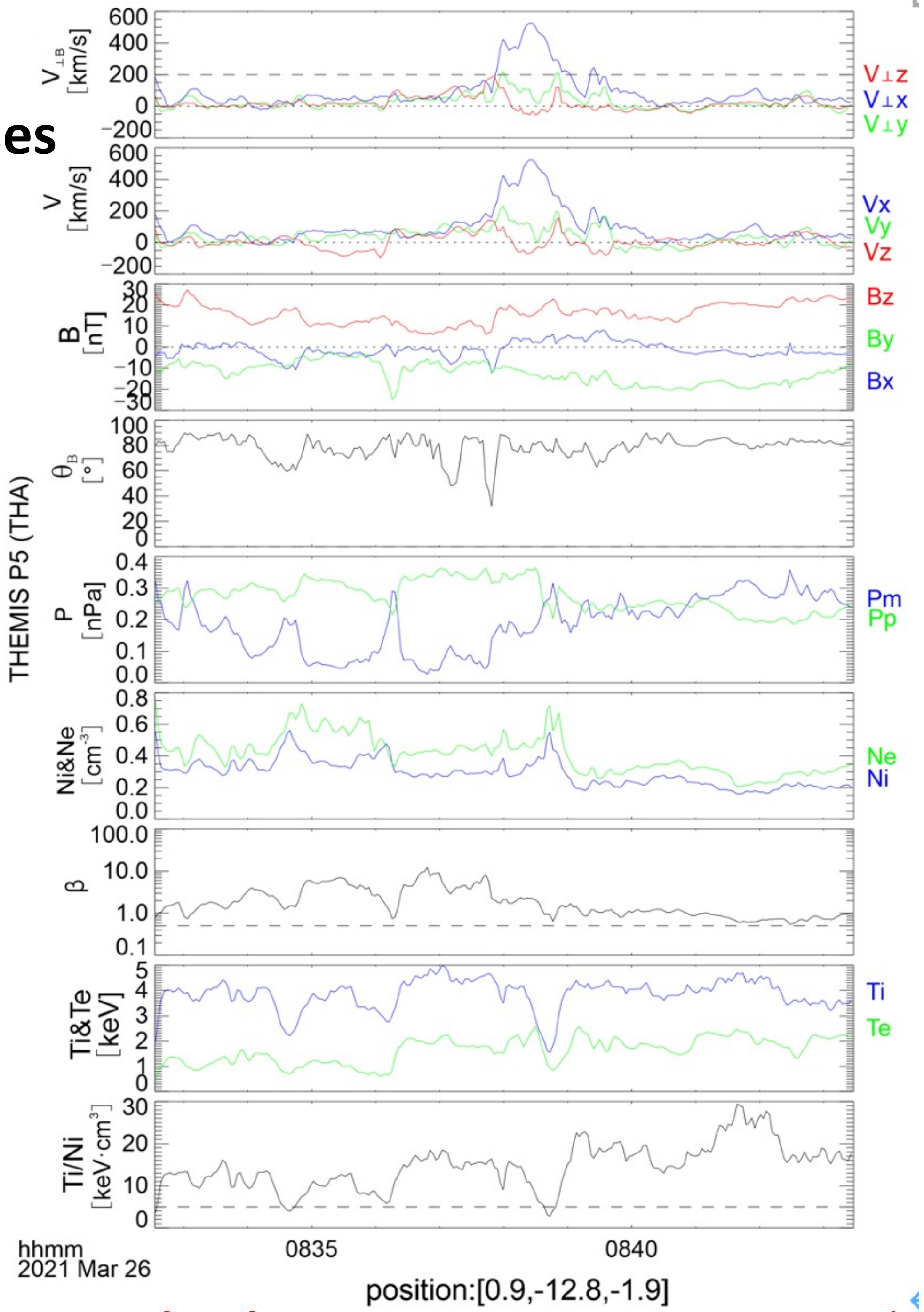


a few data points' velocities exceed 200km/s

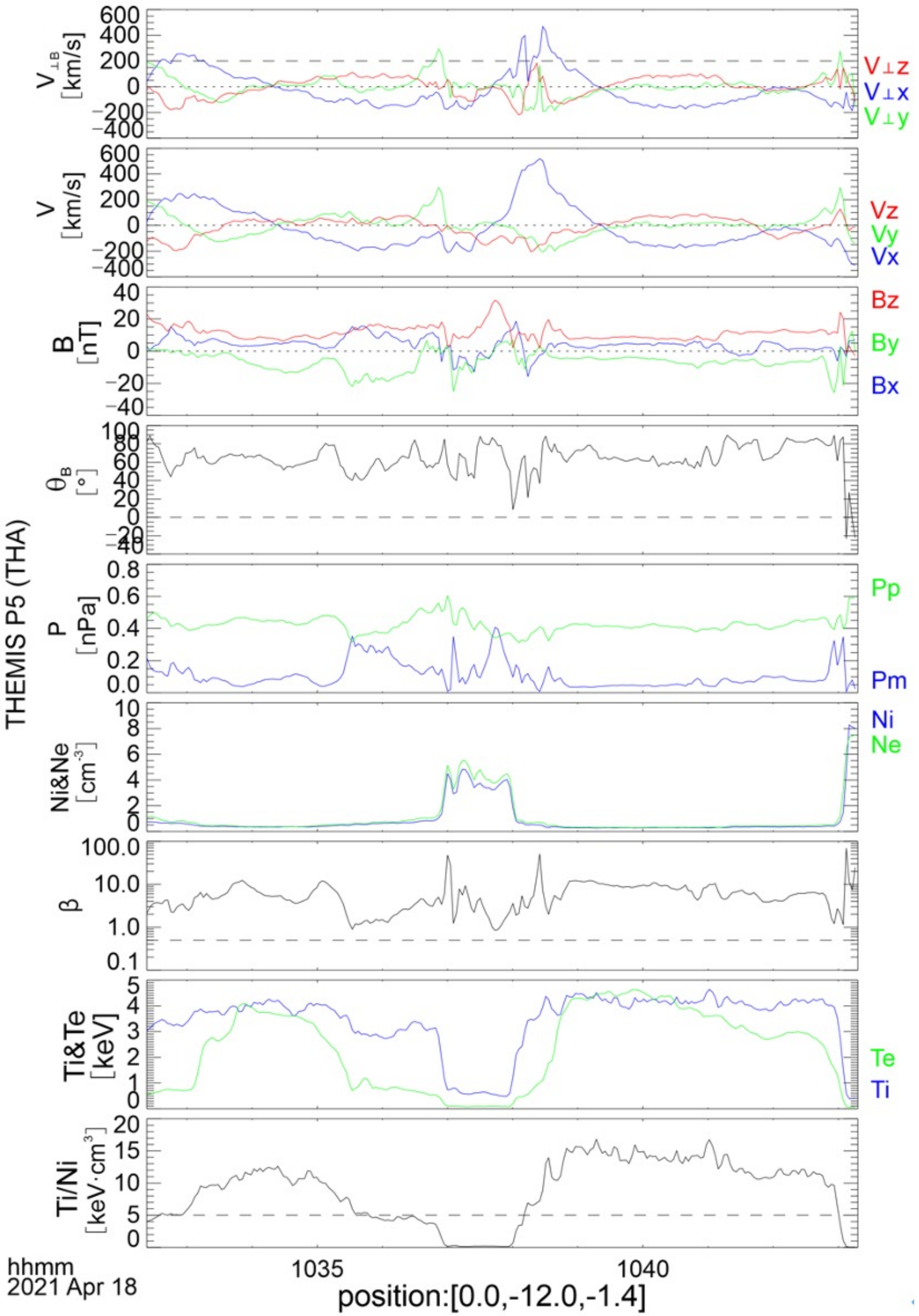


not complete event

FP cases



earthward fast-flow events,temperatures do not rise



the cool and dense magneto-sheath.