Harnessing expressive capacity of Machine Learning modeling to represent complex coupling of Earth's auroral space weather regimes

Jack Ziegler

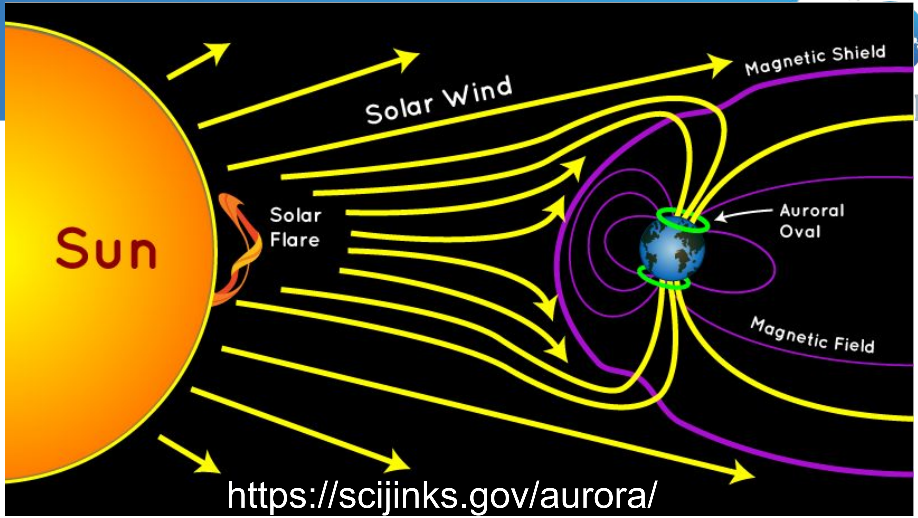
Ryan M. McGranaghan

ASTRA
(Atmospheric & Space Technology Research Associates)

ICMLA 2021

jziegler@astraspace.net
rmcgranaghan@astraspace.net

Introduction



Goal:

- Model Coupling of Solar Wind Energy with Magnetosphere and Upper Atmosphere
    - Specifically NowCast of Auroral Particles

Key Contributions:

1. Advancing the state-of-the-art ML applied to auroral particle precipitation
2. Assessment of the efficacy of increasing ML model expressive capacity (loss function engineering, multi-task learning, and transforming time series prediction to spatio-temporal prediction)
3. New lessons learned to advance ML research through application to novel science data.

Challenges:

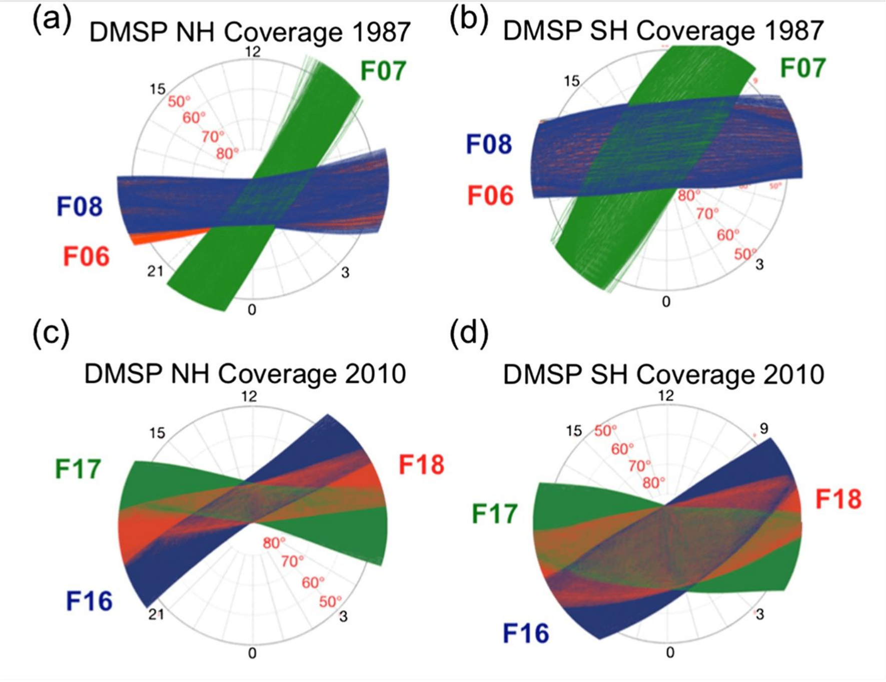
- Deconvolving spatial and temporal responses to inputs
- Extreme event predictions of rare space weather storms
- Evaluating 2D spatial performance of model
- Designing loss functions / metrics prioritizing large fluxes
- Predicting mesoscale variations and structures

Data

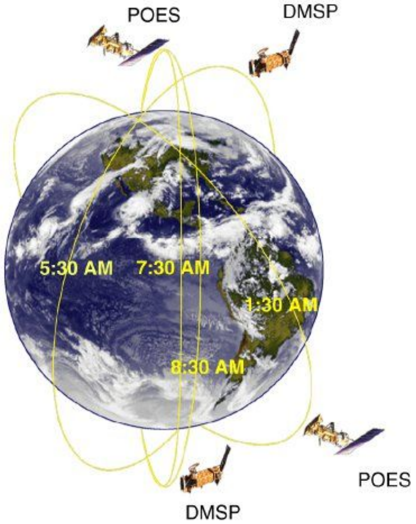
Our models use Defense Meteorological Satellite Program (DMSP) data dating back to 1987. DMSP spacecraft carry a particle detector (SSJ) which is sensitive to particles with characteristic energies between 30 eV and 30 keV.

Example Polar plots that each have three satellites, illustrating their coverages over different years.

McGranaghan, “Determining global ionospheric conductivity …”, 2016

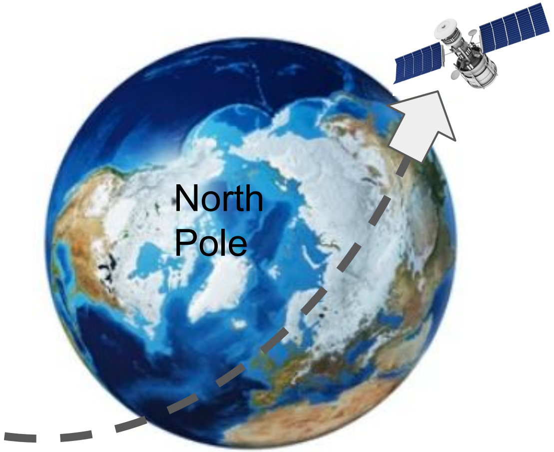


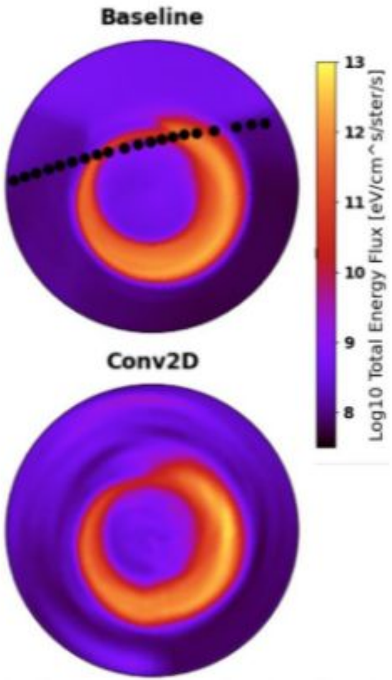
Polar Orbits



Extensibility to Similar Space Weather Problems

- Black Dots represent Measurements
- Colors represent unknown full representation





Four new areas of the ML algorithm search space

1) Multi-task learning models

2) 'Tail of the distribution'-specific custom loss functions

3) Sample distribution weighted loss function

4) Novel 2D, inverse convolutional (Conv2D) model for sparse spatial-temporal training data

- **Inputs:** *Global* solar wind observations and geomagnetic indices at five-min. Cadence

https://omniweb.gsfc.nasa.gov/OMNIWeb

(Temporal Sun/Earth Environment)

Target Output: Total electron energy flux at any given location (nowcast), MLat and MLT (Magnetic Local Time (zone))

("Aurora Strength" Spatial & Temporal)

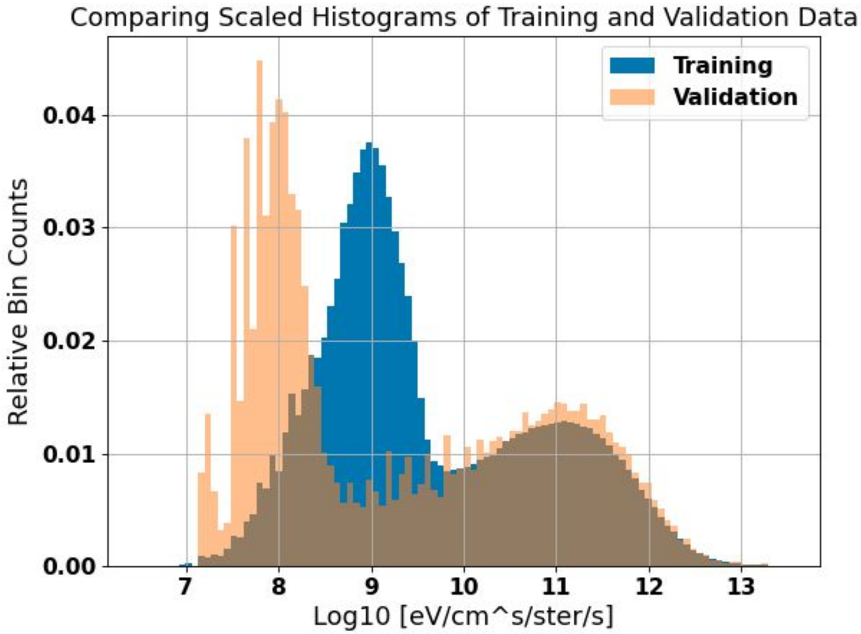
https://www.ngdc.noaa.gov/stp/satellite/dmsp/

Multidimensional Regression Problem

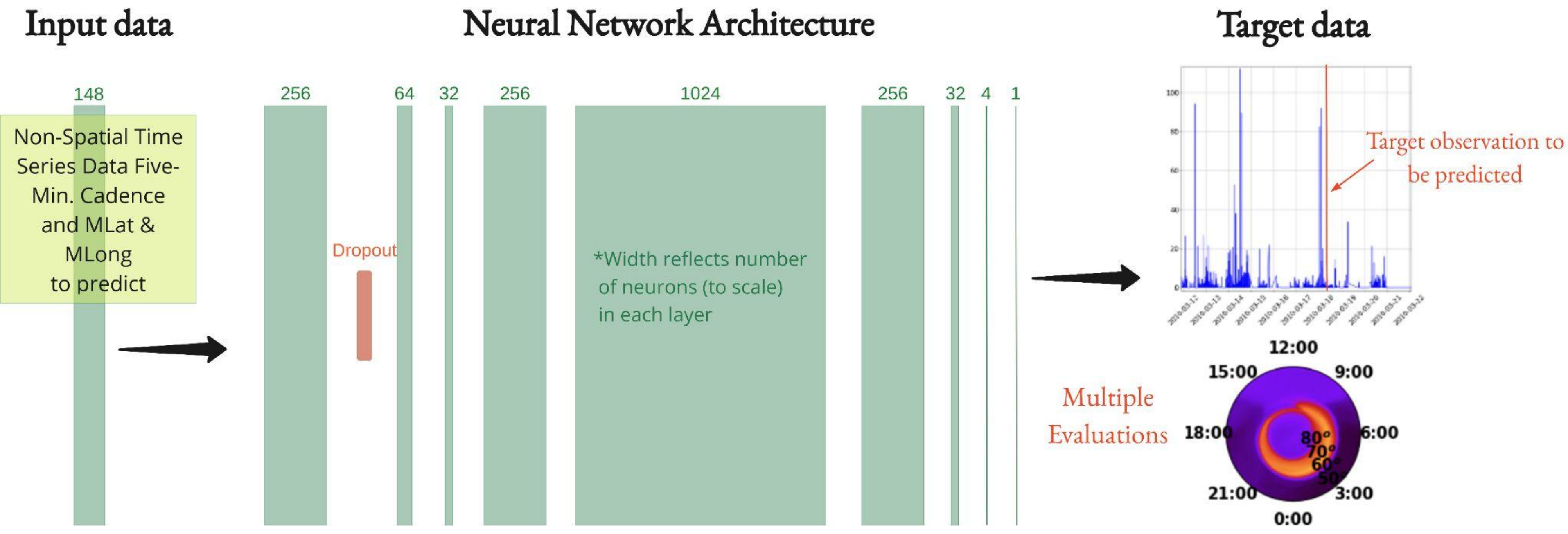
Data Challenges

- The electron energy fluxes [eV/cm 2 /s] vary over six orders of magnitude and require a log transformation.
- Imbalanced data: Large fluxes tend to occur during storms, so our target variable exhibits a large right "tail".
- Target points composed of multiple "modes" of skewed Gaussian-like distributions
- Input at 5 min Cadence, target at 1 sec cadence (translates to spatial resolution)

log10 scale transformation on Target



Baseline Model: PrecipNet

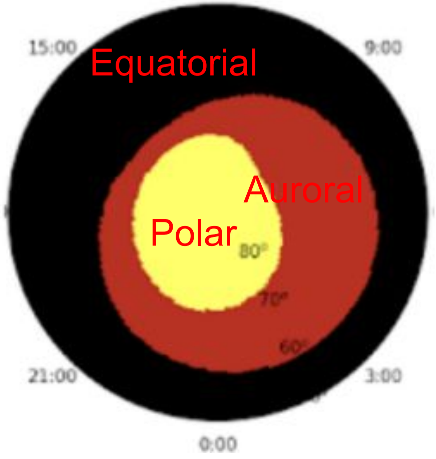


Auroral Region Model: Combining MSE and Classification

- Different Physics in Different regions
- Different Accuracies

$$\text{Loss} = \text{Mean}\Big(\big(y_{\text{flux-true}} - y_{\text{flux-pred}}\big)^2\Big) + \text{C.C.L.}\big(y_{\text{class-true}}, y_{\text{cclass-pred}}\big)$$

C.C.L. represents the categorical crossentropy loss



Region	Train samples	Test samples	Test MSE
Equatorial	925990	26821	0.45
Auroral	496715	14856	0.55
Polar	415578	13533	0.84

Training with Custom Loss Functions

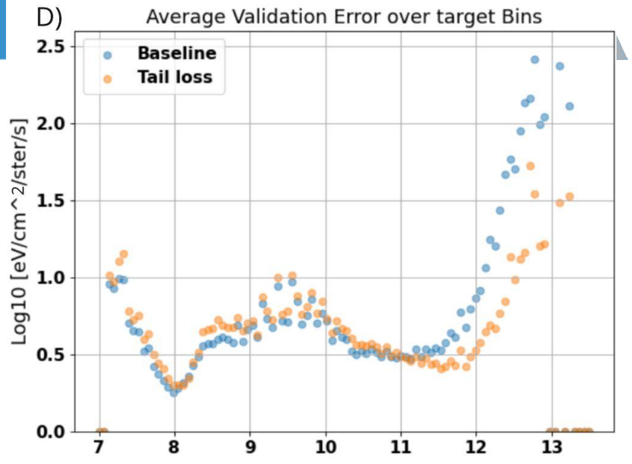
- logic-based function only activates in Right Side “tail region”

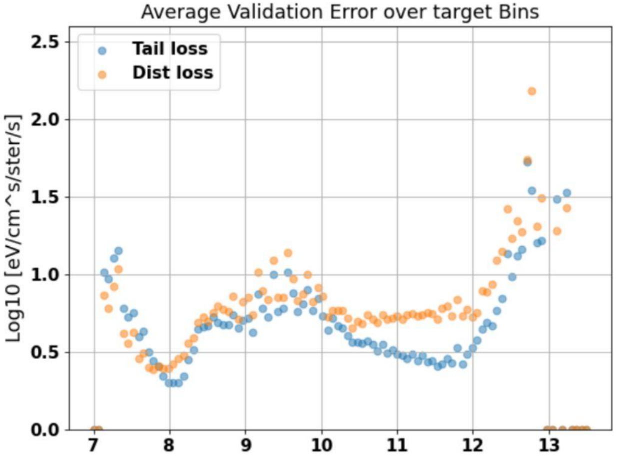
$$\text{Loss} = (y_{\text{true}} - y_{\text{pred}})^2(1 + a)$$
$$\text{for } y_{\text{true}} > y_r \ \& \ y_{\text{pred}} < y_r$$

Additional loss penalty is added to the loss, proportional by the factor “a” to the MSE

- Distribution based Loss for full Distribution
  - “Both Tails”

$$\text{bins} = \min(\log_{10}(y_{\text{train}}), \ldots, \max(\log_{10}(y_{\text{train}}),$$
$$\text{weight}_n = \frac{1}{[\text{hist}(\text{bin}_n^i) + 1] \cdot n_{\text{total}}}$$





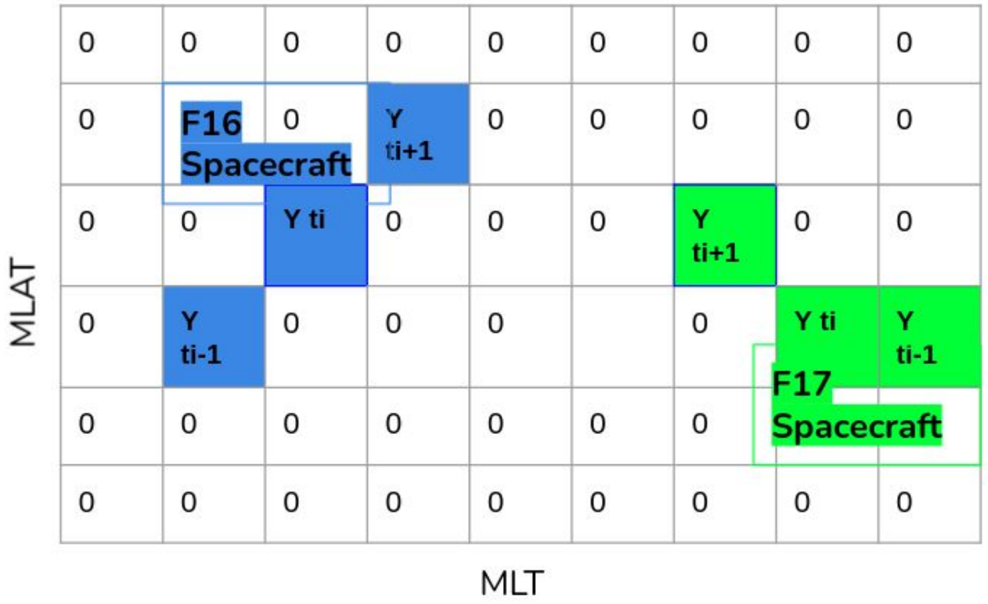
2D Convolutional (Inverse) Model

Two major enhancements:

- Couple multiple 2D points at one “time step” (training instance)
- More direct control of 2D spatial learning using CNN kernels

- Assumes a “pixelated” discretization of Magnetic Longitude and Local Time (similar to longitude)
- Assumes 5 minute window of spatial positions traced by satellites are at the same “time step”

Spatially and Temporally Sparse Training Data



$$\text{Loss} = \sum_{t} \sum_{y_{\text{true}}(t_i)\exists} \left(y_{\text{true}}(t_i) - y_{\text{pred}}(t_i)\right)^2$$

Training Inputs + Training Sample Target + Training --> Prediction Target

Cat at time t1

(orange, cute, t1)

Cat at time t2

(orange, cute, t2)

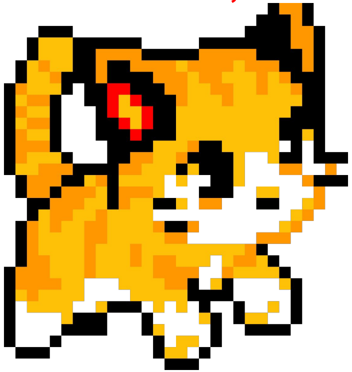
Cat at time t

(orange, cute, t)

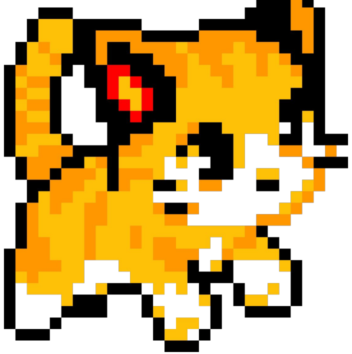
Training Sample Targets --> Training Prediction

Cat at time t, observer 1

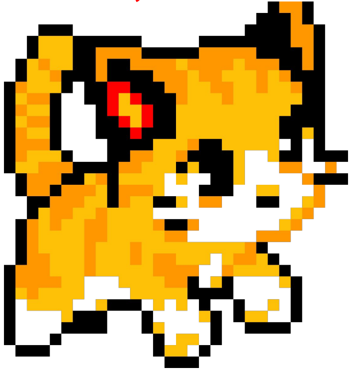
Cat at time t, observer 2



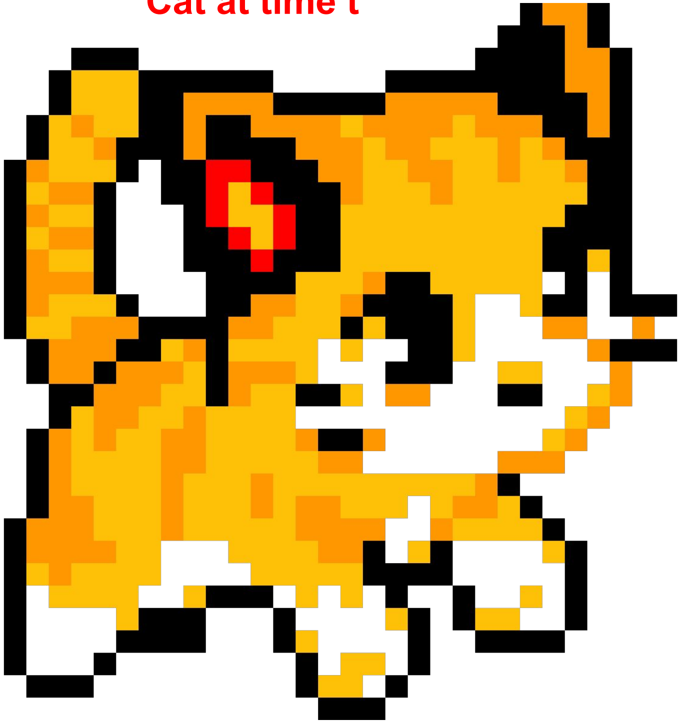
Cat at time t, observer 3



Cat at time t, observer 3



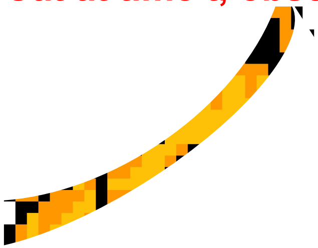
Cat at time t



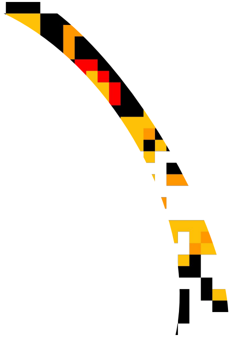
Training Sample Targets --> Training Prediction

Cat at time t, observer 1

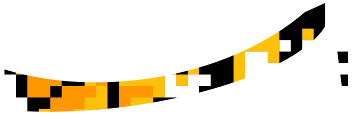
Cat at time t, observer 2



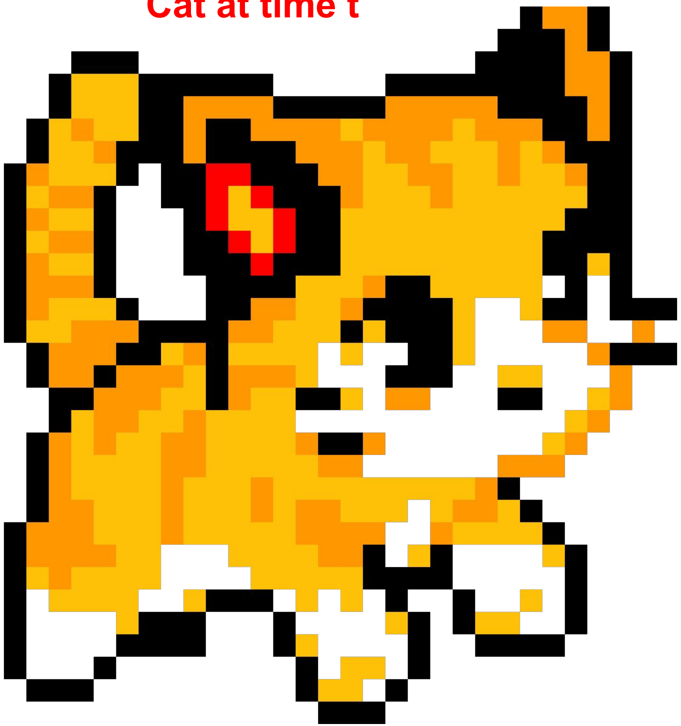
Cat at time t, observer 3



Cat at time t, observer 3



Cat at time t

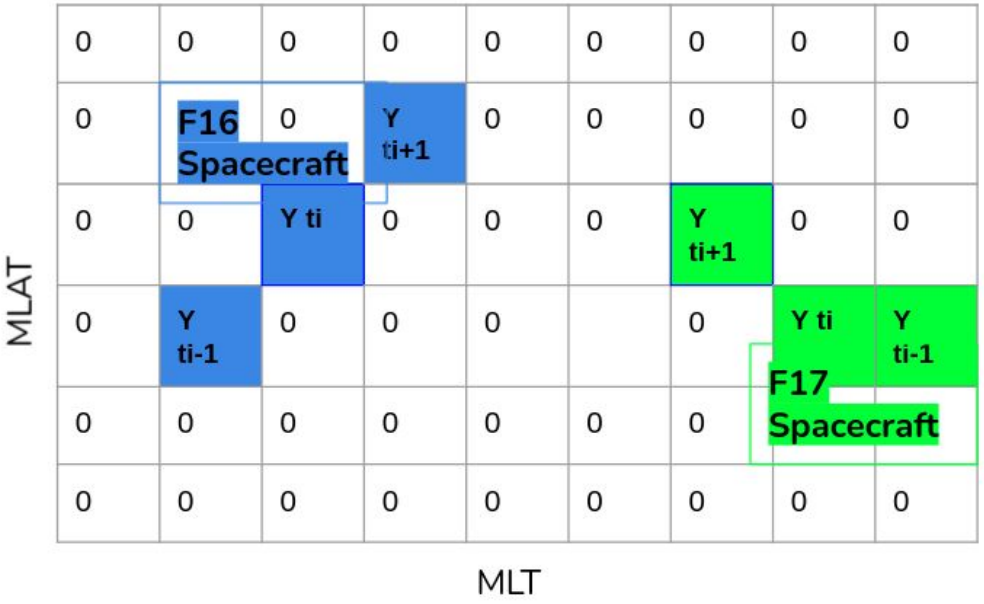


2D Convolutional (Inverse) Model

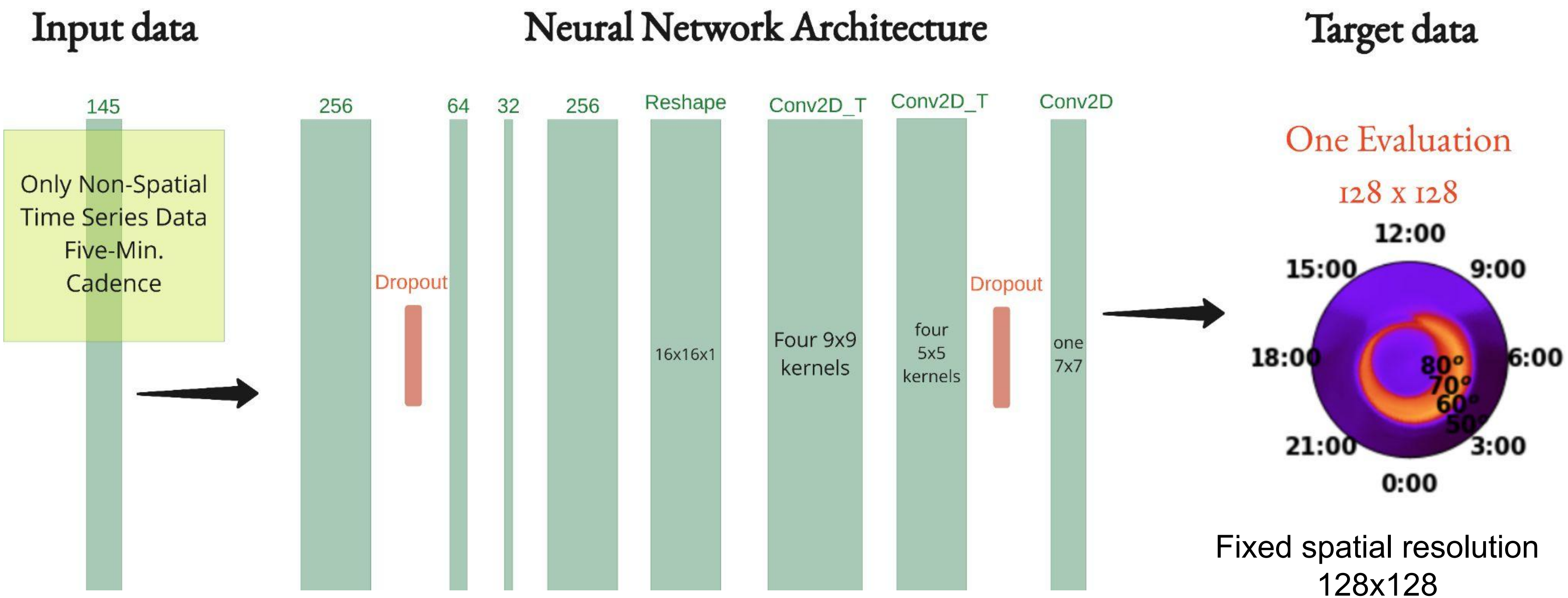
- Assumes 5 minute window of spatial positions traced by satellites are at the same "time step"

$$\text{Loss} = \sum_{t} \sum_{y_{\text{true}}(t_i)\exists} \left(y_{\text{true}}(t_i) - y_{\text{pred}}(t_i)\right)^2$$

Spatially and Temporally Sparse Training Data



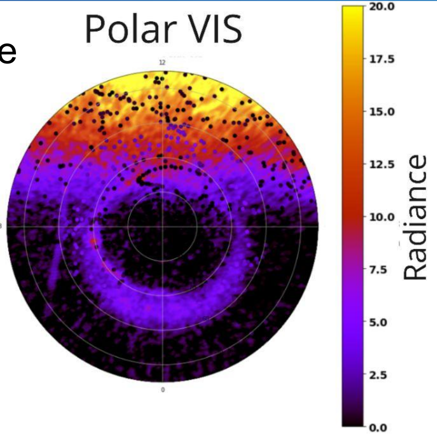
2D CNN Model

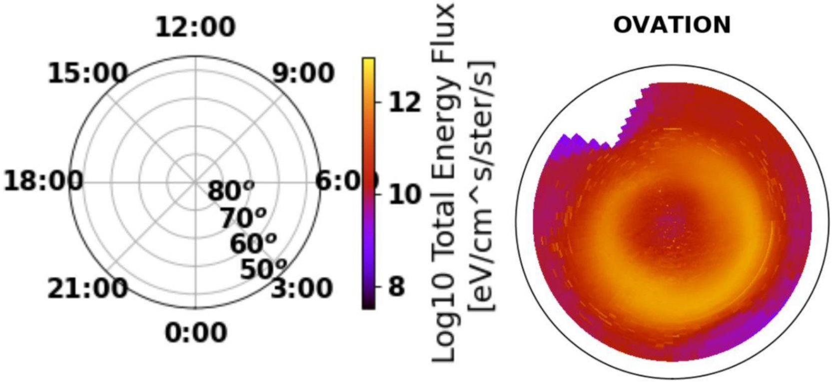


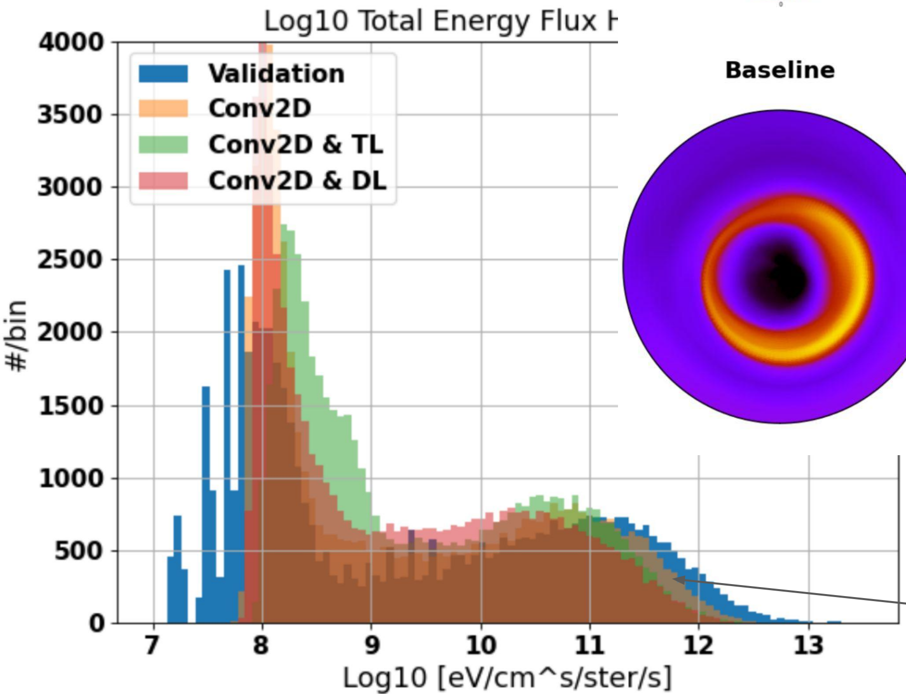
Fixed spatial resolution 128x128

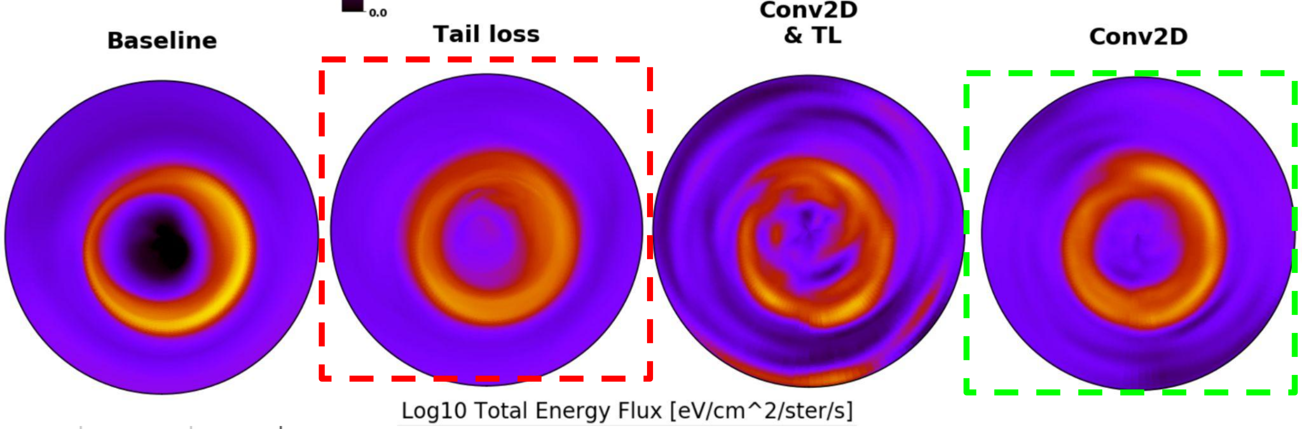
Polar Results

Northern hemisphere



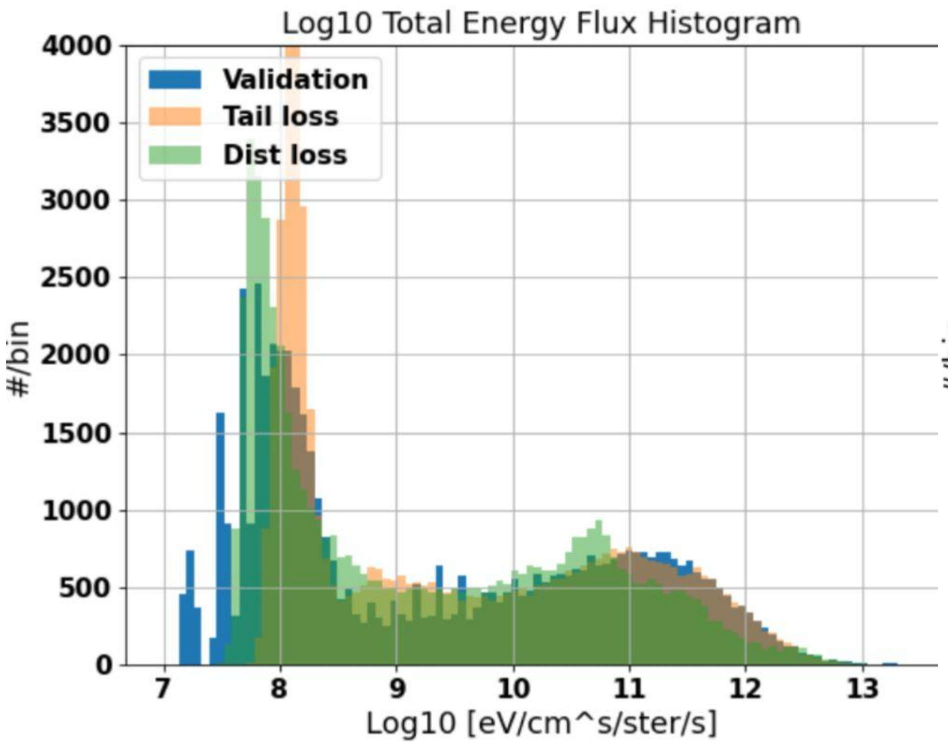


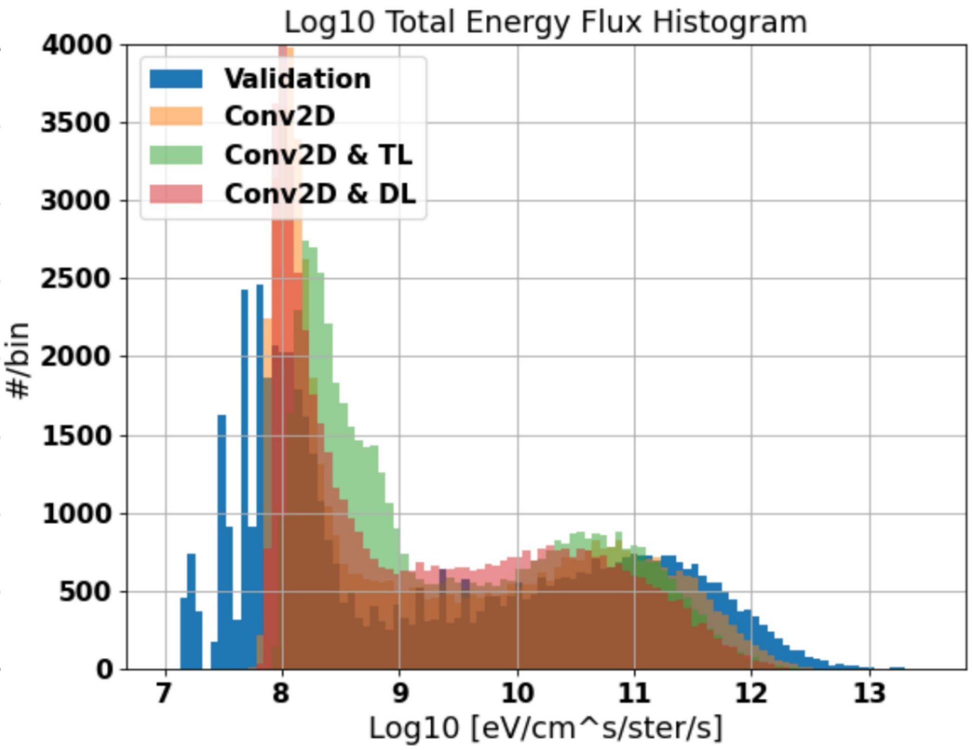


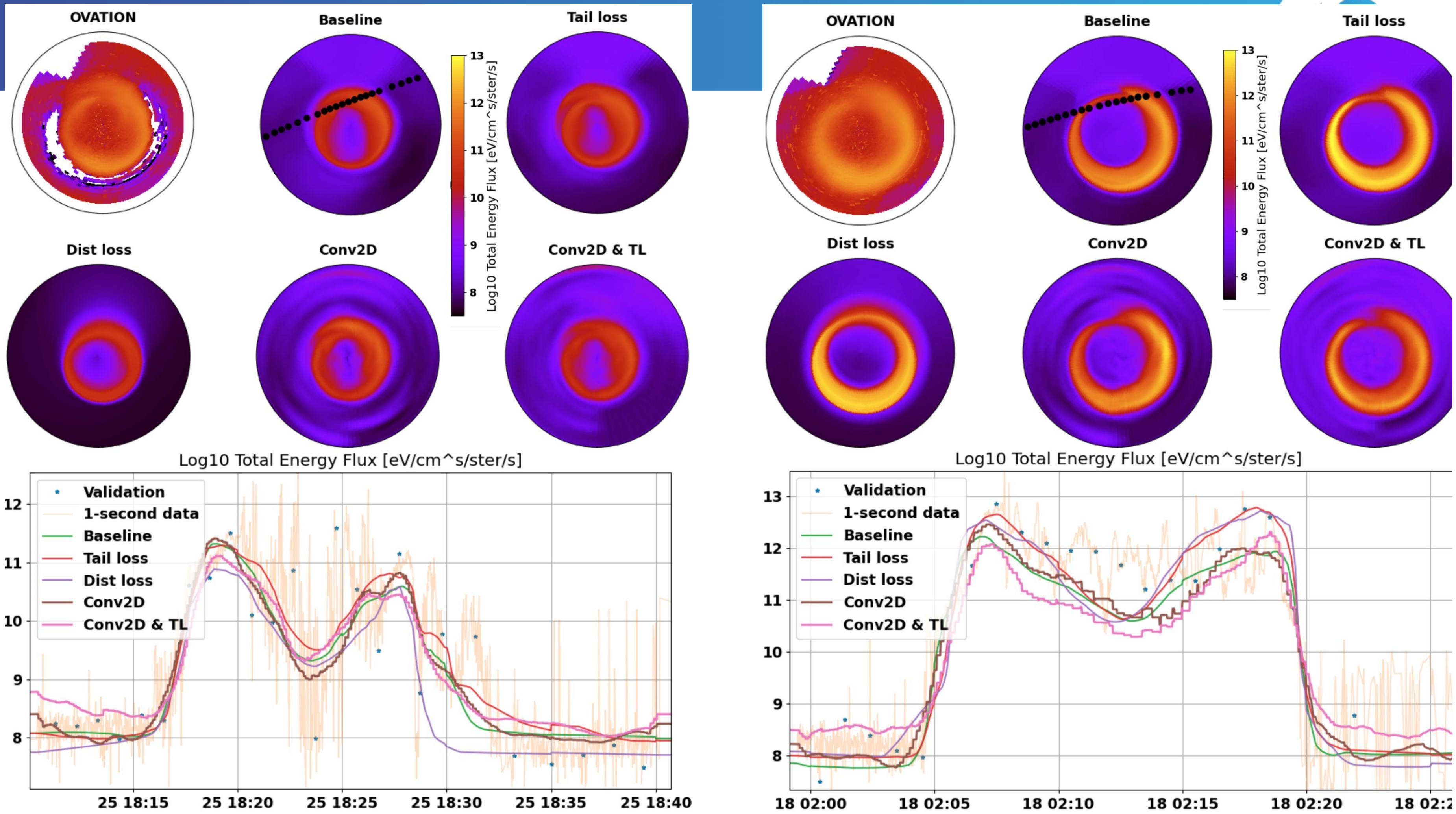


- Tail Loss model (no Conv2D)performs the best in the auroral region
- However, Conv2d predicts finer scale spatial features

Tail Loss underperformed with Conv2D






OVATION
Baseline
Tail loss
Dist loss
Conv2D
Conv2D & TL
Log10 Total Energy Flux [eV/cm^s/ster/s]
Validation
1-second data
Baseline
Tail loss
Dist loss
Conv2D
Conv2D & TL
25 18:15
25 18:20
25 18:25
25 18:30
25 18:35
25 18:40
18 02:00
18 02:05
18 02:10
18 02:15
18 02:20

Conclusions & Future Work

Finer spatial scale phenomenon is important as well as accuracy of Extreme Events:

- One-sec. data measurement cadence of DMSP has an unknown small-scale component, difficult to deconvolve from a 1D spatio-temporal trace into a 2D spatial map.
    - Our original PrecipNet approach averages out smallest scales, only medium scales captured
- Only Conv2D extension of Precipnet shows small scale features

Possibility for Large scale training:

- 1 second data spans ~120 gBytes of Target data,
    - currently training on 2 - 6 gB of data at 1 min cadence
- Increased Conv2D resolution for > 128 x 128
    - Highly Efficient Code needed for combining 1 second resolution with 2DConv

Dual Model possible

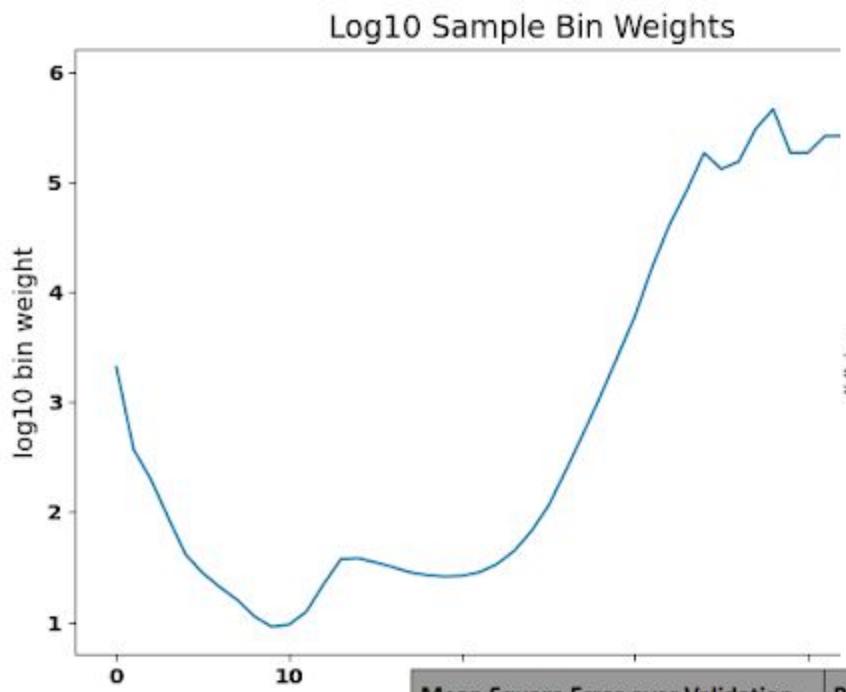
- Similar to turbulence Modeling, predict moving average (medium scales) and standard deviation (smale scales) separately

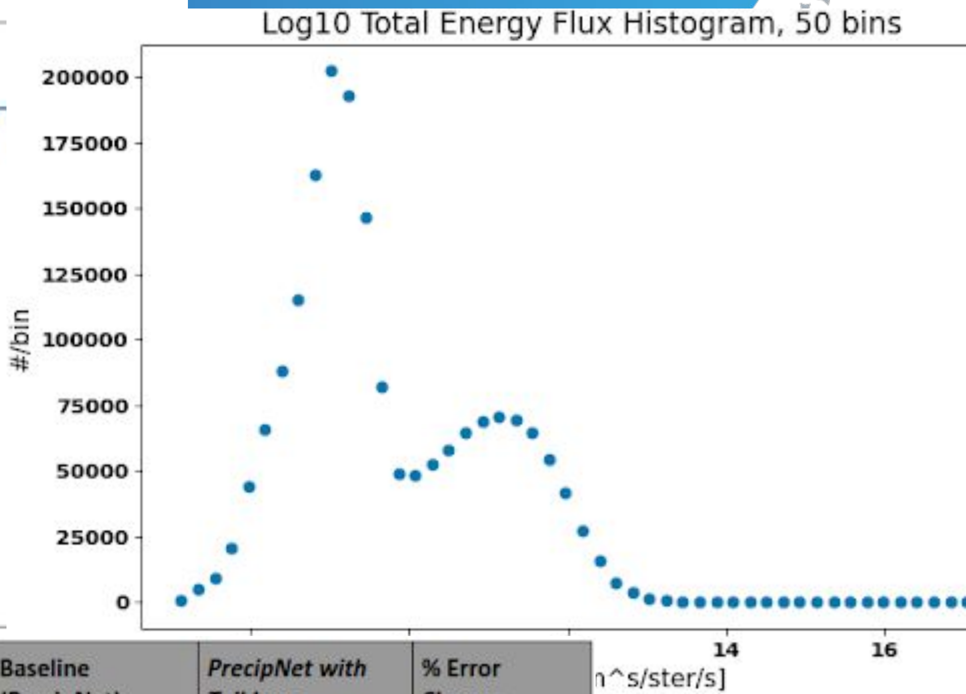
Software:

https://github.com/rmcgranaghan/HEARTBEAT , HEARTBEAT/Model_comparisons_ICMLA_Dec_2021/

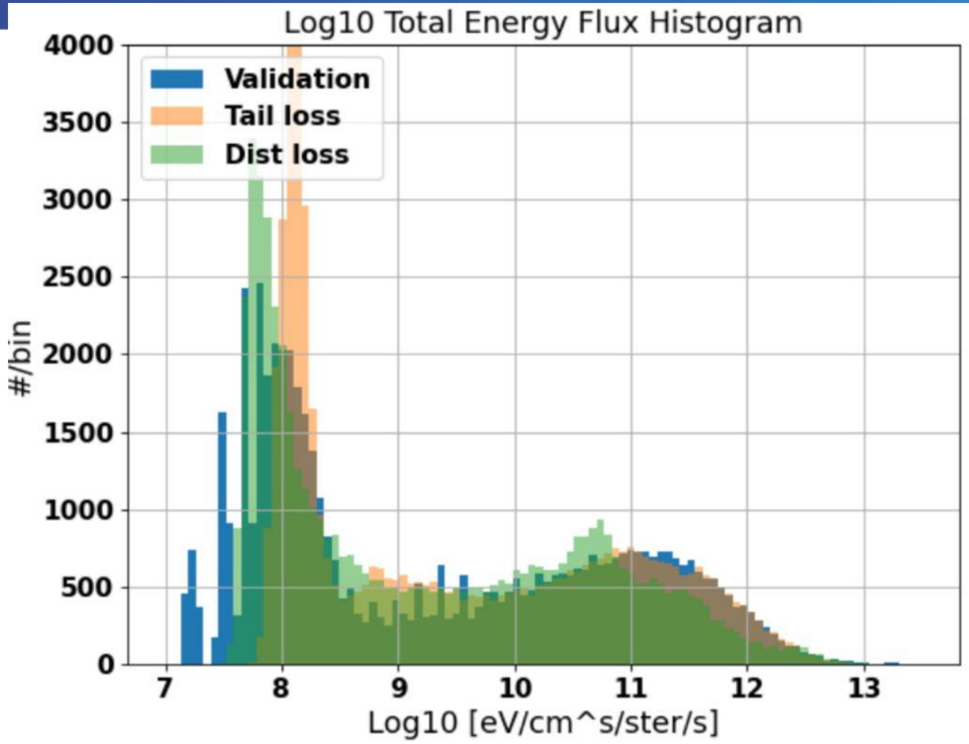
(GPU required to fully train, We used 10 gb 2080Ti)

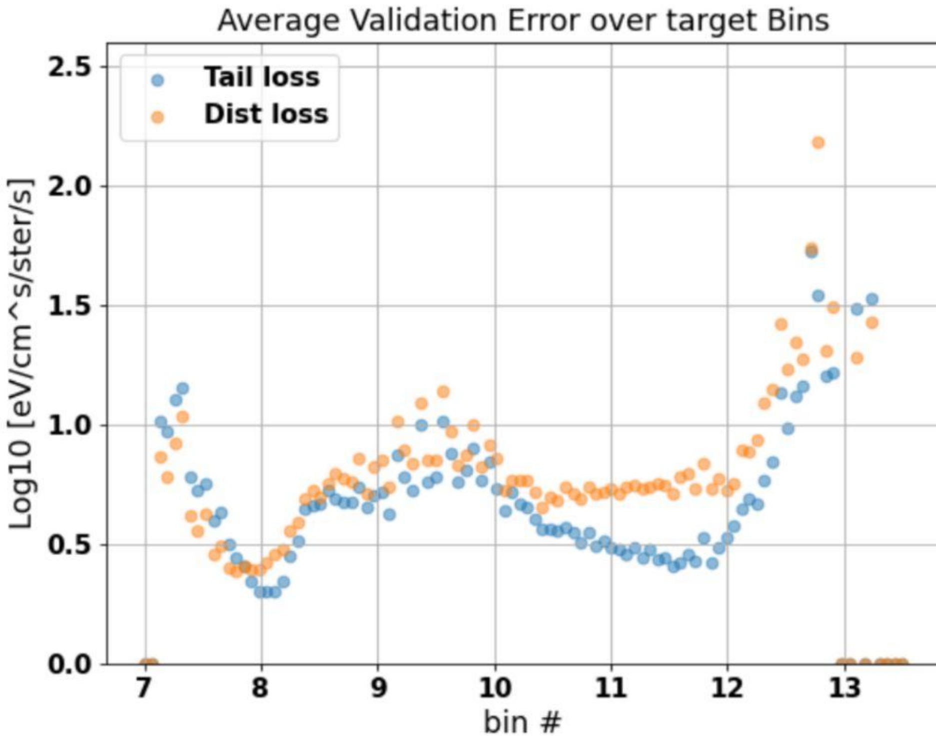
Backup

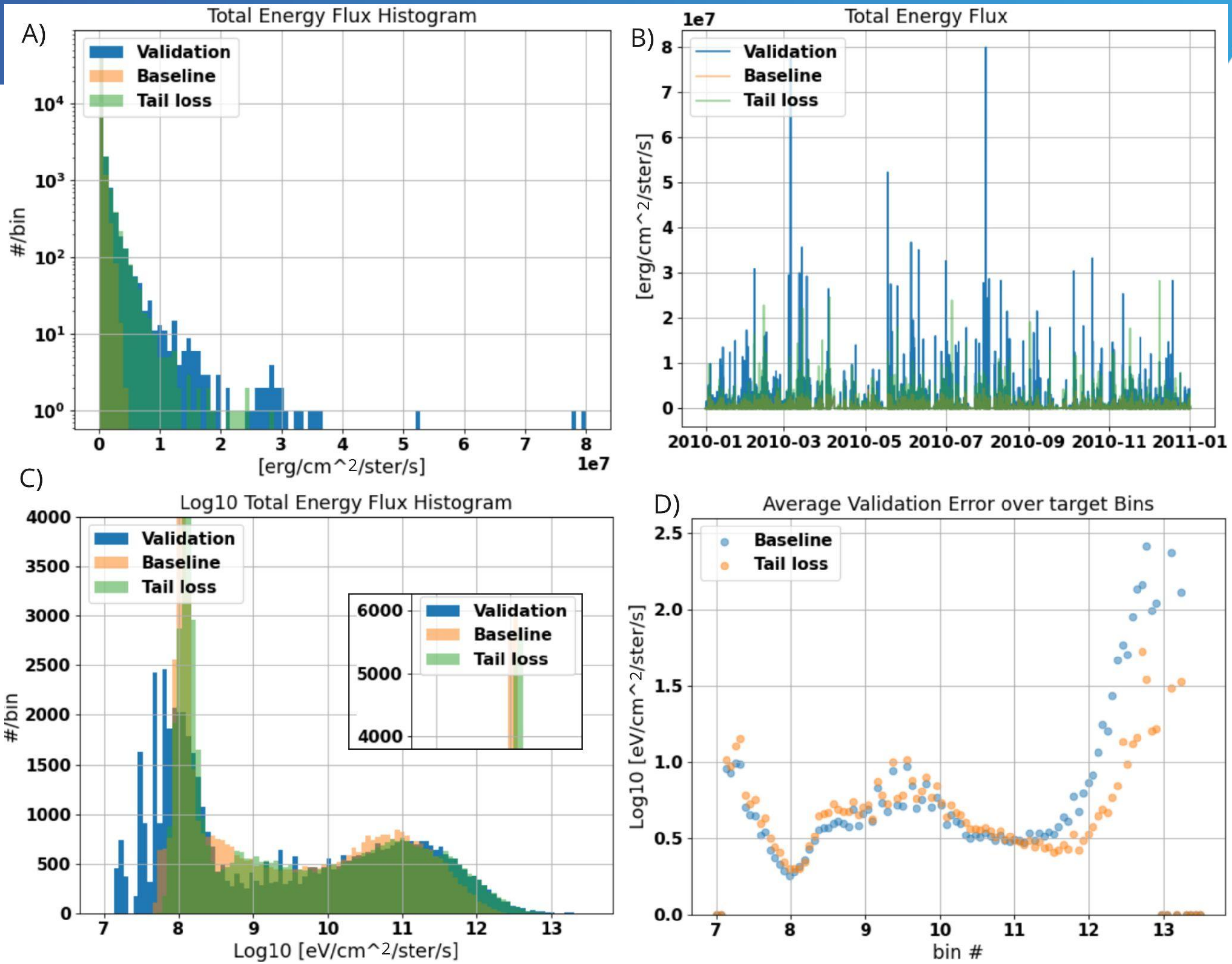




Mean Square Error over Validation Target Ranges	Baseline (PrecipNet)	*PrecipNet with Tail Loss*	% Error Change
Full Range	0.5648	0.6256	**11 %**
90th percentile (90-100%) 11.5- 13.7 log10 [eV/cm^2/ster/s]	0.7020	0.5205	**-26 %**
95th percentile (99-100%) 11.8- 13.7 log10 [eV/cm^2/ster/s]	0.9215	0.5782	**-37 %**
99th percentile (99-100%) 12.4- 13.7 log10 [eV/cm^2/ster/s]	1.9492	1.0766	**-44 %**


Log10 Total Energy Flux Histogram
Validation
Tail loss
Dist loss
#/bin
Log10 [eV/cm^s/ster/s]


Average Validation Error over target Bins
Tail loss
Dist loss
Log10 [eV/cm^s/ster/s]
bin #


A)
Total Energy Flux Histogram
Validation
Baseline
Tail loss
#/bin
[erg/cm^2/ster/s]
1e7
B)
Total Energy Flux
1e7
Validation
Baseline
Tail loss
[erg/cm^2/ster/s]
2010-01 2010-03 2010-05 2010-07 2010-09 2010-11 2011-01
C)
Log10 Total Energy Flux Histogram
Validation
Baseline
Tail loss
#/bin
Log10 [eV/cm^2/ster/s]
D)
Average Validation Error over target Bins
Baseline
Tail loss
Log10 [eV/cm^2/ster/s]
bin #