A baseline flare prediction using only event statistics

M.S. Wheatland

School of Physics
University of Sydney

Forecasting the Operational All-Clear
22-24 April 2009

The University of Sydney
Overview

Method
  Persistence
  Flare statistics
  Event statistics method

Workshop applications
  Whole-Sun, active-region prediction of GOES flares

Validation
  Accuracy of probabilistic forecasts

Summary and comments
Persistence

- Past flaring history is an indicator of future flaring
Flare statistics

- Flares obey a power-law frequency-size distribution (Drake 1971)

\[ N(S) = \lambda_1(\gamma - 1)S_1^{\gamma-1}S^{-\gamma} \]  

- \( S \): energy, or peak flux in X-ray,...
- \( N(S) \): number of flares per unit time, per unit \( S \)
- \( \gamma = \frac{3}{2} \) to \( \gamma = 2 \) (depends on specific choice of \( S \))
- \( \lambda_1 = \lambda_1(t) \): total rate above size \( S_1 \)

- Occurrence in time may be modelled as a Poisson process (e.g. Wheatland 2001)
  - If \( \lambda \) does not vary, distribution of waiting times \( \tau \):

\[ P(\tau) = \lambda \exp(-\lambda \tau) \]
Waiting-time

Frequency-size

Waiting-time

\( \log N(S) \)

\( \log P(\tau) \)

\( \log S \)

\( \tau \)

- \( S_1 \) = size of “small” event, \( S_2 \) = size of “big” event
- \( \lambda_1 \) = observed rate above \( S_1 \); PL size distribution \( \Rightarrow \)

\[
\lambda_2 = \lambda_1 \left( \frac{S_1}{S_2} \right)^{\gamma - 1} \quad (3)
\]

- even if no big events have been observed
- Probability of at least one big event in time \( T_P \) is

\[
\epsilon = 1 - \exp(-\lambda_2 T_P)
= 1 - \exp \left[ -\lambda_1 \left( \frac{S_1}{S_2} \right)^{\gamma - 1} T_P \right] \quad (4)
\]

- assuming Poisson waiting times
- If \( M \) events are involved in inferring \( \lambda_1 \) then

\[
\sigma_{\epsilon}/\epsilon \approx 1/\sqrt{M} \quad (5)
\]

- accurate if many small events observed
Method well-suited to a power-law size distribution

\[ \log N \]

Count these ones

Predict these ones

Exponential

Power law

\[ S_1 \quad S_2 \quad \log S \]
Bayesian version

- Data $D$: events $s_1, s_2, ..., s_M$ at times $t_1 < t_2 < ... < t_M$
- Infer $P_\gamma(\gamma|D)$ and $P_1(\lambda_1|D)$
- Calculate

$$P_2(\lambda_2|D) = \int_1^\infty d\gamma \int_0^\infty d\lambda_1 P_1(\lambda_1|D) P_\gamma(\gamma|D) \times \delta [\lambda_2 - \lambda_1 (S_1/S_2)^{\gamma^{-1}}] \quad (6)$$

and then

$$P_\epsilon(\epsilon|D) = P_2 [\lambda_2(\epsilon)|D] \left| \frac{d\lambda_2}{d\epsilon} \right| \quad (7)$$

where

$$\lambda_2(\epsilon) = -\ln(1 - \epsilon)/T_P \quad (8)$$

- Problem: infer power-law index and rate of small events
Inferring the power-law index $\gamma$

- Events power-law distributed, so (Bai 1993)

$$P(D|\gamma) \propto \prod_{i=1}^{M} (\gamma - 1)(s_i/S_1)^{-\gamma}$$  \hspace{1cm} (9)

- uniform prior over a range $\gamma_1 < \gamma < \gamma_2$ used
- for $M \gg 1$, peaked around

$$\gamma_{ML} = \frac{M}{\ln \pi} + 1 \quad \text{where} \quad \pi = \prod_{i=1}^{M} \frac{s_i}{S_1}$$  \hspace{1cm} (10)

- use $P_{\gamma}(\gamma|D) = \delta(\gamma - \gamma_{ML})$
Inferring the rate of small events $\lambda_1$

- Complicated by time variation
- Bayesian blocks used (Scargle 1998)
  - iterative comparison of one- versus two-rate Poisson models
  - decomposition into piecewise-constant Poisson process
  - data $D'$ in last piece (block): $M'$ events in time $T'$
  - most recent interval with constant rate

$$P_1(D'|\lambda_1) \propto \lambda_1^{M'} e^{-\lambda_1 T'} \quad (11)$$
Bayesian blocks
Whole-Sun, active-region prediction of GOES flares

- Method applied to whole-Sun prediction of GOES events (Wheatland 2005, Space Weather 3, S07003)
  - this workshop: predictions for times of MDI observations
  - magnetogram information not used
- Applied here also to specified active regions
- Illustration for 10486 and 10488
  - largest soft X-ray flare: 4 Nov 2003 20:06UT in 10486
  - 10486, 10488: flare-productive active regions on disk
  - $D$: previous events from these active regions (62 events)
  - $S_1 = 3.6 \times 10^{-6}$ W m$^{-2}$ (smallest event)
  - inference of $\gamma$: ML approximation not made
  - predictions for C1.0 (24 hr), and M1.0, M5.0 (12 hr) at 00:00UT
Bayesian blocks procedure
Frequency-size distribution of events

$\gamma = 1.64$
Posteriors for M1.0, M5.0
Table: Predictions for 10486, 10488 at 4 Nov 2003 00:00UT

<table>
<thead>
<tr>
<th>Event</th>
<th>$T_P$</th>
<th>$\varepsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>24 hrs</td>
<td>1.00 ± 0.00</td>
</tr>
<tr>
<td>M1</td>
<td>12 hrs</td>
<td>0.71 ± 0.05</td>
</tr>
<tr>
<td>M5</td>
<td>12 hrs</td>
<td>0.36 ± 0.07</td>
</tr>
</tbody>
</table>

- In fact five events $\geq$ C1 within 24 hrs, one $\geq$ M1, zero $\geq$ M5 within 12 hrs
- Workshop: applied to ARs listed in each MDI file
  - magnetogram information not used
**Accuracy of probabilistic forecasts**

- \( f = \) forecast, \( x = \) observation (0 or 1)
- Mean square error
  
  \[
  \text{MSE}(f, x) = \langle (f - x)^2 \rangle \tag{12}
  \]
- "Climatological skill score":
  
  \[
  \text{SS}(f, x) = 1 - \frac{\text{MSE}(f, x)}{\text{MSE}(\langle x \rangle, x)} \tag{13}
  \]
  
  - improvement over forecasting the average
- Reliability plots
  
  - observed frequencies versus forecast probabilities
## Table: Whole-Sun predictions for workshop

<table>
<thead>
<tr>
<th>Event</th>
<th>$T_P$</th>
<th>$\langle f \rangle$</th>
<th>$\langle x \rangle$</th>
<th>$SS(f, x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>24 hrs</td>
<td>0.94</td>
<td>0.90</td>
<td>0.13</td>
</tr>
<tr>
<td>M1</td>
<td>12 hrs</td>
<td>0.28</td>
<td>0.23</td>
<td>0.16</td>
</tr>
<tr>
<td>M5</td>
<td>12 hrs</td>
<td>0.049</td>
<td>0.049</td>
<td>0.064</td>
</tr>
</tbody>
</table>

## Table: Active-region predictions for workshop ($\geq 10$ events)

<table>
<thead>
<tr>
<th>Event</th>
<th>$T_P$</th>
<th>$\langle f \rangle$</th>
<th>$\langle x \rangle$</th>
<th>$SS(f, x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>24 hrs</td>
<td>0.89</td>
<td>0.57</td>
<td>-0.27</td>
</tr>
<tr>
<td>M1</td>
<td>12 hrs</td>
<td>0.23</td>
<td>0.15</td>
<td>0.088</td>
</tr>
<tr>
<td>M5</td>
<td>12 hrs</td>
<td>0.0835</td>
<td>0.046</td>
<td>-0.022</td>
</tr>
</tbody>
</table>
Reliability plot for whole-Sun, M5–12 hr prediction

M5_12_fdisk

Observed

0.0  0.2  0.4  0.6  0.8  1.0

Predicted

0.0  0.2  0.4  0.6  0.8  1.0

Number

10000  1000  100  10

1  1
Problem of missed small events
Summary and comments

- Bayesian method of solar flare prediction
  - uses only event statistics
  - exploits power-law size distribution
  - quantifies “persistence”
  - applications to whole-Sun, AR prediction of GOES events
  - provides “baseline” predictions: does not use the MDI data
  - very simple to apply

- Method could be improved by:
  - more complete observations of small events
  - use of data less affected by event selection problems
  - inclusion of additional information (e.g. Wheatland 2006)