PREDICTING SOLAR PROTONS: A STATISTICAL APPROACH

THESIS

Jonathan C. Spaulding, Captain, USAF

AFIT/GAP/ENP/09-M09

AFIT/GAP/ENP/09-M09

PREDICTING SOLAR PROTONS: A STATISTICAL APPROACH

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Jonathan Spaulding, BA

Captain, USAF

March 2009

PREDICTING SOLAR PROTONS: A STATISTICAL APPROACH
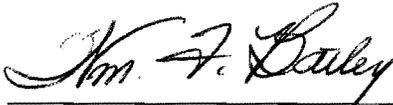
Jonathan Spaulding, BA
Captain, USAF

Approved:

_____   _____
Ariel Acebal (Chairman)   19 MAR 09
                                                                   Date

_____   _____
William Bailey (Member)   19 Mar 2009
                                                                   Date

_____   _____
David Kazizka (Member)   19 Mar 2009
                                                                   Date

AFIT/GAP/ENP/09-M09

Abstract

A small fraction of solar flares are accompanied by high energy (>10 MeV) protons. These events can cause degradation or failure of satellite systems and can be harmful to humans in space or in high altitude flight. For risk management purposes, the Air Force is interested in predicting these events. Several algorithms exist to do this operationally, but none predict when these events will occur with much accuracy. Here, we analyzed 3514 M1 and greater flares including 106 with proton events from the GOES sensors from 1 Jan 1986 to 31 Dec 2004 to produce new results, including a full scale comparison and optimization for all the algorithms. In every case, optimization leads to increased prediction ability. This research also produced a new algorithm based on the Garcia algorithm, which functions better than any other operational algorithm. This model, Garcia 2008, predicts with a skill score of .526, an improvement from .342. This new model is the best at prediction of all models measured.

Acknowledgments

First, I would like to thank my advisor, Maj A. Acebal, without whom this thesis would never have come to be. Second, I would like to thank my committee and Lt Col Smithro, who gave me the idea for this project. Without his aid, this project would be but a shadow of itself.

Next, I thank Dr. C. Balch, whose SEP database provided the information upon which all this analysis is based. His work saved tens if not hundreds of hours of data compilation and testing. I hope this analysis helps!

Finally, thanks to all the students and faculty at AFIT, for support and aid during this time.

Jonathan C. Spaulding

Table of Contents

List of Figures

List of Tables

PREDICTING SOLAR PROTONS: A STATISTICAL APPROACH

I. Introduction to Solar Energetic Protons

Solar Flares and Solar Energetic Protons

The Earth is periodically bombarded with energetic protons from space, some

with energy above 10 MeV, well above the background flux of protons (Kahler and

Vourlidas, 2005). These are known as solar energetic protons (SEP). Some of these

groups of protons arrive shortly after flares, and some after coronal mass ejections (CME)

(Balch, 2008b; Kahler, 1996). When a solar flare occurs, it has a chance to be

accompanied by protons accelerated away from the site of the flare. Not all flares are

accompanied by SEP events that are observed at Earth; however, those that do are of

great interest. It is extremely important to be able to predict when SEP will impact Earth.

These particles can cause degradation to semiconductor systems and are harmful to

humans (Getly et al., 2005). Satellites can be shutdown or reoriented to minimize

exposure. Aircraft can be grounded for the duration of the event. Particle fluxes (number

per unit area per time) and fluences (total number of protons per area received) are of

great interest to both military and civilian agencies for reasons such as high flying aircraft

(Beck et. al., 2005), manned missions to the moon or to Mars (Smart and Shea, 2003), the

International Space Station, or even tourism in space (Collins, 2005). Since neither flares

nor CMEs can be predicted, predicting which flares, once they happen, are likely to be

accompanied by SEP will allow scheduling to take advantage of times when the predicted

flux of SEP is low.

Definition of SEP Events Associated with Flares

SEP events are defined as when significant numbers of energetic protons are

measurable at Earth. The background for proton flux is .15 particle flux units (pfu)

measured with the Geostationary Operational Environmental Satellite (GOES) sensors in

orbit (Kahler and Vourlidas, 2005). One pfu is one particle per square centimeter per

steradian per second. The largest SEP events have fluxes of protons that exceed 40,000

pfu (Balch, 2008a). The National Oceanographic and Atmospheric Administration

(NOAA) classifies an event as a SEP event if proton flux exceeds 10 pfu of 10 MeV

protons for 15 minutes at the altitude of geostationary orbit. "This is one to two orders of

magnitude above background levels, and represents the lowest level where radiation

hazard analysis is needed for manned spacecraft missions," (Balch, 1999). This is the

international standard (Xiaocong, 2001; Garcia, 1994a) as well as a level the Air Force

tracks (Kahler, 1996; Cliver and Ling, 2006).

Production of SEP

The exact relation between solar flares and solar energetic proton events (SEP) is

not fully understood (Garcia, 2004a; Garcia, 2004b; Balch, 1999; Balch, 2008b). This is

complicated by the fact that flares are only half of the equation. Coronal Mass Ejections,

or CMEs, also play a strong role in determining SEP production. Studies (Kahler and

Vourlidas, 2005; Kahler, 1996; Kahler et al., 1984) have found that there was "a high but

not perfect association of prompt proton events with CMEs," (Kahler and Vourlidas,

2005). One conclusion from this research was that "a CME may be a necessary requirement for the occurrence of a flare proton event" and " the CME acts as a driver to set up a coronal shock in which protons are accelerated," (Kahler et al., 1984). In this research, CME effects will be addressed by the presence or absence of Type II and Type IV radio data, the radio signatures of a CME in the solar atmosphere. This is for consistency, as the more informative CME speed data is not always available before 1996, when the CME observational satellite SOHO LASCO starting producing data (Yashiro, 2008). With all this data, the goal of this research is to find that flares which are accompanied by SEP have certain characteristics, such as intensity, temperature, and presence of a CME, that separate them from other flares.

Damage Due to Solar Protons

It is vital to both manned and unmanned mission for SEP events to be predicted properly. Aside from the obvious radiation damage to both astronauts and high altitude pilots, spacecraft can receive damage.

SEP events are responsible for rendering many satellites inoperable. On 28 Oct 2003, the Japanese government lost contact with an experimental communications satellite, the Data Relay Test Satellite. The satellite went into 'safe' mode, shutting down all but essential functions. "The excessive signal noise coming from the Earth sensor assembly suggests the satellite was affected by a proton barrage," says Katagi, associate executive director of the Japanese Aerospace Exploration Agency, and "The most likely culprit is the solar flare," (Kallender, 2003).

The U.S. spacecraft Stardust, designed to collect comet dust, survived a hit from a SEP event 9 Nov 2000. The satellite, overloaded by the protons, shut down operations.

3

The spacecraft protected itself as designed, turning in place so that its solar panels were pointed at the sun and the cameras away from it. However, even with these protective measures, temporary loss of functionality occurred. Later inspection showed the cameras were receiving data from stray protons even in the areas normally shaded. Its main star cameras, used to locate itself in space, had both failed. Scientists left the craft in standby mode until the protons had passed, then ordered it to reboot. Images taken after the proton event showed the camera was "completely recovered from the proton hits," (Heil and Roseth, 2000).

Predicting SEP Events

No current prediction method works perfectly, and the current models take into account only a few of the many measurements recorded for each flare; such as the temperature, soft x-ray flux, or radio data. There is much more information available, so advanced statistical techniques may perform better by taking into account all the additional pieces of information.

First, I will cover the physics of flares and current practice in prediction, such as the model used by NOAA and the Air Force's own AF-Geospace. Then, I will discuss how a new model is created and its predictive ability is compared to other models. Following that, I will examine how well the new model performs versus the previous models, in both their original forms and in optimized form. Finally, I will discuss the conclusions of this analysis, the recommendations it makes for predictions, and recommendations for further study.

II. Background

Solar Flares

The sun has a strong magnetic field at the surface on the order of a few thousand Gauss, driven by unknown processes deep in the radiative zone (Aschwanden, 2004). The sun's field reverses itself over a very predictable period of about 11 years due to magnetic field lines twisting in the differential rotation (Foukal, 2004). Current theories state that the differences in rotation between the poles of the sun, the equator, and the subsurface zones cause complex patterns in the magnetic field at the surface of the sun as field lines twist about each other, yielding areas of high magnetic energy density, energy that is stored in the field (Foukal, 2004). Field lines of opposite polarity pointing into and out of the sun's surface are forced near to each other by differential rotation. The lines, originally anti-parallel to each other, break and reconnect in an X pattern instead. The new lines form an arch on the sun's surface and a curve with ends in interplanetary space that bends towards the sun, nearly meeting the top of the new arch. This reconnection releases enormous energy and is known as a solar flare (Foukal, 2004). Flares can generally only be detected by observations in the chromospheric and coronal radiations (soft and hard x-rays, radio and extreme ultra-violet), which brighten considerably more than the visible light photosphere. Exceptions to this rule are the white-light flares, where even the photosphere brightens enough to be measured in visible light on the ground. The most famous example of this was the first sighting of a flare by Carrington and Hodgson in 1859 (Foukal, 2004).

Flares and Protons

The release of energy from magnetic fields at coronal levels (high above the surface) dissipates mostly as heat, increasing the velocity of protons and electrons found in the corona. Particles are accelerated away from the reconnection site. Those electrons which are accelerated down the field lines toward the sun slow as they hit the denser regions of the chromosphere, emitting bremsstrahlung, or braking radiation, as they do. It is this bremsstrahlung which produces the hard x-rays measured by satellites. The dissipation of electron velocity as heat warms the chromosphere, increasing the radiation produced there, generating the soft x-rays observed (Foukal, 2004). If protons are present at the reconnection site, they can be accelerated along the field lines and travel out into space, and can be observed at Earth if conditions in the interstellar medium are right.

The current model of flare development, shown in Figure 1, shows a flare progressing from opposing field lines tightly packed next to each other (a), then reconnection at the X point, front (b) and side views (b'), showing the new field lines and the rising prominence that is a feature of the solar surface, and finally the late phase (c) where the new field configuration becomes stable (Aschwanden, 2004). In the final picture there are holes in the solar surface, where the chromosphere has been 'boiled' away by energetic particles, heating the surface until it produces the x-rays that characterize the flare.
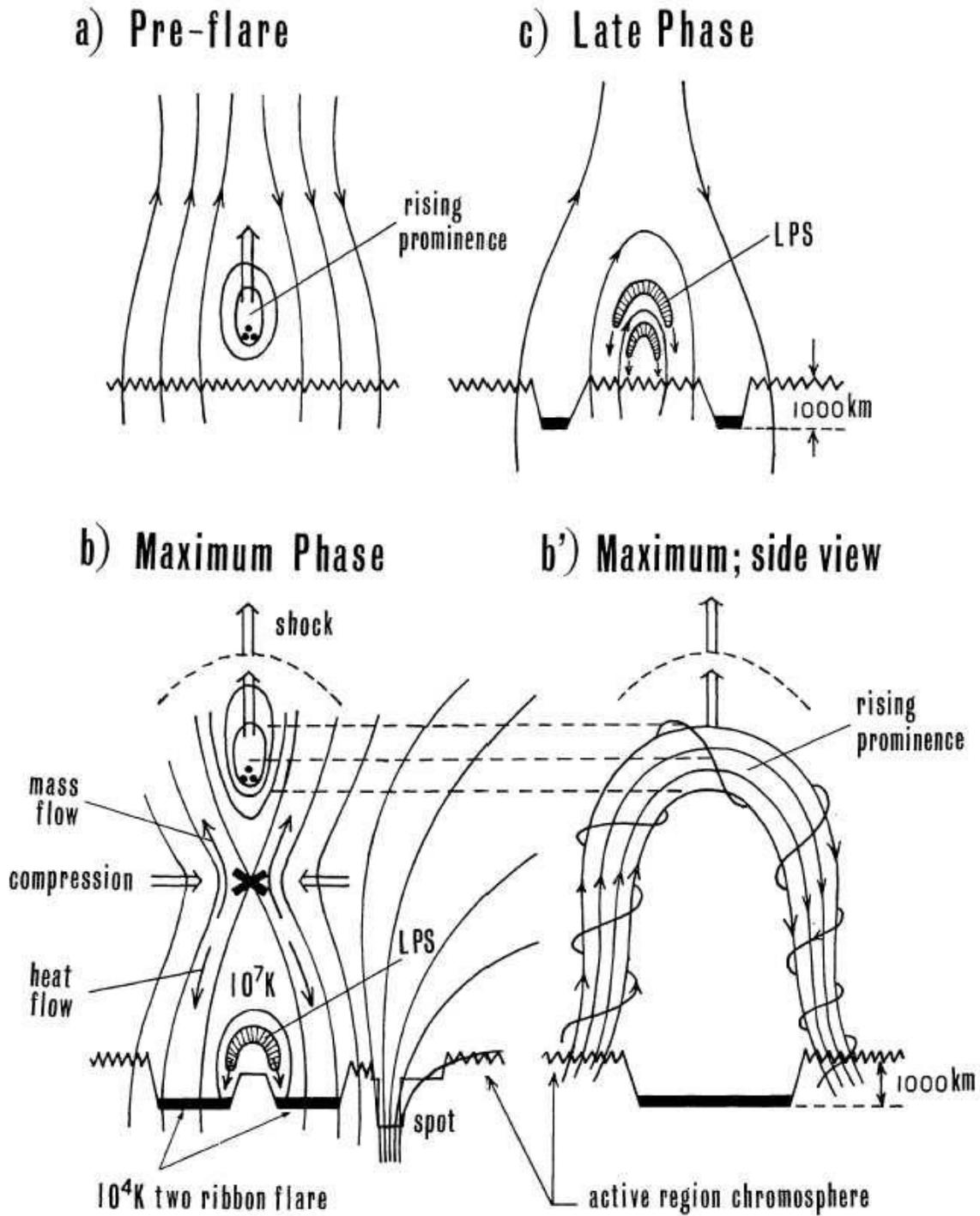
Figure 1. The Standard Model of Solar Flare Production (Used with Permission from Aschwanden, 2004)

Flare Temperature

Solar flares produce extreme temperatures in the range of 10-40 MegaKelvin (MK). Though the notion of temperature is rather vague when applied to any sparse gas, the flare has distinct features which make this determination useful. A normal gas at some defined temperature will have emission lines with a defined strength. As temperature changes, the emission lines from that gas will change in intensity relative to each other. As the temperature rises, lines which were not present before will show up as the average kinetic energy of the particles ionizing or exciting them rises. For example, the Fe VI lines (iron, ionized five times) will first show up at higher temperature than the Fe V (iron, ionized four times) lines. As the kinetic energy keeps rising, previously strong lines will weaken as fewer atoms are ionized to only that state. There will be fewer Fe V lines if most of the iron is ionized at least five times. By taking a ratio of line strengths, algorithms can generate an effective temperature for a gas and thus for a flare as well (Garcia, 1994b).

The Mewe Temperature

This temperature calculation can be simplified by looking merely at broadband x-ray flux, and taking the ratio of two bands to each other. When used on GOES data, this method results in the Mewe temperature for a flare, used in the new predictive algorithm.

The process to calculate the Mewe Temperature of a flare from the GOES data is straightforward. Take the ratio of the long x-ray flux (1 - 8 angstrom) to the short x-ray flux (.5 - 4 angstrom). This ratio R is entered into the equation

$$T(R) = A(0) + A(1)R + A(2)R^2 + A(3)R^3 \qquad (1)$$

The coefficients A (0) through A (3) are available in the literature but are different for each GOES (White et al., 2005). Each is also different for coronal and chromospheric emissions, given the different elemental abundances there, but since flares emitting primarily in the corona, the coronal abundances will be used to calculate temperature. The coefficients are reproduced in Table 1:

Table 1. Coefficients for Mewe Temperature Calculation

Satellite	A (0)	A (1)	A (2)	A (3)
GOES 6	3.83	86.2	-193.3	242.1
GOES 7	3.68	101.2	-271.3	409.3
GOES 8	4.02	100.3	-257.1	366.5
GOES 10	3.81	101.5	-270.7	404.6
GOES 12	3.90	101.2	-266.4	390.2

Thus, a flare measured by GOES 6 with long x-ray flux of $1.83*10^{-4}$ W/m^2 and short x-ray flux of $5.79*10^{-5}$ W/m^2 would have a ratio of .316 and thus a Mewe Temperature of 19.4 MK.

Integrated Flux

Another predictor used by forecasting algorithms is the integrated flux. This is a measure of how strong the flare is over time. As the GOES data is not instantaneous but available in one minute intervals, the integrated flux is the sum of the one minute intervals from flare beginning to flare end, corrected by a factor of 60 to change from minutes to seconds. Flares are classified as ending after their flux has fallen to half the peak value (Balch, 1999).

The Importance of Location

The sun's strong magnetic field penetrates out into interplanetary space well past the Earth. These magnetic field lines influence the travel of all particles, in particular SEP. As the protons leave the sun, they travel radially outward, but are accelerated around the field lines in corkscrew fashion. The field lines themselves are not straight, but rather spiral out of the sun, victim to the sun's rotation. This pattern is known as the Parker spiral (Tascione, 1994).

Since charged particles like protons are constrained to follow field lines through space, SEP can only reach Earth when field lines connect the Earth to the sun. Field lines connect the west side of the sun to the Earth (because of solar rotation), so no particles originating from flares on the east side should be expected to be seen at Earth; the magnetic field lines have led those particles off into interplanetary space. This matches observations; few flares in the extreme eastern areas of the sun produce SEP (Garcia, 2004a).

Not all flares have a well defined location. Without some imaging system observing the solar disk during the flare, such as visible light or x-ray imagers, the location remains unknown. Approximately 18% of flares in the Balch database have no known location (Balch, 2008a). Any predictive algorithm will have to be able to forecast for these flares with null location data.

X-Ray Flare Categorization

The categorization of flare by x-ray flux is based on flux. The flare's maximum flux in the x-ray region is measured in watts per meter^2 by the GOES in two channels, the .5 - 4 angstrom and 1 - 8 angstrom bands. Flares with a peak x-ray flux between 10^{-6}

W/m^2 and 10^{-5} are given the category 'C' for Common. Flares with flux between 10^{-5}

and 10^{-4} are give 'M' for Medium, and flares above 10^{-4} are called 'X' for eXtreme.

These categories are further separated by the numerical value of the flux; the category is

followed by a number specifying the flux. Therefore, a flare with maximum x-ray flux of

5*10^{-6} would be labeled a 'C5' flare. Class C flares are common, with tens or hundreds

of these per M flare. Two additional categories called B and A exist for the flares one

and two orders of magnitude smaller than C. Class C and lesser flares are also the least

likely to be associated with SEP events (Balch, 2008a).

Classification of SEP Events Associated with Flares

One difficulty with reporting SEP events lies in the fact that there is no perfect

way to associate a SEP event with a particular flare. Flares occur quite often. Sensors

(such as the GOES) detect a rise in protons. SEP have a wide range of time to impact

from the beginning of the flare, making predictions of arrival time difficult at best. Some

impacted within 40 min of the observance of the flare itself, while others range out into

the hundreds of hours (Garcia, 2004a; Balch, 1999). Thus, associating protons with a

flare to produce ground truth is a tricky business. The data used as ground truth for this

analysis comes from Dr. C. Balch at NOAA, compiled over eighteen years of flares. The

observed protons have been associated as best possible with the flares that probably

produced them.

The next problem with reporting and classifying SEP events comes when a flare

happens during a period of already high flux, whether from a previous flare, coronal mass

ejection, or other event. In this case, the flare may not have produced enough protons to

be normally classified as a SEP event, but with the previously high proton flux, the total

proton flux is over 10 pfu. The events are referred to as 'Enhancements' and are NOT

classified as SEP events in this analysis, for the event associated with that particular flare

would not be a SEP event without aid. Thus, any prediction must classify this type of

flare as non-SEP if future flares are to be correctly identified.

Coronal Mass Ejections

Coronal Mass Ejections (CMEs) appear to be important if not essential for a SEP

event to occur (Belov et al., 2005; Kahler and Vourlidas, 2005; Kahler et. al., 1984;

Kahler, 1996). This important role should not overshadow the valuable information for

predictions to be gained from flare associations with SEP. This research will focus on

those protons which are associated with flares, so the CME or absence thereof will be

used as a predictor for SEP alongside flare characteristics.

Physical Characteristics

The sun regularly erupts with huge masses of coronal plasma, called Coronal

Mass Ejections. CMEs of varying sizes occur at a variable rate of up to several per day.

Propagation speeds in the interstellar medium vary between 200-1000 km/sec. Though

widely varying in size, these eruptions of coronal material average 10^{15}-10^{16} g in mass,

with a size comparable to half the radius of the sun, about $3*10^8$ m. As much as 10% of

the mass contained in the solar wind may be the direct result of CMEs (Foukal, 2004).

Space-based coronagraphs are used to detect CMEs. To measure the minimal

brightness associated with the CME, an occulting disk is placed over the image of the

sun, allowing very low levels of scattered light with brightness on the order of 10^{-6} of the

sun's disk brightness to be observed and measured. "Statistical studies indicate that only

about half of the CME's observed can be associated with flares or filament eruptions on

the visible disk," (Foukal, 2004). Many more CMEs may come from the rear half of the sun, yet the mechanism for the production and energy associated with these events is not conclusively identified (Foukal, 2004).

Radio Data from CMEs

Thousands of CMEs have been studied since they were first filmed in 1971 (Foukal, 2004). The pressure wave of the moving front of dense mass solidifies into a shock front near .3 astronomical units (AU) distance from the sun, and sometimes up to 1 AU (the location of the Earth). This shock front is known to produce Type II meter wave bursts, also known as Type II radio sweeps. A Type II radio signature consists of two distinct bands which drift to lower frequency over a time scale of a few minutes. These bands are usually interpreted as the fundamental and first harmonic plasma oscillations due to a disturbance in the corona. The speed that these frequencies require (500-5000 m/sec) are well beyond the local speed of sound waves, so the disturbances must be shock fronts travelling outward. Type II radio data is known to be closely associated with the more energetic CMEs (Foukal, 2004).

The ejected plasma travels through space towards Earth. Near the sun, it travels through the sun's magnetic field for some tens of minutes as the burst travels outward the first few solar radii, emitting radio signals as it does. This is known as Type IV radio data. These radio burst can extend in frequency from the microwave region down to about a hundred kHz. The most studied region is between 10-100 MHz. It is generally accepted that Type IV radio data is the signature of a CME (Foukal, 2004).

This study uses Type II and Type IV radio data as a proxy for CME occurrence since continuously available CME data is only available starting in 1996 (Yoshiro, 2008).

CMEs can occur without any radio data, and radio data can in some cases occur without a CME. This study focuses on solar flares, however, and only tracks radio data within a six hour window near a flare.

The Distribution of Flares

Flares with SEP and without SEP

Flares occur at all locations and at all magnitudes on the sun. Though this analysis only tracks M1 and greater flares, flares of all magnitudes occur. The majority of flares occur without any radio data, either Type II or Type IV (Balch, 2008a). Of the 3610 flares tracked, there were only 484 Type II events and 344 Type 4 events. However, flares with radio data of either sort are more likely to be associated with SEP events. Shown in Figure 2 is a chart of all the tested flares, separated into categories by SEP event occurrence. Each category (SEP or no SEP) is separated by whether or not the flares had any radio data. Flares associated with SEP are twice as likely to have radio data as those without.
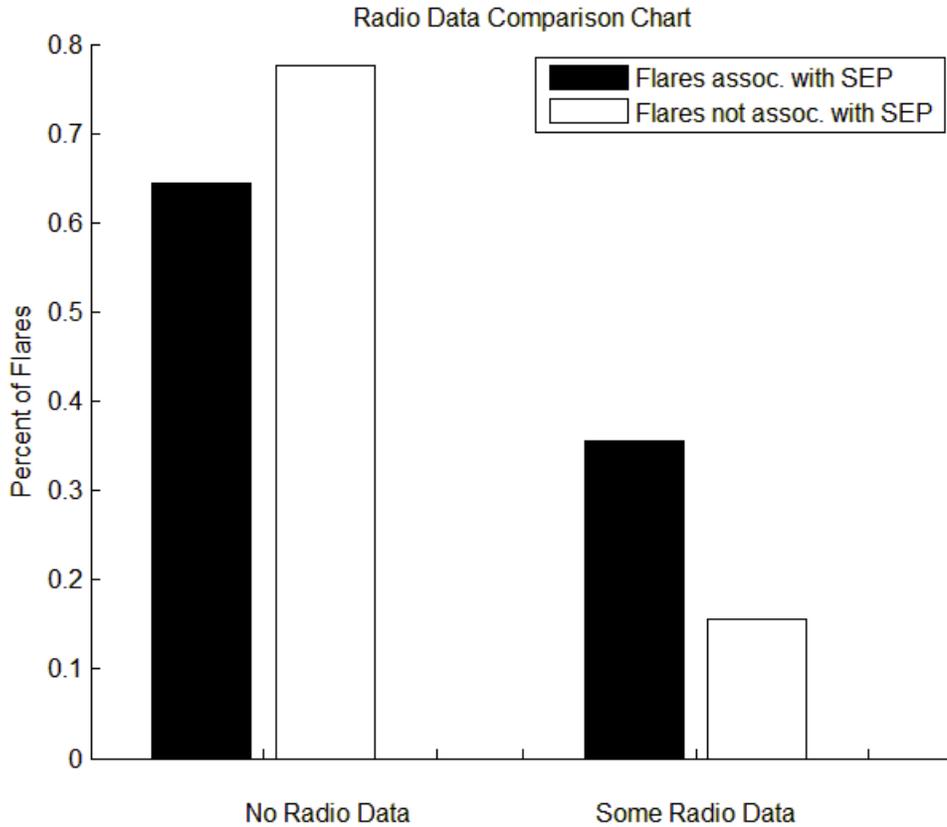
Figure 2. Fraction of Flares With and Without Any Radio Data

Flares themselves are evenly distributed across the solar disk, as is shown in Figure 3. Flares associated with SEP, however, are more prominent in the west. This shows the importance of flare location mentioned earlier, as magnetic field lines connect the west side of the sun to the Earth. Flares in the west are nearly twice as likely to have SEP events as those in the east.
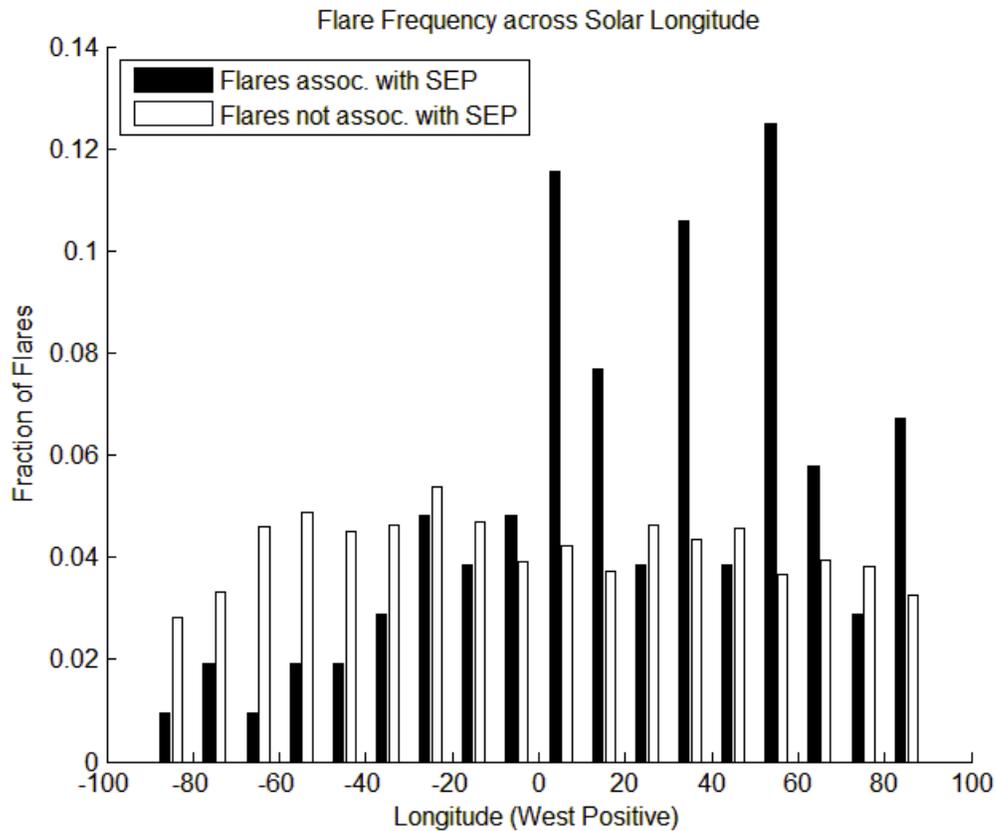
Figure 3. Flare Frequency across Solar Longitude

These graphs show two of many methods that flares associated with SEP can be separated from those without SEP. Further analysis will reveal other methods and their value for classification.

Current Practice in Prediction

There are several methods used to predict SEP events. The Air Force uses the Proton Prediction System (PPS), NOAA Space Weather Prediction Center uses the Proton Prediction Model (protons or PPM), and the Garcia model is no longer in operational use. As PPS, PPM and Garcia were all used by the Air Force for predictions,

these three will be the focus of this report. Also in the literature but not analyzed here are the JPL proton fluence model, which fits proton fluences to a log-normal distribution; and the Xapsos model, which uses the maximum entropy principle to generate a power law distribution for peak intensity of SEP events (Kahler et al., 2007).

Garcia Model

The first method of prediction to be considered is the Garcia model. The Air Force Weather Agency used the Garcia model in a web-based interface hosted by the Space Environment Center (SEC). This interface has not been available since H. Garcia's death in 2004. This way of predicting SEP events relies on the factors of flare temperature, intensity, and location (Garcia, 1994a; Garcia, 1994b; Garcia, 2004a; Balch, 1999; Balch, 2008b). First, the maximum flux of the flare in the soft and hard x-ray region is recorded (<10 keV and 10-200 keV), then the temperature is calculated via Chianti or Mewe algorithms. It is found that anomalously low apparent temperatures correspond to a higher likelihood of a SEP event, as did a western location (Garcia, 2004b; Garcia, 2004a). Garcia found these results:

> Figure 1 [Here figure 4] shows the distribution of peak flare temperature with respect to peak logarithmic X-ray flux. (Peak temperature always precedes or is concurrent with peak flux.) This plot reveals several prominent features that characterize the temperature versus X-ray intensity relationship: On average, temperature increases monotonically with increasing X-ray intensity; SEP flares (diamonds in Figure 1) occupy a lower temperature stratum than normal flares (dots in Figure 1). The partially overlapping temperature distributions, fitted with quadratic functions, appear to merge approximately at the X-ray intensities of M1 ($10^{-5}Wm^2$) and X10 ($10^{-3}Wm^2$) and diverge at midrange; the incidence (observed) density of normal flares thins out at higher X-ray intensities, while SEP flare densities remain nearly uniform over the full logarithmic intensity range (above M1) except for a pronounced weakening near the upper and lower limits; and the total number of normal flares exceeds the number of SEP flares by a large factor, roughly 40:1. (Garcia, 2004a)
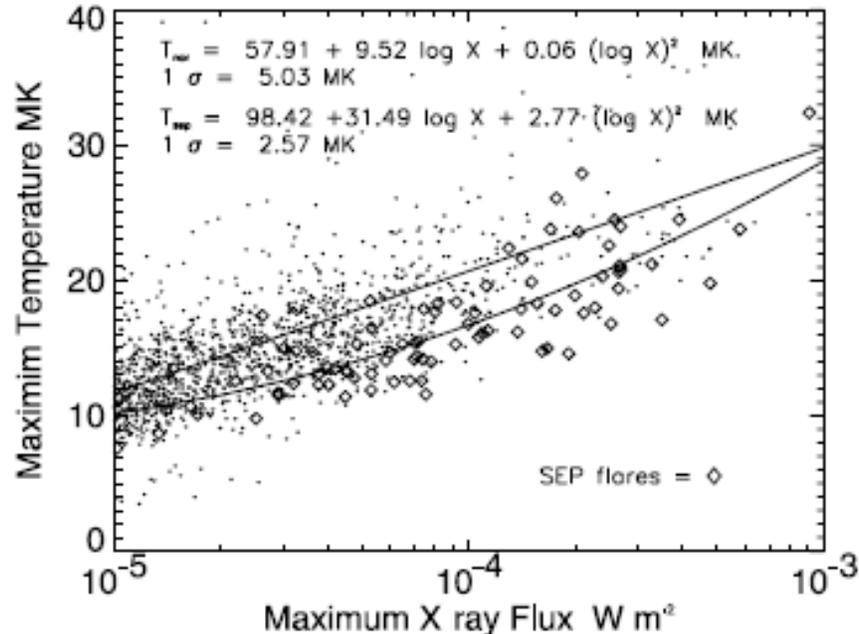
Figure 4. Flare Distribution with respect to X-Ray Flux and Temperature, Divided Between SEP Events (Diamonds) and Non-SEP Events (Dots) Showing Dependence of SEP Events on Flare Temperature (Used with Permission from Garcia, 2004a)

This data was used (Garcia, 1994a; Garcia, 2004a) to generate lines of constant probability as a method of prediction, as is shown in Figure 5. Garcia's equation for probability must be solved at a given probability for temperature as a function of x-ray peak intensity at each point across the graph to produce one curve. Solving it again at a new given probability yields more curves. Flares can be predicted as SEP or non-SEP by their location on this graph. As the maximum x-ray flux increases for a given temperature, (moving in a horizontal line across the graph) the probability of a SEP event rises continuously to the value at each numbered contour. Unfortunately, the regression is not perfect. While most SEP events are in the high probability regions, small numbers of flares without SEP fall into this region as well. These were flares that occurred either on or beyond the solar limb or at far eastern meridian distances (Garcia, 2004a). This

naturally follows, as solar location is a strong predictor of magnetic field connections

(Tascione, 1994). For Garcia's database, this predictor worked quite well. This is the

reason a longitudinal component is included in the Garcia model.
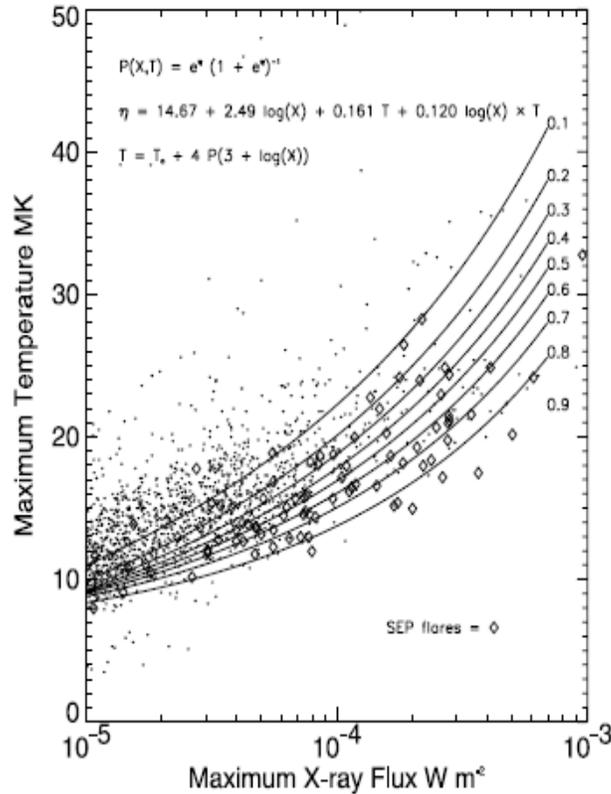
Figure 5. Curves of Constant Probability of a SEP Event, Showing SEP Events (diamonds) and Non-SEP Events in the Background (Used With Permission from Garcia, 2004a)

Though the Garcia model is no longer in operational use, the parameters to

recreate it can be obtained from the original papers. Substantial work in this thesis went

into recreating the model and verifying that it worked exactly as did the old model (see

Chapter 3: Verification of the Old Garcia Model).

Proton Prediction Model

PPM is the standard algorithm of NOAA (Balch, 1999). PPM calculates its results based on a mix of soft x-ray peak flux, predicted proton flux, and Type II and Type IV radio data and produces a percent chance that the flare will be accompanied by a SEP event (Balch, 1999). There is no direct input for the presence or absence of a CME; however the Type II and Type IV radio data serve in its place. The integrated flux parameter is calculated from the onset of the x-ray event and proceeds to the half power point on the trailing edge of the event. The prediction of maximum proton flux is based on the relationship of the log of the peak x-ray flux to the integrated flux of the associated event. It also includes historical data from the next most recent event that occurred in the same active region on the sun, as this was found to give better correlations (Balch, 2008b).

Though the physical processes that govern SEP production and transport are still an area of active research, some work with validation has been done (Balch, 2008b). All PPM data presented in this analysis comes from a modified form of the 1998 version of PPM, modified to process batch files of flares. The modifications were checked against PPM to ensure that prediction values are unchanged.

Proton Prediction System

The Air Force uses the Proton Prediction System (PPS), for the prediction of SEP events. PPS was designed by Smart and Shea in 1979 to predict the observations of SEP from observed intensity and spectra of solar events (Kahler et al., 2007). It includes inputs for onset time, peak time and location as well as one of the following sources: x-ray peak flux, radio peak flux, x-ray integrated flux and radio integrated flux. PPS differs

20

from the previous model in that it does not produce a probability of a SEP event, but

rather produces an estimated flux of protons (Kahler et al., 2007). This can easily be

converted to a prediction by applying the operational and scientific standard: >10 pfu at

10 MeV is a SEP event.

In PPS, it is assumed that SEP are accelerated by solar flares and added to the

interplanetary magnetic field within a quarter hour. The maximum flux of protons at or

above 10 MeV (J) in pfu is given (for F_{xw} being the GOES peak long x-ray flux in erg per

cm^2 per second and ΔT being the time between flare onset and peak)

$$J\,(\,E > 10\ \text{MeV}\,) = 30.67 * (F_{xw} * \Delta T)^{1.327} \tag{2}$$

PPS was also used in this analysis in batch mode, and was cross-checked with the

AF-Geospace version for individual runs to ensure correct predictions were made.

Problems with the Current Models

All current prediction algorithms evaluated here have problems, especially since

they rely primarily on peak x-ray flux (though PPS can be run solely with radio data, it

was run with x-ray peak flux during this analysis). The magnitude of x-ray flux is

insufficient to conclusively predict a SEP event. Extreme flares (x-ray flux of 10^{-4} W/m^2

(X1) or greater) have been studied since 1978, producing SEP events roughly 20% of the

time (Garcia, 2004a; Smart and Shea, 1996; Balch, 2008a). While this is much higher

than the 4% average for a flare in general, it is still inconclusive and requires additional

information to make a good prediction. Explanations for the reason that less than 100%

of extreme flares produced protons tend toward solar longitude, but 4 out of 10 occurred

at western or central locations. The likelihood of a visible disk flare "giving rise to a

significant proton event was not strongly dependent on flare location. Only the M- and

C-class proton parent flares exhibited a clear preference to occur west of the solar central meridian," (Cliver and Cane, 1989). Solar longitude is insufficiently accounted for in current models and its effects relative to flare peak x-ray flux must be examined. Neither peak x-ray flux not longitude alone can explain the sometimes erratic link between flares and SEP.

To highlight the problems in modeling, examine Figure 6, which shows the 3514 recorded flares from the database (here the ordinary flares are green squares, the flares associated with SEP are blue triangles). The figure shows how many flares there and how few are associated with SEP, just 104 out of the 3514 total flares. The difficulty in modeling this comes from the fact that no good boundary or dividing line exists between the two types.

The difficulty comes in the sheer number of flares, and how mixed the low peak x-ray flux areas are. Flares associated with SEP tend to cluster bottom right (low temperature at a given x-ray flux) as Garcia noted, but no line can divide these groups perfectly. More information is needed to classify these flares correctly.
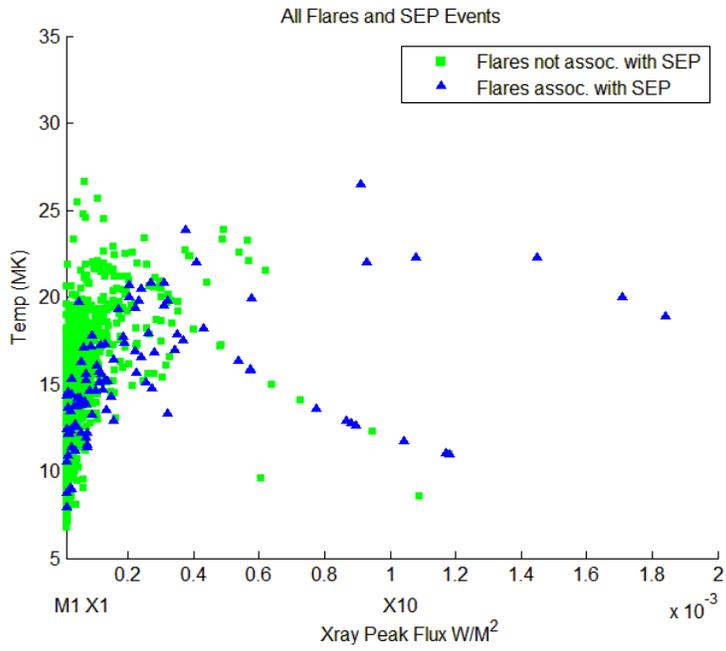
Figure 6. SEP and Control Flares in the data from 1986-2004

III. Methodology

The Data Set

The flares for this analysis are taken from the raw 1 minute GOES data available at the Space Weather Prediction Center (SWPC) website, referenced at those times when SWPC recorded that a flare had happened (SWPC, 2008). During the measured time period, 1 Jan 1986 until 31 Dec 2004, there were 3610 flares in the range M1 and above. The range M1 and above was chosen because approximately 98.5% of SEP events were associated with flares that occurred in this range, and the range misses tens of thousands of C class and below flares that rarely produce SEP (Balch, 2008b). Not all flares that occurred during this time could be used. 96 flares were removed from the dataset because there was no data recorded for these periods due to sensor overloads. Next, 708 flares with unknown locations were assigned a latitude and longitude of zero to accommodate the training algorithm. These flares were deliberately included in the dataset because the dataset needs to be as close to operational data as possible. Flares must be observed in H-alpha, white light, or other imaging system to determine flare location on the sun. Since the algorithm needs to be able to predict this sort of flare, it must be trained on it as well. The nature of the model under analysis demands a value for all the predictors under consideration. This holds true for both the location and the radio data. In the radio case, the radio data can be either observed, not observed, or not available (due to no observatory taking data at the time of the event). The net result is that some flares with no radio data actually could have had radio data that was not recorded. The effect of this change on the final algorithm is assumed to be a slight shift for lower importance for radio data. Again, as with location, this data resembles

operational data in this respect and so the shift is acceptable operationally if problematic scientifically.

Each flare's peak x-ray flux in GOES data (.5 – 4 angstrom and 1-8 angstrom) was recorded and the integrated flux was computed. The flux data is available only in 1 min intervals, so the integrated data is merely the sum of the x-ray flux from the listed beginning of the flare to the listed end time, corrected by a factor of 60 to account for the conversion from minutes to seconds. The temperature for each flare at maximum flux time was calculated as described in Chapter 2: The Mewe Temperature (White et al., 2005).

The dataset for SEP events was constructed by Dr. C. Balch, NOAA, from tracking flares from 1986 to 2004. This database contains 127 SEP events, after events are removed because of location (behind the limb) or proton flux levels too low (enhancements). These were matched up with flares from the GOES data by date and magnitude of flare. When there were two or more flares with similar data occurring soon after each other, radio data was used as the discriminator between these similar pairs of flares. Not all SEP events in the Balch database were added to the data. All SEP events associated with C class flares were removed, as were all SEP events that did not correspond to a flare in the new database. After removing these, there were 104 SEP events in the time from 1986-2004.

The Process of Verifying Models

For an analysis of the quality of each prediction algorithm, each must produce results for every flare in the database with exactly the same input data. PPS and PPM exist in operational fashion, but the Garcia model does not.

Since the death of Garcia in 2004, operational use of the Garcia model has ceased. Any usage of the Garcia model will have to start with a recreation of the model before any additional data is entered for classification. Given the new data from the database, the new model will not be exactly the same as the old. The new model should be close to the old, and analysis of how different the models are relies on the Fréchet Distance measure.

First, the Garcia Model is reproduced exactly from the coefficients published, and the resultant model is referred to as Garcia 1994 (Garcia, 1994a). The coefficients are tested against a new model produced by a Generalized Linear Model analysis when used on the new dataset, using only those predictors that Garcia chose in his 1994 paper, namely x-ray peak flux, temperature, the product of temperature and x-ray peak flux, and longitude. This is called the Reworked Garcia model. Once the two models are proven to be similar, new predictors can be added to the Reworked Garcia model in an attempt to predict SEP events better. The result of this is a new model henceforth called Garcia 2008.

Form of Garcia 1994

The original Garcia model uses a standard linear model:

$$g(P) = \eta = \sum_{j=1}^{p} \vec{x}_j \vec{\beta}_j \tag{3}$$

with $\vec{x}$ a vector of observations and $\vec{\beta}$ as a vector of empirically determined weights (Garcia, 1994a). To map this value η onto the probabilistic interval P from (0,1), Garcia used a link function $g(P) = \ln[P/(1-P)]$. The functional relationship therefore is

$$P[\log(X), T, \theta] = \frac{e^{\eta}}{1 + e^{\eta}} \tag{4}$$

26

In this model, there are 5 terms, X (x-ray peak flux in W/m^2), T (temperature in MK), an interaction between X and T, a constant, and a dependence on heliographic longitude, θ (1 if the flare occurred east of E40 central meridian distance on the sun, zero otherwise). The effectiveness of the model was determined by a standard deviation between the function P (Log(x), T, θ) and the occurrence of the event (Garcia, 1994a).

Comparison of Garcia 1994 with Reworked Garcia

In order to prove that the Garcia Model has been recreated accurately and updated properly with new flare data, I present a comparison of the two forecast algorithms, Garcia 1994 and Reworked Garcia, produced following his methodology while operating on new data. In the following figure, the lines of constant probability are plotted against data. A given flare will have a (x,y) coordinate on the graph given by its x-ray peak flux and its temperature, respectively. Its location on this graph relative to the lines of probability gives an estimate of the Garcia model's prediction of a SEP event. Any flare between the .4 line and the .5 line, for example, has a 40-50% chance to be associated with a SEP event. First, the old probability curves in Figure 7:
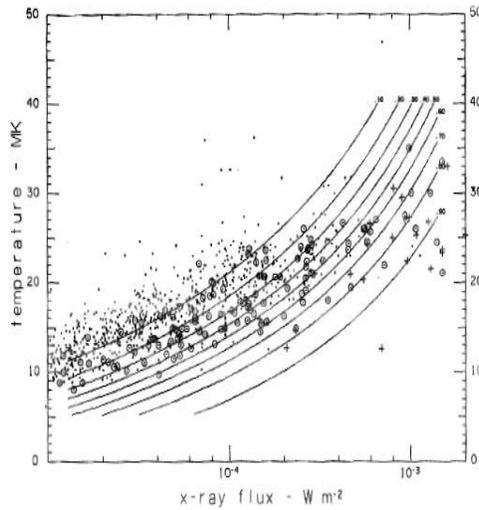
Figure 7. Garcia Model from 1994 Showing Probability Contours for the Prediction of SEP Events to Accompany Flares, Using Garcia's Original Data (Used With Permission from Garcia, 1994a)

Coefficients for these curves for Garcia 1994 (see Equation 3) are

$$\eta = 12.2558 + 1.980 Log[X] + .2049T + .1286 Log[X] * T - 1.5684 * \theta \qquad (5)$$

This holds for X the x-ray peak flux in W m^{-2}, T is the temperature in Megakelvin, and 'θ' is equal to 1 if the heliographic longitude of the flare is east of E40 on the sun, and 0 if the flare is west of E40.

For purposes of comparison, it is vital to assure ourselves that the new version of the Garcia Model is the same model quantitatively. The coefficients of both the Garcia 1994 and the Reworked Garcia models are compared. Both models have 5 terms, a constant, log of x-ray peak flux, temperature, the product of x-ray and temperature, and finally a binary term equal to zero for flares west of East 40 longitude, and equal to one for flares on the east of that dividing line. Here is a table comparing the results of the two models:

Table 2. Coefficient Comparison between Garcia 1994 and the New Model

Model	Garcia 1994	Reworked Garcia	Ratio
Constant	12.2558	11.300	1.085
Log (X-ray)	1.980	2.875	.689
Temp	.2049	.141	1.453
Log (X-ray) * Temp	.1286	.068	1.891
East / West	-1.5684	-1.706	.919

It should be noted that while the coefficients vary between models, sometimes strongly, that correlation is preserved in the sense of sign: each coefficient in the new model is the same sign as the corresponding coefficient in Garcia 1994. This is important as the sign of the coefficient declares how that observable (temperature, x-ray peak, etc.) causes the result to vary. A change in sign would imply that the models are using the data in the opposite fashion of the previous model. This would completely destroy any claims of similarity. The values for the ratio (1 represents a perfect correspondence, 0 and infinity represent no correspondence) are between .5 and 2, demonstrating imperfect alignment but good correspondence. The farthest removed, temperature x-ray interaction, is far off. Reasons for this discrepancy can be tracked to the different data sources; the Garcia model used a list of flares from Sep 1977 to May 1991. The actual reduced dataset is not available for comparison, but it was not complete. As Garcia notes:

> It [the data set] is not homogenous: few moderate-to-weak normal flares
> (<M8) are included before 1984; flares from 1984 May to 1988 September
> are mainly SEP flares; but from 1988 September to 1991 July, both
> normal and SEP type ≥ M1 flares have been included. (Garcia, 1994a)

29

This lack of data at the lower ends of the power spectrum (>M8) has an effect on the coefficients: further regression trials with a subset of the new data (>M2, >X1.5, >X5) suggest that the fewer low temperature flares (mainly flares without SEP) are included, the lower the interaction and temperature coefficients fall, with temperature venturing negative for thresholds above X5. The non-homogenous exclusion of moderate flares from the original Garcia model has had an effect on the regression.

Comparing Curves: Maximum Distance

There are several ways to determine how similar to curves are to each other. This analysis uses the Fréchet Distance (see Appendix A), which is a measure of how far apart two curves are. In words, the Fréchet Distance is "if a man is walking a dog, and the man must travel on one curve, and the dog on the other, the Fréchet Distance is the shortest possible leash the man can use," (Dumitrescu and Rote, 2004). This analysis has been accomplished for the Garcia 1994 and Reworked Garcia models for both the 50% of SEP and 90% Probability of SEP Event curves.

The Fréchet Distance for the 50% curve is a modest .629, while the 90% has a value of .838. These numbers represent actual (Euclidean) distances across the graph, and thus can be understood in a relative sense by comparison to the maximum of the data. When compared to the maximum temperature, the 50% probability curve Fréchet Distance is 1.57%, while the 90% curve distance is 2.10%. In other words, the new model's curve is 2% as far from the old model's curve as from the axis. This shows that despite their differing coefficients, these two models produce similar curves of equal probability for predicting SEP events.

As a comparison, I examine temperature measurements for Norman, OK starting in 1995 (Brooke and Doswell, 1996). The data compares the actual observed temperatures for the area with three models predicting 12-24 hours in advance. The three models are from the National Weather Service (NWS), the Limited Area Fine Mesh (LFM), and the Nested Grid Model. The Fréchet Distances for each of these curves, as compared to the observed temperature, is revealing. The NWS prediction has a Fréchet Distance of 7, while the NGM prediction has a Fréchet Distance of 7.810, and the LFM prediction has a Fréchet Distance of 7.071. These numbers are absolute (Euclidean) distance across the graphs, and can again be understood best as a percentage of maximum range (temperature). The maximum range or temperature of the data is 103 F, so each Fréchet Distance can be compared to that. In this fashion, the NWS prediction has a percentage of .068 or 6.8%, the NGM has a percentage of 7.58% and the LFM has 6.87%. Compared to the Garcia models, with Fréchet Distances in percentage form of only 1.57% and 2.10%, the weather models are much more different.

The weather data is analyzed here because single day predictions are assumed to be relatively accurate and relatively close to each other. As the Garcia curves had much smaller percentage-wise Fréchet Distances than the weather predictions did, we can assume the two Garcia models are more similar to each other than the weather models are to observations. Thus, the Fréchet Distances between the old and new Garcia models are small enough that I can confidently claim them to be similar.

Proof of the Recreation of Garcia 1994 with New Coefficients

This overall assessment shows that the new model is similar to the Garcia 1994, and the areas of greatest difference, the temperature x-ray interaction coefficient, can be

31

explained by database differences between Garcia in 1994 and the current model's data from 1986-2004. Further, as the current database includes all flares magnitude M1 and greater (less those removed as noted earlier), and no areas with reduced reporting, the new Garcia coefficients should be regarded as updated versions of the old.

From here, new coefficients can be added to the new model to increase its ability to discriminate between SEP and non-SEP events. This will be done with the Generalized Linear Model method. The result of these additions is the algorithm Garcia 2008.

Modeling Techniques

The Generalized Linear Model

One method to design an algorithm to classify and predict solar proton events with general statistical techniques is a Generalized Linear Model (GLM), which is a flexible form of a least squares regression (Hardin and Hilbe, 2001). It relates the response variable (SEP event occurrence) to the predictors (x-ray flux, temperature, etc.) by the use of a link function (in this case, the binomial distribution given by the occurrence or non-occurrence of SEP events requires the logistic equation) (Hardin and Hilbe, 2001). In this form of model, each predictor is compared to the response and the variance in the response that can be explained by the predictor is calculated via the likelihood function. Predictors with a high relevancy will show up as each predictor is assigned a p-value based on the likelihood that variance in the predictor explains variance in the response. The p-value is a measure of the likelihood that the measured variation arose by a change in the related predictor; thus, p-values below .05 are considered significant, while p-values above that arbitrary limit are considered not significant

(McClave, 2008). This value of .05 can be raised if the amount of data far exceeds the number of parameters (Hardin and Hilbe, 2001). During the modeling process, predictors are removed from the model in order of relevancy to see what the effects of removal are; removal of one "not significant" factor can remove some predictor variance that did explain the change in the response. In this case, the remaining predictors will become more significant as the variance of the response still needs to be accounted for. However, the p-values of the predictors will not change at the same rate, allowing the most useful predictors to be singled out. When all remaining variables have a p-value that indicates it is relevant to the prediction, each is evaluated.

The evaluation of coefficients in a Generalized Linear Model is done by maximum likelihood estimation or least squares estimation. Maximum likelihood estimation takes the values observed, assumes a distribution with some mean and variance, and calculates what the mean and variance must be for the observed data to be the most likely observation. Formally, this is done by setting the partial derivative of the log of the prediction equal to zero for the variable of interest (Gumbel, 1958). As the number of observations increases, the probability that this mean is the mean of all the data approaches unity. This calculated mean can then be used as a fit coefficient. Least squares method minimizes the sum of the squared deviations between the predicted value and the measured value (Bulmer, 1965).

This method of Generalized Linear Modeling has a useful trait: it allows the user to specify interaction effects as separate predictors. If two or more predictors influence each other, producing non-linear effects, this can be measured and accounted for by a

Generalized Linear Model. Higher order terms can thus be treated in nearly the same fashion as linear terms.

Logistic Regression

The result of this GLM process is a number z which is the sum of coefficients times the values of the parameters (x-ray flux, temperature, etc.). This number may be positive or negative, and is not asymptotically smooth at the boundaries [0,1], where any realistic prediction must end. These are awkward features for a prediction. A link function is used to map the results into the interval [0,1] for predictions P (z). In this case, the solution is to use a logistic regression. This maps the output function into the interval [0,1] using a logistic function as the link function:

$$z = Log(\frac{P(z)}{1-P(z)}) \tag{6}$$

Logistic regression is suited for applications where there are many predictors and only two outcomes, here SEP event or no SEP event (Brannick, 2006). Fitting to the logistic function prevents values from going beneath 0 and from going over 1, both theoretically impossible. The logistic equation is plotted below in Figure 8.
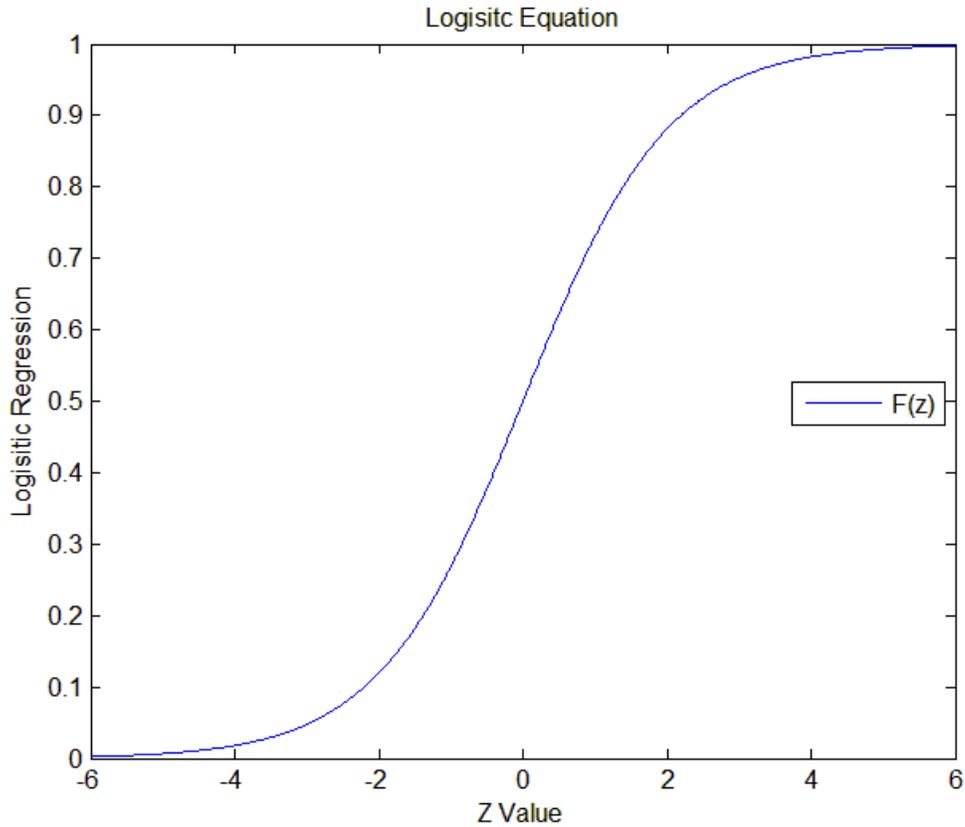
Figure 8. The Logistic Equation

Classification

This method provides some insight into the problem of class separation. The focus of this work will be to find the best process that will allow the separation of the two distinct classes: flares associated with SEP and flares not associated with SEP. To be useful for forecasting purposes, this process must use only data that is readily available, such as GOES data. This work will focus on extending and enhancing the current state of the art for prediction.

These results will be compared to the previous (Garcia) model and strengths and weaknesses of all the models will be explored.

Quantitative Measures of Success

False Alarm Rate and Missed Forecast

To determine the quality of a forecast, some definitions are needed. A correct forecast is one where the event was forecast and occurred (A). A False Alarm (FA) is where the event was forecast but did not occur (B). A Missed Forecast (MF) is where the event was not forecast but occurred anyway (C). A correct null is where the event was not forecast nor did it occur (D). The total number of events is $N = A + B + C + D$. A two way truth table would appear as such:

Table 3. Definition of the Two Way Truth Table for SEP Prediction

Event

		Yes	No	
Prediction	Yes	A	B	FA
	No	C	D	

MF

Some formulas for calculating the quality of the prediction (Balch, 2008b):

$$\text{Probability of detection (POD)} = A / (A + C)$$

$$\text{False Alarm Rate (FAR)} = B / (A + B)$$

$$\text{Percent Correct (PC)} = (A + D) / N$$

None of these measures really takes into account the quality of the forecast. A .97 percent correct prediction rate sounds great, but here it is merely the result of always predicting a non-event. Always predicting 'no SEP event' is correct 3410/3514 or 97% of the time. This is a high prediction rate, but contains no real forecast dependent on

data. Clearly, these measures of prediction lack real information on the quality of the forecast.

The Heidke Skill Score

The comparison between the prediction methods (PPM, PPS, Garcia) relies on a skill score known as the Heidke skill score (HSS). This number ranges between $-\infty$ (for incorrect predictions) to +1 (for all correct predictions). Zero is a skill score representative of a prediction method no better than guessing. The formula for calculating the Heidke skill score is based on the argument: (Balch, 2008b referencing the original work by Heidke, 1926;)

$$\text{Probability (event = Yes)} = (A + C) / N$$

$$\text{Probability (forecast = Yes)} = (A + B) / N$$

Thus, the probability for a hit by chance (e.g. the probabilities are independent) is the product of the probabilities of each:

$$P\ (\text{Event} = \text{Yes} \cap \text{Prediction} = \text{Yes}) = (A + C) * (A + B) / N^2$$

A chance correct null is similar to the positive:

$$\text{Probability (event = No)} = (B + D) / N$$

$$\text{Probability (forecast = No)} = (C + D) / N$$

So the probability of a correct null by chance is:

$$P\ (\text{Event} = \text{No} \cap \text{Prediction} = \text{No}) = (B + D) * (C + D) / N^2$$

Thus, the probability of a correct prediction by chance (both correct forecasts and nulls) is the sum of the two probabilities, or

$$((A + B) * (A + C) + (B + D) * (C + D)) / N^2$$

Multiplying by the number of trials N, we obtain the number of correct forecasts by chance:

$$E = ((A + B) * (A + C) + (B + D) * (C + D)) / N$$

We can then define the Heidke skill score by the number of correct results more than by chance per the total number of attempts less chance successes:

$$HSS = (A + D - E) / (N - E) \qquad (7)$$

As noted earlier, the Heidke skill score ranges from $-\infty$ (all wrong) to $+1$ (all correct), with 0 the equivalent of exactly any many right as might be predicted by chance.

For an example of the Heidke Skill Score, I return to the temperature forecasts for Norman, OK (Brooke and Doswell, 1996). Each of the three forecasts for the next day's temperature is a forecast and can be tracked by a skill score. In order to keep the format of the predictions the same, each prediction was assigned binary score tracking whether the prediction was for an increase in temperature or not (No change in temperature is classified with no). This was compared to the observed data, also formatted the same way. This format allowed a simple comparison of these predictions to SEP predictions by allowing for correct forecasts (the forecast predicts temperature going up, and the observed temperature rises), false alarms (the forecast predicts a rise in temperature but no rise occurs), missed forecasts (the forecast predicts a fall in temperature or no change, but the temperature rises), and correct nulls (neither the prediction nor the observation are of a rise in temperature). The results are that LFM has a HSS of .473; NGM has a HSS of .571; and NWS has a HSS of .517. By looking at the HSS, it is readily apparent that the NWS model is the best at forecasting changes in temperature. Also, we note that

even an easy forecast, a hotter/colder forecast for 12-24 hours in advance, has a HSS of no more than .571.

A second comparison is a more standard weather comparison, with correct predictions as those within 3 degrees of the observed temperature, false alarms as those predictions 3 or more degrees above the observed, and missed forecasts 3 or more degrees below the observed. In this method of analysis, there are no nulls. In this case, the HSS comes out as -.359 for LFM, -.350 for NGM and -.272 for NWS. Though it may seem trivial to change the range for a good prediction from 3 degrees to something larger, and therefore obtain a positive HSS, this does not work. The HSS rises to zero but does not become positive. The Heidke Skill Score tracks the number of predictions right by chance as well as the total number correct, so by the time the range is raised to 15 degrees, the HSS is at zero for two of the three systems—there are no more correct predictions than can be explained by chance.

This is not a comprehensive analysis of the Heidke skill score, merely a note to understand the relevance of a Heidke score above .5, and the value of finding the maximum score.

Thresholds

Each algorithm needs some method of determining its prediction. The Air Force needs a definitive answer, yes or no, SEP or Non-SEP, for each and every flare (Kahler et al., 2007). Thus, some sort of threshold must be established, such that every flare falling below that threshold will be predicted as not having a SEP event, and every flare at or above the threshold will be predicted to have a SEP event. In Garcia 1994, Garcia 2008, and PPM, the numbers come out as percentages. As the prediction is in the form of a

percentage, it has a natural threshold of 50%. PPS, which outputs a predicted flux of protons has a threshold at 10 pfu, the operational standard. However, subsequent analysis reveals that these are not the optimal thresholds to choose. Substantial improvement can be found by optimization of the threshold with respect to Heidke Skill Score.

Truth Tables

For comparison, all results will include a truth table, detailing the results into categories for correct forecasts (A), false alarms (B), missed forecasts (C), and correct nulls (D).

Table 4. Truth Table Definitions

Event

Observed

Event		Yes	No	
Forecast	Yes	A	B	FA
	No	C	D	

MF

An ideal or perfect classifier would appear as in Table 5:

Table 5. Ideal Truth Table for Flare Database

Event

Observed

Event		Yes	No	
Forecast	Yes	104	0	FA
	No	0	3410	

Events in a perfect classifier are perfectly separated into correct forecasts (A) and correct nulls (D).

Optimized Prediction Algorithms

The Process of Optimization

As noted in the previously, all of these predictive algorithms have some threshold or threshold value above which they predict a SEP event. This limit is either based on probability (.5 or greater) or a physical definition (10 pfu or greater), but these limits need not be regarded as hard limits. Indeed, as the optimization process shows, much higher HSS results can be obtained by shifting the threshold values. In this process, the threshold value is altered and a new HSS is obtained at each value. As the threshold increases, the false alarm rate decreases, but the missed forecast rate rises. The percentage correct also falls as the threshold rises, but since the vast majority of the data is nulls in the low probability region, the threshold must be high enough not to receive false alarms from this data. The false alarm rate, HSS, missed forecast rate, and correct forecast rates are all shown in Figure 9. While all these numbers are displayed as percentages, the HSS calculates from absolute numbers. Thus, the false alarm rate is more important than it seems, as the number of false alarms is high. The HSS line rises smoothly to a maximum as fewer false alarms are registered, then falls off again as the number of correct forecasts begins to fall and the corresponding number of missed forecasts rises. Maximizing the HSS is thus a simple matter of finding the maximum of this curve, as in Figure 9. While the graph does have occasional local maxima, the overall maximum can be found by perturbing the chosen threshold significantly and looking for the slope. The curve is fairly flat at the maximum, and thus it is unsurprising

41

that it jumps up and down slightly around the true maximum. This is due to the discrete

nature of the data. A small increase in threshold will not add missed forecasts and

remove false alarms at the same rate, and thus the curve jumps. These jumps are usually

small, on the order of .02. In the case where two or more maxima exist, to the limit of 3

decimal places, the lowest threshold is chosen to minimize the number of missed

forecasts. This process will be applied to all the algorithms.
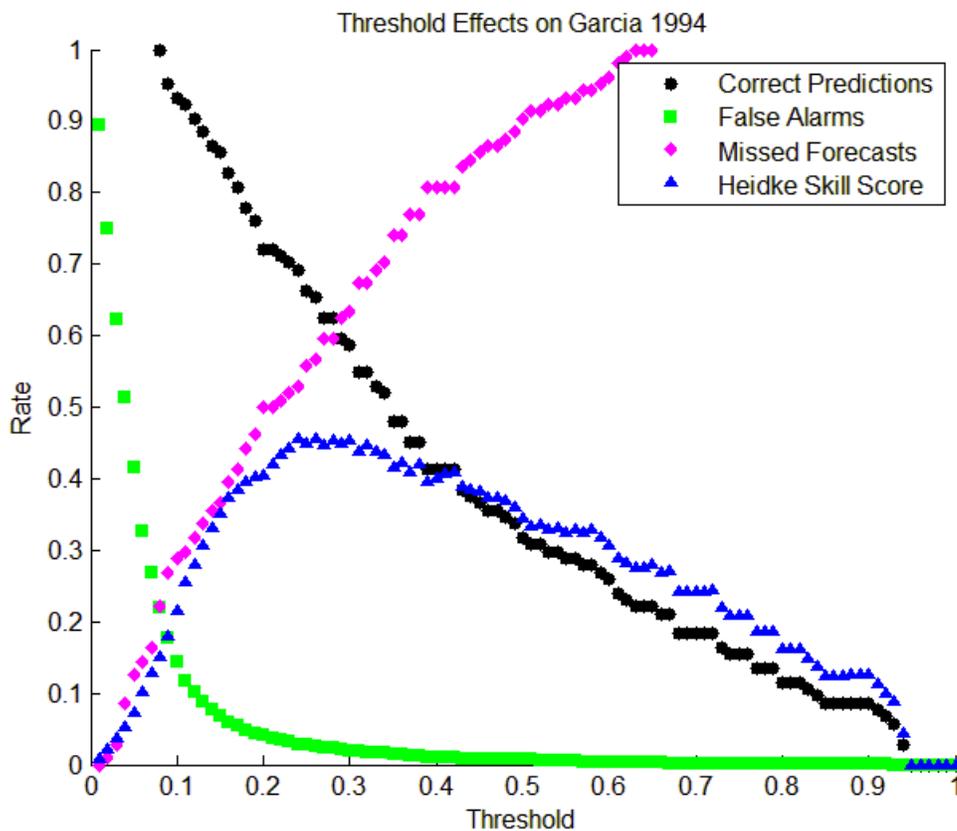
Figure 9. SEP Prediction Threshold Effects on Garcia 1994

For additional clarity, this graph is replotted in Figure 10 by absolute number of

successful forecasts, false alarms, and missed forecasts. Since the HSS maximizes the

number correct while penalizing equally for each false alarm and missed forecast, the

maximum score occurs at and around the intersection of the lines for false alarms and

missed forecasts.
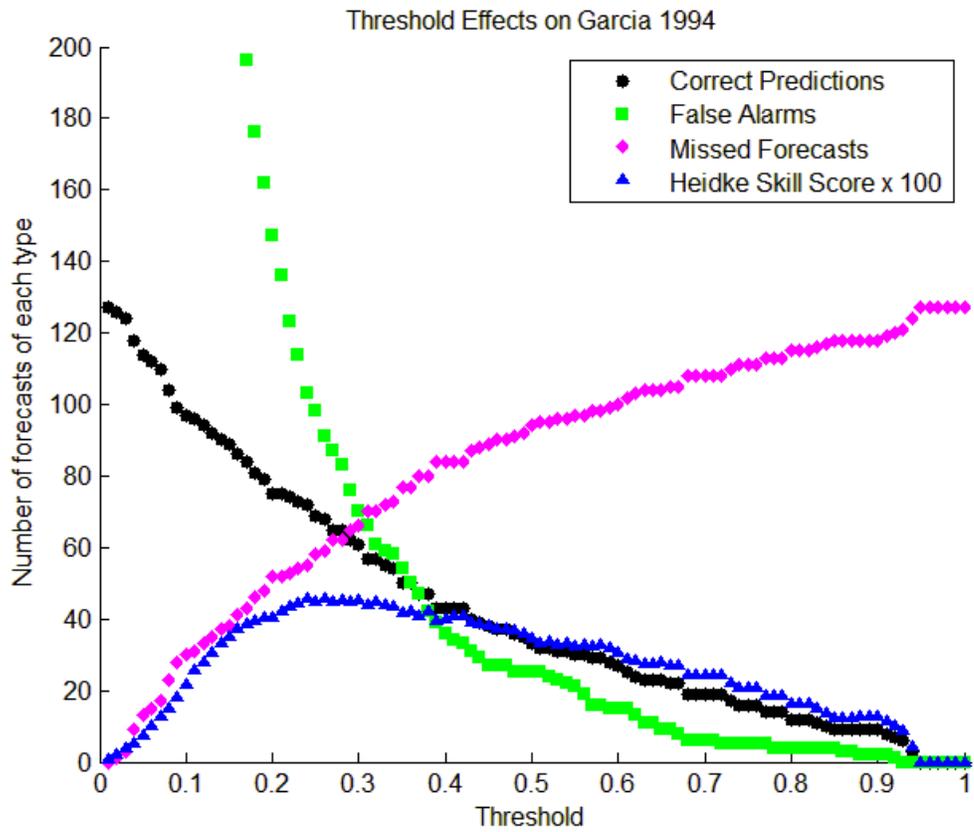
Figure 10. SEP Prediction Threshold Effects on Garcia 1994, Absolute Numbers

Original Results

All the flares in the database (3514 flares, after the removal of <M1 flares and those that overloaded the sensors) were run through each algorithm for an initial prediction. The overall accuracy of these predictions relative to the SEP truth data from NOAA will be compared to the results from the optimized algorithms and to Garcia 2008.

Garcia 1994

The Garcia 1994 algorithm is the first algorithm tested for prediction. Below is a truth table of the data, followed by a graph, showing results. Here, successful forecasts are blue triangles, missed forecasts are pink diamonds, and false alarms are green squares. This format will be standard for displaying results over the next graphs. Again, this uses a 50% or .5 threshold for prediction. The HSS for Garcia 1994 is .342. The algorithm performs well (few false alarms or missed forecasts) at high x-ray regions and poorly at low regions. The truth table follows as Table 6, and the results are shown in graphical form as Figure 11:

Table 6. Truth Table for Garcia 1994

Observed

Forecast	Yes	No	
Yes	50	119	FA
No	54	3292	

MF

44
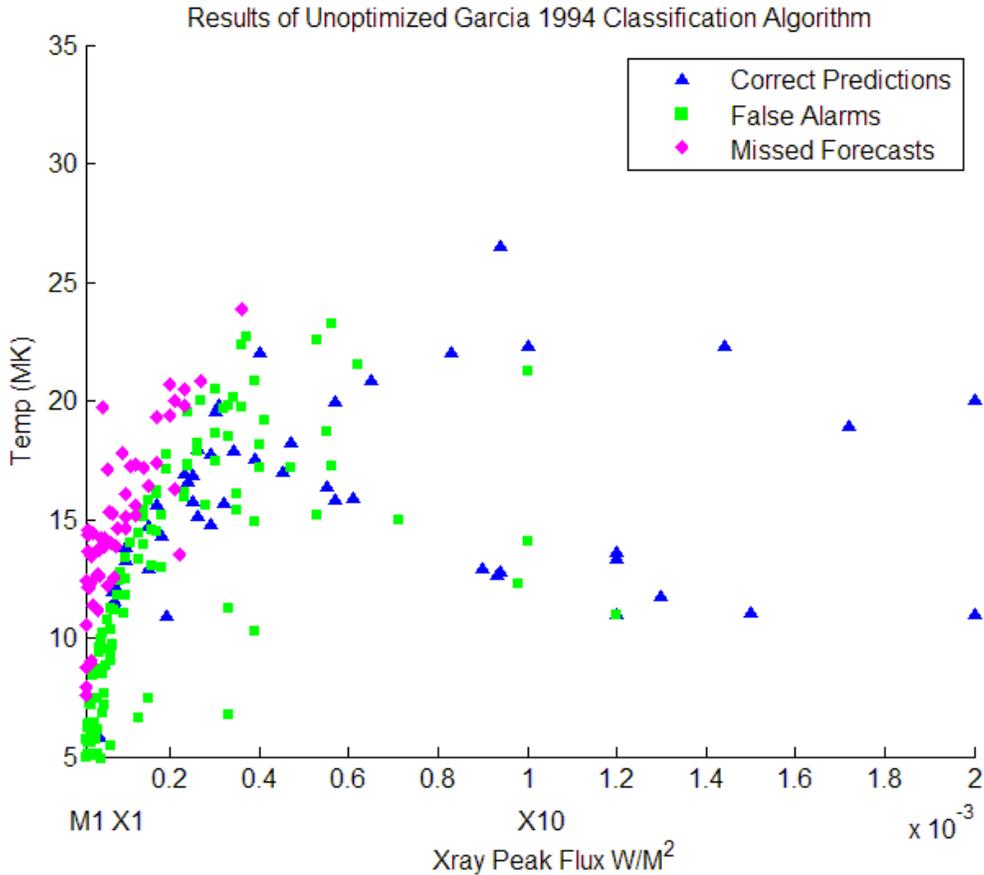
Figure 11. Garcia 1994 Forecast Results at 50% threshold

Details of the Garcia 2008 Model

After using a 5 input linear model to compare the Reworked Garcia model to Garcia 1994, I added more terms into it to improve its forecasting ability. This new model is called Garcia 2008. The Garcia 2008 model is similar to the original Garcia model in many respects. It has inputs corresponding to the log of the x-ray flux, temperature, interaction and longitude, as did Garcia 1994, but it also includes the existence of Type II and IV radio data, and the integrated x-ray flux. Also, it references the numerical value of the longitude rather than a binary expression for the flare location east of E40 on the sun. This selection of predictors was chosen as described in the

Generalized Linear Model section, by removing the predictors with highest p-value one by one until only relevant ones were left. The result of the model is a number between 0 and 1 representing a probability that the associated flare will produce SEP. It requires a threshold, just like Garcia 1994 and can be optimized with respect to that threshold. It cannot predict with no missed forecasts unless the threshold is set to zero, which makes using the algorithm useless. It performs better at lower maximum x-ray flux than does Garcia 1994, and better overall. Coefficients for Garcia 2008 are listed in Table 7:

Table 7. Coefficients for Garcia 2008

Predictor	Coefficient
Constant	-54.597
Log (Xray)	-19.930
Log (Xray)^2	-2.050
Temperature	2.525
Temperature^2	-.030
Log (Xray) * Temperature	.4375
Longitude	.013
Presence of Type II radio data	.3577
Presence of Type IV radio data	1.040
Integrated Flux	2.730
Integrated Flux^2	-.3725

Further details and usage of the model are listed in Appendix B.

Results from Garcia 2008

Garcia 2008, using a threshold of .5, produces a HSS of .387, slightly better than

Garcia 1994. Here are the results in Figure 12 and Table 8, with successful forecasts as

blue triangles, missed forecasts as pink diamonds, and false alarms as green squares:

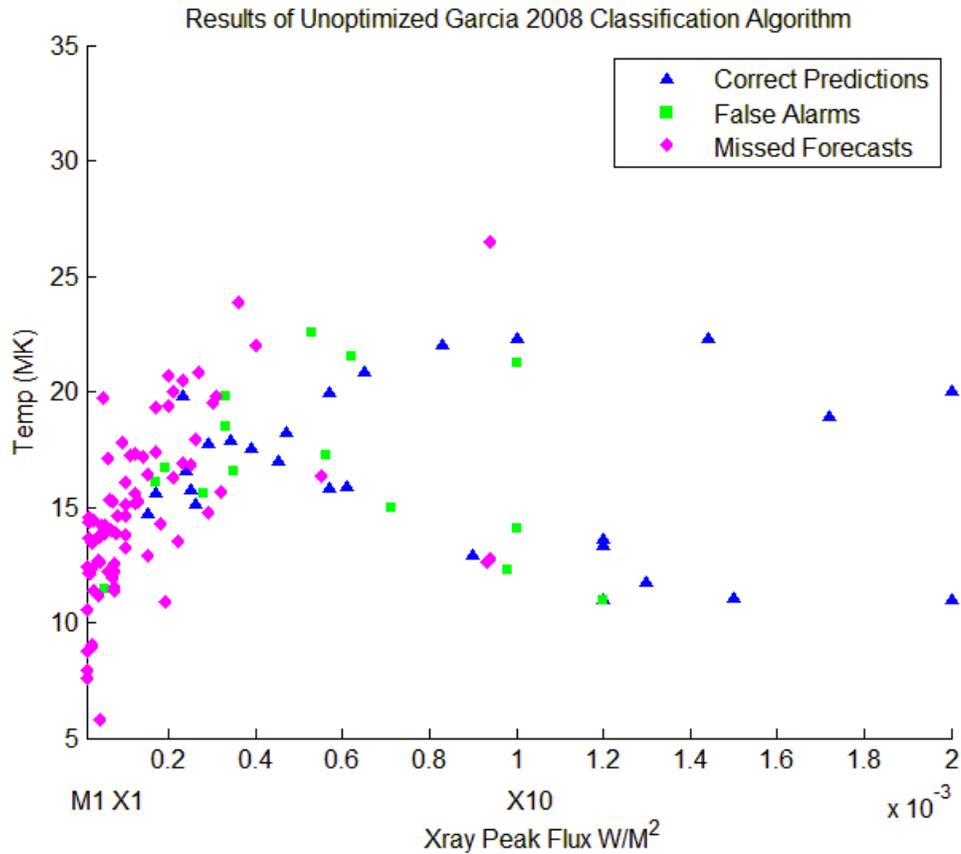**Figure 12. Results of Garcia 2008 Predictions at a 50% Threshold**

Table 8. Garcia 2008 Truth Table at a 50% Threshold

Truth Table

Event forecast		Observed	
	Yes	No	
Yes	29	13	FA
No	75	3397	

MF

47

PPS

PPS is the best algorithm when it comes to predicting SEP events with no missed forecasts, however the ability to have no missed forecasts comes at the price of too many false alarms. For a threshold at 10 pfu, PPS has 2077 false alarms, a false alarm rate of about 59%, and only 1 missed forecast. This indicates that whenever PPS makes a prediction for a SEP event, there is a 59% chance that it is wrong. This algorithm has a HSS of .036, indicating is it barely better than a random prediction. As a different way of considering it, note that an algorithm that always predicts a SEP event will have only one additional success (104) and more false alarms (3410 instead of 2077). If a prediction of a SEP event each time a flare occurs is good enough, then no algorithm is needed at all. The results are shown in Table 9 and Figure 13:

Table 9. PPS Truth Table at 10 pfu

Truth Table

event forecast   Observed

	Yes	No	
Yes	103	2077	FA
No	1	1334	

MF
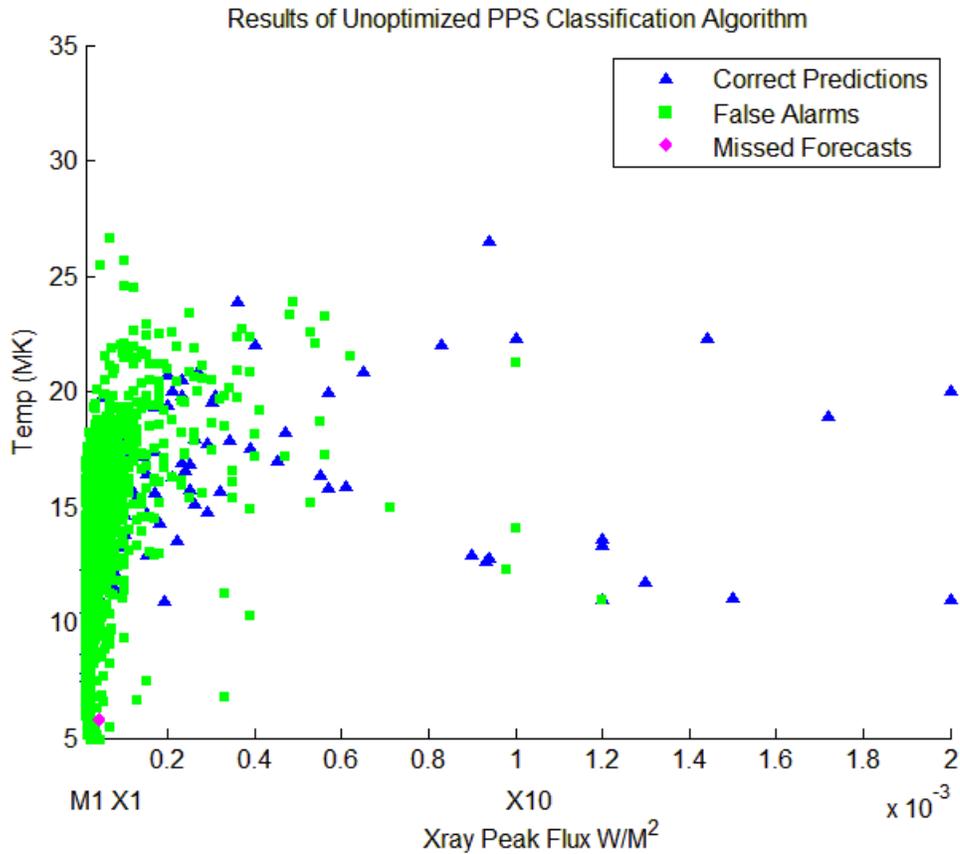
Figure 13. PPS Results at a Threshold of 10 pfu

PPM

PPM is the final algorithm to be considered. PPM includes more input information with which to make predictions, such as location, radio data and integrated flux. However, at the .5 threshold, it has problems with too many missed forecasts. Its HSS comes out at .093, as seen in Table 10 and Figure 14:

49

Table 10. PPM truth table

Truth Table

event forecast   Observed

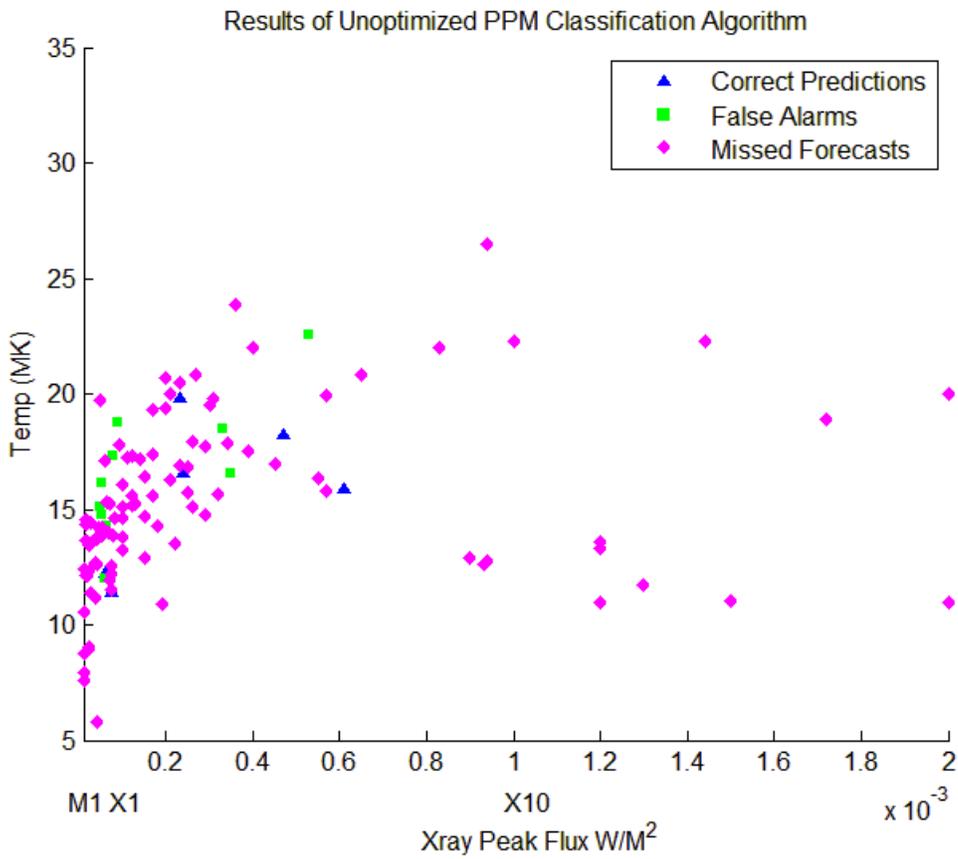|	Yes	No	
Yes	6	10	FA
No	98	3401	

MF

Figure 14. PPM Results at 50%

Comparison of all Unoptimized Prediction Algorithms

None of the prediction algorithms is particularly effective. All of them have either too many false alarms or too many missed forecasts. The results are listed below in Table 11 for easy comparison:

Table 11. Heidke Skill Score Comparison for Unoptimized Algorithms

Algorithm	HSS	POD	FAR	PC
Garcia 1994	.342	.432	.642	.923
Garcia 2008	.387	.269	.363	.936
PPS	.036	.461	.634	.923
PPM	.093	.423	.577	.929

Optimization

Following the low skill scores, each algorithm was optimized with respect to its skill score to see how well it could predict.

Optimized Garcia 1994

The Garcia 1994 model benefits from optimization. At a threshold of .58, it obtains a HSS of .371, up from its previous HSS of .342, as shown in Table 12 and Figure 15:

Table 12. Garcia 1994, Comparison Before and After Optimization

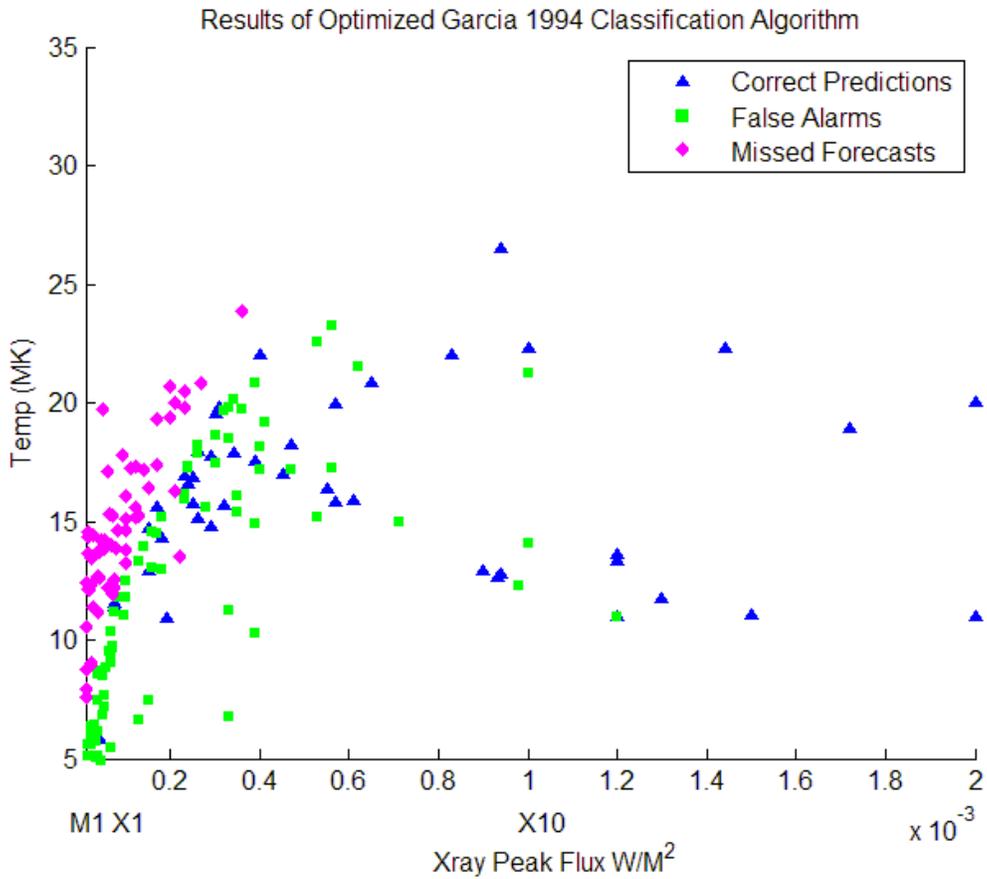| Unoptimized				Optimized			
Truth Table		Before		Truth Table		After	
Event		Observed		Event		Observed	
Forecast	Yes	No		Forecast	Yes	No	
Yes	50	119	FA	Yes	45	81	FA
No	54	3292		No	59	3330	
	MF				MF		

51

Figure 15. Garcia 1994 Forecast Results After Optimization

Optimized Garcia 2008

While the 'original' version of Garcia 2008 was merely for comparison, the working model is the optimized version. This model has a HSS of .526, well above the HSS for Garcia 1994. It has a threshold at .18. The results are seen in Table 13 and Figure 16.

Table 13. Garcia 2008, Comparison Before and After Optimization

Unoptimized Truth Table

Optimized Truth Table

Forecast	Observed Yes	Observed No	
Yes	29	13	FA
No	75	3397	
	MF		

Forecast	Observed Yes	Observed No	
Yes	60	58	FA
No	44	3352	
	MF		

Figure 16. Optimized Garcia 2008 Forecast Results

Figure 17 and Figure 18 are graphical representations of the bad forecasts of

Garcia 2008. Flares with unknown longitude, adding no information, are not plotted

here. The false alarms show little pattern, spread evenly across the sun's longitude. The only exception is between 40-60 West, where the false alarms share a peak with missed forecasts. The missed forecasts, however, show a definite pattern in the west. This is confirmation that the longitude is important to Garcia 2008, but to rely on it too much raises more false alarms in that area.
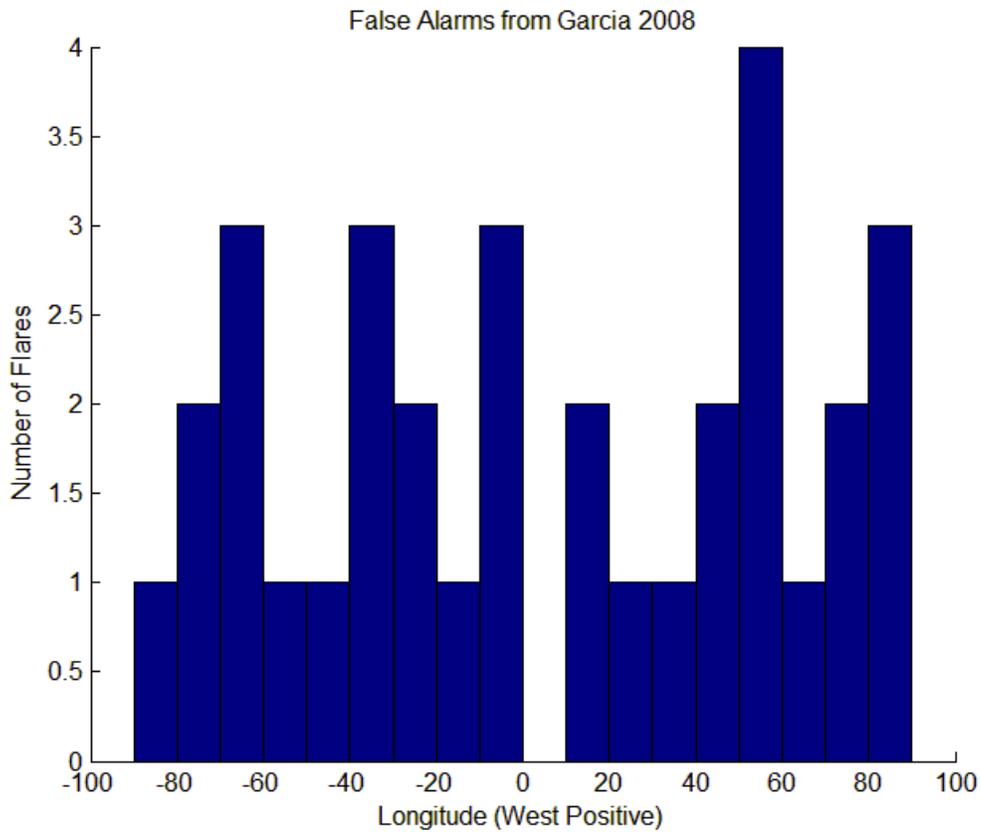
Figure 17. False Alarms from Garcia 2008 by Longitude
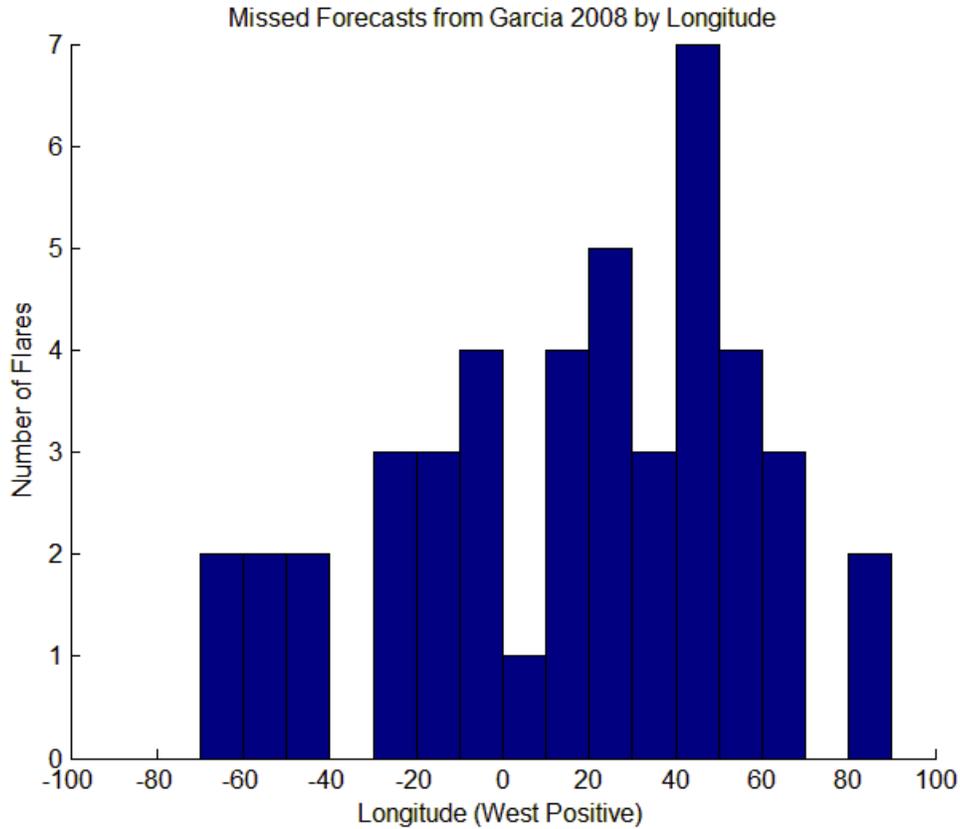
Figure 18. Missed Forecasts for Garcia 2008 by Longitude

The presence or absence of radio data for both false alarms and missed forecasts is shown in Figure 19. The largest number of bad forecasts is for missed forecasts with no radio data. The SEP events associated with these flares are extremely hard to predict. They have no radio signature, either Type II or Type IV, yet there was a SEP event. This is evidence radio data is not required for a SEP event and that there are other mechanisms at work here. While SEP events with no radio data are hard to predict, the opposite is not true: having some radio data means a much lower rate of bad forecasts, both false alarms and missed forecasts.

Radio Data From Garcia 2008, False Alarms and Missed Forecasts
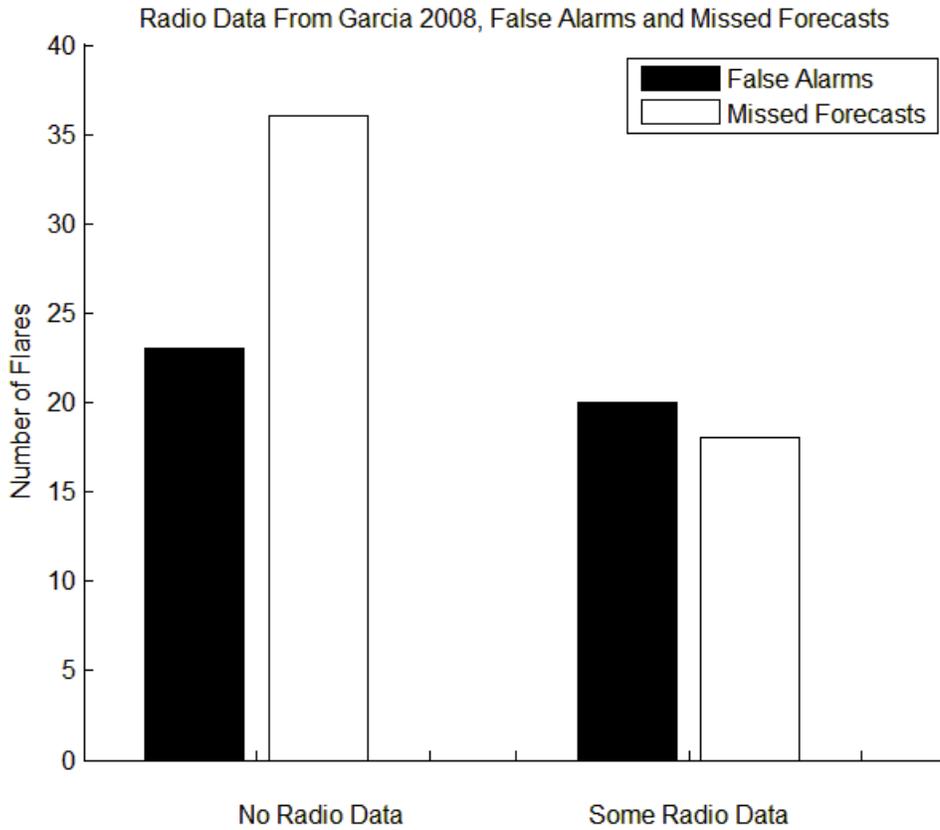
Figure 19. All Incorrect Forecasts from Garcia 2008, by Radio Data

Optimized PPS

PPS has the greatest change during the optimization process. Previously, its HSS was a mere .036, and it had 2077 false alarms. This is because it used a threshold of 10 pfu, the operational value representing a SEP event. However, since the threshold functions in the predicted measure, it can be changed at will while maintaining the operational measure for flares when they are observed. PPS is optimized with a HSS of .388 at maximum, with a threshold of 720 pfu. The results are in Table 14 and Figure 20:

Table 14. PPS Truth Table, Before and After Optimization

Unoptimized

Truth Table

Forecast   Observed

	Yes	No	
Yes	103	2077	FA
No	1	1334	

MC

Optimized

Truth Table

Forecast   Observed

	Yes	No	
Yes	48	83	FA
No	56	3328	

MC
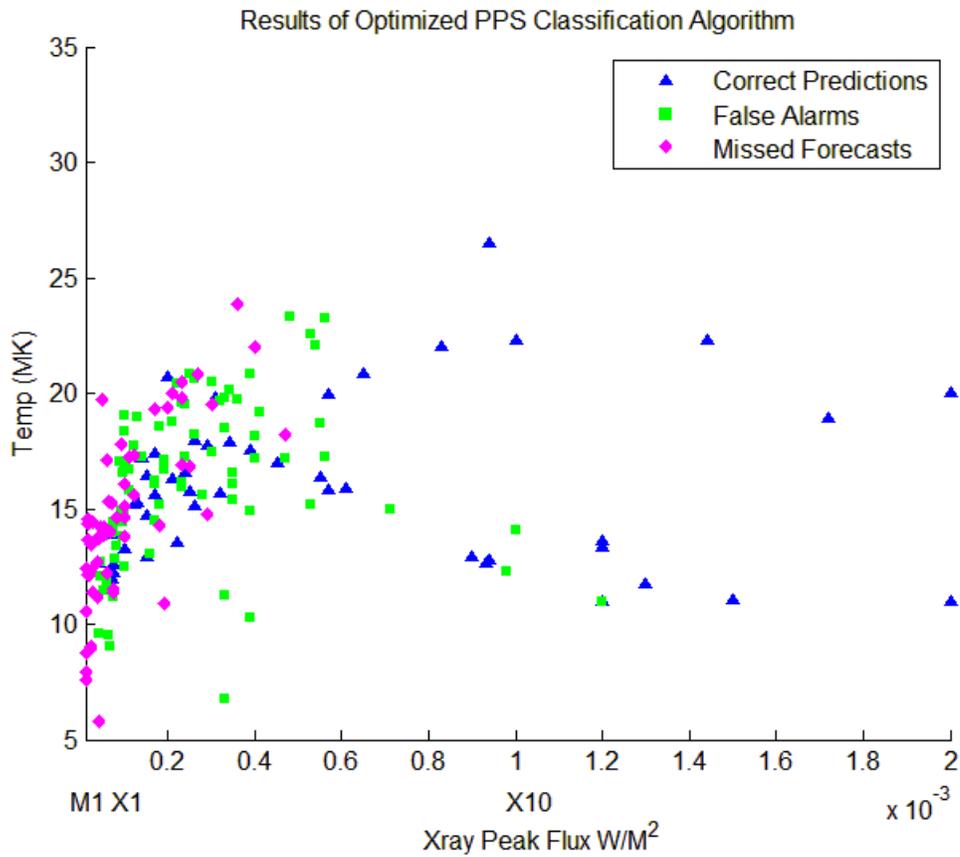
Figure 20. Optimized PPS Forecast Results

57

Optimized PPM

When optimized, PPM obtains a higher HSS. Originally at a threshold of .50, its HSS was .093. After optimization, its threshold is at .30 (the number above which a SEP event is predicted), and its HSS goes up to .405, showing a much improved forecast ability. Its comparison truth table is now

Unoptimized
Truth Table

Forecast		Observed	
	Yes	No	
Yes	6	10	FA
No	98	3401	
	MC		

Optimized
Truth Table

Forecast		Observed	
	Yes	No	
Yes	44	60	FA
No	60	3351	
	MC		

Table 15 and its graphical results as Figure 21:

Table 15. PPM Truth Table, Before and After Optimization
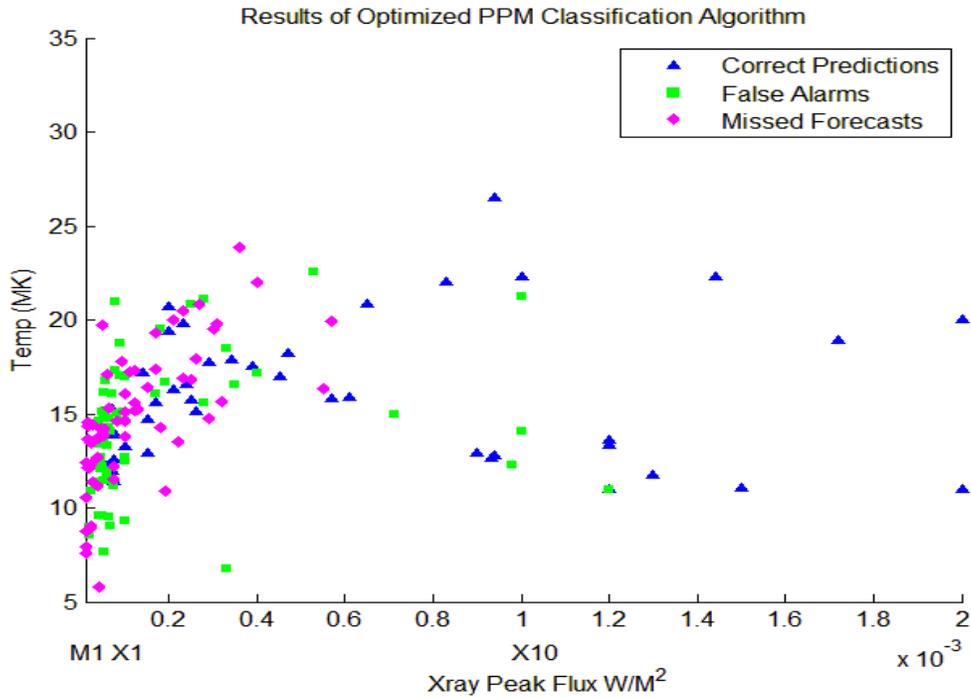
Figure 21. PPM Forecast Results after Optimization

Comparison of Heidke Skill Scores for Optimized Algorithms

Each of the algorithms listed in Table 16 has improved under the process of

optimization. The best forecasting algorithm is Garcia 2008, by more than .120 HSS

points. This is also a rise of .184 over the previous best method of prediction,

unoptimized Garcia 1994. This is a tremendous change in predictive power, particularly

over the operational algorithms of PPS and PPM.

Table 16. Heidke Skill Score Comparison for Optimized Algorithms

Algorithm	Unoptimized HSS	Optimized HSS	Threshold
Garcia 1994	.342	.371	.58
Garcia 2008	.387	.526	.18
PPS	.036	.388	720
PPM	.093	.405	.30

Verification with Modern Flares

When a model is trained and tested on the same database, a problem called overtraining can result (Irizarry, 2006). An overtrained model is not general but rather constructed so as to maximize the result for the given information. In extreme cases, the model remembers every training data point and can therefore perfectly classify that one dataset. In areas outside the trained region the model performs less well. In the worst cases, the model may begin to fail entirely when presented with the new data. In the case of solar flares, if the database were not representative of the standard types of solar flare, for example if all flares above X1 had been recorded with an accompanying SEP event, the model would be trained to predict SEP events for all X1 and greater flares, no matter what their other features.

The solution for this problem is simple: test the model on a new representative database. Fortunately, there is plenty of data available, as the flare database ends 31 Dec 2004. There have been many solar flares since then. For validation purposes, I used the flare data from NOAA's Space Weather Prediction Center (SWPC, 2008), augmented with radio data from NASA (Kaiser, 2008). This was correlated with NOAA's database of SEP events to generate a new list of flares over the span of 1 Jan 2006 - 31 Dec 2007 (SWPC, 2008). There were 138 M1 and greater flares during this period, and three SEP events. The three SEP events were clustered in December of 2006, on 5 Dec, 13 Dec and 14 Dec.

Each algorithm is tested on this new data, and each has a new HSS associated with the new data. While changing the threshold again would allow further optimization (in this case the HSS for Garcia 2008 rises from its new value to .443, equal to the best of

60

all the algorithms on this dataset), this is not performed here because this is the testing database. The purpose of the testing database is to find out how the algorithm performs on new data. It is a test of the optimization procedure. It is unlikely for any of the algorithms to do better than in the training dataset (unless the verification dataset is selectively chosen to include only 'easy' flares), and will likely do worse. This is because they were optimized to fit the training set as perfectly as they could. If the new verification dataset were identical, in a statistical sense, to the old dataset, then the new scores would be equal to the old. Otherwise, the scores will fall. This is not a problem, we are looking for proof that the algorithm still performs in this new region.

Garcia 1994 does badly with the new data, with a HSS of just .229. This is quite a change from the previous values of optimized HSS at .371.

When the data is examined with PPM, there is also a fall. The value of PPM at the optimized threshold, set at .3, is .222, a fall from the optimized value (.405). In this dataset, PPM does worse still far above the HSS it had with the original dataset before optimization (.093).

For PPS, the situation is similar to the original dataset. On the verification data, PPS has a HSS at .443. This is the only algorithm to show an increase in HSS between the training dataset and the verification dataset.

For Garcia 2008, the HSS comes out at .330 after optimization, showing that this dataset is similar but not the same as the old dataset. Garcia 2008, operating on the verification data, is still better than either of its unoptimized competitors PPS and PPM predicting on the old dataset, and functions fairly well even in this new region. Each result is listed in Table 17:

61

Table 17. Verification Heidke Skill Scores for all Algorithms

Algorithm	Unoptimized HSS	Optimized HSS	Verification HSS
Garcia 1994	.342	.371	.229
Garcia 2008	.387	.526	.330
PPS	.036	.388	.443
PPM	.093	.405	.222

Garcia 2008 and the Verification Dataset Problem

Garcia 2008 performs worse in the Verification Dataset than it did on the master dataset from 1986-2004. Table 18 is a truth table showing the bad forecasts, then Figure 22 is a graph of the data:

Table 18. Garcia 2008 Truth Table for Verification Database

Truth Table

Event Forecast	Observed		
	Yes	No	
Yes	3	11	FA
No	0	124	
	MF		
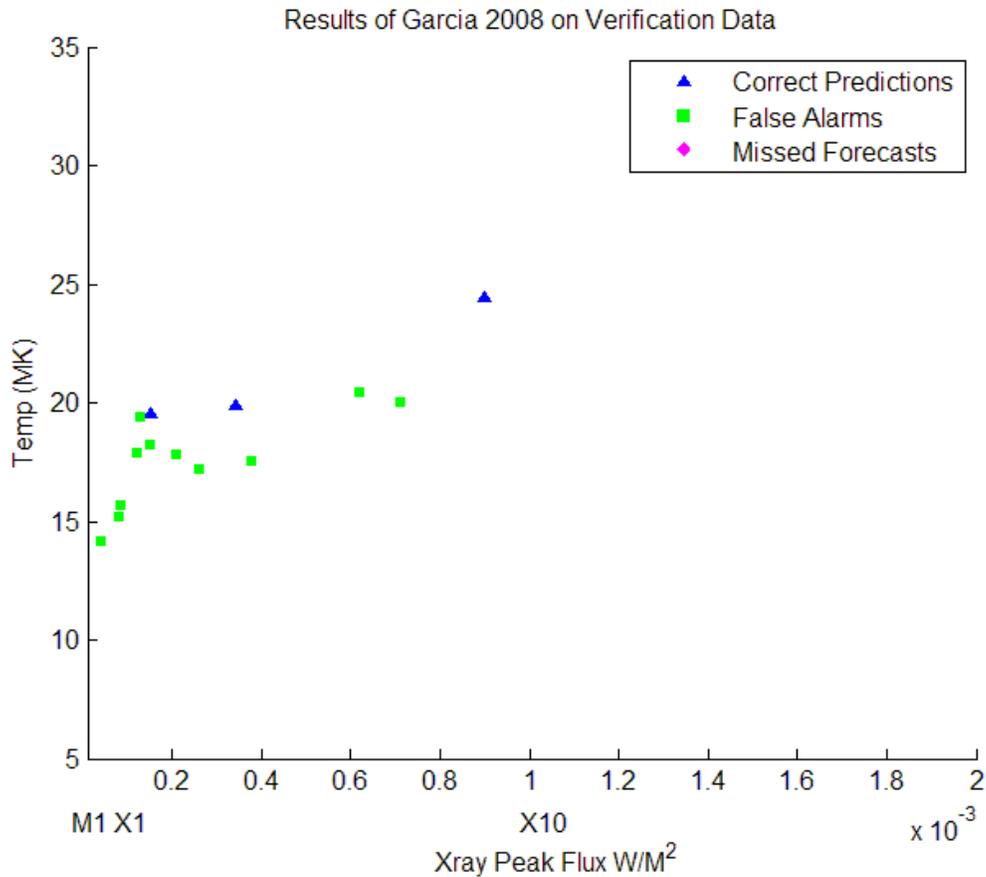
Figure 22. Prediction Results from Applying Garcia 2008 to Verification Dataset

There are several reasons for Garcia 2008's lack of skill in the verification dataset. First, the flares in the verification data are mostly small M class flares. Garcia 2008 does work with these, but this is not its best region. Garcia 2008 does best with cool X class flares, and though the SEP events in the verification data are associated with hot flares, there are several cool flares without SEP, as seen in Figure 23.
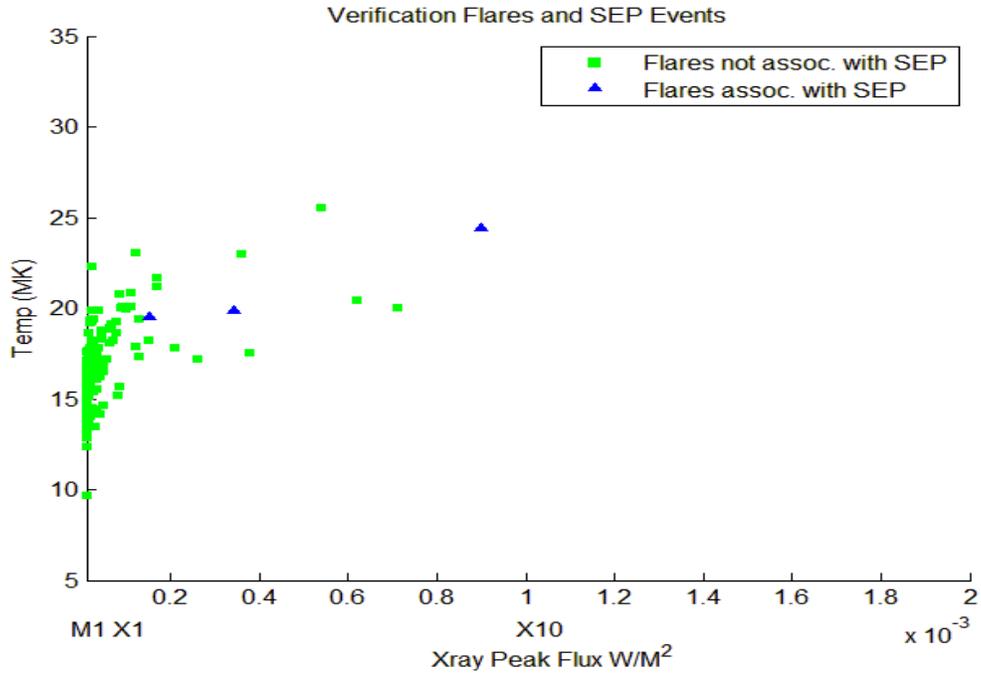
Figure 23. Verification Database Flares and SEP Events

Secondly, Garcia 2008 successfully predicts all three SEP events in the verification data. The drop in HSS comes solely from 11 false alarms. There is only a slight preference of west over east pattern to the false alarm flares in longitude, as shown in Figure 24:

Figure 24. Garcia 2008 on the Verification Dataset, False Alarms

These false alarms also come from the radio data. In all but one case of the 11 false alarms, the flare had both Type II and Type IV data associated with it. This overwhelming preference is shown in Figure 25:
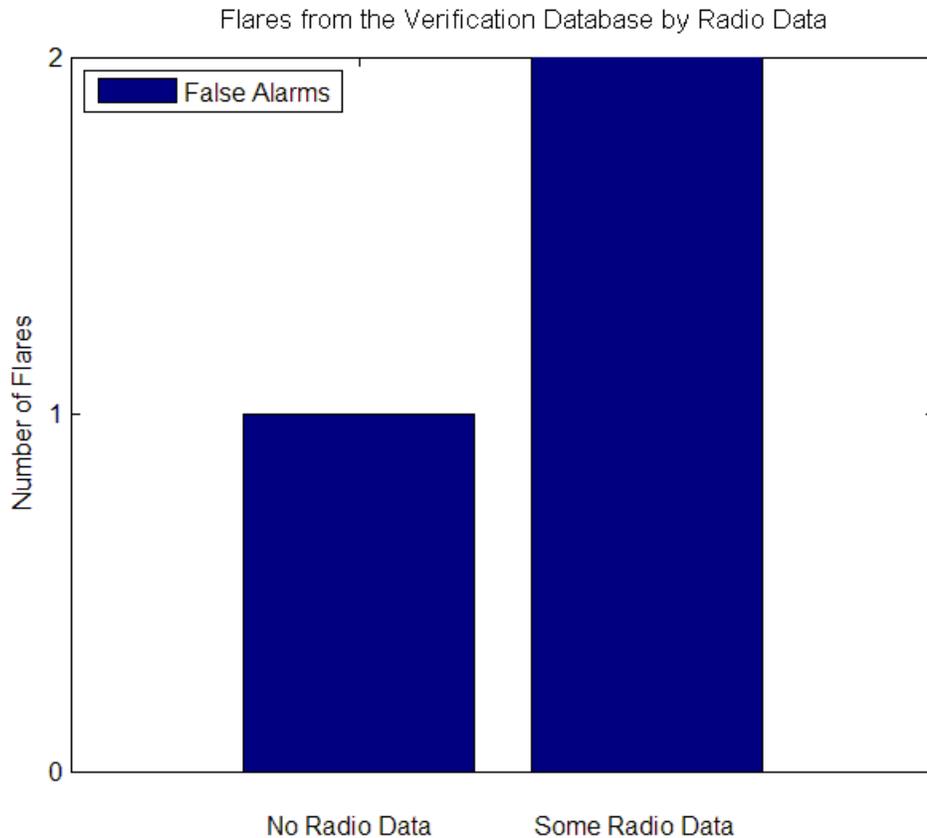
Figure 25. Garcia 2008 False Alarms in the Verification Dataset, by Radio Data

Six of these 11 flares came from western longitudes, and one was unknown. The flares were also cooler than normal for their peak x-ray flux. The false alarms w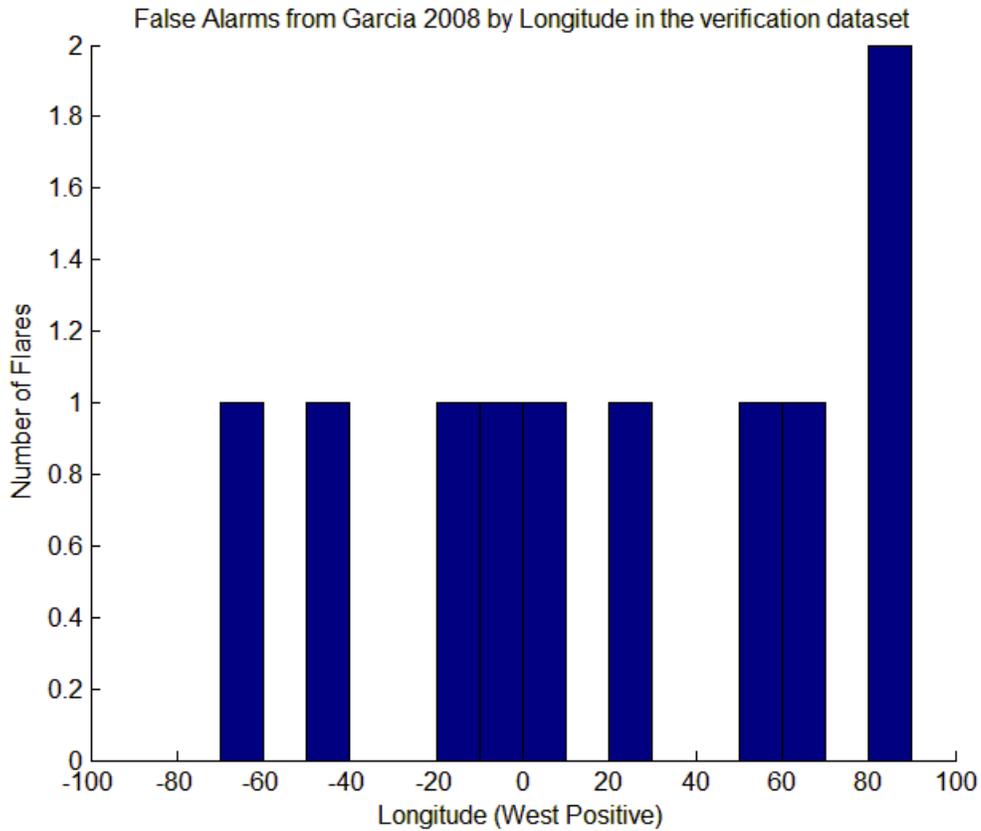ere also mostly X flares. Overall, these are exactly the kind of flares that Garcia 2008 learned to predict as associated with SEP events in the original database. As shown, there is no association of predictors that will perfectly separate these flares. Without additional predictors, these flares will not be classified correctly.

Several Flares: A Scientific Comparison

As the false alarms in the verification database appear so similar to flares with SEP, further analysis is needed on the difference between flares associated with SEP and those without. A comparison of similar flares yields interesting light on the physics

inherent in this analysis. It also explains the difficulty in telling flares that are associated

with SEP events from other flares. There is no simple dividing line, as this data shows.

Flare A, with a SEP event, from 20 Feb 2002, has similar properties to flares B (27 Oct

2003), C (19 Sep 2000), and D (29 July 99); each not accompanied by a SEP event. Here

Table 19 is a list of the easily comparable data from each flare, from the Balch

Database (x-ray class, location on the solar disk, Type II and IV radio events, duration of

these events if applicable, presence or absence of a Coronal Mass Ejection (CME),

integrated flux, temperature in MK, emission measure, truth of SEP events, and the

prediction each of the algorithms upon receiving this data):

Table 19. A Comparison of 4 Similar Flares Showing Differences and Similarities Between Flares with SEP Events and Flares without SEP Events

Title	Flare A	Flare B	Flare C	Flare D
Date	20 Feb 2002	27 Oct 2003	19 Sep 2000	29 Jul 1999
X-ray Scale	M5.1	M5.0	M5.1	M5.1
Location	N12W72	S16E26	N14W46	N25E51
Type II?	Yes	No	Yes	No
TII Duration	15	NULL	1	NULL
Type IV?	Yes	No	Yes	No
TIV duration	16	NULL	28	NULL
CME?	Yes	No	Yes	Yes
CME speed	952	NULL	766	199
Integrated Flux	2.26E-02	2.70E-02	7.09E-02	1.57E-02
Mewe Temp.	17	17.2	16.9	16.6
SEP?	Yes	No	No	No
PPM Predicts	No	No	Yes	No
PPS Predicts	Yes	Yes	Yes	Yes
Garcia Predicts	No	No	No	No

As is apparent from this data, these flares share many similarities. They have

similar x-ray maximum flux (x-ray scale), they have a similar temperature, and their

emission measures (measure of electron density per area) are exactly the same. Their

integrated flux values are very close for three of the four flares. The biggest difference is

the western longitude of Flare A. As noted earlier, longitude is a good predictor of SEP event. However, the probability of SEP event seems nearly constant in the west, so Flare C is similar to Flare A in this respect and difficult to differentiate. The next difference is the lack of the Coronal Mass Ejection (CME) for flare B, also without the radio traces that it produces. CMEs are strongly correlated with SEP events (Kahler and Vourlidas, 2005; Kahler et. al., 1984; Kahler, 1996), and the radio signatures are a function of the mass that was ejected pushing its way through the thin gas in the interplanetary medium. They are not required for a SEP event, nor are they sufficient, as flares C and D suggest. The high speed of the CME in flare A as compared to flare D is noteworthy, even given the extreme look angle of flare A. Flare A comes from 72 degrees west of solar center, so only a fraction of the true speed may be shown. CMEs do not necessarily originate at the some location as flares, so the exact number is unknown. This higher CME speed may indicate favorable conditions for the SEP event.

Both flares A and C have a Type II and Type IV radio signature. Eastern latitudes are correlated with a lesser chance of observed SEP events of certain magnitudes, due to magnetic field lines that direct particles away from our sensors (Garcia, 1994a).

Here, flare C is the closest in appearance to flare A, which has the only SEP event of the four. Indeed, predictions by both PPS and PPM anticipate this will have a SEP event, incorrectly. Clearly, neither algorithm has sufficient information to make a good forecast about this flare. Garcia 1994 gets this one right, though it fails to correctly predict the actual SEP event.

This series of comparisons serves to highlight the difficulties in forecasting SEP events. The majority of flares are below X1 and it is extremely difficult to separate flares

with SEP from those without in that region. None of the three models in operational use

can separate these 4 flares correctly.

V. Conclusions

Changes to State of the Art

There are several conclusions which can be reached from this research. The first is that the current operational algorithms, PPS and PPM, can both be improved dramatically by simple changes to threshold values.

Threshold Value Changes

Both PPS and PPM have operational thresholds assigned to them by the nature of their prediction. PPM predicts as a percentage chance of SEP events and at the 50% mark has a Heidke Skill Score (HSS) of only .093. Optimizing PPM with a threshold value of .3 leads to a HSS of .405, 96% correct total predictions and 42% correct prediction of SEP events. PPS shows the most dramatic change from original to optimized version. At the default level of 10 particle flux units (pfu), PPS has a HSS of only .039. It has only one missed forecasts at a cost of thousands of false alarms. Raising the threshold for a prediction to 720 pfu causes the HSS to rise to .388. This is a significant change. The original Garcia 1994 model shows similar improvements as optimization is applied. From as HSS of .342, it rises to a HSS of .370 when the threshold is raised from .50 to .58. Finally, the Garcia 2008 model has the best overall results at a HSS of .526, well above any other predictive algorithm.

The improvement of all these models under optimization is an important conclusion from this paper. No model is perfect as it is, and simple changes can drastically improve the ability of the algorithms to predict SEP events. All algorithms can be optimized with respect to the HSS by using the thresholds listed in Table 16, and thus obtain better results.

New Garcia Model versus Old Garcia Model

The method that generated the old Garcia model, Garcia 1994, was sound: use a Generalized Linear Model to calculate the dependence of the probability of a SEP event on the logarithm of the flux and temperature of the flare. However, since the database used to generate this model was outdated and incomplete (Garcia, 1994a), there is every reason to want a new model updated with new coefficients. I performed this analysis and then found a new model with more terms that was substantially more accurate. Both models predicted the high x-ray flux region equally well, and had similar troubles in the low x-ray regions.

The Success of the Garcia 2008

The second major conclusion from this research is that Garcia 2008 is the most successful predictive algorithm tested for the training data. It also predicts in the verification dataset. It uses a linear combination of coefficients for the log of the x-ray peak flux, the temperature, the integrated flux, radio data for Type II and Type IV radio events, and flare longitude for prediction. These factors give it an unprecedented .526 Heidke Skill Score, .12 greater than the next best algorithm. For usage of Garcia 2008, see Appendix B. For a full text version of the generating algorithm that can be used to add predictors or use new data, see Appendix C.

Recommendations for Further Research

There is room for improvement on several facets of this analysis. The association of SEP with predictors more difficult to measure, such as gradual hardening of x-rays, the interplanetary magnetic field, as represented by the geomagnetic index, and solar wind speed are all important, particularly to a physics based model. These parameters are

difficult to measure but may increase our ability to predict SEP events drastically. Second, the association of SEP with CMEs is important, but this study focused on flares. This study should be repeated using robust measures of CMEs, such as mass, speed, and location, in addition to the radio data currently used, and these predictors should be compared to observed SEP events. There is room for improvement with the statistical approach as well. In this approach, only basic logistic regression was used. A modern statistical look at SEP, for example with a neural net or support vector machines, may be able to improve results. Finally, a new type of statistical model that can take into account uncertain data such as flares with no location or unknown radio data should be able to better predict this subset of flares more efficiently.

These recommendations for further study are not easy, but they should be done for the safety of operators in space. There are enough solar flares in even a calm time like 2006-2007 (138 flares at M1 or greater, 3 proton events) to be a danger to anyone in space.

Appendix A. Fréchet Distance

There are several ways to determine how far apart two curves are. The first here illustrated is the Hausdorff Distance, defined here as (Pelletier, 2002; Dumitrescu and Rote, 2004):

$$\delta_H(A,B) = Max(\underset{a \in A}{Max}(\underset{b \in B}{Min}(d\{a,b\})), \underset{b \in B}{Max}(\underset{a \in A}{Min}(d\{a,b\}))) \qquad (8)$$

The Hausdorff distance thus finds the closest distance (here we use the Euclidean distance metric) between two points. Pick a point a that exists on curve A, and find the minimum distance to any point b on curve B. Pick the maximum value over all choices of a, then repeat the process starting from curve B. This procedure has obvious weaknesses, in that it does not take the course of the curve into consideration, merely the location of points.

The second way to measure distances between curves is the Fréchet Distance. This process can be illustrated by the example: suppose a man is walking a dog, and both man and dog are constrained to walk along curves, but may travel at their own speeds. They may not move backward. The Fréchet Distance is the minimum length of the leash the man must use (Pelletier, 2002). Mathematically, this appears as:

$$\delta_F(P,Q) = \underset{\substack{\alpha[0,1] \mapsto [0,N] \\ \beta[0,1] \mapsto [0,M]}}{Min}(\underset{t \in [0,1]}{Max}(d\{P(\alpha(t)),Q(\beta,t)\})) \qquad (6)$$

In words, this equation represents the phrase, "For every possible function α(t) and β(t), find the largest distance between the man and his dog as they walk along their respective path; finally, keep the smallest distance found among these maximum distances," (Pelletier, 2002).

For the application here, determining whether or not two probability curves are similar, the Fréchet Distance and Hausdorff Distance are the same. Fortunately, the choice of α(t) and β(t) are obvious as the two curves possess no inflection point, there will be no better match then the point of closest approach at each point. This can be found by finding the intersection of the perpendicular to the derivative of the slope with the second curve. Alternatively, for small data sets, it can be simpler to compare each and every point and simply take the minimum. Each point has a closest neighbor in the opposite set, and the maximum of these local minimums is the Fréchet Distance for the two curves.

Appendix B. Use of Garcia 2008

Garcia 2008 is easy to use as a prediction algorithm. Use this command in MATLAB, associated with variables as follows (logxray = base 10 log of the peak x-ray flux in W/m^2; temp = Mewe temp as seen in Chapter 2; long = longitude of the flare if known, positive west, zero if unknown; type2 = 1 if a type II radio event occurred, 0 otherwise; type4 =1 if a type IV radio event occurred; intflux = the integrated flux between flare beginning and end):

```
b=[-54.5973;-19.9285;-2.0495;2.5251;-0.0304;0.4375;0.0126;0.3577;1.0397;2.7282;-0.3725;];

fit=glmval(b,(logxray,logxray.^2,temp,temp.^2,logxray.*temp,long,type2,type4, intflux,intflux.^2),'logit');
```

The result stored in the variable 'fit' is the prediction for a SEP event. Apply the threshold .18 and mark as no-SEP any flare with a probability less than that number, and mark as a SEP event any flare with greater probability.

To perform this operation without MATLAB, use the coefficients listed times the appropriate predictor. This number, η, is transformed via the logistic equation into the probability. Use equation 4 with η for the probability.

Appendix C. Text of Garcia 2008

Following is a text of the MATLAB code that produced Garcia 2008. If a spreadsheet 'C:\Documents and Settings\addyourdatahere.xls', with columns in the order: flare onset, flare max time, flare end time, flare peak flux in W/m^2, blank, blank, blank, longitude, blank, blank, type II binary, type 2 duration, type 4 binary, type 4 duration, cme binary, blank, cme speed, blank, blank, integrated flux, blank, blank, temp (mewe), 9 blank columns, SEP truth binary, 9 blank columns, then lat and long (south and east negative). Output will be in the variable b (coeffiecients for GLM) and fit (prediction for each flare). A new flare can be predicted with variables loaded as shown and the command:

b=[-54.5973;-19.9285;-2.0495;2.5251;-0.0304;0.4375;0.0126;0.3577;1.0397;2.7282;-0.3725;];

fit=glmval(b,(logxray,logxray.^2,temp,temp.^2,logxray.*temp,long,type2,type4, intflux,intflux.^2),'logit');

Following is the full text of the algorithm to fit new parameters or new coefficients into the model, from a spreadsheet as described in Appendix B:

Garcia 2008

```
clear all
close all
%load data
(text, numeric,raw)=xlsread('C:\Documents and Settings\addyourdatahere.xls');
xray=();
temp=();
sep=();
rpatrol=();
type2=();
type4=();
cme=();
cmespeed=();
intflux=();
emmewe=();
tmewe=();
eMFhianti=();
type2dur=();
type4dur=();
lat=();
long=();
west=();
```

```
%load raw into columns with titles
sepraw=raw(1:end,34);
xrayraw=raw(1:end,4);
tempraw=raw(1:end,25); %chianti
rpatrolraw=raw(1:end,10);
type2raw=raw(1:end,11);
type4raw=raw(1:end,13);
cmeraw=raw(1:end,15);
cmespeedraw=raw(1:end,17);
intfluxraw=raw(1:end,20);
tmeweraw=raw(1:end,23);
emmeweraw=raw(1:end,24);
eMFhiantiraw=raw(1:end,26);
%intflux=raw(101:end,20);
type2durraw=raw(1:end,12);
type4durraw=raw(1:end,14);
latraw=raw(1:end,44);
longraw=raw(1:end,45);
westraw=raw(:,47);
for q=2:length(xrayraw)
%turn cells into doubles
        xray=(xray;xrayraw{q});
        temp=(temp ; tempraw{q});
        sep=(sep;sepraw{q});
rpatrol=(rpatrol;rpatrolraw{q});
type2=(type2;type2raw{q});
type4=(type4;type4raw{q});
cme=(cme;cmeraw{q});
cmespeed=(cmespeed;cmespeedraw{q});
intflux=(intflux;intfluxraw{q});
emmewe=(emmewe;emmeweraw{q});
tmewe=(tmewe;tmeweraw{q});
  eMFhianti=(eMFhianti; eMFhiantiraw{q});
type2dur=(type2dur;type2durraw{q});
type4dur=(type4dur;type4durraw{q});
lat=(lat;latraw{q});
long=(long;longraw{q});
west=(west;westraw{q});
end



b=();
% observation=();
% sep=();
% control=();
newrow=();
 numberofboxestemp=50;
 numberofboxesxray=200;
 xraysep=();
 xrayctl=();
 tempsep=();
 tempctl=();
 %setup for graphing
 for q=1:length(xray)
```

```
if sep(q)==1
    xraysep=(xray(q),xraysep);
    tempsep=(temp(q),tempsep);
else
    xrayctl=(xray(q),xrayctl);
    tempctl=(temp(q),tempctl);
end
 end
 (boxesnorm,centers)=hist3((xrayctl',tempctl'),(numberofboxesxray,numberofboxestemp));
 boxessep=hist3((xraysep',tempsep'),centers);
 minxray=min(xray);
 maxxray=max(xray);
 stepxray=(max(xray)-min(xray))/(numberofboxesxray);
 xcent1=(minxray:stepxray:maxxray);
 mintemp=min(temp);
 maxtemp=max(temp);
 steptemp=(max(temp)-min(temp))/(numberofboxestemp);
 xcent2=(mintemp:steptemp:maxtemp);


dev=();
b=();
stats=();
fit=();
dlo=();
dhi=();


%model
logxray=log10(xray)
%this step does all the work
b=glmfit((logxray,logxray.^2,temp,temp.^2,logxray.*temp,long,type2,type4,intflux,intflux.^2),sep,'binomial');
%this step makes a prediction for each flare, stores the prediction in
%'fit'
fit=glmval(b,(logxray,logxray.^2,temp,temp.^2,logxray.*temp,long,type2,type4, intflux,intflux.^2),'logit');



%plot everything to see how it did
%triangles are probabilities, + overlaid are the events
scatter(xray,temp,10*(fit),'r^', 'filled');
hold on
scatter(xraysep,tempsep,10, 'b+');
```

Bibliography

Aschwanden, M. *Physics of the solar corona*, Praxis Publishing Ltd. 2004

Balch, C. C. "SEC proton prediction model: Verification and analysis," *Radiation Measurements, 30*, 1999

Balch, C. C. "Database of solar flares, 1986 through 2004," Private Communication, 2008a

Balch, C. C. "Updated verification of the space weather prediction center's solar energetic particle prediction model," *Space Weather, Vol. 6, 6*, 2008b

Beck, P., M. Latocha, S. Rollet, & G. Stehno. "TEPC reference measurements at aircraft altitudes during a solar storm," *Advances in Space Research, 36*, 2005

Belov, A., H. Garcia, V. Kurt. "Proton Enhancements and their relation to the x-ray flares during the three last solar cycles," *Solar Physics*, 229, 135-159, 2005

Brannick, Michael T. "Logistic Regression," *Lectures,* USF. 2006, http://luna.cas.usf.edu/~mbrannic/files/regression/Logistic.html

Brooke, Harold E., Charles Doswell. "A Comparison of Measures-Oriented and Distributions-Oriented Approaches to Forecast Verification," *Weather and Forecasting.* 1996

Bulmer, M. G. *Principles of Statistics.* Cambridge Massachusetts, M.I.T. Press. 1965

Cliver, E., & H. V. Cane. "X-class soft X-ray bursts and major proton events during solar cycle 21," *Solar Terrestrial Prediction*, 1989

Cliver, E.W., & A.G. Ling. "Electrons and Protons in Solar Energetic Particle Events," *The Astrophysical Journal,* 658:1349-1356, 2006

Collins, P. "Space tourism: From earth orbit to the moon," *Advances in Space Research, 37*, 116-122. 2005

Dumitrescu, A., & G. Rote. "On the Fréchet distance of a set of curves," 2004, http://page.mi.fuberlin.de/rote/Papers/pdf/On+the+Frechet+distance+of+a+set+of+curves.pdf

Foukal, P. *Solar astrophysics* (2nd ed.) Wiley-VCH. 2004

Garcia, H. A. "Temperature and hard X-ray signatures for energetic proton events," *Astrophysics Journal, 420*(1), 422:432. 1994a

Garcia, H.A. "The Determination of Temperature and Emission Measure from GOES Two-Channel, Broadband, Soft X-Ray Measurements," *American Astronomical Society,* 24, p. 820. 1994b

Garcia, H. A. "Forecasting methods for occurrence and magnitude of proton storms with solar soft X rays," *Space Weather, 2*, 2004a

Garcia, H. A. "Forecasting methods for occurrence and magnitude of proton storms with solar hard x rays," *Space Weather,* 2, 2004b

Getly, I. L., M.L. Duldig, D.F. Smart, & M.A. Shea. "The applicability of model based aircraft radiation dose estimates," *Advances in Space Research, 36*, 1638:1644. 2005

Gumbel, E.J. *Statistics of Extremes.* Columbia University, NY. Columbia University Press, 1958

Hardin, James and Joseph Hilbe. *Generalized Linear Models and Extensions*. College Station TX 77845. Sata Corp. 2001

Heil, M., & B. Roseth. *Stardust spacecraft encounters solar flare.* 2000, http://stardust.jpl.nasa.gov/news/status/001121.html

Irizarry, Rafael A. "Overview of Machine Learning," *Essentials of Probability and Statistical Interference*. Johns Hopkins Univ. 2006, http://ocw.jhsph.edu/courses/EssentialsProbabilityStatisticalInference/PDFs/Lecture2.pdf

Kahler, S. W. "Coronal mass ejections and solar energetic particle events," *Reprint from AIP Conference Proceedings on High Energy Solar Physics 16-18 Aug, Greenbelt MD.* 1996

Kahler, S. W., N.R. Sheeley, R.A. Howard, M.J. Koomen, D.J. Michels, R.E.McGuire, et al. "Associations between coronal mass ejections and solar energetic proton events," *Journal of Geophysics Research, 89*, 9683-9694. 1984

Kahler, S. W., & A. Vourlidas. "Fast coronal mass ejection environments and the production of solar energetic particle events," *Journal of Geophysics Research, 110*. 2005

Kahler, S. W., E.W. Cliver, A.G. Ling. "Validating the proton prediction system," *Journal of Atmospheric and Solar-Terrestrial Physics*, 69, 43-49. 2007

Kaiser, Michael L. "The Radio and Plasma Wave Investigation on the WIND Spacecraft," 2008, http://lep694.gsfc.nasa.gov/waves/waves.html

Kallender, P. "Solar flare hobbles Japanese communications satellite," 2003, http://www.space.com/news/kodama_down_031029.html

McClave, J. T., P.G. Benson, & T. Sincich. *Statistics for business and economics* (10th ed.). Upper Saddle River NJ 07458. Pearson Education. 2008

Pelletier, S. "Computing the Fréchet distance between two polygonal curves," 2002, http://www.cim.MFgill.ca/~stephane/cs507/Project.html

Smart, D. F., & M.A. Shea. "The heliolongitudinal distribution of solar flares associated with solar proton events," *Adv. Space Res, 17*(2), 113-116. 1996

Smart, D. F., & M.A. Shea. "Comment on estimating the solar proton environment that may affect Mars missions," *Advances in Space Research, 31*(1), 45:50. 2003

Space Weather Prediction Center. "Preliminary Report and Forecast of Solar Geophysical Data," 2008, http://www.swpc.noaa.gov/weekly/index.html

Tascione, T. *Introduction to the space environment* (2nd ed.). Malabar FL: Krieger Publishing Co. 1994

White, Stephen M., Roger J. Thomas, Richard A. Schwartz. "Updated Expressions for Determining Temperature and Emission Measure from GOES Soft X-ray Measurements," *Solar Physics*, 227,231-248. 2005

Yashiro, Seiji. "SOHO LASCO CME Catalogue," *CDAW Data Center*, 2008, http://cdaw.gsfc.nasa.gov/CME_list/.

Xiaocong, L. "Solar proton events and fluctuations in the density of electrons trapped within the earth's magnetosphere," *Solar Physics, 201*(2), 393-403. 2001

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE (DD-MM-YYYY)	**2. REPORT TYPE**	**3. DATES COVERED** (From – To)
27-03-2009	Master's Thesis	06-2008 to 03-2009

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Predicting Solar Protons: A Statistical Approach	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Spaulding, Jonathan C., Capt, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology	AFIT/GAP/ENP/09-M09
Graduate School of Engineering and Management (AFIT/EN)	
2950 Hobson Way	
WPAFB OH 45433-7765	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Intentionally Left Blank	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A small fraction of solar flares are accompanied by high energy (>10 MeV) protons. These events can cause degradation or failure of satellite systems and can be harmful to humans in space or in high altitude flight. For risk management purposes, the Air Force is interested in predicting these events. Several algorithms exist to do this operationally, but none predict when these events will occur with much accuracy. Here, we analyzed 3610 M1 and greater flares including 106 with proton events from the GOES sensors from 1 Jan 1986 to 31 Dec 2004 to produce new results, including a full scale comparison and optimization for all the algorithms. In every case, optimization leads to increased prediction ability. This research also produced a new algorithm based on the Garcia algorithm, which functions better than any other operational algorithm. This model, Garcia 2008, predicts with a skill score of .526, an improvement from .342. This new model is the best at prediction of all models measured.

15. SUBJECT TERMS
Solar Energetic Protons, Solar Flares, Protons, Solar Corona, Cosmic Radiation

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Maj. A. Acebal AFIT/ENP
REPORT U	**ABSTRACT** U	**c. THIS PAGE** U	UU	93	**19b. TELEPHONE NUMBER** (Include area code) 937-255-3636x4518 Ariel.Acebal@afit.edu