Statistical Prediction of Solar Flares Using Line of Sight Magnetogram Data

J. Knoll, Mentors: K.D. Leka and G. Barnes
Northwest Research Associates, Colorado Research Associates Division

Abstract

Solar flare prediction is becoming increasingly important as humans increase their presence in space and their dependence upon flare-sensitive activities on earth. With the push for manned missions to the moon and Mars, the advent of space tourism, and the increasing numbers of polar flights, a reliable predictive measure for solar flares is sought by governments and private industry.

I analyze a large SOHO MDI dataset of line-of-sight magnetograms (prepared by J. McAteer) using the statistical technique of discriminant analysis. I investigate dataset limitations and various corrective methods, as well as how the predictive power for solar flares varies with parameters chosen, year of data collection, and distance of the data from disk center.

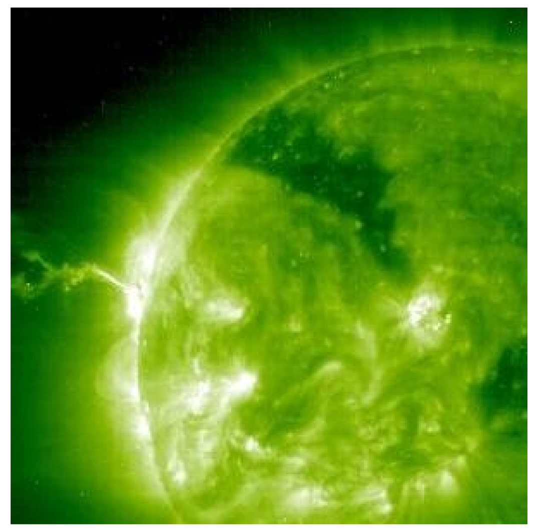
Using line of sight magnetograms is limited because various corrective methods must be employed to compensate for lack of full vector information. The two corrective models used in this study are a simple observing-angle correction and a potential field correction. Parameters that perform best with this data set are the mean of the absolute value of the gradient of the magnetic field and variations on that parameter, as well as the total flux of the active region. The angle-corrected data tend to perform better than the potential field correction, though this is ambiguous.

When analyzed on an annual basis, certain years significantly outperform others. It is unclear whether this trend is due to the number of magnetograms available from year to year or whether this outcome is related in some way to the solar cycle.

Distance from disk center has a profound effect on predictive power, with skill scores falling nearly 200% between including data only within 5 degrees of disk center and including data through 45 degrees. Further research is needed to analyze these findings in more detail.

Solar Flares

Solar flares are explosions in the sun's atmosphere that release tremendous amounts are energy in the form of energetic particles, which, upon reaching Earth, cause harmful effects to satellites, disrupt communication, and endanger astronauts. Unlike coronal mass ejections (CMEs), solar flares cannot be "now-cast" because their effects arrive at the speed of light. Flares must be forecast accurately in order for their effects to be mitigated or prevented.



Solar Flare as seen by the EIT instrument on SOHO

Discriminant Analysis

Discriminant analysis is a statistical technique used to classify events into one of two (or more) groups. Used primarily in the social sciences, it been adapted as a method both to predict solar flares and evaluate the efficacy of new parameters posited as solar flare predictors. One of the advantages of discriminant analysis is that it is able to classify objects based on multiple parameters. This is useful for flare forecasting because no single event triggers a flare – rather, it is a combination events.

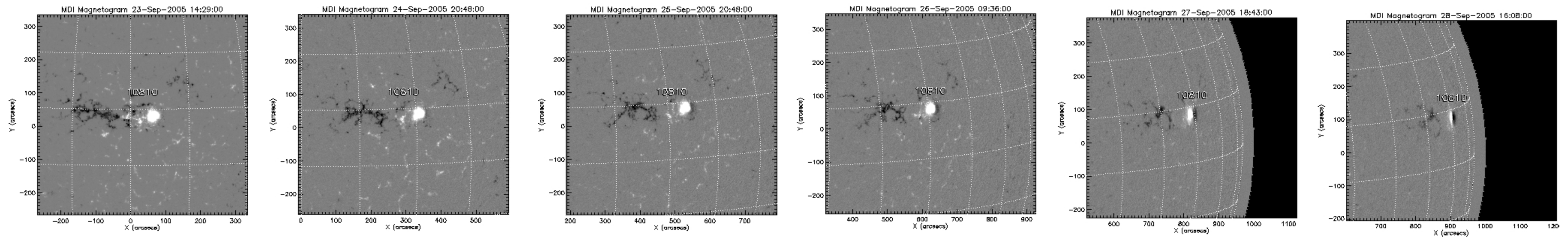
Data

The data used were line of sight magnetograms collected between 1996 and 2004 by the SOHO MDI instrument and processed by J. McAteer. Each magnetogram is a 204 x 204 pixel snapshot centered on active regions identified by NOAA.

There are several advantages and disadvantages in regards to this dataset. The large amount of raw data (nearly 20,000 magnetograms), which is eventually pared down to between six and ten thousand useful magnetograms, means that statistical methods can be more confidently applied to these data than to data previously analyzed using discriminant analysis, such as the Imaging Vector Magnetogram (IVM) data (Barnes and Leka 2003,2005, 2006, 2007).

Disadvantages of line of sight data are that corrective approximations must be used to determine the magnetic field. A standard correction is to divide the observed magnetic field by the cosine of the observing angle, assuming that the magnetic field is perpendicular to the area being imaged. This correction becomes increasingly inaccurate as the observed magnetograms move further from disk center. The series of images following this section demonstrate this clearly.

Another possible correction is to calculate a corrected magnetic field based on the simplifying assumption that every magnetogram observed is a potential field. This correction takes longer to calculate but may be a more accurate representation of the magnetic field far from disk center.

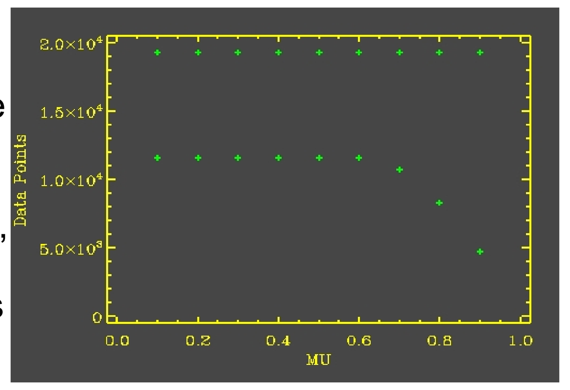


MDI Magnetograms of active region 10810 as it progresses across the disc of the sun. The increasingly large region of "negative" magnetic field at the edge of the region is an artifact of using a simple observing-angle correction.

Data Reduction

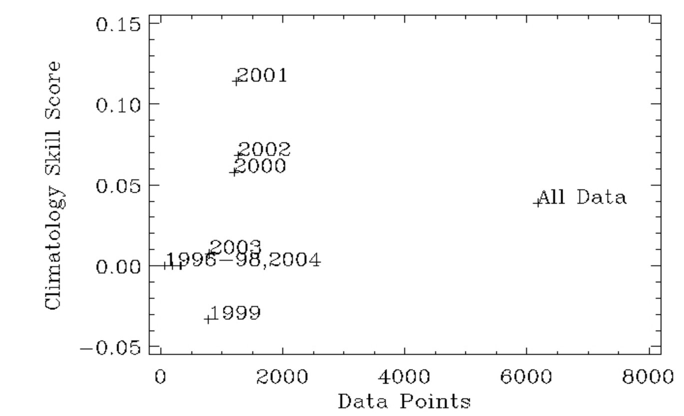
Before being useful for analysis, the dataset had to be cleansed of bad or non-useful data. There were two steps to this process. First, bad instrument data had to be eliminated. This was done by eliminating magnetograms that contained single-pixel magnetic field readings greater than an absolute value of 5000 Maxwells. This is well outside of any possible physical reading, and eliminated 7709 of our original 19296 magnetograms, or roughly 40% of the data. Magnetograms that were eliminated consisted primarily of small slivers of magnetic field where the instrument had failed in some way.

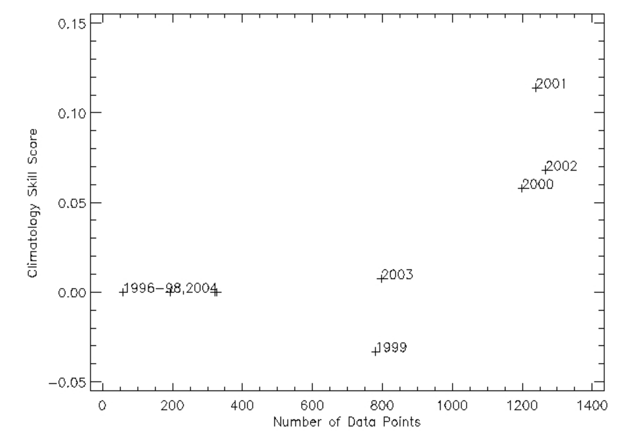
For the second step, I created two IDL keywords that allow the user to define his own unique dataset. The first limits the dataset to only those magnetograms within a chosen distance from disk center. All magnetograms outside this limit are disregarded. The plot to the right, which shows the number of magnetograms available as you move closer to disk center, demonstrates the drastic reduction in available points. The top line of points shows the original number of magnetograms before any reduction, and is therefore constant. If the user does not wish to throw out entire magnetograms, the second keyword allows him to simply zero out part of a magnetogram's data beyond a certain limit from disk center.



Results

Once data reduction and code-checking were complete, my investigation yielded three primary results. First, the predictive power of the discriminant analysis varied annually with this data set, as is demonstrated in the two plots below, which show the climatology skill score (a measure of the success of a prediction versus a null prediction) for discriminant analysis based on total magnetic flux versus the number of data points available for each year at 45 degrees from disk center.



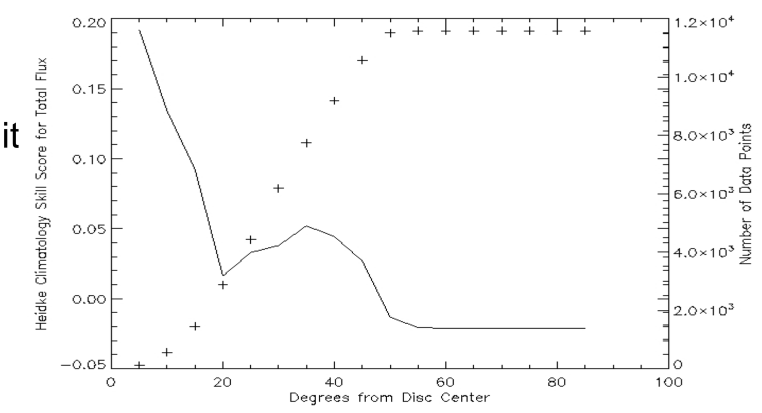


There are two possible explanations for this behavior. First, there is a weak correlation between the number of magnetograms used in the discriminant analysis and the skill score. Therefore, it could simply be that statistically you are more likely to do better with a greater number of data points. However, this trend is belied by the fact that when all of the data is used, resulting in a significant increase in data, the skill score is lower.

The other explanation is that predictive power is linked in some way to the solar cycle. Though this is an attractive explanation, more data across several solar cycles would be required to confirm this hypothesis.

My second major result was a confirmation of the fact that line of sight data becomes more unreliable the further you move from disk center. The plot below shows how the climatology skill score decreases rapidly the further you move from disk center. The scatter plot overlayed and scaled on the right y-axis is the number of magnetograms available. There is a slight rise in the middle as the increased number of magnetograms moderately compensates for the bad data, but this quickly diminishes.

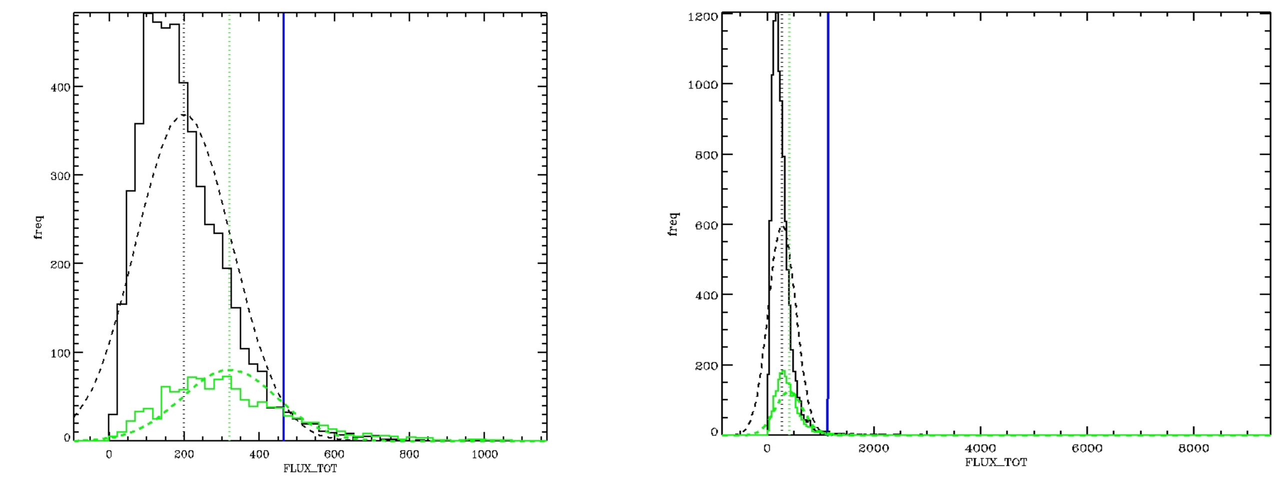
it



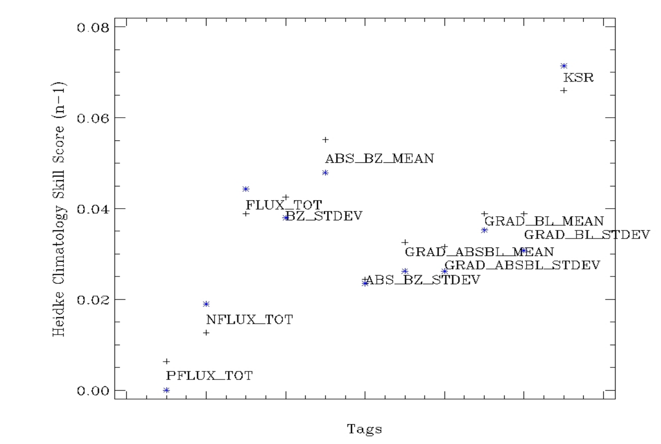
This result is particularly important because much research assumes that it is reasonable to extend line of sight magnetogram analysis to 45 or even 60 degrees using an observing angle correction. This is clearly overly ambitious and may lead to erroneous conclusions.

The following two plots demonstrate just how bad this assumption can be. They show flaring versus non-flaring populations based on the total flux. The black and green histograms are the frequency distributions of non-flaring and flaring populations, respectively, and are overlayed with Gaussian fits. Total flux is in Maxwells and has been scaled by a factor of 10^{20}.

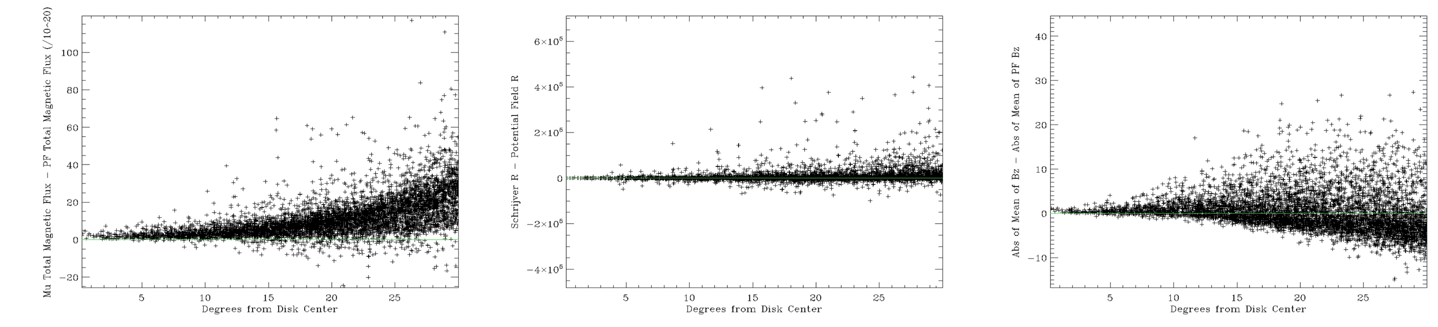
While the plot on the left demonstrates a reasonable range for total flux, the plot on the right is clearly unphysical and a poor approximation of what is actually happening. The difference? The plot on the right only includes data within 45 degrees of disk center, while the plot on the right extends to 60 degrees.



My final investigation looked at the differences between the observing angle correction and the potential field correction. Though it was hypothesized that the potential field correction would perform significantly better in terms of predictive power, the differences turned out not to be statistically significant. The plot to the right shows a comparison between the climatology skill score for the highest-performing parameters using an observing angle correction and a potential field correction. The potential field corrections are shown as blue stars and the observing angle corrections are black crosses. From this plot, it is possible to see that the potential field correction produces a higher score in some cases and that the observing angle produces a higher score in others. However, in no case is the difference large enough to make unassailable conclusions.



It is not accurate to claim that there is no difference, however. Plots of the difference between a potential field correction and an observing angle correction versus the distance from disk center show that the differences increase as you move from disk center. Further research is needed to confirm if the potential field correction results in better flare prediction outside of 45 degrees from disk center.



Summary

Line of sight magnetogram data analyzed using discriminant analysis demonstrated several unique characteristics. First, its predictive power varied annually. Second, it allowed an easy demonstration of the folly of using line of sight data at great distances from disk center. Finally, I was able to show that there are not significant differences between a potential field correction and an observing angle correction within 45 degrees of disk center.

Resources

Barnes, G., Leka, K.D., Schumer, E. A., Della-Rose, D.J. "Probabilisitic Forecasting of Solar Flares from Vector Magnetogram Data." *Space Weather,* 2007

Leka, K.D. And Barnes, G. "Photospheric Magnetic Field Properties of Flaring versus Flare-Quiet Active Regions I-IV." *The Astrophysical Journal,* 2003, 2003, 2006, 2006.

Klecka, W. "Discriminant Analysis."

Acknowledgements

K.D. Leka and Graham Barnes provided the discriminant analysis code and initial code for MDI magnetogram analysis, which I modified. The dataset used was initially constructed by James McAteer.