EXPLORING THE CAPABILITIES OF A MACHINE LEARNING ALGORITHM TO DETECT SPACE WEATHER-SIGNIFICANT EMERGING ACTIVE REGIONS

Irina N. Kitiashvili[1], Spiridon Kasapis[1], Alexander G. Kosovichev[1,2]

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

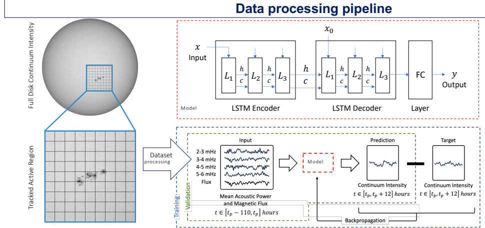
Abstract

Active regions are a source of various phenomena responsible for Space Weather disturbances; therefore, developing a technology for early warning about upcoming magnetic activity is crucial to mitigate its impact. However, observational limitations and the high nonlinearity of processes associated with the accumulation of magnetic flux and its interaction with the surrounding plasma during the emergence through the convection zone make early activity detection a challenging problem.

To address these challenges, we developed a physics-driven machine-learning model that allows us to detect active regions (ARs) before they become visible on the solar surface by analyzing the power spectra of acoustic oscillations observed by the SDO/HMI instrument. This study is based on a time series of Doppler shift maps of 31x31-degree areas tracked with the Carrington rotation rate for four days before and after the emergence. The Doppler shift time series are processed into the oscillation power maps for four frequency ranges and accompanied by line-of-sight magnetograms and the continuum intensity maps from SDO/HMI.

The resulting data are converted into a 1D time series representing the mean temporal variations of these quantities. The redacted time series are used as input to predict AR emergence using the Long Short Term Memory (LSTM) method. The training of the LSTM model is based on 40 ARs, which includes an independent analysis for each subregion that exhibits AR emergence or remains quiet. The emergence of magnetic flux (defined as a decrease of the continuum intensity) was detected with the developed LSTM algorithm from 5 to 48 hours before the reported time by NOAA. The developed model is capable of pointing to the time and location of active region formation. In this presentation, we discuss reasons that impact how early in advance the model can identify the upcoming activity and the possibility of improving the current predictive skills and steps to transition to the operational forecast.

Data processing pipeline



Target active regions before and after the emergence have been tracked over the solar disk. The resulting Dopplergrams were used to generate power maps for the four frequency ranges. The 2D-time series was converted into an ensemble of the 1D time series by averaging the values of each tile. The resulting timelines are used as input together with the continuum intensity and unsigned magnetic flux data in the training and validation/testing of LSTM models. The LSTM Models architecture is presented in the area enclosed by the red dashed line.

Testing LSTM model to predict AR emergence

Active regions used for testing

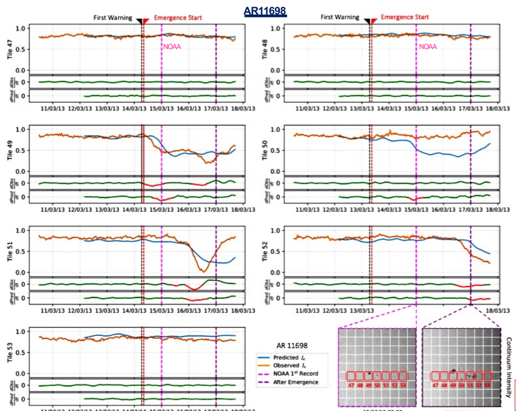
AR#	First Record	Last Record	ϕ	λ_{Carr}	λ_s	λ_e	A_s	A_e	A_{max}	A_{max} Date	McIntosh	Hale
11698	2013.03.15	2013.03.19	-19.5	117.0	29.0	86.0	20	200	200	2013.03.18	Eao	β
11726	2013.04.20	2013.04.27	13.0	327.0	-7.0	93.0	20	600	1000	2013.04.26	Fkc	$\beta\gamma\delta$
13165	2022.12.12	2022.12.18	-20.0	277.5	10.0	88.0	20	150	340	2022.12.16	Ekc	$\beta\delta$
13179	2022.12.30	2023.01.05	13.5	43.0	11.0	92.0	30	80	380	2023.01.03	Dko	β
13183	2023.01.06	2023.01.12	-16.5	309.0	8.0	91.0	30	50	200	2023.01.08	Dso	$\beta\delta$

AR# is NOAA active region number; the date of NOAA's first (First Record) and last (Last Record) record; ϕ is the AR emergence latitude; λ_{Carr} is the Carrington longitude; λ_s and λ_e are the starting and ending longitudes; A_s, A_e, and A_{max} are the starting, ending and maximum area of the target AR and the date of the maximum area (A_{max} Date).

AR prediction performance for 12 hours-ahead

AR#	T_{obs}-T_{pred}	Tile ID
11698	-8 hours	49
11726	4 hours*	41/42*
13165	12 hours	32
13179	0 hours	41
13183	-7 hours	41

AR11698



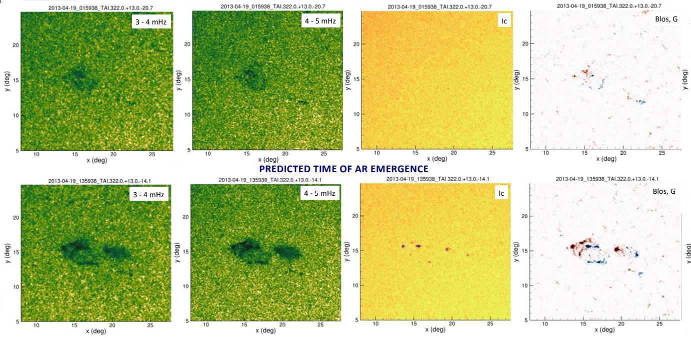
Average metrics for all tiles plotted: MAE = 0.098, MSE = 0.024, RMSE = 0.126, RMSLE = 0.079, R^2 = −5.295

Evaluation of Model 8 on predicting variations of the Ic for the selected tiles of AR11698 (tiles 47-53). The tiles' locations are marked in red squares at the bottom right continuum intensity images. Each tile's corresponding observed and predicted mean continuum intensity variations are shown as orange and blue curves. The time-derivatives of the continuum intensity (observed and predicted) are shown color-coded according to the emergence criteria. In red are the emergence periods while in green are the non-emerging (quiet) states. The rest of the tiles are quiet and correspond to non-emerging states. Vertical dashed lines identify the following moments: NOAA's first record of the active region (magenta), the time two days after NOAA's first record (purple), the time when Model 8 produces its first emergence alarm (First Warning, black) and the time when the observed emergence starts (Emergence Start, red). The tiles for which instead of dashed, the First Warning and Emergence lines are solid, are the tiles in which these events took place.

Emergence criteria:
dIc/dt < -0.01 for over 3 hours

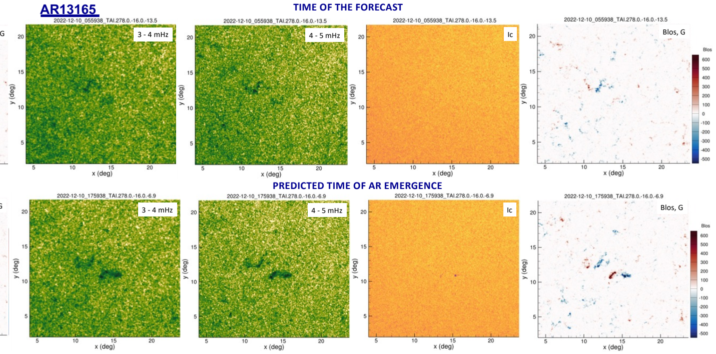
AR11726

TIME OF THE FORECAST / PREDICTED TIME OF AR EMERGENCE

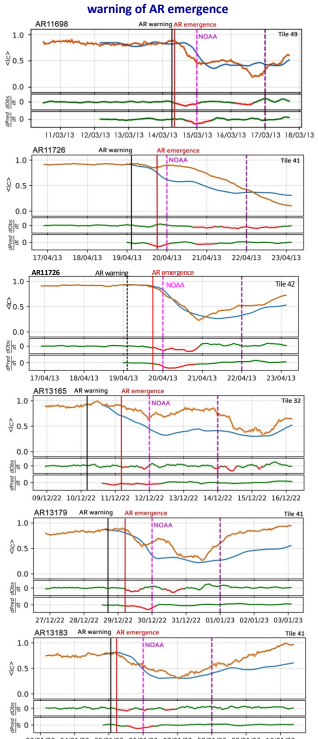


AR13165

TIME OF THE FORECAST / PREDICTED TIME OF AR EMERGENCE

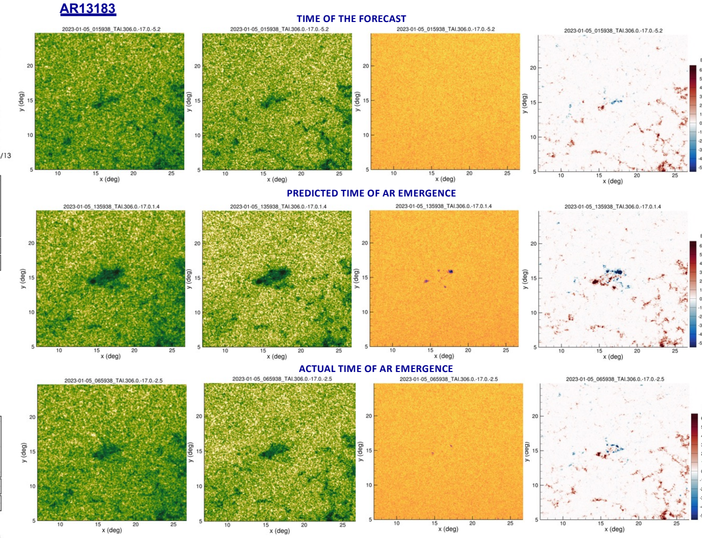


Prediction of the continuum intensity to enable early warning of AR emergence

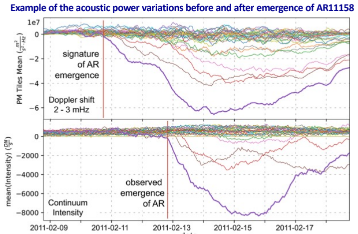


AR13183

TIME OF THE FORECAST / PREDICTED TIME OF AR EMERGENCE / ACTUAL TIME OF AR EMERGENCE



Example of the acoustic power variations before and after emergence of AR11158



Tile-averaged time-evolution of a power map for the frequency range of 2-3mHz (upper panel) reveals a decrease in the oscillation power about one day before the AR11158 emergence on the solar surface (bottom panel). Different curves correspond to the evolution of the oscillatory power for different tiles. The red vertical lines indicate the time of the power suppression in the central tile (violet curve, upper panel) and the decrease of the continuum intensity decrease in the same tile (bottom).

Conclusions

In this work we address the problem of predicting the emergence of active regions (ARs) on continuum intensity maps} by developing a dataset that includes 45 ARs tracked with solar rotation before and after emergence. This dataset was used to generate acoustic power time-series for four different frequency ranges. In this research only 45 ARs were utilized due to the presence of data gaps on the remaining 16 ARs. Using the acoustic power and unsigned magnetic flux time series as input, we developed ML models to predict decreases in the continuum intensity associated with the emergence of an AR. Despite utilizing four frequency ranges to predict AR emergence, we found that power maps for 3-4 and 4-5 mHz frequency ranges carry most of information related to coming emergence of an AR.

The analysis of the AR emergence results highlights potential improvements not only for the ML-ready dataset but also for the methods used to train and test the ML models. The 9-by-9 grid setup used here produces 81 tiles, for the majority of which (>90%) no activity can be detected. This active-quiet tile imbalance is addressed by omitting the majority of quiet tile time series during training to create a balance between the two types of data. This training technique, although adequate for training the models presented, discards a large amount of training data, which can potentially carry useful information related to AR emergence.

In this paper, we demonstrate the capabilities of an LSTM-based RNN architecture to predict the emergence of active regions on the solar surface using the local evolution of the unsigned magnetic flux and the acoustic power for four frequency ranges to predict a sharp decrease in the continuum intensity associated with the emergence of the active regions

This approach allowed us to demonstrate the model's capabilities to discriminate between regions that remain quiet over time and regions that exhibit activity related to the emergence of large active regions.

References

Hartlep T., Kosovichev A. G., Zhao, J., Mansour, N. N. 2011. Signatures of Emerging Subsurface Structures in Acoustic Power Maps of the Sun. Solar Physics, Volume 268, Issue 2, pp.321-327.

Ilonidis S., Zhao J., Kosovichev A. 2011. Detection of Emerging Sunspot Regions in the Solar Interior. Science 333, Issue 6045, pp. 993.

Ilonidis S., Zhao J., Hartlep T. 2013. Helioseismic Investigation of Emerging Magnetic Flux in the Solar Convection Zone. ApJ 777, Issue 2, article id. 138, 11 pp. (2013).

Kasapis S., Kitiashvili I.N., Kosovichev A.G., Stefan J.T. 2024. Solar Active Regions Emergence Prediction Using Long Short-Term Memory Networks. ApJ (submitted), arXiv:2409.17421

Kasapis S., Kitiashvili I., Kosovichev A., Stefan J., Apte, Bhairavi. 2024. Predicting the Emergence of Solar Active Regions Using Machine Learning. Proceedings of IAU Symposium #365 "Dynamics of Solar and Stellar Convection Zones and Atmospheres".

Ilonidis S., Zhao J., Kosovichev A. 2011. Detection of Emerging Sunspot Regions in the Solar Interior. Science 333, Issue 6045, pp. 993.

Ilonidis S., Zhao J., Hartlep T. 2013. Helioseismic Investigation of Emerging Magnetic Flux in the Solar Convection Zone. ApJ 777, Issue 2, article id. 138, 11 pp. (2013).

Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. 2020. Physica D: Nonlinear Phenomena, Volume 404, id. 132306.

Hochreiter S., Schmidhuber J. 1997. Long Short-Term Memory. Neural Comrutation 9(8), 1735-178.

Sample of the training dataset

AR#	First Record	Last Record	ϕ	λ_{Carr}	λ_s	λ_e	A_s	A_e	A_{max}	A_{max} Date	McIntosh	Hale
11130	2010.11.29	2010.12.06	12.5	330.5	0.0	94.0	60	40	250	2010.12.02	Dai	β
11149	2011.01.22	2011.01.28	17.0	346.5	5.0	90.0	70	0	250	2011.01.27	Dso	β
11158	2011.02.12	2011.02.21	-20.0	33.5	-25.0	88.0	40	200	620	2011.02.17	Ekc	$\beta\gamma\delta$
11162	2011.02.19	2011.02.25	17.5	337.0	6.0	89.0	260	0	260	2011.02.19	Dai	$\beta\gamma$
11199	2011.04.27	2011.05.02	19.5	188.5	23.0	86.0	20	210	210	2011.05.02	Dso	β
11327	2011.10.21	2011.10.28	-21.0	335.5	-10.0	81.0	10	60	200	2011.10.25	Dso	β

This work is supported by the NASA AI/ML HECC Expansion Program, NASA Heliophysics Supporting Research Program, and the NASA grants 80NSSC19K0630, 80NSSC19K0268, 80NSSC20K1870, and 80NSSC22M0162.