

DIAGNOSIS OF SOLAR FLARE PROBABILITY FROM CHROMOSPHERE IMAGE SEQUENCES

Donald C. Norquist and K. S. Balasubramaniam

30 December 2011

Interim Report

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.



**AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
Kirtland AFB, MA 87117-5776**

DTIC COPY

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 377 ABW Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RV-PS-TP-2011-0005 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//

Donald Norquist, RVBXS
Program Manager

//signed//

Joel Mozer, RVB
Division Chief

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 30-12-2011		2. REPORT TYPE Interim Report		3. DATES COVERED (From - To) 10 Jun 2009 to 31 Aug 2011	
4. TITLE AND SUBTITLE Diagnosis of Solar Flare Probability from Chromosphere Image Sequences				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62601F	
6. AUTHOR(S) Donald C. Norquist and K. S. Balasubramaniam				5d. PROJECT NUMBER 1010	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 837055	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Research Laboratory Space Vehicles Directorate 3550 Aberdeen Ave SE Kirtland AFB, NM 87117-5776				8. PERFORMING ORGANIZATION REPORT NUMBER AFRL-RV-PS-TP-2011-0005	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RVBXS	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited. (377 ABW-2011-1566, dtd 7 Nov 11)					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT We used optical observations of the solar chromosphere in the diagnosis of flare probability at the observation time. We derived eigenvectors from sequences of hydrogen-alpha (H α) images of sub-regions of selected solar active regions at one-minute intervals. A degree of flaring category was set from H α intensity and 1-8 Å x-ray flux. Leading eigenvector elements were the predictors, and flaring category the predictand, in employing multivariate discriminant analysis (MVDA) on a training set of image sequences. We applied resulting discriminant vectors to image sequence eigenvector elements to diagnose flaring category probability. Metrics of comparison with the specified flaring category determined diagnosis skill. Binary no-flare/flare predictand categories set with a specified H α rise produced too many flaring diagnoses. Setting multiple flaring categories from H α magnitude produced a large positive category bias. Assigning flaring categories according to an H α eigenvector pattern - x-ray flare intensity association improved the category diagnosis performance. The latter method will be further refined and applied to additional ISOON image sequences.					
15. SUBJECT TERMS Solar flares, optical images, x-ray flux, flare probability, principle component analysis, multivariate discriminant analysis					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT unlimited	18. NUMBER OF PAGES 54	19a. NAME OF RESPONSIBLE PERSON Donald C. Norquist
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED			19b. TELEPHONE NUMBER (include area code)

This page is intentionally left blank.

Table of Contents

1. INTRODUCTION	1
2. DATA	3
3. METHODS	3
4. RESULTS	16
4.1. H α Magnitude Flare Categorization Technique.....	19
4.2. H α Eigenvector Flare Categorization Technique	22
5. SUMMARY AND CONCLUSIONS.....	32
REFERENCES	38
APPENDICES	39

List of Figures

1. ISOON H α Image Sequence Eigenvalues for 11 June 2003, Active Region 10375	5
2. First Nine Eigenvectors of the 11 June 2003 ISOON H α Image Sequence.....	5
3. H α Intensity for 9 May 2003 ISOON Image Sequence, Unknown Active Region.....	8
4. GOES 1-8 Å X-ray Flux for the One-Minute Image Times Shown in Figure 3.....	8
5. First Nine Eigenvectors of the 9 May 2003 ISOON H α Image Sequence	9
6. H α /H α (Bkgd) vs. log {[Xr]/Xr(thrs)} for Seven Sequences of Correlation > 0.7.....	11
7. Eigenvectors from Examples of FLI (a) 0, (b) 1, (c) 2, and (d) 3 Sequences.....	14
8. Discriminant Space Plots from the MVDA/HMFC Development Algorithm: (a) Group 0, (b) Group 1, (c) Group 2, and (d) Group 3	17
9. Brier Score and Diagnosis Uncertainty from Application of the MVDA/HMFC Development Algorithm.....	20
10. Bias from Application of the MVDA/HMFC Development Algorithm, and Fraction of Flaring Times.....	21
11. Frequency Distribution from Application of the MVDA/HMFC Algorithm	23
12. Brier Score from the MVDA/HEFC Development Algorithm, and Fraction of Flare Rise Times.....	24
13. Bias from the MVDA/HEFC Development Algorithm	25
14. Diagnosis Uncertainty from the MVDA/HEFC Development Algorithm	26
15. Frequency Distribution from Application of the MVDA/HEFC Algorithm	29
16. 20031029_10486 Image Sequence (a) C _p (Dashed) and C _o (Solid) and (b) p _g from Application of the MVDA/HEFC Algorithm on Case 3	30
17. Same as in Figure 16 Except for the 20030318_10318 Image Sequence.....	31
18. 20030117_10258 Image Sequence (a) H α /H α (bg) and (b) First Nine Eigenvectors.....	33
19. 20030318 Image Sequence GOES One-Minute Average 1-8 Å X-Ray Flux	34

List of Tables

1. X-Ray Flare Class Category Table for $[Xr]/Xr(\text{Thrs})$	10
2. X-Ray Flare Class Category Table for $[H\alpha]/H\alpha(\text{Bkgd})$	11
3. Conversion Table from Flaring Categories to X-Ray Flare Class	12
4. Performance Scores for the MVDA/HEFC Development/Application Process.....	27

Acknowledgements

We express our appreciation to Tim Henry of the National Solar Observatory at Sacramento Peak, NM for his valuable contribution to building and maintaining the ISOON data archive. We obtained GOES x-ray data from the NOAA National Geophysical Data Center. This research was funded by the Applied Research program of the Air Force Research Laboratory.

1. INTRODUCTION

Intense, high-energy electromagnetic emissions and expulsion of charged particles are known to be products of solar flare events. As these events pose a potential threat to satellite navigation and communication systems, manned space operations and high altitude aviation, there is considerable interest in determining whether or not significant solar flares can be anticipated to protect valuable assets. Because of the complex physical nature of solar disturbances and mankind's current inability to comprehensively simulate them, most flare prediction efforts to date have had an observational basis. Norquist [1] briefly reviewed past attempts to use statistical methods to predict the probability of flaring in specific active regions, and summarized them in two groups: those that infer flaring probability from the observed flaring rates in historical catalogs of classified active regions, and those that use current solar observations of the solar magnetic field to deduce properties likely to be associated with flaring to predict flaring probability. A recent example of the former is the work of Contarino et al. [2], in which sunspot group characteristics historically associated with flaring were assessed to evaluate the probability of flaring. Two encouraging examples of the latter type of statistical flare probability algorithm were recently published by Falconer et al. [3] and Reinard [4], who used observations of free magnetic energy and helioseismological oscillations, respectively, to predict the probability of near-term flare occurrence.

In the current study, we investigate the utility of optical observations of the solar chromosphere in the diagnosis of flare probability. With flare probability prognoses as the ultimate goal, we first determine the feasibility of correctly inferring flaring likelihood at the observation time. If feasible, we would next investigate if there is sufficient information in the image sequence to project the flare probability estimate ahead in time. Sequences of hydrogen-alpha ($H\alpha$) images at one-minute intervals from the U. S. Air Force Improved Solar Observing Optical Network (ISOON) telescope at Sacramento Peak, NM (Neidig et al. [5]) were used as the basis for flare probability estimation. We used picture element (pixel) values of $H\alpha$ intensity at one arc second grid spacing for selected sub-regions of solar active regions for the image sequence of 8-10 hours duration (the observable daylight hours) as the optical data for this study. $H\alpha$ image sequences were subjected to principal component analysis (PCA) to derive the eigenvectors and associated eigenvalues. At each image time, the sub-region average $H\alpha$ intensity and the independently obtained whole disk 1-8 Å x-ray flux from the Geostationary Operational Environmental Satellite (GOES) were used to determine a category of degree of flaring. A subset of the leading eigenvector elements at each time served as the predictors, and the flaring category the predictand, in employing multivariate discriminant analysis (MVDA) on several image sequences making up a "training set." The consequent discriminant vectors were then applied to the eigenvector element subsets of both the training set sequences and other independent sequences to determine the probability of each flaring category at the observation time. These were compared with the specified flaring category at each time to determine the skill of the flaring category probability diagnosis.

Norquist [1] found that flares were produced from a relatively small proportion (about seven percent) of sunspot group-days over Solar Cycle 23. Because the "non-flaring" category was so likely to be the correct one, it seemed most effective to carry out a two-phased approach to diagnosing flare probability: diagnose the binary yes/no category flaring probability, then for the

“yes” times, attempt to diagnose the intensity of flaring. Thus, a discriminant analysis for two groups was implemented on the PCA eigenvectors as predictors and yes/no flaring specification predictand categories based on a minimum five percent rise in area-average H α and corresponding x-ray flux of $\geq 1.0 \times 10^{-6} \text{ W m}^{-2}$. We found that much of the time, the discriminant function was unable to correctly distinguish between flaring and non-flaring in the subsequent diagnosis. We concluded that using an arbitrary threshold to separate flaring from non-flaring compromised the ability of the discriminant analysis to distinguish between the two categories.

We next attempted an application of MVDA for multiple groups in which we derived a empirical relationship between (1) the ratio of H α area mean intensity to a background (bkgd) full-disk mean intensity from the previous day and (2) the ratio of the x-ray intensity to a “threshold” level (the next full power of 10 above its prior-day background value) to set the predictand categories. We used four categories: non-flaring (corresponding to x-ray flux below the threshold), weak flaring (within a power of 10 of the x-ray threshold), moderate flaring (between 10 and 100 times the threshold), and strong flaring (at or above 100 times the x-ray threshold). Using the derived relationship, we set the category boundaries in H α /H α (bkgd) and used them to determine the category to which each image time would be assigned based on its H α /H α (bkgd). After implementing the four-group MVDA, the consequent discriminant vectors were applied to the eigenvector elements at each image time for both training set and independent cases. We found a tendency for the diagnosis to be biased positive – that is, the most probable category determined by the algorithm was on average greater than the specified category. This result parallels those found by other investigators (e.g., Reinard [4]) in which an excessive number of false alarms are predicted.

The last attempt documented in this report involved using the characteristics of the temporal variation of the leading eigenvectors for each image sequence as the principal indicator for flaring category classification. We noticed that the time series of the primary eigenvectors appeared to be very similar in their pattern for ISOON sequences of like degrees of flaring. This varied from very smoothly varying sinusoidal shapes for non-flaring sequences to nearly constant magnitude followed by a sharp drop then quick rise above the prior level of constant magnitude at the time of flare onset in strongly flaring sequences. We used a combination of the eigenvector pattern characteristics and the x-ray flare intensity to set the category of each flare apparent in the area-average H α time series. In every case, non-zero flaring categories were assigned only during the times of the flare rise – at all other times, including flare demise, zero (non-flaring) was assigned. Once this four-category scheme was analyzed by MVDA, the application of the resulting discriminant functions to the ISOON sequence eigenvectors revealed a significantly improved ability to discriminate among flaring categories. Also compared to the H α /H α (bkgd) magnitude approach to assigning predictand categories, the eigenvector pattern method measurably reduced the positive bias of the diagnosed flaring category.

The organization of this report is as follows. Section 2 describes the data used in the study in more detail. Section 3 explains the methods used in the various attempts to develop and apply the discriminant vectors to the ISOON image sequence eigenvectors. Section 4 presents the results of comparing the diagnosed flaring category probabilities with the specified category for all image sequences. Section 5 concludes the article with a summary of the study and some expectations for extending the work to include prognostication of flaring probability.

2. DATA

The observational data used in this study to develop and apply the statistical method for diagnosing flare occurrence was the hydrogen-alpha ($H\alpha$) images at one-minute intervals from the ISOON telescope. Our archive of images extends back to late 2002 when the ISOON telescope began taking observations, primarily on weekdays. Each image is a grid of picture elements (pixels) with a spacing of 1.1 arc second covering the entire solar disk visible from Earth. Every image is gain corrected by the telescope software. Solar active regions (sunspot groups) of interest were identified from selected dates available in the ISOON data archive. The selection included dates during which flares did and did not occur at any time during each day's available solar viewing time period. $H\alpha$ intensity was extracted at each image time from a sub-region of pixels for the entire observational time period of each date, varying in duration from 4½ to 10 hours. After extraction, the images were normalized, spatially oriented and aligned. In all, sub-region image sets were prepared for 44 sunspot group-dates for this study, extending from 13 December 2002 to 6 December 2006. For one of the sunspot group-days, we extracted $H\alpha$ intensity for three separate sub-regions in a single active region, resulting in a total of 46 ISOON $H\alpha$ image sequences considered in the study.

Due to environmental seeing condition limitations and occasional ISOON operation issues, short data gaps (usually less than one hour in duration) sometimes occurred in the time series for a particular date. In this study, we ignored the data gaps and used the data whenever available. Timestamps of the image measurements were used only to extract x-ray flux data at the same times, and to plot each observed time series and resulting products in clock time for display purposes.

X-ray flux measurements in the 1-8 Å wavelength range from the soft x-ray sensor aboard the GOES satellite operational during the range of dates of extracted $H\alpha$ data were acquired from the National Geophysical Data Center (NGDC) at one-minute intervals (see <http://goes.ngdc.noaa.gov/data/avg>). The x-ray data were extracted from the NGDC archive for each entire date of interest, and for the entire prior day as well. We used the ISOON timestamps of available one-minute interval $H\alpha$ data to extract x-ray flux at the same times. This resulted in an exact match of $H\alpha$ intensity and x-ray flux one-minute interval values for each of the 46 sunspot group-date image sequences. The prior-day x-ray flux record was used to derive a background x-ray flux level that was used to establish a non-flaring/flaring threshold as described in the next section.

Lastly, full-disk mean intensity (FDMI) $H\alpha$ time series for the available observing times of the day prior (when available) to each date of interest were prepared from the respective $H\alpha$ image sequence in the ISOON data archives. As in the case of the prior-day x-ray flux data, we used these FDMI values to establish a background level of $H\alpha$ intensity with which to judge the degree of flaring as detailed in the following section.

3. METHODS

The method of principal component analysis (PCA, e.g., see Wilks [6]) was applied to each of the 46 ISOON $H\alpha$ intensity image sequences. The goal was to use PCA to analyze

spatial/temporal characteristics of the images, endeavoring to associate them with the presence or absence of flares during the sequence time period. Given the K ISOON pixel values of H α intensity in each of n image times, we formed an n row by K column matrix of H α values. Though the pixel values are oriented in a rectangular grid in the image, we simply strung the grid rows together to form a K -dimensional vector constituting each of the rows of the $n \times K$ matrix. Reversing the usual convention of treating the multiple observations at each measurement time as the “variables”, we cast the image times as the variables in the PCA and computed the $n \times n$ covariance matrix $[S]$ made up of the elements $s_{k,l}$ defined by:

$$\text{Cov}(x_k, x_l) = s_{k,l} = \frac{1}{K-1} \sum_{i=1}^K (x_{k,i} - \bar{x}_k)(x_{l,i} - \bar{x}_l); k, l = 1, n \quad [1]$$

where the x are the H α pixel values from the $n \times K$ matrix of image values, k and l are indices of the various combinations of image times, and the overbars represent the spatial average over the pixels for a particular image time. Then the n eigenvalues λ and n -dimensional orthonormal eigenvectors e were derived from the equation $[S] e = \lambda e$ for each ISOON image sequence. The eigenvalues are ranked in decreasing order of influence in characterizing the image sequence. They are shown for the ISOON image sequence of 11 June 2003, active region 10375, in Figure 1. There are n eigenvalues $m = 0, n-1$ in which the few leading (most influential) eigenvalues are several orders of magnitude greater than the lesser ones nearer the end. Figure 2 shows the nine leading eigenvectors for this same case, numbered zero to eight. Note the observation times on the abscissa indicating the duration of the image sequence of the plotted eigenvectors. Note also the gaps are included in the eigenvector plots indicating the times of unavailable ISOON images.

As previously stated, the eigenvectors represent the spatial/temporal variations of the H α images during the course of the ISOON observations. We wish to associate them with the level of flare activity, if any, during the observation period. As such, they serve as “predictors” against which a corresponding “predictand” (in this case, a level of flaring intensity) is assigned. At each observation time, a vector having as its elements the leading eigenvectors (the predictor vector) is associated with an assigned predictand. Once the predictor vector-predictand pair is established for each observation time, they are submitted to a statistical algorithm, such as a regression scheme, to derive the coefficients of relationship(s) for the predictand as a function of the linear combination of eigenvector elements (predictor vector). If the predictands are categorical instead of continuous as in this study, there is always one fewer linear predictand – predictor function than number of categories.

The statistical method used in this project to relate predictands with predictor vectors, and thus to establish functional relationships between them with which diagnose flaring probability, is discriminant analysis. When using this technique, the derived coefficients of the predictand – predictor relationships are called discriminant vectors, and their dot product with the predictor vectors are called discriminant functions. This technique is most effective in distinguishing among the designated predictand categories when the category, or group, mean predictor vectors projected in discriminant space (that is, when dotted with each discriminant vector) have the greatest separation distance among them. In addition, if the scatter of the projection of the individual predictor vectors is small and there is little overlap in discriminant space between the

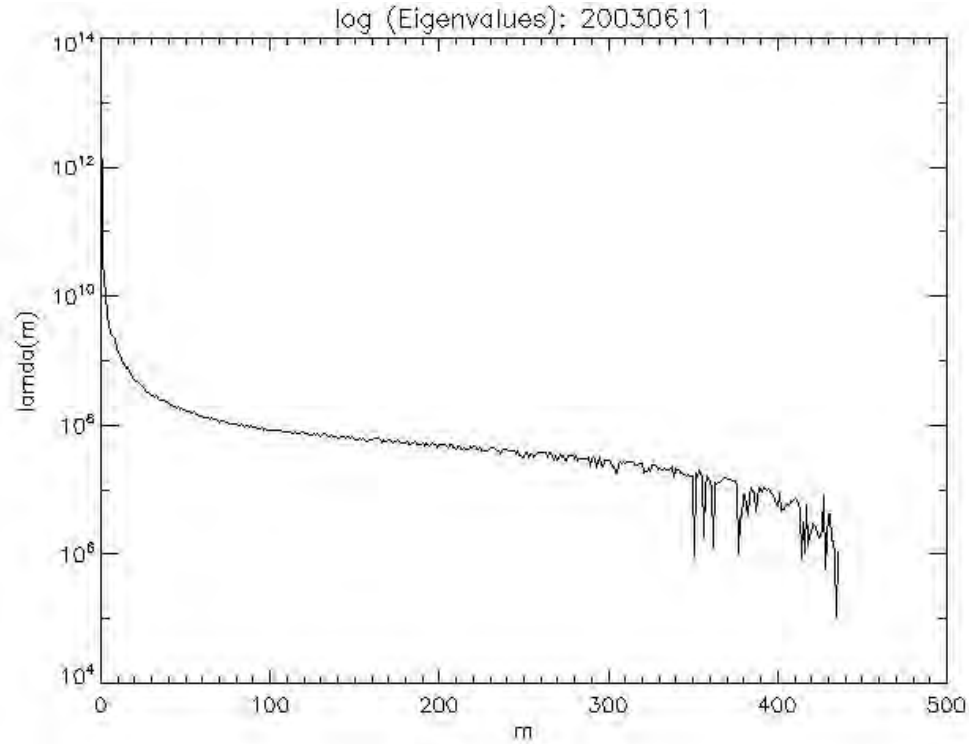


Figure 1. ISOON H α Image Sequence Eigenvalues for 11 June 2003, Active Region 10375

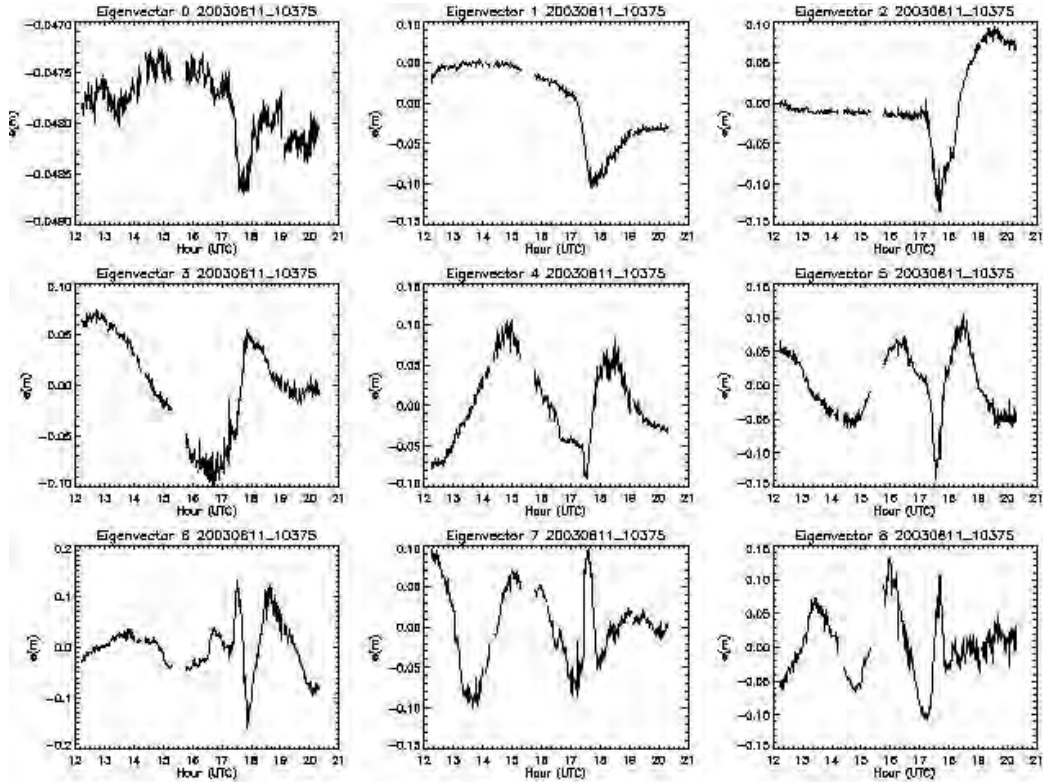


Figure 2. First Nine Eigenvectors of the 11 June 2003 ISOON H α Image Sequence

groups, this also improves the category distinction. When the discriminant vectors are applied to the leading eigenvector elements (predictor vectors) at each observation time from independent measurement cases, the group mean lying closest in discriminant space is indicative of the most likely category to which the observation should be assigned. In fact, the relative distances to the group means can be used to estimate the probabilities of each category being the correct one for a diagnosed observing time. So it can be seen that, if the group means are close together and/or if there is considerable overlap among the discriminant space domains of the groups, there will be considerable ambiguity in the category assignment of the independent observation. In this case, the categorical probabilities will be more similar in magnitude, without a clearly dominant one that would designate the most probable correct category.

As previously stated, the leading eigenvectors derived from the PCA of each ISOON H α image sequence served as the elements of the predictor vectors at each observation time. Solution of the eigenvalue-eigenvector equation results in n eigenvalues and eigenvectors, the latter each having n elements. Because it is unwieldy to carry all eigenvectors into the discriminant analysis, it is necessary to truncate the number of eigenvectors used while ensuring that all of the essential information is retained. To do this, we took advantage of the fact that the eigenvalue magnitudes signify the influence of the eigenvectors in characterizing the H α variations. As shown in Figure 1, their magnitudes quickly declined with increasing eigenvalue number, so we computed the sum of their magnitudes over all n eigenvalues, and then computed the cumulative percentage contribution to the total sum for each successive eigenvalue. We truncated the eigenvectors at the eigenvalue number where this contribution reached 99.9% of the total eigenvalue magnitude. In this way, we can ensure that virtually all of the explained variance is captured in the truncated set of eigenvectors. For the 46 individual cases investigated in this study, the number of included (leading) eigenvectors varied from 28 to 52 extracted from an average $n = 450$. When a “training set” of several image sequences were used collectively to develop the discriminant functions, the largest number of eigenvalues required to reach the 99.9% threshold among the individual sequences was the truncation level used for them all.

Having specified the predictor vectors for the discriminant analysis, it remains to designate the criteria for assigning each observation to a category in order to develop the discriminant vectors. Generally, with discriminant analysis it is beneficial to keep the number of categories to a minimum in order to maximize their distinction. In the case of solar flares as the phenomena of interest, Norquist [1] found that in the vast majority (approximately 93%) of sunspot group-days of Solar Cycle 23, no significant flares (x-ray class C5 or greater) occurred. This suggests that non-flaring is a very likely flaring category, and all levels of actual flaring are contained in the other seven percent of the cases. Therefore, our initial effort to establish predictand categories for flaring probability diagnosis was to use a binary, non-flaring/flaring approach. While we acknowledge that our limited 46 sunspot group-day data set is not representative of the solar cycle as a whole, in any given multi-hour observation period it is usually the significant minority of times during which flaring is actually occurring, if at all.

To assign individual observing times to either no-flare or flare categories, we used the time series of both the sub-region area average H α intensity and the x-ray flux. The former is specific to just the extracted sub-region, which is our area of interest for flare diagnosis, while the latter is a measure of the full disk emission in the 1-8 Å wavelength range. Therefore, H α intensity must be

the primary basis for establishing the predictand category pertinent to the sub-region of interest. However, no flaring can be declared for any sub-region unless it is apparent in the x-ray observations at the same time. For the H α basis, we employed a simple algorithm that set a no flare/flare flag at each measurement time based on the following criterion: the (five-point) smoothed area-average H α intensity time series for a given sub-region rises continually (allowing for brief decreases of $< 10\%$ of the rise to that point) by at least 5% more than the value at the start of the rise (the “base” value) and maintains a value greater than the base value. The entire period of the rise was considered flaring. Then the x-ray flux for any designated flaring times was checked to see if it was $\geq 1 \times 10^{-6} \text{ Wm}^{-2}$. If not, the flag was changed to non-flaring.

We used the Fisher’s Linear Discriminant for two groups (FLDF2G) algorithm (Wilks [6]) to develop the discriminant vector from the predictor vectors and designated predictand category at each observation time in a multiple-sequence training set. A summary of the algorithm is presented in Appendix A. We then tested the discriminant vector by applying it to the predictor vectors for just the training set sequences. At each diagnosis time, we assigned the binary category whose group mean lie closest to the resulting location on the line connecting the projections of the two group means (of the training eigenvector elements). The relative distance to either group mean also provided an estimate of the probability of either of the two groups being the correct one (see Appendix A). If the diagnosed discriminant space location lie beyond either of the group mean locations (a probability of the respective category > 1), then the category probability was set to 1. The resulting diagnosed categories and their probabilities were evaluated against the category specified in the predictands used as input to the FLDF2G algorithm.

In applying the two-category technique to the development data from which the discriminant vector was derived, we found that an excessive number of measurement times had a flaring diagnosis. The FLDF2G algorithm was unable to distinguish between flaring and non-flaring very well. The overall performance was poor for the sequences tested. In evaluating the reasons for the poor performance of using the dual x-ray and H α no flare/flare criteria for setting the binary predictand, we considered as an example the sequence dated 9 May 2003 extracted from an active region of unknown number. Time series of the sub-region area-averaged H α intensity and whole disk x-ray flux for the sequence are shown in Figures 3 and 4 respectively. In the H α intensity plot, the original values are normalized by dividing each one by 600, then the resulting one-minute normalized values are plotted as the dotted line. The solid line represents the five-point smoothed values. In comparing the smoothed H α and observed x-ray flux time series, it is clear that both begin to rise at about 15 UTC and hit a peak shortly after 16 UTC, then irregularly taper down until around 20 UTC, when they both begin to rise again. Yet because of the H α ($\geq 5\%$ rise) and x-ray ($\geq 1 \times 10^{-6} \text{ Wm}^{-2}$) flare criteria, the entire period is designated as non-flaring in the predictand category inputs to the FLDF2G algorithm. This discrepancy between the temporal pattern of H α (characterized by the eigenvectors used as predictors) corroborated by the x-ray temporal variation on one hand, and declaring that no flaring is occurring on the other hand, is apparently confusing the FLDF2G algorithm and consequently obscuring the distinction between non-flaring and flaring in the outcome. This seems to be confirmed when evaluating the time series of the leading H α eigenvectors for this case, shown in Figure 5. Several of the leading eigenvectors have clear inflection points in the period 15-17 UTC, seemingly associated with the

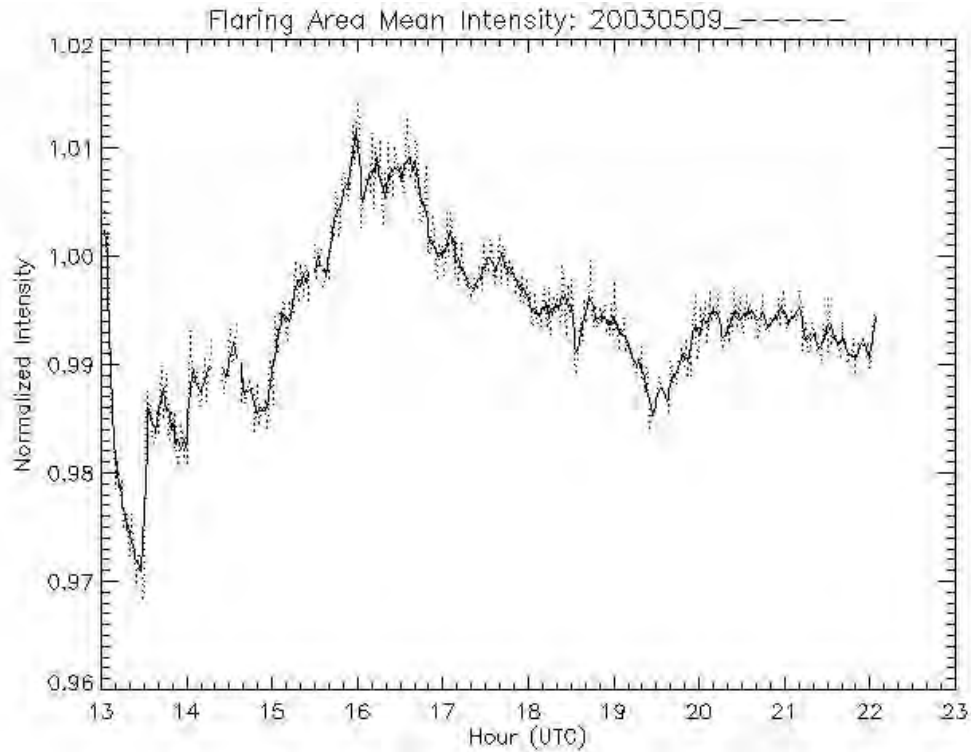


Figure 3. $H\alpha$ Intensity for 9 May 2003 ISOON Image Sequence, Unknown Active Region

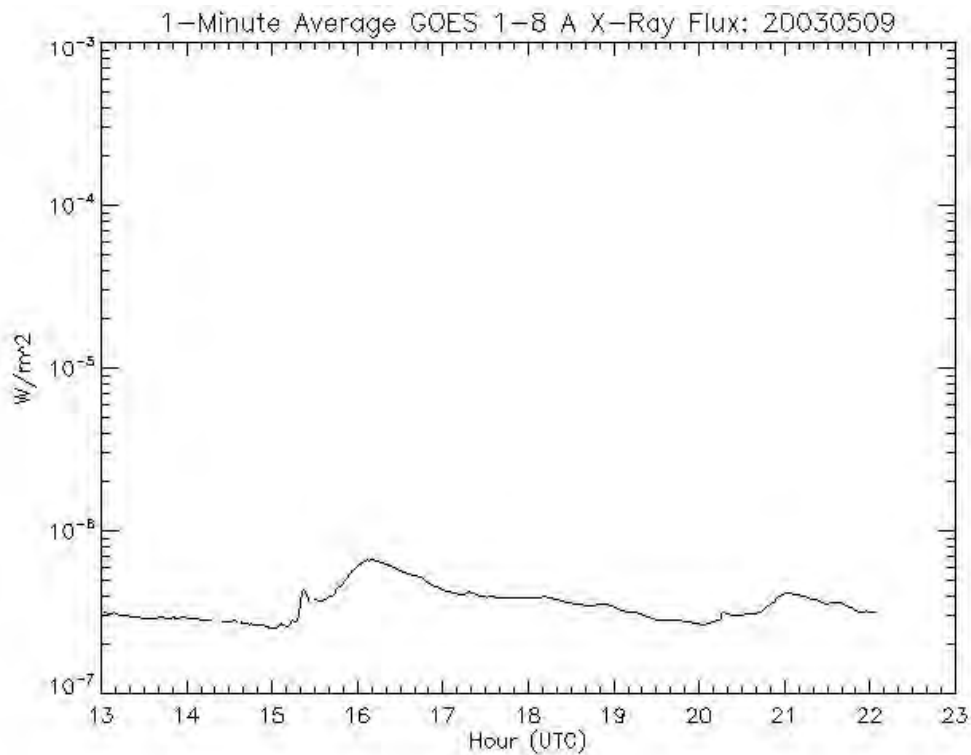


Figure 4. GOES 1-8 Å X-Ray Flux for the One-Minute Image Times Shown in Figure 3

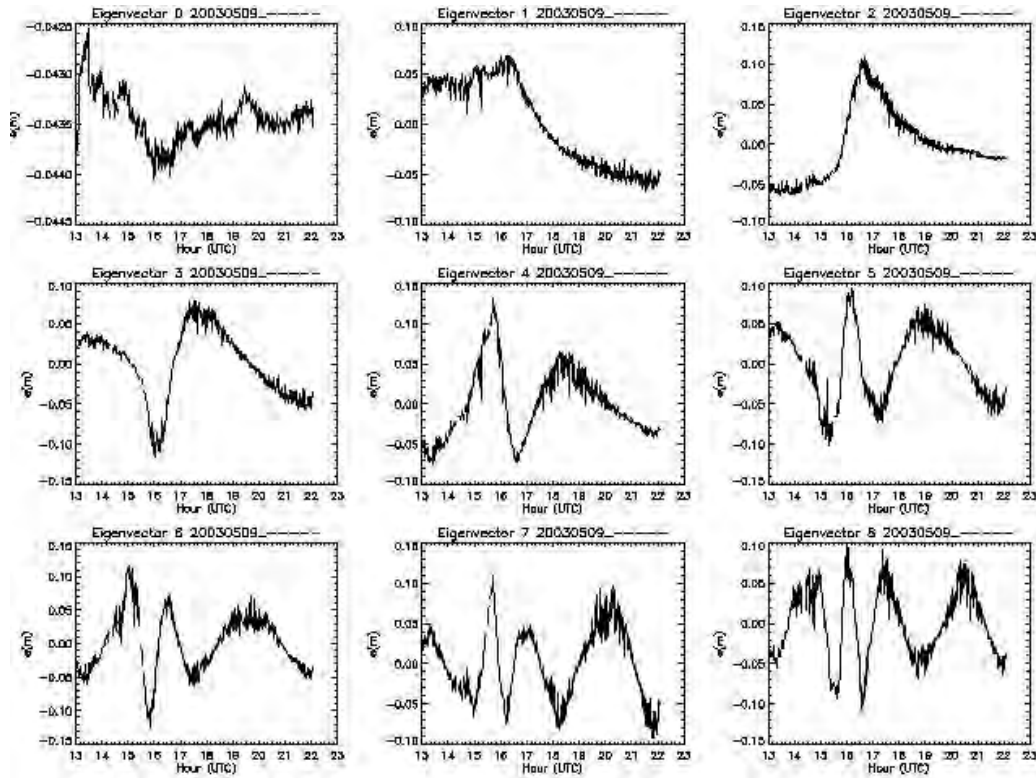


Figure 5. First Nine Eigenvectors of the 9 May 2003 ISOON H α Image Sequence

H α rise during this time. These leading eigenvectors are primary predictors for the binary no-flare/flare predictands. It is only their relationship to H α rises that will allow discrimination between flaring and non-flaring.

The use of two groups (binary predictand) necessitates the use of thresholds to assign the data to the groups. For example, if H α rise percentage is used as the basis for group assignment, how much of a rise constitutes flaring? If one sees significant rises but they are not considered great enough for flaring, this rising might be reflected in the eigenvector predictors, and thus a non-flaring declaration could “confuse” the predictors and reduce their ability to discriminate flaring and non-flaring. To preserve this ability, it seems that there must be consistency between trends in the predictors and predictands. A way to insure that such trends are preserved is to consider the “degree” or relative magnitude of the data upon which the predictand is based. Thus, for small relative H α rises (in which H α remains less than some baseline value) a category of non-flaring is assigned. But for gradually increasing H α intensities, categories of flaring may be assigned that reflect the degree of flaring. Since space weather operations uses “classes” of flare intensity as the standard, maybe degrees of x-ray flux increase from a baseline value could serve as a way to determine the demarcation among corresponding H α departures from a background value. However, it is only fair to try to associate H α and x-ray changes in sequences where there is clearly a match between the rises and falls of H α and x-ray magnitude in the coincident time series.

With this reasoning in mind, we wrote an alternative predictand category assignment algorithm to distinguish flaring level among four groups: non-flaring, weak flaring, moderate flaring, and severe flaring. However, instead of using pre-selected magnitude thresholds for these categories, the ratio of H α intensity and x-ray flux to their respective “background” values was used as the measure of the degree of flaring. First, the algorithm computes the background values for each ISOON sequence using the NOAA SWPC “X-ray Bkgd Flux” algorithm (see <http://www.swpc.noaa.gov/wwire.html>) applied to the x-ray flux measurements at one-minute intervals on the day prior to the date of the ISOON H α sequence to be analyzed. This is also done for the ISOON image sequence H α intensity values for the prior day. In this case, we used the “full disk mean intensity” (FDMI) of H α at the available one-minute observation times of the prior day to compute background H α , or H α (bkgd) using an algorithm similar to the SWPC algorithm. In cases when prior day FDMI values were not available, we used the current day FDMI values to compute H α (bkgd). Then for the ISOON sequence for the analysis date, the correlation between the five-point smoothed, one-minute interval, area-averaged H α intensity divided by H α (bkgd) and the reported one-minute interval x-ray flux divided by x-ray(bkgd) was computed. A total of seven of the 46 sequences were found to have correlations > 0.7 . These sequences were deemed to have a sufficient match of H α intensity and x-ray flux rises and falls in their time series to be used to establish a relationship between H α and x-ray flaring degree categories.

For these H α and x-ray sequences, the five-minute averages of the five-point smoothed, normalized H α intensity, divided by its sequence background value, called $[H\alpha]/H\alpha(bkgd)$, and the five-minute averages of the reported x-ray flux, divided by its sequence threshold value (where the threshold value is the next whole power of 10 higher than the x-ray background value), called $[Xr]/Xr(thrs)$, were computed. The reason that a threshold value was used is so that multiples of ten-fold increase over the threshold would correspond to the standard x-ray classes. Table 1 shows the x-ray flare class category table for $[Xr]/Xr(thrs)$. Values shown for “Lower Bound” And “X-Ray Threshold” are in $W m^{-2}$. We establish flaring categories for $[Xr]/Xr(thrs)$, as follows: category 0 (non-flaring), $[Xr]/Xr(thrs) < 1$; category 1, $1 \leq [Xr]/Xr(thrs) < 10$; category 2, $10 \leq [Xr]/Xr(thrs) < 100$; category 3, $[Xr]/Xr(thrs) \geq 100$.

Table 1. X-Ray Flare Class Category Table for $[Xr]/Xr(Thrs)$

X-ray Flare Class	X-ray Threshold			
	Lower Bound	1.0e-07	1.0e-06	1.0e-05
B	1.0e-07	1	-	-
C	1.0e-06	10	1	-
M	1.0e-05	100	10	1
X	1.0e-04	1000	100	10

To find the corresponding category boundaries for H α , we computed the best fit linear relationship $[H\alpha]/H\alpha(bkgd) = A + B \log \{[Xr]/Xr(thrs)\}$ from the five-minute average $[H\alpha]/H\alpha(bkgd)$, $[Xr]/Xr(thrs)$ pairs from the seven highly correlated sequences. The scatter plot and best fit line are shown in Figure 6. The overall correlation was 0.58, with $A = 1.03057$ and $B = 0.10903$. Using this linear relationship between $[H\alpha]/H\alpha(bkgd)$ and $\log [Xr]/Xr(thrs)$, Table 2

shows the $[H\alpha]/H\alpha(\text{bkgd})$ boundaries for the four flaring categories corresponding to Table 1 for $[Xr]/Xr(\text{thrs})$.

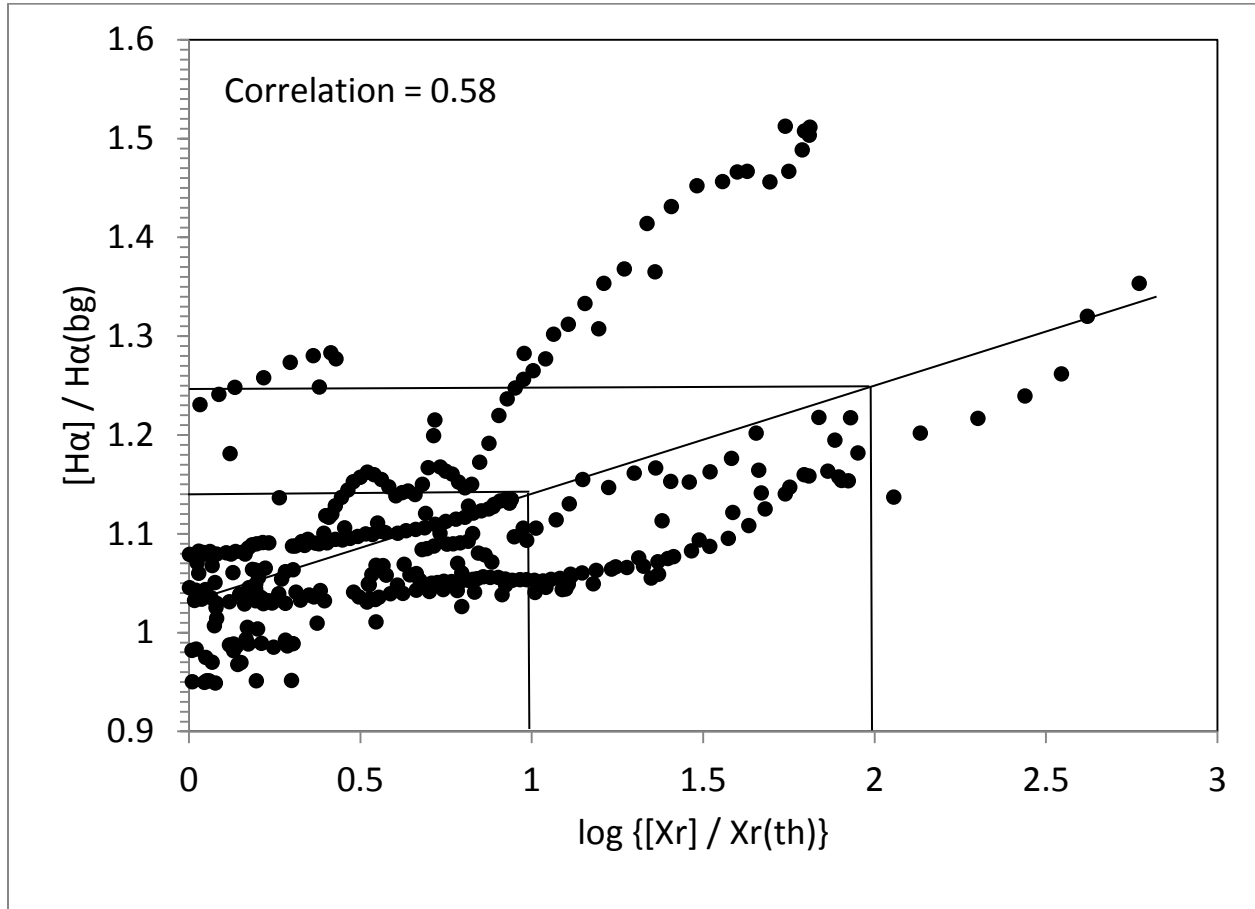


Figure 6. $H\alpha/H\alpha(\text{Bkgd})$ vs. $\log \{[Xr]/Xr(\text{thrs})\}$ for Seven Sequences of Correlation > 0.7

Table 2. X-Ray Flare Class Category Table for $[H\alpha]/H\alpha(\text{Bkgd})$

X-ray Flare Class	Lower Bound	X-ray Threshold		
		1.0e-07	1.0e-06	1.0e-05
B	1.0e-07	1.03057	-	-
C	1.0e-06	1.13960	1.03057	-
M	1.0e-05	1.24863	1.13960	1.03057
X	1.0e-04	1.35766	1.24863	1.13960

Using these values as boundaries for the four flaring categories, we set the following to designate the corresponding categories of $H\alpha$ flaring: category 0 (non-flaring), $[H\alpha]/H\alpha(\text{bkgd}) < 1.03057$; flare category 1, $1.03057 \leq [H\alpha]/H\alpha(\text{bkgd}) < 1.13960$; flare category 2, $1.13960 \leq [H\alpha]/H\alpha(\text{bkgd}) < 1.24863$; flare category 3, $[H\alpha]/H\alpha(\text{bkgd}) \geq 1.24863$. In other words, to be considered flaring in $H\alpha$, the area-averaged intensity value must be about 3% above the previous

day's background value. Which flaring category is assigned is determined by how high above the background the area-averaged H α intensity is: between 3 and 14% greater is category 1, between 14 and 25% greater is category 2, and more than 25% greater is category 3. In recognition of the uniqueness of only H α , and not necessarily x-ray, as the indicator of flaring in an active region subregion of interest, we use only H α /H α (bkgd) as the determiner of the categorical predictand values 0, 1, 2, 3 in the algorithm.

As in the previous discussion on assigning the binary predictand categories, it is necessary to insure that H α flaring in a specific sub-region is corroborated by the simultaneous evidence of x-ray flaring on the solar disk. Without the x-ray flare signal there can be no assignment of an associated flare from the active region sub-region in question. Therefore, in determining the flaring category predictands from the training data, the algorithm examined both the H α and the x-ray time series from the period of interest to confirm any suggestion of flaring. Only at times when x-ray flux exceeded the threshold flux was the predictand allowed to be set to flaring (categories 1-3) as so indicated by the H α intensity.

We used the Fisher's Linear Discriminant for multiple groups (FLDFMG) algorithm (Wilks [6]) for implementing multivariate discriminant analysis (MVDA) on the predictor vectors and associated four-category predictands. The algorithm is summarized in Appendix B. For four groups, three discriminant vectors are derived. For each group g , the sum of the squares of their dot products with each diagnostic time's independent predictor vector ("observation vector") minus the group mean of the training predictor vectors yields the squared distance (D_g^2 , in discriminant space) between the observation vector and the group mean. The group with the smallest D_g^2 is assigned as the most likely flaring category. As described in Appendix B, the four D_g^2 values also can be used to compute the probability p_g that each category is the correct one.

Once a most likely category is diagnosed, it can be converted to an equivalent x-ray flaring class using the conversion table shown in Table 3. This is essentially an inversion of Table 1, where now the diagnosed flaring category is translated into the corresponding x-ray flare class using the x-ray threshold appropriate for the date in which the diagnosis is conducted. Thus, for a given diagnosis date, the four flaring categories correspond to no flaring and three x-ray classes from B- through X-class, depending on the x-ray threshold for the date. The category diagnosis performance using this approach to assigning flaring categorical predictands, which we will refer to as H α Magnitude Flare Categorization (HMFC), was evaluated against the specified predictand category values using the same technique. The results are discussed in the next section.

Table 3. Conversion Table from Flaring Categories to X-Ray Flare Class

Flaring Category	X-ray Threshold		
	1.0e-07	1.0e-06	1.0e-05
0	No Flare	No Flare	No Flare
1	B	C	M
2	C	M	X
3	M	X	X

Another approach to assigning flaring predictand categories based on ISOON H α intensity image sequences is termed the H α Eigenvector Flare Categorization (HEFC). This technique was based on a study of the temporal characteristics of the leading H α eigenvectors of the 46 sequences. We noticed a similarity in the temporal pattern of the leading eigenvectors among sequences depicting like levels of flaring. We found that the patterns of the leading eigenvectors seem to fall into four groups as shown by the examples of each type in Figure 7. Qualitatively, we describe the temporal variations of the four flaring level indicators (FLIs) as: FLI = 0 for no flare above x-ray background and smoothly varying (sinusoidal) eigenvectors, e.g., Figure 7(a); FLI = 1 for weak flares (in the x-ray decade of the background value) with spiked otherwise smoothly varying (sinusoidal) eigenvectors, e.g., Figure 7(b); FLI = 2 for moderate x-ray flares and smoothly curved (non-sinusoidal) eigenvectors before and after the flare spike, e.g., Figure 7(c); FLI = 3 for strong x-ray flares with non-curving eigenvectors before and smoothly curving after sharp spikes, e.g., Figure 7(d). In FLI = 1-3, the spike occurs at the same time as the sharp rise in the area-average smoothed H α intensity. We predetermined the FLI for each ISOON sequence based on a subjective assessment of its eigenvector patterns and the associated x-ray flux peak value. If there was more than one qualifying flare in a sequence, each flare got its own FLI. The FLI and the bounding hours of the flare rise were input into the MVDA development algorithm for each pre-identified flare. In setting the categorical predictand at each observation time in each of the sequences, the algorithm identifies the maximum rise in H α intensity and x-ray flux between the bounding hours. At each time during the H α rise, the predictand is set to the predetermined FLI – at all other observation times, the predictand is set to category 0. By restricting the flaring designation to just the flare rise times, we felt we could maximize the association between flaring apparent in the area-average H α intensity (times when the predictand indicates flaring) and the H α eigenvectors (times when the predictors indicate flaring).

It should be mentioned that, in the maximum H α intensity and x-ray flux rise identified in the pre-determined time segment, we impose a requirement that x-ray flux maximum exceed its background value. There are two reasons why we did not require the same for H α intensity. First, we found that in some sequences the H α intensity from the prior-day sequence of FDMI, used to set the current day's background value, was actually greater than the sequence of the current day's sub-region area-average H α intensity. Area-average H α intensity failing to exceed its background in such events would mean setting predictands to zero at the H α rise times, eliminating what were indicated as flares in the corresponding H α eigenvector pattern. As such, flare indicators in the predictors would be inconsistent with indication of no flaring in the predictands. Second, the peak value of area-averaged H α intensity should not be used as an indicator of flare magnitude because it does not represent the brightening in the specific area of flaring. Instead, it is the average H α over the entire sub-region of the active region, and as such is expected have a lesser magnitude than the flare itself. Using it would under-represent the actual intensity of the flare. Therefore, the algorithm checked only the peak value of the x-ray flux against its background, and the non-zero FLIs were set at the H α rise times only if it exceeded the background value.

The MVDA application algorithm simply ingests the discriminant vectors as generated by the development algorithm, and applies them to predictor vectors of leading eigenvector elements for any sequence to be diagnosed. Nominally, this should be any sequence not included in the training set – that is, the independent cases. However, we also applied the discriminant vectors to

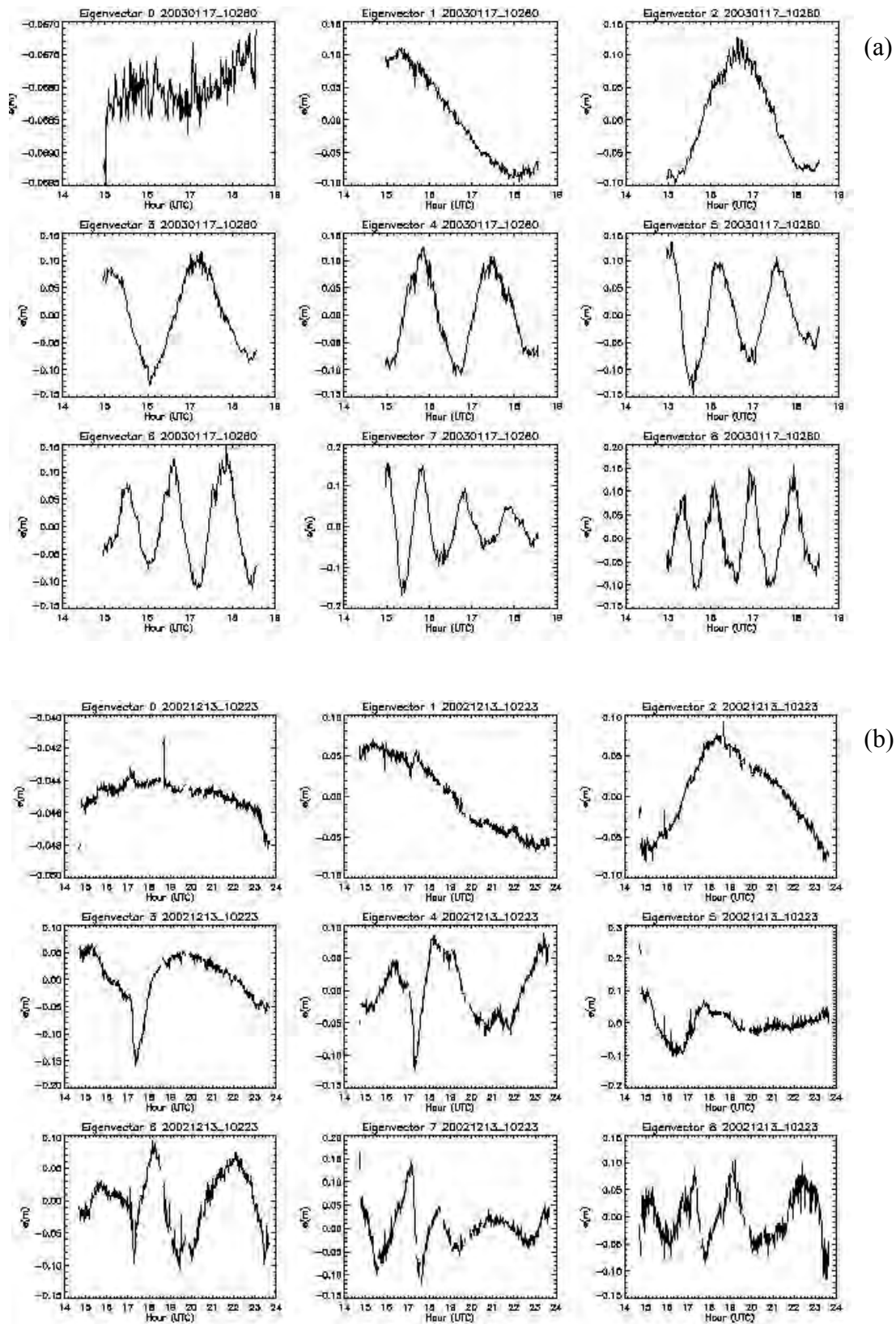


Figure 7. Eigenvectors from Examples of FLI (a) 0, (b) 1, (c) 2, and (d) 3 Sequences

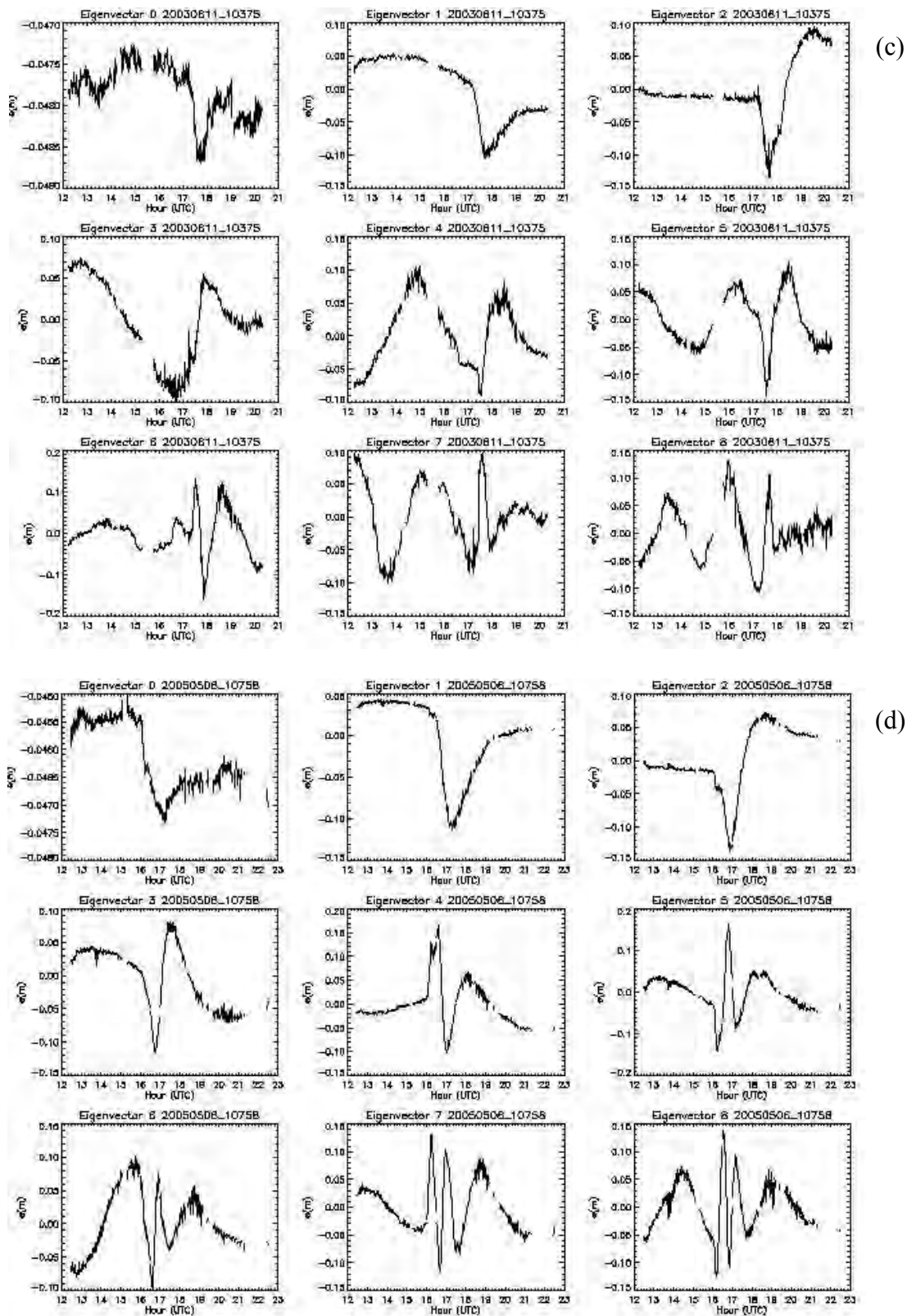


Figure 7. (Cont.)

the dependent (training set) sequences in order to see if there was a discernable difference in the diagnosis skill. After computing the D_g^2 and the p_g for each of group (category) g at each observation time, the group with the largest p_g (or equivalently, the smallest D_g^2) was declared the most likely category. In order to compare the diagnosed category with the specified one at each time, the same categorization scheme (HMFC or HEFC) that was used to set the predictands in the development algorithm is used to specify the category in the application algorithm. Thus, the comparison between diagnosed and specified flaring categories at each ISOON image time is an evaluation of the MVDA development algorithm's ability to discriminate among the categories given the category assigning method. This discrimination ability is key to determining whether or not discriminant analysis is a viable approach to diagnosis, and eventually prediction, of solar flare activity based on time series of ISOON H α intensity images.

4. RESULTS

The comparison between diagnosed and specified flaring categories constitute the discussion of results in this study. However, to get a preview of how discriminating the MVDA scheme is likely to be, we show the results of the discriminant vectors dotted with each of the predictor vectors of a selected training set of ISOON cases. Since our two categorization schemes HMFC and HEFC have four predictand categories, both have three discriminant vectors. Dotting each one with the training set predictor vectors, and plotting the resultant scalar value in a separate three-dimensional discriminant space grid for each of the four groups gives an idea of the group-to-group separation and spread of each group's discriminant function values. As mentioned above, maximizing the separation and minimizing the spread of the points in each group optimizes the ability of the multivariate discriminant analysis to provide the most unambiguous category diagnosis.

In Figure 8, we show the discriminant space plots for each of the four groups when a training set called "Case 1" was used in the MVDA development algorithm along with the HMFC predictand categorization approach. The Case 1 training set consisted of eight ISOON H α image sequences made up of two sequences of each of the four flaring level indicator classes of the HEFC categorization scheme. We chose to train on the same training set for both the HMFC and HEFC schemes in the MVDA algorithm in order to allow a more direct comparison between the diagnostic performance of the two approaches. Because the scale of the discriminant space plots are significantly different for the HMFC and HEFC schemes, it is difficult to compare them qualitatively.

In the Case 1 HMFC plots in Figure 8, most apparent is the sharp drop in the number of occurrences of flaring categories (1-3) compared to the non-flaring category (0). That would be generally true regardless of what ISOON sequences we selected for the training set. In the HMFC scheme, every point in the time series in which the smoothed, area-averaged H α intensity was more than about 3% above its background value and had an x-ray flux level above the x-ray threshold was considered flaring (that is, was in category 1-3). That produced a lot more flaring occurrences than did the HEFC scheme (discriminant plots not shown) which had flaring predictands only during the identified H α flare rise phase when x-ray flux was above its background level. While it is difficult to pinpoint the location of the group mean within each of

the four plots (which is just the dot product of each of the discriminant vectors and the group mean predictor vector), it is clear from the plots that though each group “centroid” is located in a different place in discriminant space, there is not a great deal of distance among them. The greatest group-to-group shift in the group mean’s position is in the first discriminant function (left to right axis). In the second and third discriminant function dimensions, the points seem largely confined to the back fifty units and bottom 100 units respectively. Importantly, the flaring groups (1-3) spread into the domain of the non-flaring points (0), which does not bode well for the basic distinction between a non-flaring and flaring diagnosis. However, because these are just subjective evaluations of the discriminant function values, we should not draw any definitive conclusions about the prospects for successful flaring category diagnosis from these plots. The

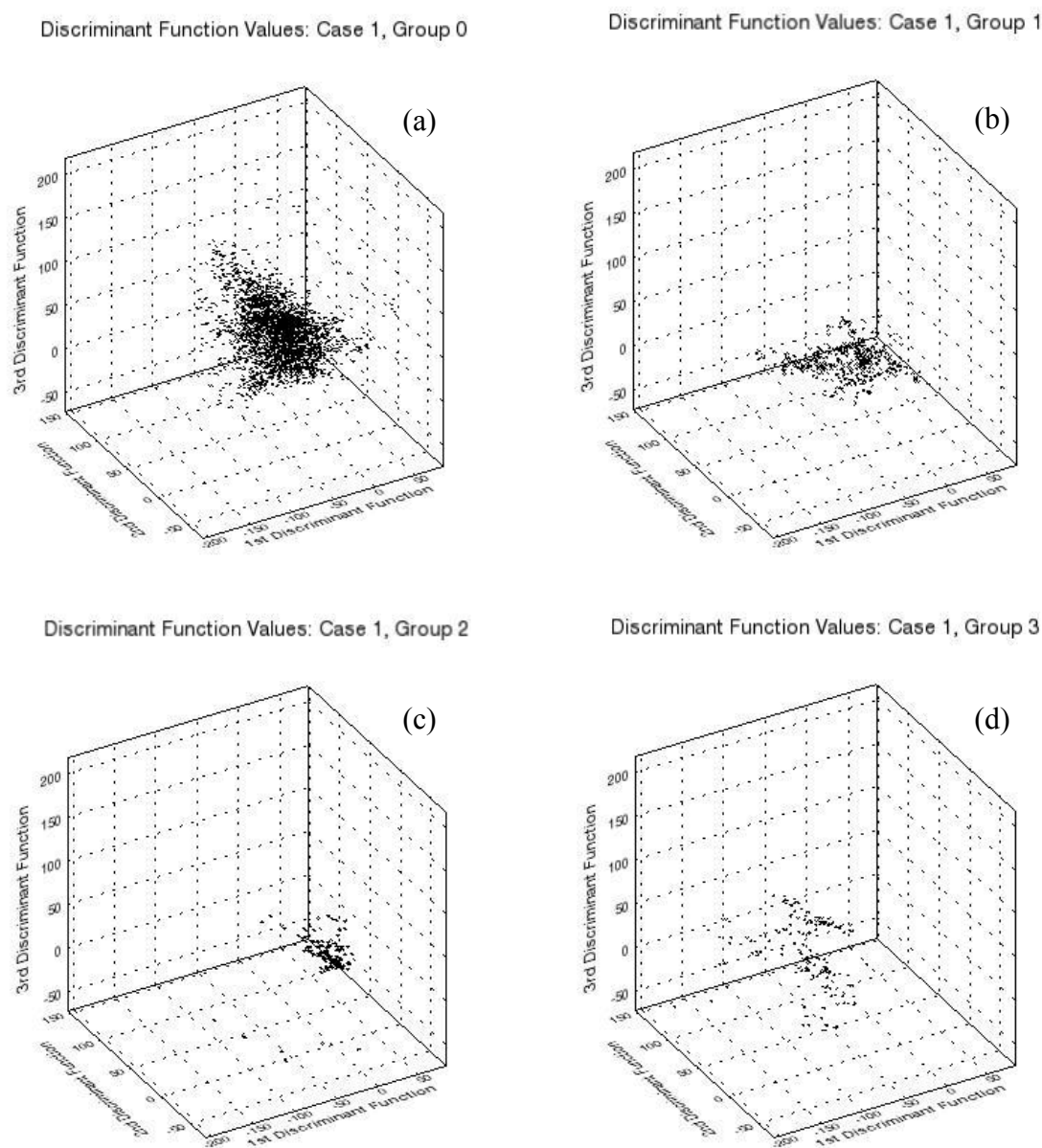


Figure 8. Discriminant Space Plots from the MVDA/HMFC Development Algorithm: (a) Group 0, (b) Group 1, (c) Group 2, and (d) Group 3

proof of discriminant analysis' ability to differentiate among prescribed categories lies in the application of the discriminant vectors to the independent predictor vectors and the comparison of the resulting category diagnoses with the specified category.

To perform this evaluation quantitatively, we computed several statistical metrics to draw out the performance characteristics of the flare category diagnoses. Because the categorical method of MVDA we are using allows for computing the probability of each of the categories being the correct one at each observation time, we can calculate a probability-weighted diagnosed flaring category C_p given by

$$C_{p_i} = \sum_{g=0}^{G-1} g p_{g_i} \quad [2]$$

at each observation time i , where in our study there are $G = 4$ groups $g = 0, 1, 2$, and 3 in the HMFC and HEFC predictand schemes. While earlier we mentioned that the group with the largest probability p_g is objectively the most likely correct category, if the probabilities are similar then we are “hedging” our diagnosis in saying that all for categories are nearly equally likely. That is generally an undesirable outcome, because we want an unambiguous diagnosis, and ultimately prognosis, upon which to stand for decision-making purposes. Therefore, a quantitative measure of the uncertainty of the most probable category is the diagnosis uncertainty

$$DU = \frac{1}{N} \sum_{i=1}^N [1 - \max_g (p_{g_i})] \quad [3]$$

which is the likelihood that the correct category is not the maximum probability category, averaged over the total number of observation times evaluated N . Since we evaluate the MVDA diagnosis performance separately for each ISOON H α sequence, N represents the number of image times in each sequence. DU is an important metric because it measures decisiveness the category diagnosis. The smaller the value of DU for a given image sequence diagnosis, the more certain the flaring category diagnoses over the full time series. But though it indicates how sure one is of the diagnosed category, it doesn't convey any information about its correctness. The metric that best measures overall skill of the diagnoses for each image sequence evaluated is the Brier Score

$$BS = \frac{1}{N(G-1)^2} \sum_{i=1}^N (C_{p_i} - C_{o_i})^2 \quad [4]$$

where C_o is the “observed”, or specified category at each image time i . In the usual computation of the Brier Score for binary predictands (e.g., Wilks [6]), $G = 2$ so the Brier Score is just the mean square error of the diagnosed probability of the event occurring ($C_o = 1$) or not occurring ($C_o = 0$). However, for $G > 2$ as in our study, C_o can take on more than just the binary values of 0 and 1, and C_p has the potential for the full range of values from 0 to $G-1$. The factor $(G-1)^2$ in the denominator ensures that BS remains in the range of 0 to 1. Another metric of interest is the bias of the diagnosed flaring categories, given by

$$Bias = \frac{1}{N(G-1)} \sum_{i=1}^N (C_{p_i} - C_{o_i}) \quad [5]$$

which measures the under- or over-diagnosis of the flaring category, resulting in too many missed flares or too many false alarms, respectively. Bias represents the systematic error of diagnosis in our study, whereas BS quantifies the random component of the error. As a reference

for bias, we also show the fraction of flaring times (FFTs) in each image sequence, so that we can tell if there is an association between the bias and the prevalence of flaring activity.

We can also compute the statistics above for all diagnosis times in all sequences, giving an overall assessment of the performance of the diagnostic method. This is useful in comparing the ability of differing techniques, but is also informative in determining the sensitivity of a given scheme to different combination of image sequences used in the training set for the MVDA. A metric to measure the ability of a diagnosis scheme to reproduce the frequency distribution of the observed predictand categories is the frequency distribution fit (FDF, Norquist [7])

$$FDF = \frac{1}{N} \sum_{g=0}^{G-1} \left| \frac{m_g}{n_g} - 1 \right| n_g \quad [6]$$

where m_g is the number of image times in which group g is the most likely category (greatest diagnosed probability), n_g is the number of image times specified (“observed”) to be in group g , and N is the total number of image times over all sequences evaluated. A perfectly reproduced frequency distribution yields an FDF = 0. While FDF could be computed for each sequence separately, the collection of image sequences offers a more robust sample size for the determination of the reproduction of the frequency distribution.

4.1 H α Magnitude Flare Categorization Technique

We show graphically the metrics computed from the C_p values resulting from application of the discriminant vectors from the MVDA with the HMFC predictand specification technique employed on the training set Case 1 in Figures 9 and 10. In both figures, the date_active region ISOON image sequences whose eigenvector predictors and specified predictand categories were used in the MVDA development are indicated by an asterisk. We might expect that the subsequent application of the derived discriminant vectors might result in more skillful flaring category diagnoses for the “dependent” sequences than the independent ones, since they are the ones on which they were derived. However, since each sequence is so different in the temporal signature of the H α intensity, it is difficult to discern any noticeable difference in the metrics between the dependent and independent sequences. One could only quantify this effect if more than one training set were used, in which a particular sequence was used in one set and not in another. For now, we focus our attention on the relative magnitudes of the several metrics to make sense of the diagnostic scheme’s performance.

In Figure 9, we see that the DU remains in the vicinity of 0.4 – 0.5 for all of the sequences, largely independent of the prevalence of flaring (see FFTs in Figure 10). Thus, there is an only slightly better than even chance that the most probable category is the correct one in comparison to the other three categories for most of the sequences. That is, it is almost equally likely that one of the unselected categories is the correct one, even though which one it might be is unknown. As mentioned earlier, this is an undesirable outcome since it is important to maximize the certainty of the diagnosis to optimize its usefulness. The ambiguity of the diagnosed probabilities stems from the inadequacy of the discriminant analysis method’s ability to clearly distinguish among the predictand categories. This is reflected in the overlap in discriminant space among the

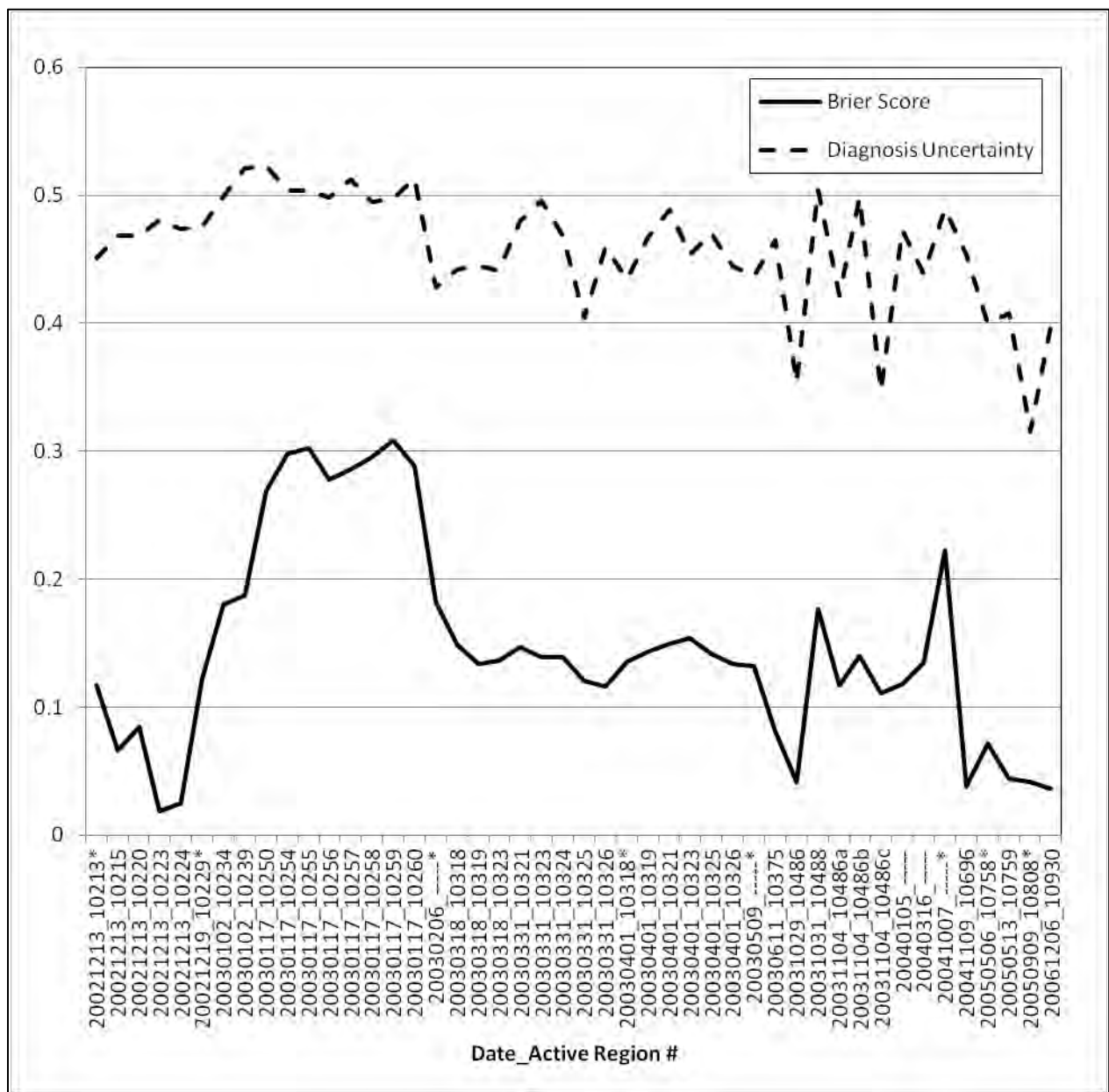


Figure 9. Brier Score and Diagnosis Uncertainty from Application of the MVDA/HMFC Development Algorithm

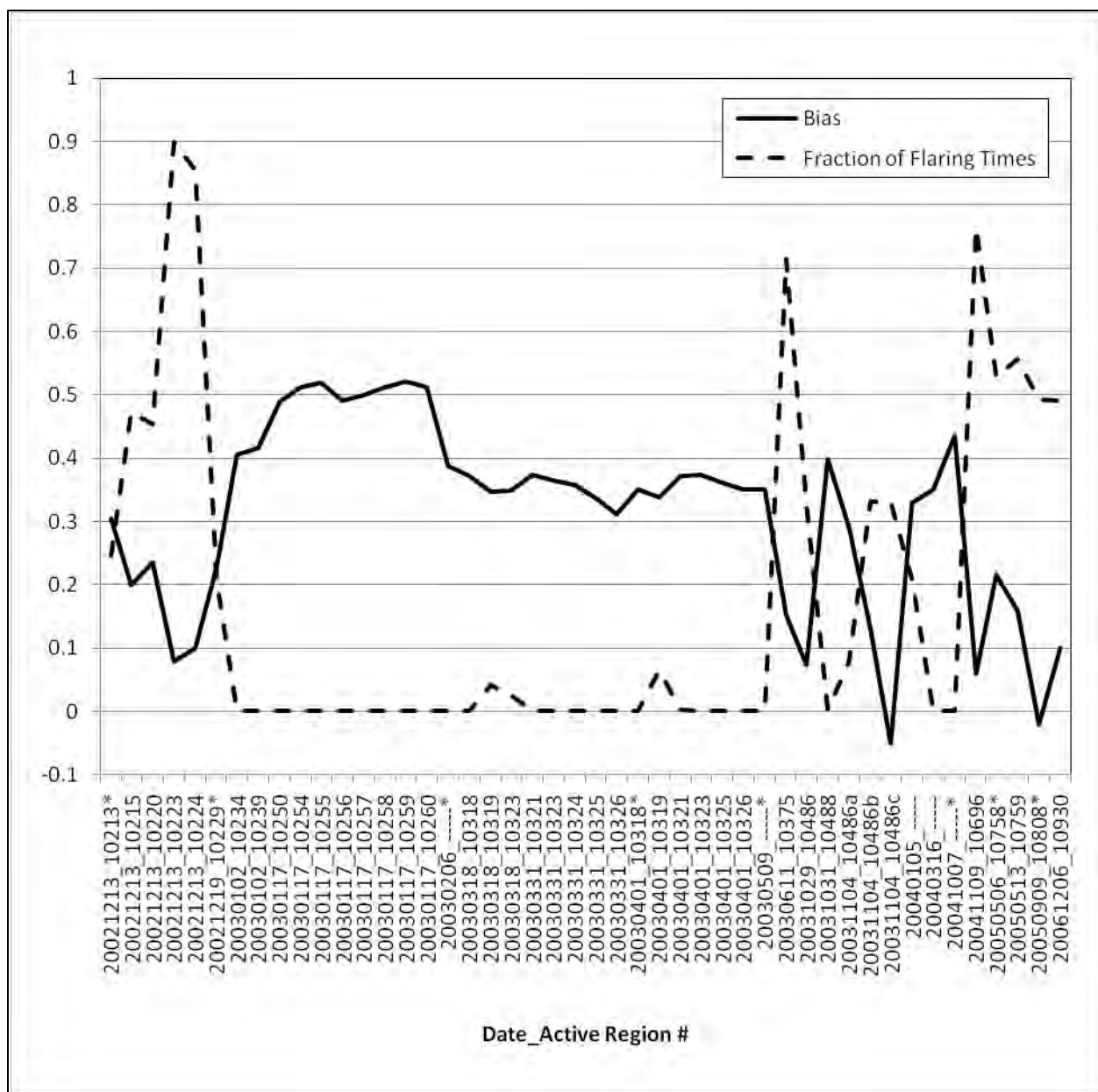


Figure 10. Bias from Application of the MVDA/HMFC Development Algorithm, and Fraction of Flaring Times

discriminant functions as plotted in Figure 8, showing the application of the discriminant vectors to the training set predictor vectors.

Next, we consider the skill of the probability-weighted category diagnosis as reflected in the BS (Figure 9) and bias (Figure 10). For these metrics, there is an apparent association with FFTs in the image sequences. Both BS and positive bias are greatest in the least flare-active sequences, particularly in those of 20030117. Even in the non-flaring sequences on and after 20030509, there is a strong relationship between lack of flare activity and magnitude of positive flare category bias. This is generally true of BS as well, though not as pronounced. For example, from sequence 20031104_10486c (highly flaring) through 20040316_----- (no flaring), there is a great rise in positive bias but only a minor increase in BS. The MVDA with the HMFC method for setting predictand category seems to be producing an excessive number of flaring false alarms, particularly in inactive sequences. Overall, a minority of image times are specified as flaring (19.4%) according to the HMFC method, yet the MVDA/HMFC technique diagnose one of the flaring categories (1-3) as most probable in 49.7% of the image times.

Figure 11 depicts the frequency distribution of diagnosed vs. specified flaring categories resulting from the HMFC predictand specification method. The chart shows that the diagnostic scheme over-represents the number of times of moderate and strong flaring categories by a factor of 8 to 10. More than twice as many flaring times are diagnosed as are specified. The FDF for this technique is 0.644. This compares with an FDF = 0.214 using the frequency of occurrence of no flares (category 0 here) and the same for C5+, M-, and X-class flares (corresponding to the current categories 1, 2, 3 respectively) from Tables 1 and 5 respectively of Norquist [7] and comparing them with the frequency distribution of specified flaring categories from this study. This suggests that the performance of the MVDA/HMFC diagnostic technique is inferior to climatology in terms of reproducing the frequency distribution of flaring occurrence.

4.2 H α Eigenvector Flare Categorization Technique

The alternative method of specifying multiple predictand categories in this study is the H α Eigenvector Flare Categorization technique. We employed the leading eigenvectors as predictor vectors for each ISOON sequence used in the training set in the same way that was done for the MVDA/HMFC development and application. But for the MVDA/HEFC scheme, we assigned the predictand categories at each image time in each sequence according to the character of the leading eigenvectors for the corresponding sequence and the peak x-ray flux of the flare. Recall also that we set flaring categories (1-3) only during the H α rise times of the pre-identified flares. This, combined with the requirement that peak x-ray flux exceed its background value in declaring flaring, resulted in only 3.1% of the image times for all 46 cases specified as flaring (categories 1-3) compared to 19.4% for the HMFC predictand specification method.

We used the Case 1 training set of image sequences for the MVDA/HEFC development and application so that we could make a direct comparison with the MVDA/HMFC performance. In order to determine the sensitivity of the performance of the diagnosis to the training data used, we also used two other training sets, called Case 2 and Case 3. All three training sets included eight image sequences, and each included the full range of FLI values. Figures 12-14 show the performance metrics for all 46 sequences resulting from employing MVDA/HEFC with the three

different training sets. In the three figures, the date_active region ISOON image sequences whose eigenvector predictors and specified predictand categories were used in the MVDA development for each training set are indicated in parentheses by case number.

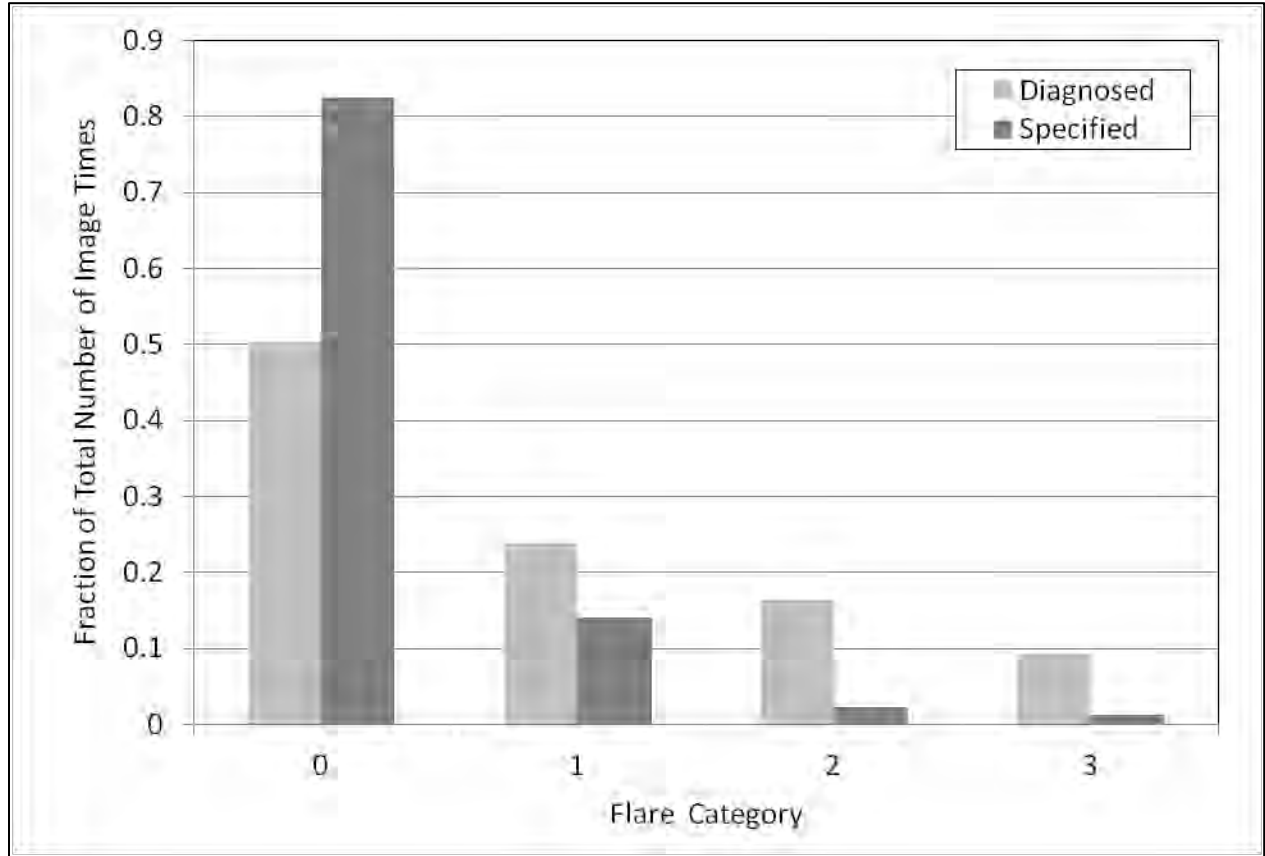


Figure 11. Frequency Distribution from Application of the MVDA/HMFC Algorithm

Figure 12 shows the BS for each case and the fraction of specified flaring times in each ISOON sequence resulting from the HEFC method. We can directly compare the plot of Case 1 in Figure 12 with the BS plot in Figure 9. For most (78%) of the image sequences, the BS is considerably smaller (better) for the HEFC method than for the HMFC scheme. The overall BS for the 46 sequences is 27.3% less for HEFC than for HMFC in Case 1. The two curves have a similar shape, suggesting that the diagnosis skill relative to the various sequences is alike. However, the variation of the fraction of specified flaring times is significantly different between HMFC and HEFC, as evidenced by the respective curves in Figures 9 and 12. This difference combined with the close association between fractional flaring and BS in the HMFC results suggests that there is little fractional flaring – BS association with HEFC. For example, in Figure 12, in the image sequences from 20030117_10250 to 20030117_10260, the BS remains large in spite of the fractional flaring ranging from 0 to 10% among the sequences. Comparing 20040316 to 20041007, fractional flaring is constant at zero while BS varies between 0.068 to 0.128.

Bias as shown in Figure 13 even more clearly shows this lesser dependence of flaring diagnosis on fractional specified flaring for HEFC. Unlike in Figure 10 (HMFC) where for most of the

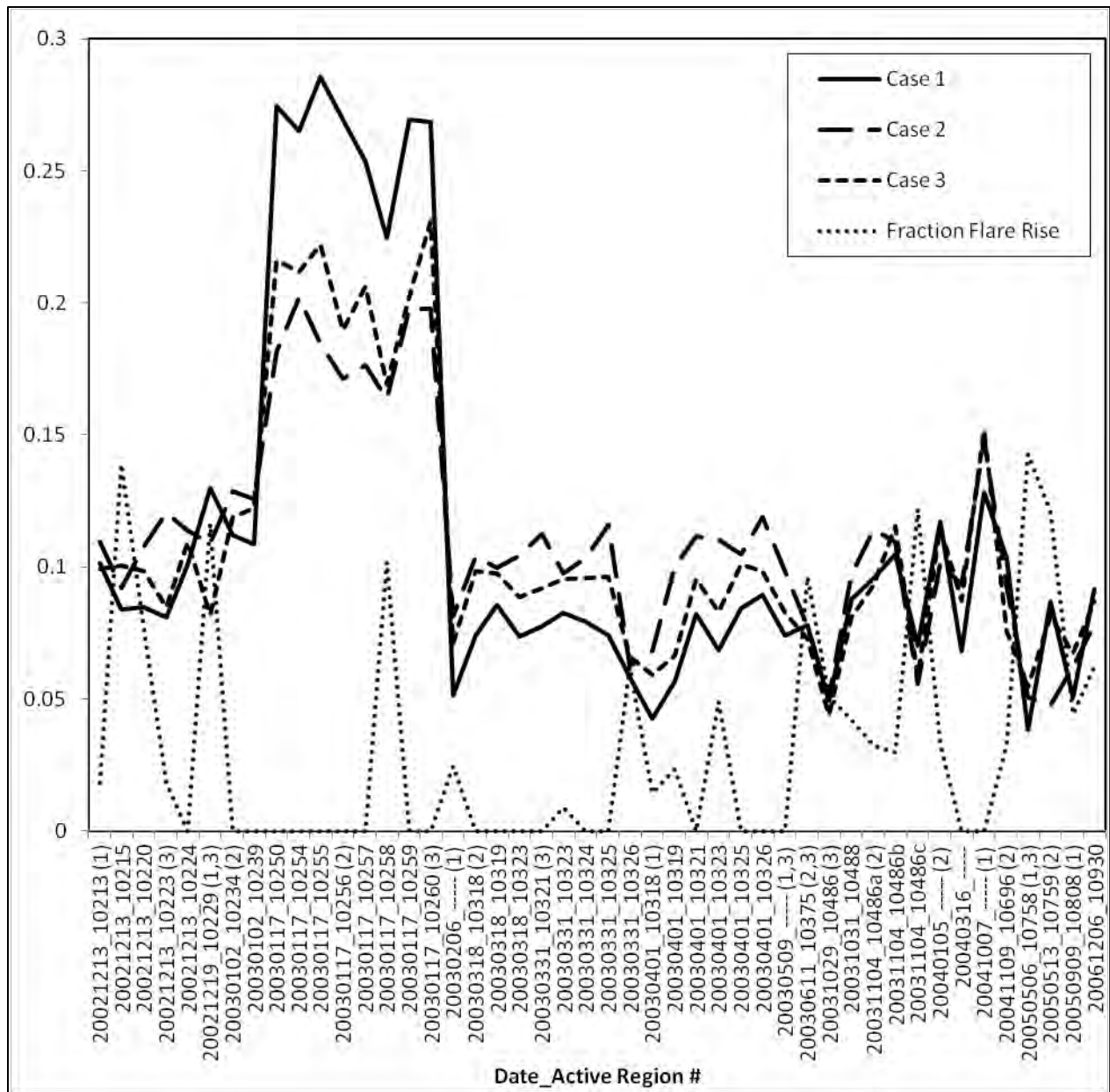


Figure 12. Brier Score from the MVDA/HEFC Development Algorithm, and Fraction of Flare Rise Times

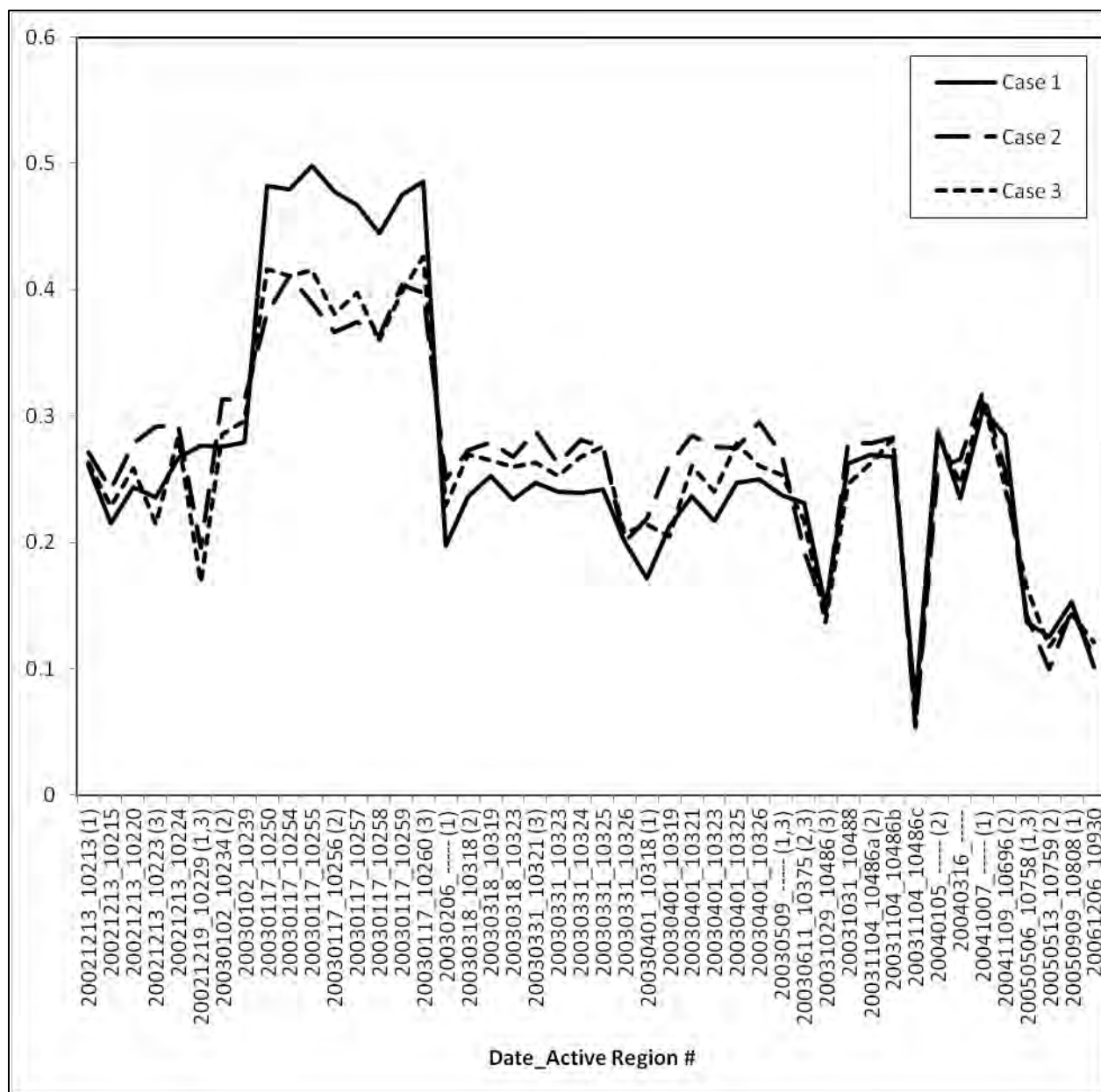


Figure 13. Bias from the MVDA/HEFC Development Algorithm

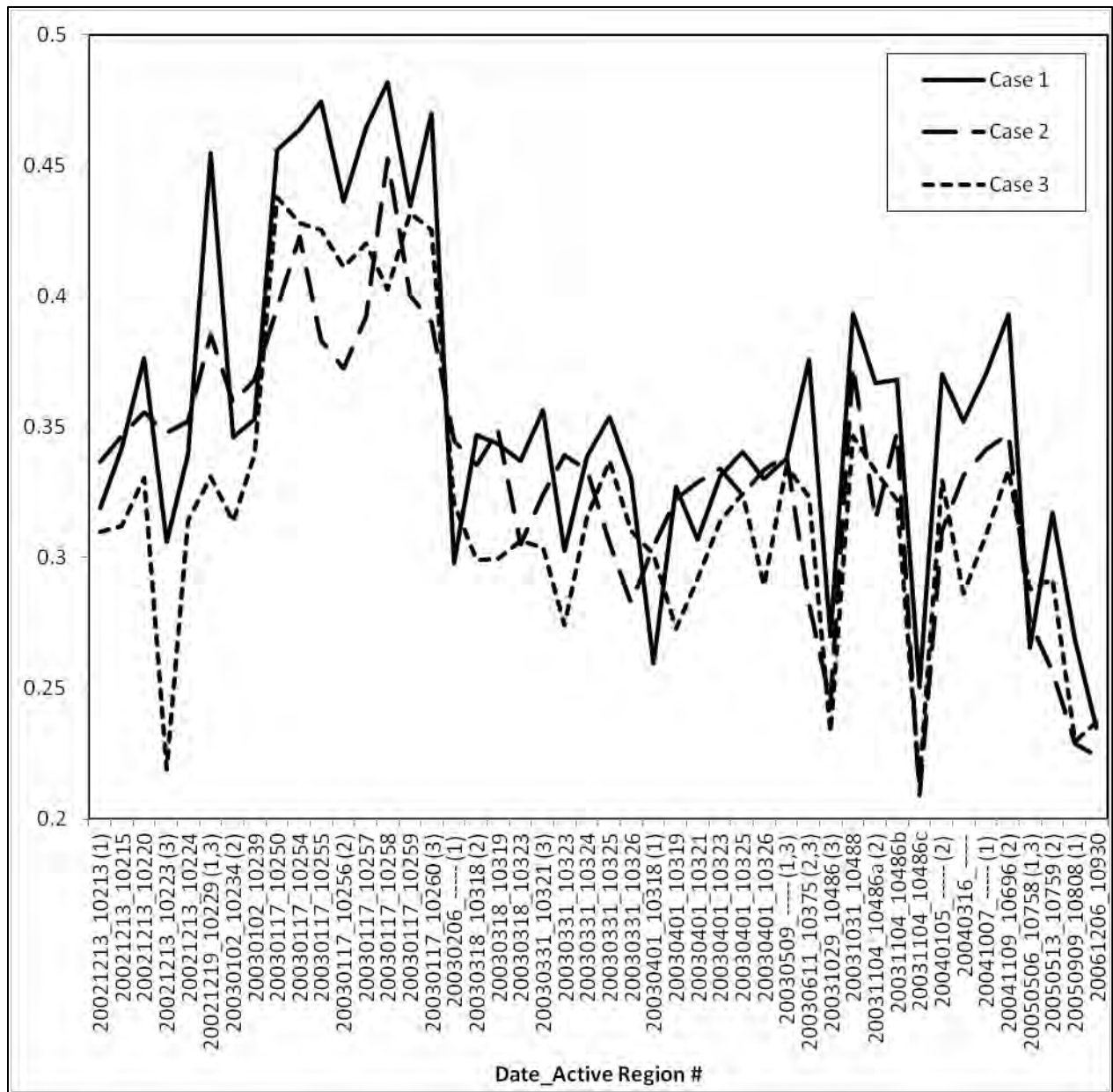


Figure 14. Diagnosis Uncertainty from the MVDA/HEFC Development Algorithm

sequences the bias is inversely related to the flaring activity, in HEFC there are just a few sequences (notably 20031029_10486, 20031104_10486c and 20050506_10758) in which greater flaring activity is linked to appreciable drops in positive bias. Overall, the positive bias is 16.5% less for HEFC than for HMFC. This is all the more remarkable considering that there are more than six times as many specified flaring times in the 46 sequences for HMFC than for HEFC, and the greater the prevalence of observed flaring, the less the positive bias for a given diagnostic scheme.

Figure 14 shows the DU values resulting from the MVDA/HEFC development and application on Cases 1, 2 and 3. Comparing just the curve for Case 1 with the same from the HMFC scheme shown in Figure 9, we see that there are two major differences. First, there is greater degree of variation in the HEFC plot. Whereas the HMFC curve is largely confined to values between 0.4 and 0.5, the HEFC curve is much more irregular, and varies from 0.23 to 0.48. Second, with the exception of the performance in the 20030117 sequences, the DU for HEFC is considerably smaller than it is for HMFC. As a result, the HEFC Case 1 shows a greater contrast between the DU values of the 20030117 sequences and those of the other sequences than do the HMFC DU values. Considering all 46 sequences together, the DU using the HEFC method is 24.6% less than for the results of the HMFC technique for Case 1. Thus, the HEFC produced flaring category diagnoses that were not only more accurate, but also more certain.

For at least for Case 1 training set, the performance of the HEFC technique is superior in all three metrics to that of HMFC in the development and application of the MVDA on the ISOON image sequences. We now look at the sensitivity of the HEFC method to the choice of training data.

Looking back at Figure 12, we can compare the BS computed from the application of the MVDA/HEFC technique for all three training sets. A first impression is that the diagnosis skill is quite similar among the three cases. All three display most of the same major maxima and minima BS values, for example in the 20030117 sequences (maximum BS) and 20050506_10758 (minimum BS). Overall, there is not a great deal of difference in the flaring diagnosis performance stemming from the use of three different training sets in the MVDA/HEFC scheme. Table 4 lists the overall assessments (results from all 46 sequences) of the metrics for all three training sets from the MVDA/HEFC development and application. The biggest difference for BS is between Case 1 and Case 2, where the former is about 8% less than the latter. Given that we are using only 46 ISOON sequences, there are a limited number of training set combinations that can be tried. With a greater number of sequences, we could get a better assessment of the sensitivity of the MVDA/HEFC performance to the training set used.

Table 4. Performance Scores for the MVDA/HEFC Development/Application Process

	BS	Bias	DU	FDF
Case 1	0.09530	0.24504	0.34075	0.27558
Case 2	0.10362	0.25688	0.32475	0.24716
Case 3	0.09821	0.24695	0.30955	0.25973

Considering now the bias as shown in Figure 13, we see the same major maxima and minima in the positive bias that we saw in the BS for all three cases. From Table 4, we can see that the Bias² (that is, the mean error squared) makes up from 62-64% of the BS (that is, the mean square error), showing that the systematic component is greater than the random component (diagnosis error standard deviation) in the overall diagnosis error. Again, the greatest bias difference among the results from the three cases is between Case 1 and Case 2, where the former is 4.6% less than the latter.

We next compare the DU resulting from MVDA development/application using the HEFC scheme for the three training sets as shown in Figure 14. As with BS and bias, the overall shape of the curves are similar for all three cases. In fact, the same major features of the BS and bias curves are present in the DU curves except for the sharp DU increase in the 20021219_10229 sequence not evident in BS or bias. Whereas BS and bias deal with diagnostic skill and DU reflects diagnosis certainty, these results suggest that there are links between them. For example, when skill is poor in the 20030117 sequences, uncertainty is also elevated. When skill is good, as in the 20031029_10486, 20031104_10486c and 20050506_10758 sequences, diagnosis certainty is much greater. Logically, this makes sense since when the probabilities of all four categories are similar, there is a better chance of a misdiagnosis of the correct category than when the probability of a certain category dominates. Overall, however, Case 1 has the best BS and bias but the worst DU, as seen in Table 4. Case 3 DU is about 9% smaller than Case 1, comparable to the Case 1-2 difference in BS. It appears that the link between diagnosis skill and certainty evident from sequence to sequence is not apparent in the overall statistics.

The last entry in Table 4 is FDF, a measure of how well the diagnosed flaring category (chosen as the one with the largest probability at each diagnosis time) frequency distribution fits that of the specified flaring categories. We see that for all three cases, when HEFC was employed the fit to the specified categories (FDF = 0.276) was much better than for HMFC (FDF = 0.644) on Case 1. In fact, the FDF for all three cases are competitive with the fit derived from the flaring climatology of Solar Cycle 23 (FDF = 0.214). Figure 15 shows graphically the frequency distribution for the three cases and for the specified flaring categories. Considering the counts of non-flaring and flaring corresponding to this bar chart, there were 4.8 – 5.4 times more flaring times diagnosed than specified. This corresponds to the consistent positive bias seen in Figure 13 that is less than that from the HMFC scheme in spite of the fact that HMFC specifies over six times more flaring times than HEFC. All three cases have their non-zero category maxima in Category 2 (moderate flaring). The number of Category 2 flaring diagnoses exceeds those specified by more than 10 times for all three training sets.

We complete the analysis of the results by showing some sample flaring diagnoses from the MVDA/HEFC development/application scheme for selected image sequences. This demonstrates the kind of variability we saw in the probability-weighted flaring category values as compared with the specified categories. Figures 16 and 17 show a “good” and a “bad” example of flaring diagnosis from the HEFC method operating on the Case 3 training set. In Figure 16(a), we show

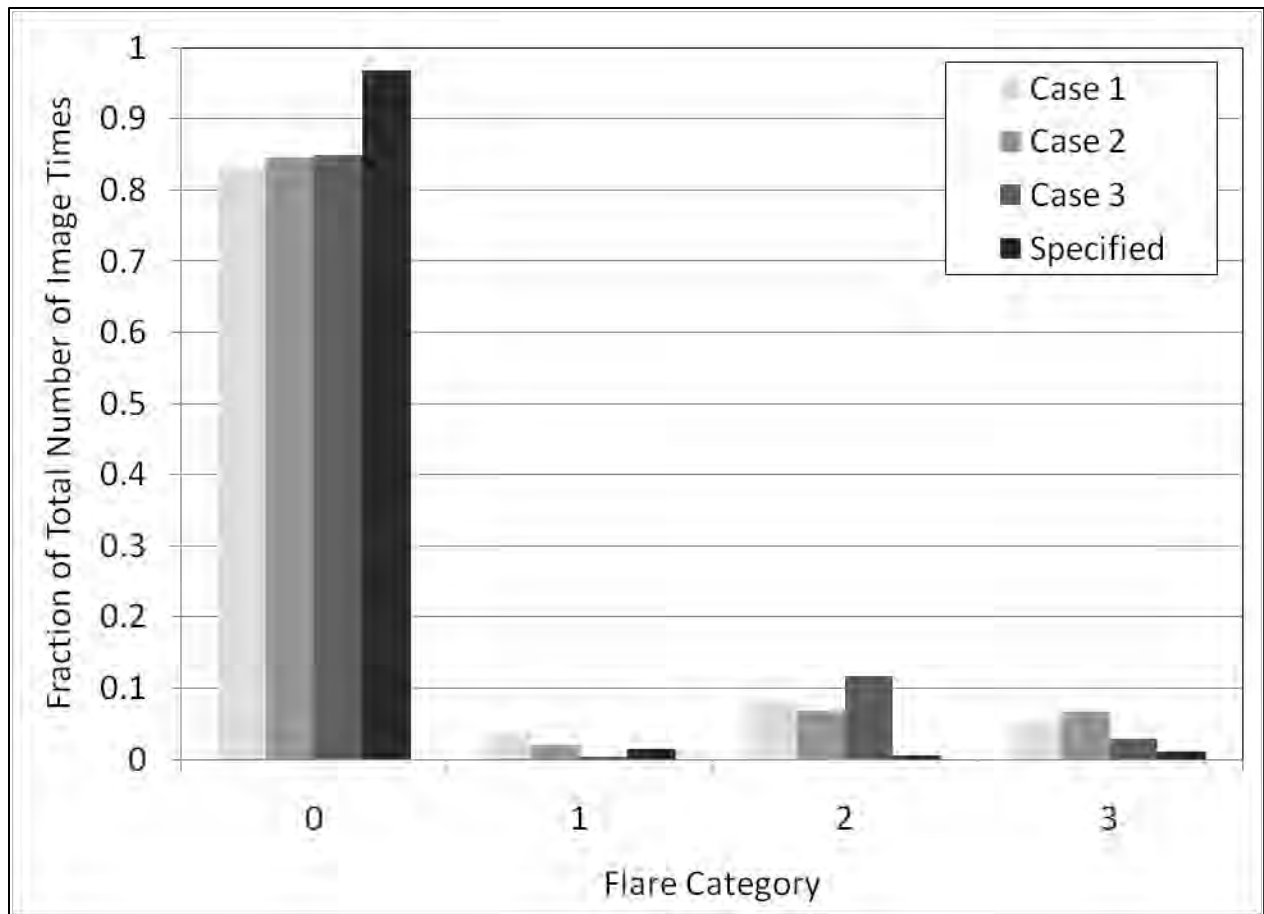


Figure 15. Frequency Distribution from Application of the MVDA/HEFC Algorithm

the result of the Case 3-based diagnosis for the 20031029_10486 image sequence. For this sequence, the diagnosis captures the occurrence of the category 3 flare (given the x-ray background value indicated, this would be an X-class flare) quite well, achieving a probability-weighted flaring category of about 2.5. A note about the flare duration – recall that in the HEFC scheme flaring is only specified during flare rise, so the solid curve does not represent the full duration of actual flaring. The diagnosed flaring has to transition from non-flaring to severe flaring and back to non-flaring, so the diagnosis appears to infer a time span of flaring. Since specification does not represent duration, neither should diagnosis be expected to do so. The diagnosis could at most signal flare magnitude and time of flare rise. Figure 16(b) shows the probability of all four categories for this sequence. It is encouraging to see that the probability of category 0 (non-flaring) is large for most of the specified non-flaring times. This means that the diagnosed uncertainty during these times was small, making the analyst more certain that flares would not be occurring before and after the large flare actually occurred.

The “bad” diagnosis example is shown in Figure 17. Figure 17(a) shows the probability-weighted diagnosed flaring categories vs. the specified categories for the 20030318_10318 sequence. The diagnoses produce probability-weighted diagnosed flaring of category 1 at 17, 20 and 21 UTC, and then nearly category 2 at the end of the sequence. Yet the specification shows no sign of any flaring in the entire duration of the sequence. The plot of the four category

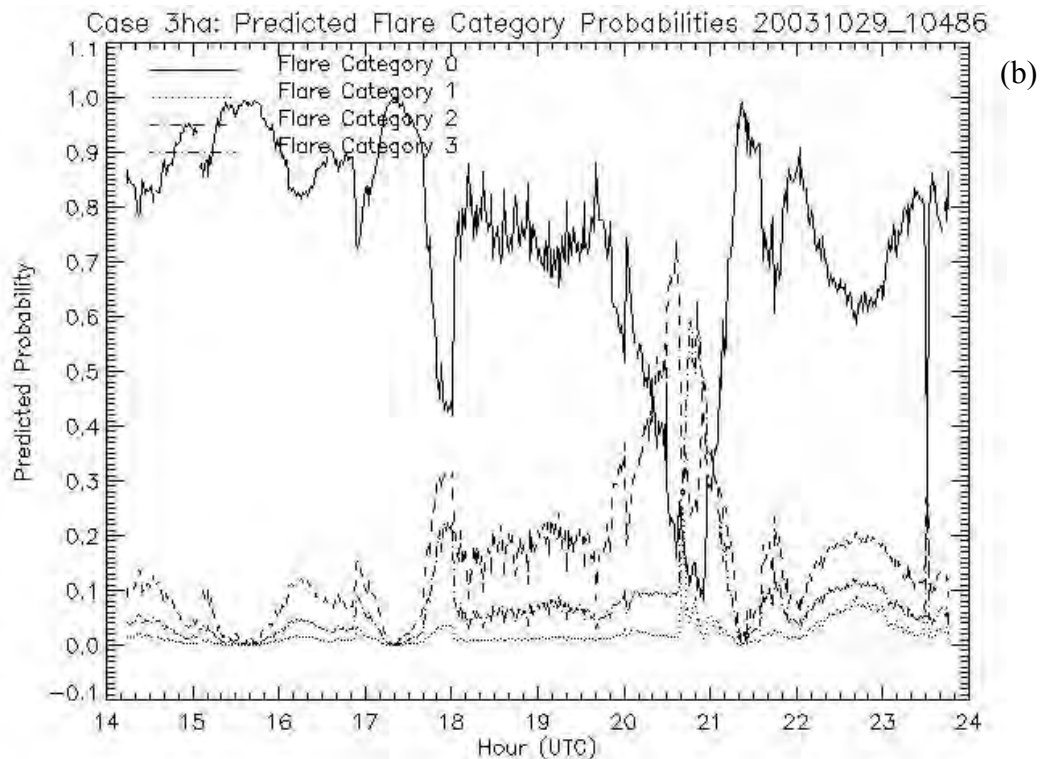
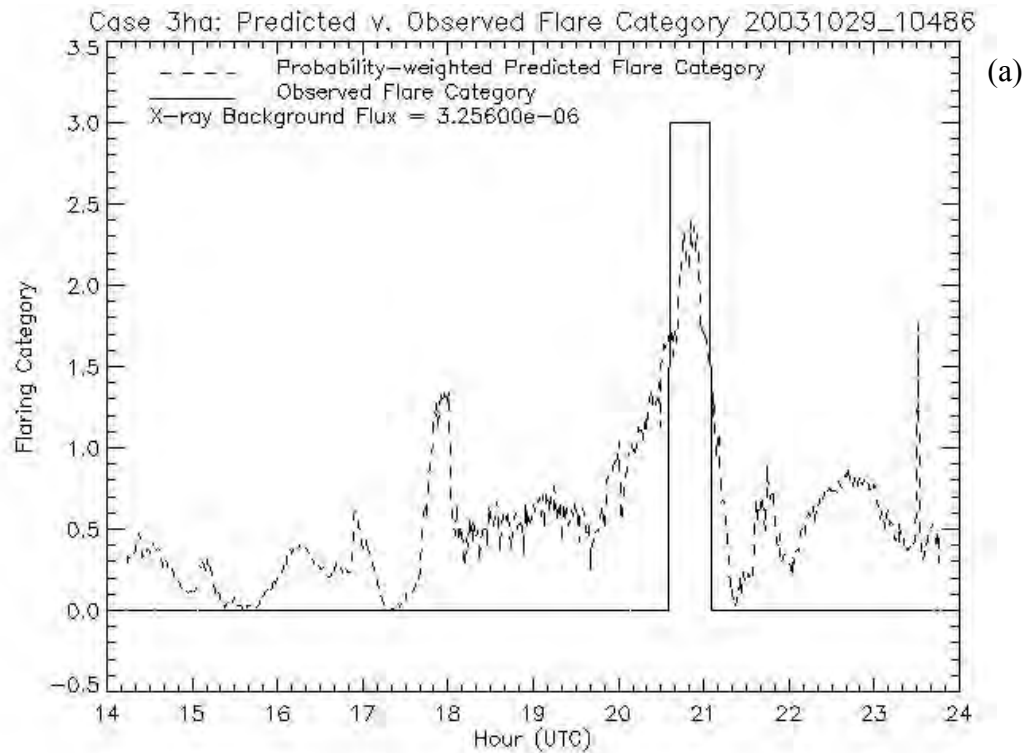


Figure 16. 20031029_10486 Image Sequence (a) C_p (Dashed) and C_o (Solid) and (b) p_g from Application of the MVDA/HEFC Algorithm on Case 3

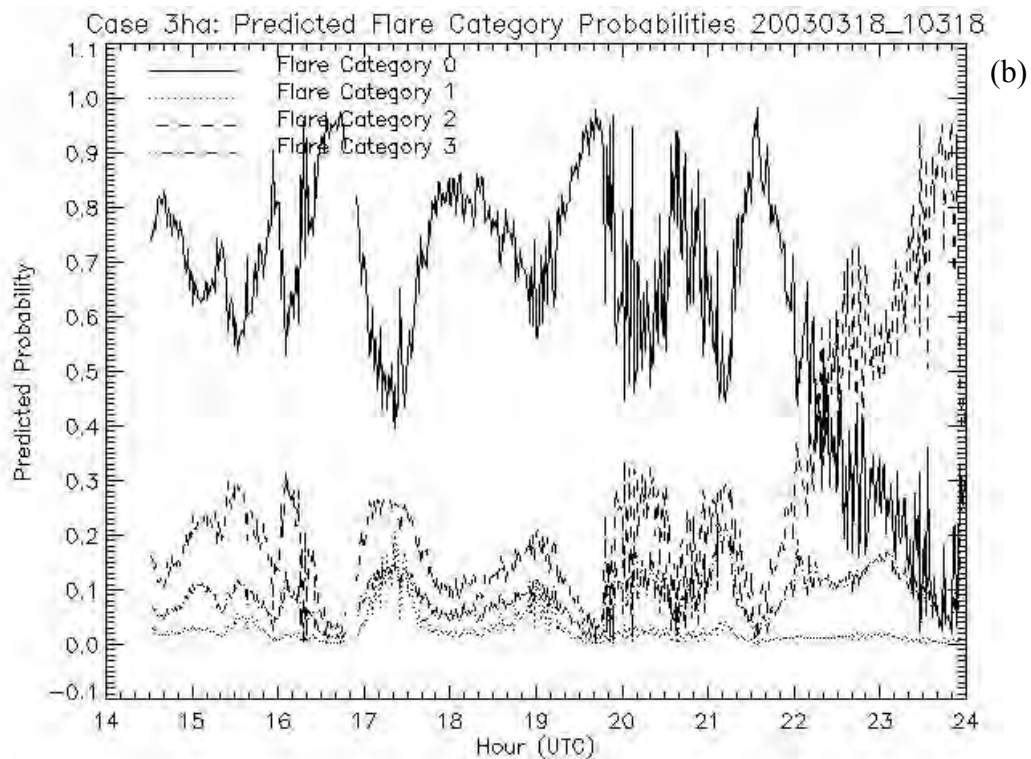
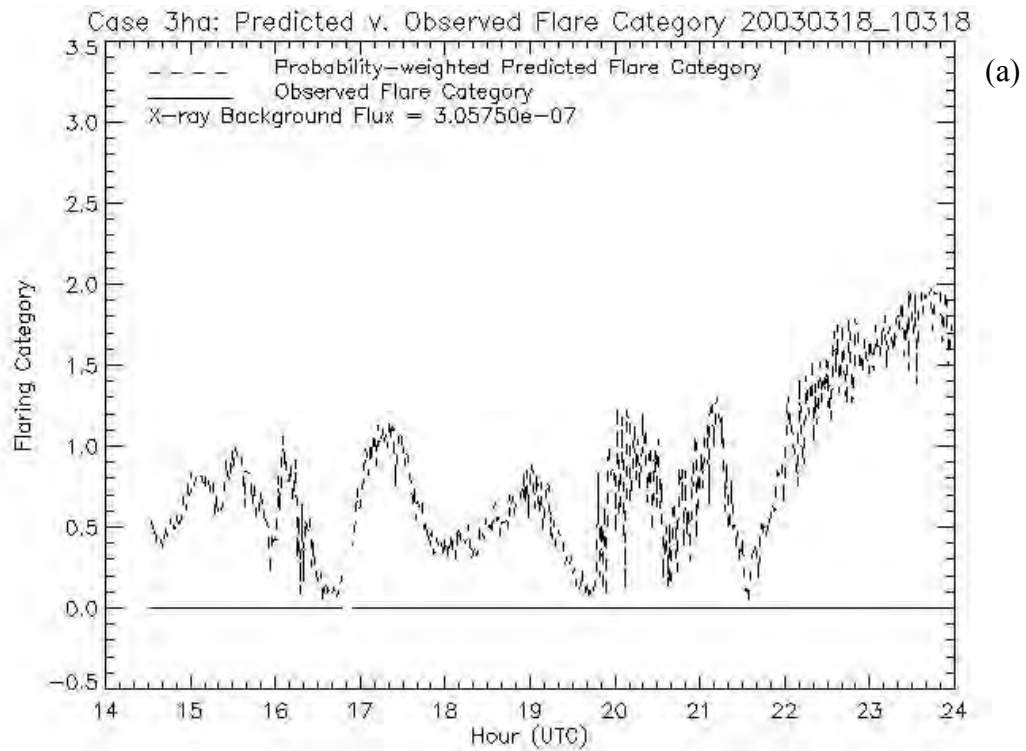


Figure 17. Same as in Figure 16 Except for the 20030318_10318 Image Sequence

probabilities (Figure 17(b)) shows that category 0 has the greatest probability through 22 UTC. However, it dips below 0.5 at the times of the probability-weighted diagnosis of category 1. This suggests a lack of diagnosis certainty that could hamper any decision-making that depends on such guidance. But the clear deficiency in the diagnosis occurs after 22 UTC, when the category 2 probability becomes dominant in place of category 0. In the last hour of the sequence, category 2 probability exceeds 0.9. Here is an example of the diagnostic scheme being apparently “sincerely wrong.” How might this happen? To see if there is any presumptive justification for such a diagnosis, we look at the area-average $H\alpha/H\alpha(\text{bkgd})$ plot and the leading eigenvectors plot for this sequence as shown in Figure 18. Figure 18(a) shows no distinct rise associated with flaring at any time after $H\alpha/H\alpha(\text{bkgd}) > 1$, including after 22 UTC. Figure 18(b) depicts the smooth sinusoidal characteristics of non-flaring (compare with Figure 7(a)) in the leading eigenvectors at all times after 16 UTC. These two pieces of evidence, specific to the sub-region of active region 10318 from which the ISOON $H\alpha$ images were drawn, suggest that if there were flares occurring during the analysis time period of this sequence they were not originating from this sub-region. Now examining the x-ray flux for 20030318 in Figure 19, we see indications of weak (C-class) flaring at several times during the sequence period. The x-ray background flux for this date is $3.06 \times 10^{-7} \text{ W m}^{-2}$, and the x-ray flares apparent in Figure 19 do exceed this level, so they can legitimately be declared C-class flares. However, the convention of the HEFC technique in assigning flaring predictand categories requires that clear evidence of flare rises exist in the area-average $H\alpha$ and the xray-flux time series, and that the eigenvectors show the characteristics of the FLI=1-3 for each such flare. Since neither area-average $H\alpha$ nor the eigenvectors show any such evidence, we must conclude that the C-class flares apparent in the x-ray flux time series originated from elsewhere on the solar disk. Therefore, the rise of category 2 probability resulted from discriminant vectors placing diagnosed points in discriminant space closer to the category 2 group mean and away from the category 0 group mean. What it is about the leading eigenvector elements (the predictors) that would cause this drift is unknown. Such errant behavior in the performance of flaring diagnosis in the MVDA algorithm is a subject for future investigation.

5. SUMMARY AND CONCLUSIONS

In this study, we explore the potential of ISOON $H\alpha$ imagery sequences in diagnosing the presence of flaring in selected solar active regions. The objective of this investigation was to determine if there is sufficient information in the characteristics of the $H\alpha$ image pixels of a sub-region of interest to detect flaring observed in optical and x-ray observations. While ultimately the goal is to be able to anticipate flaring at least several hours before the event, we know that there is no point in attempting prognosis if diagnosis is not attainable.

We used ISOON $H\alpha$ image sequences at one-minute intervals for 46 multi-hour observations of targeted sub-regions of selected solar active regions. Also used were the whole solar disk measurements of x-ray flux as an additional indicator of solar flares. Taking the image times rather than the individual pixel values as the “variables”, we performed a principal component analysis of each the 46 image sequences, producing eigenvalues corresponding to eigenvector

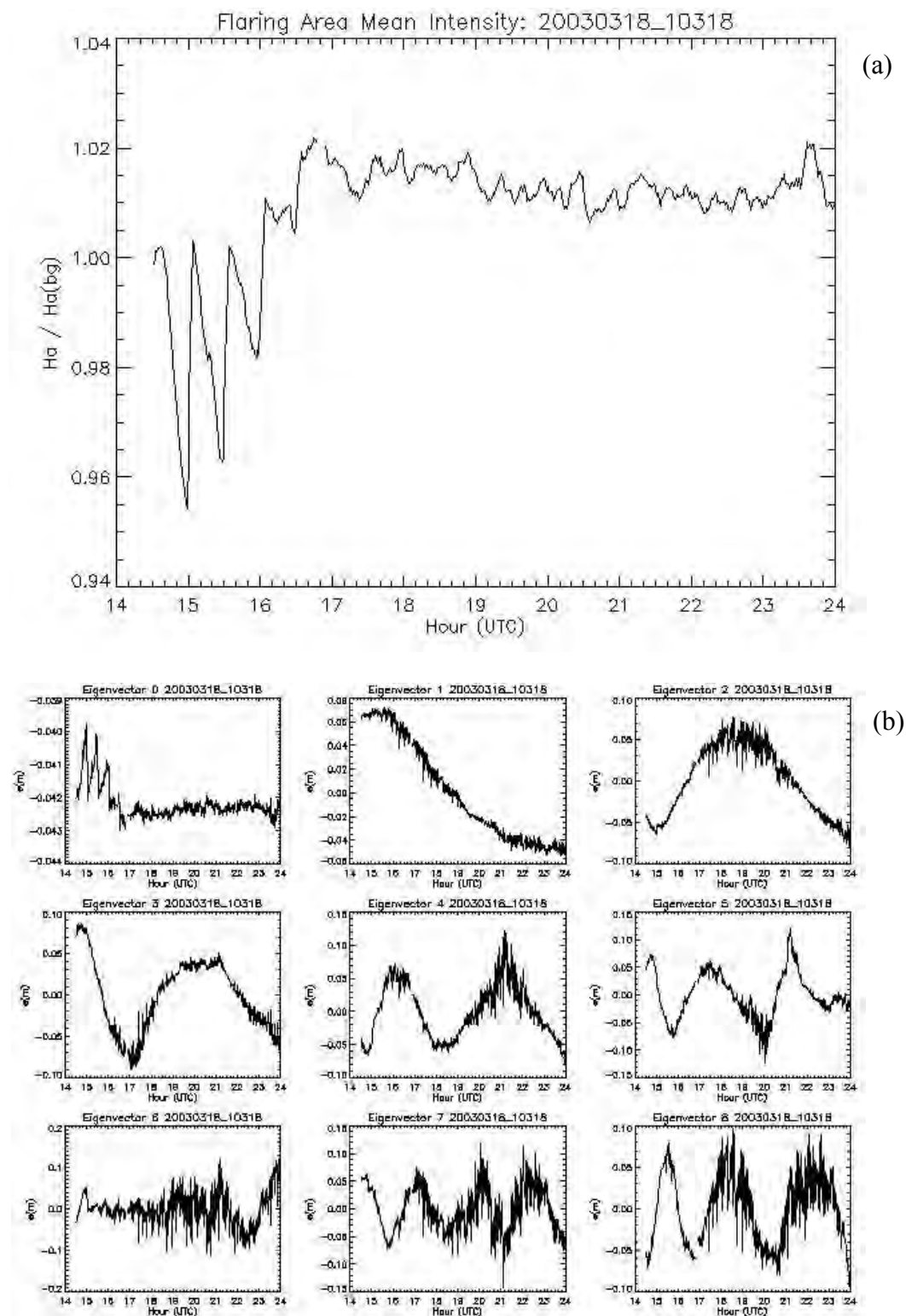


Figure 18. 20030117_10258 Image Sequence (a) $H\alpha/H\alpha(bg)$ and (b) First Nine Eigenvectors

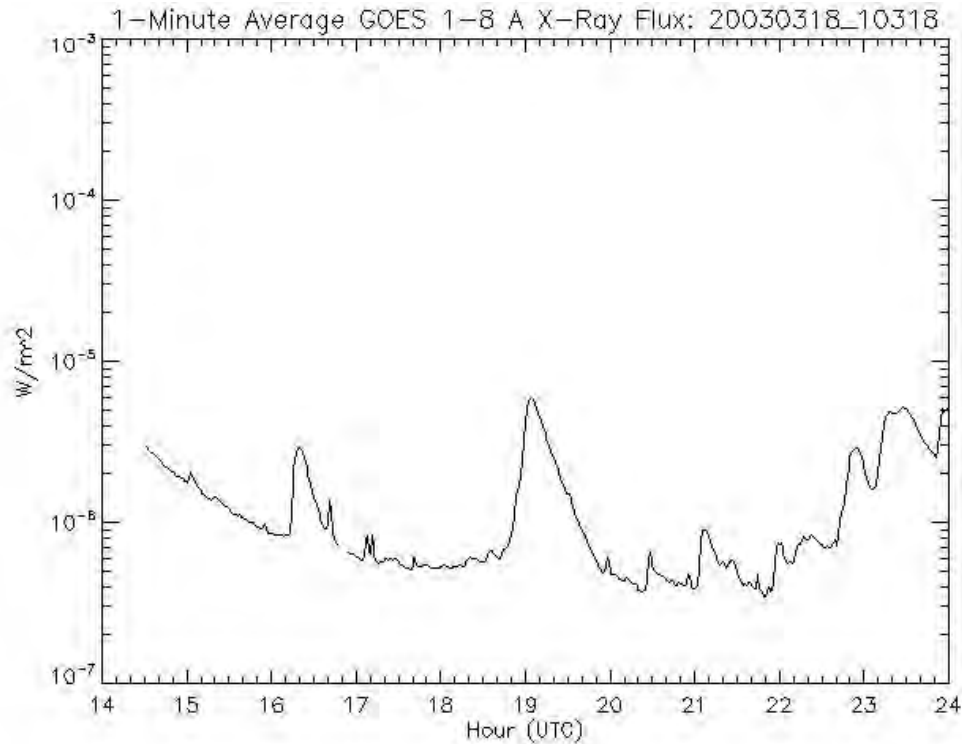


Figure 19. 20030318 Image Sequence GOES One-Minute Average 1-8 Å X-Ray Flux

elements at each image time. Using the eigenvalues to determine the cumulative explained variance of the eigenvector elements, we truncated the number of eigenvector elements well short of the full set while still ensuring that 99.9% of the variance was retained. The eigenvector elements were then used as the predictor variables at each image time in discriminant analysis with accompanying binary (non-flaring/flaring) and later multiple (non-flaring and levels of flaring intensity) categorical predictands. After using training sets of predictor vectors/predictands from all available image times of selected image sequences to develop discriminant vectors, we then applied the discriminant vectors to both dependent (the training sets) and independent predictor vectors. This resulted in probabilities of each binary or multiple predictand category being the correct one, which were compared with the specified (“observed”) category at each image time. Several metrics were deduced from the comparisons as measures of the performance of the diagnostic technique.

We first tried a binary no-flare/flare approach to setting the predictand. We used a simple algorithm in which we set the category to 1 (for flaring) for every image time during a 5% or greater rise in the area-average H α and for subsequent times as long as H α exceeded the pre-rise base value. The x-ray flux also had to be $\geq 1 \times 10^{-6} \text{ Wm}^{-2}$ at all designated flaring times. At all other times, the category was set to 0 (non-flaring). When these predictands were used with the predictor vectors in developing the discriminant vector (singular since there were only two categories), which was then applied to the same image sequences used in development, we found that an excessive number of flaring times were diagnosed. In investigating the cause, we found that in at least one sequence in the training set, there was clear indication of flaring in the eigenvectors used as predictors as well as the x-ray flux. But because there were no rises of the

area-averaged H α of at least five percent above a base value, the predictand was set at non-flaring for the entire sequence. We believe this conflict between the predictors and predictands obscured the distinction between flaring and non-flaring categories in the development of the discriminant vector, causing the discrimination between the two groups to be compromised. We concluded that using an arbitrary threshold in area-averaged H α to define what constitutes flaring when there are separate indications of non-flaring/flaring present in the associated eigenvectors can lead to such conflicts in the information supplied to the discriminant analysis.

Our second method to define flaring category predictands sought to ensure consistency between trends in the predictors and predictands. To do this, we attempted to consider the degree, or relative magnitude of the H α rise and to represent this in multiple flaring categories. We defined corresponding H α categories so that when each was diagnosed it could be easily mapped to a standard x-ray flare class. To do this, we first computed “background” values of H α intensity and x-ray flux for each image sequence, based on the previous day’s observations (when available) and using the NOAA/SWPC “X-ray Bkgd Flux” algorithm. We then computed the correlation of five-point smoothed one-minute area-averaged H α intensity and the reported one-minute interval x-ray flux when each is divided by its respective background (bkgd) value. For the seven image sequences with the highest correlations (all > 0.7), we computed five-minute averages of both area-averaged H α intensity and x-ray flux, dividing H α averages by H α (bkgd) and x-ray averages by their threshold (thrs, full power of 10 greater than the background flux). The best fit linear relationship of the form $[H\alpha]/H\alpha(bkgd)=A + B \log \{[Xr]/Xr(thrs)\}$ was then determined from the five-minute average $[H\alpha]/H\alpha(bkgd)$, $[Xr]/Xr(thrs)$ values. Next, we established four flare categories in terms of $Xr/Xr(thrs)$: Category 0, non-flaring, $Xr/Xr(thrs) < 1$; Category 1, weak flaring, $1 \leq Xr/Xr(thrs) < 10$; Category 2, moderate flaring, $10 \leq Xr/Xr(thrs) < 100$; Category 3, strong flaring, $Xr/Xr(thrs) \geq 100$. Using the best fit relationship, we computed the boundary values of the four categories in terms of $[H\alpha]/H\alpha(bkgd)$. Thus, the category assigned at each image time of the area-averaged H α depends on its ratio to the background H α for that sequence. In this predictand category assignment algorithm, which we called the H α Magnitude Flare Categorization (HMFC) method, only area-averaged H α (and not x-ray flux) was used to set the predictand category at each image time.

Once the multivariate discriminant analysis was conducted on the predictor vectors and associated assigned predictands from a training set of eight image sequences, the resulting discriminant vectors were applied to all 46 image sequences. The dot product of each predictor vector with the three discriminant vectors results in a projection in the three dimensions of discriminant space. The point’s location with respect to the four group means projections of the predictor vectors used in development indicate the likelihood that it falls into each of the respective categories. After computing the category probabilities for each application image time, we derived the probability-weighted diagnosed flaring category. We calculated several performance metrics by comparing this with the specified (by the HMFC scheme), or “observed” category. Using the HMFC technique in the MVDA, we found that the most probable category was only slightly more likely to be the correct one than not. There was considerable overlap among the development discriminant functions in discriminate space, indicating that clear distinction was lacking. We found a substantial positive bias in the probability-weighted diagnosed flaring categories, which was greater for non-flaring than for flaring sequences. This was also true to a lesser extent than for Brier Score (mean square error). Over all sequences,

categories diagnosed as indicating flare occurrence exceeded specified flare event categories by almost a factor of three, indicating an excessive number of false alarms. The number of diagnosed moderate and strong flaring image times was 8 – 10 times the number specified. The fit of the frequency distribution of the most probable flaring category to the frequency distribution of the specified categories was not nearly as good as for the flaring climatology of Solar Cycle 23.

We tried another predictand assignment algorithm, called H α Eigenvector Flare Categorization (HEFC), in an attempt to improve the performance of the MVDA on the ISOON image sequences. A comparison of the leading H α eigenvectors and the x-ray flux time series for the 46 sequences showed similar eigenvector patterns for the sequences that had like x-ray flare characteristics. On the basis of these flare intensity groups, we assigned each H α flare in each image sequence to one of the four flaring categories of non-flaring, and weak, moderate and strong flaring that corresponded to the associated eigenvector pattern/x-ray flare group. After experimenting with several ways of including the image times of the H α flare in the flaring category assignment, we settled on only the rise times as receiving a non-zero flaring category designation. This was done to make the strongest association possible between the predictand flaring indicators and the flaring inflections in the leading eigenvector patterns used as the predictor vectors. However, only when the x-ray flux, and not the area-averaged H α intensity, rose above its respective background level was the non-zero flaring assignment made.

Subsequent MVDA development and application of discriminant vectors using the HEFC technique showed considerable improvement in flaring category diagnosis over the HMFC method when both were compared with categories specified according to their respective algorithms. Using the same training set of image sequences and evaluating the metrics over all 46 image sequences, we found that the Brier Score was about 27% less (better) for HEFC than for HMFC and was less associated with flaring prevalence in a given sequence. Even though there were many fewer (less than 1/6th) specified flaring times in the HEFC than in the HMFC, the HEFC had a 17% smaller positive bias overall than did HMFC. The diagnosed uncertainty from HEFC was about 25% less than from HMFC, indicating that the HEFC method produces more skillful and more certain diagnoses of flaring category.

Given that the performance of HEFC in MVDA was better than that of HMFC, at least for the single training set used, we next assessed the sensitivity of MVDA/HEFC to the choice of training set. Using three training sets of eight image sequences each (with some overlap among the sets), we found similar sequence-to-sequence trends in the assessment of the diagnosis performance. The best of the three cases was 8% better in Brier Score and 5% in bias than the worst. We found that the systematic component of the diagnosis error was greater than the random component, comprising 62 – 64% of the total among the three cases. There was a clear link between greater diagnosis certainty and better skill among the results for three cases. Image sequences that had less positive bias and smaller Brier Score also had a reduced diagnosed uncertainty. This seems logical in light of the fact that diagnoses of similar probability among the four categories are more likely to result in an incorrect category choice than are diagnoses where one category has a predominant probability. Diagnosis certainty was 9% greater for the best case than for the worst case. Finally, unlike the single HMFC case, the frequency distribution fit of the diagnosed categories to the frequency distribution of specified categories in the three HEFC

cases were competitive with the fit for the Solar Cycle 23 climatological flaring category frequency distribution. Despite its better frequency distribution fit than the HMFC method, HEFC still diagnosed about five times more occurrences of flaring than were specified. The over-diagnosis factor exceeded the 2 – 3 factor of HMFC due to the fact that HEFC specified 16% of the number of flaring times than did HMFC.

From the best of the three cases when the MVDA/HEFC technique was employed, we examined plots of sequences of both “good” and “bad” diagnosis performance. In the sequence that showed desirable skill, we noted that the certainty of the non-flaring diagnosis was large both before and after the flare diagnosis, which was correct in both magnitude and timing. Thus there was clear distinction between times of non-flaring and flaring in this sequence. In the “bad” performance sequence, incidences of moderate flaring were diagnosed with great certainty, whereas no flaring was specified in the sequence. Upon investigation, there was no evidence in the H α intensity, either the area-area average or the eigenvectors, of flaring of any kind from the active region sub-region from which the ISOON images were taken. The cause of the erroneous diagnosis of flaring at the end of the sequence period is unknown, and is a subject for further investigation.

We conclude from these results that the MVDA/HEFC technique, with possible further modifications, may have the potential to perform flaring diagnoses of appreciable skill and certainty based on adequate training sets of H α image sequences and coincident x-ray flux time series. There are two future directions that we desire to take in this work. The first is to increase the number of H α image sequences and x-ray time series available for development and application. These should include approximately equal number of sequences that are non-flaring and flaring. Increasing the development and application sample should allow us to more definitively determine the sensitivity of the diagnosis procedure to the choice of image sequences in the training set. The second is to consider additional types of observations of solar conditions known to be associated with flaring, notably the magnetic fields of the photosphere. For example, fast cadence and highly resolved magnetograms could complement the optical data used to develop the predictor set. Ultimately, we wish to determine if such quick cadence observations can be used in conjunction with MVDA to anticipate the occurrence of flaring with suitable skill and confidence. We hope to conduct such an investigation in the future.

REFERENCES

- [1] Norquist, D. C., “An Analysis of the Sunspot Groups and Flares of Solar Cycle 23,” *Solar Phys.*, **269**, 2011, pp. 111-127.
- [2] Contarino, L., Zuccarello, F., Romano, P., Spadaro, D., Guglielmino, S. L., and Battiato, V., “Flare Forecasting Based on Sunspot-Groups Characteristics,” *Acta Geophys.*, **57**, 2009, pp. 52-63.
- [3] Falconer, D., Barghouty, A. F., Khazanov, I., and Moore, R., “A Tool for Empirical Forecasting of Major Flares, Coronal Mass Ejections, and Solar Particle Events from a Proxy of Active-Region Free Magnetic Energy,” *Space Weather*, **9**, S04003, 2011, p. 12.
- [4] Reinard, A. A., Henthorn, J., Komm, R., and Hill, F., “Evidence That Temporal Changes in Solar Subsurface Helicity Precede Active Region Flaring,” *Astrophys. J. Lett.*, **710**, 2010, pp. L121-L125.
- [5] Neidig, D. et al., “The USAF Improved Solar Observing Optical Network (ISOON) and Its Impact on Solar Synoptic Data Bases,” In *Synoptic Solar Physics*, **140**, Astronomical Society of the Pacific Conference Series, K. S. Balasubramaniam, J. W. Harvey and D. M. Rabin, eds., 1998, pp. 519-528.
- [6] Wilks, D. S., **Statistical Methods in the Atmospheric Sciences, Second Edition**, Academic Press, Burlington, MA, 2006, p. 627.
- [7] Norquist, D. C., “Cloud Predictions Diagnosed from Global Weather Model Forecasts,” *Mon. Wea. Rev.*, **128**, 2000, pp. 3538-3555.

APPENDIX A

LINEAR DISCRIMINATION AND CLASSIFICATION OF SAMPLE DATA IN TWO GROUPS (SUMMARIZED FROM SECTIONS 13.1 AND 13.2.1 OF WILKS [6])

A1. Discrimination and Classification

Consider a K -dimensional vector $\mathbf{x}$ of observations of two or more variables sampled from a larger population of measurements that belong to two or more groups that are distinct from each other in some identifiable manner. The groups might represent distinguishable geographical regions, months of year, climatological realms, etc. What is known about the sample in advance are: (1) the number G of distinct groups, (2) each measurement vector $\mathbf{x}$ belongs to one and only one group, (3) that all measurement vectors belong to one group, none are excluded, (4) a set of training data is available, a subset of the sample of observations, in which the group membership of each of the data vectors $\mathbf{x}_i$, $i=1, \dots, n$ is known.

Discrimination is the process of determining functional forms of the training data that best distinguish the features of the groups from each other. Classification is the act of applying these functional forms to an independent subset of the sample of observations $\mathbf{y}_j$, $j=1, \dots, m$ in order to assign them to their respective groups. The probability $p_g(\mathbf{y}_j)$, $g=1, \dots, G$ that each observation vector $\mathbf{y}_j$ belongs to each group g can then be determined.

If the groups to which the current measurements $\mathbf{y}_j$ are to be assigned refer to future event categories, classification is used as a means of assigning each measurement vector to a category that is predicted to occur. In this case discrimination and classification can be used as a forecasting tool, and the probabilities $p_g(\mathbf{y}_j)$ of future occurrence of the G event categories can be estimated.

A2. Linear Discrimination for $G = 2$

Suppose we wish to develop the functional form that will distinguish between two groups based on an available set of n K -dimensional vectors of observations $\mathbf{x}$ for which the group membership of each observation vector is known. The K dimensions of the vectors may be, for example, different variables (e.g., wind speed, temperature, etc.), while n might be, for example, the number of samples in a time series of observations of those variables. The elements of each vector $\mathbf{x}_i$ are $x_{i,1}, x_{i,2}, \dots, x_{i,K}$, which are the observed values of the variables for the i^{th} observation.

Let n_1 be the number of observations in group $g=1$ and n_2 the number in group $g=2$. Then there are two sets of training vectors $\mathbf{x}^g$: $\mathbf{x}^1_i$, $i=1, \dots, n_1$, and $\mathbf{x}^2_i$, $i=1, \dots, n_2$. The goal of linear discrimination is to determine the vector of coefficients $\mathbf{a}$ of the linear combination of K elements $\mathbf{a}^T \mathbf{x}$, the discriminant function, that optimizes the classification of independent K -dimensional observation vectors $\mathbf{y}_j$ into either $g=1$ or $g=2$. R. A. Fisher devised an approach to this problem by determining the vector of coefficients $\mathbf{a}$ that is directed in K -dimensional space

so as to maximize the separation of the means of the observations when the observation vectors $\mathbf{x}$ are projected onto $\mathbf{a}$. He showed that this criterion is equivalent to finding $\mathbf{a}$ to maximize the expression

$$\frac{(\mathbf{a}^T \bar{\mathbf{x}}^1 - \mathbf{a}^T \bar{\mathbf{x}}^2)^2}{\mathbf{a}^T [S_{pool}] \mathbf{a}}. \quad [\text{A-1}]$$

The two mean vectors $\bar{\mathbf{x}}^1$ and $\bar{\mathbf{x}}^2$ are calculated separately for each group, as the vector of means of the observations of each element (variable):

$$\bar{\mathbf{x}}^g = \frac{1}{n_g} \begin{bmatrix} \sum_{i=1}^{n_g} x_{i,1}^g \\ \sum_{i=1}^{n_g} x_{i,2}^g \\ \vdots \\ \sum_{i=1}^{n_g} x_{i,K}^g \end{bmatrix}; g=1,2. \quad [\text{A-2}]$$

The common sample covariance matrix for the entire sample of training data (i.e., over both groups) is $[S_{pool}]$, which is given by

$$[S_{pool}] = \frac{n_1-1}{n_1+n_2-2} [S^1] + \frac{n_2-1}{n_1+n_2-2} [S^2], \quad [\text{A-3}]$$

the weighted average of the two sample covariance matrices of the respective groups, the elements of which are

$$S_{k,l}^g = \frac{1}{n_g-1} \sum_{i=1}^{n_g} (x_{i,k}^g - \bar{x}_k^g)(x_{i,l}^g - \bar{x}_l^g); k,l = 1, \dots, K. \quad [\text{A-4}]$$

Note that the $\bar{x}_k^g, \bar{x}_l^g$ are the means of the observations of each of the combinations of variables k, l . The sample covariance matrices of the respective groups are then

$$[S^g] = \begin{bmatrix} S_{1,1}^g & S_{1,2}^g & \dots & S_{1,l}^g & \dots & S_{1,K}^g \\ \vdots & \vdots & & \vdots & & \vdots \\ S_{k,1}^g & S_{k,2}^g & \dots & S_{k,l}^g & \dots & S_{k,K}^g \\ \vdots & \vdots & & \vdots & & \vdots \\ S_{K,1}^g & S_{K,2}^g & \dots & S_{K,l}^g & \dots & S_{K,K}^g \end{bmatrix}. \quad [\text{A-5}]$$

Note that since $S_{k,l}^g = S_{l,k}^g$, $[S^g]$ will be a symmetric matrix. Thus, calculating below and to the diagonal gives the corresponding elements above the diagonal.

All that is assumed in the formulation to this point is that the covariance matrices of the populations of the two groups from which the sample of training data were drawn are equal. Maximizing the first expression leads to the vector $\mathbf{a}$ of the coefficients in Fisher's linear discriminant function, represented by $\delta_1 = \mathbf{a}^T \mathbf{x}$. Application of this function to independent observations $\mathbf{y}$ of the same K -dimensional parameters results in two groups of scalar values with

different means and equal variances distributed along the $\mathbf{a}$ axis. The vector $\mathbf{a}$ locating the direction of maximum separation between the two groups is given by

$$\mathbf{a} = [S_{pool}]^{-1}(\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2). \quad [\text{A-6}]$$

Then Fisher's linear discriminant function is

$$\delta_1 = \mathbf{a}^T \mathbf{x} = \left[[S_{pool}]^{-1}(\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2) \right]^T \mathbf{x} = (\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2)^T [S_{pool}]^{-1} \mathbf{x}. \quad [\text{A-7}]$$

Applying $\mathbf{a}$ to the difference between the two groups' means maximizes the distance between their projections on $\mathbf{a}$, which is the Mahalanobis distance D^2 between the groups:

$$D^2 = \mathbf{a}^T (\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2) = (\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2)^T [S_{pool}]^{-1} (\bar{\mathbf{x}}^1 - \bar{\mathbf{x}}^2). \quad [\text{A-8}]$$

A3. Classification using Fisher's Linear Discriminant Function

Fisher's linear discriminant function can be used to classify any number M of independent observations $\mathbf{y}$ of the same K -dimensional parameters as in the training vectors $\mathbf{x}$:

$$\delta_1(\mathbf{y}_i) = \mathbf{a}^T \mathbf{y}_i ; i = 1, \dots, M. \quad [\text{A-9}]$$

This scalar quantity, the dot product of $\mathbf{a}^T$ and the independent observation vectors, is the projection of the vectors onto the direction of maximum separation, and represents a new variable. The observation vector is assigned to $g=1$ if $\delta_1(\mathbf{y}_i)$ is closer to the projection of the Group 1 mean onto the direction $\mathbf{a}$, or to Group 2 if it's closer to the Group 2 mean projection. The way this is determined is by computing the midpoint between the means of the two groups, which is the projection of the average of the two mean training vectors onto $\mathbf{a}$:

$$\hat{m} = \frac{1}{2}(\mathbf{a}^T \bar{\mathbf{x}}^1 + \mathbf{a}^T \bar{\mathbf{x}}^2) = \frac{1}{2} \mathbf{a}^T (\bar{\mathbf{x}}^1 + \bar{\mathbf{x}}^2). \quad [\text{A-10}]$$

The independent observation vector $\mathbf{y}_i$ is assigned to one of the two groups as follows:

$$\begin{aligned} \mathbf{Y}_i & \text{ assigned to Group 1 if } \delta_1(\mathbf{y}_i) \geq \hat{m}, \\ \mathbf{Y}_i & \text{ assigned to Group 2 if } \delta_1(\mathbf{y}_i) < \hat{m}. \end{aligned}$$

A probability of the observation lying within Group 1 can be estimated by:

$$p_1(\mathbf{y}_i) = 1 - \frac{\mathbf{a}^T \bar{\mathbf{x}}^1 - \delta_1(\mathbf{y}_i)}{D^2} \quad [\text{A-11}]$$

so that for $\delta_1(\mathbf{y}_i) > \mathbf{a}^T \bar{\mathbf{x}}^1$, $p_1(\mathbf{y}_i) > 1$ which can be set to 1 (100% probability of being in Group 1). For $\delta_1(\mathbf{y}_i) < \mathbf{a}^T \bar{\mathbf{x}}^2$, $p_1(\mathbf{y}_i) < 0$ which can be set to zero (0% probability of being in Group 1, or 100% probability of being in Group 2).

APPENDIX B

FISHER'S LINEAR DISCRIMINANT FOR MULTIPLE GROUPS (SUMMARIZED FROM SECTION 13.3.1 OF WILKS [6])

- This generalization of Fisher's Linear Discriminant for two groups is called "multiple discriminant analysis."
- Problem: allocate n K -dimensional data vectors $\mathbf{x}$ to one of $G > 2$ groups.
- Basis vectors for group discrimination are called discriminant vectors, of which there are $J = \min(G-1, K)$, denoted $\mathbf{a}_j, j=1, \dots, J$, which are K -dimensional.
- The projection of the data onto these vectors gives J discriminant functions

$$\delta_j = \mathbf{a}_j^T \mathbf{x}; j=1, \dots, J. \quad [\text{B-1}]$$

- The discriminant vectors are determined from G data matrices of the training data, each dimensioned $n_g \times K$:

$$\begin{bmatrix} x_{1,1} & \dots & x_{1,K} \\ \vdots & & \vdots \\ x_{2,1} & \dots & x_{2,K} \\ \vdots & & \vdots \\ x_{n_g,1} & \dots & x_{n_g,K} \end{bmatrix}; g=1, \dots, G. \quad [\text{B-2}]$$

- The rows of the data matrices are the training set of observation vectors $\mathbf{x}$, each of which is classified by group g according to a categorical assignment that ensures sufficient distinction or separation among the groups. n_g is the number of observation vectors assigned to group g .
- A $K \times K$ sample covariance matrix is computed from each data group:

$$[S_g] = \begin{bmatrix} S_{1,1}^g & S_{1,2}^g & \dots & S_{1,l}^g & \dots & S_{1,K}^g \\ \vdots & \vdots & & \vdots & & \vdots \\ S_{k,1}^g & S_{k,2}^g & \dots & S_{k,l}^g & \dots & S_{k,K}^g \\ \vdots & \vdots & & \vdots & & \vdots \\ S_{K,1}^g & S_{K,2}^g & \dots & S_{K,l}^g & \dots & S_{K,K}^g \end{bmatrix}; g=1, \dots, G \quad [\text{B-3}]$$

In which the matrix elements are the covariances of the observation vectors for each group:

$$S_{k,l}^g = \frac{1}{n_g-1} \sum_{i=1}^{n_g} (x_{i,k}^g - \bar{x}_k^g)(x_{i,l}^g - \bar{x}_l^g); k, l = 1, \dots, K. \quad [\text{B-4}]$$

Note that the group mean vectors are given by

$$\bar{\mathbf{x}}^g = \frac{1}{n_g} \begin{bmatrix} \sum_{i=1}^{n_g} x_{i,1}^g \\ \sum_{i=1}^{n_g} x_{i,2}^g \\ \vdots \\ \sum_{i=1}^{n_g} x_{i,K}^g \end{bmatrix}; g=1, \dots, G \quad [\text{B-5}]$$

and that the $\bar{x}_k^g, \bar{x}_l^g$ are the elements of the group mean vectors that are the means of the observations of each of the combinations of variables k, l .

- The weighted average of the G sample covariance matrices gives a pooled estimate of the common covariance matrix:

$$[S_{pool}] = \frac{1}{n-G} \sum_{g=1}^G (n_g - 1) [S_g], \quad [B-6]$$

a $K \times K$ matrix, also called the within-group covariance matrix, where

$$n = \sum_{g=1}^G n_g. \quad [B-7]$$

- The between-groups covariance matrix is also required for computation of the multiple discriminant functions:

$$[S_B] = \frac{1}{G-1} \sum_{g=1}^G (\bar{x}^g - \bar{x})(\bar{x}^g - \bar{x})^T \quad [B-8]$$

where

$$\bar{x} = \frac{1}{n} \sum_{g=1}^G n_g \bar{x}^g \quad [B-9]$$

is the grand mean, or overall vector mean of all of the n observation vectors. The between groups covariance matrix can also be expressed as

$$[S_B] = \frac{1}{G-1} \sum_{g=1}^G \begin{bmatrix} (\bar{x}_1^g - \bar{x}_{1,\cdot})(\bar{x}_1^g - \bar{x}_{1,\cdot}) & (\bar{x}_1^g - \bar{x}_{1,\cdot})(\bar{x}_2^g - \bar{x}_{2,\cdot}) & \dots & (\bar{x}_1^g - \bar{x}_{1,\cdot})(\bar{x}_K^g - \bar{x}_{K,\cdot}) \\ (\bar{x}_2^g - \bar{x}_{2,\cdot})(\bar{x}_1^g - \bar{x}_{1,\cdot}) & (\bar{x}_2^g - \bar{x}_{2,\cdot})(\bar{x}_2^g - \bar{x}_{2,\cdot}) & \dots & (\bar{x}_2^g - \bar{x}_{2,\cdot})(\bar{x}_K^g - \bar{x}_{K,\cdot}) \\ \vdots & \vdots & \ddots & \vdots \\ (\bar{x}_K^g - \bar{x}_{K,\cdot})(\bar{x}_1^g - \bar{x}_{1,\cdot}) & (\bar{x}_K^g - \bar{x}_{K,\cdot})(\bar{x}_2^g - \bar{x}_{2,\cdot}) & \dots & (\bar{x}_K^g - \bar{x}_{K,\cdot})(\bar{x}_K^g - \bar{x}_{K,\cdot}) \end{bmatrix} \quad [B-10]$$

- The $J = \min(G-1, K)$ K -dimensional discriminant vectors $\mathbf{a}_j, j=1, \dots, J$ are derived from the first J eigenvectors (corresponding to the non-zero eigenvalues) of the $K \times K$ matrix

$$[S_{pool}]^{-1} [S_B]. \quad [B-11]$$

- The $\mathbf{a}_j$ are the scaled (to unit magnitude) eigenvectors $\mathbf{e}_j$. However, if the eigenvectors are not of unit magnitude, the discriminant vectors $\mathbf{a}_j$ can be computed from the eigenvectors using the relation

$$\mathbf{a}_j = \frac{\mathbf{e}_j}{(\mathbf{e}_j^T [S_{pool}] \mathbf{e}_j)^{1/2}}. \quad [B-12]$$

-The first discriminant vector $\mathbf{a}_1$, associated with the largest eigenvalue, makes the largest contribution to separating the group means. Conversely, the J^{th} eigenvector $\mathbf{a}_J$, is associated with the smallest of the first J eigenvalues, and makes the smallest contribution.

- Once the discriminant vectors $\mathbf{a}_j$ are derived from the training set, they can be used to assign independent observation vectors $\mathbf{x}_0$ one of the G groups. This is done by determining the group whose group mean is closest to the observation in discriminant space. To do this, compute the squared distances between the observation vector and each of the group means in discriminant space:

$$D_g^2 = \sum_{j=1}^J [\mathbf{a}_j^T (\mathbf{x}_o - \bar{\mathbf{x}}^g)]^2; g=1, \dots, G \quad [\text{B-13}]$$

$$= \sum_{j=1}^J \{ [a_{1,j} \quad a_{2,j} \quad \dots \quad a_{K,j}] \begin{bmatrix} x_{1,o} - \bar{x}_1^g \\ x_{2,o} - \bar{x}_2^g \\ \vdots \\ x_{K,o} - \bar{x}_K^g \end{bmatrix} \}^2; g=1, \dots, G. \quad [\text{B-14}]$$

Then assign the observation vector $\mathbf{x}_o$ to the group with the smallest value of D^2 , that is, to the group with the closest group mean.

(This ends the discussion of the multiple discriminant analysis algorithm of Section 13.3.1 in Wilks [6]).

- An estimate of the probability that each group g is the correct group to which to assign the observation vector can be derived as follows. First, compute the fraction of each group's squared distance to the sum of the squared distances of all groups:

$$f_g = \frac{D_g^2}{\sum_{g=1}^G D_g^2}. \quad [\text{B-15}]$$

The probabilities for the G groups must sum to one:

$$\sum_{g=1}^G p_g = 1. \quad [\text{B-16}]$$

We seek a probability of group membership p_g that is inversely proportional to the fractional distance of the observation from the respective group mean:

$$p_g = \frac{C}{f_g} \quad [\text{B-17}]$$

where C is an arbitrary constant. Using the b^{th} group mean (that is, the group mean with the smallest f_g) as the basis, we can relate the other group probabilities to the basis group probability p_b by

$$p_l f_l = p_b f_b \quad [\text{B-18}]$$

for all $l=1, \dots, b-1, b+1, \dots, G$ groups. Inserting all p_l from these relationships into the summation of probabilities to one gives

$$p_b + \sum_{l=1}^{b-1} \left(\frac{f_b}{f_l} p_b \right) + \sum_{l=b+1}^G \left(\frac{f_b}{f_l} p_b \right) = 1 \quad [\text{B-19}]$$

or, solving for p_b ,

$$p_b = \left[1 + f_b \left(\sum_{l=1}^{b-1} \frac{1}{f_l} + \sum_{l=b+1}^G \frac{1}{f_l} \right) \right]^{-1}. \quad [\text{B-20}]$$

Then the other group probabilities can be obtained from their relationships to the basis group probability. For example, for $G = 4$ and $f_1 = 0.1, f_2 = 0.2, f_3 = 0.3$, and $f_4 = 0.4$ (note that the

fractional distances always sum to one) so that $b = 1$, we get $p_1 = 0.48$, $p_2 = 0.24$, $p_3 = 0.16$, and $p_4 = 0.12$, which sum to one as required.

DISTRIBUTION LIST

DTIC/OCP 8725 John J. Kingman Rd, Suite 0944 Ft Belvoir, VA 22060-6218	1 cy
AFRL/RVIL Kirtland AFB, NM 87117-5776	2 cys
Official Record Copy AFRL/RVBXS/Donald Norquist	1 cy