AFRL-RV-PS-
TR-2012-0043

AFRL-RV-PS-
TR-2012-0043

VALIDATION OF THE NOAA SPACE WEATHER PREDICTION CENTER'S SOLAR FLARE FORECASTING LOOK-UP TABLES AND FORECASTER ISSUED PROBABILITIES (PREPRINT)

Misty D. Crown

30 March 2012

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 AIR FORCE BASE, NM 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-TR-2012-0043 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//	//signed//
Misty D. Crown Project Manager, RVBX	Joel B. Mozer Chief, RVB

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

1. REPORT DATE ***(DD-MM-YYYY)***
30-03-2012

2. REPORT TYPE
Interim Report

3. DATES COVERED *(From - To)*
01 Apr 2007 – 01 Dec 2011

4. TITLE AND SUBTITLE
Validation of the NOAA Space Weather Prediction Center's Solar Flare Forecasting Look-Up Tables and Forecaster Issued Probabilities (Preprint)

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63401F

6. AUTHOR(S)
Misty D. Crown

5d. PROJECT NUMBER
5021

5e. TASK NUMBER
PPM00004258

5f. WORK UNIT NUMBER
EF004416

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-TR-2012-0043

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. (377ABW-2011-1625 dtd 01 Dec 2011)

13. SUPPLEMENTARY NOTES
Accepted for publication in Space Weather Journal, Projected publication date 12 April 2012, American Geophysical Union, 2000 Florida Avenue N.W., Washington, DC 20009-1277. Government Purpose Rights.

14. ABSTRACT
This paper provides an assessment of the operational solar flare look-up table currently in use at the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) during solar cycle 23 (May 1996 – December 2008). To assess the value of human interaction, a validation of subjective flare probability forecasts was conducted and compared to the results obtained from the look-up tables. Probabilistic flare forecasts were evaluated using the Brier Skill Score, then discretized and entered into contingency tables from which a variety of verification measures were calculated. These scores and statistics will be used as the basis for future evaluation of models to be presented to the operational community.

15. SUBJECT TERMS
Forecasting, CME, Solar Flare

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Unlimited	18	Misty Crown
					19b. TELEPHONE NUMBER *(include area code)*

This page intentionally left blank.

Table of Contents

This page is intentionally left blank.

1. Introduction

Solar active region flare forecasting has been a challenging task since space environmental forecasting efforts began in 1962 [Lanzerotti et al., 2006]. Active region flare forecasting has evolved very little, even with the advent of new instrumentation that allows an increasing better view of solar features. Operational forecasters at NOAA SWPC still rely on the use of a look-up table coupled with climatology, persistence, and forecaster know-how or expertise to create the daily probabilities. One way to evaluate numerical models is to compare them to operational forecasts. The validation study presented here is conducted to provide an operational standard (or baseline) for comparison.

The forecasts used in this effort were made available by staff at the Space Weather Prediction Center. The dataset provided includes the subjective forecast for each active region visible daily for 24-hour, 48-hour, and 72-hour intervals, as well as forecaster name. For the study presented here only the 24-hour forecasts are validated. In the situation where the forecaster's name was not listed in the dataset or a forecast was not in the dataset (approximately 125 active region flare forecasts of the over 31,000 available forecasts), this information was gathered from the daily synoptic drawing available on the National Geophysical Data Center's website.

2. Forecast Validation Concepts

In order to do a proper validation of forecasting methods, relevant and robust metrics need to be selected. The simplest measure to use is the Brier Score (BS) for probabilistic forecasts:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2 \qquad [1]$$

where values found in P_i are the forecaster issued probabilities or the probabilities found in the look-up table with values between 0 and 1.0. The actual observation value found in O_i is either 1 or 0, O_i = 1 if an event was observed, and O_i = 0 if an event was not observed. The Brier Score ranges from 0 to 1, and the closer the result approaches 0 the more accurate the forecast [Brier, 1950].

Validation is also performed here using measures derived from contingency tables. Table 1, shows a basic two-by-two contingency table. The value found in A is the total of where both the event was forecasted and observed. The value in B is the count where the event was forecast, but not observed. C contains the count when the event was not forecasted, however was observed. Lastly D includes the sum of where the event was neither forecasted, nor observed. The appendix lists the various scores that will be evaluated from the contingency tables.

Typically, a forecasted event or "yes" forecast in the contingency table is associated with a forecast probability of >0.50. However, in the prediction of solar flare probabilities, it is necessary to adjust the probability required for a "yes" forecast. According to Wilks (2006), there are two methods that are most widely accepted operationally. Both methods require building two-by-two contingency tables at user-defined increments of probabilities. Once the contingency tables are constructed the user computes values both for the bias and the critical success index (CSI) for each of the two-by-two contingency tables. To select the probability threshold characterizing a “yes” event, the user must choose the contingency table in which either the bias is closest to unity or the critical success index is maximized. The critical success index approach is used here, since this method is most often used in the meteorological community.

3. Validation of Subjective Forecasts

Before a forecaster issues their daily region flare forecasts, they have to analyze a great deal of data that comes into the forecast center. Typically, at the most simple level an active region is classified using data from the United States Air Force Solar Optical Observing Network (SOON). Observatories are located to provide twenty-four hour coverage of the Sun. The data sent to SWPC, from each of the observing stations, provides information on each active region such as: modified Zurich classification [McIntosh, 1990], magnetic classification [Smith et al., 1968], sunspot count, areal coverage, location (in both hemispheric coordinates and Carrington coordinates), and areal extent. On a day where all of the observing stations have clear seeing and no equipment issues, a forecaster will get reports from each of the SOON observatories. The forecaster takes active region information from the observatories and chooses the best report (or an average of them). From this generalized region classification the forecaster assigns a probability based on the following considerations: the climatologically based lookup table with flare probability as a function of modified Zurich class, flaring history, growth/decay in spot and areal coverage of the active region, and lastly and probably most importantly a forecasters expertise.

Table 2 shows the Brier skill scores for solar cycle 23 for subjective forecasts and the look-up table forecasts where flares were observed for the first six rows, an overall score is in the seventh and eighth row combining regions where flares might or might not have been observed. The first column is all types of active regions, columns two and three are broken down by magnetic complexity of the active region for the day. For the subjective forecasts, overall the results are as expected, i.e., the forecast performance was inversely related to region complexity.

An issue for any forecasting technique that requires intervention by an observer is the role of forecaster expertise in the predictions. In order to study this, the forecasters were binned into three categories based on their experience level. The first bin was chosen based on the least amount of experience a forecaster would have. The second and third bins were based on the average experience level of the Space Weather Prediction Center, accounting for forecasters with less than the average experience level, and those that had more than the average level of experience. There were thirty-two forecasters that were with the Space Weather Prediction Center during solar cycle 23, with an average experience level of approximately eleven years. Table 3 shows that a difference in experience level seemed to have very little effect on the Brier skill score.

Table 4 shows the contingency table for the subjective forecast probabilities broken down by X-ray flaring class. For the subjective forecast probabilities there were over thirty-one thousand active region records, more than ninety-three thousand forecasts (split evenly between flaring event type) analyzed for this study and summarized in Table 5. The bias was within several tenths of one in all cases for the forecaster issued probabilities, indicating that X-ray events are forecast somewhat more often than they are observed. As illustrated in Table 5, the critical success index (CSI) and equitable threat score (ETS) are fairly closely related. CSI has a bias against rare events, such as X-class flares. ETS compensates for climatology by using the term a_r that equates to the number of forecasts correct due to chance. As expected the scores with relatively uncommon events are similar, and more common events have CSI and ETS scores with greater differences. Probability of detection (POD) is not affected by false alarms, so over-forecasting (forecasting more events) will result in higher POD scores approaching 1. For C-class events roughly 5/8 of the observed events were predicted, and for X-class events roughly 1/2 of all observed events were predicted. False Alarm Ratio (FAR) and Probability of False Detection (POFD) are both sensitive to event climatology and do not consider missed events. POFD can be improved by decreasing the number of "yes" forecasts to cut the amount of false alarms. According to the FAR 2/5 of the forecasted C-Class events and roughly 3/5 of the forecasted M- and X-Class events, were in fact non-events. POFD results indicate that, of all the forecast periods in which flares of their respective classes did not occur, flare forecasts were issued in six, two, and less than one percent of them for C-, M-, and X-classes respectively. Proportion Correct (PC) according to Wilks (2006) "does not distinguish between correct forecasts of an event…and correct forecasts of the nonevent." As the event gets rarer, such as X-Class flares, as seen in Table 5, the PC improves approaching the perfect score of 1. This improvement occurs due to the PC being so heavily biased by the correct forecast of the nonevent. Lastly the Heidke Skill Score (HSS) computes the percentage of correct forecasts after the portion correct due to chance has been removed. With this score as the event gets rarer, as seen in Table 5, the score decreases. Nearly 56% of the forecasts are found to be correct in the case of C-Class subjective forecasts, and the number of correct forecasts decreases to almost 47% and 46% for M- and X-Class subjective forecasts respectively.

4. Validation of Look-up Table Forecasts

The flare climatology look-up table is used by operational forecasters as the starting point for assigning a flare probability to an observed active region. To investigate the sensitivity of just this component to the overall prediction process, we compared the Brier Score resulting from the flare probability indicated by the look-up table. The results are shown by X-ray flare class in Table 2. The Brier Scores are most noticeably different from the subjective forecasts when the active region is complex (has a delta component to the magnetic class).

The contingency table statistics gives a different outlook of how the look-up table is performing. Table 6 shows the contingency tables for the look-up tables and Table 5 shows the calculated scores from these contingency tables. Equitable Threat Score (ETS) was not calculated for the look-up table due to the look-up table being a climatology based. ETS as stated earlier compensates for climatology so the values calculated would not be valid. The Probability of False Detection (POFD) in the C- and M-Class flare categories has the forecasters performing 49% and 57% better than the look-up table (0.059 and 0.022 respectively for the forecaster, and 0.115 and 0.051 for the look-up table). The Probability of Detection (POD) has the forecasters performing 48% better than the look-up table at the X-Class flare category (0.490 for the forecaster and 0.253 for the look-up table). Most notably of all the contingency table scores the Heidke Skill Score has the forecasters performing 67% better than the look-up table at the X-Class flare category (0.455 for the forecasters and 0.151 for the look-up table).

5. Conclusions

For the period studied here, the 24-hour subjective forecasts issued by the Space Weather Prediction Center forecasters are found to be better than the climatology-based look-up table overall. While the Brier Score does not show marked improvement, unless the active region is complex, the contingency table statistics show that there is a significant improvement over using the climatology based look-up table. The False Alarm Rate (FAR) is reduced by 35% in subjective forecasting at the X-Class flaring category when compared to the look-up table. The other important contingency table score to look carefully at is the Probability of Detection (POD) which shows subjective forecasts perform 48% better than the look-up table.

6. Acknowledgments

The author would like to thank Christopher Balch and many others of the National Oceanic and Atmospheric Administration Space Weather Prediction Center for providing the forecaster issued probabilities and region data necessary to conduct this validation.

References

Brier, G.W., "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, **78**, 1950, pp. 1 – 3.

World Climate Research Programme, "Forecast verification: issues, methods, and FAQ," http://www.cawcr.gov.au/projects/verification/, last updated: 26 January 2012.

Lanzerotti, L.J., D.N. Baker, T.E. Jernigan, D.J. Knipp, R. Williamson, S.P. Worden, "Report of the assessment committee for the National Space Weather Program," *NSWP Assessment Report*, FCM-R24-2006, June 2006.

McIntosh, P.S., "The classification of sunspot groups," *Solar Physics*, **125**, February 1990, pp. 251 – 267.

Smith, S.F. and Howard, R., "Magnetic classification of active regions, Structure and Development of Solar Active Regions," *International Astronomical Union*, Symposium no. 35, held in Budapest, Hungary, 4-8 September 1967, 1968, pp. 33-42.

Wilks, D.S., **Statistical Methods in the Atmospheric Sciences**, Academic Press, Burlington, MA, 2006, pp. 255 – 335.

Appendix

LIST OF TABLES

Bias (B) = $\frac{A+B}{A+C}$ Range is 0 to 1, with 1 being a perfect score

Probability of Detection (POD) = $\frac{A}{A+C}$ Range is 0 to 1, with 1 being a perfect score

Probability of False Detection (POFD) = $\frac{B}{B+D}$ Range is 0 to 1, with 0 being a perfect score

Critical Success Index (CSI) = $\frac{A}{A+B+C}$ Range is 0 to 1, with 1 being a perfect score

Proportion Correct (PC) = $\frac{A+D}{A+B+C+D}$ Range is 0 to 1, with 1 being a perfect score

False Alarm Ratio (FAR) = $\frac{B}{A+B}$ Range is 0 to 1, with 0 being a perfect score

Equitable Threat Score (ETS) = $\frac{A - a_r}{A+B+C - a_r}$ where $a_r = \frac{(A+B)(A+C)}{A+B+C+D}$

Range is -$\frac{1}{3}$ to 1, with 1 being a perfect forecast

Heidke Skill Score (HSS) = $\frac{2(AD-BC)}{[(A+C)(C+D)+(A+B)(B+D)]}$

Range is - ∞ to 1, with 1 being a perfect forecast. A negative HSS is indicative that a chance forecast is better, and a 0 HSS is deemed an unskilled forecast.

Tables

Table A1: Two-by-Two Contingency Table

		Event Observed	
		Yes	No
Event Forecast	Yes	A	B
	No	C	D

Table A2: Brier Skill Scores

	All Region Types	Beta, Beta-Gamma, and Gamma Region Types	Beta-Delta, Gamma-Delta, and Beta-Gamma-Delta Region Types
C-Class Flares Observed Subjective Forecasts	0.100	0.123	0.183
C-Class Flares Observed Look-up Table	0.111	0.139	0.193
M-Class Flares Observed Subjective Forecasts	0.031	0.034	0.190
M-Class Flares Observed Look-up Table	0.037	0.042	0.229
X-Class Flares Observed Subjective Forecasts	0.004	0.002	0.067
X-Class Flares Observed Look-up Table	0.005	0.003	0.080
Combined Flaring and Non-flaring Subjective Forecasts	0.045	0.053	0.147
Combined Flaring and Non-flaring Look-up Table	0.051	0.061	0.167

Table A3: Brier Skill Scores for Subjective Forecast Probabilities by Years of Experience

	≤ 3 Years	> 3 and ≤ 10 Years	> 10 Years
Number of Forecasters	17	6	9
C-Class Flares Observed Subjective Forecasts	0.103	0.102	0.098
M-Class Flares Observed Subjective Forecasts	0.033	0.025	0.031
X-Class Flares Observed Subjective Forecasts	0.004	0.004	0.004
Combined Flaring and Non-flaring	0.047	0.043	0.044

Table A4: Contingency Tables for Subjective Forecast Probabilities by X-Ray Flare Class

Yes Forecast > 0.50		C-Class Observed Yes	C-Class Observed No
C-Class Forecast	Yes	2476	1630
	No	1458	25920

Yes Forecast > 0.35		M-Class Observed Yes	M-Class Observed No
M-Class Forecast	Yes	511	685
	No	406	29882

Yes Forecast > 0.25		X-Class Observed Yes	X-Class Observed No
X-Class Forecast	Yes	50	67
	No	52	31315

Table A5: Contingency Table Statistics by X-Ray Flare Class

	Bias	CSI	POD	POFD	PC	FAR	ETS	HSS	Records
Perfect Score	1	1	1	0	1	0	1	1	
C-Class Flare Subjective Forecasts	1.043	0.445	0.629	0.059	0.902	0.397	0.389	0.560	31484
C-Class Flare Predictions Look-up Table	1.099	0.366	0.563	0.115	0.829	0.488	---	0.431	21634
M-Class Flare Subjective Forecasts	1.304	0.319	0.557	0.022	0.965	0.573	0.304	0.466	31484
M-Class Flare Predictions Look-up Table	1.559	0.185	0.400	0.051	0.926	0.743	---	0.276	21634
X-Class Flare Subjective Forecasts	1.147	0.296	0.490	0.002	0.996	0.573	0.294	0.455	31484
X-Class Flare Predictions Look-up Table	2.222	0.085	0.253	0.009	0.988	0.886	---	0.151	21634

Table A6: Contingency Tables for Look-Up Table Probabilities by X-Ray Flare Class

Yes Forecast > 0.25		C-Class Observed	
		Yes	No
C-Class Forecast	Yes	2141	2042
	No	1665	15786

Yes Forecast > 0.25		M-Class Observed	
		Yes	No
M-Class Forecast	Yes	362	1049
	No	543	19680

Yes Forecast > 0.15		X-Class Observed	
		Yes	No
X-Class Forecast	Yes	25	195
	No	74	21340

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/RVBX/Misty Crown	1 cy

This page is intentionally left blank.