**Introduction**

Solar activity can have deleterious effects on both terrestrial and space-based systems. Powerful eruptions on the surface of the sun known as solar flares eject high energy charged particles. Solar flares are categorized as A, B, C, M, or X on a logarithmic scale according to their peak x-ray flux (W/m2) in the 0.1 - 0.8 nm wavelength range. M and X class flares are the most powerful and pose the largest threat to infrastructure. Charged particles from such events can interact with Earth’s magnetic field and cause slowly varying currents in the upper atmosphere (ionosphere). Currents in the ionosphere can, in turn, induce damaging direct currents in man-made systems such as transcontinental pipelines and electricity transmission networks. The ability to predict solar flares from observable solar features, such as sunspots, would allow protective measures to be taken on critical systems prior to the arrival of the flare.

It has been shown that sunspot activity is closely related to solar flare emission [1]. Sunspots are regions on the sun’s photosphere where intense magnetic fields cool the surrounding plasma to temperatures that are several thousand Kelvin less than the surrounding photosphere gasses. As a result of this temperature gradient sunspots appear as dark spots. Sunspots consist of a dark core region, the *umbra*, and a surrounding less dark area of higher temperature, the *penumbra*. The international solar physics community classifies sunspot types using the three component McIntosh classification scheme [2]. McIntosh classifications consist of a three letter code typically abbreviated as *Zpc* where *Z* describes the Zurich group type, *p* represents the penumbra type, and *c* represents spot compactness. Examples of the McIntosh classification scheme are shown in figure 1 and allowed classifications are tabulated in table 1. Sunspot locations are given by their National Oceanographic and Atmospheric Administration (NOAA) active region number. Sunspots must be observed by two separate observatories to be assigned a NOAA active region number.

This work attempted to answer two questions; given a sunspot group type will an M or X class flare occur, and if a sunspot group is known to flare what flare classification will be produced. The data was provided by the NOAA National Geophysical Data Center (NGDC). The Solar-Terrestrial Physics (STP) division at NGDC has made Mount Wilson Observatory sunspot group data and Geostationary Operational Environmental Satellite (GOES) solar flare event data available on their ftp site.

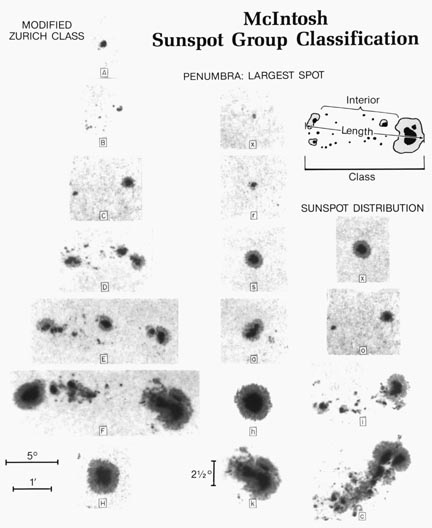


Figure 1: The 3-component McIntosh classification, with examples of each category [2].

|  |  |  |  |
| --- | --- | --- | --- |
| Group Type | Penumbra of Largest Spot | Spot Compactness | Number of Types |
| A | x | x | 1 |
| B | x | o, i | 2 |
| C | r, s, a, h, k | o, i | 10 |
| D, E, F | r | o, i | 6 |
| D, E, F | s, a, h, k | o, i, c | 36 |
| H | r, s, a, h, k | x | 5 |

Table 1: Allowed types of groups in McIntosh sunspot classification. There are a total of 60 allowed classifications.

**Data Preparation**

A Python script (goes\_sunspot\_parser.py) was used to gather data for both measurements from 1982 to 2014. X-ray data consisted of measurement date, x-ray flare class, and NOAA active region number where as sunspot group data consisted of measurement date, the three component McIntosh classification, and NOAA active region number. Two separate data frames containing sunspot group data and x-ray event data were created.

Raw x-ray event and sunspot group data were cleaned using a Python script entitled sun\_x\_ray.py. It was decided to only investigate M and X class solar flares, as these events pose the most severe threat to infrastructure. The sunspot data was parsed so that only sunspot observations with NOAA active region numbers present within the x-ray event data frame were kept. A sunspot group tended to have multiple observations over the course of a day. Each solar flare observation, however, was unique. A data frame of solar flare measurements and associated sunspot observations was created by matching sunspot groups and flare events using the NOAA active region number. If a flare and a sunspot group shared a common NOAA active region number, the time difference between the flare event and each sunspot measurement in the group was calculated. Sunspots that are observed within six hours of a flare event were considered to be associated with the flare. If multiple sunspot observations were made within six hours of a flare measurement, only the observation with the minimum time difference was considered to be associated.

The associated sunspot group observations were merged with the appropriate solar flare observations and written to a data frame, sunspot groups that were not associated with a solar flare observation were written to a separate data frame. The final products were two data frames, the first consisting of the associated solar flare class and the three component McIntosh sunspot classification (Table 2), and the second consisting of the three McIntosh labels for unassociated sunspots.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Zurich | Penumbra | Compactness | Flare Class |
| 1981-12-28 20:28:00 | D | r | o | M |
| 1982-01-08 13:54:00 | E | k | o | M |

Table 2: Associated solar flare/sunspot data example.

**Analysis**

Analysis is currently in progress, both naïve Bayes and logistic regression have been applied to the dataset to predict if a flare will occur given a McIntosh sunspot classification. See the script entitled flare\_prediction.py for further details.

**References**

1. H. Zinn and M.A. Liggett, *Sol. Phys.* **113**, 267 (1987)
2. P.S. McIntosh, *Sol. Phys.* **125**, 251 (1990)