A statistical solar flare forecast method

M. S. Wheatland

School of Physics, University of Sydney, Sydney, New South Wales, Australia

[1] A Bayesian approach to solar flare prediction has been developed which uses only the event statistics of flares already observed. The method is simple and objective and makes few ad hoc assumptions. It is argued that this approach should be used to provide a baseline prediction for certain space weather purposes, upon which other methods, incorporating additional information, can improve. A practical implementation of the method for whole-Sun prediction of Geostationary Observational Environment Satellite (GOES) events is described in detail and is demonstrated for 4 November 2003, the day of the largest recorded GOES flare. A test of the method is described on the basis of the historical record of GOES events (1975−2003), and a detailed comparison is made with U.S. National Oceanic and Atmospheric Administration (NOAA) predictions for 1987−2003. Although the NOAA forecasts incorporate a variety of other information, the present method outperforms the NOAA method in predicting mean numbers of event days for both M-X and X events. Skill scores and other measures show that the present method is slightly less accurate at predicting M-X events than the NOAA method but substantially more accurate at predicting X events, which are important contributors to space weather.

1. Introduction

[2] Large solar flares are associated with a variety of space weather effects. Those effects which occur promptly motivate flare prediction, or forecasting. For example, soft X-ray enhancements due to large flares cause increased ionization of the upper atmosphere, which can cause short-wave radio fade-outs, and Australia's Ionospheric Prediction Service (IPS) issues flare predictions on this basis (see http://www.ips.gov.au). Delayed space weather effects allow the possibility of physical modeling, given the knowledge that the flare has occurred. In this case prediction of the flare itself may be less important.

[3] Solar flare prediction remains in its infancy, because of a lack of detailed understanding of the physical processes underlying flares [e.g., *Priest and Forbes*, 2002]. Existing prediction methods are probabilistic. One popular approach relies on the McIntosh optical classification of sunspots, which divides sunspot groups into 60 classes on the basis of three parameters [*McIntosh*, 1990; *Bornmann and Shaw*, 1994]. The historical rate of flaring for a given classification provides an initial estimate for the expected flaring rate of an observed sunspot group. This approach is the basis for predictions published by the Space Environment Center of the U.S. National Oceanic and Atmospheric Administration (NOAA) (see http://www.sec.noaa.gov/ftpdir/latest/daypre.txt) as well as NASA [*Gallagher et al.*, 2002] (see http://beauty.nascom.nasa.gov/arm/latest/) and IPS. NOAA/SEC uses an ''expert system'' developed by *McIntosh* [1990], and adopted in 1987. The associated code begins with the McIntosh

classification but also incorporates additional information, including dynamical properties of spot growth, rotation and shear, magnetic topology inferred from sunspot structure, magnetic classification, and previous (large) flare activity. The method involves more than 500 decision rules including ''rules of thumb'' provided by human experts.

[4] A variety of properties of active regions are known to correlate with flare activity, and in principle could be incorporated into predictions. Attention has focused on photospheric magnetic field measurements. For example, *Sammis et al.* [2000] confirmed that most large flares occur in large, magnetically complex regions. Studies of vector magnetograms have suggested the length of the strongly sheared strong field region along a neutral line as a predictor of flares and coronal mass ejections (CMEs) [e.g., *Hagyard*, 1990; *Falconer*, 2001]. Rapid emergence of new magnetic flux is also often identified as being associated with flare occurrence [e.g., *Schmieder et al.*, 1994]. Recently *Leka and Barnes* [2003] examined the relationship between moments of quantities constructed from vector magnetic field maps and flaring.

[5] There are many problems with existing methods of flare prediction. One problem with classification-based approaches is that they tend to ignore the variability in flaring rate within a class. A related difficulty is that choices for classes are ad hoc, and possibly subjective. For example, the McIntosh classification is an arbitrary construction—other choices of the three parameters could be made, and different observers might disagree with a given classification. Similar criticisms apply to the inclu-

sion of additional information, for example, properties of an active region, in existing methods of prediction. The choice of properties is essentially arbitrary, and it is unclear how the information can be included in an objective way.

[6] The best indicator to future flaring activity is past flaring activity [e.g., *Neidig et al.*, 1990]. (In the prediction literature, the tendency of an active region which has produced large flares in the past to produce large flares in the future is termed persistence.) Recently *Wheatland* [2004] presented a Bayesian approach to solar flare prediction which uses the observed history of large and small flares together with the known phenomenological rules of flare occurrence to make a prediction for flaring. This "event statistics" approach has the advantage that it depends only on past flaring activity, and so avoids the ad hoc choices implicit in other prediction methods. The method has been developed into a practical automated prediction scheme for whole-Sun prediction of soft X-ray flares based on NOAA solar event lists for the Geostationary Observational Environmental Satellites (GOES). The method has been tested on the historical catalog of GOES events, and a brief account of this test was given by *Wheatland* [2005].

[7] Because of its simplicity and relative objectivity, the event statistics method is well suited to providing a baseline forecast for whole-Sun flaring for space weather purposes. Other methods of prediction, incorporating additional information, may then be applied to improve upon this baseline. In this paper the method and its implementation are described in detail. Section 2 reiterates the simple theory of the method [*Wheatland*, 2004]. Section 3 describes the practical implementation for whole-Sun prediction of GOES events, and section 4 gives a detailed account of the results of the test of this implementation on historical GOES data, including comparison with NOAA predictions for 1987−2003. Section 5 presents a brief summary and discussion.

2. Event Statistics Method

[8] It is well known that the size distribution of flares (i.e., the distribution of some measure of flare size, such as peak flux in soft X-rays) obeys a power law distribution [e.g., *Crosby et al.*, 1993]. For a given choice S of the measure of size the distribution may be written

$$N(S) = \lambda_1(\gamma - 1)S_1^{\gamma-1}S^{-\gamma}, \qquad (1)$$

where $N(S)dS$ is the number of events per unit time with size in the range S to $S + dS$, the quantity λ_1 is the total rate of events above the size S_1, and γ is the power law index. The value of the power law index depends on the measure of size S. For flare peak fluxes in X-ray bands, γ is generally found to be in the range 1.7−1.9 [e.g., *Drake*, 1971; *Hudson et al.*, 1969; *Hudson*, 1991; *Lee et al.*, 1995; *Shimizu*, 1995; *Feldman et al.*, 1997; *Aschwanden et al.*, 1998].

[9] Typically we are interested in the rate of occurrence of large events. If we denote the large size of interest S_2, then the corresponding rate may be denoted λ_2, and according to the distribution (1), this rate is given in terms of the rate λ_1 by

$$\lambda_2 = \lambda_1\left(\frac{S_1}{S_2}\right)^{\gamma-1}. \qquad (2)$$

[10] The power law size distribution is one phenomenological rule of flare occurrence. A second rule is that on short timescales, flares appear to occur as a Poisson process in time. On longer timescales flare occurrence may be described as a time-dependent Poisson process [e.g., *Biesecker*, 1994; *Wheatland*, 2001]. Assuming Poisson statistics, the probability of at least one large flare within a time ΔT is
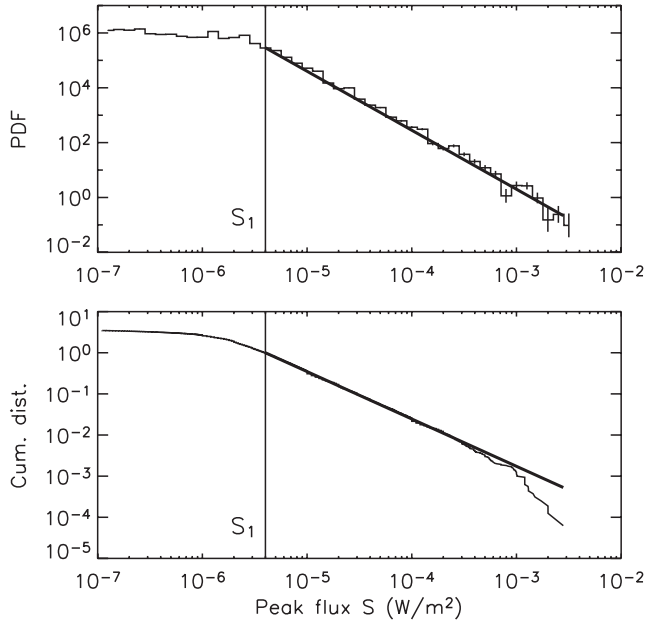
$$\epsilon = 1 - \exp(-\lambda_2\Delta T). \qquad (3)$$

[11] Equations (2) and (3) provide a naive prediction. If the power law index γ is estimated, and the current rate λ_1 of small events is estimated, then ϵ is the required probability. The Bayesian generalization involves calculating a posterior probability distribution $P(\epsilon)$ for the unknown parameter ϵ, given the available data and relevant background information [e.g., *Box and Tiao*, 1992]. Specifically we assume that the data are a sequence of M events with sizes $s_1, s_2, ., s_M$ (where $s_i \geq S_1$ for each i), observed during an observation interval from $t = 0$ to $t = T$. The events are assumed to occur at times $0 \leq t_1 < t_2 < . < t_M \leq T$. We also assume that an interval of time from $t = T - T'$ to $t = T$ has been identified during which the mean rate of events appears to be constant. A number M' of events is assumed to occur during this interval, which has duration T'. A practical solution to identifying this interval is provided by the Bayesian blocks procedure [*Scargle*, 1998] which is a Bayesian change point algorithm designed to identify times of rate variation. This procedure is discussed in more detail below. As shown by *Wheatland* [2004], the power law index may be approximated from the data by the maximum likelihood value

$$\gamma^* = \frac{M}{\ln \pi} + 1, \quad \text{where} \quad \pi = \prod_{i=1}^{M} \frac{s_i}{S_1}, \qquad (4)$$

and then the posterior distribution for ϵ, based on γ^*, M' and T' is

$$P(\epsilon) = C[-\ln(1-\epsilon)]^{M'}(1-\epsilon)^{(T'/\Delta T)(S_2/S_1)^{\gamma^*-1}-1}$$
$$\times \Lambda\left[-\frac{\ln(1-\epsilon)}{\Delta T}\left(\frac{S_2}{S_1}\right)^{\gamma^*-1}\right], \qquad (5)$$

Figure 1. (top) Probability density function (PDF) for the peak flux (1−8 Å) of GOES events 1975−2003, together with the threshold S_1 (vertical line) and the power law model (thick line). (bottom) Cumulative distribution.

where $\Lambda(\lambda_1)$ is the prior distribution for λ_1, and C is the normalization constant, determined by the requirement $\int_0^1 P(\epsilon)d\epsilon = 1$. The prior distribution $\Lambda(\lambda_1)$ describes the values we would assign to λ_1 in the absence of any data. For a given prior distribution, the mean of the posterior distribution provides an estimate for ϵ, and the width of the distribution provides an estimate of the associated uncertainty.

[12] The approximation involved in using equation (4) is valid provided the power law index γ is narrowly defined by the data, which is true for large M. For smaller numbers of events it is necessary to simultaneously infer γ and ϵ, and the details are in work by *Wheatland* [2004]. For the applications in sections 3 and 4, equation (5) is found to be sufficiently accurate.

## 3. Application to GOES Events
3.1. Implementation

[13] The method has been applied to event lists constructed from Geostationary Observational Environmental Satellite (GOES) observations at 1−8 Å. The relevant measure of size, S is the peak X-ray flux in the 1−8 Å channel, and it should be noted from the outset that this flux includes background. The GOES peak flux is routinely used to classify flares, with moderate-sized flares labeled M class (i.e., having a peak flux exceeding $S_2 = S_M = 10^{-5}$ W m^{-2}), and big flares labeled X class (peak

flux exceeding $S_2 = S_X = 10^{-4}$ W m^{-2}). Hence we are interested in predicting M class and X class events. Predictions are made for the occurrence of at least one event within $\Delta T = 1$ day of the prediction time. We take the time of events to correspond to the time of the peak flux recorded in the NOAA event lists.

[14] It is necessary to choose a threshold size S_1 above which event sizes are assumed to be power law distributed. Figure 1 shows the peak flux distribution of all GOES events for 1975−2003 from the historical event lists available from the NOAA National Geophysical Data Center (see ftp://ftp.ngdc.noaa.gov). The top plot shows the data plotted as a probability density function, (PDF, with bins of one tenth of a decade), and the bottom plot shows the data plotted as a cumulative distribution (without binning). Figure 1 shows that the data are power law distributed for large peak fluxes. At low peak fluxes there is a departure from power law behavior, which is due to the problem of identifying small events against the time varying soft X-ray background. A nominal threshold $S_1 = 4 \times 10^{-6}$ W m^{-2} for power law behavior has been chosen, and is indicated by the vertical solid line. Also shown by a thick line is the power law model distribution $P(S) = (\gamma^* - 1)S_1^{\gamma^*-1}S^{-\gamma^*}$, with the maximum likelihood index γ^* given by equation (4). For these data we find $\gamma^* \approx 2.15 \pm 0.01$. We note that this power law index is significantly larger than indices quoted in the literature for soft X-ray peak fluxes [e.g., *Aschwanden et al.*, 1998], which are in the range 1.7−1.9. The reason is that the data used here are not background subtracted. For the prediction purposes it is preferable not to background subtract, because we wish to make a prediction about the flux including background. The method outlined here requires only that the quantity S is power law distributed above the chosen threshold, which is confirmed by Figure 1. We will return to this point in section 5.

[15] Predictions are made using data within a window of duration T, which is taken to span one year prior to the prediction time. Equation (4) is applied to the year of events to determine γ^*. Then the Bayesian blocks procedure [*Scargle*, 1998] is applied to the year of data. This procedure takes the sequence of event times $t_1, t_2, ., t_M$ and returns a sequence of change point times $t_{B0} < t_{B1} < . < t_{BK}$ at which the rate is determined to change (where t_{B0} and t_{BK} are the start and end time of the data window), and a corresponding sequence $\lambda_{B1}, \lambda_{B2}, ., \lambda_{BK}$ of rates. The change points define K "Bayesian blocks," that is, intervals described by a single rate. The last Bayesian block provides T' and M' according to $T' = t_{BK} - t_{B(K-1)}$ and $M' = \lambda_{BK}T'$.

[16] The Bayesian blocks procedure involves Bayesian hypothesis testing, in which a single rate Poisson model is compared with a dual rate Poisson model for the data, for all possible choices of change points coincident with event times. If the dual rate model is more likely (by a factor PRIOR, nominally taken to have the value two) the section of data is segmented, and the two segments are them-
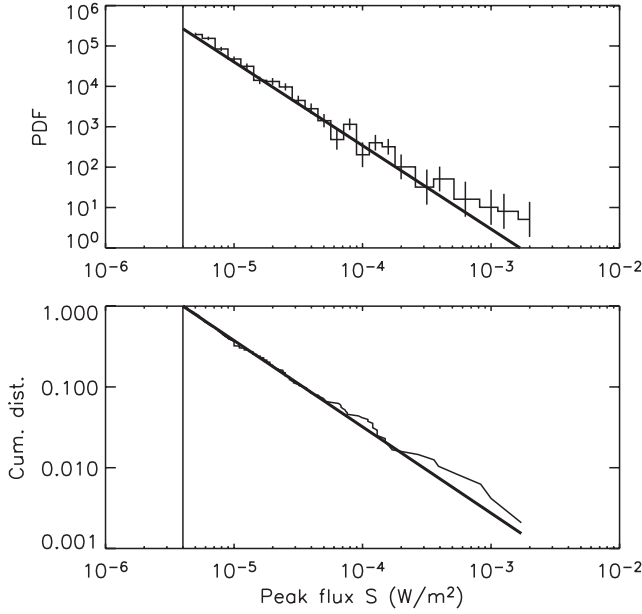
Figure 2. (top) PDF for peak flux of 1−8 Å GOES events above threshold for one year prior to 4 November 2003. (bottom) Cumulative distribution.

selves subject to the test. This process is iterated. A segment is deemed complete when the single rate model is more likely, or when there is only one event in the segment. It should also be noted that the Bayesian blocks procedure requires event times in discrete time steps, and to this purpose the peak times are rounded to the nearest minute.

[17] The rates in the Bayesian blocks before the last block contain information about how the flaring rate varies, and so are used to construct the prior $\Lambda(\lambda_1)$ in equation (5). The model form

$$\Lambda(\lambda_1) = a \exp\left(-b\lambda_1^c\right) \quad (6)$$

was chosen for the prior on the basis of inspection of a rate distribution constructed from a Bayesian blocks decomposition of the GOES event lists for 1976−2003. For each prediction the parameters a, b, and c are determined, for the given one-year window of data, by requiring that the first three moments of the model match the first three moments of the data, estimated from the Bayesian blocks results. Specifically we require

$$\int_0^\infty d\lambda_1 \Lambda(\lambda_1) = 1$$
$$\int_0^\infty d\lambda_1 \lambda_1 \Lambda(\lambda_1) = \sum_{i \neq K} N_{Bi} / \sum_{i \neq K} T_B i \quad (7)$$
$$\int_0^\infty d\lambda_1 \lambda_1^2 \Lambda(\lambda_1) = \sum_{i \neq K} \lambda_{Bi}^2 T_{Bi} / \sum_{i \neq K} T_{Bi},$$

where N_{Bi} and $T_{Bi} = t_{Bi} - t_{B(i-1)}$ denote the number of events in, and duration of, the ith Bayesian block respectively, where $\lambda_{Bi} = N_{Bi}/T_{Bi}$, and where the summations exclude the last block. As shown in Appendix A, these three conditions uniquely determine values of a, b, and c.

[18] Once γ^*, M' and T' have been determined, and the prior has been constructed, equation (5) is used (with $S_2 = S_M$ and $S_2 = S_X$) to construct posterior distributions $P_M(\epsilon)$ and $P_X(\epsilon)$ for the probability of occurrence of events above M size and above X size respectively. The means of these distributions are taken as suitable estimates ϵ_M and ϵ_X of the probabilities of at least one M class event (or larger), and at least one X class event within $\Delta T = 1$ day:

$$\epsilon_M = \int_0^1 d\epsilon \, \epsilon P_M(\epsilon)$$
$$\quad (8)$$
$$\epsilon_X = \int_0^1 d\epsilon \, \epsilon P_X(\epsilon).$$

Corresponding uncertainties σ_M and σ_X may be obtained in the usual way from the first and second moments of the posterior distributions.

[19] Another quantity of interest is the probability of getting at least one flare of M class, that is, a flare with peak flux greater or equal to 10^{-5} W m^{-2}, but less than 10^{-4} W m^{-2}. To avoid ambiguity, we will refer to events with peak flux in this range as M-X flares. If ϵ' describes the probability of at least one M class event, or larger, and ϵ'' describes the probability of at least one X class event, or larger, then the desired probability is $\epsilon = \epsilon' - \epsilon''$. If we denote the corresponding posterior distribution $P_{MX}(\epsilon)$ then we have

$$P_{MX}(\epsilon) = \int_0^1 d\epsilon' \int_0^1 d\epsilon'' \delta(\epsilon - \epsilon' + \epsilon') P_M(\epsilon') P_X(\epsilon'')$$
$$= \int_0^1 d\epsilon' \, P_M(\epsilon') P_X(\epsilon' - \epsilon). \quad (9)$$

The estimate for this quantity is taken to be the mean of the posterior distribution, which is denoted ϵ_{MX}, and similarly the associated uncertainty is denoted σ_{MX}.

3.2. Example: Application to 4 November 2003

[20] To illustrate the application of the method, we consider a prediction for the time 0000 UT on 4 November 2003. The largest soft X-ray flare of the period of GOES observations (1975−2004) occurred starting at 1929 UT on that day. The event saturated the GOES detectors, but was estimated by NOAA to be X28, that is, to have a peak flux of 2.8×10^{-3} W m^{-2} in the 1−8 Å band. The enhanced X-ray flux due to this flare caused a substantial lowering of the D region of the ionosphere, suggesting an even higher
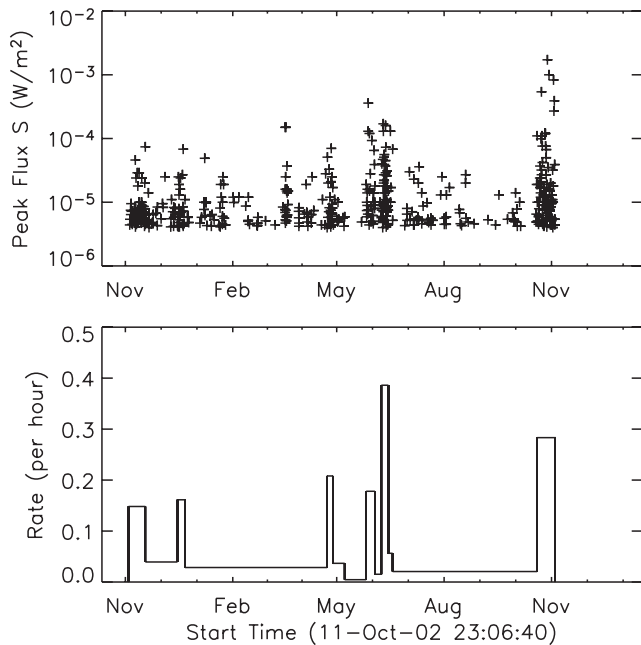
Figure 3. (top) Peak flux of $1-8$ Å GOES events (crosses) above threshold versus time for one year prior to 4 November 2003. (bottom) Bayesian blocks decomposition of rate versus time.

peak flux of around X45 [*Thomson et al.*, 2004]. Figures 2, 3, and 4 illustrate the prediction process for this day.

[21] Figure 2 shows the size distribution of the one-year window of data (prior to 4 November 2003 0000 UT) used in the prediction. Only events above the threshold flux $S_1 = 4 \times 10^{-6}$ W m^{-2} (which is indicated by a vertical line) are shown, and there are 480 events in all. The top plot shows the PDF, and the bottom plot shows the cumulative distribution. The power law model (with maximum likelihood index $\gamma^* \approx 2.07 \pm 0.05$) is indicated by a thick line. Once again we note that this value is significantly larger than other values for the distribution of soft X-ray peak flux quoted in the literature, because of the lack of subtraction of background flux.

[22] Figure 3 illustrates the inference on the rate. The top plot shows the 480 events (indicated by crosses) above the threshold during the one-year window, as a plot of peak flux versus time. The bottom plot shows the result of applying the Bayesian blocks procedure to these data. The procedure decomposed the data into 13 blocks. The last block had a duration $T' \approx 15.3$ days and contained $M' = 104$ events.

[23] Figure 4 shows the posterior distributions $P_{MX}(\epsilon)$ (top) and $P_X(\epsilon)$ (bottom). The estimates ϵ_{MX} and ϵ_X obtained from the means of the distributions are shown as vertical lines. The result is that the probability of at least one M-X event within one day is $\epsilon_{MX} \approx 0.73 \pm 0.03$, and the probability of at least one event of X size within one day is $\epsilon_X \approx 0.19 \pm 0.02$. These values are

quite high, reflecting the high rate of flaring immediately prior to 4 November. However, the results also show the limitations of probabilistic forecasting—the prediction for an X class event is only around 20%, yet the largest flare of the last three decades is imminent.

4. Test of the Method

4.1. Results

[24] To test the method, predictions were performed for each day of GOES event data for $1975-2003$. Each prediction was compared with the historical fact of whether flares did or did not occur on the given day. A brief description of the test was given by *Wheatland* [2005]. Here the test is described in more detail, including comparison, in section 4.2, of the prediction statistics with published NOAA predictions for $1987-2002$.

[25] Predictions were made according to the procedure outlined in section 3.1. Since the predictions use a one-year window of data, the results are only considered for $1976-2003$; that is, the predictions for 1975 are omitted from consideration because they are based on less than a year of previous data.

[26] Figure 5 summarizes the results, as plots of yearly numbers of observed event days, that is, days with one or more events (solid histograms), and yearly numbers of predicted event days (diamonds). The top plot shows the results for M-X events, and the bottom plot shows the results for X events. The numbers of predicted event days are the sums of the prediction estimates ϵ_{MX} and ϵ_X over all days in a given year. Representative error bars are
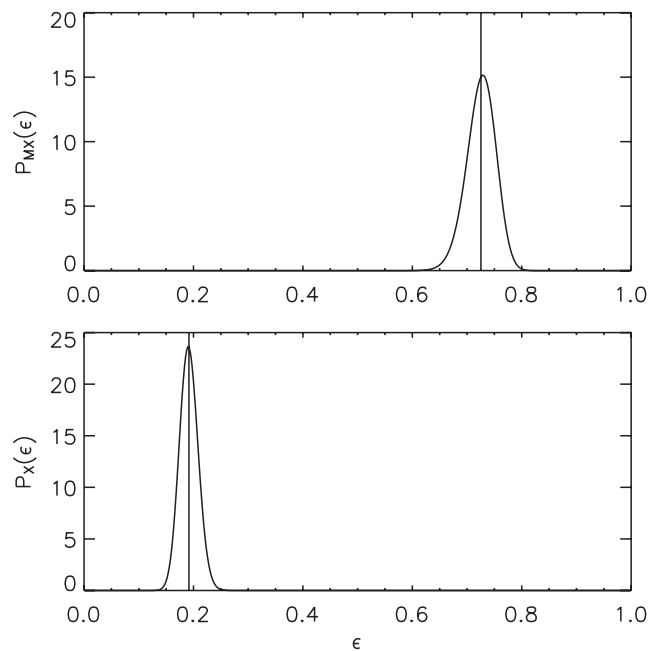
Figure 4. (top) Posterior distribution for probability of M-X events for 4 November 2003. (bottom) Posterior distribution for X events for the same day.
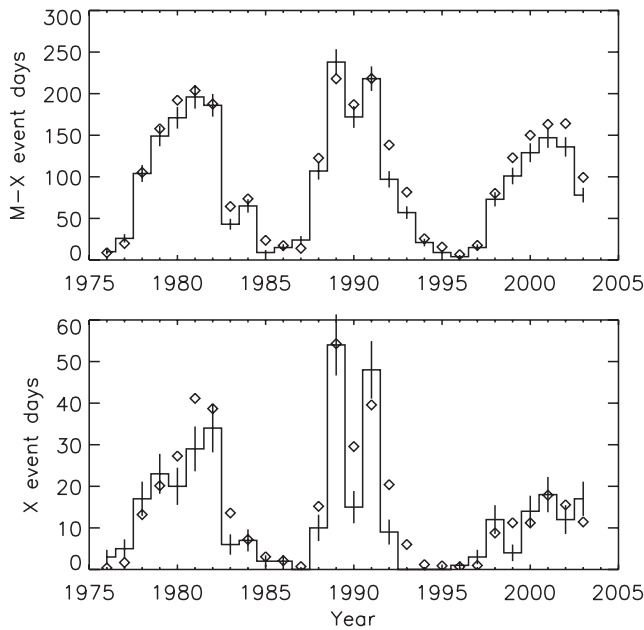
Figure 5. Comparison of predictions and observations for 1976−2003: observed, histogram; predicted, diamonds. (top) Event days for M-X events. (bottom) Event days for X events.

shown for the observations, corresponding to the square root of the number of events. These error bars reflect the expected variability in the observed numbers. These plots suggest that the method does quite well in predicting overall event numbers, although some systematic overprediction is apparent.

[27] Table 1 summarizes the statistics of the predictions for 1976−2003, in a notation following meteorological practice [e.g., *Murphy and Winkler*, 1987]. To assess probabilistic predictions, the joint probability distribution for forecasts (denoted f) and observations (denoted x) may be constructed, and properties of this distribution examined. In the present context we have $f = \epsilon_{MX}$ or $f = \epsilon_X$, for predictions for M-X or X events respectively. The value of x for each day is zero or one, according to whether an event did or did not occur. Averages over all days are denoted by $\langle \cdots \rangle$. For example, $\langle f \rangle$ is the average of the forecast probability over all days. As a second example, $\langle f | x = 1 \rangle$ is the average of the forecast probability over all days on which at least one flare did occur. Standard deviations over all days are denoted by σ. MAE denotes the mean absolute error

$$\text{MAE}(f, x) = \langle |f - x| \rangle, \qquad (10)$$

and MSE denotes the mean square error

$$\text{MSE}(f, x) = \langle (f - x)^2 \rangle. \qquad (11)$$

The linear association is the correlation of f and x.

[28] Table 1 also gives the climatological skill score [e.g., *Murphy and Epstein*, 1989], defined by

$$\begin{aligned} \text{SS}(f, x) &= 1 - \text{MSE}(f, x)/\text{MSE}(\langle x \rangle, x) \\ &= 1 - \text{MSE}(f, x)/\sigma_x^2, \qquad (12) \end{aligned}$$

which is a measure of the improvement of the forecasts over a constant forecast given by the average. Perfect prediction ($f = x$) corresponds to a skill score of unity. A positive skill score indicates better performance, and a negative skill score worse performance, with respect to the average.

[29] Table 1 indicates that the method performs quite well in describing the overall frequency of occurrence of flares: we have $\langle \epsilon_{MX} \rangle \approx 0.282$, whereas M-X events occurred on a fraction 0.254 of days; and $\langle \epsilon_X \rangle \approx 0.040$, whereas X class events occurred on a fraction 0.036 of days. However, these values also confirm that there is a tendency for overprediction.

[30] The method also has good discrimination; that is, it assigns substantially higher values to f on days on which events occurred, compared with nonevent days. The skill scores for the method are 0.272 (for M-X events) and 0.066 (for X events).

[31] Another way to summarize the results is in terms of "reliability plots," which show the success of the predictions as a function of ϵ_{MX} or ϵ_X. Figure 6 shows the reliability plot for M-X events. This diagram is constructed as follows. The predictions ϵ_{MX} for all days are sorted into bins of width 0.05. For each bin, the observed number of those days on which at least one event did occur is used to estimate the underlying probability of an event on those days, and this is the vertical value for the bin. Specifically, the estimate used is the Bayesian estimate assuming binomial statistics and a uniform prior: if there are R days with at least one event out of a total of S days, then the

Table 1. Verification Statistics for the Prediction of M-X and X Events for 1976−2003

	M-X	X
Total days	10,226	10,226
Event days	2600	365
$\langle f \rangle$	0.282	0.040
$\langle x \rangle$	0.254	0.036
Median f	0.218	0.017
σ_f	0.251	0.058
σ_x	0.435	0.186
$\langle f | x = 1 \rangle$	0.508	0.120
$\langle f | x = 0 \rangle$	0.204	0.038
Median $f	x = 1$	0.565	0.105
Median $f	x = 0$	0.116	0.016
SD $f	x = 1$	0.216	0.084
SD $f	x = 0$	0.212	0.055
MAE(f, x)	0.277	0.068
MSE(f, x)	0.138	0.032
Linear association	0.528	0.262
SS(f, x)	0.272	0.066
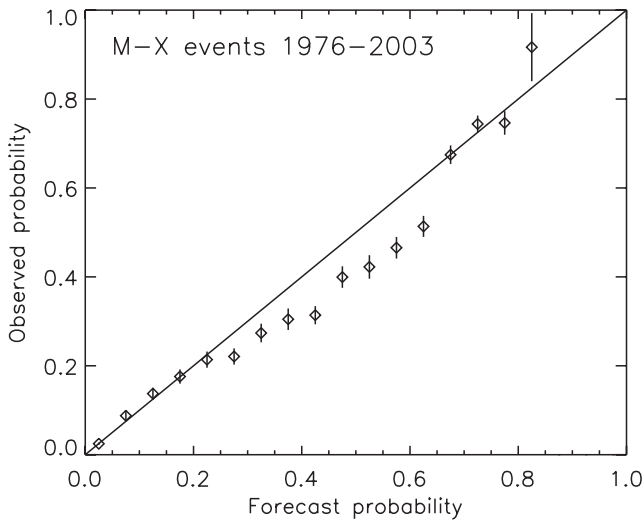
Figure 6. Reliability plot for prediction of M-X events for 1976−2003. The horizontal axis shows the probabilities assigned in the predictions, and the vertical axis shows probabilities derived from the observed frequencies of event occurrence.

estimate for the probability is $p = (R + 1)/(S + 2)$ (Laplace's rule of succession), and the associated uncertainty is $[p(1 − p)/(S + 3)]^{1/2}$ [e.g., *Jaynes*, 2003]. On a reliability plot, perfect prediction corresponds to a 45 degree line, which is indicated by the solid line in Figure 6. Figure 6 shows that the method has performed quite well for prediction of M-X events. There is some overprediction (the points fall below the perfect prediction line) for days on which the forecast has moderate values ($\epsilon_{MX} \approx 0.25−0.65$).

[32] Figure 7 shows the reliability plot for prediction of X events for 1976−2003. The predictions are conservative, in that ϵ_X is less than about 0.5 for all days. Once again the method appears to perform quite well, with some tendency to overprediction.

[33] Although we have argued that the present prediction method involves relatively few ad hoc choices, there are a number of free parameters, in particular PRIOR, the prior ratio involved in the Bayesian blocks routine, S_1, the threshold for the power law model, T, the length of the data window used in the predictions, as well as the functional form (6) for the prior rate distribution. It is interesting to briefly examine the effect on the predictions of varying some of these choices. Table 2 shows the effect on $\langle \epsilon_{MX} \rangle$ and $\langle \epsilon_X \rangle$ of varying these choices one by one. The observed means are 0.254 and 0.036 for M-X events and X events, respectively. Table 2 suggests that the method is relatively insensitive to the choices PRIOR and T. If the threshold S_1 is halved, the predictions are higher, and hence worse. This effect is probably due to the departure from power law behavior at small sizes (see Figure 1) causing the inferred power law index to

be too small. Table 1 also indicates that if the prior for the rate is ignored (corresponding to the choices $a = 1$, $b = 0$, $c = 1$ in equation (6)) the predictions are worse. This shows that the Bayesian blocks before the last are providing useful information for prediction.

4.2. Comparison With NOAA Predictions

[34] As described in section 1, the NOAA uses the McIntosh expert system [*McIntosh*, 1990] to make flare predictions. The NOAA publishes web pages with tables of statistics describing the reliability of its flare forecasts for the period 1987−2003 (see http://www.sec.noaa.gov/ forecast_verification/). Using these tables, it is possible to compare the NOAA predictions with those of the present method.

[35] Table 3 compares the predictions of the two methods for 1987−2003. Table 3 lists the means of the forecast probabilities $f = \epsilon_{MX}$ and $f = \epsilon_X$ for the present method and for the NOAA method, as well as the means of the observed values x. The present method gives mean prediction probabilities closer to the observations for both M-X and X events. For M-X events, both methods show similar overprediction, although the present method is slightly better. For X events the present method gives substantially improved mean prediction probabilities compared with the NOAA method.

[36] The average of f is an incomplete measure of the success of a prediction method, since it ignores, for example, whether high predictions are assigned on event days, and low predictions on nonevent days. Hence Table 3 also lists the average forecasts for event days and nonevent days, the mean absolute error, the mean square error, and the skill scores, for the two methods. The values of $\langle f|x = 1 \rangle$ and $\langle f|x = 0 \rangle$ show that the NOAA method is somewhat more discriminating than the present method, for both M-X and X event prediction. How-
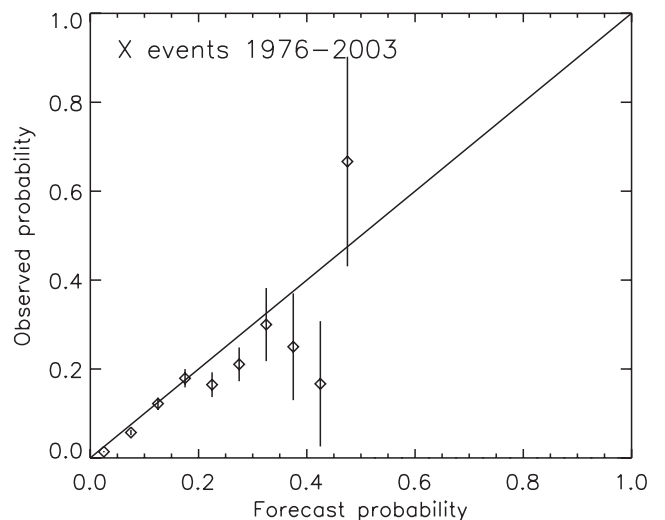
Figure 7. Same as Figure 6 but for X events.

Table 2. Dependence of Prediction of M-X and X Events on Free Parameters

PRIOR	S_1, W m^{-2}	T, years	Rate Prior?	$\langle \epsilon_{MX} \rangle$	$\langle \epsilon_X \rangle$
2	4×10^{-6}	1	yes	0.282	0.040
4	4×10^{-6}	1	yes	0.280	0.040
2	2×10^{-6}	1	yes	0.286	0.052
2	4×10^{-6}	2	yes	0.285	0.041
2	4×10^{-6}	1	no	0.289	0.069

ever, the present method has a lower mean prediction for nonevent days for X class flares. The mean absolute and mean square errors suggest that the NOAA method is slightly more accurate for M-X event prediction, but less accurate for X event prediction. The overall skill scores also support this. Notably the NOAA skill score for X event prediction is negative.

[37] It is also interesting to compare the predictions and observations on a year by year basis. Figure 8 shows the predicted numbers of event days for the present method (diamonds), the predicted numbers of event days for the NOAA method (asterisks), as well as the observed number of event days (solid histograms), for each year in the period 1987−2003. The top plot shows the results for M-X events, and the bottom plot shows the results for X events. The predictions for M-X events for the two methods show a comparable scatter around the observed values. The predictions for X events are much better in the case of the present method, in particular for cycle 23.

[38] Figure 9 compares the skill scores (equation (12)) for the predictions by the two methods, on a year by year basis. The scores for the present method are shown by diamonds, and the scores for the NOAA method by asterisks, with the top plot showing the results for M-X events, and the bottom plot showing the results for X events. The NOAA method is seen by this measure to be slightly better at predicting M-X events but worse at predicting X events (and in particular had two years with very poor X event predictions).

5. Discussion

[39] A practical implementation of an event statistics approach to solar flare prediction [*Wheatland*, 2004] is

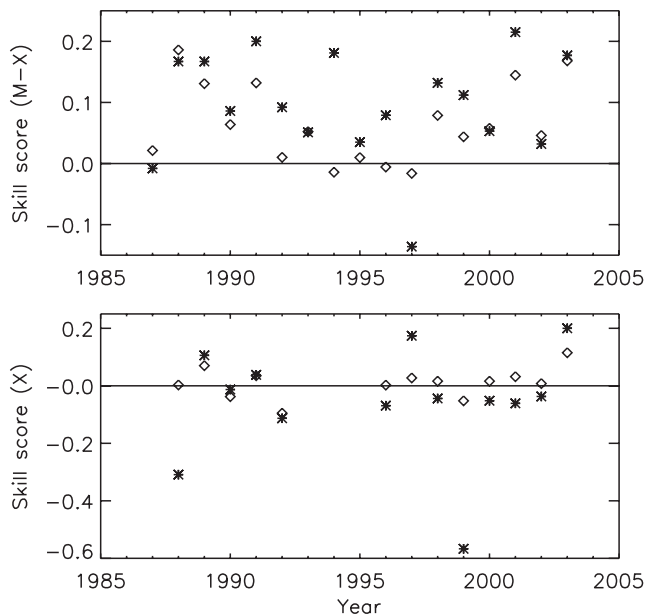**Table 3.** Comparison of Predictions of Present Method with NOAA Predictions, 1987−2003

	Present Method		NOAA Method	
	M-X	X	M-X	X
$\langle f \rangle$	0.294	0.040	0.298	0.064
$\langle x \rangle$	0.262	0.035	0.262	0.035
$\langle f \vert x = 1 \rangle$	0.510	0.122	0.551	0.244
$\langle f \vert x = 0 \rangle$	0.217	0.037	0.208	0.057
MAE(f, x)	0.289	0.066	0.271	0.081
MSE(f, x)	0.143	0.031	0.139	0.032
SS(f, x)	0.258	0.078	0.262	−0.006

demonstrated for daily whole-Sun prediction of GOES soft X-ray flares, and is illustrated by application to 4 November 2003, the day of the largest recorded GOES flare (Figures 2, 3, and 4). The method makes predictions based only on the observed history of flaring by exploiting the phenomenological rules of flare occurrence, in particular the power law distribution of flares in size (Figure 1). The method is simple, both conceptually and in terms of implementation, and involves few ad hoc assumptions. It requires no specialized observations, only the NOAA solar event lists. As such it is well suited to providing a first guess for the probability of flare occurrence. In principle additional observations of physical parameters related to flaring could be used to improve upon the basic prediction provided by the present method, and approaches to this problem will be considered in future work.

[40] The method has been tested on prediction of GOES events for each day in the period 1976−2003. The test was briefly described by *Wheatland* [2005], and a detailed account is given here. The method is found to perform well in assigning probabilities to the occurrence of events in the range M to X (M-X events) and to the occurrence of X events, although there is a tendency to overprediction, in particular for M-X events. A number of measures of success of the method are examined, including predicted numbers of event days (Figure 5) reliability plots (Figures 6 and 7), and verification statistics (Table 1). In particular, the skill scores for M-X event prediction and X event prediction are found to be 0.272 and 0.066 respectively.
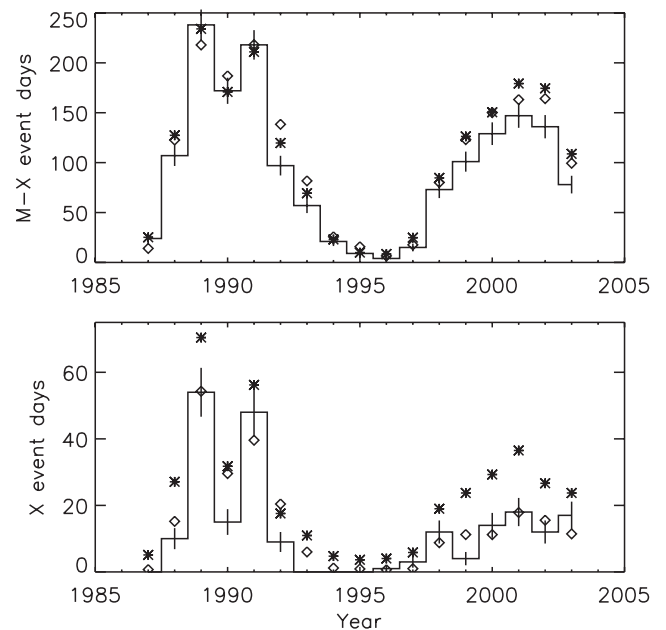
Figure 8. Comparison of observed event days (histograms), predictions of the present method (diamonds), and NOAA predictions (asterisks) for 1987−2003: (top) M-X events and (bottom) X events.

Figure 9. Comparison of skill scores for the present method (diamonds) and for NOAA predictions (asterisks) for 1987−2003: (top) M-X events and (bottom) X events.

[41] The results of the test are found to only weakly depend on a number of chosen parameters in the method, in particular a prior ratio for segmentation in the Bayesian blocks procedure used to determine the current rate, and the length of the data window used (see Table 2). However, the predictions require an accurate choice of the threshold for the power law size distribution, and benefit from the inclusion of prior information on the rate.

[42] The present method is also compared with the long-standing NOAA prediction method, using the NOAA's published prediction statistics for 1987−2003 (Table 3 and Figures 8 and 9). The event statistics method is found to outperform the NOAA method in predicting overall numbers of M-X and X event days. In particular the NOAA method is observed to seriously overpredict X class events (Figure 8, bottom plot). The NOAA method is found to be slightly more accurate in M-X event prediction but less accurate in X event prediction, on the basis of skill scores and other validation statistics (Table 3). Overall the present method provides improved prediction of X class flares (e.g., compare the skill scores SS(f, x) in Table 3). This is significant because X flares, although infrequent, are the most important flares from the point of view of space weather. It is perhaps surprising that the present method fares as well as it does, given the amount of background information incorporated into the NOAA forecasts [*McIntosh*, 1990]. These results support the contention that the event statistics method is well suited to providing a baseline forecast, upon which other methods can improve.

[43] The tendency to overpredict moderate sized flares is still being investigated, but it is likely that it stems from the determination of the current rate. The Bayesian blocks procedure is always trying to "catch up" with variations in the Sun's flaring rate. If the Bayesian blocks method is systematically late in detecting a sudden decline in rate (e.g., due to the decay of an active region, or its rotation off the disk), then there will be a period of overprediction. A specific issue with the implementation of the Bayesian blocks procedure is that there must be at least one event in a block, so that the inferred rate is never identically zero, even if the true rate is zero. This is expected to lead to overestimation of the rate at times of low activity. A related problem is that it is intrinsically difficult to accurately determine low rates because of the absence of events. In future work these questions will be examined more carefully. It should also be noted that the Bayesian blocks procedure is not guaranteed to find the optimal decomposition [*Scargle*, 1998]. Recently Scargle described a new, optimal Bayesian blocks algorithm [*Scargle*, 2004], and in future the new method will be applied to flare prediction.

[44] In section 3 it was noted that the 1−8 Å GOES peak fluxes used here are not background subtracted. For many applications of soft X-ray data, for example, determining intrinsic properties of flares, it is essential to perform accurate background subtraction [e.g., *Bornmann*, 1990]. It is also important in statistical studies where the concern is with the distribution of the intrinsic quantity, for example in studies investigating what the distributions of peak flux reveal about underlying physical processes. However, in the present context background subtraction is unnecessary. As noted in section 3, provided the peak fluxes (including background) are power law distributed, they are suitable for predictive purposes. This is an advantage of the method, in that readily available data (GOES event lists) may be used as the basis for a prediction. However, the lack of background subtraction means that the peak flux distributions constructed here cannot be readily compared with other published distributions. In particular, the power law indices obtained here are typically larger than two, whereas a large body of literature reports that the intrinsic distribution of X-ray peak flux has an index in the range 1.7−1.9 [e.g., *Drake*, 1971; *Hudson et al.*, 1969; *Hudson*, 1991; *Lee et al.*, 1995; *Shimizu*, 1995; *Feldman et al.*, 1997; *Aschwanden et al.*, 1998].

[45] Although convenient, the GOES event lists also have a specific shortcoming for predictive purposes. Events are selected against a substantial and time-varying soft X-ray background. At times of high solar activity, more events are missed because of the increased background [e.g., *Wheatland*, 2001], and this leads to a relatively large threshold for power law behavior of the peak fluxes (see Figure 1), which is a disadvantage for the present method. In future other data sets will be considered. However, the advantages of the GOES event lists are their availability, and the close relationship of soft X-ray peak

flux to an important space weather effect (ionization of the upper atmosphere).

[46] Predictions made using the method described in this paper are now published daily on the Web (see http://www.physics.usyd.edu.au/~wheat/prediction/). The Web pages also include running measures of how accurate the published predictions are, in the form of automatically updated plots of reliability and skill scores. The codes used to make the predictions are written in the Interactive Data Language (IDL)-based SolarSoft system [*Freeland and Handy*, 1998], which has become the de facto standard for solar data analysis. All codes are available on request from the author.

Appendix A: Calculation of the Prior for λ_1

[47] The moments of the model distribution (6) are given by

$$
\langle \lambda_1^\alpha \rangle \equiv a \int_0^\infty \lambda_1^\alpha \exp\left(-b\lambda_1^c\right) d\lambda_1
$$
$$
= \frac{a}{b^{(\alpha+1)/c} c} \Gamma\left(\frac{\alpha+1}{c}\right), \tag{A1}
$$

where $\Gamma(x)$ is the Gamma function. Hence the three moment equations (7) may be written

$$
\frac{a}{b^{1/c} c} \Gamma(1/c) = 1
$$
$$
\frac{a}{b^{2/c} c} \Gamma(2/c) = \overline{\lambda_1} \tag{A2}
$$
$$
\frac{a}{b^{3/c} c} \Gamma(3/c) = \overline{\lambda_1^2},
$$

where $\overline{\lambda_1}$ and $\overline{\lambda_1^2}$ denote the right hand sides of the second and third of equation (7).

[48] Eliminating a between the first and second parts of equation (A2) gives

$$
b^{-1/c} \frac{\Gamma(2/c)}{\Gamma(1/c)} = \overline{\lambda_1}, \tag{A3}
$$

and eliminating a between the first and third parts of equation (A2) gives

$$
b^{-2/c} \frac{\Gamma(3/c)}{\Gamma(1/c)} = \overline{\lambda_1^2}. \tag{A4}
$$

Eliminating b between equations (A3) and (A4) gives the transcendental equation for c:

$$
[\Gamma(2/c)]^2 \overline{\lambda_1^2} - \left(\overline{\lambda_1}\right)^2 \Gamma(1/c)\Gamma(2/c) = 0. \tag{A5}
$$

This equation needs to be solved for c, for example, using Newton-Raphson, for a suitable initial guess, and then equation (A3) gives

$$
b = \left[\Gamma(1/c)\overline{\lambda_1}/\Gamma(2/c)\right]^{-c}, \tag{A6}
$$

and the first part of equation (A2) gives

$$
a = b^{1/c} c / \Gamma(1/c). \tag{A7}
$$

References

Aschwanden, M. J., B. R. Dennis, and A. O. Benz (1998), Logistic avalanche processes, elementary time structures, and frequency distributions in solar flares, *Astrophys. J.*, *497*, 972–993.

Biesecker, D. A. (1994), On the occurrence of solar flares observed with the Burst and Transient Source experiment, Ph.D. thesis, Univ. of N. H., Durham, Dec.

Bornmann, P. L. (1990), Limits to derived flare properties using estimates for the background fluxes: Examples from GOES, *Astrophys. J.*, *356*, 733–742.

Bornmann, P. L., and D. Shaw (1994), Active region classifications, complexity, and flare rates, *Sol. Phys.*, *150*, 127–146.

Box, G. E. P., and G. C. Tiao (1992), *Bayesian Inference in Statistical Analysis*, John Wiley, Hoboken, N. J.

Crosby, N. B., M. J. Aschwanden, and B. R. Dennis (1993), Frequency distributions and correlations of solar X-ray flare parameters, *Sol. Phys.*, *143*, 275–299.

Drake, J. F. (1971), Characteristics of soft solar X-ray bursts, *Sol. Phys.*, *16*, 152–185.

Falconer, D. A. (2001), A prospective method for predicting coronal mass ejections from vector magnetograms, *J. Geophys. Res.*, *106*, 25,185–25,190.

Feldman, U., G. A. Doschek, and J. A. Klimchuk (1997), The occurrence rate of soft X-ray flares as a function of solar activity, *Astrophys. J.*, *474*, 511–517.

Freeland, S. L., and B. N. Handy (1998), Data analysis with the Solar-Soft system, *Sol. Phys.*, *182*, 497–500.

Gallagher, P. T., Y.-J. Moon, and H. Wang (2002), Active-region monitoring and flare forecasting, *Sol. Phys.*, *209*, 171–183.

Hagyard, M. J. (1990), The significance of vector magnetic field measurements, *Mem. Soc. Astron. Ital.*, *61*, 337–357.

Hudson, H. S. (1991), Solar flares, microflares, nanoflares, and coronal heating, *Sol. Phys.*, *133*, 357–369.

Hudson, H. S., L. E. Peterson, and D. A. Schwartz (1969), The hard solar X-ray spectrum observed from the third Orbiting Solar Observatory, *Astrophys. J.*, *157*, 389–415.

Jaynes, E. T. (2003), *Probability Theory: The Logic of Science*, 165 pp., Cambridge Univ. Press, New York.

Lee, T. T., V. Petrosian, and J. M. McTiernan (1995), The Neupert effect and the chromospheric evaporation model for solar flares, *Astrophys. J.*, *448*, 915–924.

Leka, K. D., and G. Barnes (2003), Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis, *Astrophys. J.*, *595*, 1296–1306.

McIntosh, P. S. (1990), The classification of sunspot groups, *Sol. Phys.*, *125*, 251–267.

Murphy, A. H., and E. S. Epstein (1989), Skill scores and correlation coefficients in model verification, *Mon. Weather Rev.*, *117*, 572–581.

Murphy, A. H., and R. L. Winkler (1987), A general framework for forecast verification, *Mon. Weather Rev.*, *115*, 1330–1338.

Neidig, D. F., P. H. Wiborg, and P. H. Seagraves (1990), The role of persistence in the 24-hour flare forecast, in *Solar-Terrestrial Predictions: Proceedings of a Workshop in Leura, Australia, October 16–20, 1989*, vol. 1, edited by R. J. Thompson et al., pp. 541–545, NOAA Environ. Res. Lab., Boulder, Colo.

Priest, E. R., and T. G. Forbes (2002), The magnetic nature of solar flares, *Astron. Astrophys. Rev.*, *10*, 313–377.

Sammis, I., F. Tang, and H. Zirin (2000), The dependence of large flare occurrence on the magnetic structure of sunspots, *Astrophys. J.*, *540*, 583–587.

Scargle, J. D. (1998), Studies in astronomical time series analysis. V. Bayesian blocks, a new method to analyze structure in photon counting data, *Astrophys. J.*, *504*, 405–418.

Scargle, J. D. (2004), Data analysis through segmentation: Bayesian blocks and beyond, paper presented at 8th High Energy Astrophysics Division Meeting, Am Astron. Soc., New Orleans, La.

Schmieder, B., M. J. Hagyard, A. Guoxiang, Z. Hongqi, B. Kalmán, L. Györi, B. Rompolt, P. Démoulin, and M. E. Machado (1994), Relationship between magnetic field evolution and flaring sites in AR 6659 in June 1991, *Sol. Phys.*, *150*, 199–219.

Shimizu, T. (1995), Energetics and occurrence rate of active-region transient brightenings and implications for the heating of the active-region corona, *Publ. Astron. Soc. Jpn.*, *47*, 251–263.

Thomson, N. R., C. J. Rodger, and R. L. Dowden (2004), Ionosphere gives size of greatest solar flare, *Geophys. Res. Lett.*, *31*, L06803, doi:10.1029/2003GL019345.

Wheatland, M. S. (2001), Rates of flaring in individual active regions, *Sol. Phys.*, *203*, 87–106.

Wheatland, M. S. (2004), A Bayesian approach to solar flare prediction, *Astrophys. J.*, *609*, 1134–1139.

Wheatland, M. S. (2005), Initial test of a Bayesian approach to solar flare prediction, *Publ. Astron. Soc. Aust.*, *22*, 153–156.

————————————

M. S. Wheatland, School of Physics, University of Sydney, Sydney, NSW 2006, Australia. (m.wheatland@physics.usyd.edu.au)