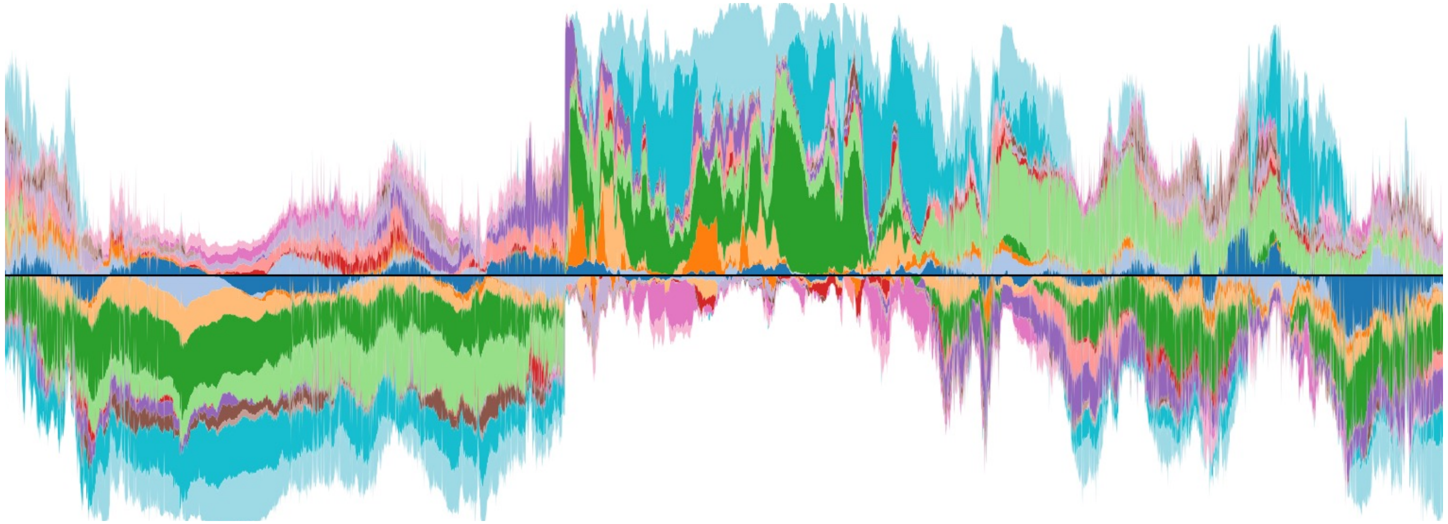
Ground Magnetic Perturbation Predictions Using Machine Learning and Interpretability Methods



Mike Coughlan, Amy Keesee, Victor Pinto, Jose P. Marchezi, Raman Mukundan, Jeremiah Johnson, Hyunju Connor, Don Hampton

MAGICIAN

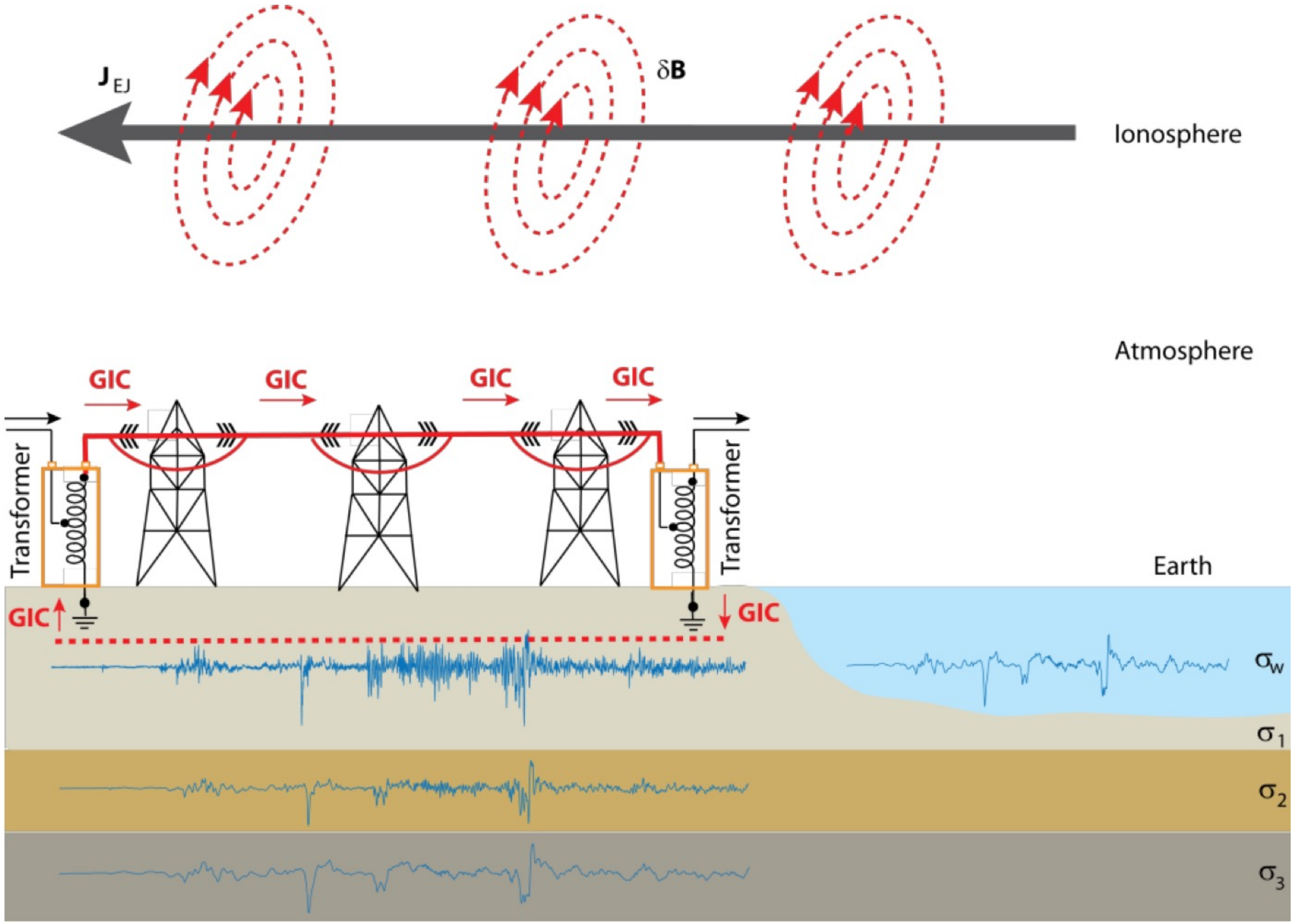
Machine Learning Algorithms for Geomagnetically Induced Currents in Alaska and New Hampshire

Outline

- Introduction
  - Geomagnetically Induced Currents (GICs)
  - dB/dt Localization – A Problem for Forecasting
- Our Methodology
  - A Classification Problem
  - Very Brief Introduction to Convolutional Neural Networks (CNNs)
  - ML Interpretability - SHapley Additive exPlanation (SHAP) Values
- Results
  - A Whole Lot of Figures

Geomagnetically Induced Currents (GIC)

Or: How I Learned to Stop Worrying and Love the Sun



[Rosenqvist (2018)]

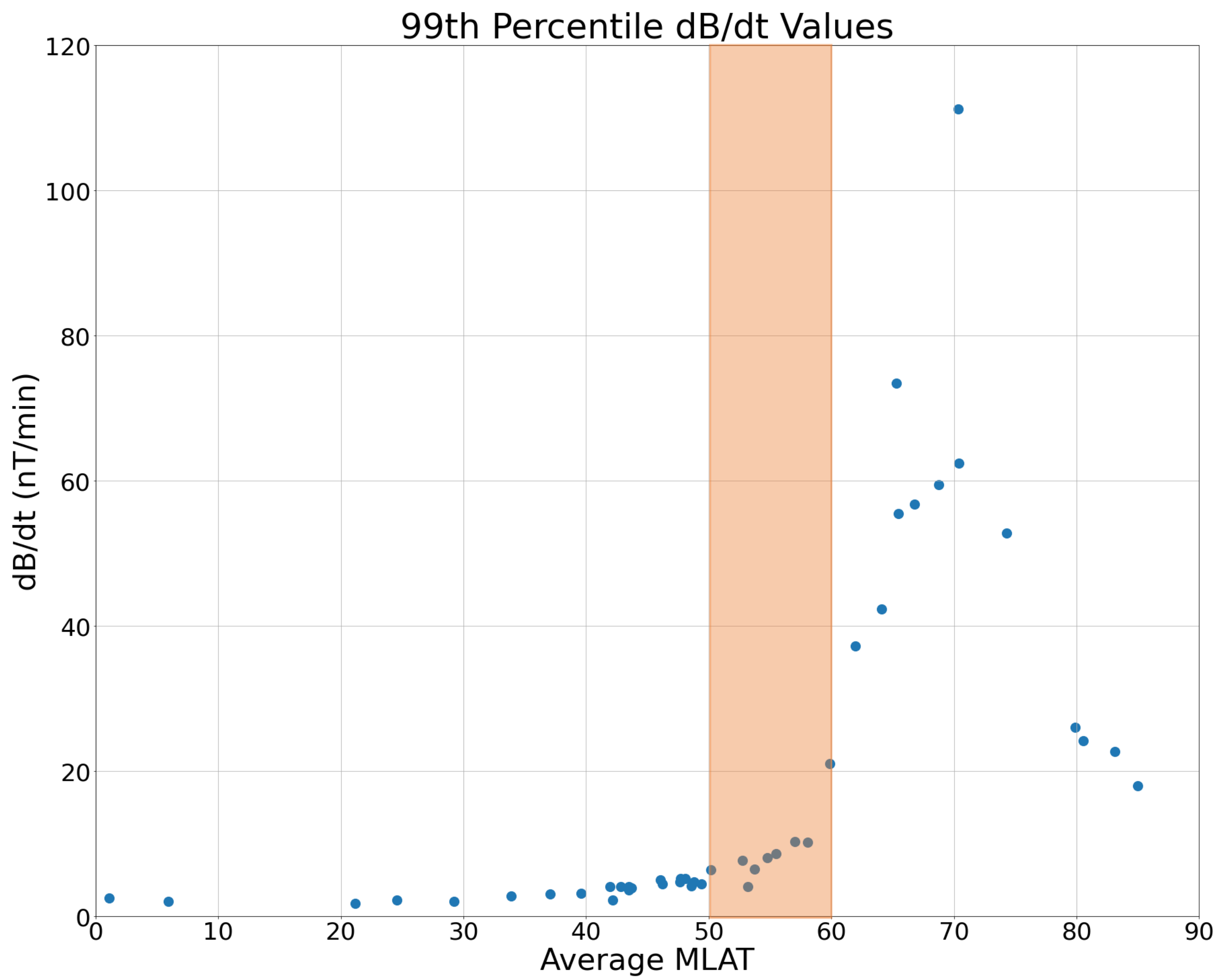
$$\frac{dB}{dt}$$

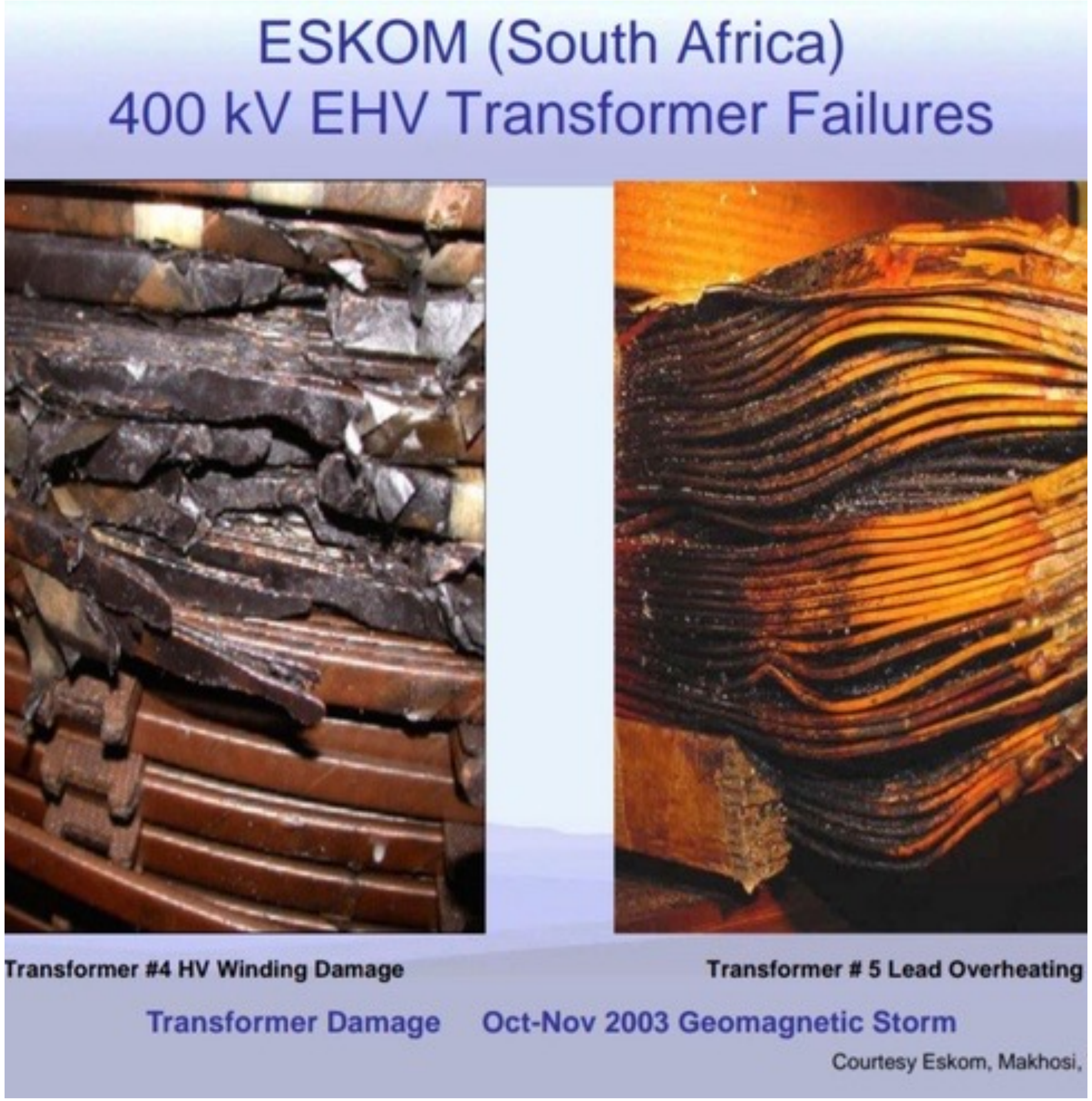
$$\downarrow$$

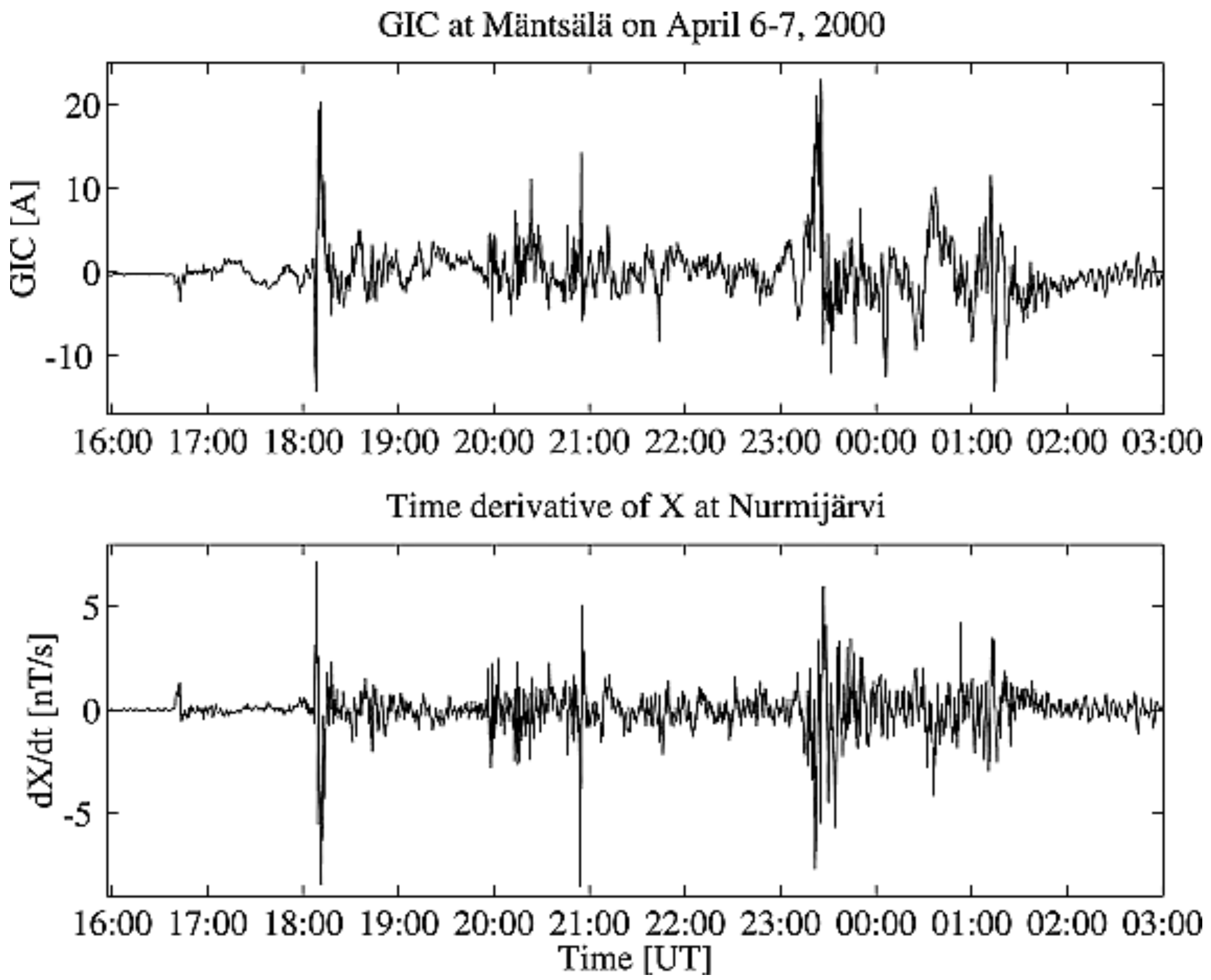
$$E_{x,y} = \pm \frac{Z}{\mu_0} \frac{dB_{y,x}}{dt}$$

$$\downarrow$$

$$GIC(t) = aE_x(t) + bE_y(t)$$


99th Percentile dB/dt Values
dB/dt (nT/min)
Average MLAT
0
20
40
60
80
100
120
0
10
20
30
40
50
60
70
80
90

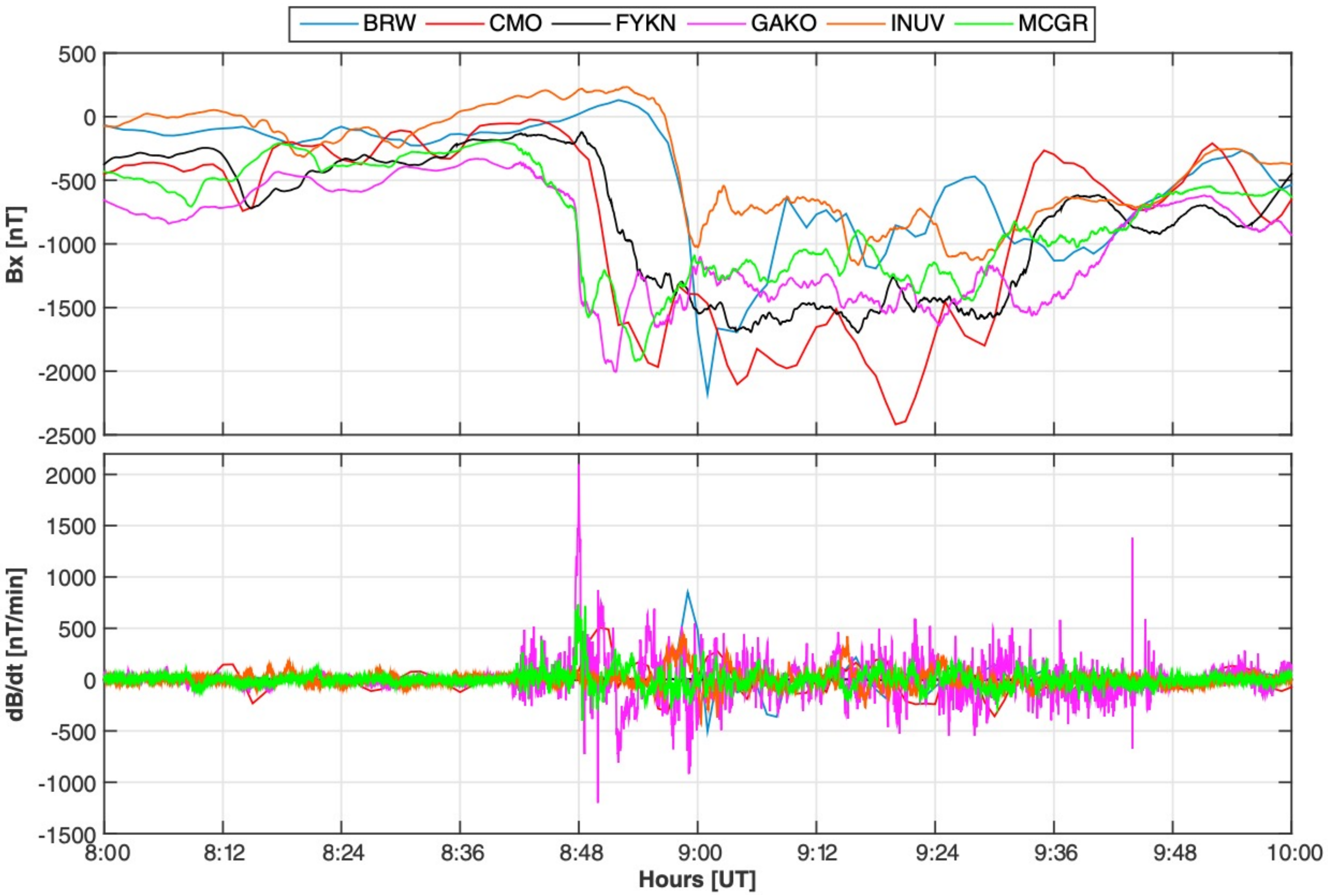

ESKOM (South Africa)
400 kV EHV Transformer Failures
Transformer #4 HV Winding Damage
Transformer # 5 Lead Overheating
Transformer Damage Oct-Nov 2003 Geomagnetic Storm
Courtesy Eskom, Makhosi,



[Huttunen (2002)]

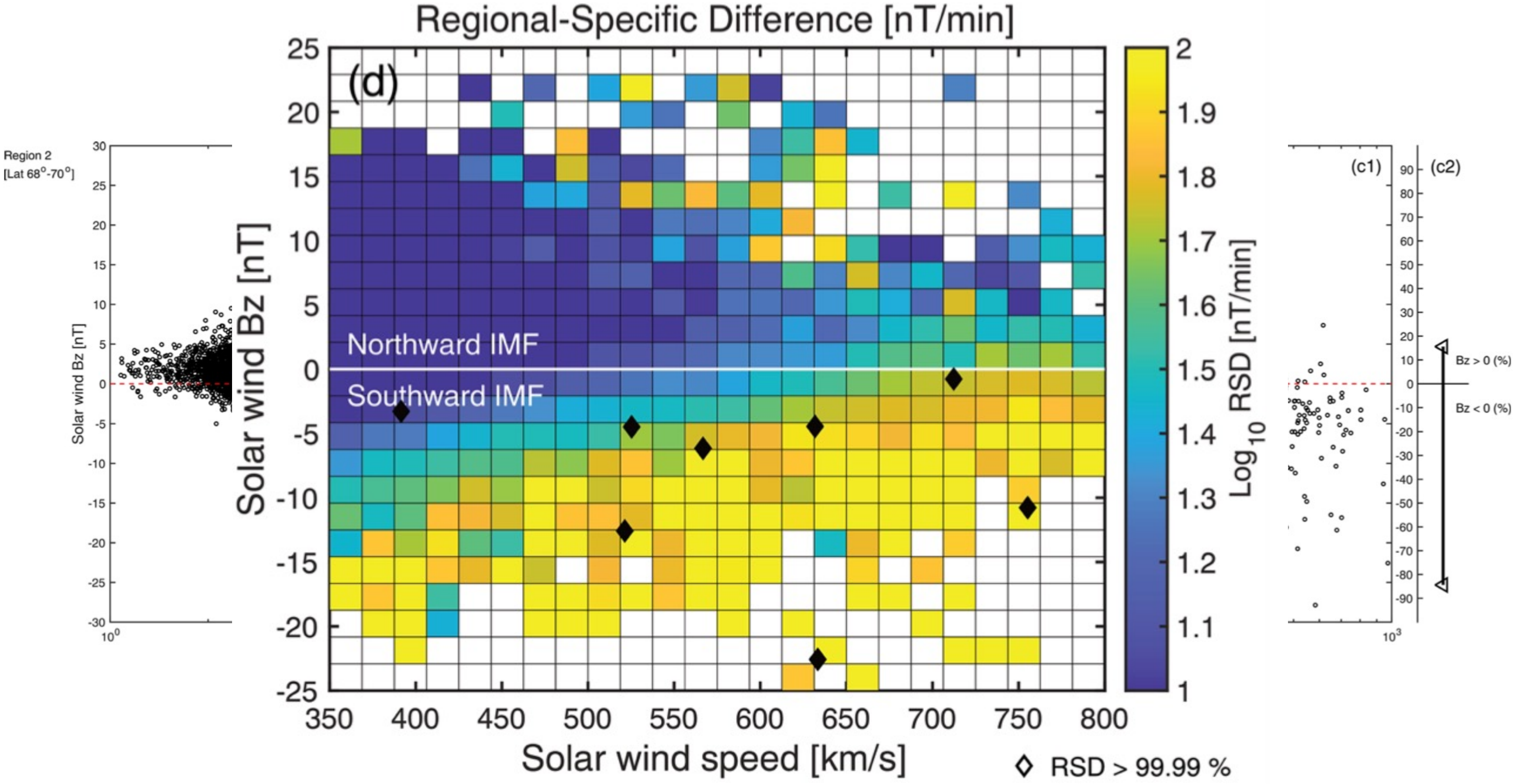
dB/dt Localization – A Problem for Forecasting

March 09, 2012 Storm



- The localized nature of dB/dt has been an observed phenomenon for some time.
- Ngwira et al. (2018) found peak in $dB_x/dt \sim 4x$ as high as the next highest peak (Alaska & Western Canada).

Ngwira C. M., et al. (2018). A study of intense local dB⁄dt variations during two geomagnetic storms. Space Weather, 16, 676–693. https://doi.org/10.1029/2018SW001911



Dimmock, et al. (2020). On the regional variability of dB/dt and its significance to GIC. Space Weather, 18, e2020SW002497.
https:// doi.org/10.1029/2020SW002497

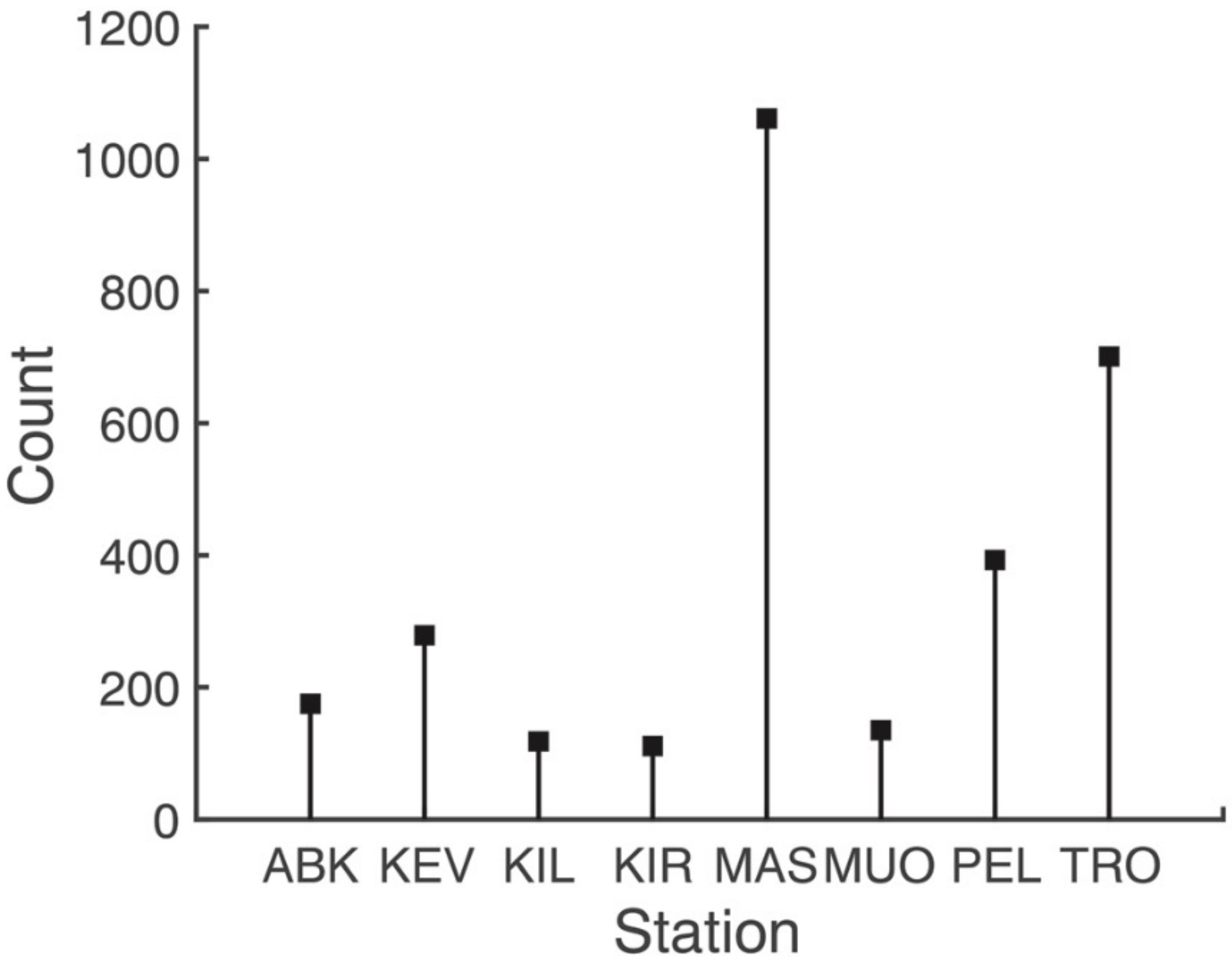


Figure 11. Histogram showing the stations which were responsible for the largest *RSD*. Note that the MAS station appears to be statistically significant.

Dimmock, et al. (2020). On the regional variability of dB/dt and its significance to GIC. Space Weather, 18, e2020SW002497. https:// doi.org/10.1029/2020SW002497

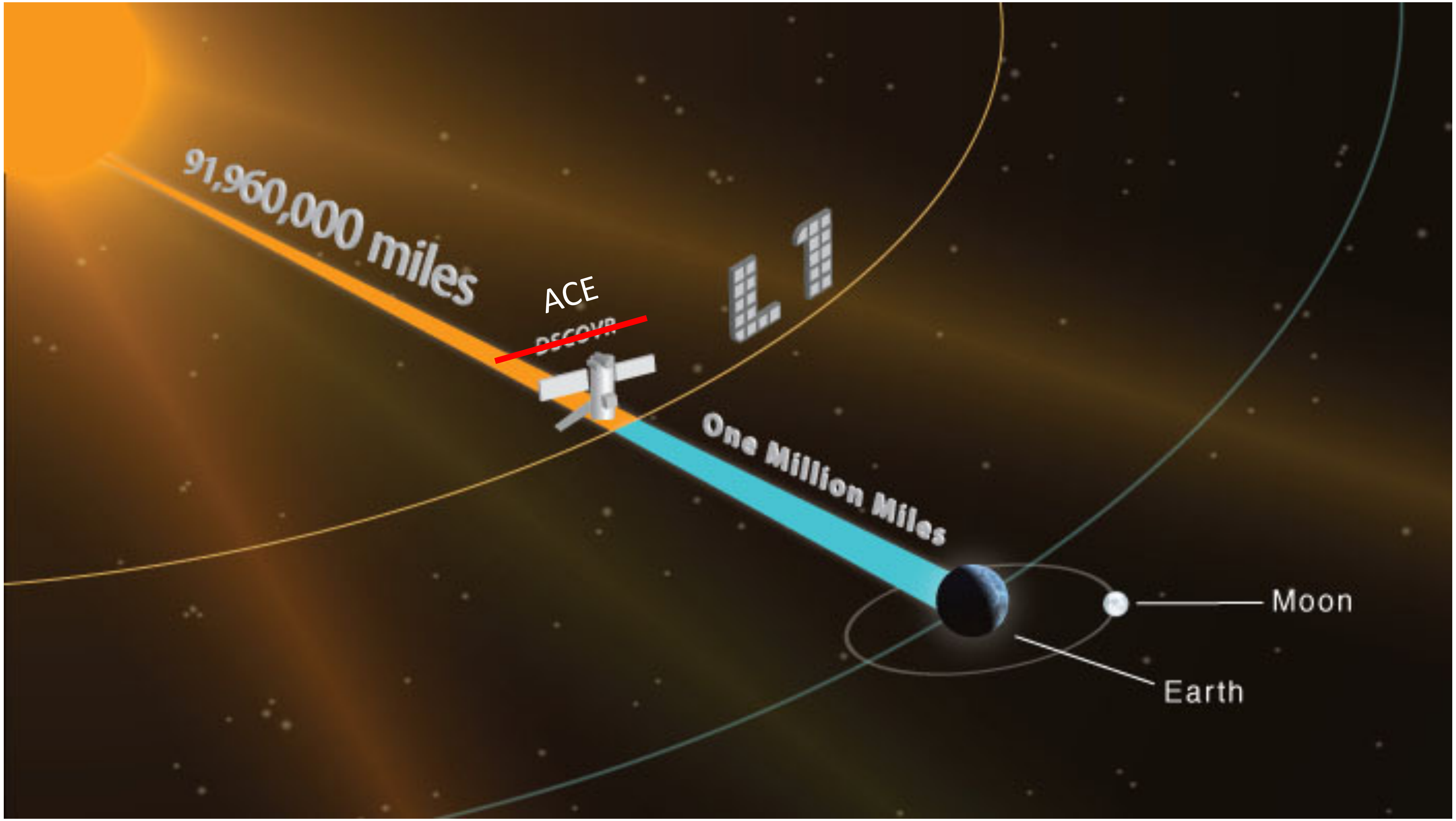


Image Credit: NOAA

Solar wind data alone could be insufficient to capture localization

A Classification Problem

- Rapid variations in dB/dt complicate direct prediction

- Can often capture general shape of storm, but misses higher resolution variations and overall magnitude.

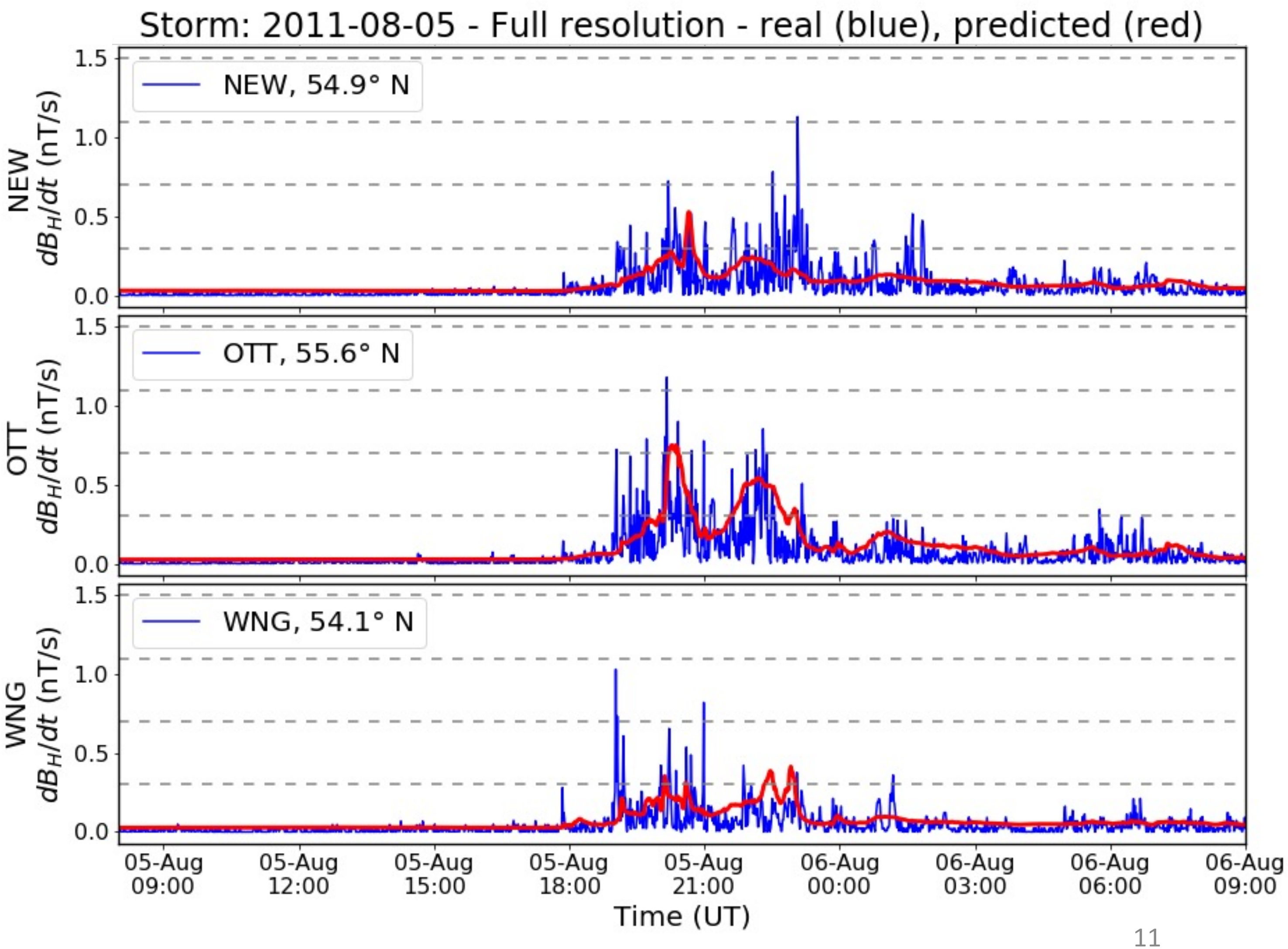


Image credit: V. Pinto

A Classification Problem

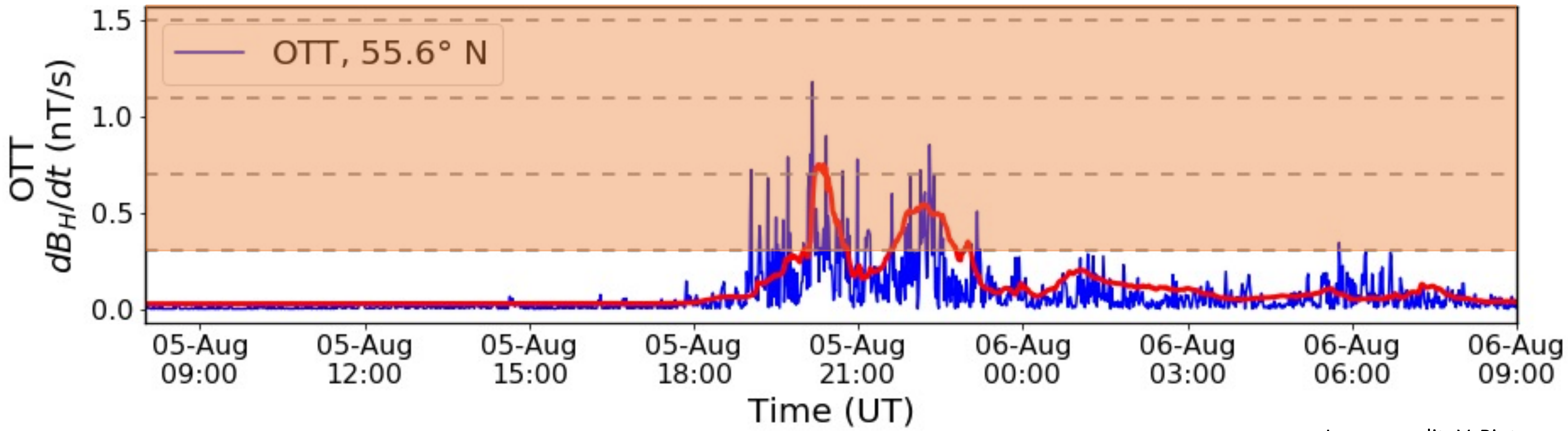


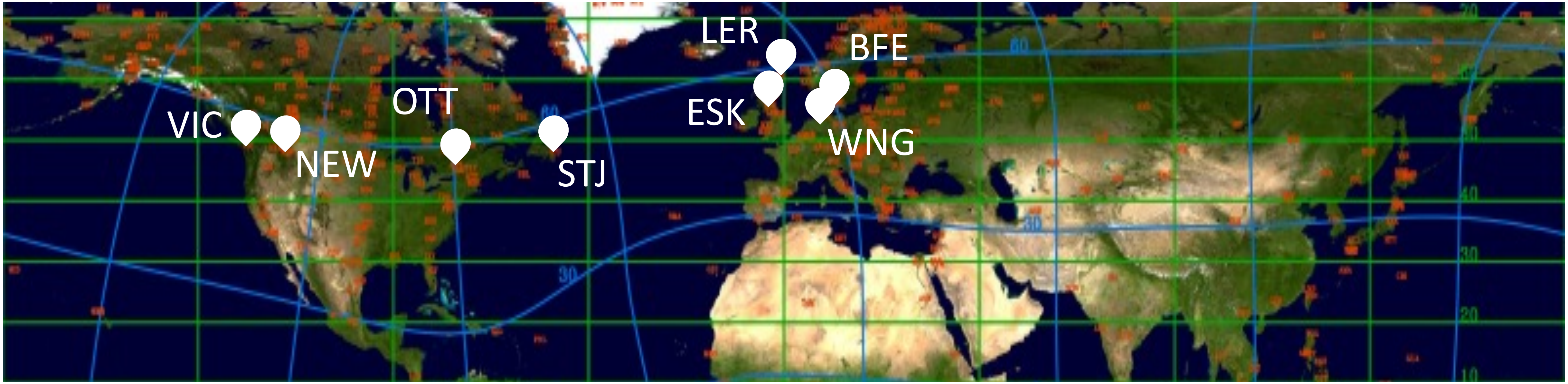
Image credit: V. Pinto

A Classification Problem

Time History | Prediction Window

$t-31, \boxed{t-30, t-29, \ldots, t-1}, t, t+1, \ldots, t+29, \boxed{t+30, t+31, \ldots, t+58, t+59}, t+60$

- 30-minutes of time history
- 30-minute window -- 30 minutes in the future
- Probabilistic prediction of dB/dt going over a threshold within that prediction window.



- Average MLAT falls between 50° - 60°
- Span ~ 9 MLT

18 nT/min

BFE: 99.88th

WNG: 99.94th

LER: 99.74th

ESK: 99.89th

STJ: 99.93rd

OTT: 99.82nd

NEW: 99.86th

VIC: 99.91st

99th Percentile

BFE: 6.80 nT/min

WNG: 5.59 nT/min

LER: 8.05 nT/min

ESK: 6.52 nT/min

STJ: 4.94 nT/min

OTT: 7.15 nT/min

NEW: 6.65 nT/min

VIC: 5.46 nT/min

Inputs

Model	Input parameters
Solar wind	B_y^{GSM}, B_z^{GSM}, B_{total}^{GSM}, V_x, V_y, V_z, ρ_{sw}, T, sin(MLT), cos(MLT)
Combined	B_y^{GSM}, B_z^{GSM}, B_{total}^{GSM}, V_x, V_y, V_z, ρ_{sw}, T, sin(MLT), cos(MLT), B_N, B_E, B_H, $\boxed{dB/dt,}$ AE Index, SZA

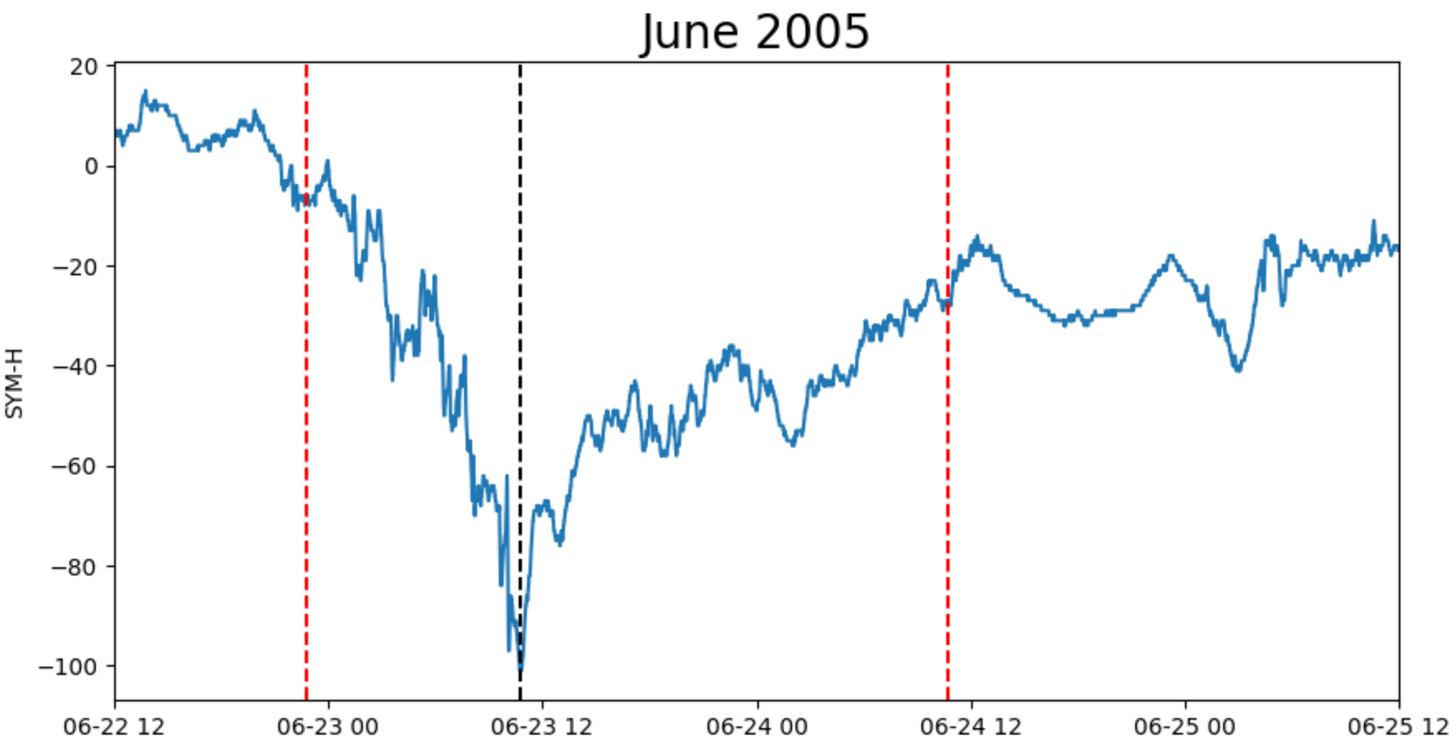
A Concern: Could including the variable used to make the target array turn this model into something close to a persistence model?

The Persistence Model: 1 if the dB/dt input exceeds the threshold in the input window, 0 otherwise.

$$t-31, \boxed{t-30, t-29, \dots, t-1,} t, t+1, \dots, t+29, \boxed{t+30, t+31, \dots, t+58, t+59}, t+60$$

Data Prep

- Identify time periods with SYM-H < -50 nT for at least 2 hours
- Identify point of minimum SYM-H
- Add 12 hours of lead time and 24 hours of recovery time to min point
- Results in 397 training storms, 8 test storms
- At 1-minute time resolution 850,000 training samples



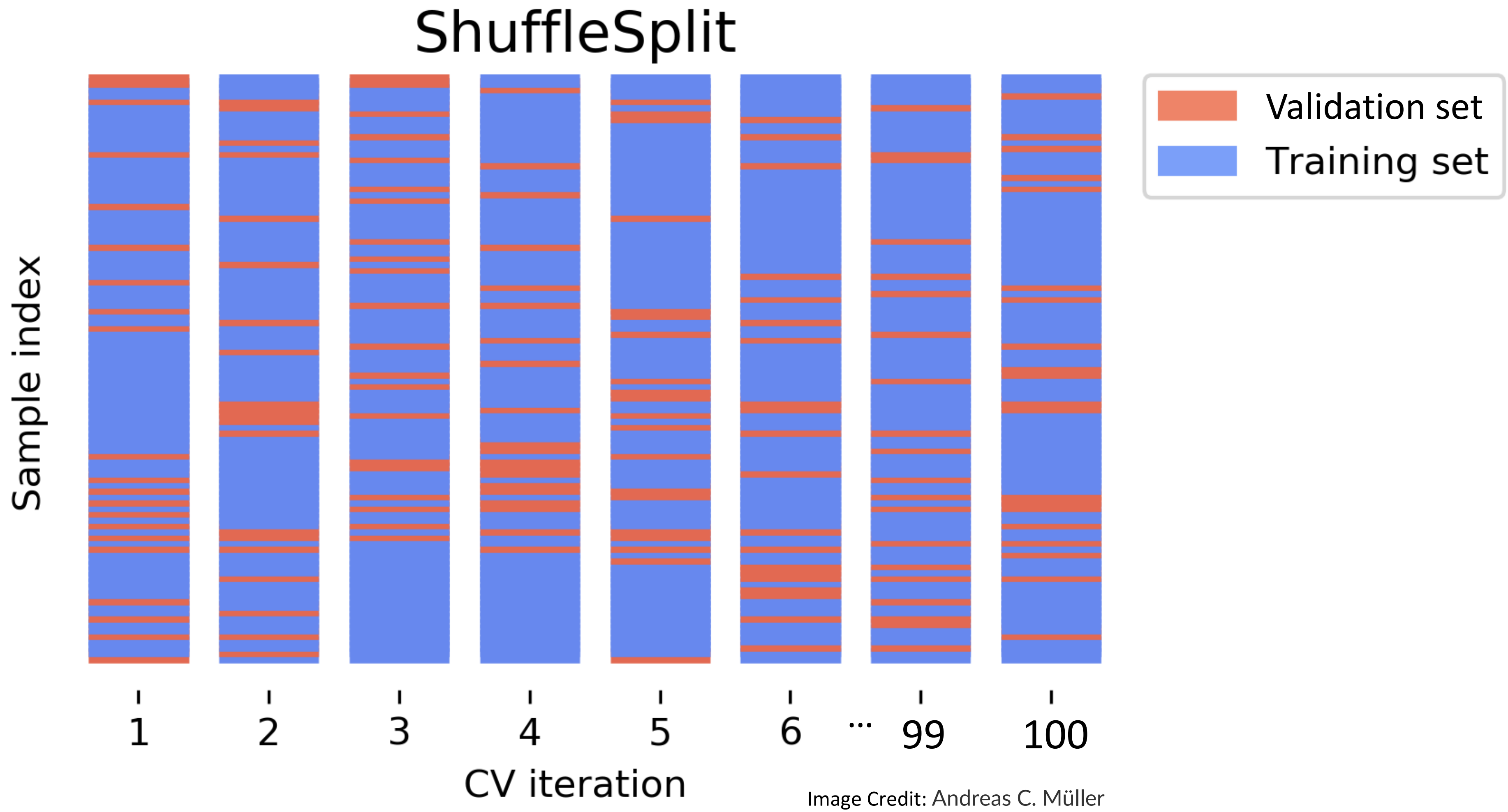
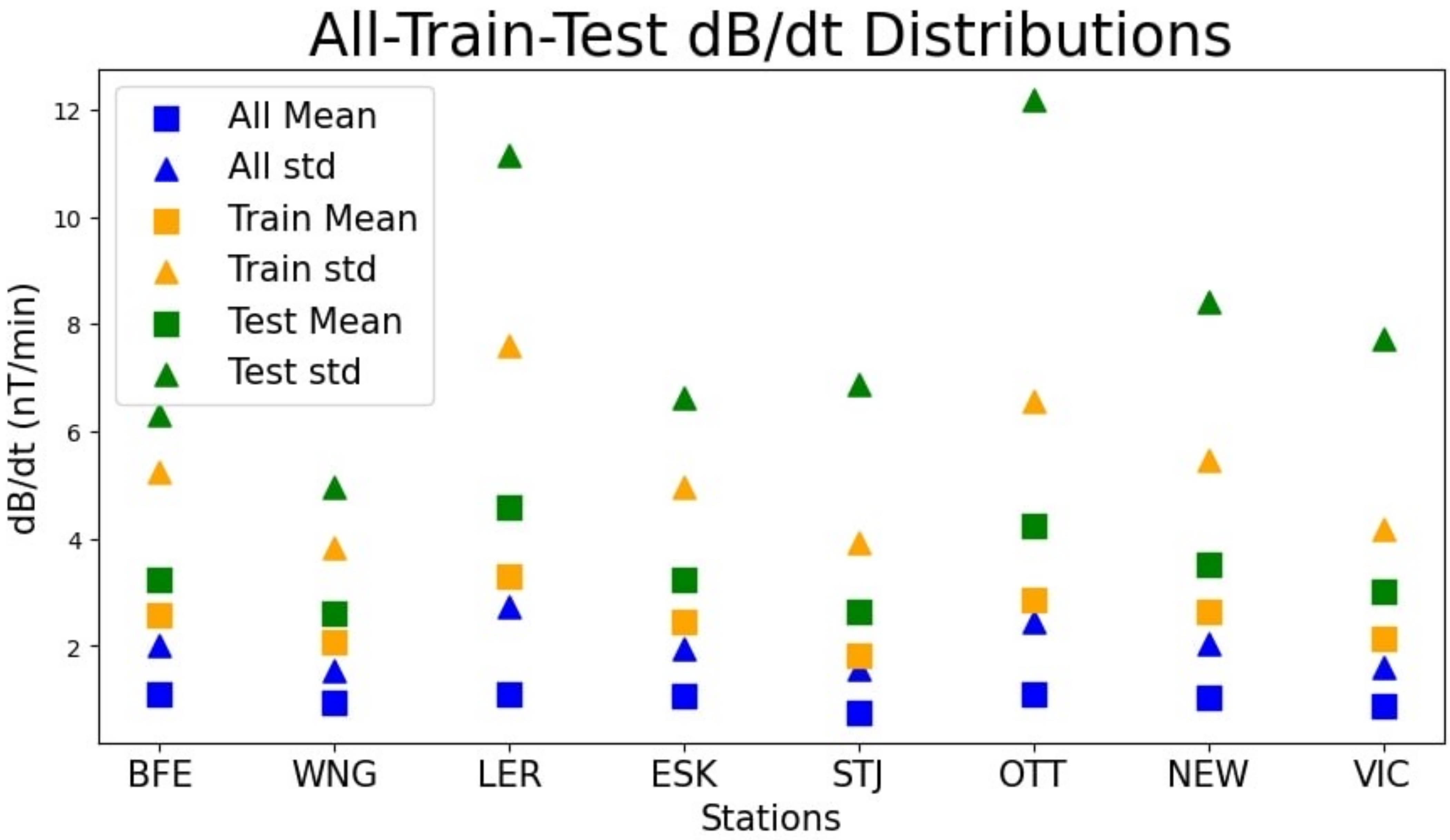
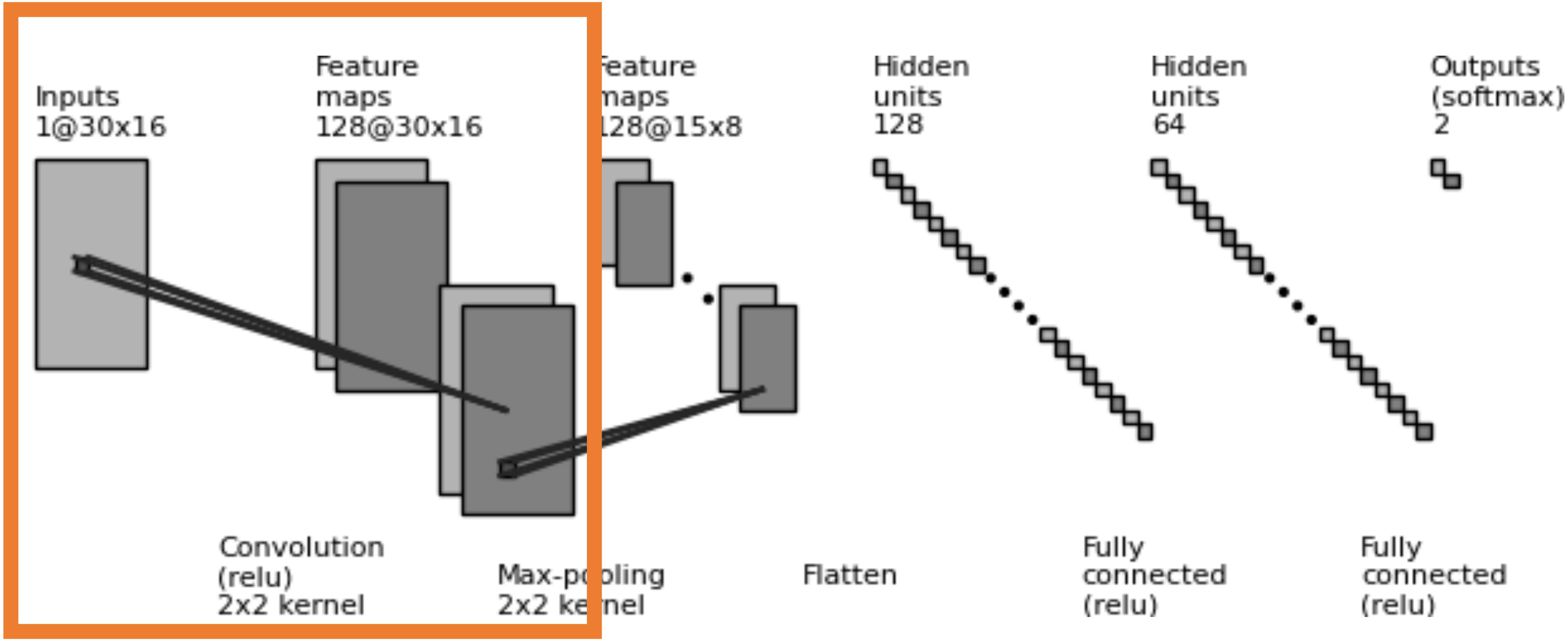


Image Credit: Andreas C. Müller

- Ensemble was created using 100 model trained on unique training-validation pairs with 80%-20% training–validation splits.


All-Train-Test dB/dt Distributions
All Mean
All std
Train Mean
Train std
Test Mean
Test std
dB/dt (nT/min)
12
10
8
6
4
2
BFE
WNG
LER
ESK
STJ
OTT
NEW
VIC
Stations

A Very Brief Introduction to Convolutional Neural Networks (CNNs)



-5	-2	10	0
0	-1	4	9
2	-4	-1	4
3	9	1	-7

1	0
-1	1

relu: $f(x) = max(0,x)$

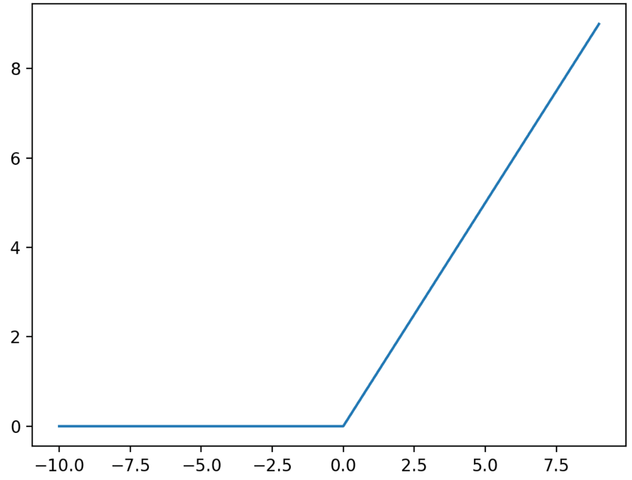


Image Credit: Jason Brownlee

$$\mathrm{Relu}(a \cdot w + b)$$

5	-2	10	0
0	-1	4	9
2	-4	-1	4
3	9	1	-7

1	0
-1	1

5		

$$(5 * 1) + (-2 * 0) + (0 * -1) + (-1 * 1) + 1 = 5$$

$$Relu(5) = 5$$

5	-2	10	0
0	-1	4	9
2	-4	-1	4
3	9	1	-7

1	0
-1	1

5	4	

$$(-2 * 1) + (10 * 0) + (-1 * -1) + (4 * 1) + 1 = 4$$

$$Relu(4) = 4$$

5	-2	10	0
0	-1	4	9
2	-4	-1	4
3	9	1	-7

1	0
-1	1

5	4	16
0	3	10
9	0	0

Different dimensions

0	0	0	0	0
0	-5	-2	10	0
0	0	-1	4	9
0	2	-4	-1	4
0	3	9	1	-7

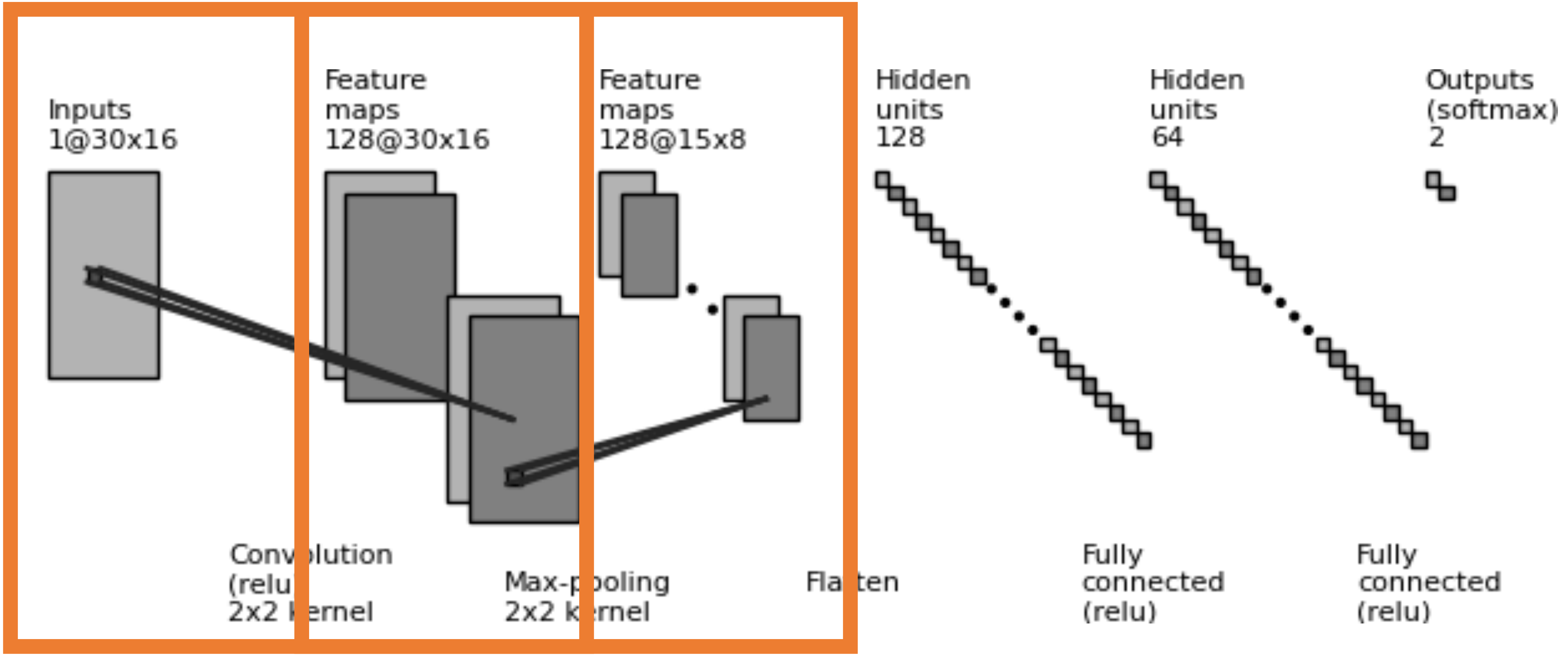
"Padding" – retains the initial dimensions

1	0
-1	1

5	4	16
0	3	10
9	0	0

0	4	13	0
0	5	4	16
3	0	3	10
4	9	0	0

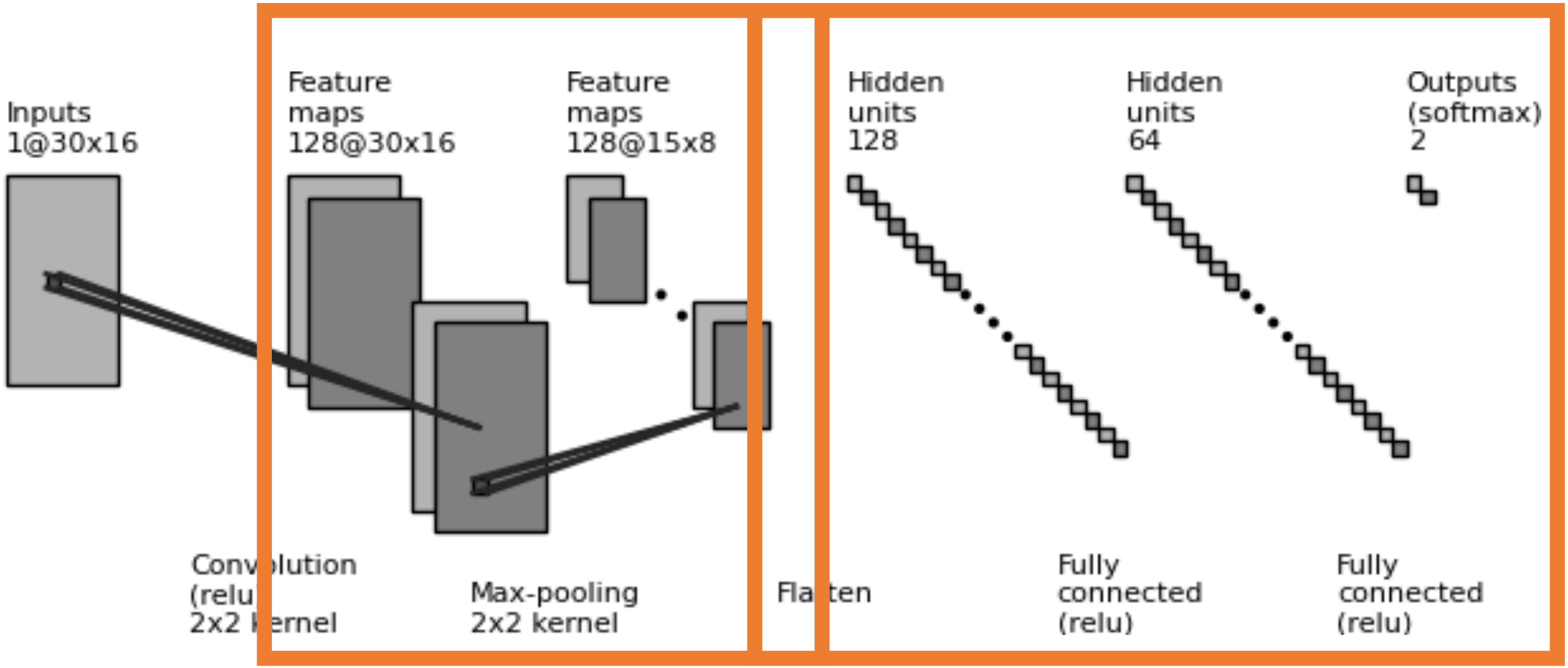
A Very Brief Introduction to Convolutional Neural Networks (CNNs)

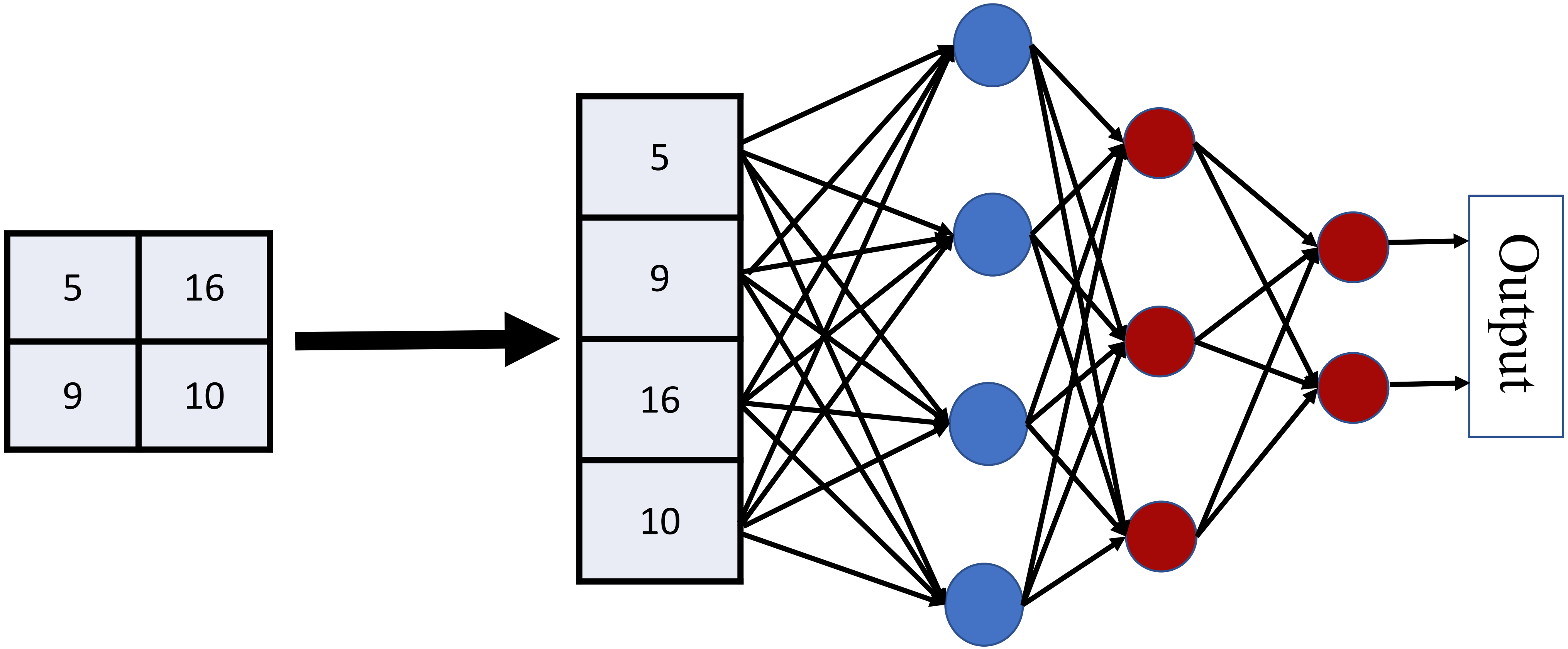


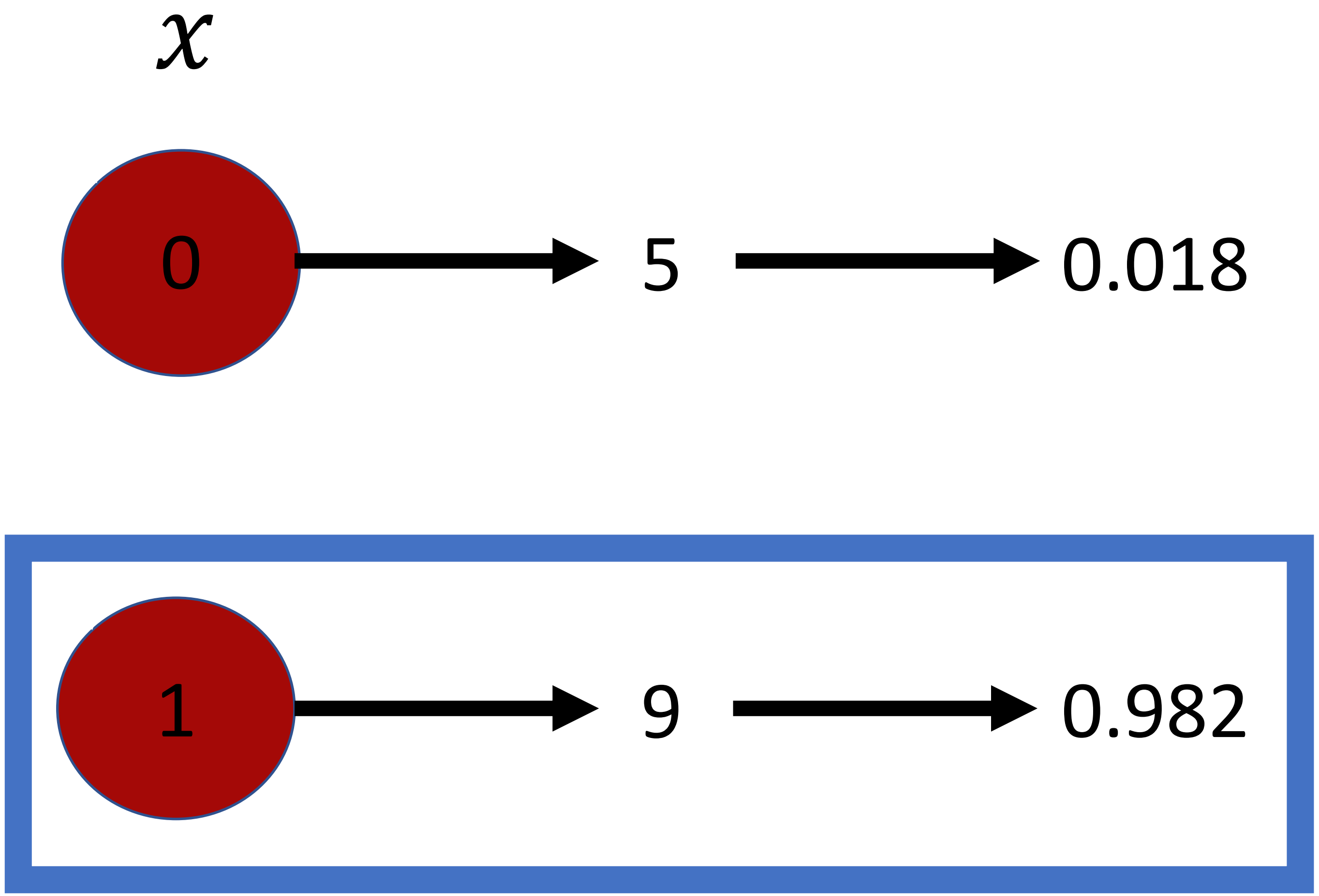
0	4	13	0
0	5	4	16
3	0	3	10
4	9	0	0

5	16
9	10

A Very Brief Introduction to Convolutional Neural Networks (CNNs)

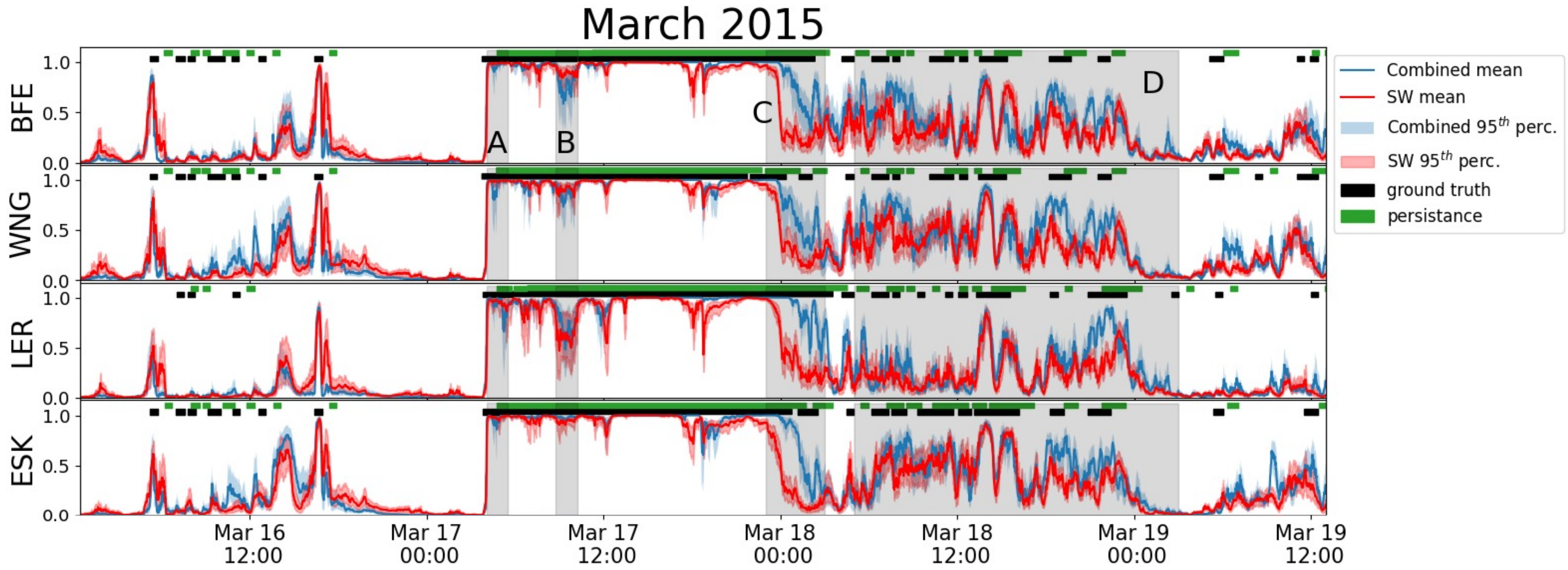



5
16
9
10
5
9
16
10
Output

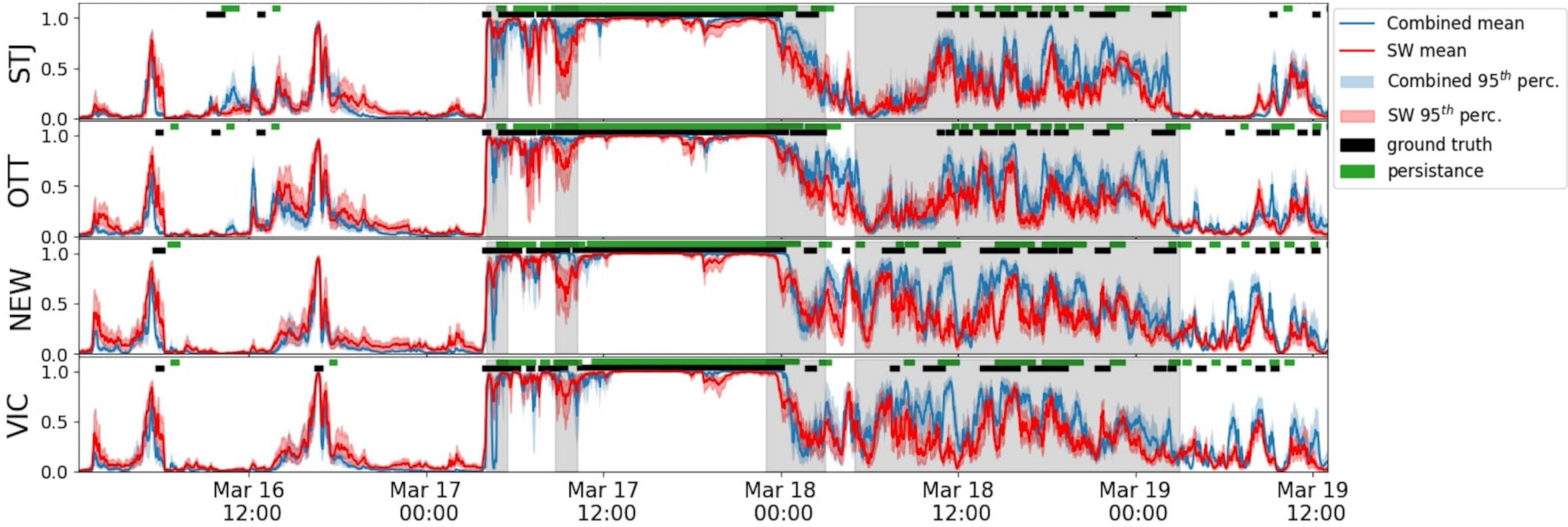


$$softmx(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

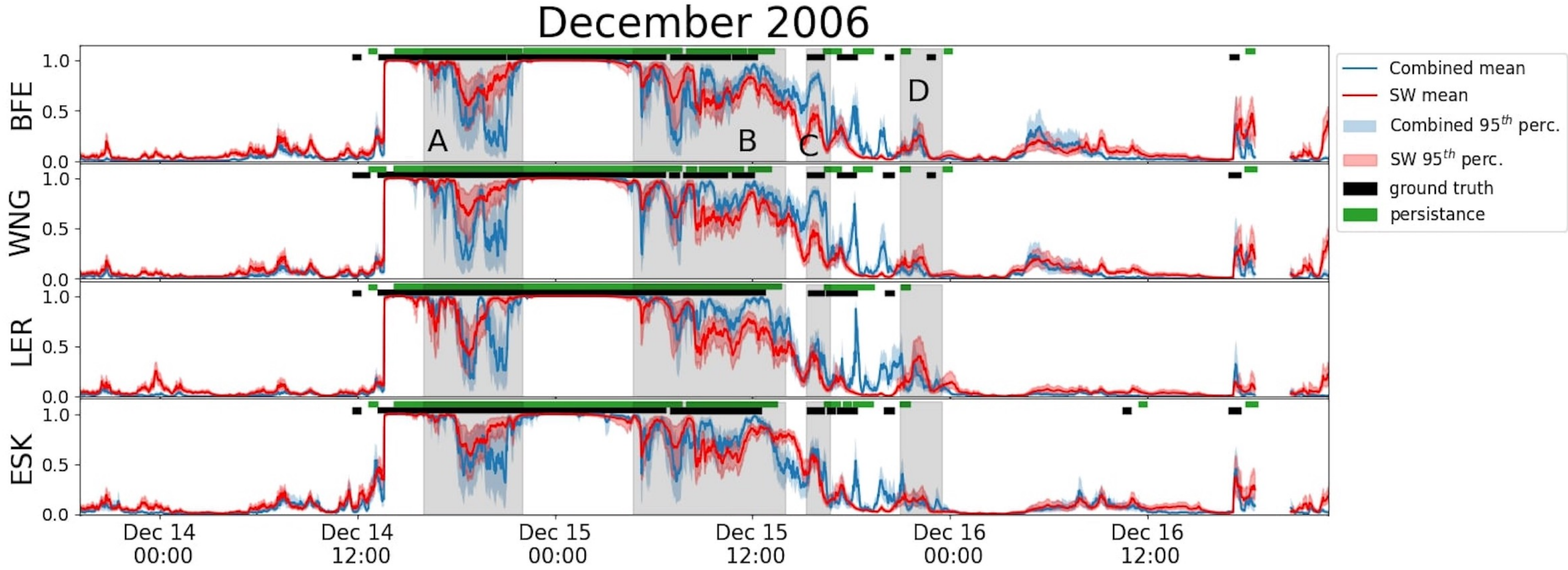
2015 St. Patrick's Day Storm



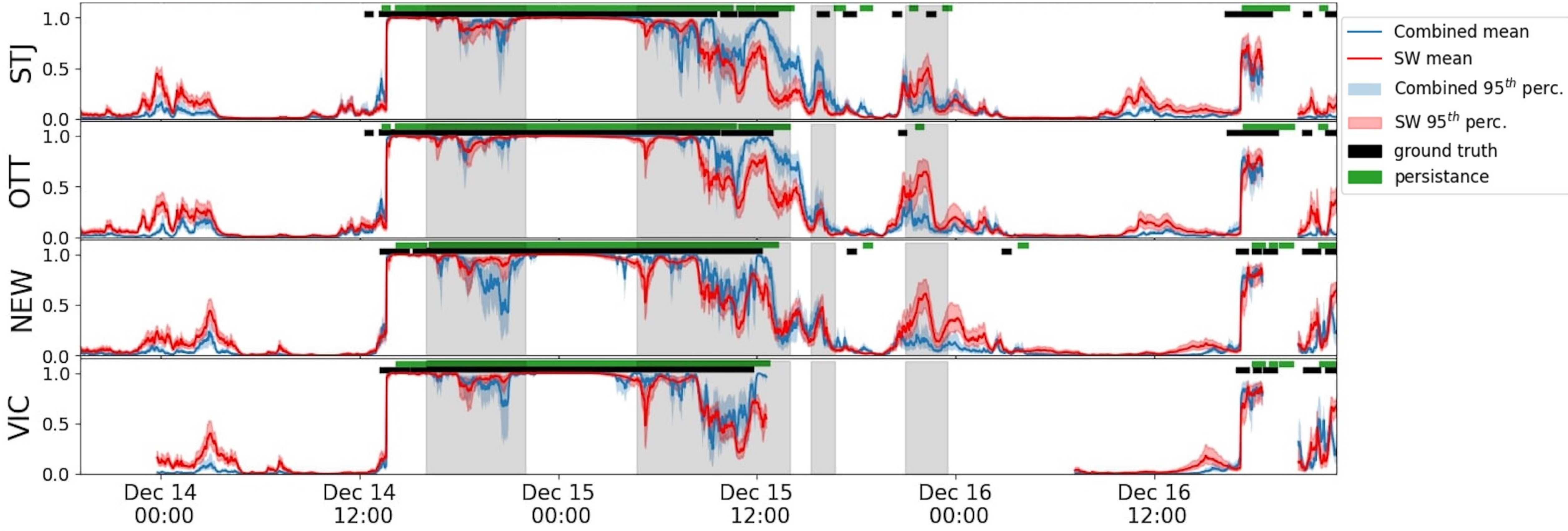
2015 St. Patrick's Day Storm



December 2006 Storm



December 2006 Storm



ML Interpretability - SHapley Additive exPlanation (SHAP) Values

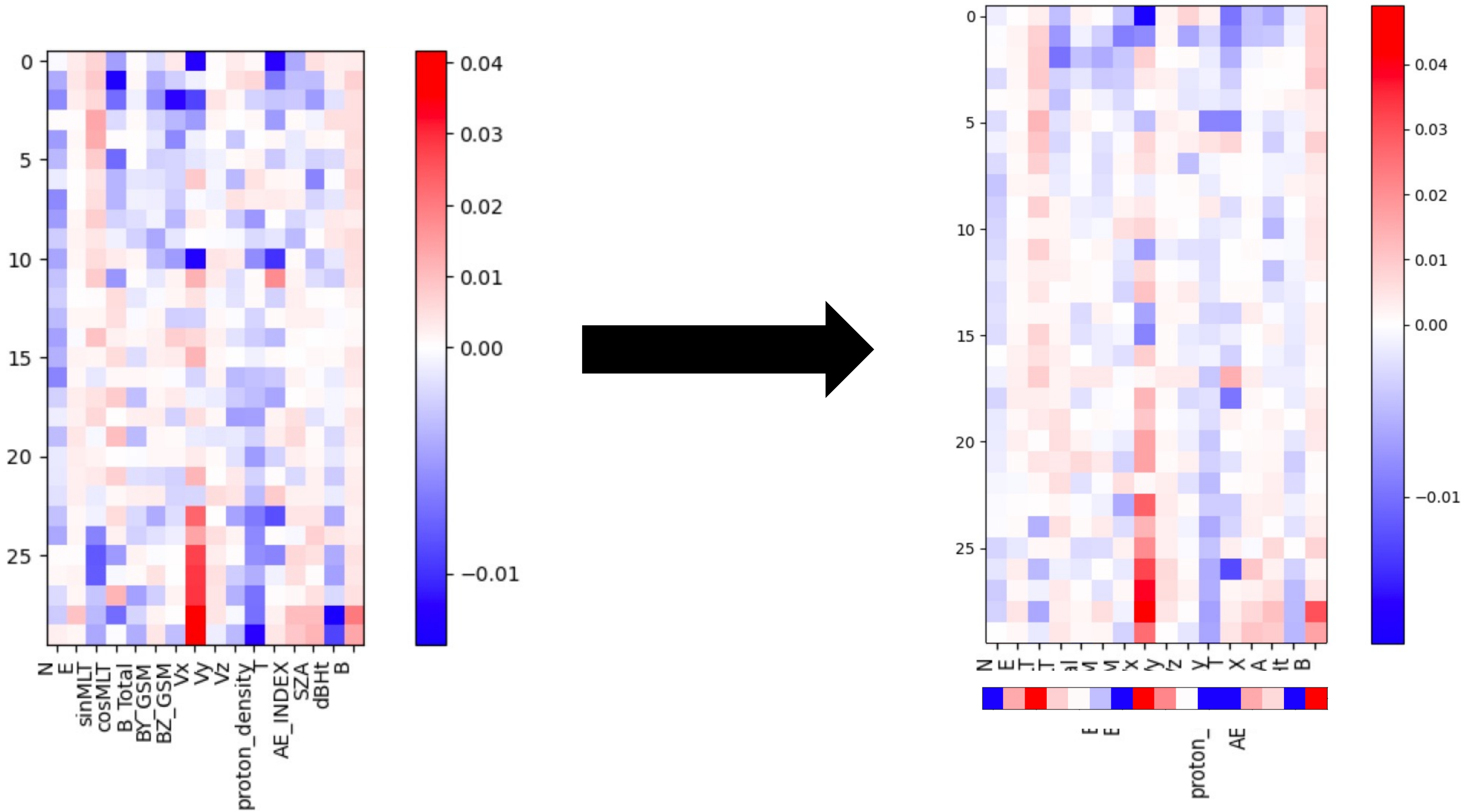
- Way of approxi)
  the model's out
- Based on Shapl
- See Donglai Ma
  explanation!

A VALUE FOR n-PERSON GAMES

L. S. Shapley

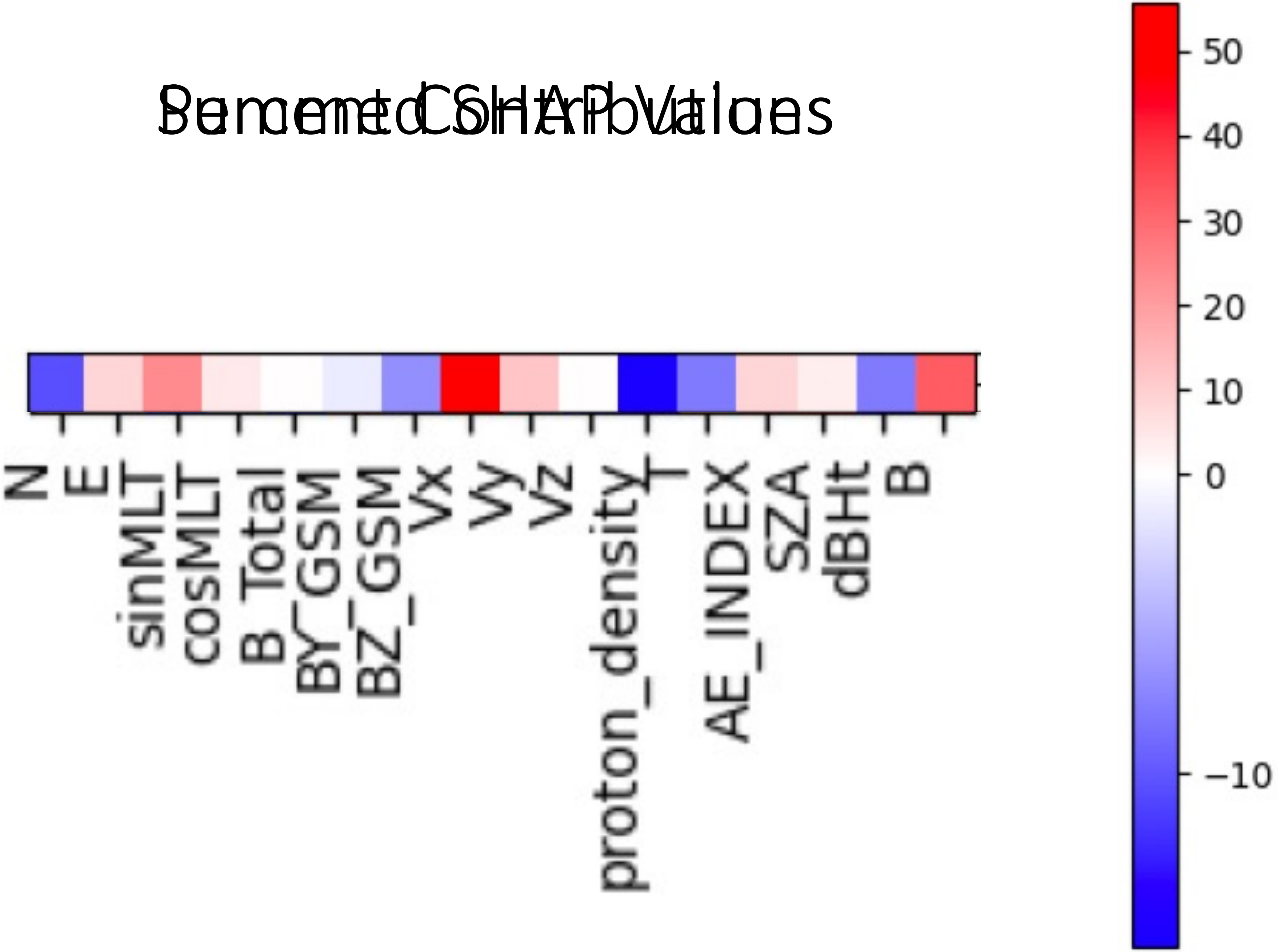
P-295

18 March 1952

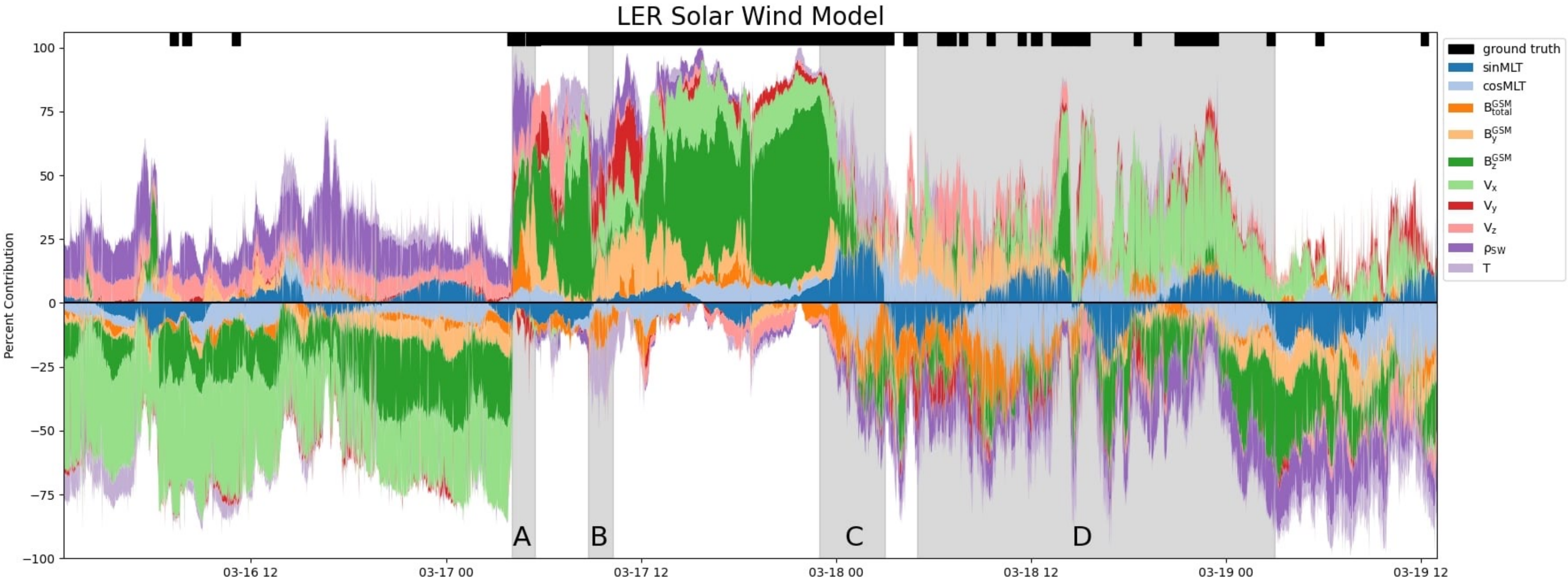

0
5
10
15
20
25
0.04
0.03
0.02
0.01
0.00
−0.01
N
E
sinMLT
cosMLT
B_Total
BY_GSM
BZ_GSM
Vx
Vy
Vz
proton_density
T
AE_INDEX
SZA
dBHt
B

Summed SHAP Values

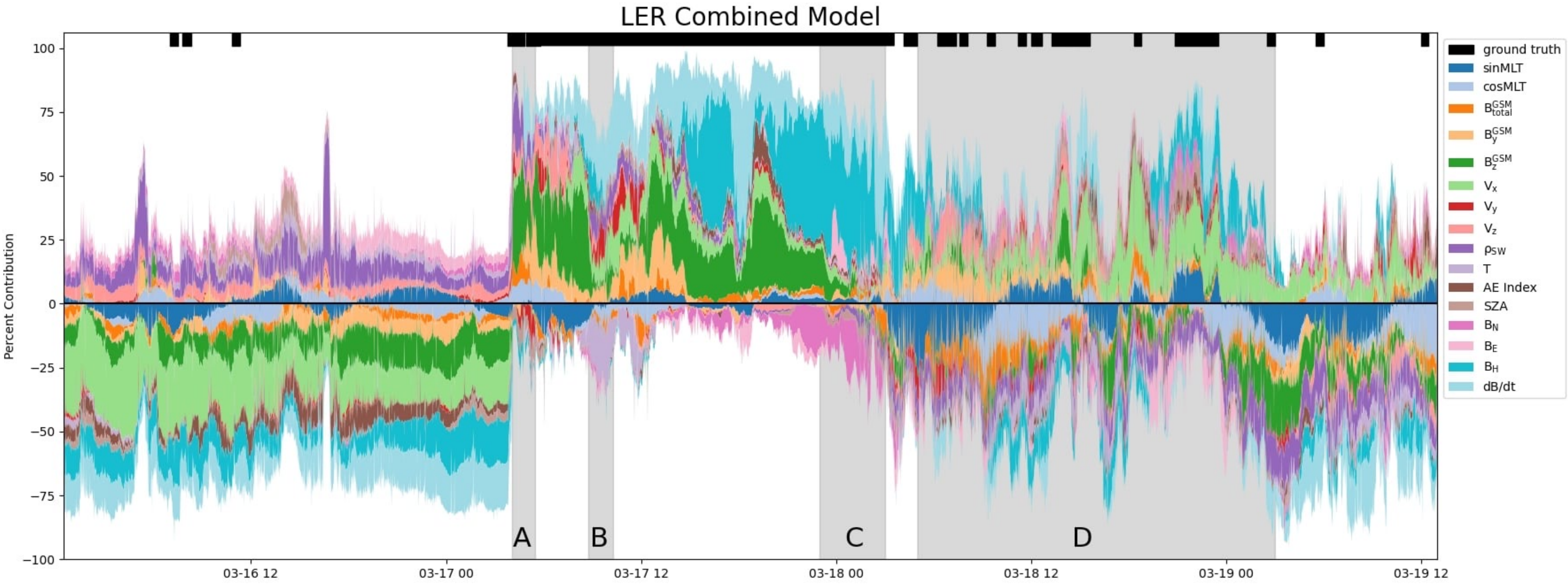
Percent Contribution



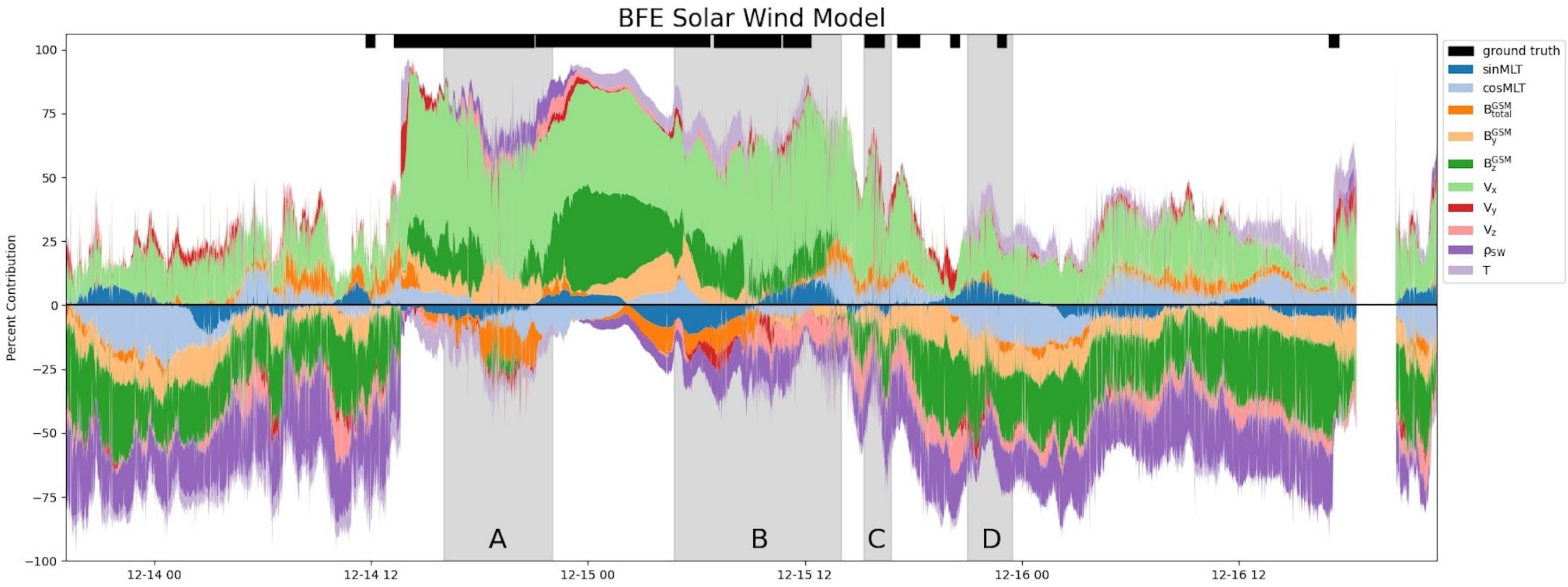
SHAP Stack Plots – March 2015 Storm



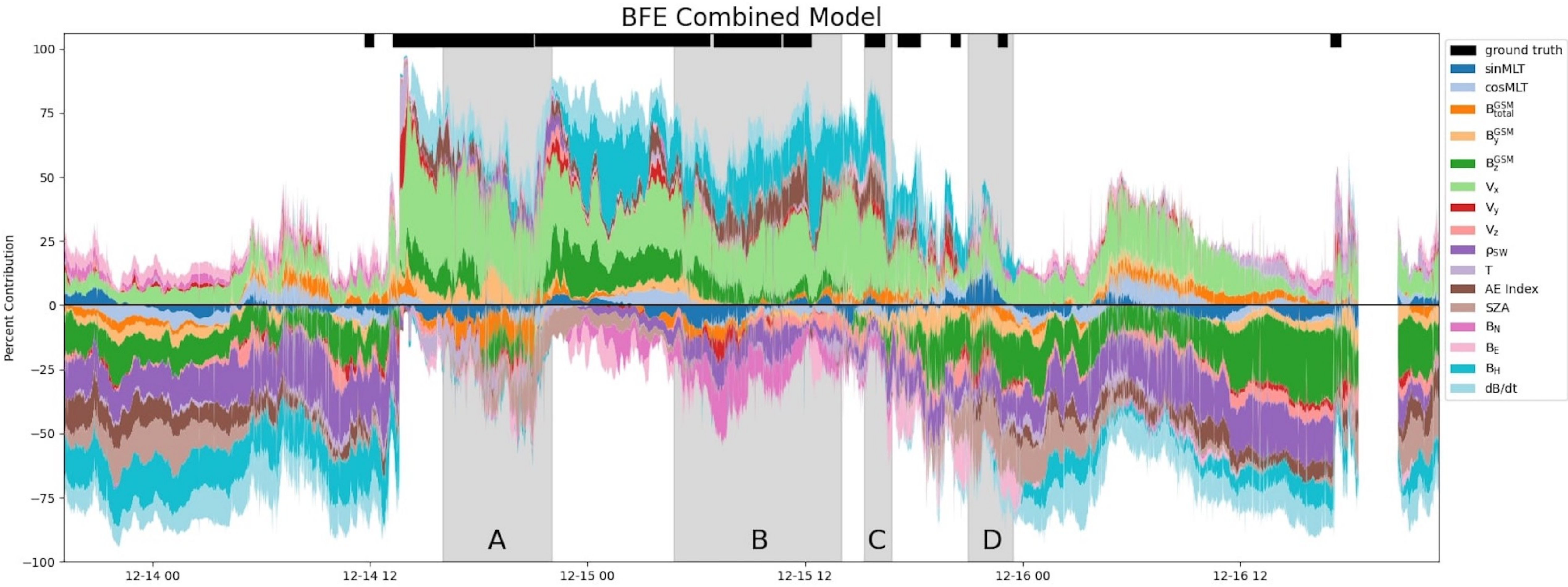
SHAP Stack Plots – March 2015 Storm

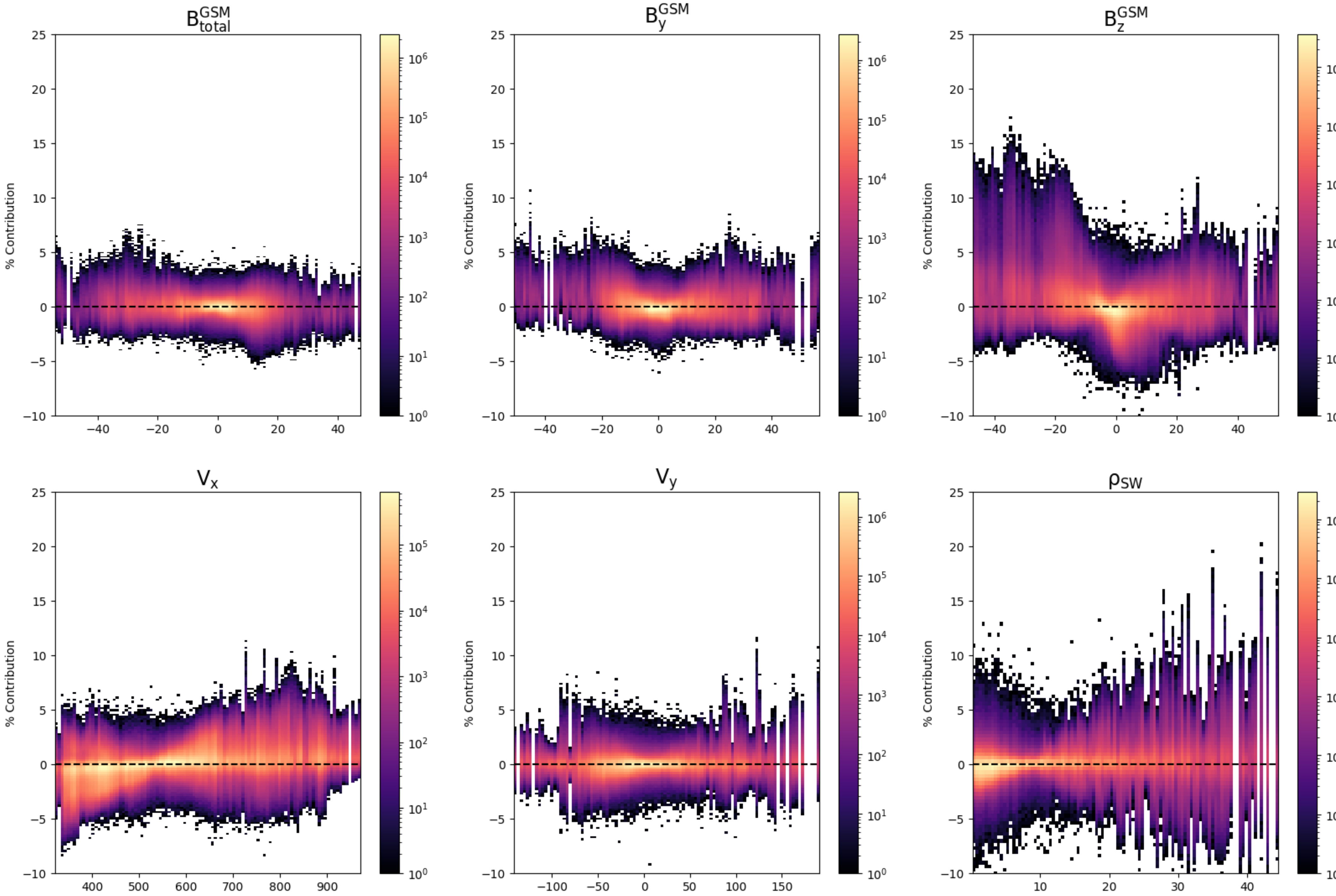


SHAP Stack Plots – December 2006 Storm

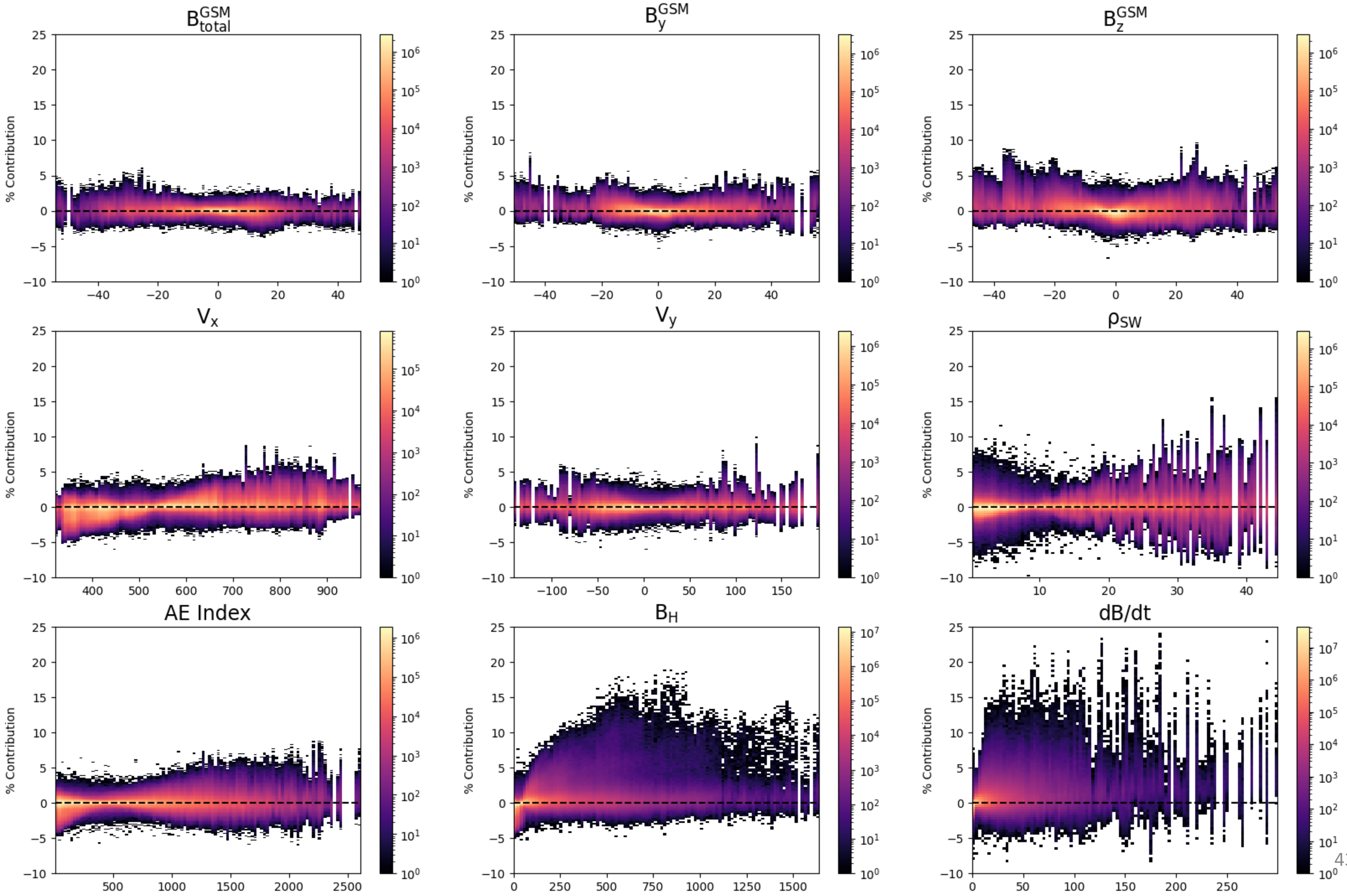


SHAP Stack Plots – December 2006 Storm





B_{total}^{GSM}
B_y^{GSM}
B_z^{GSM}
V_x
V_y
ρ_{SW}
% Contribution


B_{total}^{GSM}
B_y^{GSM}
B_z^{GSM}
V_x
V_y
ρ_{SW}
AE Index
B_H
dB/dt
% Contribution

Metrics Plots

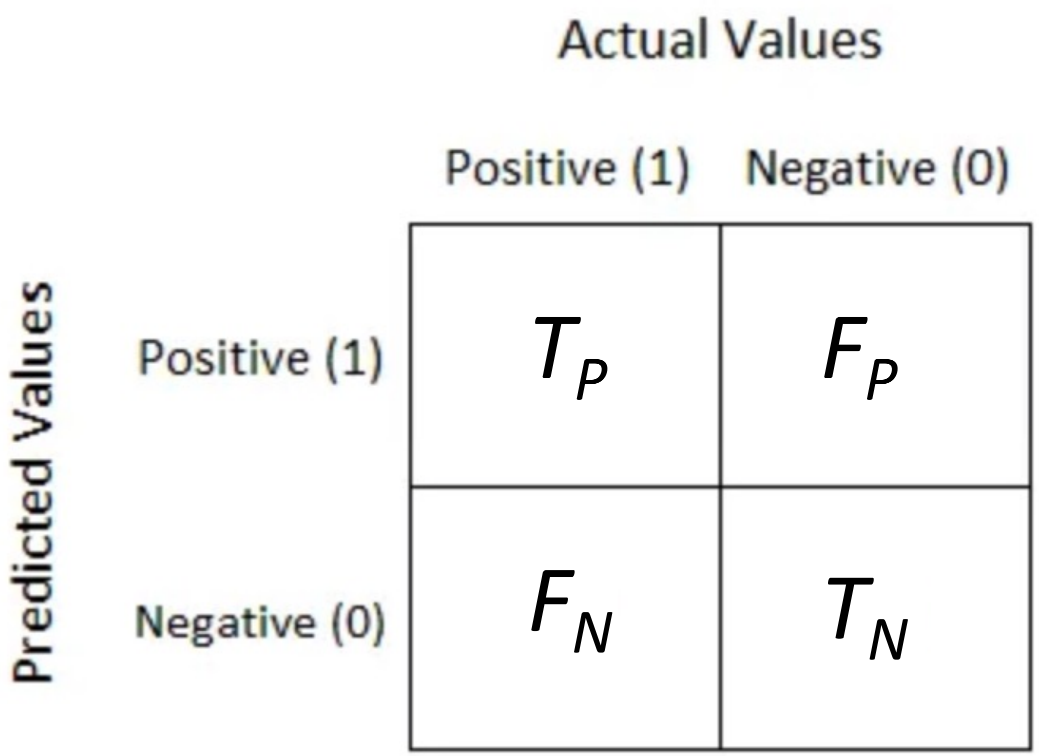


Image Credit: Sarang Narkhede – Towards Data Science

Metrics Plots

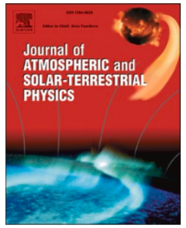
Journal of Atmospheric and Solar–Terrestrial Physics 218 (2021) 105624

Contents lists available at ScienceDirect

Journal of Atmospheric and Solar-Terrestrial Physics

journal homepage: www.elsevier.com/locate/jastp

ELSEVIER



Pre

Check for updates

RMSE is not enough: Guidelines to robust data-model comparisons for magnetospheric physics

Michael W. Liemohn [a,*], Alexander D. Shane [a], Abigail R. Azari [b], Alicia K. Petersen [a,c], Brian M. Swiger [a], Agnit Mukhopadhyay [a]

[a] *Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA*
[b] *Space Sciences Laboratory, University of California, Berkeley, CA, USA*
[c] *Now at Air Force Research Laboratory, Kirtland Air Force Base, Albuquerque, NM, USA*

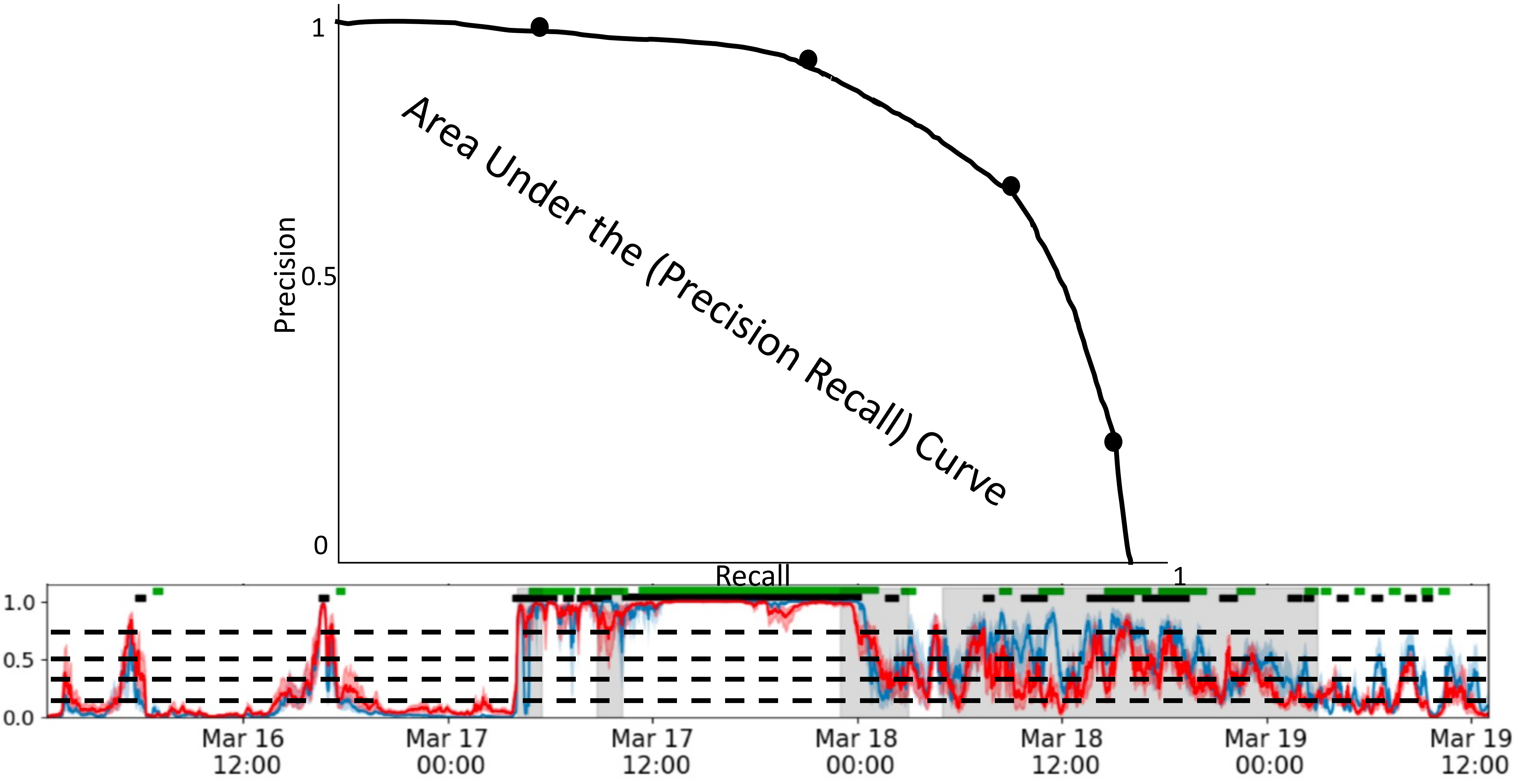
A R T I C L E I N F O

Keywords:
Data-model comparisons
Metrics
Fit performance
Event detection
Space weather
Magnetospheric physics
Forecasting

BS

A B S T R A C T

The magnetospheric physics research community uses a broad array of quantitative data-model comparison methods (metrics) when conducting their research investigations. It is often the case, though, that any particular study will only use one or two metrics, with the two most common being Pearson correlation coefficient and root mean square error (RMSE). Because metrics are designed to test a specific aspect of the data-model relationship, limiting the comparison to only one or two metrics reduces the physical insights that can be gleaned from the analysis, restricting the possible findings from modeling studies. Additional physical insights can be obtained when many types of metrics are applied. We organize metrics into two primary groups: 1) fit performance metrics, often based on the data-model value difference; and 2) event detection metrics, which use a discrete event classification of data and model values determined by a specified threshold. In addition to these groups, there are several major categories of metrics based on the aspect of the data-model relationship that the metric assesses: 1) accuracy; 2) bias; 3) precision; 4) association; 5) and extremes. Another category is skill, which is a measure of any of these metrics against the performance of a reference model. These can be applied to a subset of



Area Under the (Precision Recall) Curve
Precision
Recall
1
0.5
0
1
1.0
0.5
0.0
Mar 16
12:00
Mar 17
00:00
Mar 17
12:00
Mar 18
00:00
Mar 18
12:00
Mar 19
00:00
Mar 19
12:00

Metrics Plots

Unachievable Region in Precision-Recall Space and Its Effect on Empirical Evaluation

Kendrick Boyd,
University of Wisconsin – Madison, 1300 University Avenue, Madison, WI 53706 USA

Vítor Santos Costa,
CRACS INESC-TEC & FCUP, Rua do Campo Alegre, 1021/1055, 4169 - 007 PORTO, Portugal

Jesse Davis, and
KU Leuven, Celestijnenlaan 200a, Heverlee 3001, Belgium

C. David Page
University of Wisconsin – Madison, 1300 University Avenue, Madison, WI 53706 USA

Kendrick Boyd: boyd@cs.wisc.edu; Vítor Santos Costa: vsc@dcc.fc.up.pt; Jesse Davis: jesse.davis@cs.kuleuven.be; C. David Page: page@biostat.wisc.edu

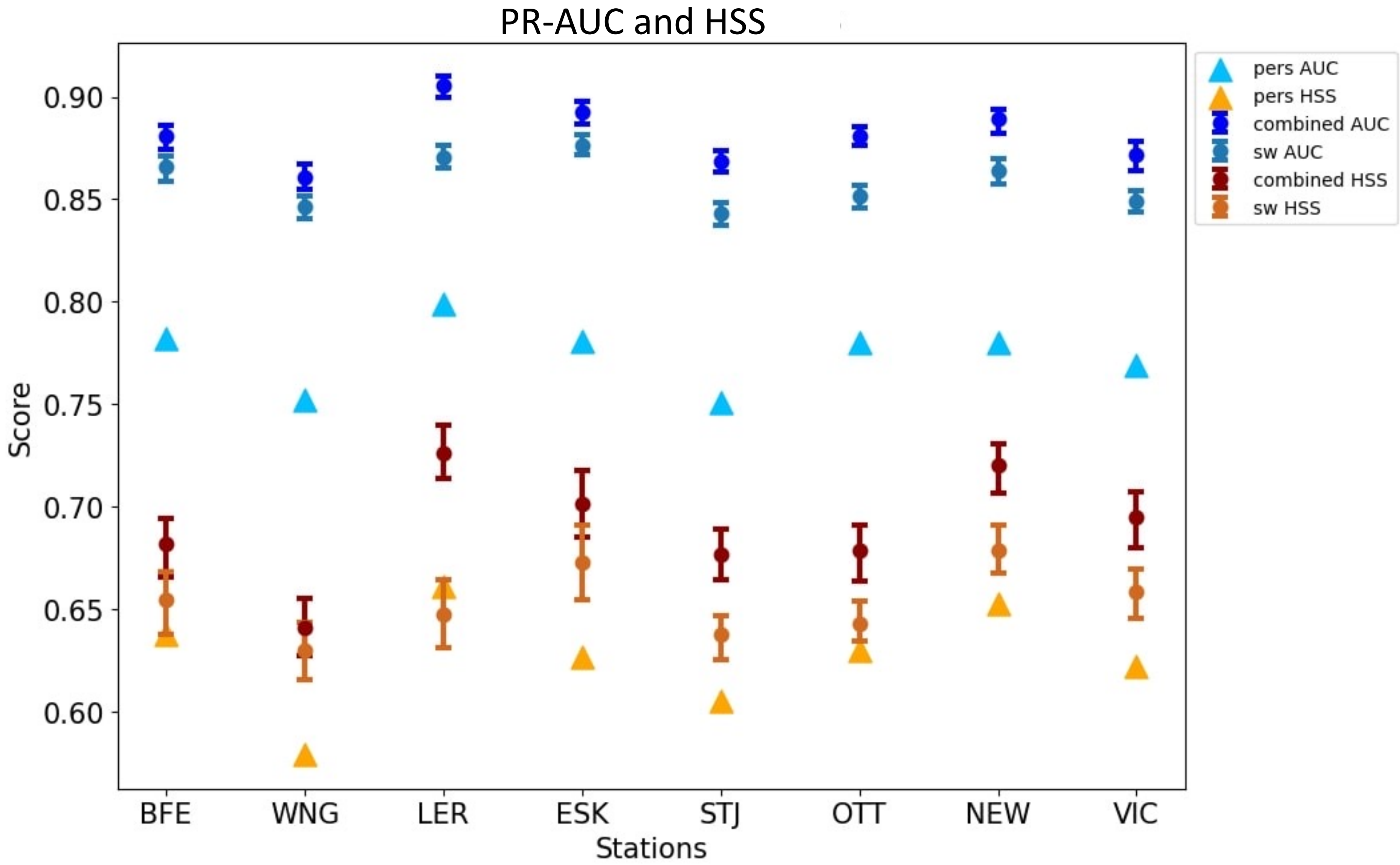
Abstract

Precision-recall (PR) curves and the areas under them are widely used to summarize machine learning results, especially for data sets exhibiting class skew. They are often used analogously to ROC curves and the area under ROC curves. It is known that PR curves vary as class skew changes. What was not recognized before this paper is that there is a region of PR space that is completely unachievable, and the size of this region depends only on the skew. This paper precisely characterizes the size of that region and discusses its implications for empirical evaluation methodology in machine learning.

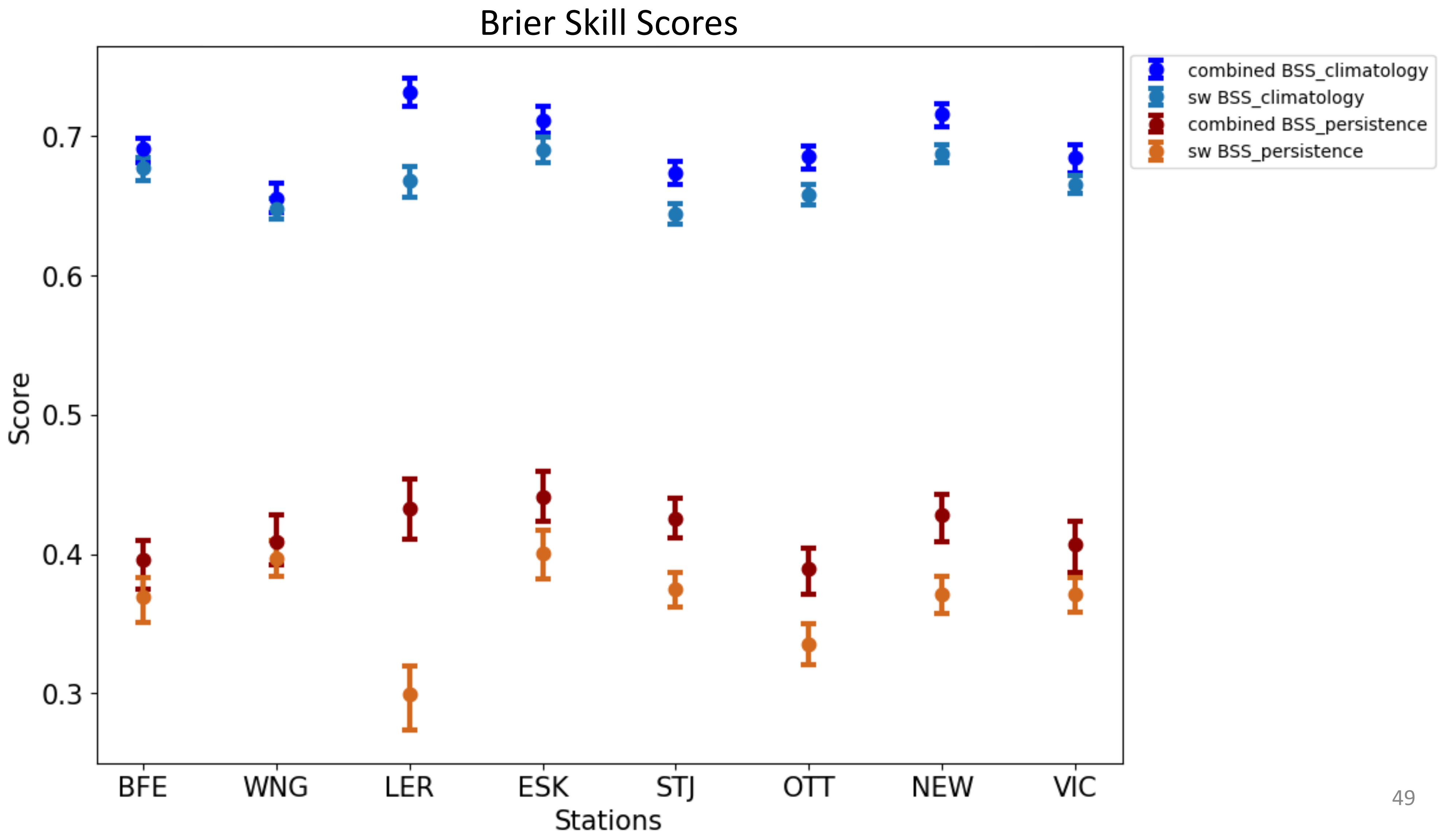
1. Introduction

Precision-recall (PR) curves are a common way to evaluate the performance of a machine learning algorithm. PR curves illustrate the tradeoff between the proportion of positively labeled examples that are truly positive (precision) as a function of the proportion of

Metrics Plots



Metrics Plots



Summary

- Rapid changes in the ground magnetic field can lead to Geomagnetically Induced Currents.
- The localized nature of dB/dt can lead to an issue with forecasts that only use solar wind data.
- Machine Learning models can do probabilistic forecasting of dB/dt threshold crossings.
- The SHAP method allows us to open the neural network black-box and gain insight into model decision making.
- Ground magnetic field data enables better predictions in later parts of the storm.
- Models conform to our current understanding of the solar wind – magnetosphere system.
- Models have good metric scores, outperforming persistence and climatology.

Future work

- Incorporate TWINS ion temperature maps.
- Predict RSD directly.
- Focus on known localization events.

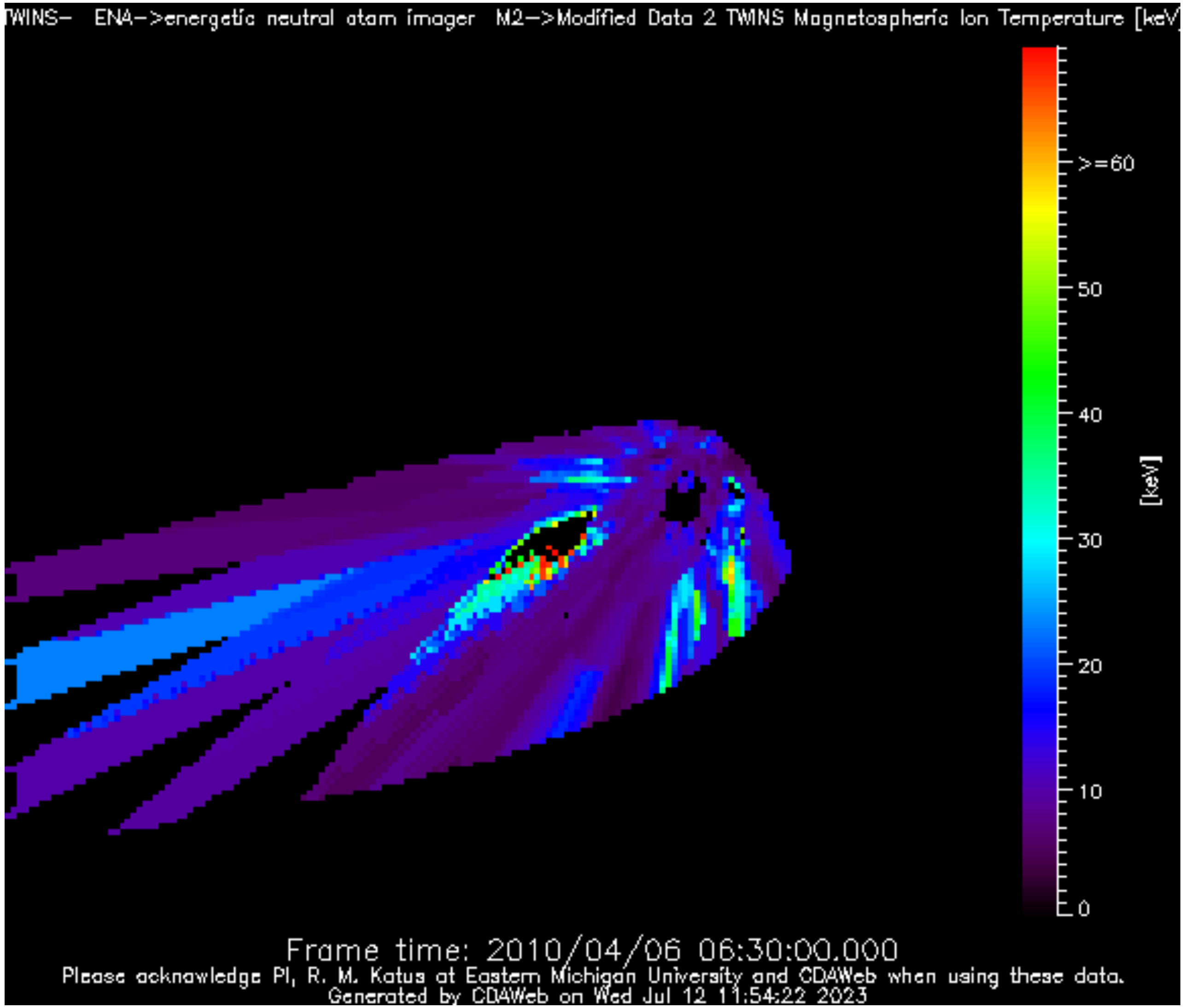


Image Credit: NASA SPDF

Publication in Space Weather

ADVANCING EARTH AND SPACE SCIENCE

Space Weather

RESEARCH ARTICLE

10.1029/2023SW003446

Special Section:
Machine Learning in Heliophysics

Key Points:

- Machine learning based probabilistic predictions can provide risk forecasts of extreme *dB*/*dt* events at mid-latitude magnetometer stations
- Models utilizing both solar wind and past ground magnetometer data outperform solar wind only and *dB*/*dt* based persistence models
- Input feature analysis shows solar wind velocity and IMF B_Z^{GSM} strongly influence model output, demonstrating alignment with current understanding

Supporting Information:

Supporting Information may be found in the online version of this article.

Probabilistic Forecasting of Ground Magnetic Perturbation Spikes at Mid-Latitude Stations

Michael Coughlan[1,2], Amy Keesee[1,2], Victor Pinto[3,4], Raman Mukundan[1,2], José Paulo Marchezi[1,2], Jeremiah Johnson[5], Hyunju Connor[6], and Don Hampton[7]

[1]Department of Physics & Astronomy, University of New Hampshire, Durham, NH, USA, [2]Institute for the Study of Earth, Oceans, & Space, University of New Hampshire, Durham, NH, USA, [3]Departamento de Fisica, Universidad de Santiago de Chile, Santiago, Chile, [4]Center for Interdisciplinary Research in Astrophysics and Space Exploration (CIRAS), Universidad de Santiago de Chile, Santiago, Chile, [5]Department of Electrical & Computer Engineering, University of New Hampshire, Manchester, NH, USA, [6]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [7]Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA

Abstract The prediction of large fluctuations in the ground magnetic field (*dB*/*dt*) is essential for preventing damage from Geomagnetically Induced Currents. Directly forecasting these fluctuations has proven difficult, but accurately determining the risk of extreme events can allow for the worst of the damage to be prevented. Here we trained Convolutional Neural Network models for eight mid-latitude magnetometers to predict the probability that *dB*/*dt* will exceed the 99th percentile threshold 30–60 min in the future. Two model frameworks were compared, a model trained using solar wind data from the Advanced Composition Explorer (ACE) satellite, and another model trained on both ACE and SuperMAG ground magnetometer data. The models were compared to examine if the addition of current ground magnetometer data significantly improved

Thank you! Questions?

Acknowledgements

This work was supported by NSF EPSCoR Award OIA-1920965. We thank all members of the MAGICIAN team at UNH and UAF that participated in the discussions leading to this article. We also thank the OMNIWeb, SuperMAG and ACE teams for providing the data. Additionally, the authors would like to thank Mayowa Adewuyi for discussions about model output visualizations.

GitHub Link

github.com/mikecoughlan/multi-station-dbdt-risk-assessment/