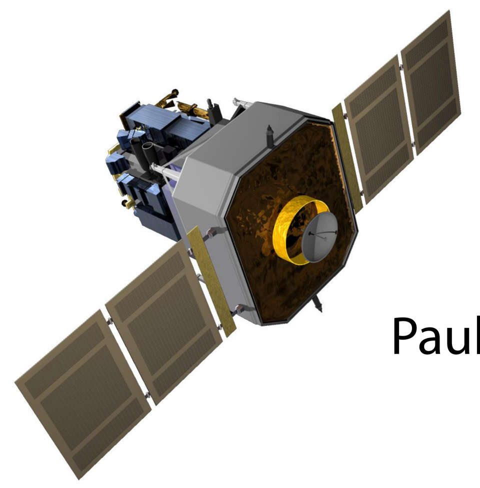
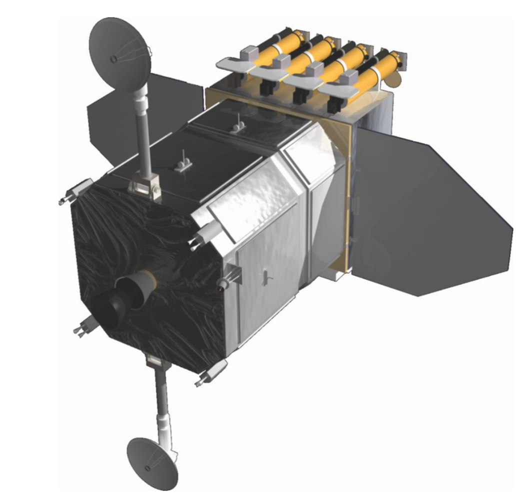


Time Series of Magnetic Field Parameters Extracted from Merged Space-Weather MDI/HMI Active Region Patches as Potential Tool for Solar Flare Forecasting



Paul Kosovich[1], Viacheslav Sadykov[2], Vincent Oria1, Patrick O'Keefe[1], Chun-Jie Chong[1], Fraila Francis[1], Aatiya Ali[2], Russell Marroquin[2], Alexander Kosovichev[1], Irina Kitiashvili[3] , Gelu Nita[1]

[1]New Jersey Institute of Technology, [2]Georgia State University, [3]NASA Ames Research Center

ABSTRACT

Space-Weather MDI Active Region Patches (SMARPs) and Space-Weather HMI Active Region Patches (SHARPs) are two recently developed data products, which have been used for solar flare prediction studies. The present work is an effort to expand the application of SMARP and SHARP summary heliomagnetic parameters to the forecasting of solar flares. A new data product was derived by filtering, rescaling, and merging the SMARP and SHARP summary parameter data series, which were further converted into two-dimensional arrays by selecting time slices corresponding to R-value maxima, where R-value is a measure of the unsigned magnetic flux near polarity inversion lines. The resulting combined MDI-HMI time series currently span the period between April 4, 1996 and December 13, 2022, and can be extended to a more recent date, providing an opportunity to correlate and compare them with other solar activity parameters, such as the daily solar flare index, which is computed as a sum of the product of GOES X-ray flare magnitude and flare duration, for all M- and X-class flares during a day. Preliminary results demonstrate a significant overall correlation, with Pearson coefficients between 0.339 and 0.627. In addition, an oscillating pattern is seen in the daily-averaged sliding-window correlation coefficient. Time-lagged cross-correlation indicates that a leader-follower dynamic exists in some parameters, especially R-value, where they lead the flare index by at least several days, which may have potential for further application in space weather forecasting.

INTRODUCTION

Space-Weather MDI Active Region Patches (SMARPs) and Space-Weather HMI Active Region Patches (SHARPs) include summary space-weather parameters (keywords) for tracked and evolving solar active regions (ARs), and are developed and maintained at Stanford University. The keywords from the two data series of interest here, mdi.smarp_cea_96m and hmi.sharp_cea_720s, are derived from photospheric magnetic field maps in Lambert cylindrical equal-area coordinates, at 96-minute and 12-minute cadences, respectively. The overlap in high-quality MDI and HMI observations from May 1, 2010 to October 28, 2010 provides an opportunity to rescale and merge the data from these instruments.

Keyword	Description
USFLUXL	Total line-of-sight unsigned flux, Maxwells.
MEANGBL	Mean value of the line-of-sight field gradient, Gauss/Mm.
R_VALUE	Unsigned flux R near polarity inversion lines, Maxwells.
CMASKL	CEA (cylindrical equal-area) pixels in the active region.
MEANGBZ	Mean value of the vertical field gradient, Gauss/Mm.
USFLUXZ	Vertical component of the total unsigned flux, Maxwells.

Summary of the keywords referred to in this study.

DATA PROCESSING

The data are retrieved from Stanford's Joint Science Operations Center (JSOC) database using the Lookdata tool, accessed through the JSOC web portal[1]. The higher-cadence SHARP keywords are retrieved from the database in four steps, due to the query size limit, and are then merged with SMARP keywords using a relatively complex Matlab script. Before running this script, the tab-delimited text files from JSOC first need to be converted into UTF-8 .csv format. Then, the script looks for HMI SHARP .csv keyword data files in a specified source directory, imports them, and merges them into one table. An MDI SMARP .csv file is imported, any duplicate records from both data series are removed, and units are converted from Gauss/pixel to Gauss/Mm. The R-value is recomputed by finding the common antilogarithm of the original R-value. Additional filtering is applied to the data, in order to exclude low-quality data values[2], as well as any data below -65 and above +65 degrees Stonyhurst longitude. The vertical components of the unsigned flux and mean field gradient are absent from the SMARP data series. Therefore, they are recalculated from the line-of-sight keywords as follows[3]:

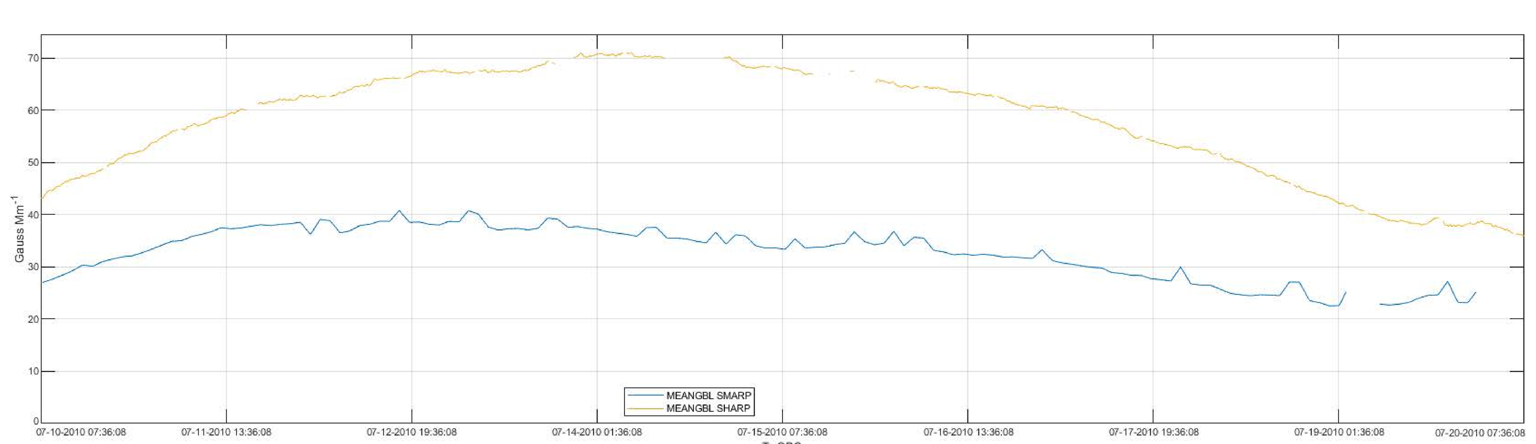
$$\Phi_z = \frac{\Phi_{LoS}}{\cos(\varphi_c - \beta_0)\cdot\cos(\lambda_c)} \qquad \overline{|\nabla_z|} = \frac{\overline{|\nabla_{LoS}|}}{\cos(\varphi_c - \beta_0)\cdot\cos(\lambda_c)}$$

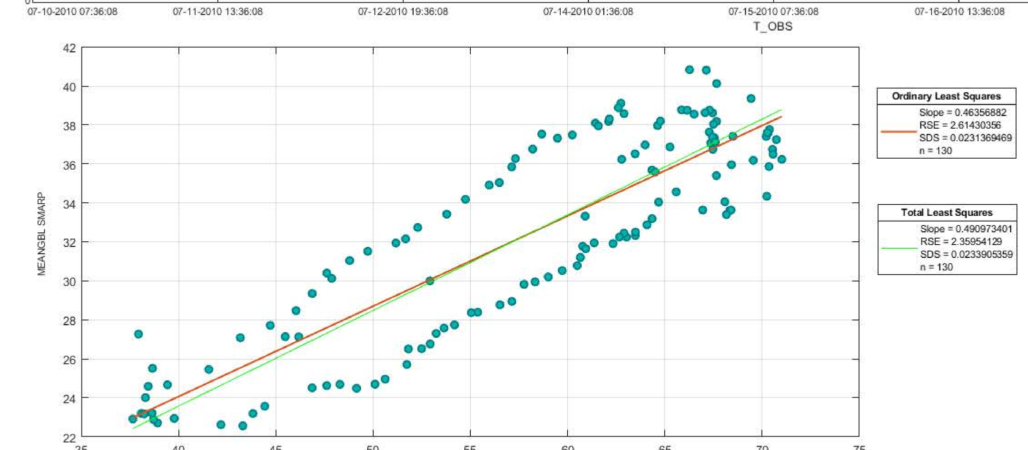
In the second stage, the script re-scales the SMARP keywords using pre-determined scaling parameters (discussed below), and merges them with SHARP time series at May 1, 2010, 00:00:00. As a third and final step, all records are grouped by time stamp, the records with the highest R-value are selected, missing records are replaced with NaN values, and the output is exported into a new .csv file.

RESCALING PARAMETERS

Ten of the most active ARs in the SHARP-SMARP overlap period were selected, based on a combination of peak unsigned flux, sum of the unsigned flux, duration of observation, and data quality. Correlating the data from HMI and MDI involved plotting time-matched SHARP and SMARP values against each other, and applying linear regression using Ordinary Least Squares (OLS) and Total Least Squares (TLS) methods, of which ultimately the latter was selected for rescaling. The Standard Error (RSE) and Standard Deviation of the Slope (SDS) of each regression was computed. Finally, the rescaling parameters (slope and intercept) for each of the ten ARs were tabulated, and their weighted averages were determined for each keyword, using the formula:

$$\mu' = \sum\left(\frac{x_i}{\sigma_i^2}\right)\left[\sum\left(\frac{1}{\sigma_i^2}\right)\right]^{-1}$$





$$RSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{(n-2)}}$$

$$RSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - b - mx_i)^2}{(n-2)(1+m^2)}}$$

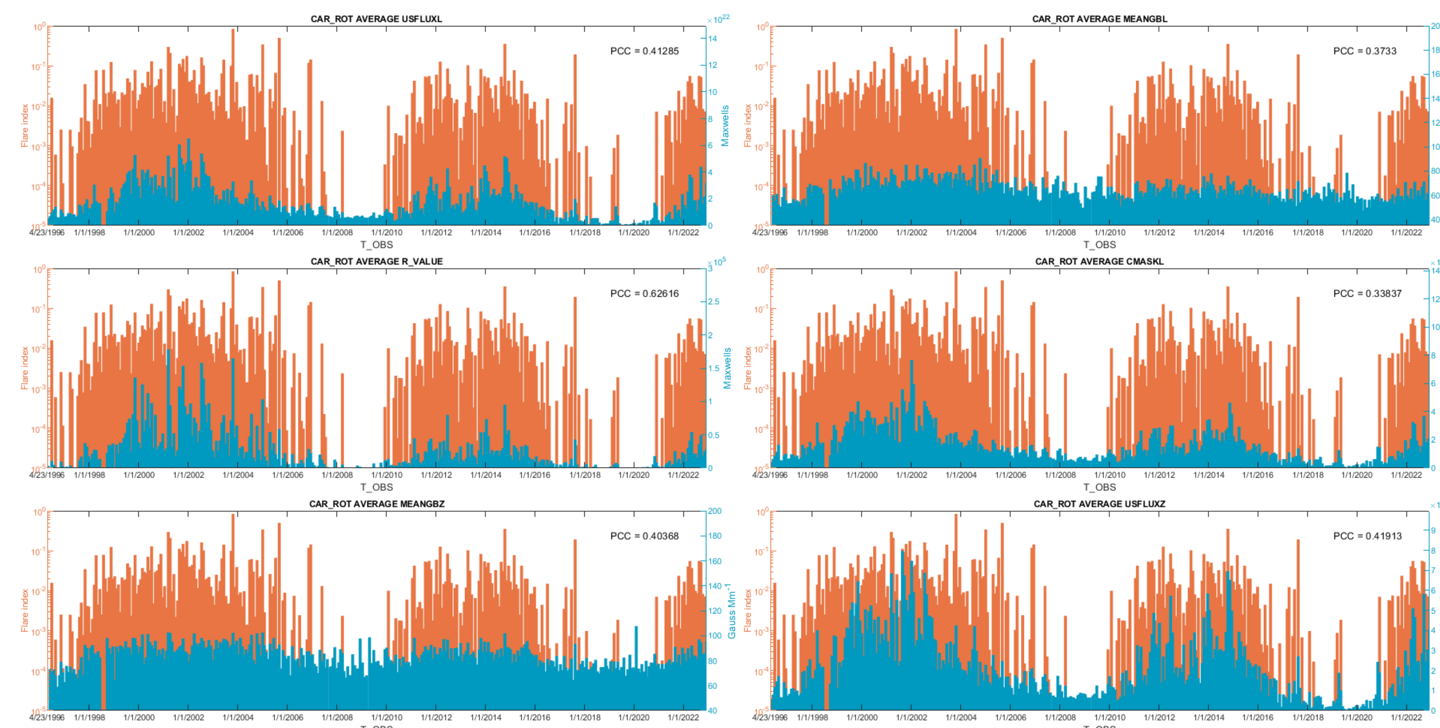
The SHARP and SMARP mean field gradients for NOAA active region 11087 plotted as a time series and on a scatter diagram, with the OLS first-order polynomial regression as a solid red line, and TLS as a green line. The slope, RSE and SDS for both regression methods are also shown. The SDS calculation for TLS is based on Tellinghuisen, 2020.

1. http://jsoc.stanford.edu/ajax/lookdata.html
2. Filtering is based on the QUALITY keyword; see Bobra et al, 2021.
3. φ_c, λ_c = Carrington latitude and longitude of the flux-weighted center of the active pixels.
   β_0 = Carrington latitude of the disk center relative to the observer.

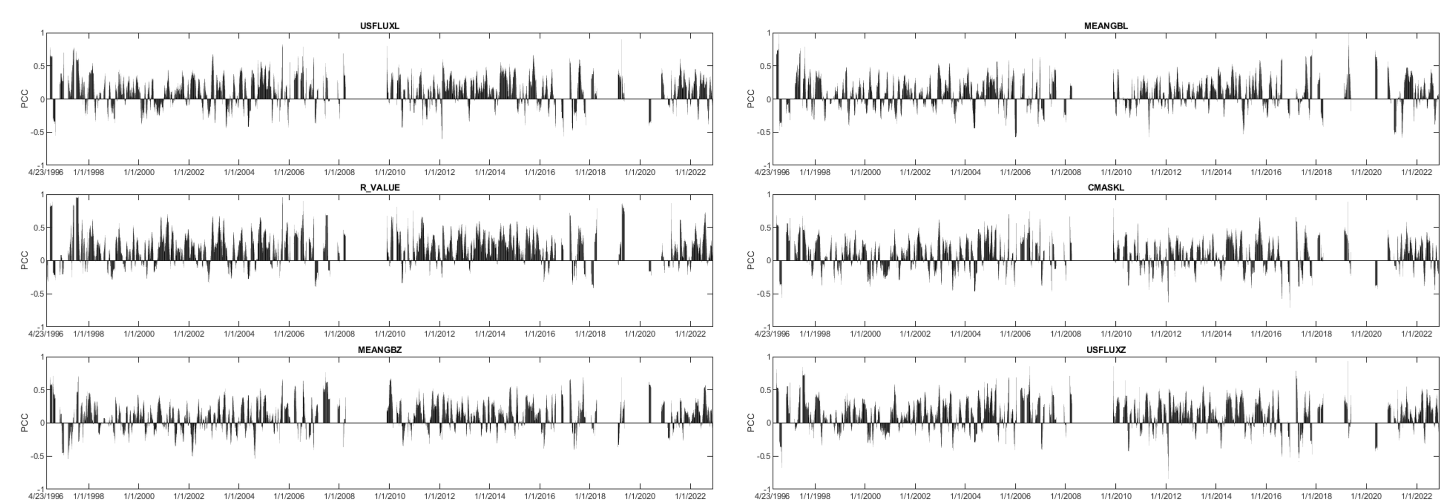
REFERENCES

1. Bobra, M.G., Wright, P.J., Sun, X., Turmon, M.J., 2021. SMARPs and SHARPs: Two Solar Cycles of Active Region Data, Astrophysical Journal Supplement Series, 256(2).
2. Cheong, J. H., 2020. Four ways to quantify synchrony between time series data. https://doi.org/10.17605/OSF.IO/BA3NY
3. Tellinghuisen, J.B., 2020. Least Squares Methods for Treating Problems with Uncertainty in x and y. Analytical Chemistry, *92*, p. 10863-10871.
4. https://www.swpc.noaa.gov/products/solar-and-geophysical-activity-summary

RESULTS



Carrington Rotation averages of the rescaled and merged SMARP and SHARP time series superimposed on the daily flare index, which was calculated from NOAA Solar and Geophysical Activity Summary data[4], and is in units of Joule/m[2]. Solar Cycles 23 and 24 are clearly visible, as well as the start of Solar Cycle 25. The Pearson Correlation Coefficient (PCC) values indicate a significant overall correlation between the keywords and flare activity. The flare index.

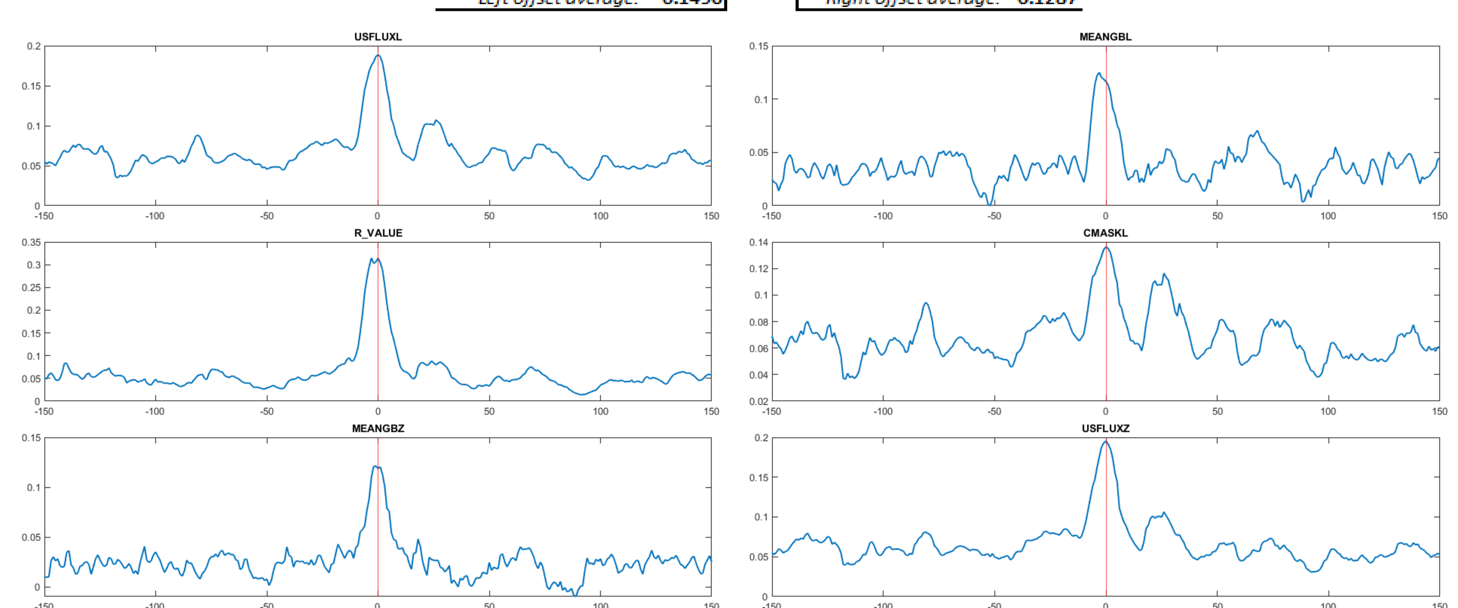


Rolling-Window Pearson Correlation (RWPCC) between daily-averaged keywords and the flare index, computed after performing proximal interpolation, and with a constant window size of 11 days, showing how the correlation score varies with time. The complex oscillating patterns suggest that peak correlation does not necessarily occur at the peaks of the solar cycles, and warrants further investigation.

	-8	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6	+7	+8	*Mean*
USFLUXL	0.0970	0.1181	0.1391	0.1508	0.1637	0.1708	0.1756	0.1819	0.1842	0.1816	0.1736	0.1589	0.1400	0.1258	0.1025	0.0946	0.0813	0.1435
MEANGBL	0.0574	0.0841	0.1099	0.1273	0.1381	0.1433	0.1438	0.1391	0.1426	0.1342	0.1197	0.1049	0.0991	0.0873	0.0813	0.0670	0.0458	0.1073
R_VALUE	0.1469	0.1874	0.2366	0.2688	0.2950	0.3146	0.3024	0.3054	0.3138	0.3037	0.2867	0.2522	0.2111	0.1766	0.1495	0.1332	0.1116	0.2350
CMASKL	0.0824	0.0949	0.1058	0.1081	0.1140	0.1177	0.1226	0.1274	0.1292	0.1278	0.1249	0.1182	0.1094	0.1027	0.0860	0.0833	0.0747	0.1076
MEANGBZ	0.0748	0.0767	0.0811	0.0972	0.1139	0.1383	0.1570	0.1503	0.1579	0.1493	0.1413	0.1254	0.0990	0.0969	0.0728	0.0643	0.0643	0.1094
USFLUXZ	0.1020	0.1163	0.1316	0.1411	0.1577	0.1713	0.1845	0.1889	0.1916	0.1869	0.1812	0.1694	0.1508	0.1394	0.1050	0.0959	0.0869	0.1471
Mean	0.0934	0.1129	0.1340	0.1489	0.1637	0.1760	0.1810	0.1822	0.1866	0.1806	0.1712	0.1548	0.1349	0.1214	0.0995	0.0897	0.0774	0.1417

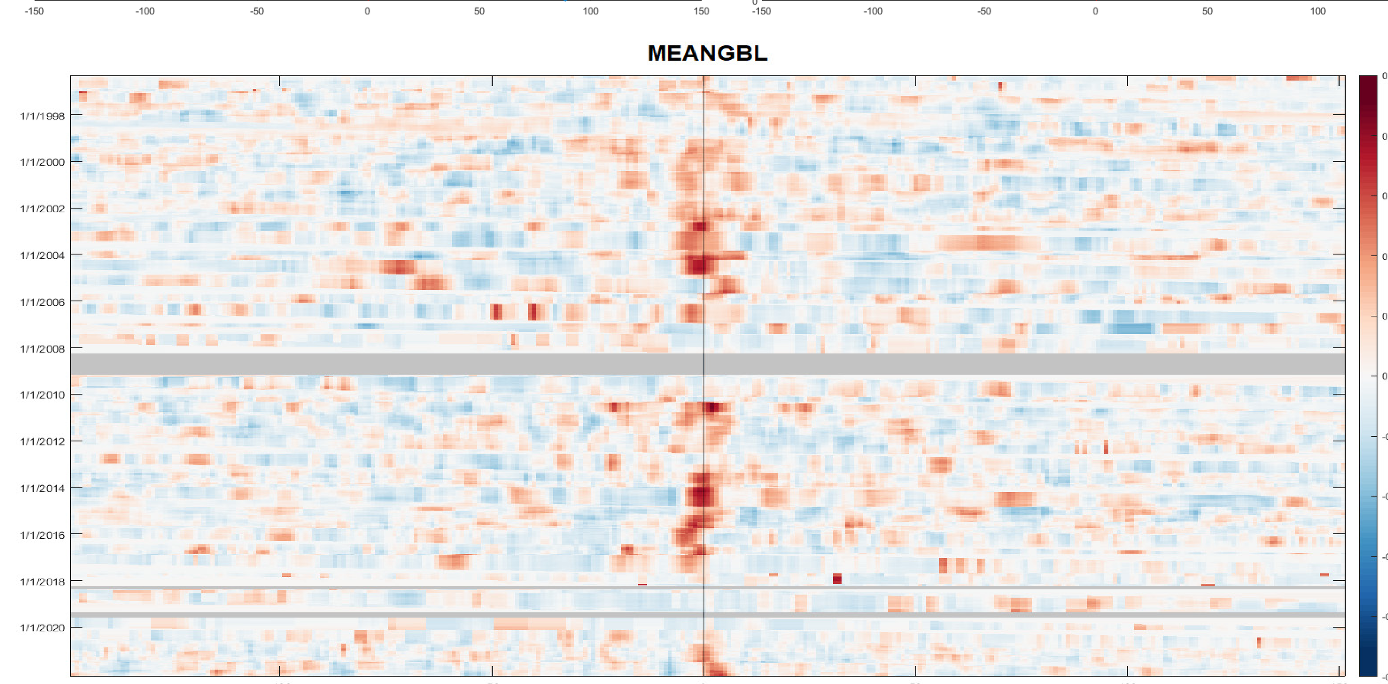
Left offset average: 0.1490 *Right offset average:* 0.1287

The PCC of daily-averaged keywords and the time-lagged daily flare index, which is offset in one-day intervals, as indicated by the top row. When negative time lags are introduced, the average PCC scores tend to be higher, suggesting that the keywords lead the flare index.



In order to examine the keyword-flare dynamic more closely, Time-lagged Cross-Correlation (TLCC) of daily-averaged keywords and the daily flare index is computed by applying the xcorr() function in Matlab, after performing proximal interpolation and subtracting the means of each time series. The horizontal axis shows the time lags in one-day intervals with respect to zero, which is marked by the vertical red line for reference.

A negative peak offset and asymmetry is evident, which indicates that a leader-follower dynamic exists, where the keywords lead the flare index. It is particularly strong with MEANGBL, R_VALUE, and MEANGBZ.



Windowed Time-Lagged Cross-Correlation (WTLCC) can be used in order to assess how the leader-follower relationship varies with time. This method repeats the time-lagged cross-correlation along a time-series pair (in this example, MEANGBL and the flare index), which is divided into sequential time frames, or epochs. It can be thought of as an adaptation of the rolling window technique to the TLCC method. The result suggests that, although MEANGBL leads the flare index for the majority of the active solar cycle, this relationship is not constant.

ACKNOWLEDGEMENTS

Joel Tellinghuisen – *Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37240, USA.*

This research was supported by NASA Early Stage Innovation program grant 80NSSC20K0302, NASA LWS grant 80NSSC19K0068, NSF EarthCube grants 1639683, 1743321 and 1927578, and NSF grant 1835958. VMS acknowledges the NSF FDSS grant 1936361 and NSF grant 1835958.