

Probabilistic forecasting of solar flares from vector magnetogram data

G. Barnes,¹ K. D. Leka,¹ E. A. Schumer,² and D. J. Della-Rose³

Received 23 January 2007; revised 16 April 2007; accepted 1 June 2007; published 7 September 2007.

[1] Discriminant analysis is a statistical approach for assigning a measurement to one of several mutually exclusive groups. Presented here is an application of the approach to solar flare forecasting, adapted to provide the probability that a measurement belongs to either group, the groups in this case being solar active regions which produced a flare within 24 hours and those that remained flare quiet. The technique is demonstrated for a large database of vector magnetic field measurements obtained by the University of Hawai'i Imaging Vector Magnetograph. For a large combination of variables characterizing the photospheric magnetic field, the results are compared to a Bayesian approach for solar flare prediction, and to the method employed by the U.S. Space Environment Center (SEC). Although quantitative comparison is difficult as the present application provides active region (rather than whole-Sun) forecasts, and the present database covers only part of one solar cycle, the performance of the method appears comparable to the other approaches.

Citation: Barnes, G., K. D. Leka, E. A. Schumer, and D. J. Della-Rose (2007), Probabilistic forecasting of solar flares from vector magnetogram data, *Space Weather*, 5, S09002, doi:10.1029/2007SW000317.

1. Introduction

[2] It is generally accepted that the energy to produce a solar flare is the stored magnetic energy of an active region in which the solar magnetic field departs from a simple potential configuration. A solar flare can result in such a region if a trigger occurs to tap the stored magnetic energy [e.g., Priest and Forbes, 2002].

[3] Many efforts to build a dependable objective forecasting method have relied upon white light images of active region morphology. Because morphological characteristics are typically governed by the evolution and state of the local magnetic field, the morphological interpretations essentially act as a proxy for the magnetic field. Several flare prediction methods currently employed are not completely objective and rely with varying degree on human involvement. These methods often use an active region's McIntosh classification as a means of obtaining an initial flare rate. The McIntosh classification scheme [McIntosh, 1990] uses a visual evaluation of an active region's size and character in white light to assign it to one of 60 classes, with which are associated historical flare-production rates.

Thus an assignment provides a predicted rate of flaring in and of itself. Refinements to this approach have been developed; for example, Bornmann and Shaw [1994] applied multiple linear regression to the historical McIntosh flare-rate data in order to determine to what extent each of the three McIntosh-class parameters contributes to the rate of flaring. Gallagher *et al.* [2002] have implemented a system (<http://www.solarmonitor.org>) in which historical flare rates from the McIntosh classifications determine an active region's initial flare-production probability, and then for new forecasts the probability is modified according to Poisson statistics.

[4] An objective method of flare prediction which does not rely on the McIntosh classification has been proposed by Wheatland [2004, 2005]. It accounts for an active region's previous flare activity combined with historical flare statistics [see Moon *et al.*, 2001] to determine future flare production. However, since flares are thought to be a result of an MHD instability or magnetic reconnection, observations of the magnetic fields may be expected to yield more accurate forecasts than white light observations or flare persistence alone.

[5] One approach which has been applied to active region photospheric vector magnetic field measurements is discriminant analysis, a statistical technique for assigning a new set of measurements to one of several mutually exclusive populations [Kendall *et al.*, 1983; Anderson, 1984].

¹Colorado Research Associates Division, NorthWest Research Associates, Inc., Boulder, Colorado, USA.

²Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, USA.

³Department of Physics, United States Air Force Academy, Colorado Springs, Colorado, USA.

One of the strengths of discriminant analysis is the ability to simultaneously consider multiple parameters.

[6] *Leka and Barnes* [2003a] demonstrated discriminant analysis applied to a small sample of photospheric magnetic data as a means of distinguishing between flare-producing and flare-quiet active regions on timescales of order an hour. While individual magnetic parameters had little ability to identify an active region primed for flare-production, it was shown that certain multivariate discriminant functions might be able to distinguish between flaring and nonflaring active regions, although the small sample size precluded definite conclusions.

[7] *Leka and Barnes* [2007] demonstrated discriminant analysis applied to a much larger sample size of individual vector magnetograms, with the flare event window defined as 24 hours after the observation, consistent with the National Oceanic and Atmospheric Administration Space Environment Center's (NOAA SEC) established forecasts. Again, no single magnetic parameter could distinguish flaring from flare-quiet active regions at a dramatically improved rate above that which would arise from simply predicting all regions to be flare-quiet. As with the original small-sample study, some improvement was demonstrated with multivariate discriminant functions.

[8] However, the standard implementation of discriminant analysis as used in the work of *Leka and Barnes* [2007] produces a binary categorization of a region as flaring/flare-quiet. Here, the approach is extended to a probability forecast. As such, the results using solely vector magnetic field data as input can be compared to the results of both the SEC and *Wheatland* [2005], by way of forecast reliability plots and verification statistics described by *Wheatland* [2005].

2. Data

[9] The vector magnetic field data used here were obtained by the Mees Solar Observatory Imaging Vector Magnetograph (IVM) [*Mickey et al.*, 1996; *LaBonte et al.*, 1999]. This instrument was designed for rapid imaging spectropolarimetric observations of entire active regions with moderate (1") spatial resolution, and has been in synoptic operation for over a decade. The initial sequence of the instrument's nominal observing mode includes a "survey" of each numbered NOAA active region present on the solar disk. An accumulation of these maps over the time period 2001 to 2004 yields a statistically significant sample of observations with which to test flare forecasting methods. A longer time period of IVM data was not considered due to the possibility of systematic changes resulting from instrument upgrades. For the present analysis, we removed data that suffered from obvious defects, those that were close to the limb (centered beyond $\mu = \cos(\theta) \approx 0.5$) and those for which there were fewer than 64 pixels above the 2σ noise level. All image-plane data were resolved of the inherent 180° ambiguity in the transverse component using the University of Hawai'i

approach [*Canfield et al.*, 1993; *Metcalfe et al.*, 2006], to determine the heliographic components of magnetic field. Additional details can be found in the work of *Leka and Barnes* [2007]. No further selection for size, bipolar nature, complexity, or flaring history was imposed; the final tally is 1212 magnetograms of 496 different active regions on 430 days.

[10] An active region was classified as "flaring" if it produced at least one Soft X-ray event with a peak emission level in a specified range in the 24-hours post-magnetogram, and "flare-quiet" otherwise. The flare events were determined using the event logs for the Geostationary Operational Environmental Satellite (GOES) available through the National Geophysical Data Center (<http://www.ngdc.noaa.gov>). The peak-level thresholds were either C1.0 (1.0×10^{-6} W m $^{-2}$ peak X-Ray flux in the 1–8 Å GOES bandpass), M1.0 (1.0×10^{-5} W m $^{-2}$ peak flux level) or X1.0 (1.0×10^{-4} W m $^{-2}$ peak flux level). It is assumed that undersampling due to background X-ray contamination levels is statistically minimal. No region-associated H α or radio-burst flares were included if they did not also register as a GOES event. There are no further distinctions herein concerning the character of the flare. With these criteria, the database contains 359 magnetograms of flaring regions (29.6% of the total), with 111 of those having produced at least one M-flare or greater (9.2%), and 20 having produced at least one X-flare (1.7%); no flares of at least C class were recorded for the remaining 853 magnetograms (70.4% of the total).

2.1. Data Parameters

[11] The state of the photospheric magnetic field can be quantitatively described in numerous ways. The focus was on deriving parameters which were both analogous to the morphological descriptions used by the McIntosh classification, and which also elucidate the energy stored in the active region's magnetic field. The variables, described in detail in the work of *Leka and Barnes* [2003b], are each evaluated over the field of view and essentially fall into eight categories, describing the distribution of (1) the magnetic field vector components, B_z and B_h , (2) the inclination angle, $\gamma = \tan^{-1}(B_z/B_h)$, (3) the horizontal gradients of the magnetic fields, $|\nabla_h B|$, $|\nabla_h B_z|$, $|\nabla_h B_h|$, (4) the vertical current density, $J_z \sim \partial B_y / \partial x - \partial B_x / \partial y$, (5) the force-free parameter, $\alpha \sim J_z / B_z$, (6) the vertical portion of the current helicity density, $h_c \sim J_z B_z$, (7) the shear angle from potential, $\Psi = \cos^{-1}(\mathbf{B}^p \cdot \mathbf{B}^o / B^p B^o)$, (8) the photospheric excess magnetic energy density, $\rho_e = (\mathbf{B}^p - \mathbf{B}^o)^2 / 8\pi$, where \mathbf{B}^p is the potential field whose vertical component matches the observed B_z at the photosphere. Thus variables based on the horizontal gradients of the magnetic field distribution will directly relate to the "compactness," or third measure in the McIntosh class, while those relating to the force-free parameter and the current density describe how far the active region deviates from a potential-field

configuration, and have no direct analogy in the McIntosh class.

[12] The goal of this analysis of vector magnetic field data is to quantitatively and objectively produce a set of parameters that will be useful in describing the flare-readiness of an active region on the Sun. For an algorithm such as this to be useful in real-world applications, it is imperative that no human subjectivity or intervention be required for evaluating, identifying, or classifying the data. Thus the spatially distributed physical variables outlined above must be reduced to single-number parameters. Following our previous work, a spatially distributed variable x is parameterized by its first four moments: mean, \bar{x} , standard deviation, $\sigma(x)$, skew, $\varsigma(x)$, and kurtosis, $\kappa(x)$ [Leka and Barnes, 2003b, 2003a, 2007], given by:

$$\text{mean } \bar{x} = \frac{\sum_i w_i x_i}{\sum_i w_i} \quad (1)$$

$$\text{standard deviation } \sigma = \left[\frac{\sum_i w_i (x_i - \bar{x})^2}{\sum_i w_i} \right]^{1/2} \quad (2)$$

$$\text{skew } \varsigma = \frac{\sum_i w_i (x_i - \bar{x})^3 / \sigma^3}{\sum_i w_i} \quad (3)$$

$$\text{kurtosis } \kappa = \frac{\sum_i w_i (x_i - \bar{x})^4 / \sigma^4}{\sum_i w_i} - 3 \quad (4)$$

where w_i is a weighting factor that, unless otherwise specified, is set to the area of the pixel. This is necessary because the calculations are carried out on a regular grid in image coordinates, so the physical area of the pixels varies across the field of view.

[13] The mean and standard deviation are familiar to most readers, giving the typical value of the distribution, and the spread about that typical value. The skew describes the asymmetry of the distribution, indicating the presence of a one-sided tail. The kurtosis is normalized to zero for a Gaussian distribution, and deviations from zero indicate whether the distribution has long or short tails in comparison to a Gaussian. The skew and kurtosis are sensitive to small patches of extreme values. Thus for example, a highly twisted δ -class sunspot within an otherwise potential region should appear as a significant non-zero skew in the distribution of the force-free parameter α . In some cases, the moments are supplemented by the total and/or net value of the variable. An example is the total unsigned flux $\Phi_{\text{tot}} = \sum |B_z| dA$, which affords a direct measure of the active region's size, and reflects the first and second components of the McIntosh class. When considering the distribution of the magnetic

shear angle, we also consider the total area of strong shear, and the length of strongly sheared neutral lines. The latter differs slightly from the implementation as originally proposed by *Hagyard et al.* [1990] and incorporated into more recent studies [e.g., *Falconer et al.*, 2006], as we do not subjectively select the "primary" magnetic neutral line but rather automatically evaluate the magnetic shear along all neutral lines which also have strong horizontal magnetic fields.

3. Discriminant Analysis and Probabilistic Forecasts

[14] Discriminant analysis is a statistical tool for classifying a measurement into one of several predetermined, exclusive groups or populations based on the values of m parameters [e.g., *Kendall et al.*, 1983; *Anderson*, 1984]. For the case of solar flare forecasting, the measured parameters characterize an individual active region's magnetic field as alluded to above; the groups consist of those regions which produced one or more flares in the 24 hours subsequent to the observation, versus those regions which produced no such flares. The discriminant function is constructed to maximize the rate of correct predictions. If the probability density function for each of the populations is known, then a given measurement is assigned to the group with the highest probability at that point. Thus the location of the classification boundary is given by equal probabilities for flaring or being flare-quiet, corresponding to a 50% flare probability forecast.

[15] In order to compare the performance of the discriminant analysis flare prediction method to other flare warning systems, flare probability forecasts must also be produced. Using Bayes's theorem, the probability that a region belongs to the flaring population when it is observed to have properties \mathbf{x} is

$$P_f(\mathbf{x}) = \frac{q_f f_f(\mathbf{x})}{q_f f_f(\mathbf{x}) + q_q f_q(\mathbf{x})}. \quad (5)$$

In this expression, $f_j(\mathbf{x})$ is the probability density function for population j , meaning that the probability of a measurement belonging to population j falling between \mathbf{x}_a and \mathbf{x}_b is given by

$$P(\mathbf{x}_a < \mathbf{x} < \mathbf{x}_b) = \int_{\mathbf{x}_a}^{\mathbf{x}_b} f_j(\mathbf{x}) d\mathbf{x} \quad (6)$$

where $j = f$ refers to the flaring population, while $j = q$ refers to the flare-quiet population, while q_j is the prior probability of belonging to population j , which is the probability that any region, whose properties are unknown, will belong to population j .

[16] This expression is valid for any well behaved probability density function f . An application of equation (5) is shown in Figure 1 for two populations with Gaussian probability density functions. Typically, the true popula-

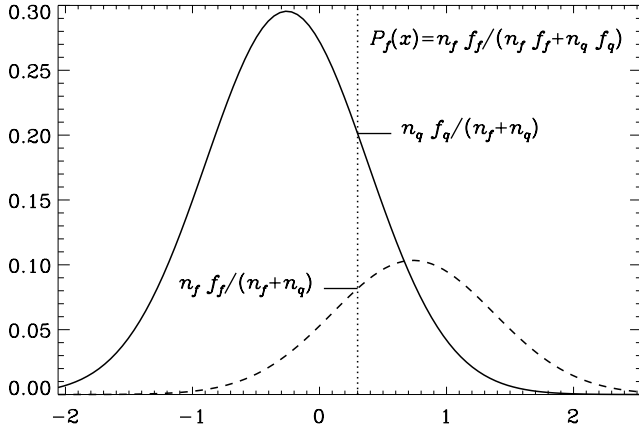


Figure 1. Illustration of how a probability forecast is determined for two populations with Gaussian distributions and an unequal a priori probability of membership. The 50% probability forecast, equivalent to the discriminant boundary, occurs where the two curves intersect, but in general, the probability forecast is determined by the relative probability of belonging to each population. For the example shown, the probability that a measurement at $x = 0.3$ belongs to the flare-quiet population is 0.201, and to the flaring population is 0.081, so the flaring probability is $P_f(x = 0.3) = 0.081/(0.081 + 0.201) = 29\%$.

tion probability density will not be known, and so must be estimated from the samples. For the results presented here, we shall assume the probability densities are Gaussian distributions, and that the flaring and flare-quiet populations have equal covariance matrices. Figure 2 shows an example for two variables discussed by *Leka and Barnes [2007]*, under the above assumptions. In the work of *Leka and Barnes [2007]*, the effect of relaxing these assumptions was investigated, and it was shown that even using nonparametric estimates for the probability density function generally results in relatively small changes to the discriminant function and corresponding error rate. Thus, the results seem to be fairly robust to deviations from a Gaussian. In any case, making a better estimate of the probability density function is only likely to improve the performance of the discriminant.

[17] The multivariate Gaussian probability distribution of m -dimensions is given by

$$f_j(\mathbf{x}) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{m/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu^{(j)})' \Sigma^{-1} (\mathbf{x} - \mu^{(j)}) \right] \quad (7)$$

where $\mu^{(j)}$ is the vector of mean parameter values, Σ is the population covariance matrix (assumed equal for the two populations), and \mathbf{x} is the vector of parameter values for the new active region to be classified. The population mean is estimated from the sample mean, $\mu^{(j)} =$

$\bar{\mathbf{x}}^{(j)} = \sum_{i=1}^{n_j} \mathbf{x}_i^{(j)} / n_j$, and the population covariance matrix is estimated from the sample covariance matrices

$$\mathbf{C} = \frac{(n_f - 1)\mathbf{C}^{(f)} + (n_q - 1)\mathbf{C}^{(q)}}{n_f + n_q - 2}, \quad (8)$$

where

$$\mathbf{C}^{(j)} = \frac{1}{n_j - 1} \sum_{k=1}^m (\mathbf{x}_k^{(j)} - \bar{\mathbf{x}}^{(j)}) (\mathbf{x}_k^{(j)} - \bar{\mathbf{x}}^{(j)})'. \quad (9)$$

Assuming that the a priori probability of membership in a population is proportional to the sample size gives $q_j = n_j / (n_f + n_q)$. Substituting these expressions into equation (5) gives the forecast probability at any point in parameter space.

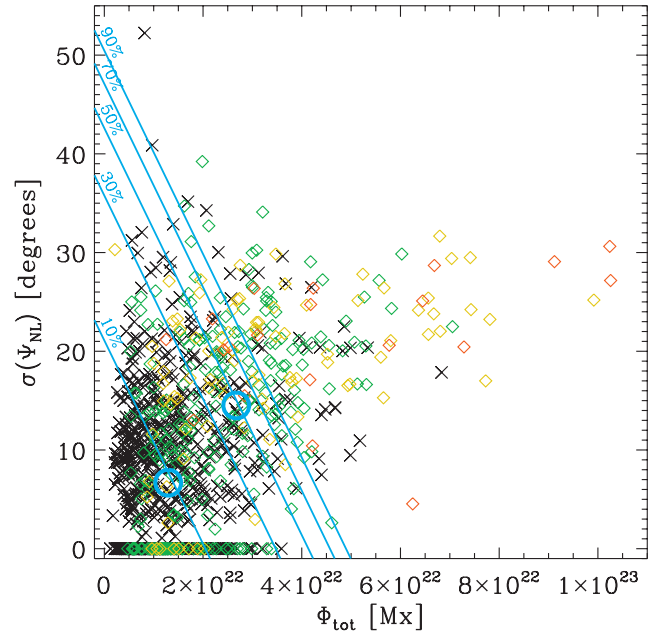


Figure 2. An example of probability forecasts for two variables characterizing the photospheric magnetic field: the total unsigned flux, and the standard deviation of the neutral line shear angle. Flare quiet regions are shown as crosses; flaring regions are shown as diamonds, with the color determined by the largest flare: green for C-flares, yellow for M-flares, and red for X-flares. The mean of each sample is shown as a blue circle, and selected contours of probability forecasts for C-flares are shown as blue lines and annotated with the probability of flaring, under the assumptions described in the text. Many of the regions fall close to the 10% forecast probability contour, which is in accord with the median flare forecast of $f = 0.14$ (see Table 1). The points with $\Phi_{\text{tot}} \sim 10^{23}$ Mx are region NOAA AR 10486.

[18] A probability forecast is the focus here for reasons of comparing our results with other methods, as described in detail below. However, this formulation provides a degree of flexibility which might be useful in future applications, in the following sense. The monetary costs of inaccurate forecasts can impact different industries in quite different ways. The precautions taken by users of space-based systems in light of impact from a solar flare can be costly; system shutdowns and data interruptions may or may not be balanced by the costs of extra shielding and redundancy. An accounting for a customer's unique inequality in the costs of a "false-alarm" instead of a "miss" can be incorporated by way of including a ratio of these two in equation (5): essentially, the cost enters like a prior probability. More specific discussions on the application of this flexibility are beyond the scope of this paper as the forecast must be tailored to specific customer requirements.

3.1. Error Rates and Verification Statistics

[19] To evaluate the performance of the probabilistic forecast, the same verification statistics presented by *Wheatland* [2005] are employed. These include considering the average forecast probability for all observed active regions, $\langle f \rangle$, and the average of all observations, $\langle x \rangle$ (where x is either one or zero depending on whether the region flared or was flare-quiet). Other statistics include the average forecast probability over the flaring regions, $\langle f|x=1 \rangle$, and over the flare-quiet regions, $\langle f|x=0 \rangle$. These averages are supplemented by the mean absolute error,

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

and the mean square error,

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

Finally, the climatological skill score [e.g., *Murphy and Epstein*, 1989], defined by,

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

is calculated. The skill score indicates the improvement of the forecasts over a constant forecast given by the average observed rate, $\langle x \rangle$. Positive scores indicate better performance, with a maximum score of 1.0 for perfect forecasting, while negative scores indicate worse performance.

4. Results and Comparison of Forecast Accuracy to Other Methods

[20] In order to gain insight into what properties of the photospheric magnetic field are most closely related to

flaring, several approaches were considered in the work of *Leka and Barnes* [2003a, 2007] to determine first the minimum number of variables needed to achieve the majority of the discriminatory power, and second, given that number, which variable combinations offered the best insight. The emphasis in those studies was on determining the physical basis for the differences between active regions in flare productivity. In the present investigation, the focus is on empirical flare prediction. Thus throughout we simply use all 74 variables considered by *Leka and Barnes* [2007]. This approach is likely to produce the best forecasts, even though many of the variables included may contribute almost nothing to its performance.

[21] In comparing a Bayesian approach to solar flare prediction with the method employed by the SEC, *Wheatland* [2005] made use of forecast reliability plots, as presented on the SEC's Web page (www.sec.noaa.gov/forecast_verification/mFlare.html), and also verification statistics as described in section 3.1. It is important to note that our approach differs from both the other methods in that it makes forecasts for individual active regions rather than for the whole Sun; additionally, our database only covers a subset of days in the years 2001–2004 as compared to the much longer periods covered by the data used by both the SEC and *Wheatland* [2005]; finally, our results are for events greater than or equal to a given peak emission, whereas the other methods consider events within a range of peak emission. These differences result in a very different mean flaring rate for a given peak emission, nonetheless, we present similar forecast reliability plots in Figure 3 for the C-flare threshold, Figure 4 for the M-flare threshold, and Figure 5 for the X-flare threshold. The results for the last of these categories can be compared most directly to the other methods, but it is also the one with by far the smallest sample size, and thus the least statistically significant.

[22] The approach here can also be used to forecast for several categories of flares simultaneously if multiple populations are considered. For single variables at a time, we have considered the case of three populations, consisting of nonevents, M-X flares and X-flares, to match the forecasts of *Wheatland* [2005]. Unlike the two population case discussed by *Leka and Barnes* [2007], the assumption of Gaussian distributions with equal covariance matrices leads to significantly different results than using a nonparametric density estimator. Since the nonparametric approach requires even larger sample sizes than the parametric approach, it can only reasonably be done with a single variable at a time for our data set, and even this is subject to issues of small sample sizes, since at least the X-flare sample may be considered small. We have therefore chosen not to include potentially misleading results using more than two populations here.

[23] A reliability plot is constructed by first dividing the forecasts into probability bins. For the smaller data set available here, the bin width is set to 0.10; for the results of the SEC and *Wheatland* [2005] methods, the bin width is

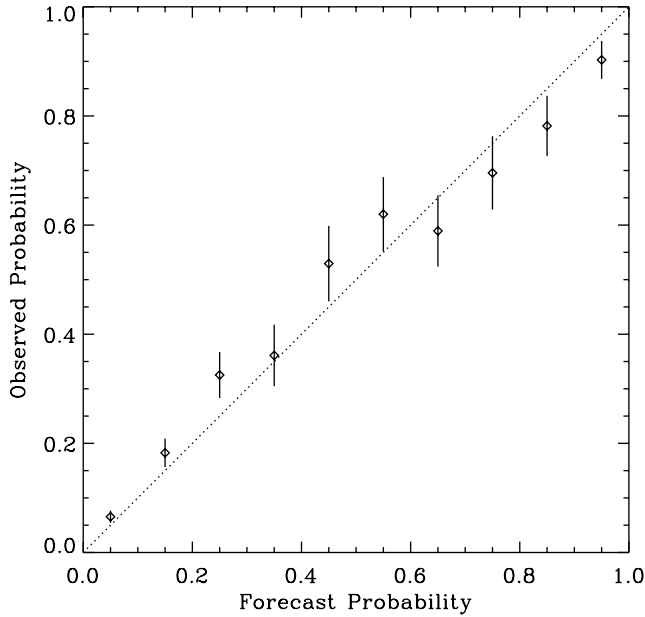


Figure 3. Reliability plot for the vector magnetic field probabilistic forecasts with a threshold of C1.0, in the same format as the SEC. For a perfect forecast, all points lie along the line; points lying above the line indicate an underprediction, points lying below the line indicate an overprediction. Error bars reflect the number of data points in each bin. There is a tendency to underpredict for small forecast probabilities.

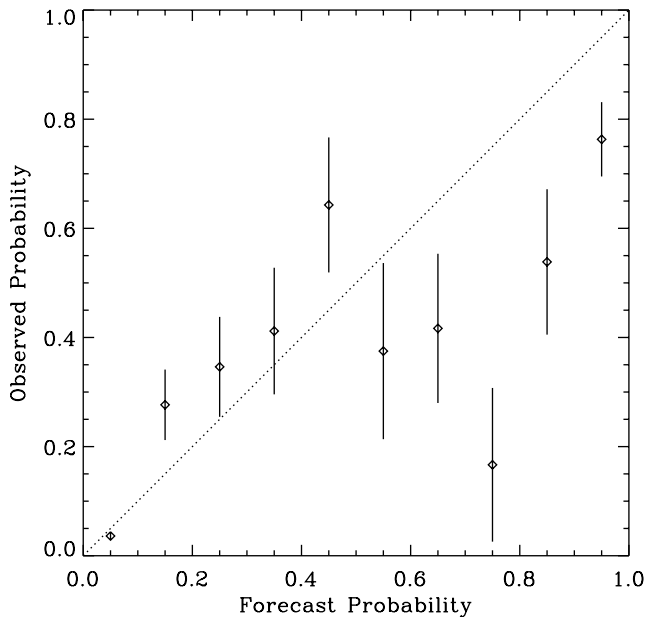


Figure 4. Reliability plot for vector magnetic field forecasts with an M1.0 threshold, in the same format as Figure 3. There is a tendency to overpredict for large forecast probabilities, but most of the points are in the smallest probability bin.

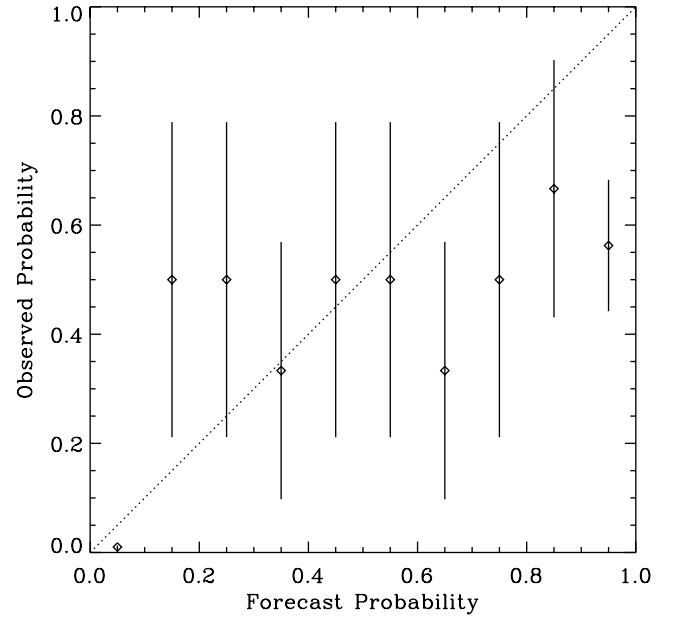


Figure 5. Reliability plot for vector magnetic field forecasts with an X1.0 threshold, in the same format as Figure 3. The vast majority of the points are in the smallest probability bin, making it difficult to determine if there is any tendency to either overprediction or underprediction.

0.05, so only a subjective comparison can be made; given the other differences listed above, this is not a great limitation. For a bin containing S total forecasts, of which R were observed to have at least one event, then the observed probability is $p = (R + 1)/(S + 2)$, with an associated uncertainty $\delta p = [p(1 - p)/(S + 3)]^{1/2}$ [Wheatland, 2005]. The reliability plot shows the observed probability as a function of the forecast probability bin. With this definition, a perfect forecasting scheme would result in the diagonal line shown in Figure 3. For the C-flare threshold, the forecasts perform quite well qualitatively, with a slight tendency to underpredict (points lying above the line) for most probability bins except for the largest probability bins, where there is some overpredicting (points lying below the line). For larger flares (M-flares and larger: Figure 4, X-flares and larger: Figure 5), most of the points fall in the smallest probability bin, while all the other bins have large uncertainties due to the small number of points. Thus it is difficult to determine if there is any tendency to under/overpredict, but the trend appears similar to C-flares, with underprediction at low forecast probabilities turning to overprediction at large forecast probabilities. In comparison, for both the SEC and the Wheatland [2005] approaches, there is a tendency for overpredicting at most probabilities.

[24] For a more objective measure of the performance of the discriminant analysis forecasts, consider the verification statistics given in Tables 1, 2, and 3, and the compar-

Table 1. Verification Statistics (C1.0 Threshold)

Statistic	Value
Total AR-days	1212
Total AR-events	359
$\langle f \rangle$	0.279
$\langle x \rangle$	0.296
Median f	0.140
σ_f	0.291
σ_x	0.457
$\langle f x=1 \rangle$	0.544
$\langle f x=0 \rangle$	0.167
SD $f x=1$	0.313
SD $f x=0$	0.193
MAE(f, x) = $\langle f - x \rangle$	0.253
MSE(f, x) = $\langle (f - x)^2 \rangle$	0.136
SS(f, x)	0.346

Table 3. Verification Statistics (X1.0 Threshold)

Statistic	Value
Total AR-days	1212
Total AR-events	20
$\langle f \rangle$	0.013
$\langle x \rangle$	0.017
Median f	1.1×10^{-8}
σ_f	0.111
σ_x	0.127
$\langle f x=1 \rangle$	0.450
$\langle f x=0 \rangle$	0.006
SD $f x=1$	0.500
SD $f x=0$	0.073
MAE(f, x) = $\langle f - x \rangle$	0.015
MSE(f, x) = $\langle (f - x)^2 \rangle$	0.014
SS(f, x)	0.123

ison of a selection of the statistics to the other forecasting methods given in Table 4. Before comparing the results of the different methods, note that in the present case, the flare event rate is much lower than for the other data. For example, for M-flares, the present data has $\langle x \rangle = 0.092$ compared with $\langle x \rangle = 0.262$ for the SEC's tabulated data. This is likely a result of considering individual active regions rather than the whole sun. In terms of flare event rate, the C-flare data with $\langle x \rangle = 0.296$ in the present case are similar to the M-flare data in the other data, hence we include both in our comparisons.

[25] To understand how the various methods behave, consider first the average forecast probabilities on event and nonevent days. For both the SEC and *Wheatland* [2005] methods applied to M-X events, $\langle f|x=1 \rangle > 0.5$, while the present method for the C-flare threshold has a comparable score, meaning it provided a similar average forecast for flare-event data (see Table 4). However, the present method for M-flare threshold has $\langle f|x=1 \rangle = 0.422$, indicating that it typically produced lower forecast probabilities for flare-event data. For both the SEC and *Wheatland* [2005] methods, $\langle f|x=0 \rangle > 0.2$, while the

present method for both thresholds is lower, indicating that it typically produces lower forecasts for data for which no event occurs. The better low probability forecasts combined with the worse high probability forecasts in the present method for M-flares compared to the SEC and *Wheatland* [2005] methods for M-X flares lead to extremely similar skill scores ($0.252 \leq SS(f, x) \leq 0.262$); the comparable high probability forecasts with the C-flare threshold combined with the improved low probability forecasts gives a significantly improved skill score for the present method ($SS(f, x) = 0.346$). It should be noted that all the methods perform better for lower flare thresholds, where the event rate is higher, but we include the C-flare threshold for our results because its event rate is comparable to the event rate for M-X events in the other methods.

[26] For the X-flare threshold, the present method typically had much higher forecast probabilities for the flare-event data, $\langle f|x=1 \rangle = 0.450$ compared to the other approaches, which had $\langle f|x=1 \rangle < 0.250$. In addition, the present method had lower forecast probabilities for the flare-quiet data, resulting in a larger skill score than either of the other approaches ($SS(f, x) = 0.123$ for the

Table 2. Verification Statistics (M1.0 Threshold)

Statistic	Value
Total AR-days	1212
Total AR-events	111
$\langle f \rangle$	0.076
$\langle x \rangle$	0.092
Median f	0.006
σ_f	0.206
σ_x	0.289
$\langle f x=1 \rangle$	0.422
$\langle f x=0 \rangle$	0.041
SD $f x=1$	0.391
SD $f x=0$	0.134
MAE(f, x) = $\langle f - x \rangle$	0.090
MSE(f, x) = $\langle (f - x)^2 \rangle$	0.062
SS(f, x)	0.252

Table 4. Comparison of Verification Statistics for M Versus M-X Flares

	Present Method		Wheatland		SEC	
	M	X	M-X	X	M-X	X
$\langle f \rangle$	0.076	0.013	0.294	0.040	0.298	0.064
$\langle x \rangle$	0.092	0.017	0.262	0.035	0.262	0.035
$\langle f x=1 \rangle$	0.422	0.450	0.510	0.122	0.551	0.244
$\langle f x=0 \rangle$	0.041	0.006	0.217	0.037	0.208	0.057
MAE(f, x)	0.090	0.015	0.289	0.066	0.271	0.081
MSE(f, x)	0.062	0.014	0.143	0.031	0.139	0.032
SS(f, x)	0.252	0.123	0.258	0.078	0.262	-0.006

present method compared with $SS(f, x) < 0.08$ for the other approaches).

[27] When comparing the verification statistics, it is important to note that there is a great deal of variability in forecast success. For example, over the course of approximately one solar cycle (1995–2006), the annual SEC skill scores for M-X flares ranged from a low of -0.157 to a high of 0.322 , while for X flares the lowest annual skill score was -0.567 while the highest was 0.242 . Thus, even though the present method has a skill score of 0.123 for forecasting X-flares, it is not at all clear that this represents a real improvement over the skill scores of 0.078 for *Wheatland* [2005] or -0.006 for the SEC over the time interval 1987–2003.

5. Discussion

[28] A probabilistic forecasting method has been demonstrated for a variety of parameters characterizing the photospheric vector magnetic field. Although it has been shown using discriminant analysis that the majority of predictive power can be contained in just a few variables [Leka and Barnes, 2007], the approach here focuses on extracting the greatest success rate rather than understanding the underlying physics. Thus probability forecasts were constructed with all of the photospheric variables considered in our on-going series. Three cases are considered here, with the threshold of whether to consider a data point as “flaring” or “flare-quiet” set at either the GOES C1.0, M1.0, or X1.0 levels.

[29] We employ the simple assumption that the probability density function of each variable is Gaussian, and the parameters characterizing the Gaussian are determined from the samples. The flaring probability for any value of the variables is determined based on the estimated probability of belonging to the flaring population. It was shown by Leka and Barnes [2007] that even using nonparametric techniques to estimate the probability density does not greatly improve the performance of the discriminant function, so the results of making the assumption of Gaussian distributions are likely to be quite robust when considering only two populations.

[30] The all-variable probability forecast results presented here were compared in detail to two other forecasting methods, one based on event statistics and flare persistence [Wheatland, 2004, 2005] and the other the method used for the SEC’s flare forecasts, based essentially on the flare rates according to the classification of white light images [McIntosh, 1990]. With appropriate caveats due to the differences between the methods in what is forecast (active region versus whole disk, flare threshold levels) the performance of the present method is comparable.

[31] In particular, for the M-flare threshold, the skill score for the present method is 0.252 , comparable to the performances of the SEC and *Wheatland* [2005] methods for M-X flares (see Table 4). For X-flares, none of the methods perform particularly well, the present method’s

skill score of 0.123 is higher than either of the other approaches, although this is likely to be a statistically insignificant result. The positive values of the skill scores indicate that all the approaches indeed add information above a uniform forecast probability. The present method still tends to overpredict for at least the large forecast probability bins, similar to the performance of the other methods (see Figures 3 and 4).

[32] There are ways in which incremental improvements can be made in the present method: in particular, larger sample sizes would help in several ways. First, a larger sample would better constrain the performance of the analysis. Second, with a larger sample, it is possible to relax the assumption of Gaussian distributions and apply nonparametric approaches for small numbers of variables considered simultaneously; the sample size needed for a nonparametric representation grows very rapidly with the number of variables considered [see Table 4.2 of Silverman, 1986] so it will be extremely difficult to consider large numbers of variables simultaneously. The present analysis also includes only the declining phase of one solar cycle, and it may well be the case that there are variations over the solar cycle that have not been captured. Furthermore, we have been consistent in considering only the parameters described in our previous studies; it may be that including other characteristics of the photospheric magnetic field, such as its fractal dimension or the power spectrum of spatial scales present [Abramenko, 2005; McAtter et al., 2005], or parameters characterizing the coronal magnetic topology inferred from the photospheric field [Barnes and Leka, 2006] may significantly improve the results. Finally, including a measure of flare persistence, such as forms the basis for the *Wheatland* [2005] method, in combination with the magnetic field characterization may bring together the most pertinent information for empirical flare forecasting [Wheatland, 2007]. Nonetheless, it is clear that an automated forecasting approach based on photospheric vector magnetic field measurements can perform comparably to existing flare forecast approaches.

[33] **Acknowledgments.** The data used for this work are photospheric vector magnetograms obtained at the University of Hawaii Mees Solar Observatory, and the authors would like to thank Don Mickey and the staff of the Mees Solar Observatory for their work in creating the archive of IVM data. This work was conducted at the Colorado Research Associates Division of NorthWest Research Associates, Inc. and at the Air Force Institute of Technology with funding provided by the Air Force Office of Scientific Research under contracts F49620-00-C-0004 and F49620-03-C-0019.

References

- Abramenko, V. I. (2005), Relationship between magnetic power spectrum and flare productivity in solar active regions, *Astrophys. J.*, **629**, 1141–1149, doi:10.1086/431732.
- Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York.

- Barnes, G., and K. D. Leka (2006), Photospheric magnetic field properties of flaring versus flare-quiet active regions. III. Magnetic charge topology models, *Astrophys. J.*, 646, 1303–1318, doi:10.1086/504960.
- Bornmann, P. L., and D. Shaw (1994), Flare rates and the McIntosh active-region classifications, *Sol. Phys.*, 150, 127–146.
- Canfield, R. C., et al. (1993), The morphology of flare phenomena, magnetic fields, and electric currents in active regions I. Introduction and methods, *Astrophys. J.*, 411, 362–369, doi:10.1086/172836.
- Falconer, D. A., R. L. Moore, and G. A. Gary (2006), Magnetic causes of solar coronal mass ejections: Dominance of the free magnetic energy over the magnetic twist alone, *Astrophys. J.*, 644, 1258–1272, doi:10.1086/503699.
- Gallagher, P., Y.-J. Moon, and H. Wang (2002), Active-region monitoring and flare forecasting, *Sol. Phys.*, 209, 171–183, doi:10.1023/A:1020950221179.
- Hagyard, M. J., P. Venkatakrishnan, and J. B. J. Smith (1990), Nonpotential magnetic fields at sites of gamma-ray flares, *Astrophys. J. Suppl. Ser.*, 73, 159–163, doi:10.1086/191447.
- Kendall, M., A. Stuart, and J. K. Ord (1983), *The Advanced Theory of Statistics*, vol. 3, 4th ed., Macmillan, New York.
- LaBonte, B., D. L. Mickey, and K. D. Leka (1999), The imaging vector magnetograph at Haleakalā II: Reconstruction of stokes spectra, *Sol. Phys.*, 189, 1–24.
- Leka, K. D., and G. Barnes (2003a), Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis, *Astrophys. J.*, 595, 1296–1306, doi:10.1086/377512.
- Leka, K. D., and G. Barnes (2003b), Photospheric magnetic field properties of flaring versus flare-quiet active regions. Data, I, general approach, and sample results, *Astrophys. J.*, 595, 1277–1295, doi:10.1086/377511.
- Leka, K. D., and G. Barnes (2007), Photospheric magnetic field properties of flaring versus flare-quiet active regions. IV. A statistically significant sample, *Astrophys. J.*, 656, 1173–1186, doi:10.1086/510282.
- McAteer, R. T. J., P. T. Gallagher, and J. Ireland (2005), Statistics of active region complexity: A large-scale fractal dimension survey, *Astrophys. J.*, 631, 628–635, doi:10.1086/432412.
- McIntosh, P. S. (1990), The classification of sunspot groups, *Sol. Phys.*, 125, 251–267.
- Metcalf, T. R., et al. (2006), An overview of existing algorithms for resolving the 180° ambiguity in vector magnetic fields: Quantitative tests with synthetic data, *Sol. Phys.*, 237, 267–296, doi:10.1007/s11207-006-0170-x.
- Mickey, D. L., R. Canfield, B. J. LaBonte, K. D. Leka, M. F. Waterson, and H. M. Weber (1996), The imaging vector magnetograph at Haleakalā, *Sol. Phys.*, 168, 229–250.
- Moon, Y.-J., G. S. Choe, H. S. Yun, and Y. D. Park (2001), Flaring time interval distribution and spatial correlation of major X-ray solar flares, *J. Geophys. Res.*, 106, 29,951–29,962.
- Murphy, A. H., and E. S. Epstein (1989), Skill scores and correlation coefficients in model verification, *Mon. Weather Rev.*, 117, 572–582, doi:10.1175/1520-0493(1989)117.
- Priest, E. R., and T. G. Forbes (2002), The magnetic nature of solar flares, *Astron. Astrophys. Rev.*, 10, 313–377, doi:10.1007/s001590100013.
- Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.
- Wheatland, M. S. (2004), A Bayesian approach to solar flare prediction, *Astrophys. J.*, 609, 1134–1139, doi:10.1086/421261.
- Wheatland, M. S. (2005), A statistical solar flare forecast method, *Space Weather*, 3, S07003, doi:10.1029/2004SW000131.
- Wheatland, M. S. (2007), Reconstruction of nonlinear force-free fields and solar flare prediction, in *Advances in Geosciences*, World Sci., Singapore, in press.

G. Barnes and K. D. Leka, Colorado Research Associates Division, NorthWest Research Associates, Inc., 3380 Mitchell Lane, Boulder, CO 80301, USA. (graham@cora.nwra.com; leka@cora.nwra.com)

D. J. Della-Rose, Department of Physics, United States Air Force Academy, Colorado Springs, CO 80840, USA. (devin.della-rose@usafa.af.mil)

E. A. Schumer, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433, USA. (evelyn.schumer@afit.edu)