Data provided/archived by
DoD and NOAA

An unsupervised learning approach to superstorm signature identification in precipitating particle data

Hannah Marlowe, PhD
Principal Data Scientist
Amazon Web Services

Janet Kozyra, PhD
Program Scientist
NASA HQ Heliophysics Division

Delores J. Knipp, PhD
Research Professor
Smead Aerospace Engineering Sciences Department
University of Colorado, Boulder

Liam Kilcommons
Research Software Engineer
University of Colorado, Boulder

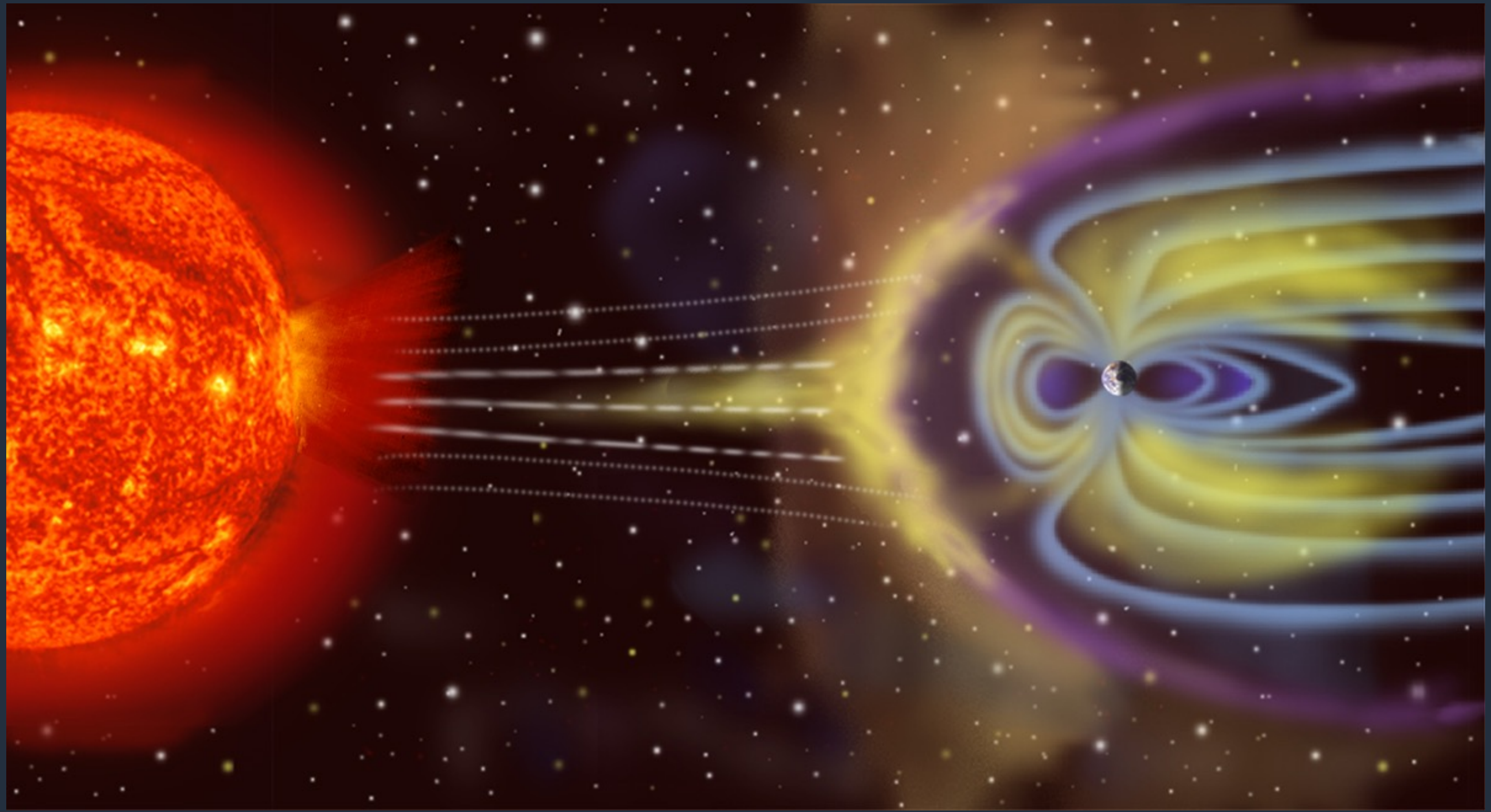
Related content!

Attend Delores Knipp's session for "Part II":

Geophysical interpretations from machine learning superstorm signature identification in satellite precipitating particle data

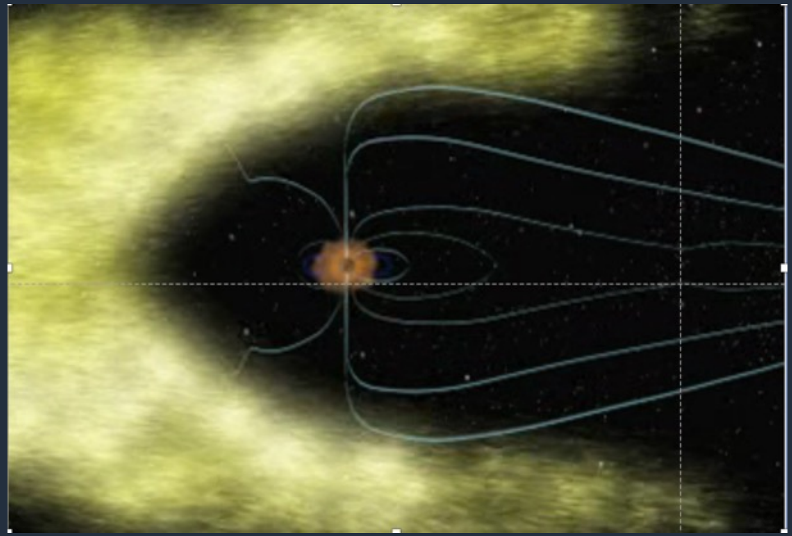
🕚 Friday, March 25 at 11:00-11:20AM

What is a Storm in Space?

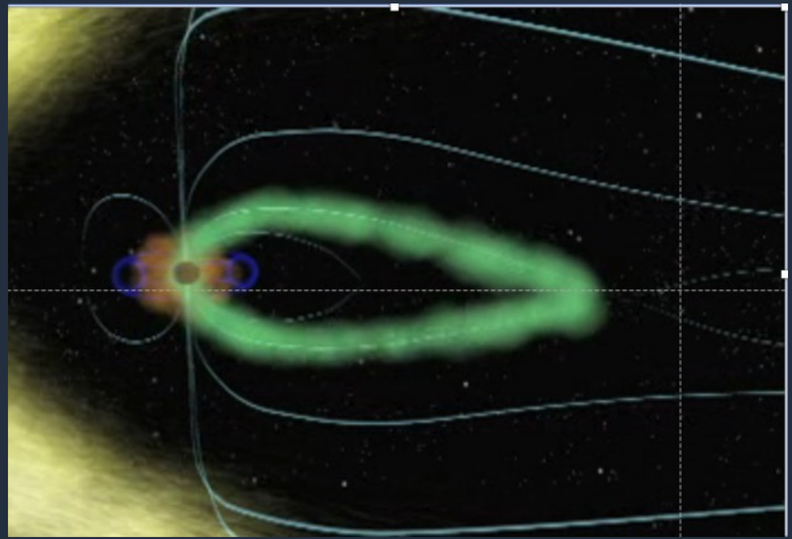


NASA artist's rendition of a space storm

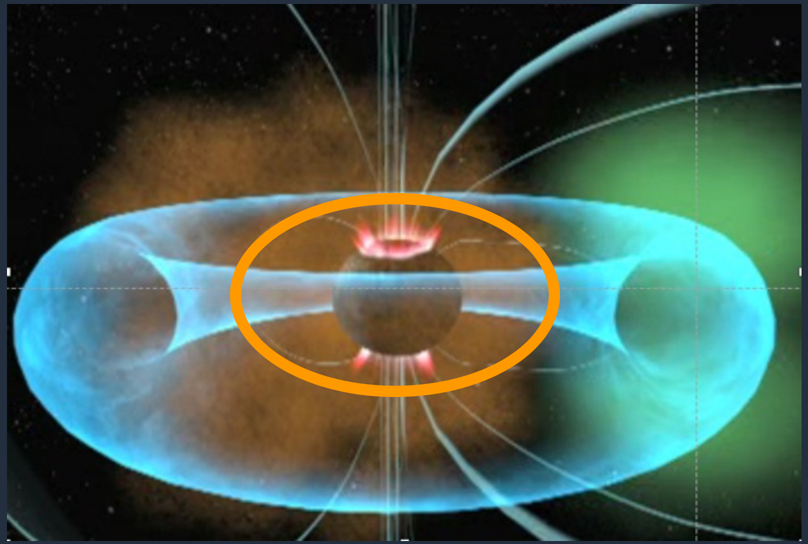
- High energy ions (green & brown) drift in a torus around the Earth producing a current which depresses the surface magnetic field in an amount proportional to the total energy of the radiation.
- A commonly used measure of storm intensity is called the Dst index



Solar eruption arrives at Earth. Dayside magnetic field lines become interconnected with those of solar wind. Energy and plasma enter

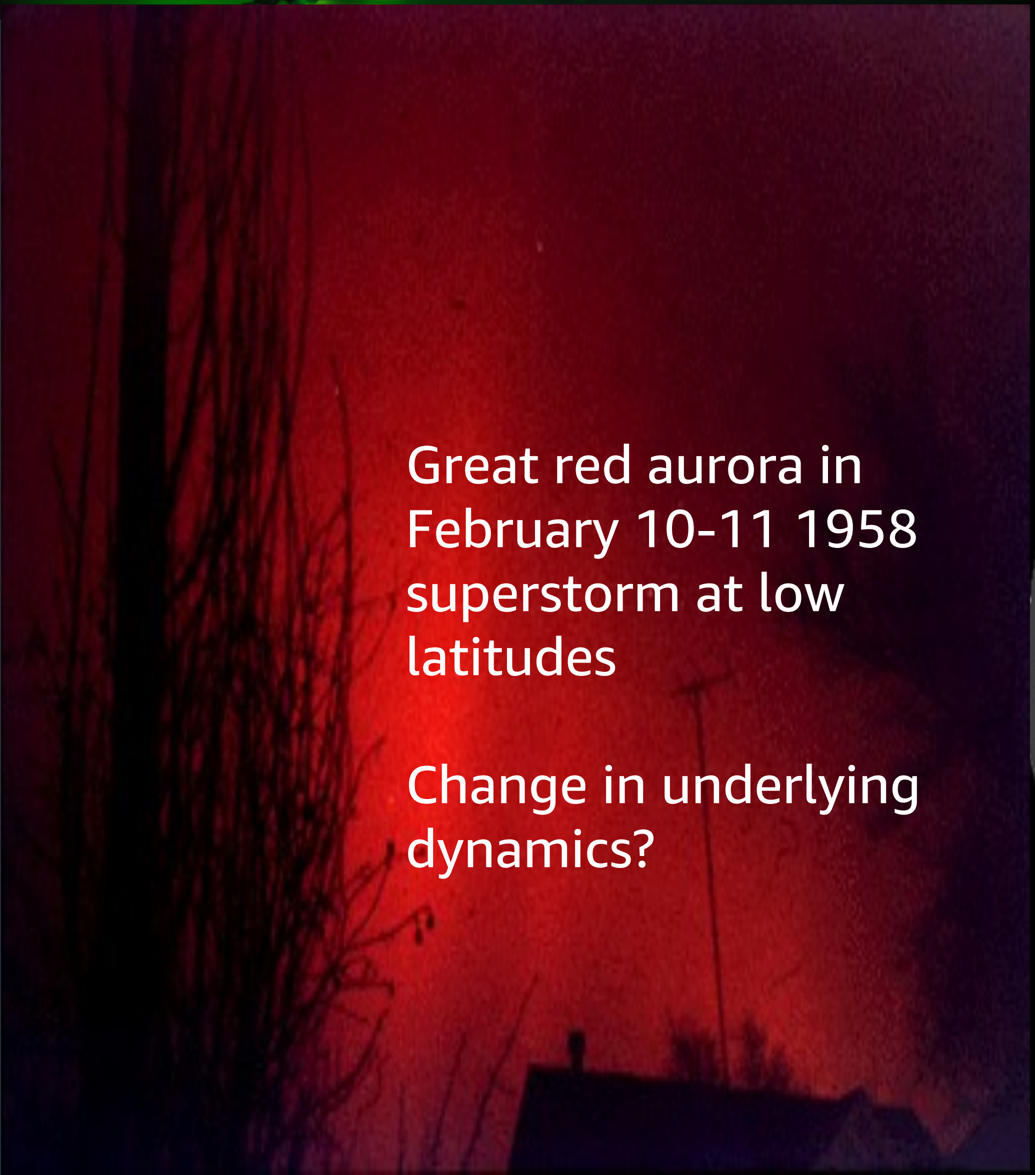


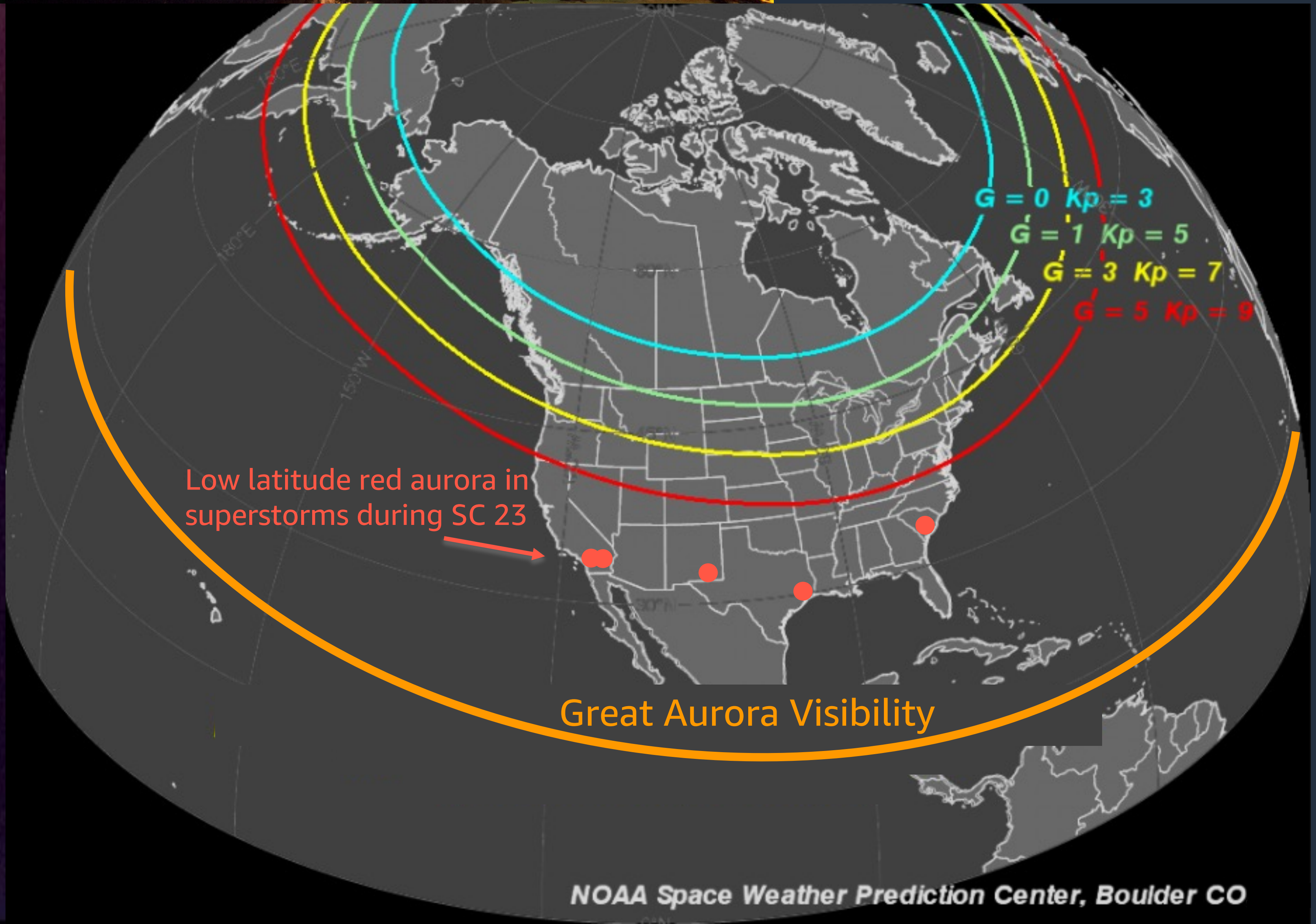
Oxygen ions (green) heated in the atmosphere flow out into geospace & are further accelerated intensifying the storm



Electrons from the magnetotail precipitate (not shown) creating auroras

Anomalous features of storms



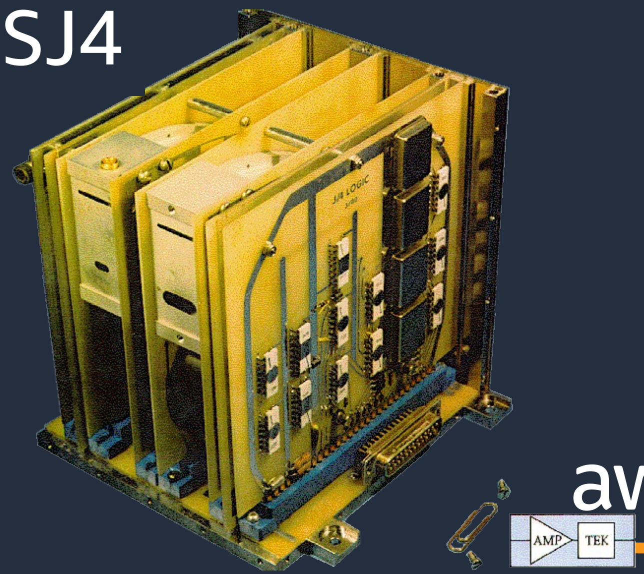


Defense Meteorological Satellite Program (DMSP)

- **Operational Air Force spacecraft (850 km)**
  - 3-4 spacecraft in operation at any one time
- **Archive of space weather data (since 1980s)**
  - Data not originally intended for science use
- **Auroral Particle Instrument (SSJ4) since 1990's**
  - Measures electrons and ions (protons) entering the atmosphere from magnetosphere along geomagnetic field lines
  - Can detect particles with kinetic energies between 100 eV to 30,000 eV

https://www.ospo.noaa.gov/Operations/DMSP/index.html

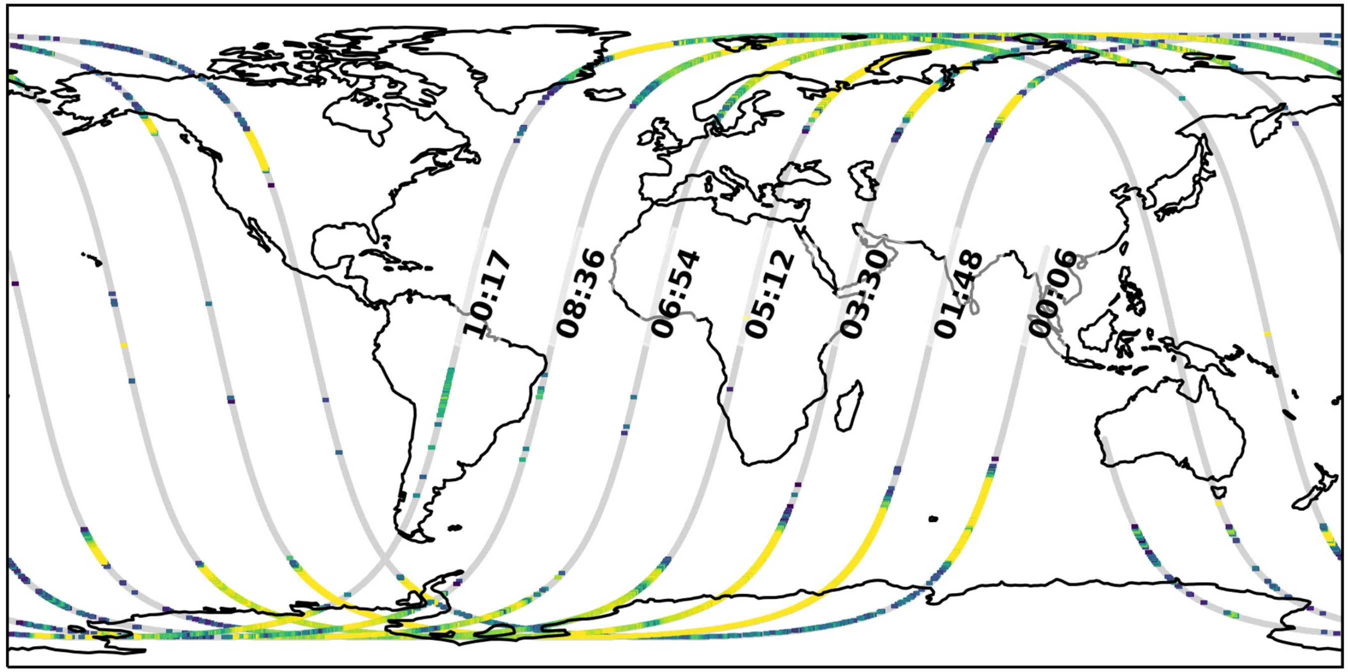
SSJ4



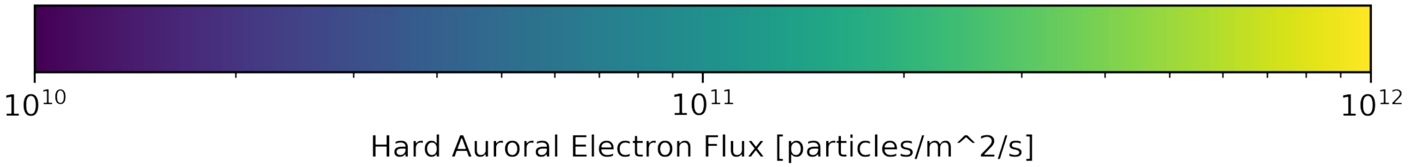
aws

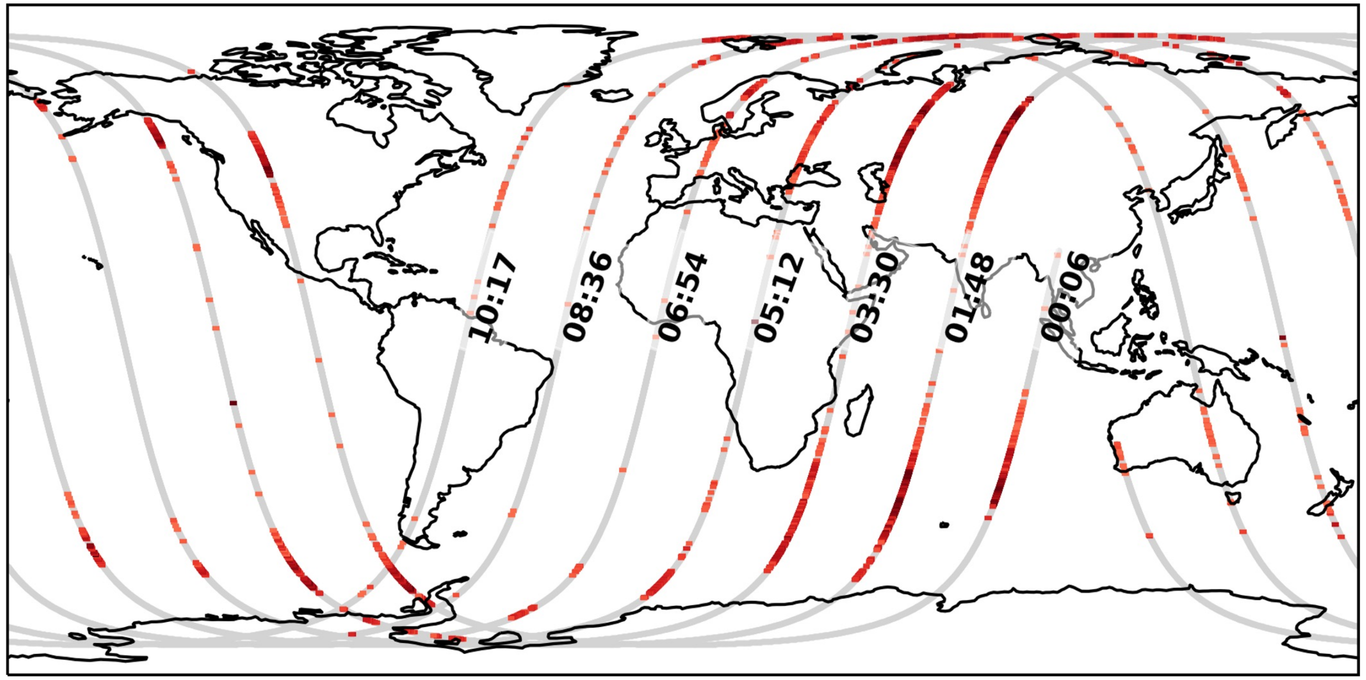
Particle Precipitation During a Superstorm:

Electron precipitation at higher latitudes than Ion precipitation

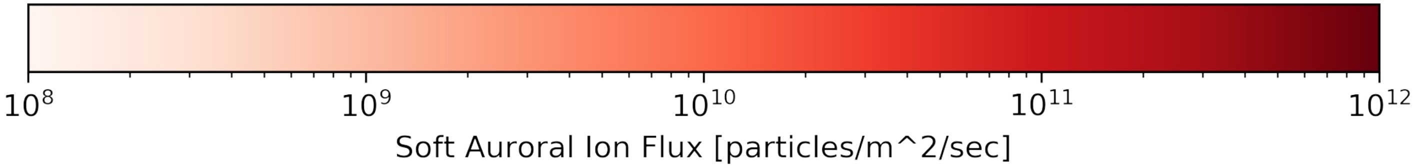


Creates 'usual' green aurora, but more equatorward





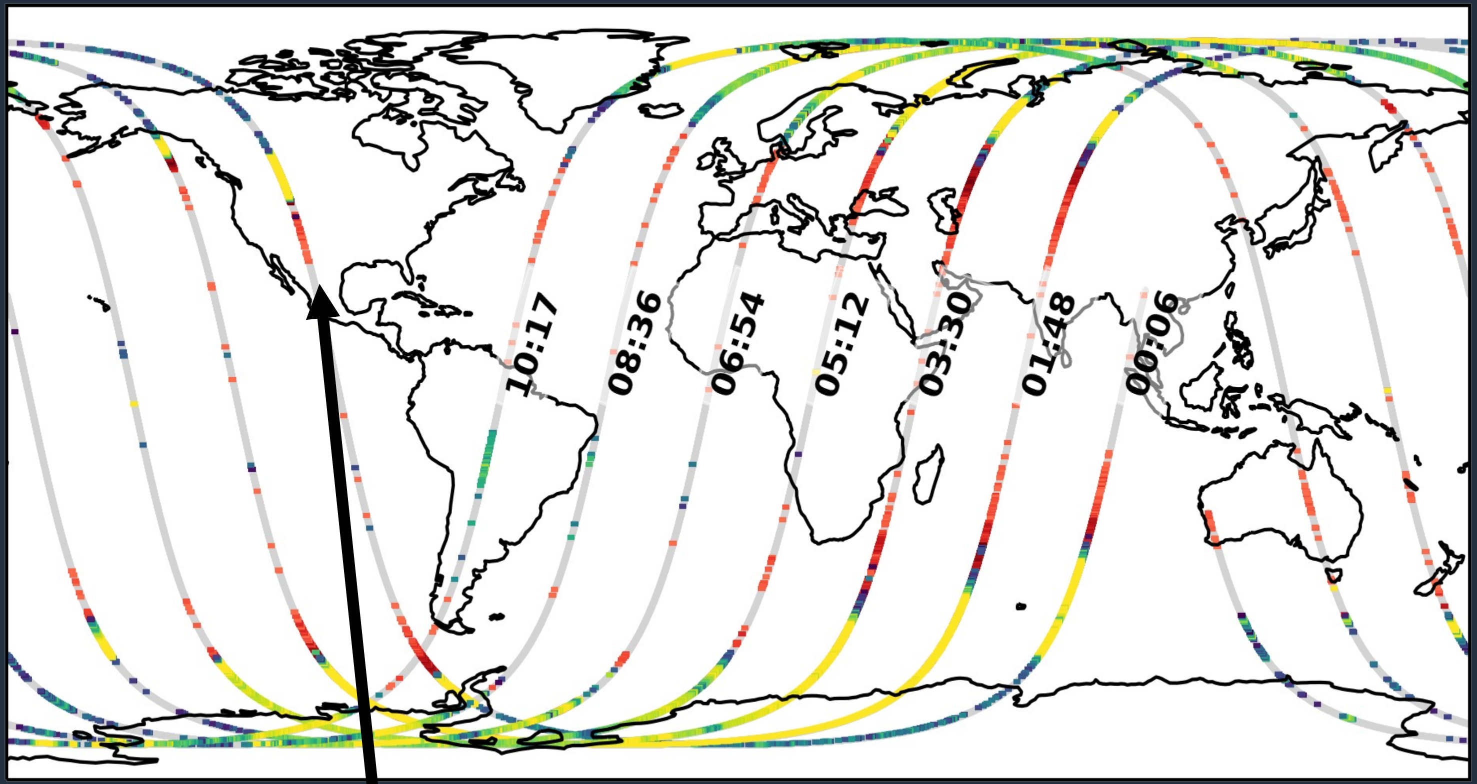
Related to (may not directly cause) 'unusual' aurora, SAR arcs, red aurora



DMSP F13 2003-10-30

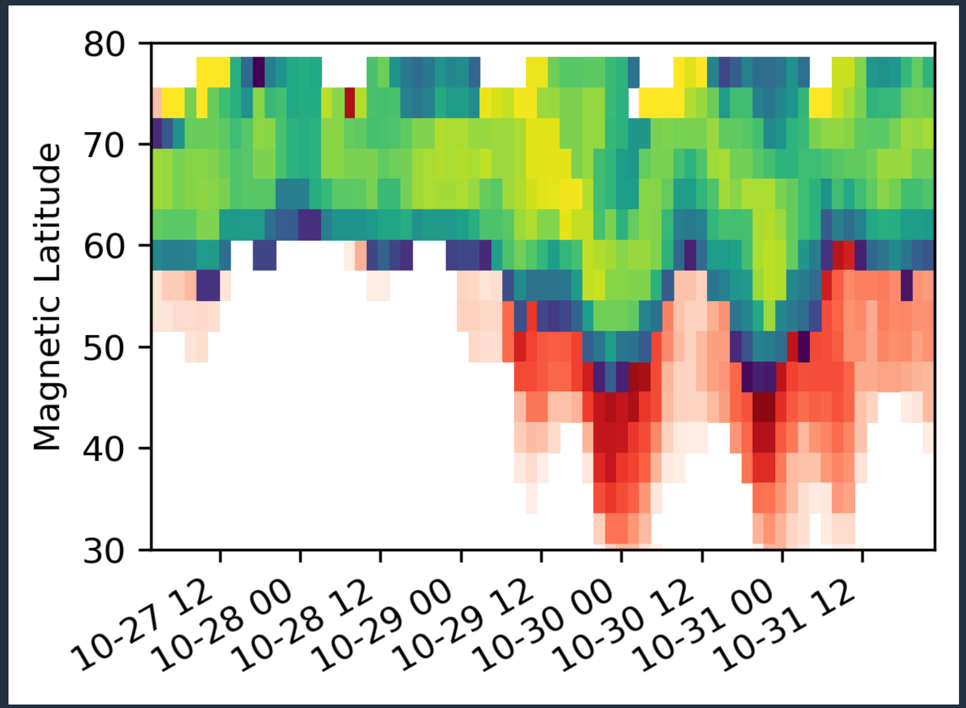
- Specifically look at hard electrons and soft ions

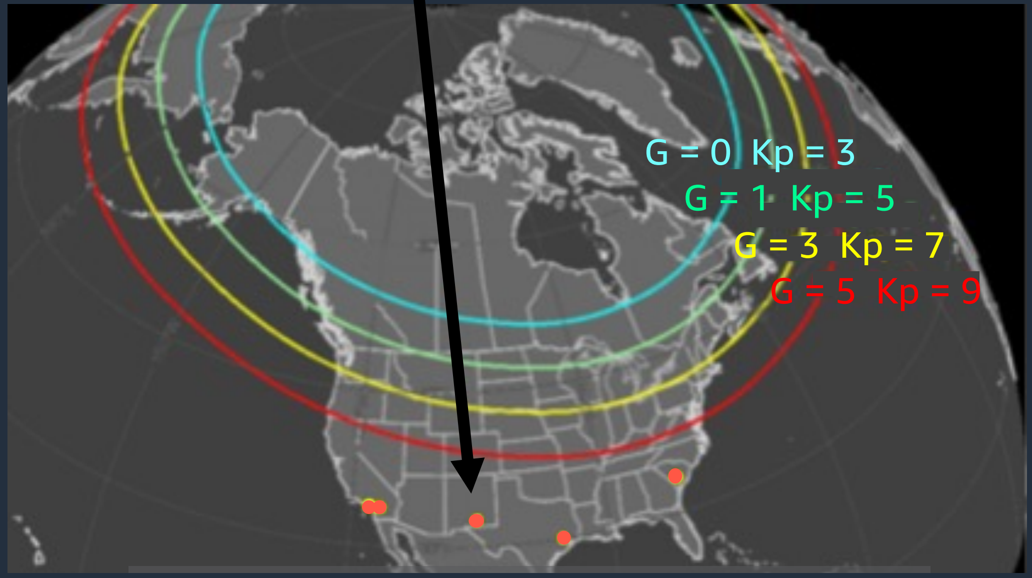
Ion precipitation reaches lower latitudes than Electron precipitation during superstorms



19 channels, of ions and electrons per second (millions / day)

Each orbit becomes one time step



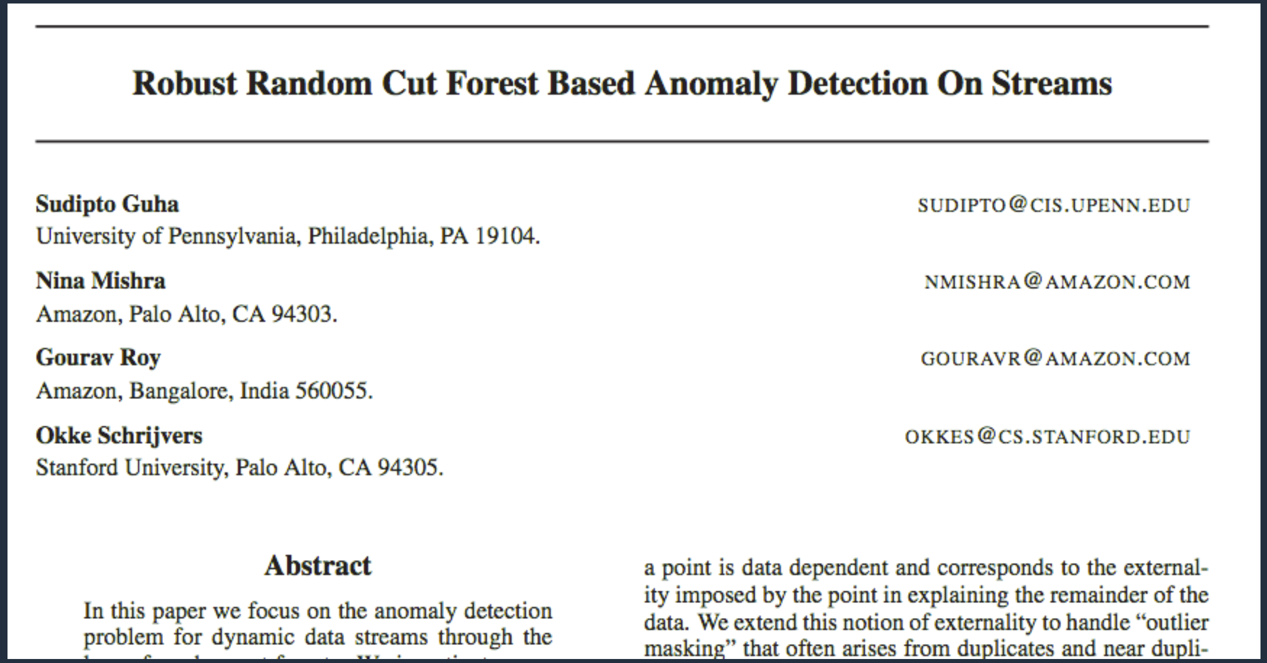


Ion precipitation reaches as far south as red aurora observations (red dots)

- **Need ML-ready data**
  - Impractical to use full data stream
  - Remove detector anomalies and nuisance signals—median filtering
  - Careful data reduction (averaging/filtering) to preserve signals of interest

Categorizing storms – unsupervised anomaly detection

- Intense storms are rare events and we expect this signal to be anomalous compared to 'fair weather' conditions
- Example: tree-based methods
    - E.g., Robust Random Cut Forest and Isolation Forest
    - Optimized for outlier isolation versus methods that focus on profiling the normal distribution
    - Binary trees are constructed from the data through cuts along data dimensions
        - Outliers will be close to the root
    - Suitable for high-dimensional data
        - RRCF scales well with respect to number of features, data set size, and number of training compute instances



Robust Random Cut Forest Based Anomaly Detection On Streams

Sudipto Guha SUDIPTO@CIS.UPENN.EDU
University of Pennsylvania, Philadelphia, PA 19104.

Nina Mishra NMISHRA@AMAZON.COM
Amazon, Palo Alto, CA 94303.

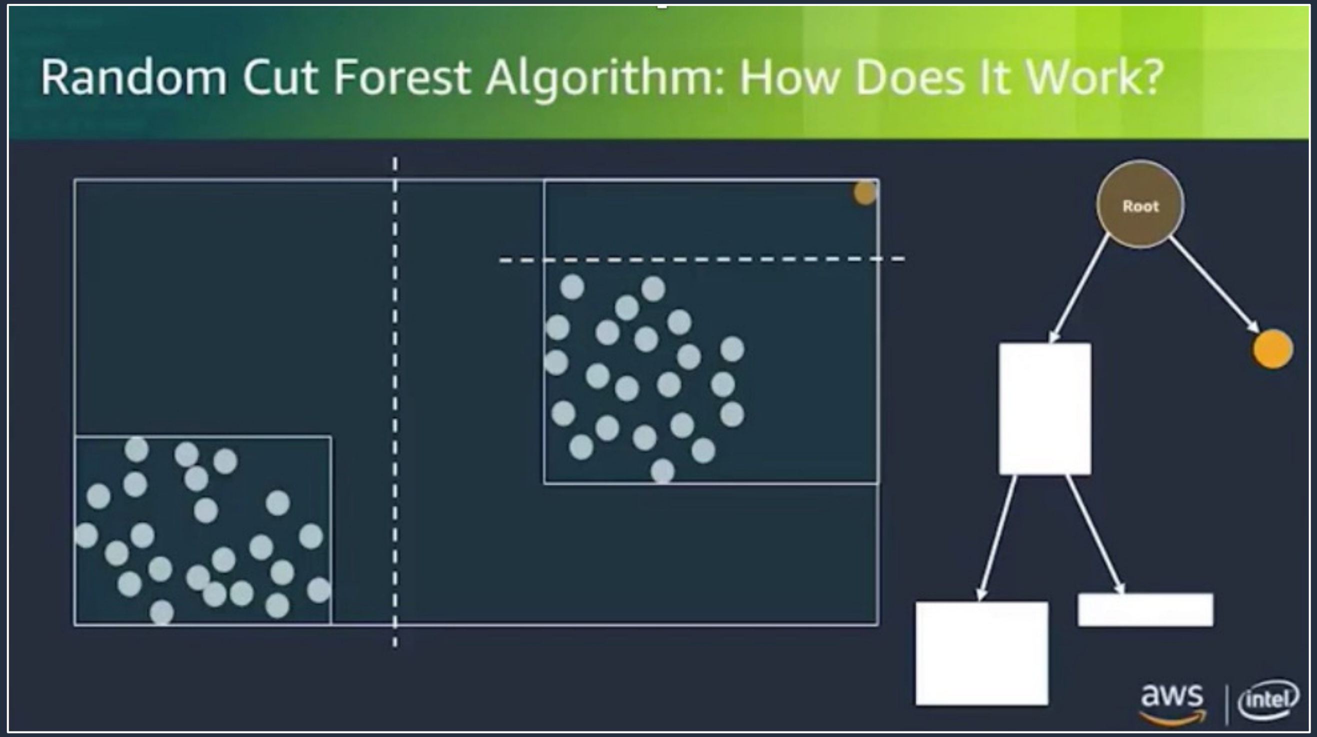
Gourav Roy GOURAVR@AMAZON.COM
Amazon, Bangalore, India 560055.

Okke Schrijvers OKKES@CS.STANFORD.EDU
Stanford University, Palo Alto, CA 94305.

Abstract

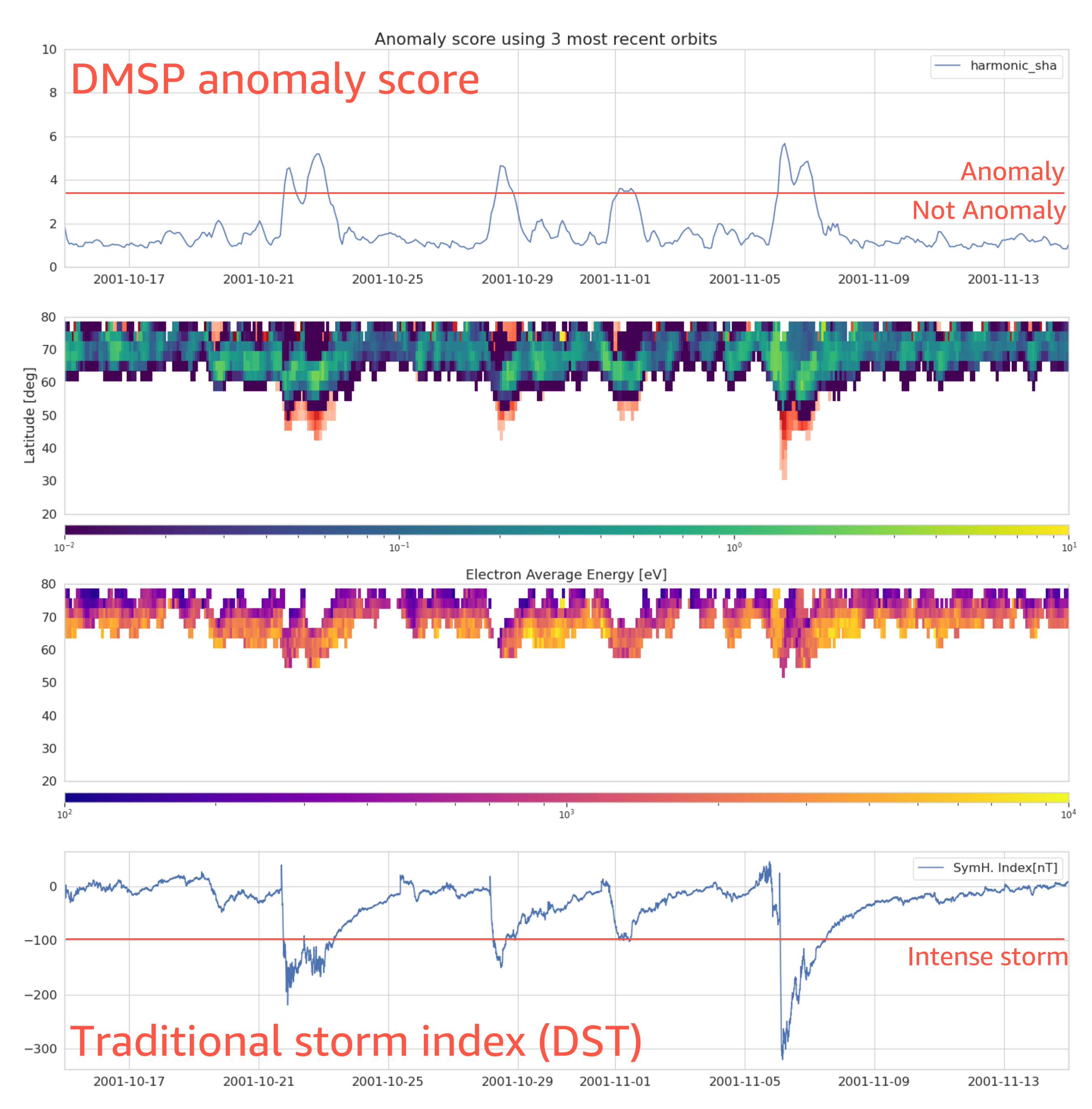
In this paper we focus on the anomaly detection problem for dynamic data streams through the

a point is data dependent and corresponds to the externality imposed by the point in explaining the remainder of the data. We extend this notion of externality to handle "outlier masking" that often arises from duplicates and near dupli-



Anomaly detection: Results

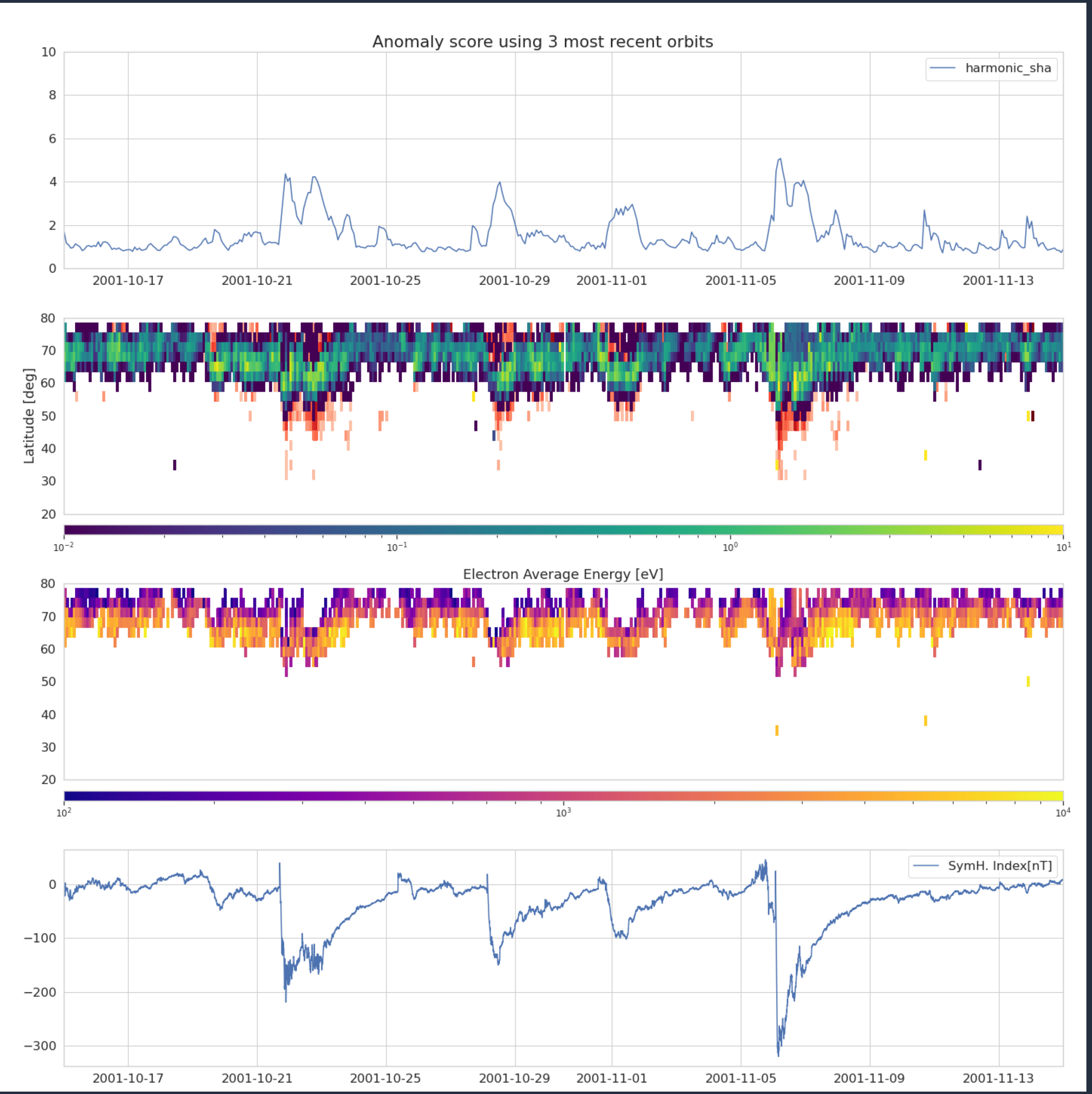
- Trained RCF model using 2000-2002 DMSP data
  - Input includes soft ion flux (red), hard electron flux (green/blue) and average electron energy (bottom plot)
  - Included time-lags (past 3 orbits)
- Model outputs continuous anomaly score value for every data point
  - Set threshold of 3std above mean as “anomaly”



Typical cutoff of DST<-100 nT for intense storms

Types of anomalies

Model detects physical and also detector anomalies



*Note that this is an operational satellite and wasn't meant to provide science input

Comparing Particle Anomalies to Storm Indices

Interested in how particle signals relates to commonly used solar storm indicators

- Compared particle anomalies to the symmetric ring current H index (SYMH)
- "Storm" defined if storm index, SYMH, goes below threshold
- "Anomaly" defined if anomaly score goes above 3 std threshold

Note – this is an unsupervised model trained only with DMSP data, not the comparison SYMH

True Positive (TP) = Anomaly within +/-12 hr of storm

False Positive (FP) = Anomaly with no storm in +/-12 hr

False Negative (FN) = Storm with no anomaly

Evaluation metric: F1 score

F1 score is the harmonic mean of precision and recall

Useful for imbalanced problems where accuracy is less helpful

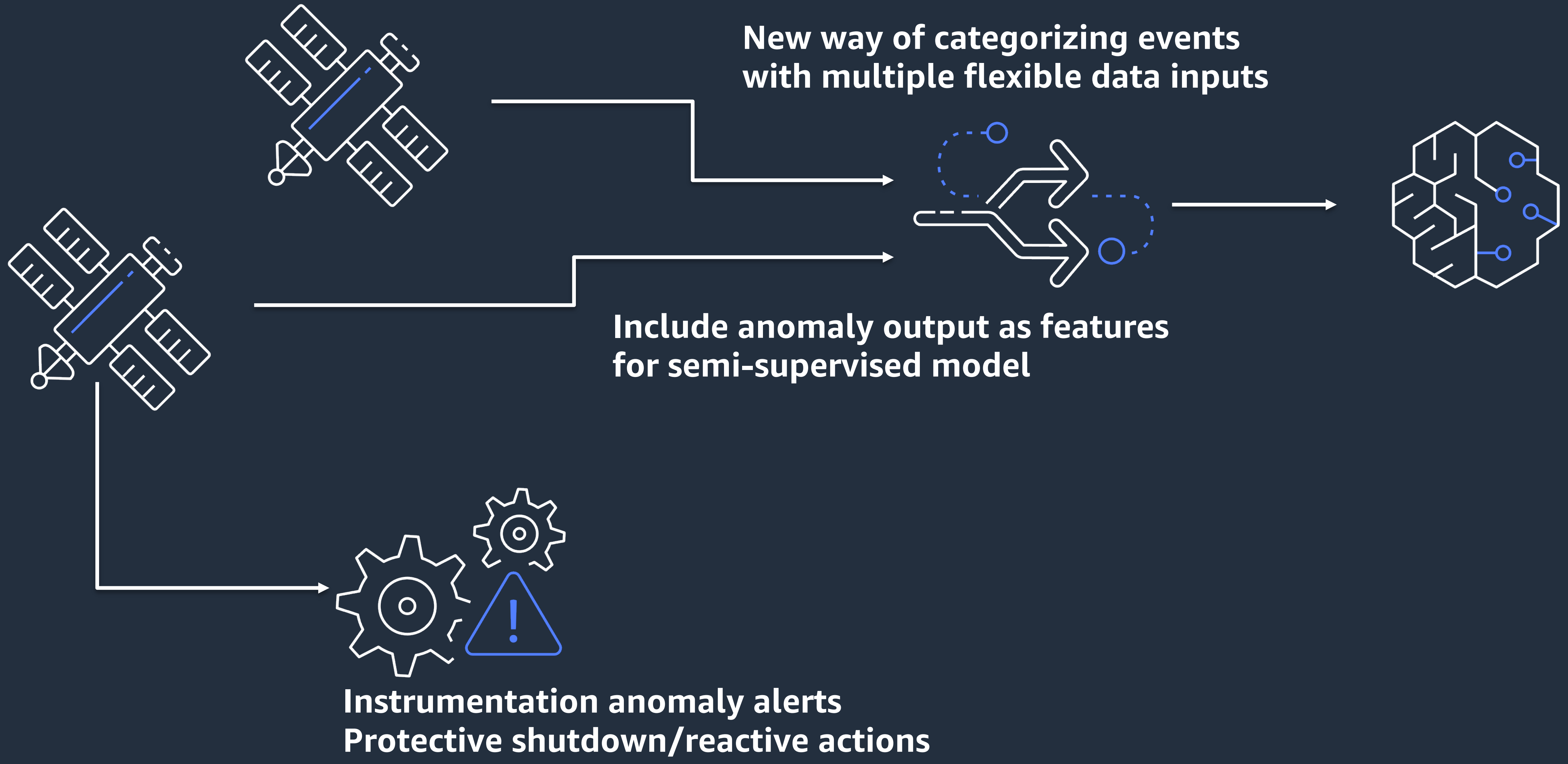
F1 score = 2TP / (2TP + FP + FN)

	Storm	Non-storm
Anomaly	42 (TP)	19 (FP)
Non-Anomaly	8 (FN)	

F1 score = 0.76

Events of interest also at edge between anomalous/non-anomalous and storm/superstorm

Interpretation and next steps: other science applications



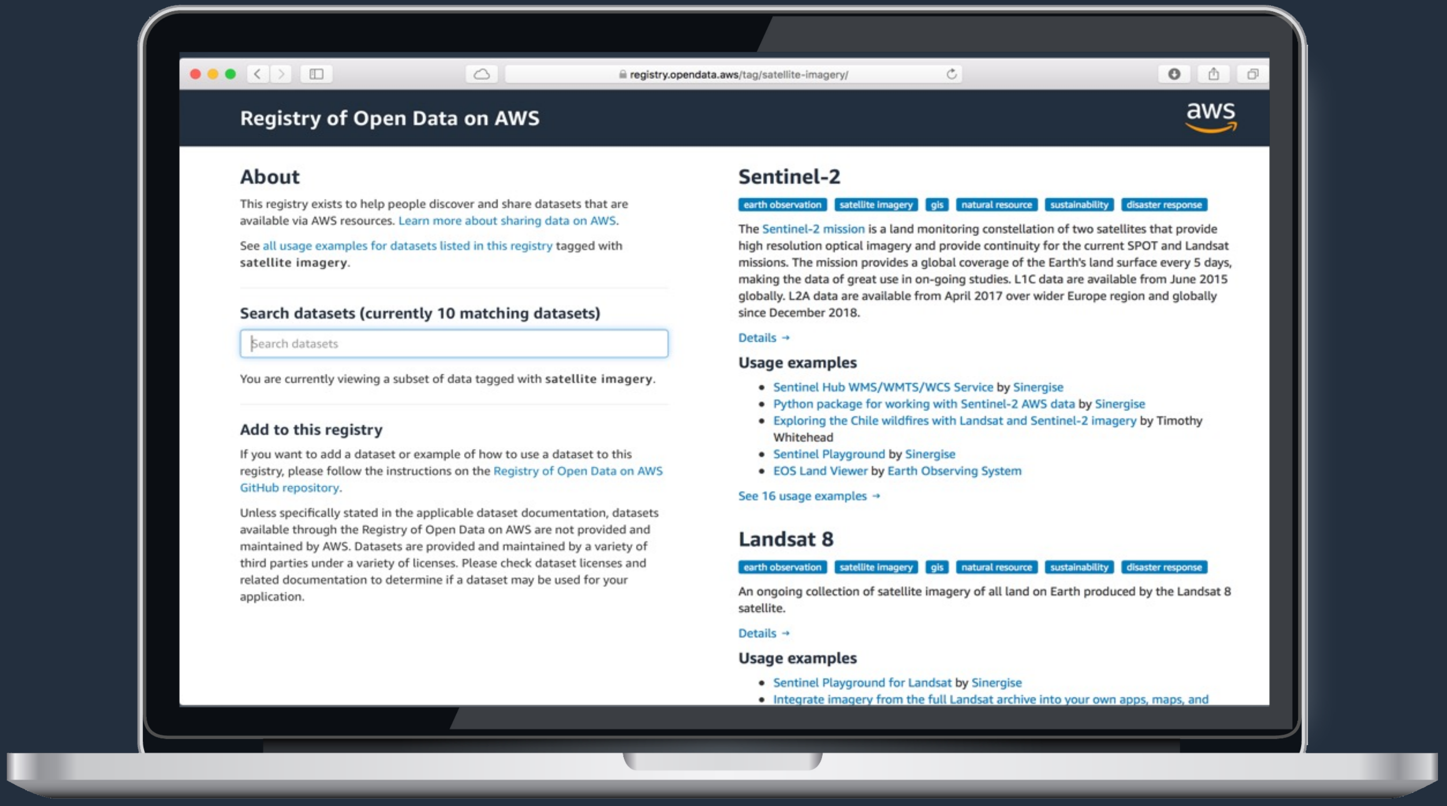
Access to the data and code

Raw CDF DMSP data is currently available

https://cdaweb.gsfc.nasa.gov/pub/data/dmsp/dmspf13/ssj/precipitating-electrons-ions/

Raw and ML-ready data with example model code also be available through the Registry of Open Data on AWS in the coming weeks

https://registry.opendata.aws/



Thank You!

marloweh@amazon.com