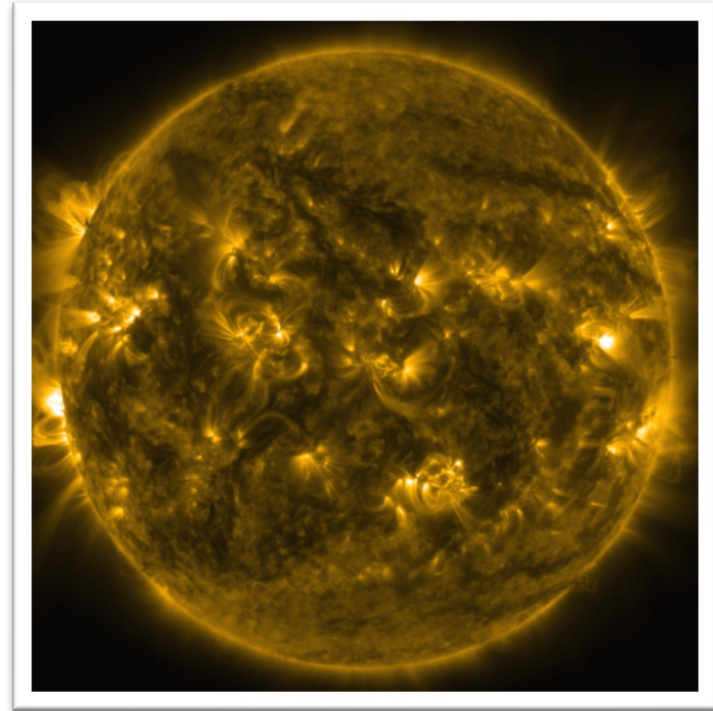


Predicting the Appearance of Sunspots on the Solar Surface Using Machine Learning

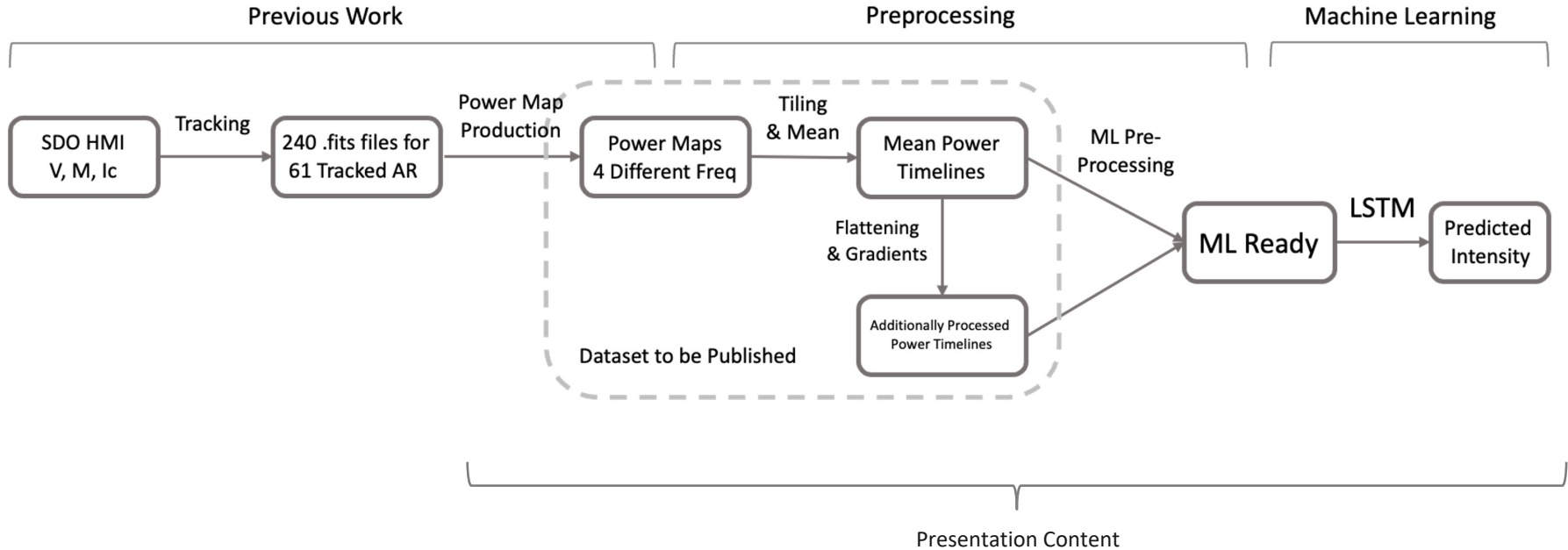
Spiridon Kasapis & Irina Kitiashvili
NASA Ames Research Center

What are Solar Active Regions?

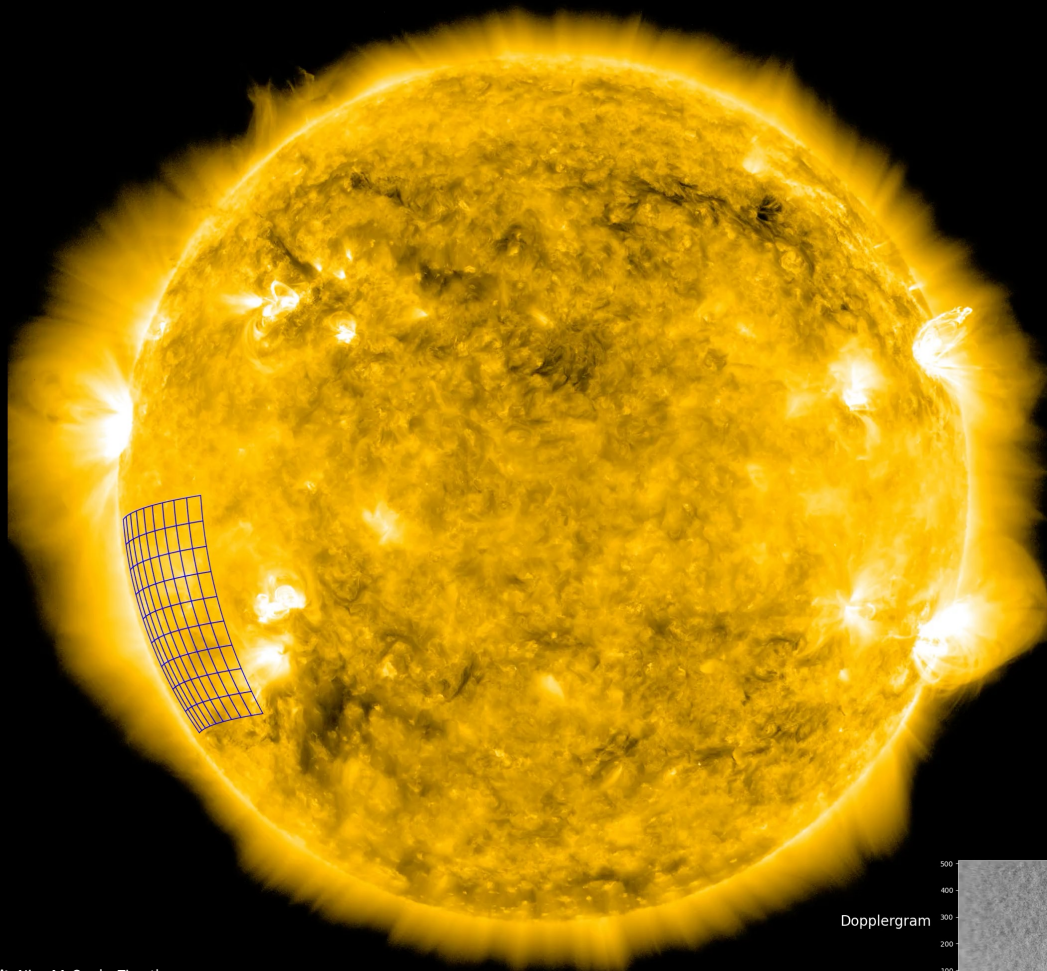
- **Areas on the Sun's surface (sunspots) with intense magnetic activity.**
- **Primary sites for the emergence of solar flares and coronal mass ejections (CMEs).**
- **The number and location is dependent on the solar cycle.**



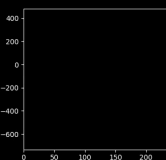
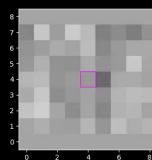
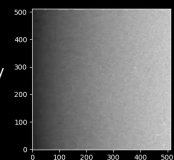
Project Data Pipeline



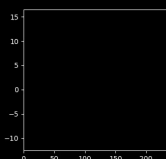
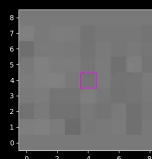
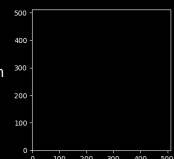
2011-02-08 18:59:37.500000



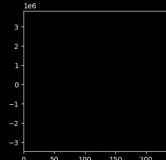
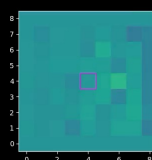
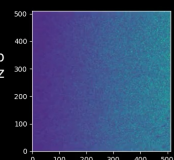
Continuum Intensity



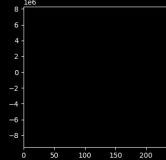
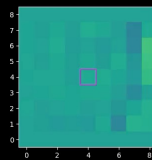
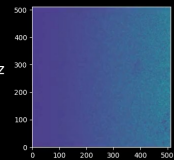
Magnetogram



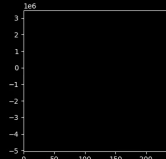
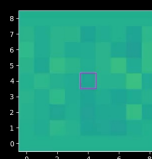
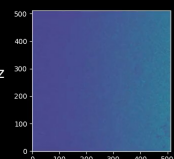
Acoustic Power Map
2-3 mHz



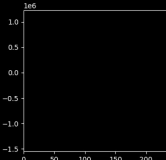
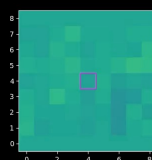
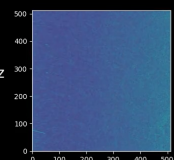
3-4 mHz



4-5 mHz



5-6 mHz



Dopplergram

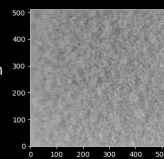
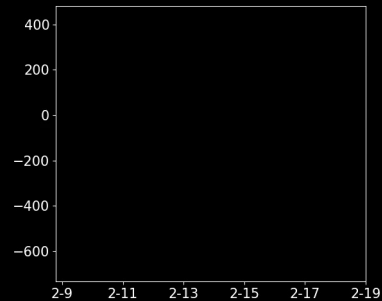
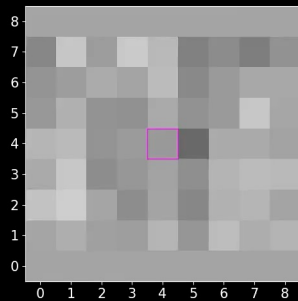
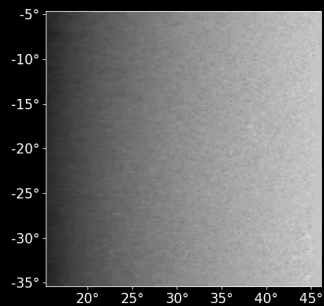


Image Credit: Nina McCurdy, Timothy Sandstrom, Spiridon Kasapis

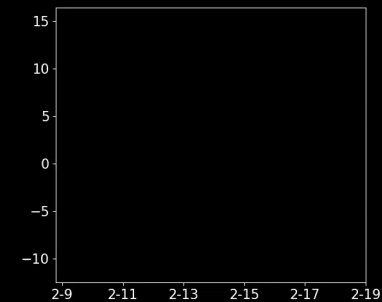
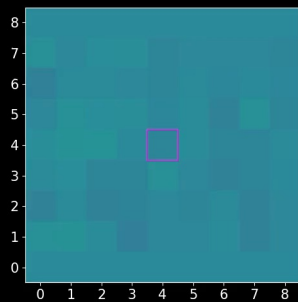
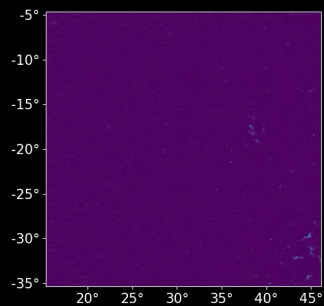
SDO/AIA 171A 2011-02-08 19:00:00:34 UTC

2011-02-08 18:59:37.500000 UTC

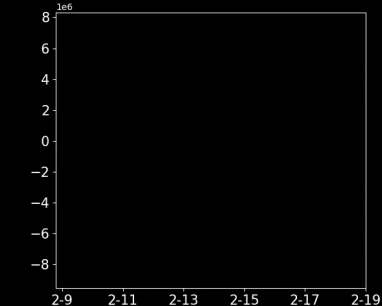
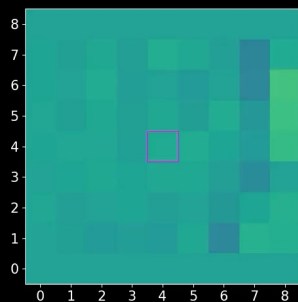
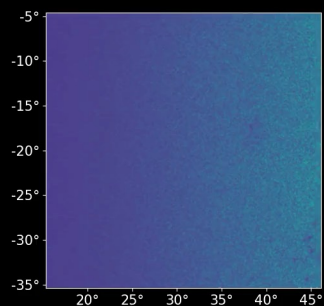
Continuum Intensity



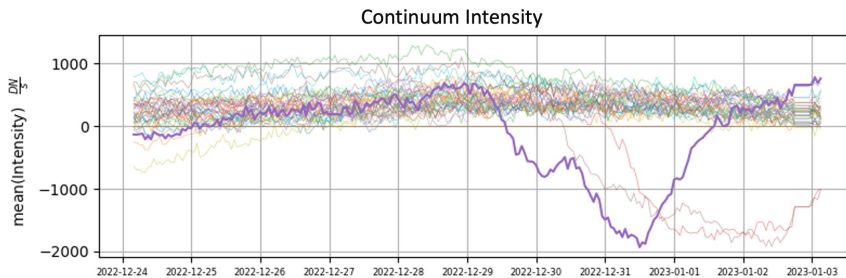
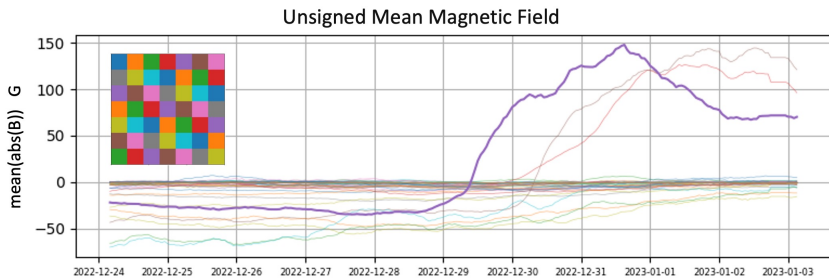
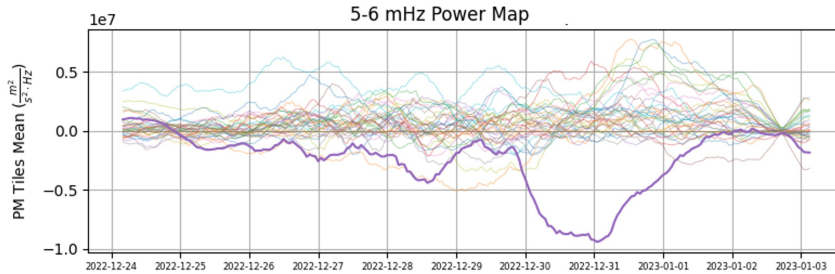
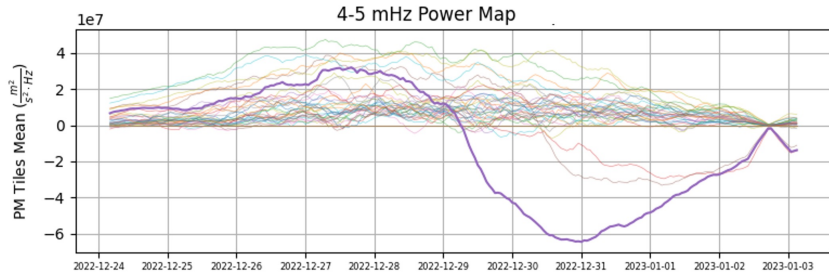
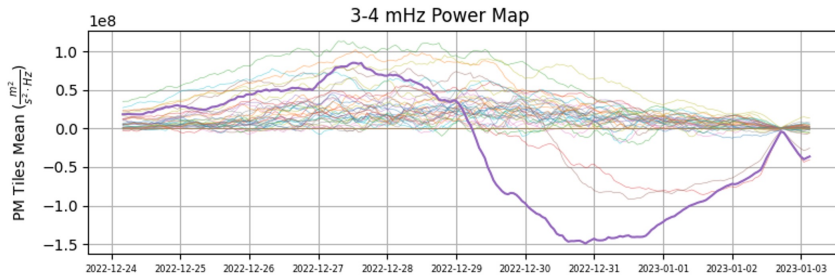
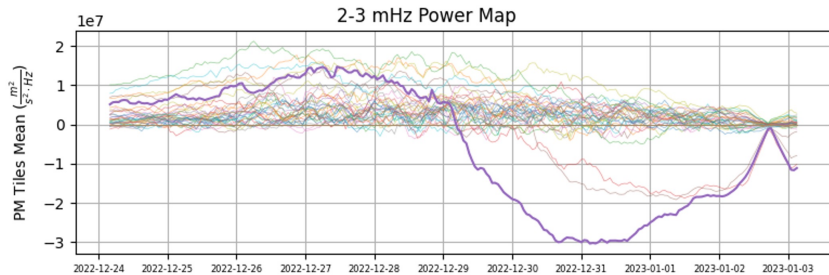
Magnetogram



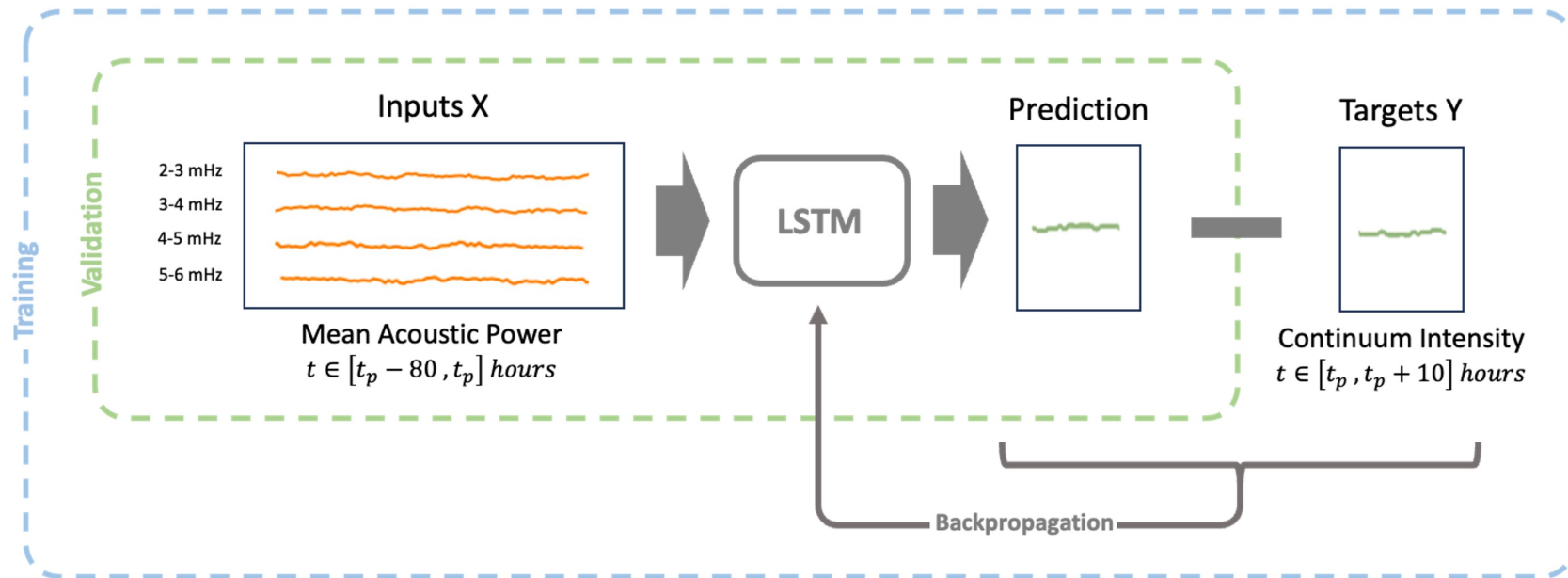
Acoustic Power Map
3-4 mHz



Machine Learning Ready Data for AR 13179



Machine Learning



Data Split

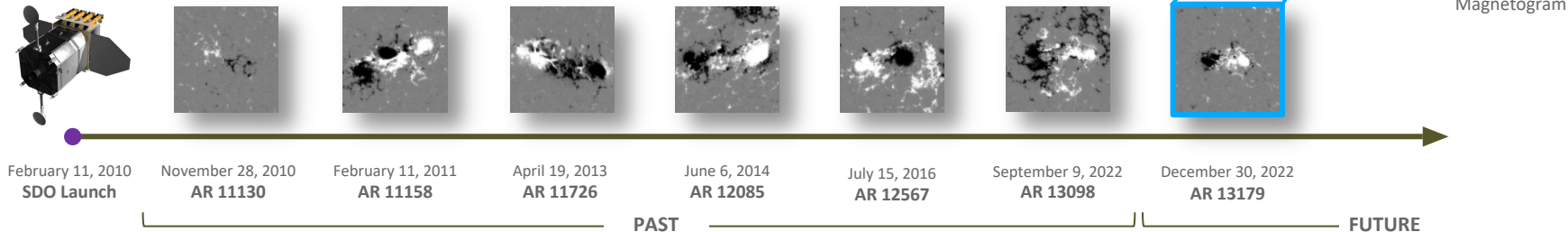
Training:

- Active Regions: **58** Active Regions (November 2010 to December 2022)
- Physical Parameters: **4** Frequency Range Acoustic Power Maps
- Map Split: **63** Tiles Each Power Map (Omitting Top and Bottom Rows)
- Sliding Windows: **140** Different Intervals
- Points in Each Window: **80** Points (1 hour cadence)
- Total Parameters: $58 \times 4 \times 63 \times 140 \times 80 = \mathbf{164 \text{ Million}}$ Data Points

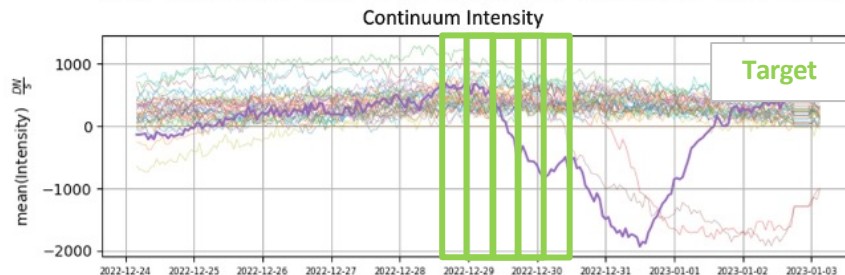
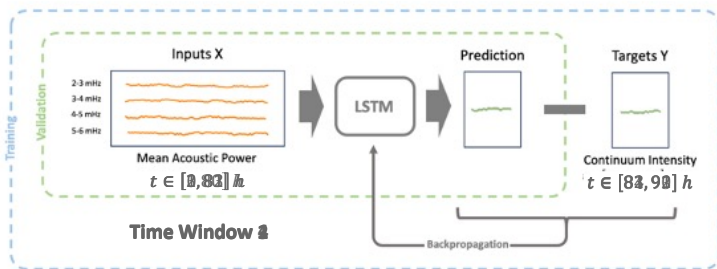
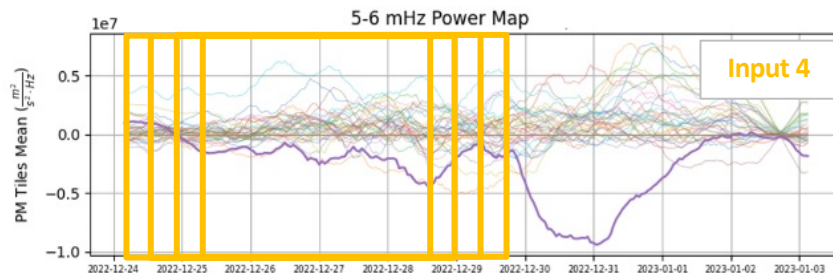
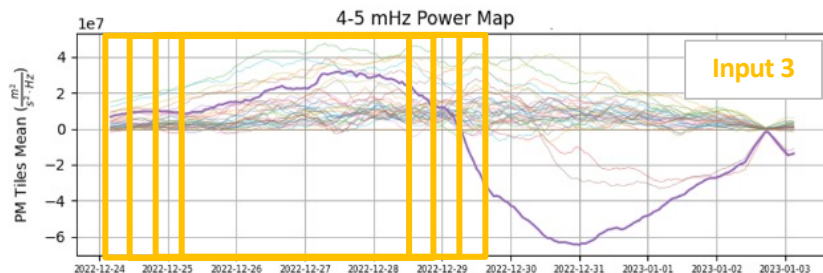
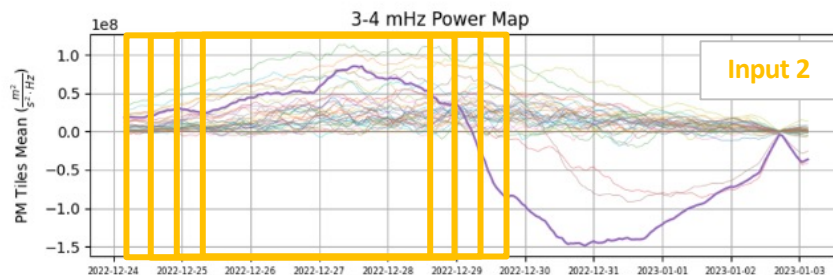
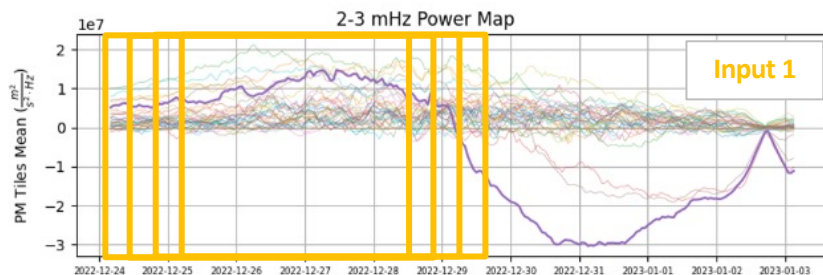
Testing:

- Active Regions: **3** Active Regions (December 2022 to April 2023)

GPU Acceleration: CPU takes **9 Days** to process all data, GPU takes **30 Hours** (x7.1 Speed-Up)

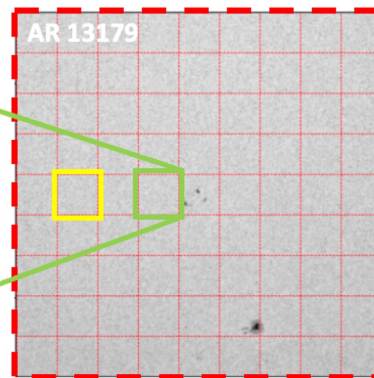
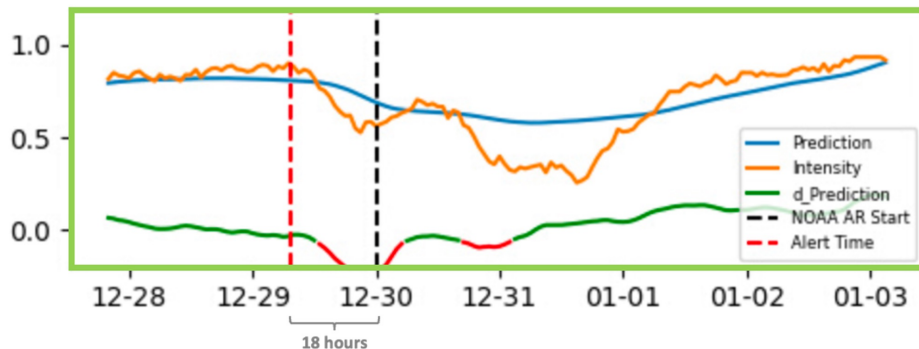


Machine Learning Ready Data for AR 13179



Predictions of LSTM Trained on 5 Active Regions (500 Epochs, 0.005 Learning Rate, 3 LSTM Layers, 16 Hidden Size)

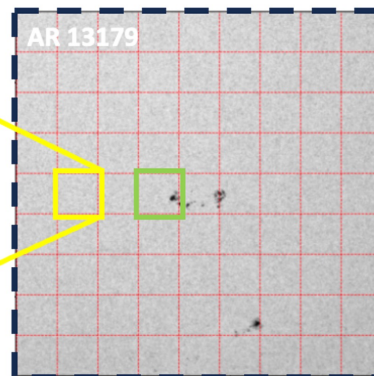
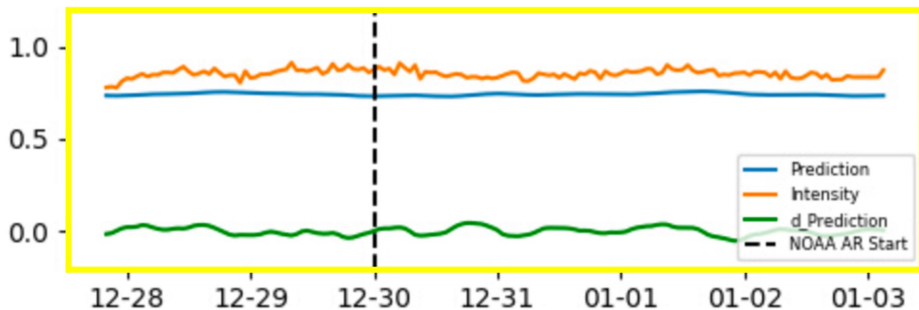
Prediction for Emerging Tile **(Alert – 18h in Advance)**



2022-12-29 06:00:00

Continuum Intensity Maps

Prediction for Non-Emerging Tile **(No Alert)**



2022-12-30 00:00:00

Challenges Addressed Using HECC Systems



- Data Transfer

- The HMI instrument takes **2TB** of data per day/ **10 PB** raw data since SDO launch in 2010.
- The data for every AR analyzed in this project takes up 150GB of space.
- A total of **9.15 TB** worth of data (150GB x 61ARs) was transferred from Stanford/JSOC and save on Lou Front End (lfe) using the NASA Secure Unattended Proxy (sup shiftc command).

- Data Preprocessing

- The 9.15TB AR data was processed on Pleiades Front End (pfe) and was reduced down to 350GB creating our ML ready datasets.
- Such a process could not be carried out on a personal computer due to storage and computational limits.

- Machine Learning (full scale ML model run time):

- Personal Computer: Apple M1 Pro CPU (up to 10 cores) would take **>2 Weeks**.
- Aitken CPU node: Intel Xeon Gold 6248 (40 cores) takes **9 Days**.
- Sky Electra GPU node: NVIDIA Volta V100 take **30 Hours** (x7.1 Speed-Up).
- Further optimization by utilizing all GPUs in a node (4 or 8 in total) is possible. Code needs to be optimized for parallel processing. For a shallow network like ours might not be worth it.



Conclusions

- We create a machine learning ready dataset of emerging solar active regions aiming to:
 - Understand the physical processes on the surface of the sun
 - Predict emergence of active regions: A drop in intensity can be predicted in a tile 18 hours in advance without false positives.
- Calculated acoustic power maps carry precursor information associated with the decrease in continuum intensity on the solar surface
- Machine learning models (LSTM/Transformers) can capture this information creating early warning systems for the emergence of solar active regions

Acknowledgements

NAS Visualization Team: Nina McCurdy, Timothy Sandstrom, Christopher Henze
New Jersey Institute of Technology: Alexander Kosovichev, John Stephan, Apte Bhairavi
HECC AI Expansion Grant



Thank you for your attention!