A Machine Learning Ready Dataset of Acoustic Power Maps for Detection of Active Region Emergence

Spiridon Kasapis[1], Irina N. Kitiashvili[1], Alexander G. Kosovichev[2], John T. Stefan[2], Bhairavi Apte[2]

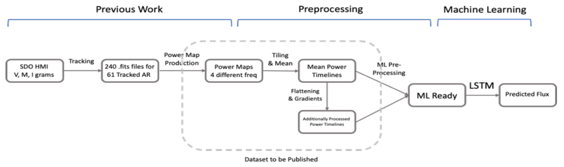
[1]NASA Ames Research Center, [2]New Jersey Institute of Technology

Overview

The development of an accurate forecast for solar eruptive activity has become increasingly important in order to prevent any potential impact on activities in space and the Earth's environment. It is therefore crucial to detect active regions before they appear on the solar surface and create early warning capabilities for upcoming Space Weather disturbances.

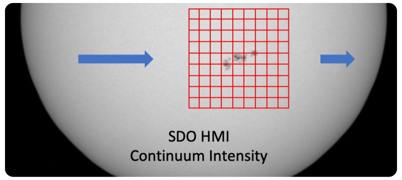
In this work, 9TB of solar data (SDO/HMI dopplergrams, magnetograms and continuum intensity maps) involving the emergence of 61 NOAA solar active regions since 2010 were processed using the NASA HECC capabilities. An acoustic power maps time-series dataset was created (for four different frequency ranges and processed to take into account the solar sphere geometric effect) which can be used for understanding the dynamics of the solar surface and train a variety of ML models. The calculated acoustic power maps carry precursor information associated with the decrease in continuum intensity on the solar surface, verifying older helioseismology research.

Our results show that a Long Short-Term Memory (LSTMs) model, with a modest layer depth and the right hyperparameters tuned, when trained on this solar acoustic power maps dataset can predict without false negatives a drop in intensity (associated with the emergence of the active region), up to 18 hours in advance.



We acknowledge the work of the NAS Visualization Team (Nina McCurdy, Timothy Sandstrom and Christopher Henze) along with HECC AI Expansion Grant for the funding provided for this research.

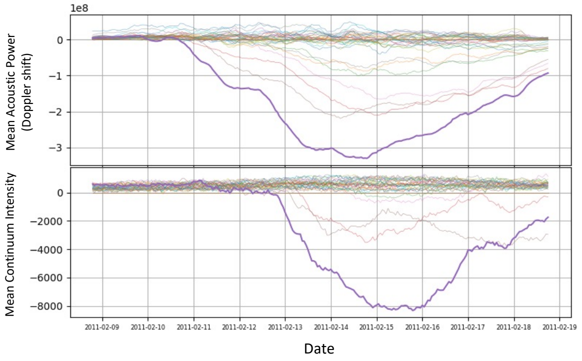
Tracked Active Regions Processing



VIDEO

For each one of the 61 largest active regions that have emerged on the visible solar surface:

- Each Aris tracked before, during and after its emergence on a 512x512 square (tracking for ~7 days/90° longitude).
- Obtained power maps for four frequency ranges.
- The mean value of all the pixels in a tile is calculated.
- Correction for the geometric effect is done by assuming that the top and bottom tile rows are non-active tiles.
- Active regions with missing or corrupt information are being further processed.

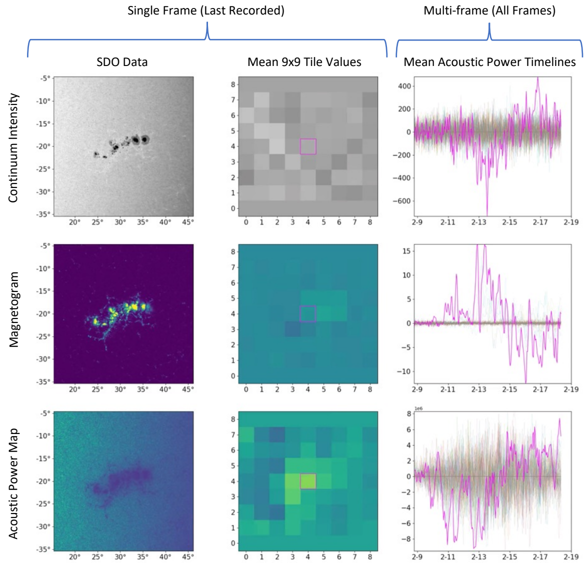


The data product shows the delayed decrease in mean continuum intensity compared to the decrease in 3-4 mHz range acoustic power during the emergence of NOAA AR 11158.

Dataset Description

The dataset we use for our ML models in numbers:

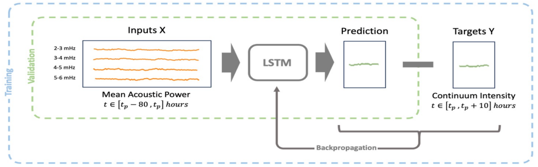
- Active Regions: **61** total Active Regions (November 2010 to December 2022) with a recommended 90/10 training-testing split (**55** training – **6** testing ARs)
- Physical Parameters: **4** Frequency Range Acoustic Power Maps (2-3, 3-4, 4-5 and 5-6 mHz), **240** step timelines.
- Map Split: **63** Tiles for Each Acoustic Power Map. We use a 9 by 9 grid (81 tiles) but we omit the top and bottom when getting rid of geometric effect.
- ML Model Normalization: We globally normalize the timelines and bring them to the 0-1 range.



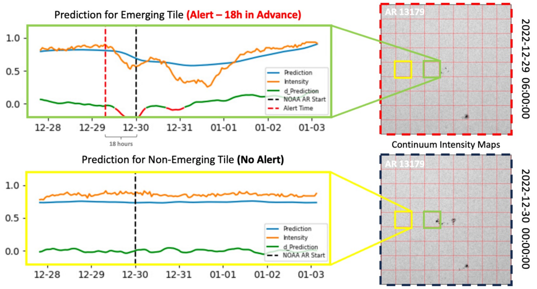
Frame comparison of the SDO/HMI original and processed data along with the tiling and mean timelines created using all frames.

Machine Learning Results

Training and testing schematic of the LSTM Method. The inputs to the network are the four frequency range mean acoustic power timelines and the target is the corresponding continuum intensity timeline.



Continuum intensity predictions of a 3-Layer LSTM (Epochs = 500, LR = 0.005, 16 Hidden Size) trained on 5 tracked ARs and evaluated on an active and non-active tile of NOAA AR13179.



Conclusions

- We create a machine learning ready dataset of emerging solar active regions aiming to a) understand the physical processes on the surface of the sun and b) 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) can capture this information creating early warning systems for the emergence of solar active regions.