Space Weather

RESEARCH ARTICLE
10.1029/2020SW002670

A Summary of National Oceanic and Atmospheric Administration Space Weather Prediction Center Proton Event Forecast Performance and Skill

H. M. Bain[1,2], R. A. Steenburgh[2], T. G. Onsager[2], and E. M. Stitely[3]

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, CO, USA, [2]Space Weather Prediction Center, NOAA, Boulder, CO, USA, [3]Millersville University, Millersville, PA, USA

Key Points:

- For Solar Cycle 24, Space Weather Prediction Center day 1 probabilistic proton forecasts have a Brier Skill Score of 0.25 over persistence
- The ≥10 MeV proton Warnings have a Probability of Detection (POD) of 91% and a False Alarm Ratio (FAR) of 24% with a median lead time of 88 min
- The ≥100 MeV proton Warnings have a POD of 53% and a of FAR 38% with a median lead time of 10 min

Correspondence to:
H. M. Bain,
hazel.bain@noaa.gov

Citation:
Bain, H. M., Steenburgh, R. A., Onsager, T. G., & Stitely, E. M. (2021). A summary of National Oceanic and Atmospheric Administration Space Weather Prediction Center proton event forecast performance and skill. *Space Weather*, *19*, e2020SW002670. https://doi.org/10.1029/2020SW002670

Received 29 OCT 2020
Accepted 20 MAY 2021

Abstract The effects of solar radiation storms at Earth are felt across a number of technology-based industries. Energetic particles present during these storms impact electrical components on spacecraft, disrupt high frequency radio communications, and pose a radiation risk for passengers and crew on polar flight routes, as well as for astronauts. An essential aspect of space weather forecasting is therefore to predict the occurrence and properties of a solar proton event before it occurs. In this study, we review radiation storm products issued by the National Oceanic and Atmospheric Administration's Space Weather Prediction Center (SWPC) during Solar Cycles 23 and 24. These include three-day probabilistic proton event forecasts and short-term Warning and Alert hazard products. We present performance metrics and forecast skill scores for SWPC probabilistic forecasts and Warning products, which can be used as a benchmark for assessing the performance of radiation storm forecast models.

Plain Language Summary Energetic events occurring at the Sun, such as solar flares and coronal mass ejections, can accelerate protons, electrons and ions to high energies. These energetic particles can travel towards Earth where they are observed as a solar radiation storm. These storms can impact a number of technology-based systems and industries. For example, the energetic particles present during these storms impact electrical components on spacecraft, disrupt high frequency radio communications, and pose a radiation risk for passengers and crew on polar flight routes, as well as for astronauts. An essential aspect of space weather forecasting is therefore to predict the occurrence and properties of a solar proton event before it occurs. The National Oceanic and Atmospheric Administration Space Weather Prediction Center (SWPC) issues space weather forecasts for solar radiation storms. In this study, we compare these forecasts with observations to determine how accurate the forecasts were. In particular the paper reviews forecasts between January 1996 and December 2019, a time period covering the last two Solar Cycles. The results of this paper can be used to test new forecasting models to determine if they would be capable of improving SWPC forecasts.

1. Introduction

The National Oceanic and Atmospheric Administration's Space Weather Prediction Center (NOAA SWPC) operates 24/7, to continuously monitor the near-Earth space environment. SWPC forecast, Warning, and Alert products provide advance information and real-time situational awareness of solar and geophysical events and their impacts at Earth.

Solar energetic particle (SEP) events, consisting of protons, electrons and heavy ions, are a major component of space weather. These radiation storms have the capacity to: damage electrical hardware on spacecraft (Smart & Shea, 1992); disrupt long distance high frequency (HF) radio communications; and can pose a radiation hazard to astronauts, as well as passengers and crew on high-flying aircraft over the poles (Beck et al., 2005; Posner, 2007; Schrijver & Siscoe, 2010; Schwadron et al., 2010). An essential aspect of space weather forecasting is therefore to predict the occurrence of SEPs at Earth and the continuation of an event once it is in progress.

SWPC radiation storm products are based on proton intensity levels in geostationary orbit, as observed by particle sensors on the Geostationary Operational Environmental Satellite (GOES) (Kress et al., 2020; Rodriguez et al., 2014; Sauer, 1989). The NOAA Solar Radiation Storm Scale (S-scale: https://www.swpc.noaa.

Scale	Description	Effect	Physical measure (Flux level of >= 10 MeV particles)	Average Frequency (1 cycle = 11 years)
S 5	Extreme	**Biological:** Unavoidable high radiation hazard to astronauts on EVA (extra-vehicular activity); passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** Satellites may be rendered useless, memory impacts can cause loss of control, may cause serious noise in image data, star-trackers may be unable to locate sources; permanent damage to solar panels possible. **Other systems:** Complete blackout of HF (high frequency) communications possible through the polar regions, and position errors make navigation operations extremely difficult.	10^5	Fewer than 1 per cycle
S 4	Severe	**Biological:** Unavoidable radiation hazard to astronauts on EVA; passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** May experience memory device problems and noise on imaging systems; star-tracker problems may cause orientation problems, and solar panel efficiency can be degraded. **Other systems:** Blackout of HF radio communications through the polar regions and increased navigation errors over several days are likely.	10^4	3 per cycle
S 3	Strong	**Biological:** Radiation hazard avoidance recommended for astronauts on EVA; passengers and crew in high-flying aircraft at high latitudes may be exposed to radiation risk. **Satellite operations:** Single-event upsets, noise in imaging systems, and slight reduction of efficiency in solar panel are likely. **Other systems:** Degraded HF radio propagation through the polar regions and navigation position errors likely.	10^3	10 per cycle
S 2	Moderate	**Biological:** Passengers and crew in high-flying aircraft at high latitudes may be exposed to elevated radiation risk. **Satellite operations:** Infrequent single-event upsets possible. **Other systems:** Small effects on HF propagation through the polar regions and navigation at polar cap locations possibly affected.	10^2	25 per cycle
S 1	Minor	**Biological:** None. **Satellite operations:** None. **Other systems:** Minor impacts on HF radio in the polar regions.	10	50 per cycle

Figure 1. The National Oceanic and Atmospheric Administration Space Weather Prediction Center Solar Radiation Storm Scale (S-scale) relating the intensity of Geostationary Operational Environmental Satellite ≥ 10 MeV integral proton flux (in units of p.f.u.) and to biological impacts and effects on technological systems.

gov/noaa-scales-explanation), shown in Figure 1, is used to communicate the severity of proton events to the general public and customers. The S-scale is based on the ≥10 MeV integral proton flux observed by GOES, and relates the intensity of an event to the impacts on satellite systems, HF communications, navigation systems and biological impacts to astronauts, in addition to crew and passengers on aircraft.

An event is defined as the time when the GOES 5-min averaged ≥10 MeV integral proton flux exceeds 10 particle flux units (1 p.f.u. = 1 particle/(cm^2 s sr)) that is, the threshold for an S1 storm, for at least three consecutive 5-min readings. SWPC issues three kinds of products relating to proton events. Long-term, probabilistic forecasts indicate the likelihood of a proton event occurring in the next 3 days. Short-term (hours-minutes) hazard products are deterministic and indicate an imminent threat (Warning) or observed onset (Alert) of an event. Warnings and Alerts are also issued for the ≥100 MeV integral proton flux exceeding 1 p.f.u.. Finally, summaries are issued when an event concludes.

In this study, we assess the performance and skill of NOAA SWPC proton event products for the years 1996–2019, covering Solar Cycles 23 and 24. In Section 2 we assess the center's probabilistic forecasts and in Section 3 we assess the Warning and Alert products.

2. Probabilistic Proton Event Products

SWPC forecasters calculate whole solar disk proton event probabilities for upcoming days 1, 2, and 3 (where a probability is issued for each individual day) as follows:

1. Each active sunspot region is assigned a McIntosh Class based on its white light characteristics (McIntosh, 1990).
2. Historic proton event rates for days 1–3 are retrieved for each region, based on the assigned McIntosh Class.
3. The historic probabilities associated with each region's McIntosh class are then adjusted, taking into account: current activity and trends over the past few days; magnetic class and the regions magnetic structure/morphology; proton event history. If data are available, morphology of coronal loops (e.g.,

potential vs. non-potential magnetic field configurations), structure of fibrils, magnetic shear, etc. are also considered (Toriumi & Wang, 2019).

4. Whole-disk proton event probabilities are then calculated based on the individual active region probabilities. (Note: individual region event probabilities are archived in the daily SWPC solar synoptic drawings, see https://www.swpc.noaa.gov/products/solar-synoptic-map)

A host of real-time observational data are available to SWPC forecasters to support adjustments to the proton event climatological probabilities described in step 3. Examples include solar active region magnetogram data from GONG (Hill, 2018), products from the Solar Dynamic Observatory (SDO: Pesnell et al., 2012) Helioseismic and Magnetic Imager (HMI: Schou et al., 2012), imagery from the SDO Atmospheric Imaging Assembly (Lemen et al., 2012) and the GOES Solar UltraViolet Imager (Seaton et al., 2020; Vasudevan et al., 2019), as well as GOES X-ray and proton data. Input from the United States Air Force solar region and activity reports, containing plain text information remarking on active region characteristics, is also considered.

The resulting event probabilities are communicated via SWPC forecast products, including the 3-day forecast product, the Report of Solar and Geophysical Activity (RSGA), and the Solar Synoptic Drawing mentioned above. As the 3-day product only dates back to 2010, we will limit our focus to the RSGA.

The RSGA, also referred to as the Joint United States Air Force and NOAA Solar Geophysical Activity Report and Forecast, is issued daily at 2200 UTC and is valid from 0000 UTC of the following day. It includes day 1, 2, and 3 forecasts. The product is available from the SWPC website at https://www.swpc.noaa.gov/products/report-and-forecast-solar-and-geophysical-activity and by email subscription (https://www.swpc.noaa.gov/content/subscription-services). An archive of RSGA products since 1966 is hosted by the NOAA National Centers for Environmental Information (NCEI) and is available at https://www.ngdc.noaa.gov/stp/spaceweather.html under Daily Reports. The RSGA is broken into six sections:

I Analysis and Forecast of Solar Active Regions and Activity
II Geophysical Activity Summary and Forecast
III Event Probabilities
IV Penticton 10.7 cm Solar Flux
V Geomagnetic A Indices (Fredericksburg, Virginia and Planetary)
VI Geomagnetic Activity Probabilities at Mid and High Latitudes

Proton event probabilities are listed in Section III of the RSGA as for example, "Proton 60/50/45", indicating a 60%, 50%, and 45% chance of an S1 event or higher occurring on days 1, 2, and 3 respectively, see Figure 2 right panel (note: day 1 begins at 0000UT, two hours after the RSGA is issued).

The end of Solar Cycle 24 in December 2019 provides an opportunity to calculate and compare forecast performance metrics and skill scores for cycles 23 and 24, gleaned from RSGA products from 1996 to 2019. Figure 3 (top) shows the progression of Solar Cycles 23 and 24 via the monthly mean V2 total sunspot number, archived by the World Data Center Sunspot Index and Long-term Solar Observations, Royal Observatory of Belgium (http://www.sidc.be/silso/datafiles). The bottom panel of Figure 3 shows the corresponding S1 proton event rate as a rolling mean of the previous 120 days. This 120 day climatological event rate will be used later as a reference event rate for no skill forecasts. This follows similar climatological averaging by Sharpe and Murray (2017), and used in Leka et al. (2019) for flare forecast model verification. The 120 day mean is used to capture the relatively sharp onset of events at the start and end of the Solar Cycle after several years of few events during solar minimum. Horizontal dashed lines indicate the mean S1 proton event rate for each Solar Cycle, 0.058 and 0.033 events per day for Solar Cycle 23 and 24 respectively.

2.1. Performance Metrics

RSGA 3-day proton event forecasts allow users to optimize their risk strategy by incorporating the forecast uncertainty conveyed by the event probability (in contrast to deterministic forecasts which give a simple Yes/No forecast). However, thresholding probabilistic forecasts to frame them as deterministic Yes/No forecasts can provide a quick assessment of forecast performance. Applying a decision threshold of P_{th}, such

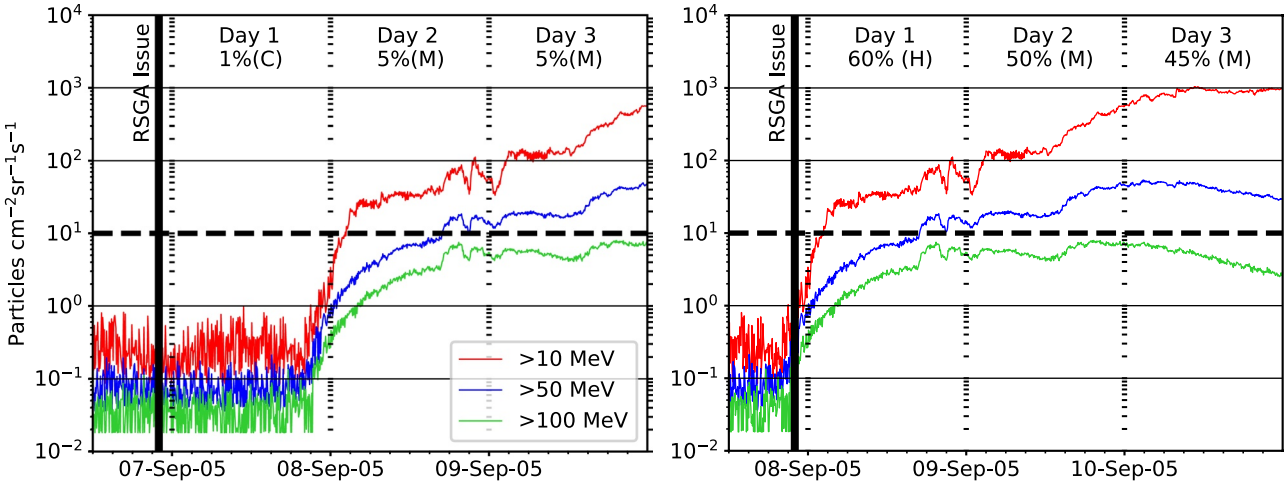

Figure 2. Examples of the 3-day Space Weather Prediction Center S1 proton forecasts as distributed in the Report of Solar and Geophysical Activity (RSGA) on two consecutive days during September 2005. Vertical line shows the RSGA issue time and the corresponding day 1–3 forecasts. Portions of this paper apply a probability threshold of $P_{th} = 50\%$ to RSGA probabilistic forecasts to label outcomes as the categories of a contingency table, see Table 1. Parentheses show whether the event was considered a hit (H), miss (M) or correct null (C).

that forecast probabilities $>P_{th}$ indicate a *Yes* forecast and $\leq P_{th}$ indicate a *No* forecast, and comparing with observations generates a 2×2 contingency table of forecast and observation outcomes, see Table 1.

TP is the number of true positives, the number of yes forecasts that coincided with an observed proton event, that is, a hit. FP is the number of false positives, the number of yes forecasts which did not result a proton event, that is, a false alarm. FN is the number of false negatives, the number of proton events

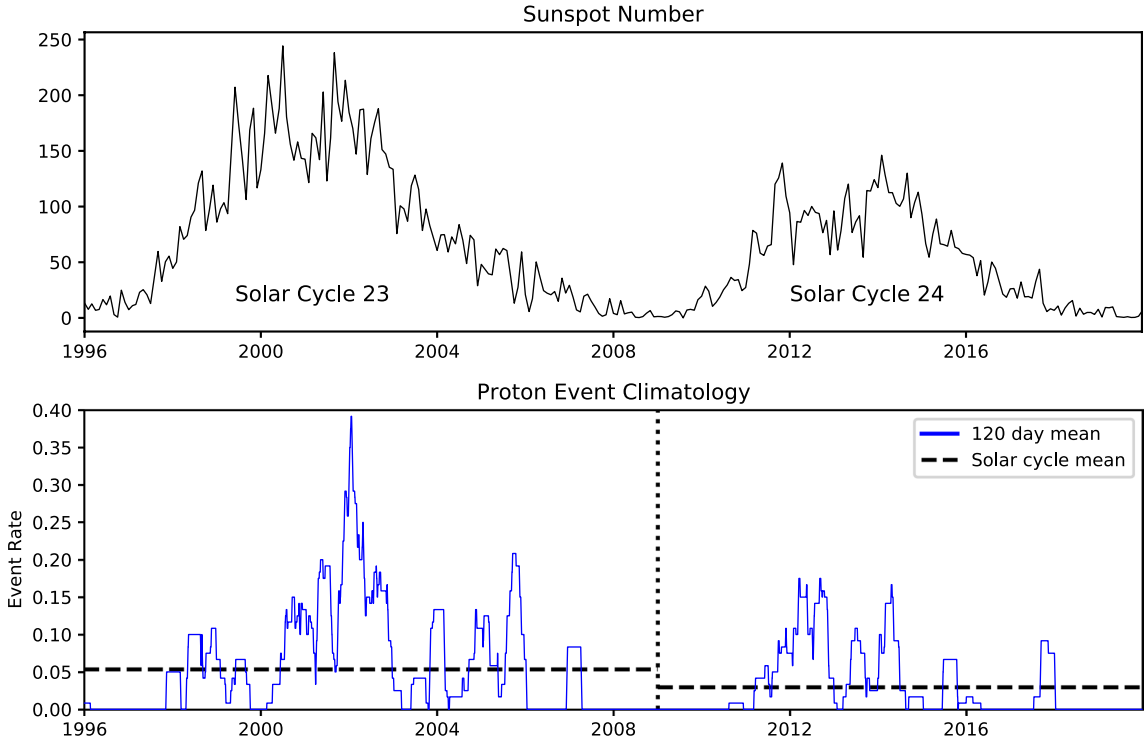

Figure 3. Top: Monthly mean total sunspot number depicting the progression of Solar Cycles 23 and 24 (Source: World Data Center Sunspot Index and Long-term Solar Observations, Royal Observatory of Belgium, Brussels). Bottom: 120 day rolling mean S1 proton event rate climatology. Horizontal dashed lines indicate overall Solar Cycle climatological S1 event rate.

Table 1
Contingency Table

		Observed	
		Yes	No
Forecast	Yes	TP (Hits)	FP (False Alarms)
	No	FN (Misses)	TN (Correct Nulls)

Abbreviations: FN, number of false negatives; FP, number of false positives; TN, number of true negatives; TP, number of true positives.

that occurred with a corresponding no forecast having been issued before the event onset, that is, a missed event, and TN indicates the number of true negatives, that is, the number of correct no forecasts of correct nulls. From this contingency table, we can generate well known performance metrics, such as the Probability of Detection (POD),

$$POD = TP / (TP + FN) \quad (1)$$

the Probability of False Detection (POFD), also known as the False Alarm Rate,

$$POFD = FP / (TN + FP) \quad (2)$$

the False Alarm Ratio (FAR),

$$FAR = FP / (TP + FP) \quad (3)$$

and the Critical Success Index (CSI)

$$CSI = TP / (TP + FP + FN) \quad (4)$$

The RSGA forecasts are compared to GOES particle sensor data, which is publicly available at the NOAA National Centers for Environmental Information (NCEI). It is important to point out the identification of proton events and the forecast products that are analyzed in this study are all based on measurements from the GOES satellites. Over the two solar cycles covered by this study, measurements are obtained from as many as eight different GOES satellites, from GOES-8 through GOES-15. Consequently, it is unavoidable that some variability has occurred in the measurements from satellite to satellite. Nonetheless, all of the energetic proton detectors used on these GOES satellites have had identical design, and it has been shown that the relative responses agree to within +/−20%, sometimes better than 1% (Rodriguez et al., 2014). The differences in detector responses could impact the identification of an event and its timing in cases where the flux levels are near the event threshold, but the impact will be small for events that significantly exceed the thresholds. The effect on real-time proton event Alerts is <10%.

Figure 4 shows how select performance metrics, as well as values for the Heidke Skill Score (HSS) (see Section 2.2 later) vary as a function of P_{th}. Such plots provide a guide for forecasts users in determining a decision threshold based on the optimization of a particular metric. A record of the corresponding contingency table, metrics and skill score values plotted in Figure 4 is reported in Appendix A and Tables A1 and A2 for further assessment by forecast users and modelers. From Figure 4 an increase in FAR is observed for day 3 forecasts around the 80% threshold, this is discussed in detail in the following sections.

Table 2 highlights the SWPC day 1, 2, and 3 contingency table results (TP, FP, TN, and FN) and the associated POD, POFD, FAR and CSI forecast performance metrics for $P_{\text{th}} = 50\%$. Figure 5 shows the corresponding metrics by year. The most obvious trend is the decreasing performance level from day 1 to day 3 forecasts, with the POD decreasing the further the forecast is extended into the future. Overall there is some improvement from Solar Cycle 23 to 24, with the Solar Cycle total POD increasing from 47% to 62% for day 1, with improvements also seen for day 2 and 3 forecasts. Interestingly, both Solar Cycles show an increasing POD in the declining phase of the cycle. This can be traced to a relationship with event duration during these years. Figure 6 shows the number of events lasting 1, 2 and ≥ 3 days in length, normalized to the total number of events per year. The year 2006 has the highest POD of either Solar Cycle and was also characterized by a single period of elevated proton flux lasting 10 consecutive days from December 6 to 15. In cases where the proton flux remains consistently above threshold during extended radiation storms, it is easier to forecast the continuation of the event for the upcoming days. For other years there are a number of one day events, the onset of which are hard to forecast, as we will see in the following Sections. This relationship with event duration, is also reflected in other performance metrics such as CSI, as well as the HSS and Receiver Operating Characteristic Skill Score (ROCSS) introduced later.

For both Solar Cycles there is a very low, <1%, POFD, which results from a tendency to issue an event probability of >50% only after the S1 threshold has been crossed. Infrequent exceptions to this include,

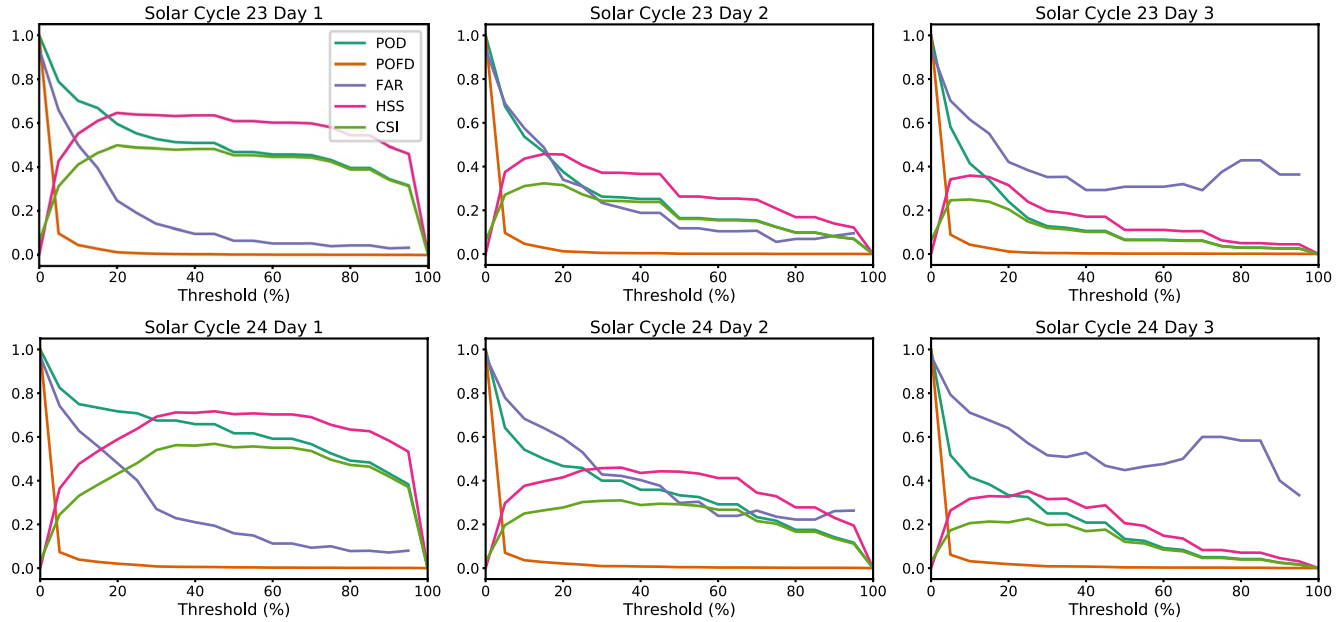

Figure 4. Probabilistic proton event forecast performance metrics and skill as a function of decision threshold for Solar Cycles 23 (top) and 24 (bottom), day 1 (left), day 2 (middle) and day 3 (right).

for example, when an active region has been particularly active and already produced a number of proton events at Earth, a forecaster may keep the probability of an event high in anticipation of another event occurring, even after the previous event has ended. In other cases the proton flux is seen to increase before the upcoming RSGA, but has yet to cross threshold. Instead, SWPC Warning products are used as a short term, high confidence forecast of an imminent proton event, see Section 3.

Table 2
Probabilistic Proton Event Forecast Performance Metrics of POD, POFD, FAR and CSI, With HSS, TSS, BSS (With Respect to 120-Day Climatological and Persistence Reference Forecasts), and ROCSS Skill Scores (See Section 2.2) for Solar Cycles 23 and 24

	Solar Cycle 23 (1996–2008)			Solar Cycle 24 (2009–2019)		
Day	1	2	3	1	2	3
TP	129	45	18	74	40	16
FP	9	6	8	14	17	13
TN	4233	4237	4234	3869	3866	3870
FN	146	229	257	46	80	104
POD	0.47	0.16	0.07	0.62	0.33	0.13
POFD	0.002	0.001	0.002	0.004	0.004	0.003
FAR	0.07	0.12	0.31	0.16	0.30	0.45
CSI	0.45	0.16	0.06	0.55	0.29	0.12
HSS	0.61	0.26	0.11	0.70	0.44	0.21
TSS	0.47	0.16	0.06	0.61	0.33	0.13
BSS (clim.)	0.45	0.22	0.12	0.46	0.20	0.08
BSS (pers.)	0.13	0.25	0.30	0.25	0.29	0.31
ROCSS	0.75	0.61	0.51	0.80	0.60	0.47

Note. A decision threshold of 50% was used to convert probabilities to deterministic binary forecasts. Column colors correspond to trend colors used in Figure 5 for day 1 (green), day 2 (orange) and day 3 (purple) forecasts.

Abbreviations: BSS, Brier Skill Score; CSI, Critical Success Index; FAR, False Alarm Ratio; FN, number of false negatives; FP, number of false positives; HSS, Heidke Skill Score; POFD, Probability of False Detection; POD, Probability of Detection; ROCSS, Receiver Operating Characteristic Skill Score; TN, number of true negatives; TP, number of true positives; TSS, True Skill Statistic.

2.2. Skill Scores

While performance metrics provide a commonly used overview of forecast performance, to determine the true quality, or *skill* of a forecast, a variety of *skill scores* can be employed. Specifically, the term *skill score* refers to a measure of forecast performance relative to a reference or no skill forecast, such as a random forecast, persistence or climatology.

2.2.1. Heidke Skill Score and True Skill Statistic

The Heidke Skill Score (HSS) (Heidke, 1926)

$$HSS = \frac{2[(TP \times TN) - (FN \times FP)]}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)} \quad (5)$$

and the Peirce Skill Score (Peirce, 1884), also known as the Hanssen and Kuipers score (Hanssen et al., 1965) or the True Skill Statistic (TSS) (Flueck, 1987)

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \quad (6)$$

are common tools for assessing forecast skill. The HSS considers the ability of a system to correctly forecast events relative to random chance. While the TSS is comprised of the POD and the POFD. Both TSS and HSS are considered equitable measures of performance, where random forecasts (including systems which only ever forecast one outcome) receive

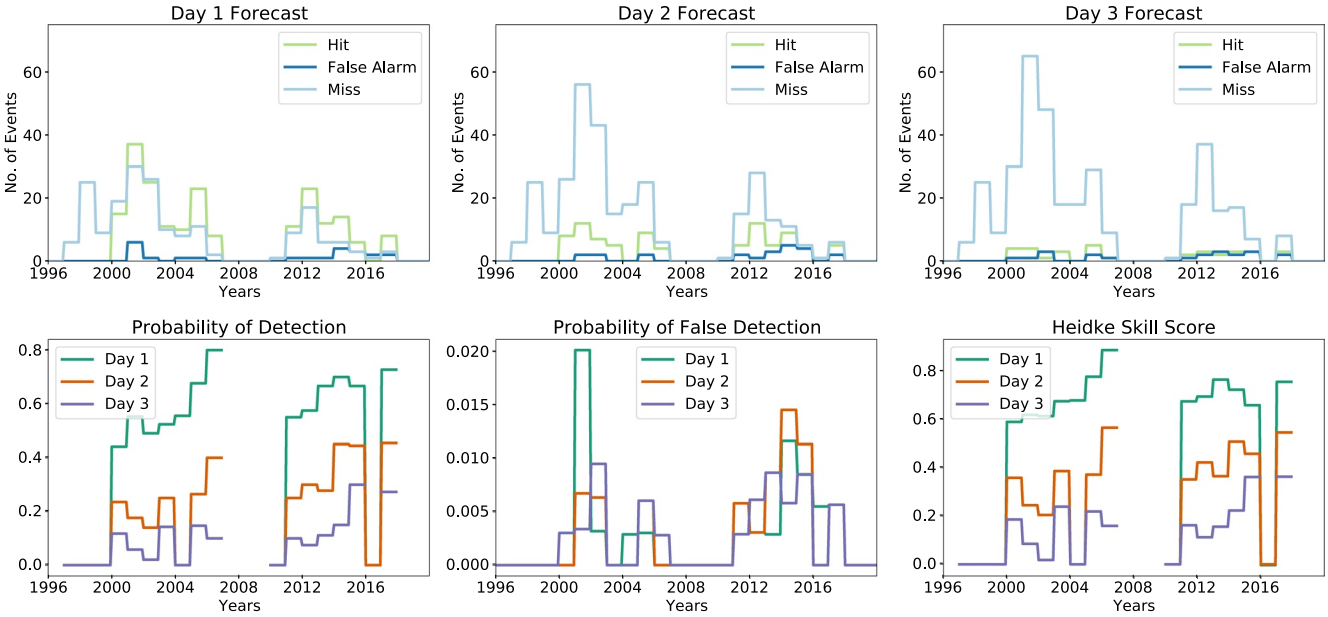

Figure 5. Top: Number of hits (light green), false alarms (dark blue) and missed events (light blue), for day 1 (left), 2 (middle) and 3 (right) thresholded proton event probability forecasts ($P_{th} = 50\%$). Bottom: shows the corresponding values of Probability of Detection (left), Probability of False Detection (middle) and Heidke Skill Score (right) for day 1 (green), 2 (orange), and 3 (purple) forecasts.

the same expected score. In this case HSS = TSS = 0 for a random forecast and 1 for perfect forecasts. A negative value indicates a forecast that is less skillful than random chance. The TSS has an advantage for verification across different forecast systems, in that it can be used for unbiased comparisons between data sets with different event to non-event ratios (Bloomfield et al., 2012). For an in depth discussion of forecast performance metrics and skill score properties, the reader is directed to Chapter 3 of (Jolliffe & Stephenson, 2012). In this study, we provide contingency table numbers for reproducibility and for comparison with other studies which may use different contingency table-based metrics.

Table 2 states the 3-day HSS and TSS results for Solar Cycles 23 and 24, while in Figure 5 we choose to highlight the HSS broken down by year. With such low POFD scores, TSS tracks closely with POD. For both Solar Cycle 23 to 24, day 1 forecast products show considerably better skill than random chance with HSS of 0.61 and 0.70. This skill decreases for days 2 and 3.

2.2.2. Brier Skill Score

The Brier Score (BS) is described as analogous to the mean square error for probability forecasts

$$BS = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2 \tag{7}$$

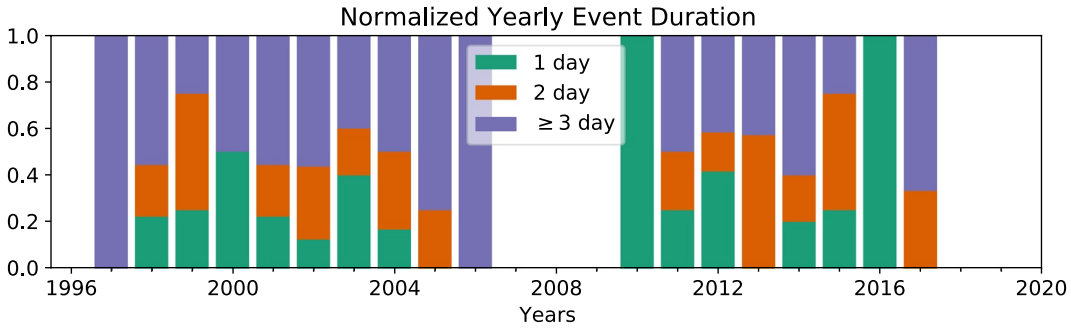

Figure 6. Fraction of proton events of length 1 day (green), 2 days (orange), and ≥3 days (purple) per year for Solar Cycles 23 and 24.

where n is the number of forecast-observation pairs, y_k is the kth probability forecast and o_k it's corresponding observation. $o_k = 1$ for days when the S1 threshold is crossed and $o_k = 0$ for days when no proton event occurs. The Brier Score has a range of $0 \leq \mathrm{BS} \leq 1$, where BS = 0 for a set of perfect forecasts.

The BS can be expressed relative to a set of reference forecasts, thus defining the Brier Skill Score (BSS)

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \tag{8}$$

$$BS_{ref} = \frac{1}{n}\sum_{k=1}^{n}(c - o_k)^2 \tag{9}$$

where BS is the Brier Score for the RSGA proton event probabilistic forecasts using Equation 7, and BS_{ref} is the corresponding Brier Score for the reference forecast, c. BS_{ref} is generated using either the 120-day climatological event rate, plotted in Figure 3, or a record of persistence. In the case of the RSGA forecast products, a persistence forecast refers to whether or not there was a proton event observed on the day that the RSGA product is prepared and issued, that is, the day preceding the day 1 forecast. The persistence forecast is given a value of $c = 1$ for days with an observed proton event and $c = 0$ for days when no event occurred. A BSS score of 0 indicates a forecast system with skill similar to the reference forecast, while a negative score indicates a system that is performing worse than the reference forecast. A BSS of 1 represents a perfect score.

Table 2 shows BSS values for Solar Cycle 23 and 24, using both the 120-day climatological event rate and persistence. For both Solar Cycles, the persistence forecast outperformed the climatological forecast on day one. This is as expected, since it is much easier to predict an event that is already in progress than to anticipate one. In the absence of further particle events, forecasters can usually extrapolate a reasonable estimate for the end of the event from the decaying rate of intensity. The situation is reversed by day 3, when climatological forecasts improve dramatically, compared to persistence. This is an important message for forecasters as they attempt to correctly weight the influence of climatology and persistence when composing the forecast.

2.3. Forecast Discrimination—Receiver Operating Characteristic (ROC)

The Receiver (or Relative) Operating Characteristic (ROC) diagram is a graphical representation of a forecast system's ability to discriminate between two possible outcomes based on a probability forecast and a probability decision threshold, P_{th}, that is, in this case, the ability to discriminate between the solar conditions likely to produce or not produce an S1 event. ROC diagrams plot the POFD against the POD as a function of P_{th}, see Figure 7. Each ROC curve is created by varying P_{th} used to classify a forecast as an event or non-event from 0% to 100%, generating a corresponding contingency table and associated values for POD and POFD. A perfect forecast is indicated by a curve which passes through the top left corner of the diagram, the point (0,1) that is, a set of forecasts which captures all proton events, with no false alarms. The diagonal $x = y$ indicates a set of forecasts with no skill. Randomly generated forecasts would fall along this diagonal. Forecasts with better ability to discriminate between event and non-event conditions will lie in the top left portion of the diagram while forecast systems with little ability to discriminate will lie close to the diagonal.

Figure 7 (top right) shows aggregated ROC diagrams for Solar Cycle 23 and 24. Plotted together in this panel are the results for the day 1, 2, and 3 forecasts. The middle and bottom rows of Figure 7 show ROC trends broken out by year for Solar Cycle 23 and 24 respectively. Each year is indicated by a different ROC color and can be matched to the progression of each cycle using the plot of sunspot number shown in the top left panel.

The ability of a forecast system to discriminate between outcomes, as displayed in an ROC diagram, can be summarized by the additional Area Under the Curve (AUC) metric. A perfect set of forecasts passing through the point (0,1) will have an $\mathrm{AUC}_{\mathrm{perfect}} = 1$, while forecasts lying along the no-skill diagonal with produce an AUC of 0.5, that is, $\mathrm{AUC}_{\mathrm{random}}$. The AUC metric can also be expressed as a skill score, that is, the ROCSS, when compared to the area under the curve arising from a reference forecast from random chance

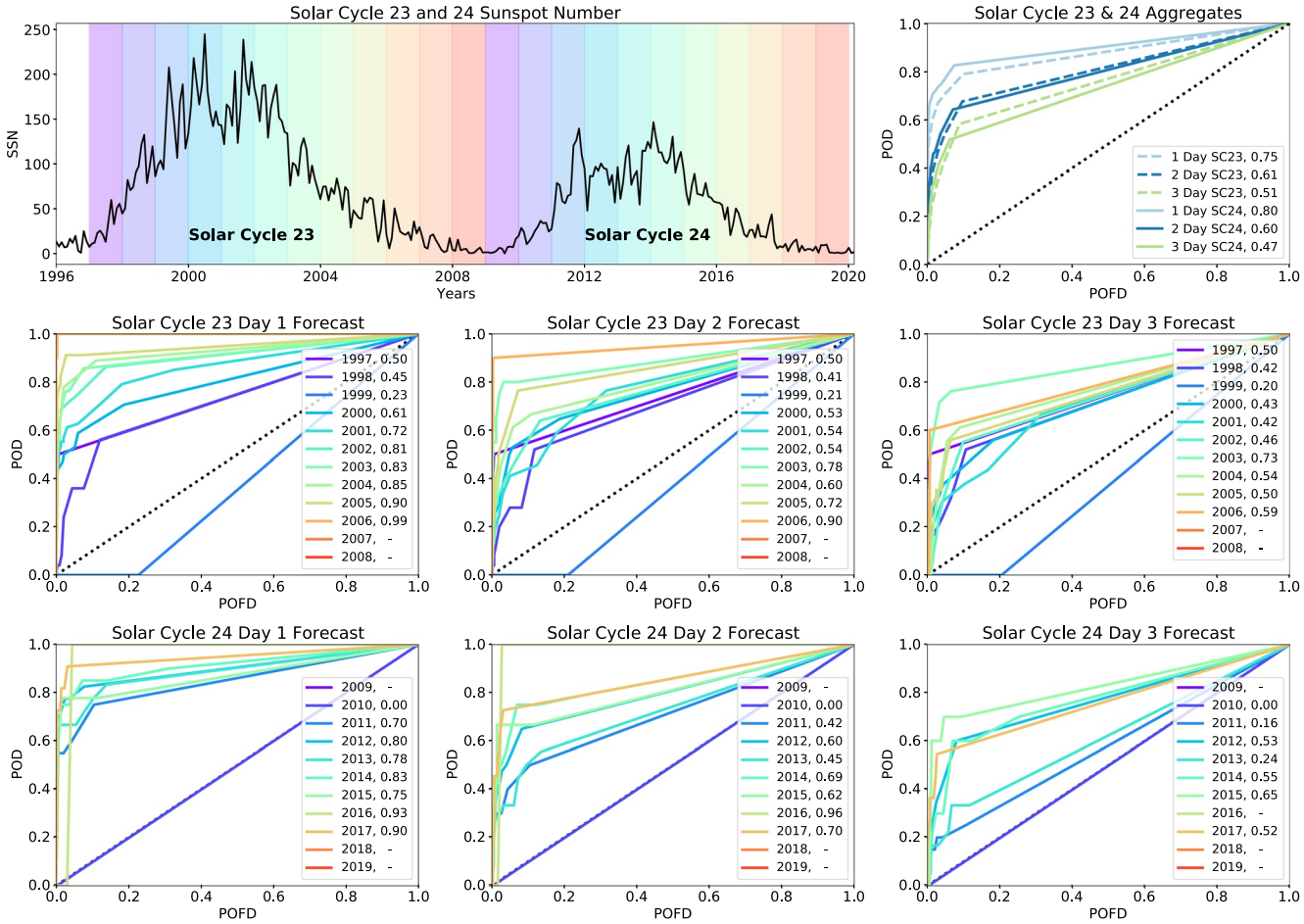

Figure 7. Top left: Monthly mean sunspot number for Solar Cycle 23 and 24 with color coding references to Receiver Operating Characteristic (ROC) curve colors in the middle and bottom rows. Top right: Solar Cycle 23 (dashed) and 24 (solid) ROC curve totals for day 1 (green), 2 (orange), and 3 (purple) forecasts. Middle row: Solar Cycle 23 ROC curves broken out by year for day 1 (left) and 2 (middle) and 3 (right) forecasts. Bottom row: Same as middle row format, for Solar Cycle 24. Values of Receiver Operating Characteristic Skill Score (ROCSS) are stated in the legend for each curve. For years where there are no events from which to construct an ROC curve the ROCSS is left blank.

$$ROCSS = \frac{AUC - AUC_{random}}{AUC_{perfect} - AUC_{random}} \tag{10}$$

$$= \frac{AUC - 1/2}{1 - 1/2} \tag{11}$$

$$= 2AUC - 1 \tag{12}$$

The ROCSS has a range of $(-1,1)$, where ROCSS $= 0$ indicates a forecast with no skill. For each ROC trend, the associated ROCSS value is reported in the legend.

For the Solar Cycle aggregated ROC curves (Figure 7 top right) there is little change between Solar Cycles, as highlighted by the close to overlapping curves and similar values for ROCSS, see also Table 2. However breaking out forecasts per year shows considerable variability from year to year. Note, during certain years in solar minimum there are no events from which to construct an ROC curve, and the value for ROCSS is left blank.

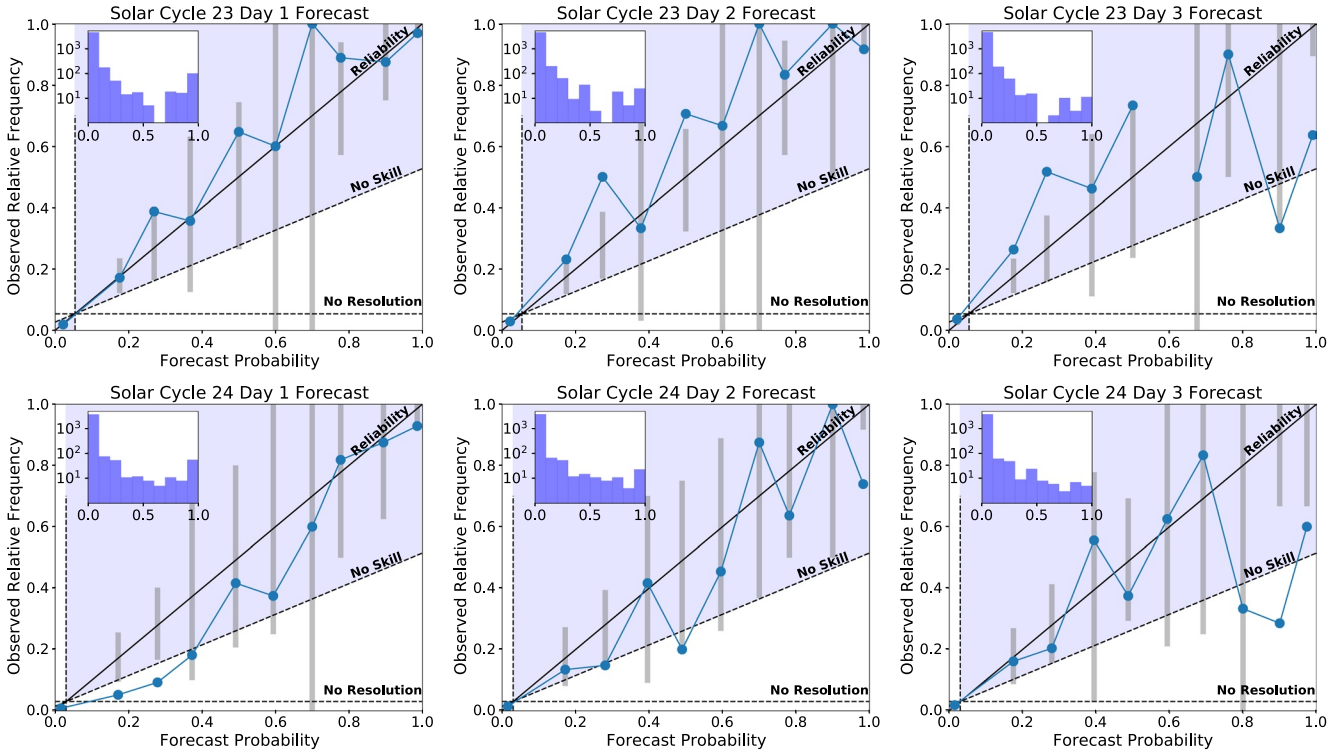

Figure 8. Reliability Diagrams for Solar Cycle 23 (top row) and Solar Cycle 24 (bottom row), for day 1 (left), day 2 (middle) and day 3 (right) forecasts. The diagonal black line depicts perfect reliability. Gray vertical bars visually display 5%–95% quantile consistency bars for guidance when assessing the reliability of the forecast. Horizontal and vertical dashed lines depict the climatological probability of an event for the sample time interval and the corresponding no resolution line, respectively. A bisector between no resolution and perfect reliability denotes a "no-skill" threshold. Shaded areas indicate regions of the parameter space which contribute to a positive Brier Skill Score. Histograms displayed in the figure insets depict the number of forecasts in each bin.

2.4. Forecast Reliability Diagram

The BS, discussed earlier, can be decomposed into three components to measure forecast reliability, resolution, and uncertainty

$$BSS = \frac{Resolution - Reliability}{Uncertainty} \tag{13}$$

all of which can be expressed graphically using a reliability (or attributes) diagram, Figure 8. For further details of BSS decomposition see Jolliffe and Stephenson (2012) and Wilks (2011).

As a complement to the numerical BSS, a Reliability Diagram displays a system's forecast probabilities in comparison to the observed relative frequency of an event, thus giving a visual representation of the forecast *reliability*. For example, a forecast with a 60% probability of an event occurring should result in an event being observed 60% of the time. A well calibrated forecast should therefore lie close to the diagonal reliability line, where the probability of an event occurring is equal to the observed relative frequency of an event. However, due to limited counting statistics, even a forecast with perfect reliability may not lie exactly on the diagonal. For a detailed discussion of how reliability diagrams are generated and why a set of perfectly reliable forecasts can deviate from the one-to-one diagonal between forecasts and observations the reader is referred to Bröcker and Smith (2007). To visualize how far from the diagonal the expected observed relative frequencies can deviate for a set of reliable forecasts, consistency bars (gray vertical bars) are plotted for guidance, generated using a resampling technique and covering the 5%–95% quantiles, see Bröcker and Smith (2007). The reliability of a forecast system can be determined from where the observed relative frequency points (blue circular markers) lie in relation to these consistency bars, rather than their vertical distance from the diagonal alone. Although points may fall towards the extremes of the consistency bars, this is not inconsistent with what is expected from a reliable forecast.

Points falling outside the consistency bars and beneath the diagonal reliability line represent an over forecast, where the average forecast consistently overestimated the probability of an event occurring relative to the fewer average number of events that were actually observed. Alternatively points lying outside the consistency bars and above the diagonal reliability line represent an under forecast, where the average forecast consistently underestimates the probability of an event occurring relative to the larger average number of events that were actually observed.

Figure 8 shows reliability plots for Solar Cycle 23 and 24. Forecasts are divided into 0.1 bins from 0.0 to 1.0, and points are plotted as the average forecast and event frequency within each bin. Histogram insets indicate the number of forecasts present in each bin that is, the forecast sharpness. As expected, as proton events are infrequent occurrences, the majority of forecasts are issued in the lowest probability range, indicating no event is likely. A relative increase is observed in the 90%–100% range. These forecasts are issued when an event is ongoing and there is a high confidence that the proton flux will remain elevated in the coming days. This relative increase in the 90%–100% range is not present in the Solar Cycle 24 day 3 forecasts where it can be considered hard to make a definitive forecast, 3 days into the future. There is also a slightly higher number of forecasts in the 70%–80% range. This particular probability range includes the threshold for an event to be described as "Likely" or "Expected" in SWPC forecast commentary. The following wording is assigned to specific probability ranges. Slight chance 10%–<25%, Chance 25%–<50%, Likely 50%–<75% and Expected 75%–100%. This change of categorization at 75% is weighed against current active region activity and a forecaster may decide to err on a specific side of this threshold, based on their expectation of whether an event is Likely or Expected.

The dashed horizontal and vertical lines depict the aggregated Solar Cycle climatological probability of an event occurring. A forecast system producing points close to the horizontal climatology "no resolution" line shows poor *resolution*, indicating that the system cannot discriminate between an event occurring and not occurring, any better than the climatological probability. Points occurring further away from the no resolution line, vertically, display an ability for the forecast system to discriminate between events occurring at different relative frequencies.

The "no-skill" dashed line denotes a bisector between no resolution and reliability. Points above this line contribute to a positive BSS as the uncertainty term is always positive (see Equation 13). Points contributing to a positive value of the BSS fall within the shaded area of Figure 8.

There are several things to note from Figure 8. Generally most forecast probabilities are found to be consistent with the observations, lying within the vertical extent of the gray consistency bars. However, there are a couple of notable exceptions that show a lack of reliability below the no skill line. In contrast to Solar Cycle 23, day 1 forecasts for Solar Cycle 24 show a tendency to over forecast probabilities ≤40% relative to the true observed frequency of events. This range can sometimes be used to indicate the possibility of an active region, which is not yet on disk, producing a proton event. The Sun Earth Connection Coronal and Heliospheric Investigation on board the STEREO spacecraft, launched in 2006 (Howard et al., 2008), provides extreme ultraviolet images taken from vantage points off the Sun-Earth line. These images of, as yet, unnumbered or recurring ARs behind the limb, do not provide sufficient information for forecasters to assign an event probability based on historical McIntosh class event rates. As such, the probabilistic forecast associated with these regions is assigned at the discretion of the on-duty forecaster. Such forecasts are used as a means to communicate and prepare customers for the non-zero possibility of a proton event in the coming days. This practice is one possible contribution to the day 1 over forecast. Furthermore, during the transition from Solar Cycle 23 to 24, several forecasters retired and SWPC lost over a century of experience. Most new forecasters came from military backgrounds where the penalties for a missed event typically outweighed those for a false alarm.

Day 3 forecasts for Solar Cycle 24 also show a tendency to over forecast, with points >80% falling below the no skill line. This can be a symptom of an event ending earlier than expected (e.g., October 2, 2013) or of a region which was previously producing proton events no longer continuing to do so (e.g., March 13, 2013). For Solar Cycle 24, 9 days contributed to incorrect forecasts in the ≥70% probability range, resulting in false alarms for day 3. All of these day 3 RSGA forecasts were issued on days when the proton flux was already (in one case almost) elevated above the S1 threshold at the time of the forecast. Manually inspecting

Table 3
All Possible Forecast and Outcome Pairs for Consecutive Two Day Forecasts

Event history		Forecast		Outcome			Day 1/Day 2				Day 2/Day 3			
1st day	2nd day	1st day	2nd day	1st day	2nd day	Label	SC23		SC24		SC23		SC24	
No Event	Event	Yes	Yes	False Alarm	Hit	F-H	0.00	(0)	0.05	(2)	0.00	(0)	0.07	(3)
		Yes	No	False Alarm	Miss	F-M	0.01	(1)	0.00	(0)	0.03	(2)	0.00	(0)
		No	**Yes**	**Correct Null**	**Hit**	**C-H**	0.00	(0)	0.00	(0)	0.00	(0)	0.00	(0)
		No	No	Correct Null	Miss	C-M	0.99	(77)	0.95	(39)	0.99	(77)	0.93	(38)
Event	Event	**Yes**	**Yes**	**Hit**	**Hit**	**H-H**	0.23	(45)	0.47	(37)	0.09	(18)	0.16	(13)
		Yes	No	Hit	Miss	H-M	0.26	(51)	0.15	(12)	0.09	(17)	0.19	(15)
		No	Yes	Miss	Hit	M-H	0.00	(0)	0.01	(1)	0.00	(0)	0.00	(0)
		No	No	Miss	Miss	M-M	0.51	(100)	0.37	(29)	0.82	(161)	0.65	(51)
Event	No Event	Yes	Yes	Hit	False Alarm	H-F	0.05	(4)	0.24	(10)	0.08	(6)	0.17	(7)
		Yes	**No**	**Hit**	**Correct Null**	**H-C**	0.37	(29)	0.37	(15)	0.05	(4)	0.12	(5)
		No	Yes	Miss	False Alarm	M-F	0.00	(0)	0.00	(0)	0.00	(0)	0.00	(0)
		No	No	Miss	Correct Null	M-C	0.58	(46)	0.39	(16)	0.86	(68)	0.71	(29)

Note. The bolded rows indicate the perfect forecast combination for each Event History category. Results in the Day 1/Day 2 and Day 2/Day 3 columns are normalized to each event category to show the relative weighting of event outcome within each category group while the numbers in parentheses show the number of pairs that fall into each outcome. Short hand labels correspond to those used in Figure 9.

each of these time periods reveals no obvious erroneous forecast, taking into account the situation at the time of the RSGA forecast. Instead, these time periods reflect the uniqueness of proton event time intensity profiles, which have the potential to for example, increase suddenly due to renewed flaring activity or decrease suddenly as magnetic connectivity to the proton source is lost. On the other hand, forecasts for days 1 and 2 which exceed 50% are found to be reliable. This speaks to the relative ease of continuing a positive forecast, for the relatively shorter period of one to two days into the future, once an event has started. This is discussed further in Section 2.5.

As discussed earlier, the BSS for both Solar Cycles 23 and 24, and each of the 3 day forecasts, is positive, as confirmed by almost all points plotted in Figure 8 falling within the shaded parameter space. For each Solar Cycle, the forecasts show good resolution, being able to discriminate between events and non events, better than climatological probability.

2.5. Consecutive Two-Day Forecasts

Thus far SWPC probabilistic forecast skill has been evaluated by considering forecasts/observation pairs as statistically independent. However, each event can span a number of days and in active periods, several events can cascade on top of one another, which in turn influences the forecast, particularly as persistence plays a large part in forecasting. Transitions from quiet to active periods, and visa versa, can be investigated using a technique demonstrated in Park et al. (2020). Looking at a subset of two consecutive day forecasts for which at least one of the two days is associated with a proton event, we can investigate two day event sequences falling into the following event history categories: no event/event, event/event, and event/no event, therefore assessing the performance of the forecast system at the start, during and at the end of proton events.

Table 3 shows all possible combinations for the first and second day forecasts for two consecutive days and the associated outcomes (i.e., hit, miss, false alarm, and correct null) for each of the three event history categories. For each event history there are four possible forecast outcome pairs. Short hand labels for each outcome are added as a key for Figure 9. For the example of two consecutive days in the no event/event category that is, a day with no proton event followed by a day with a proton event, the four possible outcomes are: F-H, F-M, C-H, and C-M, as indicated in Table 3. The perfect forecast for this scenario is "No Event" for day 1 and "Event" for day 2, resulting in an outcome, Correct Null for day 1 and Hit for day 2

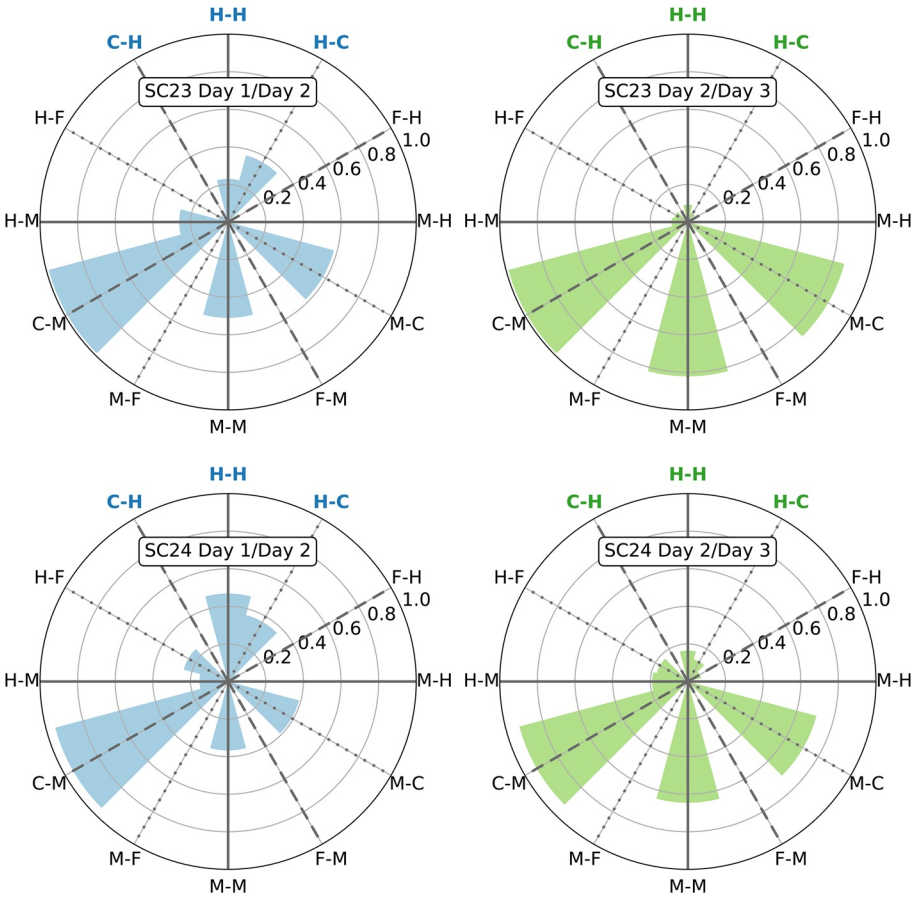

Figure 9. Radar plots depicting the performance of two consecutive days of probabilistic forecasts for which there was a proton event observed on at least one of the two days. The top and bottom rows show results for Solar Cycle 23 and 24 respectively. Left and right plots show results for the day 1/day 2 and day 2/day 3 two-day combinations respectively. Points around the dial of the plot indicate hit (H), false alarm (F), miss (M), and correct null (C) combinations. The radial extent of each cone indicates the relative number of observations that fall into each category. Outcomes associated with the same event history category are connected with similar line styles, No Event/Event (dashed), Event/Event (solid), Event/No Event (dotted). Labels for perfect forecast combinations are highlighted in bold and colored blue or green.

that is, a C-H label. In Figure 2, the example on the left shows a day 1/day 2 forecast outcome of C-M for an event that starts on day 2. Meanwhile, the example on the right shows the following day's RSGA forecast, with a day 1/day 2 outcome of H-M for the same event which now starts on the corresponding day 1 of the RSGA forecast.

Categorization is again made using a decision threshold of $P_{th} = 50\%$. Rows with bold entries in Table 3 indicate the perfect forecast combination for each event category. For each event category, the outcomes are normalized to indicate the relative number of two day periods that fall into each outcome pair, for a given category. This analysis is carried out for combinations of day 1/day 2 forecasts and for day 2/day 3 forecasts, separately for Solar Cycle 23 and 24. See Table 3 for the numerical results while Figure 9 represents this information using radar plots, adapted from Park et al. (2020), where the authors looked at a similar verification for flare forecasting.

Labels around the outside of the plot indicate each possible outcome combination of hits, false alarms, misses, and correct nulls. The radial extent of each cone indicates the relative number of events for each outcome pair in each category. For example, the radial extent of the H-H cone is $N_{H-H}/(N_{H-H} + N_{H-M} + N_{M-H} + N_{M-M})$, where N_i is the number of events in each category i. Groups of outcomes associated with each event history category are placed at orthogonal points to one another on the plot, as depicted by line style connections

across the plot. For example, all outcome pairs associated with the No Event/Event category are connected by dashed lines.

For the transition from No Event on day 1 to Event on day 2, it is clear that no perfect forecasts (C-H) were issued for either Solar Cycle. Instead, the majority of the category fell into the outcome C-M, correctly indicating the first day having no event, but missing the event on the second day, as per the example shown on the left in Figure 2. This result indicates the difficulty in forecasting the onset of proton events greater than 24 h in advance. It also highlights again the tendency to only issue a forecast of ≥50% after the threshold has been crossed. For events that span multiple days, this trend propagates into the M-M outcome for day 2/day 3, as seen in Figure 2 (left). Alternatively, if the proton event only lasts one day, the day 2/day 3 outcome will be an M-C. The M-C outcome is typically associated with short one day events, where the day 1 event was not anticipated but the following No Event was consistent with persistence at the time of the forecast issue.

It is noted that some day 2/day 3 forecasts do show an increased event probability of activity to come. This is likely to occur for situations where a region has been previously active, has an increasing growth rate, is growing in complexity or is rotating into a region of the disk which has a more favorable magnetic connection to Earth. It should also be noted that there are cases, such as that in the right hand panel of Figure 2, where the event begins on day 1, thus falling into the Event/Event category, or Event/No Event for one day events. A subset of these categories represent a correct day 1/day 2 forecast (either H-C for one day events or H-H for multiple day events). As we will see in Section 3, while the SWPC probabilistic forecasts may miss the start of a proton event, Warning forecast products work to supplement the probabilistic forecast with a shorter term, but higher confidence warning for events.

Of particular interest to some SWPC customers, for example, users planning rocket launches or astronaut extravehicular activities, is the ability to forecast days where no proton event occurs, so-called all-clear forecasting. The Event-No Event category is of interest here as it covers the end of an event and the return to quiet conditions. For Solar Cycle 24 Day 1/Day 2, 37% of the two day periods (i.e., 15/41 periods) fall into the correct H-C outcome, however 24% (i.e., 10/41) of the periods fall into the incorrect H-F outcome. In addition to efforts aimed at correctly forecasting event onsets, correctly identifying the time at which the particle flux drops back below S1 threshold is also an important factor for all-clear forecasting. Discussed so far are two day periods during which an S1 event occurred on at least one of the days. However the fourth possible two day event category, No Event to No Event, also deserves a mention with regards to all-clear forecasting. With the low values for POFD discussed in Section 2.1 the majority of two day periods in this category fall into the correctly forecast C-C outcome.

3. Warning and Alert Products

In addition to the three day probabilistic forecast, SWPC issues proton event Warning and Alert hazard products. A *Warning* is issued in response to the occurrence of a solar flare and/or coronal mass ejection (CME), when a forecaster believes that an associated proton event is imminent or highly likely. These Warning products provide a short lead time (typically several minutes to several hours), high confidence prediction. Warnings are issued to signal the likelihood of the ≥10 MeV proton flux exceeding 10 p.f.u. or the ≥100 MeV proton flux exceeding 1 p.f.u. An *Alert* product is issued when a flux threshold has been exceeded, and remains continuously above threshold for three consecutive 5-min readings, indicating the onset of an event. If a forecaster is certain the event is real, they may send the Alert after the first 5-min value exceeding threshold is observed. These procedures are applicable for all Proton Alerts, for all energies and thresholds. Alerts are issued when each of the ≥10 MeV S-scale threshold levels are exceeded, confirming the onset and progression of the event, and for the ≥100 MeV integral flux exceeding 1 p.f.u..

Figure 10 shows an example of a proton event observed by GOES and the corresponding SWPC Warning and Alert products that were issued throughout the event. Horizontal lines indicate the times for which the ≥10 MeV and ≥100 MeV Warnings are valid. Vertical lines indicate the corresponding Alert onset times for each proton energy.

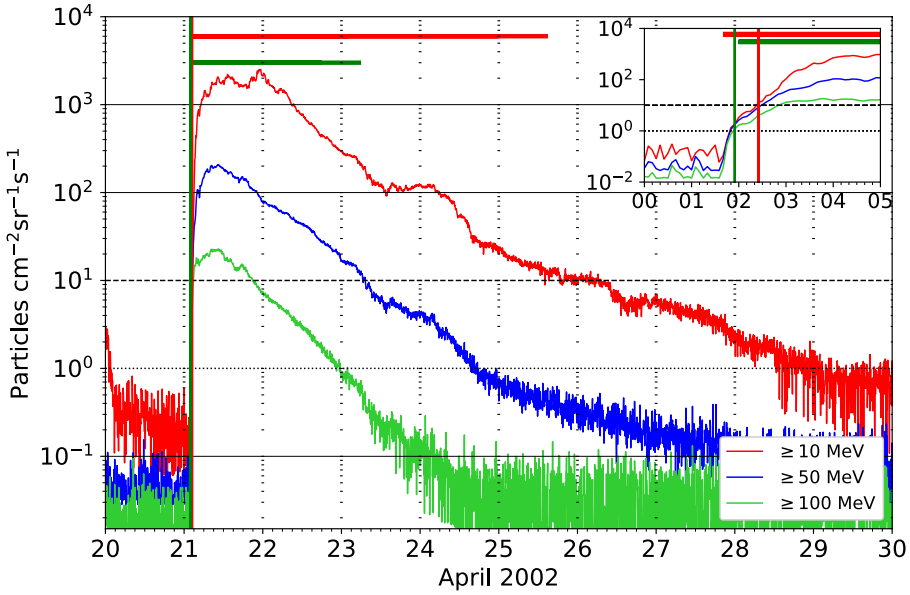

Figure 10. Example of a proton event observed at Geostationary Operational Environmental Satellite. Red and green vertical lines indicate Alerts for the ≥10 MeV integral proton flux exceeding 10 p.f.u. and ≥100 MeV integral proton flux exceeding 1 p.f.u., respectively. Red and green horizontal lines at the top of the plot indicate corresponding Warning product valid times for ≥10 MeV and ≥100 MeV events respectively. The inset shows a close up of the event onset between April 21, 2002 00:00 and 05:00 UTC.

Currently forecasters at SWPC use the Proton Prediction Model or PROTONS model (Balch, 1999, 2008) to guide their SEP Warning forecast products. The PROTONS tool is a statistical model based on the association of flares with SEPs. The model output includes a probability of the GOES ≥10 MeV integral proton flux exceeding 10 p.f.u., the predicted maximum flux level for ≥10 MeV protons and the time of peak flux. As an input, the model uses the GOES (1–8 Å) channel peak soft X-ray flux, integrated soft X-ray flux and the occurrence of metric Type II and Type IV radio bursts. As with the probabilistic forecasts, forecasters use the model as guidance, using their own judgment in regards to the model output and corresponding real time situational awareness from observations, to decide whether or not to issue a proton Warning.

The Preliminary Report and Forecast of Solar Geophysical Data report issued weekly and produced jointly by NOAA SWPC and the US Air Force Weather Agency, contains a record of SWPC Warning and Alert products from 1997 to present. An archive of these reports is hosted by NOAA NCEI and is available at (https://www.ngdc.noaa.gov/stp/spaceweather.html), listed under Periodic Reports and under Weekly Reports. The products are assessed using the following guide.

1. Warning and Alert products are based on the proton flux recorded by the West facing particle sensor on the operational GOES spacecraft at that time. In cases where the proton flux is close to a SWPC defined event threshold, this may lead to the flux crossing the threshold at one spacecraft but not at the other due to slight differences in the radiation environments experienced by spacecraft at different physical locations.
2. An event is counted as a Hit if the Warning is issued any time before the corresponding Alert and as a Miss if the Warning is issued at the same time as the Alert or any time after it, or if there is no Warning issued at all.
3. A False Alarm occurs for a period when a Warning has been issued but the proton flux did not exceed the threshold at the operational spacecraft.
4. The procedure with which Warning and Alerts are issued and extended has evolved over the years. In Solar Cycle 23 it was common for both Warning and Alert products to be extended throughout an event, to indicate continuing activity. In recent years, only Warning products are extended, while the Alert product is used to indicate a threshold crossing only. For this investigation we consider only the first Warning

and Alert product for each event as an indication of forecast skill, we do not consider the extension of a Warning as another "Hit" as the event is already under way.

5. Warning products have an associated Issue time and Begin time, where the Begin time indicates the time from which the Warning is valid. In many cases the Issue and Begin times are concurrent or have little difference. For this investigation we focus on the *Warning Lead Time* that is, the time between the Warning Issue Time and the associated Alert time.
6. For this assessment of skill we adhere strictly to the SWPC definition of a proton event, for which the Warning and Alert system was designed, as the period of time when the proton flux is above a predefined threshold continuously for three consecutive 5-min readings. As such, multiple distinct particle flux increases, due to that is, new eruptions, which may be considered as two events in other studies, will be considered as one event here if the particle flux does not drop back below the threshold between increases. In such cases the Warning is extended to indicate that the particle flux remains above threshold for longer than originally thought. This is more common when considering ≥10 MeV protons. In one example (November 24–27, 2000), one ≥ 10 MeV Warning and Alert pair is issued at the start of the event and is considered as having covered the full period that the ≥10 MeV proton flux is above threshold. But there are also two corresponding ≥100 MeV events during this time. In this example, we record the events as one ≥10 MeV and two ≥100 MeV events and score them according to the issued Warnings.

As previously stated, SWPC forecasters ultimately decide whether or not a hazard product is issued, and do so with end users and SWPC customers in mind. For example, in the case when the proton flux rises gradually or hovers around the threshold for some time, a forecaster may hold off on issuing a Warning until they are confident an event is clearly imminent (instead of being about to turn over and decline). Judgment plays a role in forecast operations and should be acknowledged, particularly when it comes to product lead time. There is a desire to not over or under alert the customer, particularly when Warning and Alerts can be actionable for the customer. As for any system with a human-in-the-loop, making decisions in real-time can occasionally lead to inconsistencies or errors. Additionally, procedures, personnel, and support requirements evolve over time and can influence the resulting hazard products. We have done our best to interpret the historical record of products with this in mind. An example of this is a product with a clearly erroneous timestamp, likely resulting from a mistake that is, a Warning time which was significantly different from the product issue time and an event in question. In cases where it was not possible to correct the record based on the information available, the record was discounted from the study as this kind of error is not related to a determination of forecast skill. Products with these kinds of inconsistencies are rare and it is not expected that their removal impacts the results of this study a significant manner.

Appendix B and Tables B1 lists each Warning and Alert product used in this verification study. Product times are given for each ≥10 MeV (columns a–c) and ≥100 MeV (columns f–h) proton event. Warning lead times are stated for ≥10 MeV (column d) and ≥100 MeV (column i) events separately and corresponding categorizations of Hit, Miss and False Alarm are given in columns e and j respectively. For time periods where an initial Warning is extended over multiple events, the timing of the initial Warning is stated and then listed as EXTENDED. This also captures time periods where there were multiple distinct ≥100 MeV events while the ≥10 MeV proton flux remained continuously above threshold (item 6. above). In the case of false alarms or missed events, when no Warning was issued at all, the Alert and Warning times are left blank. For missed events when the Warning came after the Alert, the lead time is listed as negative value.

Thsere is one time period in July 2005 where there were three distinct events occurring back to back. An Alert was issued for the first threshold crossing, but not for the following two. However the Warning product was extended to cover all three events. From a user perspective, the Warnings were verified and so we consider these as three hits and do not penalize for the lack of separate Alert products.

Figure 11 and Table 4 show the number of hits, false alarms and missed ≥10 MeV and ≥100 MeV proton events for Solar Cycles 23 and 24 and associated metrics for POD, FAR and CSI. Note that determination of POFD is not possible here as there no associated true negatives (TNs) for Warning products. While ≥10 MeV and ≥100 MeV Warning products can be considered as though they are independent, all proton events in this study which cross the ≥100 MeV 1 p.f.u. threshold also have a corresponding S1 proton event. Therefore we have also included an overall Event lead time and categorization (Appendix B table columns k and l), based on the first energy range to cross threshold and the corresponding first Warning issued, regardless of

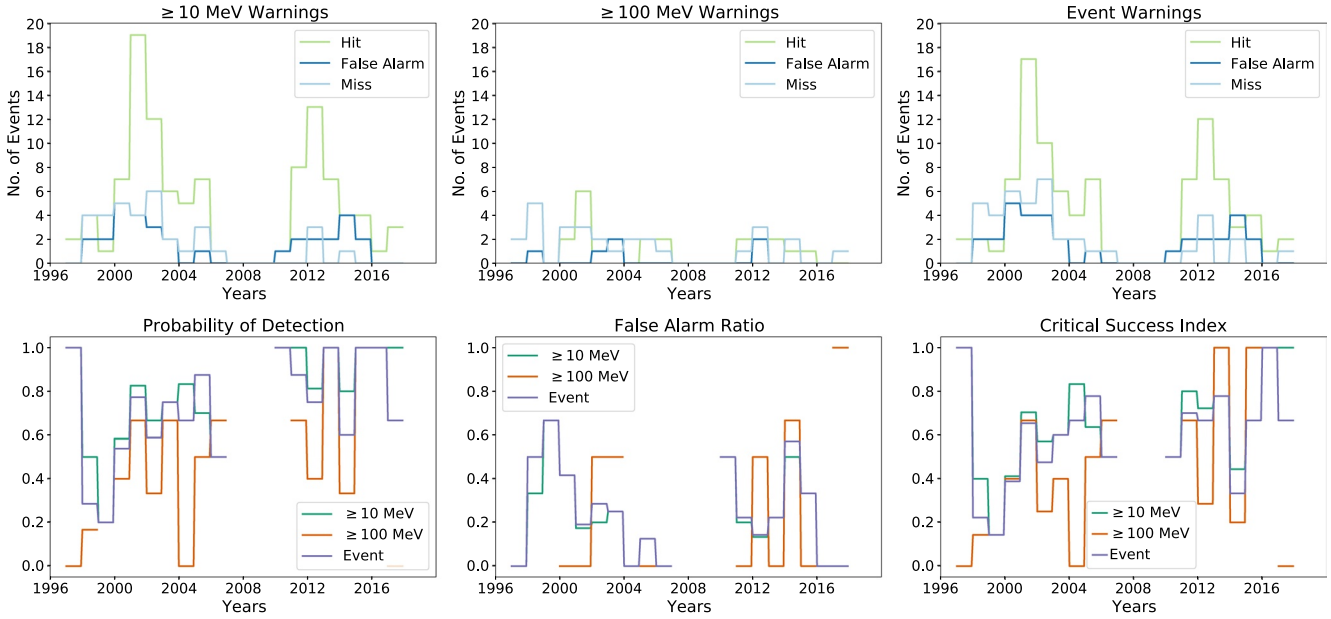

Figure 11. Top row shows National Oceanic and Atmospheric Administration's Space Weather Prediction Center ≥10 MeV (left), ≥100 MeV (middle), and overall Event (right) proton Warning hits (light green), misses (light blue) and false alarms (dark blue). Bottom shows ≥10 MeV (green), ≥100 MeV (orange) and Event (purple) probability of detection (left), false alarm ratio (middle), and Critical Success Index (right) metrics.

whether it was for ≥ 10 MeV or ≥ 100 Mev protons. Considering these products as related to one another gives a overall reflection of lead time and skill for a user who is looking for a indication of a proton event, regardless of particle energy and intensity. However this comes with the caveat that while an above threshold ≥100 MeV event indicates that the ≥10 MeV protons will exceed the threshold at some point, a ≥10 MeV Warning is no indication of whether a ≥100 MeV event will follow or not. Overall Event performance metrics are also shown in Figure 11 and in Table 4.

The POD shows considerable improvement between Solar Cycles, increasing from 68% to 91% for ≥10 MeV proton Warnings, significantly higher than for the ≥10 MeV ($P_{th} = 50$) thresholded day 1 probabilistic forecasts shown earlier (POD = 60% for SC24). However, this comes at the expense of lead time, see Figure 12. Where the probabilistic forecasts aim to give a pre-eruption indication of an proton event, the Warning products are issued in response to an eruptive event on the Sun or when a forecaster observes an increase in the proton flux at GOES, which limits the potential lead time. The median lead time for each product is stated in titles of Figure 12 plots.

Table 4
Proton Warning Performance Metrics of POD, FAR and CSI Metrics for Solar Cycles 23 and 24

SC23 1997–2008

	TP	FP	FN	POD	FAR	CSI
≥ 10 MeV	64	19	30	0.68	0.23	0.57
≥ 100 MeV	16	4	21	0.43	0.20	0.39
Event	57	20	33	0.63	0.26	0.52

SC24 2009–2019

	TP	FP	FN	POD	FAR	CSI
≥ 10 MeV	41	13	4	0.91	0.24	0.71
≥ 100 MeV	8	5	7	0.53	0.38	0.40
Event	37	13	8	0.82	0.26	0.64

Note. Row colors correspond to trend colors used in Figure 11 for ≥ 10 MeV (green), ≥ 100 MeV (orange) and Event (purple) metrics.

Abbreviations: CSI, Critical Success Index; FAR, False Alarm Ratio; FN, number of false negatives; FP, number of false positives; POD, Probability of Detection; SC, Solar Cycle; TP, number of true positives.

As indicated by the POD and lead time results, from remote sensing observations alone it is difficult to determine how energetic a proton event might be at Earth, before seeing the in-situ particle flux start to increase. This is reflected in the number of events with a negative lead time, from the Warning product being issued after the Alert (light blue bars in Figure 12 histograms).

4. Summary

In this study, we have presented a forecast verification of NOAA SWPC proton event three day probabilistic forecasts and Warning products for Solar Cycles 23 and 24. The main takeaways from this study are as follows:

1. SWPC probabilistic forecasts have improved from Solar Cycle 23 to 24. With True Skill Scores increasing for day 1 (0.47–0.61), day 2 (0.16–0.34) and day 3 (0.06–0.13) forecasts.

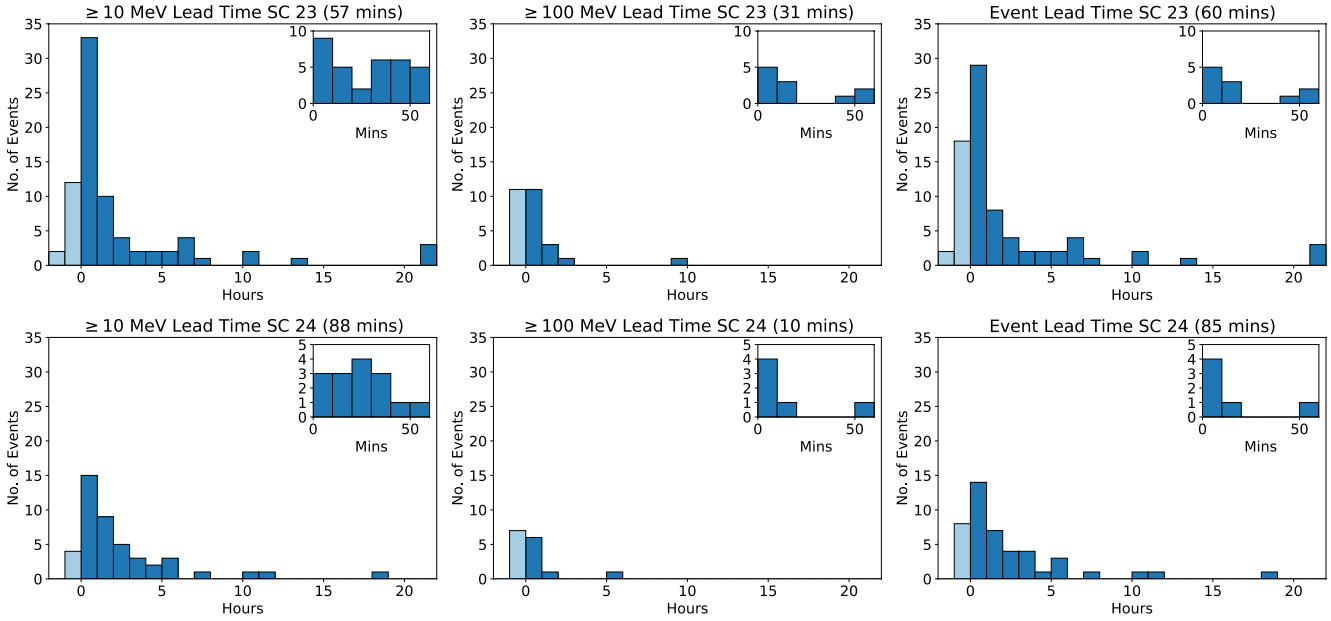

Figure 12. Lead times for ≥10 MeV (left column), ≥100 MeV (middle column) and overall Event (right column) proton Warnings in Solar Cycle 23 (top row) and 24 (bottom row). Inset shows events with lead times with less than one hour in each category. Events with negative lead times (light blue) are labeled as Misses. Events with positive lead times (dark blue) are labeled as Hits. Median lead times for each product are stated in the plot titles.

2. SWPC probabilistic forecasts struggle to accurately forecast the onset of S1 storms ahead of time. However, in general, once the threshold has been crossed SWPC day 1 forecast probabilities are reliable.
3. With respect to climatology, SWPC probabilistic forecasts have a Solar Cycle 24 BSS of 0.46 for day 1, 0.20 day 2, and 0.08 for day 3.
4. Persistence contributes greatly to proton event probabilistic forecasts. With a BSS between 0.25 and 0.31 for day 1–3 forecasts in Solar Cycle 24.
5. Reliability diagrams have revealed a tendency to over issue forecast probabilities in the <40% range for day 1 forecasts during Solar Cycle 24. With the true observed frequency in this range occurring roughly a third to half as often as the probability issued. Similarly there is a tendency to over forecast in the ≥80% probability range for day 3 forecasts in Solar Cycle 24. We attribute this bias to the desire of a forecaster to alert customers to a possible event rather than risking a missed event.
6. In Solar Cycle 23, SWPC ≥ 10 MeV proton Warnings have a POD of 0.68, a FAR 0.23 and CSI of 0.57, with a median lead time of 57 min. While the ≥100 MeV proton Warnings have a POD of 0.43, a FAR 0.2, and a CSI of 0.39 with a median lead time of 31 min.
7. For the most recent Solar Cycle, SWPC ≥10 MeV proton Warnings have a POD of 0.91, a FAR of 0.24, and a CSI of 0.71, with a median lead time of 88 min. While the ≥100 MeV proton Warnings have a POD of 0.53, a FAR 0.38, and a CSI of 0.40, with a median lead time of 10 min. This shows an improvement in performance for SWPC Warning products, however this is for a shorter lead time for the ≥100 MeV Warnings.

The results support the well recognized narrative that it is difficult to determine ahead of time precisely when an active region will erupt and whether a resulting flare or CME will produce a proton event near Earth, as evidenced by SWPC pre-event probabilistic forecasts. It is also difficult to determine ahead of time how intense and energetic the event will be, as seen by the decreasing performance from ≥10 MeV to ≥100 MeV Warning products and limited POD at ≥100 MeV. Lead times for short term hazard products remain within the tens of minutes to a few hours range, highlighting the difficulty in providing a Warning with significant lead time, without incurring a number of false alarms. Improvements between Solar Cycle 23 and 24 are likely a combination of improved observations and resulting situational awareness, as well as the relative number of events that persist for more than one day. However, new models are needed to support forecast operations, particularly in the area of pre-event probabilistic forecasting, in order to increase lead times for customers.

There are a number of physics-based models currently being developed which aim to advance our scientific understanding of energetic particle acceleration and transport in the Heliosphere (e.g., Luhmann et al., 2010; Marsh et al., 2015; Schwadron et al., 2010). However, these models are not yet suitable for real-time operations. Alternatively there are a number of empirical models which use statistical modeling or machine learning (e.g., Dierckxsens et al., 2015; Kahler et al., 2007, 2017; Laurenza et al., 2009; Richardson et al., 2018; Smart & Shea, 1992) to forecast proton events and their characteristics in lieu of a full physics-based model. It is hoped that this verification study can be used as a benchmark for any model running in an operational setting. At this point, it appears as though no single model adequately addresses all proton forecasting requirements, including forecasting the event onset, peak and end times for a range of energies with significant lead time, and robust performance metrics and skill. However, should a model demonstrate increased forecast skill in a particular area, this could prove valuable for forecast operations. Validating these models against the current SWPC operational forecasting baseline forms the next step in determining how a model's performance compares with the specific procedures employed in operations. While model validation may demonstrate an increased forecast skill, work may be required to determine whether or not that forecast skill would still be achievable in a real-time operational setting when model observational inputs may be delayed or unavailable, thus impacting the forecast lead time and potentially the skill. A cost/benefit analysis may also be beneficial to quantify the added value a new model could provide versus the resources needed to employ it.

Appendix A: SWPC Probabilistic Forecast Contingency Tables

Table A1
Performance Metrics and Skills Score for Solar Cycle 23 Probabilistic Forecasts as a Function of Decision Threshold

	Solar Cycle 23 1996–2008										
	%	TP	FP	TN	FN	POD	POFD	FAR	CSI	HSS	TSS
Day 1	0	275	4,242	0	0	1.00	1.00	0.94	0.06	0.00	0.00
	10	193	194	4,048	82	0.70	0.05	0.50	0.41	0.55	0.66
	20	164	54	4,188	111	0.60	0.01	0.25	0.50	0.65	0.58
	30	145	24	4,218	130	0.53	0.01	0.14	0.48	0.64	0.52
	40	140	15	4,227	135	0.51	0.00	0.10	0.48	0.64	0.51
	50	129	9	4,233	146	0.47	0.00	0.07	0.45	0.61	0.47
	60	126	7	4,235	149	0.46	0.00	0.05	0.45	0.60	0.46
	70	125	7	4,235	150	0.45	0.00	0.05	0.44	0.60	0.45
	80	109	5	4,237	166	0.40	0.00	0.04	0.39	0.54	0.40
	90	95	3	4,239	180	0.35	0.00	0.03	0.34	0.49	0.34
	100	0	0	4,242	275	0.00	0.00	NaN	0.00	0.00	0.00
Day 2	0	274	4,243	0	0	1.00	1.00	0.94	0.06	0.00	0.00
	10	147	199	4,044	127	0.54	0.05	0.58	0.31	0.44	0.49
	20	103	53	4,190	171	0.38	0.01	0.34	0.31	0.46	0.36
	30	72	22	4,221	202	0.26	0.01	0.23	0.24	0.37	0.26
	40	69	16	4,227	205	0.25	0.00	0.19	0.24	0.37	0.25
	50	45	6	4,237	229	0.16	0.00	0.12	0.16	0.26	0.16
	60	43	5	4,238	231	0.16	0.00	0.10	0.15	0.25	0.16
	70	42	5	4,238	232	0.15	0.00	0.11	0.15	0.25	0.15
	80	27	2	4,241	247	0.10	0.00	0.07	0.10	0.17	0.10
	90	22	2	4,241	252	0.08	0.00	0.08	0.08	0.14	0.08
	100	0	0	4,243	274	0.00	0.00	NaN	0.00	0.00	0.00

Table A1
Continued

		Solar Cycle 23 1996–2008									
	%	TP	FP	TN	FN	POD	POFD	FAR	CSI	HSS	TSS
Day 3	0	275	4,242	0	0	1.00	1.00	0.94	0.06	0.00	0.00
	10	114	182	4,060	161	0.41	0.04	0.61	0.25	0.36	0.37
	20	66	48	4,194	209	0.24	0.01	0.42	0.20	0.31	0.23
	30	35	19	4,223	240	0.13	0.00	0.35	0.12	0.20	0.12
	40	29	12	4,230	246	0.11	0.00	0.29	0.10	0.17	0.10
	50	18	8	4,234	257	0.07	0.00	0.31	0.06	0.11	0.06
	60	18	8	4,234	257	0.07	0.00	0.31	0.06	0.11	0.06
	70	17	7	4,235	258	0.06	0.00	0.29	0.06	0.10	0.06
	80	8	6	4,236	267	0.03	0.00	0.43	0.03	0.05	0.03
	90	7	4	4,238	268	0.03	0.00	0.36	0.03	0.04	0.02
	100	0	0	4,242	275	0.00	0.00	NaN	0.00	0.00	0.00

Note. Days with bad GOES data or missing forecast records are discounted from this study.

Abbreviations: CSI, Critical Success Index; FAR, False Alarm Ratio; FN, number of false negatives; FP, number of false positives; GOES, Geostationary Operational Environmental Satellite; HSS, Heidke Skill Score; POFD, Probability of False Detection; POD, Probability of Detection; TN, number of true negatives; TP, number of true positives; TSS, True Skill Statistic.

Table A2
Performance Metrics and Skills Score for Solar Cycle 24 Probabilistic Forecasts as a Function of Decision Threshold

		Solar Cycle 24 2009–2019									
	%	TP	FP	TN	FN	POD	POFD	FAR	CSI	HSS	TSS
Day 1	0	120	3,883	0	0	1.00	1.00	0.97	0.03	0.00	0.00
	10	90	152	3,731	30	0.75	0.04	0.63	0.33	0.48	0.71
	20	86	79	3,804	34	0.72	0.02	0.48	0.43	0.59	0.70
	30	81	30	3,853	39	0.68	0.01	0.27	0.54	0.69	0.67
	40	79	21	3,862	41	0.66	0.01	0.21	0.56	0.71	0.65
	50	74	14	3,869	46	0.62	0.00	0.16	0.55	0.70	0.61
	60	71	9	3,874	49	0.59	0.00	0.11	0.55	0.70	0.59
	70	68	7	3,876	52	0.57	0.00	0.09	0.54	0.69	0.56
	80	59	5	3,878	61	0.49	0.00	0.08	0.47	0.63	0.49
	90	52	4	3,879	68	0.43	0.00	0.07	0.42	0.58	0.43
	100	0	0	3,883	120	0.00	0.00	NaN	0.00	0.00	0.00
Day 2	0	120	3,883	0	0	1.00	1.00	0.97	0.03	0.00	0.00
	10	65	140	3,743	55	0.54	0.04	0.68	0.25	0.38	0.51
	20	56	82	3,801	64	0.47	0.02	0.59	0.28	0.42	0.45
	30	48	36	3,847	72	0.40	0.01	0.43	0.31	0.46	0.39
	40	43	29	3,854	77	0.36	0.01	0.40	0.29	0.44	0.35
	50	40	17	3,866	80	0.33	0.00	0.30	0.29	0.44	0.33
	60	35	11	3,872	85	0.29	0.00	0.24	0.27	0.41	0.29
	70	28	10	3,873	92	0.23	0.00	0.26	0.22	0.34	0.23
	80	21	6	3,877	99	0.17	0.00	0.22	0.17	0.28	0.17

Table A2
Continued

	Solar Cycle 24 2009–2019										
	%	TP	FP	TN	FN	POD	POFD	FAR	CSI	HSS	TSS
	90	17	6	3,877	103	0.14	0.00	0.26	0.13	0.23	0.14
	100	0	0	3,883	120	0.00	0.00	NaN	0.00	0.00	0.00
Day 3	0	120	3,883	0	0	1.00	1.00	0.97	0.03	0.00	0.00
	10	50	123	3,760	70	0.42	0.03	0.71	0.21	0.32	0.38
	20	40	71	3,812	80	0.33	0.02	0.64	0.21	0.33	0.32
	30	30	32	3,851	90	0.25	0.01	0.52	0.20	0.32	0.24
	40	25	28	3,855	95	0.21	0.01	0.53	0.17	0.28	0.20
	50	16	13	3,870	104	0.13	0.00	0.45	0.12	0.21	0.13
	60	11	10	3,873	109	0.09	0.00	0.48	0.08	0.15	0.09
	70	6	9	3,874	114	0.05	0.00	0.60	0.05	0.08	0.05
	80	5	7	3,876	115	0.04	0.00	0.58	0.04	0.07	0.04
	90	3	2	3,881	117	0.03	0.00	0.40	0.02	0.05	0.02
	100	0	0	3,883	120	0.00	0.00	NaN	0.00	0.00	0.00

Note. Days with bad GOES data or missing forecast records are discounted from this study.

Abbreviations: CSI, Critical Success Index; FAR, False Alarm Ratio; FN, number of false negatives; FP, number of false positives; GOES, Geostationary Operational Environmental Satellite; HSS, Heidke Skill Score; POFD, Probability of False Detection; POD, Probability of Detection; TN, number of true negatives; TP, number of true positives; TSS, True Skill Statistic.

Appendix B: SWPC Warning and Alerts 1997–2019

Table B1
SWPC ≥10 and ≥100 MeV Event Proton Warnings and Alerts

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
11/04/97 08:45	11/04/97 07:01*	11/04/97 10:51	104	TP	11/04/97 07:25*	–	–	–	FN	24	TP
11/06/97 13:05	11/06/97 12:15*	11/06/97 12:55	50	TP	11/06/97 12:45*	–	–	–	FN	30	TP
04/20/98 14:00*	–	–	–	FN	04/20/98 17:10	–	–	–	FN	–	FN
–	04/27/98 12:19*	04/27/98 19:00	–	FP	–	–	–	–	–	–	FP
05/02/98 14:20	05/02/98 14:05*	05/02/98 18:00	15	TP	05/02/98 14:05*	–	–	–	FN	0	TN
05/06/98 08:35	05/06/98 08:25*	05/06/98 12:00	10	TP	05/06/98 08:30*	–	–	–	FN	5	TP
08/24/98 23:55	08/24/98 23:22*	08/24/98 23:30	33	TP	08/24/98 23:10*	08/24/98 23:22*	08/24/98 23:19	−12	FN	−12	FN
09/25/98 00:10*	–	–	–	FN	–	–	–	–	–	–	FN
09/30/98 15:25	09/30/98 14:31*	09/30/98 16:31	54	TP	09/30/98 14:40*	09/30/98 14:36	09/30/98 15:36	4	TP	9	TP
11/08/98 02:45*	–	–	–	FN	–	–	–	–	–	–	FN
11/14/98 08:10	–	–	–	FN	11/14/98 07:55*	–	–	–	FN	–	FN
–	11/22/98 07:50*	11/22/98 12:00	–	FP	–	11/22/98 07:50*	11/22/98 11:00	–	FP	–	FP
01/23/99 11:05*	–	–	–	FN	–	–	–	–	–	–	FN
04/24/99 18:40*	04/24/99 17:42*	04/24/99 18:00	58	TP	–	–	–	–	–	58	TP
–	05/03/99 08:50*	05/03/99 15:00	–	FP	–	–	–	–	–	–	FP
05/05/99 18:20*	–	–	–	FN	–	–	–	–	–	–	FN

Table B1
Continued

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
–	05/27/99 12:55*	05/27/99 14:00	–	FP	–	–	–	–	–	–	FP
06/02/99 02:45*	–	–	–	FN	–	–	–	–	–	–	FN
06/04/99 09:25*	–	–	–	FN	–	–	–	–	–	–	FN
–	01/18/00 19:37*	01/19/00 06:00	–	FP	–	–	–	–	–	–	FP
02/18/00 11:30*	02/17/00 21:43*	02/18/00 02:00	827	TP	–	–	–	–	–	827	TP
–	03/02/00 09:23*	03/02/00 10:00	–	FP	–	–	–	–	–	–	FP
04/04/00 20:55*	–	–	–	FN	–	–	–	–	–	–	FN
–	05/15/00 09:20*	05/15/00 09:20	–	FP	–	–	–	–	–	–	FP
06/07/00 13:35*	06/06/00 15:52*	06/06/00 17:00	1,303	TP	–	–	–	–	–	1,303	TP
06/10/00 18:05	–	–	–	FN	06/10/00 17:50*	–	–	–	FN	–	FN
–	07/11/00 14:40*	07/11/00 14:38	–	FP	–	–	–	–	–	–	FP
–	07/12/00 12:45*	07/12/00 13:00	–	FP	–	–	–	–	–	–	FP
07/14/00 10:50	07/14/00 10:50*	07/14/00 10:55	0	FN	07/14/00 10:40*	07/14/00 10:51	07/14/00 10:55	−11	FN	−10	FN
–	–	–	–	???	–	–	–	–	–	–	FN
07/28/00 10:50*	07/28/00 04:00*	07/28/00 04:05	410	TP	–	–	–	–	–	410	TP
08/11/00 16:50*	–	–	–	FN	–	–	–	–	–	–	FN
09/12/00 15:55*	09/12/00 14:38*	09/12/00 14:37	77	TP	–	–	–	–	–	77	TP
10/16/00 11:25*	10/16/00 09:38*	10/16/00 09:40	107	TP	–	–	–	–	–	107	TP
10/26/00 00:45*	10/25/00 18:43*	10/25/00 18:42	362	TP	–	–	–	–	–	362	TP
11/08/00 23:50*	11/08/00 23:53*	11/09/00 00:00	−3	FN	11/08/00 23:50*	11/08/00 23:56	11/09/00 00:00	−6	FN	−3	FN
11/24/00 15:20*	11/24/00 07:26*	11/24/00 08:00	474	TP	11/24/00 17:20	11/24/00 17:19	11/24/00 17:30	1	TP	474	TP
11/24/00 15:20*	EXTENDED	EXTENDED	–	–	11/26/00 16:40	11/26/00 14:15	11/26/00 14:15	145	TP	–	–
01/28/01 20:25*	01/28/01 19:53*	01/28/01 19:51	32	TP	–	–	–	–	–	32	TP
03/29/01 16:35*	03/29/01 13:52*	03/29/01 15:00	163	TP	–	–	–	–	–	163	TP
04/02/01 23:40*	04/02/01 12:56*	04/02/01 13:00	644	TP	04/03/01 01:20	04/03/01 00:35	04/03/01 00:45	45	TP	644	TP
–	04/09/01 16:40*	04/09/01 18:00	–	FP	–	–	–	–	–	–	FP
04/10/01 08:50*	04/10/01 08:46*	04/10/01 08:44	4	TP	–	–	–	–	–	4	TP
04/10/01 08:50*	EXTENDED	EXTENDED	–	–	04/12/01 13:05	04/12/01 12:07	04/12/01 12:10	58	TP	–	–
04/15/01 14:10	04/15/01 14:08	04/15/01 14:08	2	TP	04/15/01 14:05*	04/15/01 14:07*	04/15/01 14:06	−2	FN	−2	FN
04/18/01 03:15	04/18/01 03:47	04/18/01 03:50	−32	FN	04/18/01 02:55*	04/18/01 03:11*	04/18/01 03:12	−16	FN	−16	FN
–	04/26/01 14:10*	04/26/01 21:00	–	FP	–	–	–	–	–	–	FP
04/28/01 04:30*	04/28/01 03:40*	04/28/01 03:45	50	TP	–	–	–	–	–	50	TP
05/07/01 19:15*	05/07/01 15:09*	05/07/01 15:08	246	TP	–	–	–	–	–	246	TP
06/15/01 17:50*	06/15/01 19:39*	06/15/01 19:37	−109	FN	–	–	–	–	–	−109	FN
08/10/01 10:20*	08/10/01 09:32*	08/10/01 10:00	48	TP	–	–	–	–	–	48	TP
08/16/01 01:35	08/16/01 01:32	08/16/01 01:35	3	TP	08/16/01 01:05*	08/16/01 01:03*	08/16/01 01:05	2	TP	2	TP
–	08/25/01 18:21*	08/26/01 08:00	–	FP	–	–	–	–	–	–	FP
09/15/01 14:35*	09/15/01 13:57*	09/15/01 14:15	38	TP	–	–	–	–	–	38	TP
09/24/01 12:15*	09/24/01 12:02*	09/24/01 12:15	13	TP	09/24/01 14:40	09/24/01 13:41	09/24/01 13:45	59	TP	13	TP
10/01/01 02:55*	10/01/01 02:22*	10/01/01 02:25	33	TP	–	–	–	–	–	33	TP
–	10/19/01 05:02*	10/19/01 05:15	–	FP	–	–	–	–	–	–	FP

Table B1
Continued

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
10/19/01 22:25*	10/19/01 17:14*	10/19/01 17:30	311	TP	–	–	–	–	–	311	TP
10/22/01 19:10*	10/22/01 17:22*	10/22/01 18:00	108	TP	–	–	–	–	–	108	TP
11/04/01 17:05	11/04/01 17:01*	11/04/01 17:00	4	TP	11/04/01 16:50*	11/04/01 21:28	11/04/01 21:26	−278	FN	−11	FN
11/19/01 12:30*	11/19/01 05:43*	11/19/01 08:00	407	TP	–	–	–	–	–	407	TP
11/22/01 23:20	11/22/01 21:42*	11/22/01 21:45	98	TP	11/22/01 22:50*	11/22/01 21:44	11/22/01 21:45	66	TP	68	TP
12/26/01 06:23	12/26/01 06:16	12/26/01 06:15	7	TP	12/26/01 06:13*	12/26/01 06:02*	12/26/01 06:02	11	TP	11	TP
12/29/01 05:10*	EXTENDED	EXTENDED	–	FN	–	–	–	–	–	–	–
12/30/01 02:45*	12/30/01 02:56*	12/30/01 03:00	−11	FN	–	–	–	–	–	−11	–
12/31/01 00:15*	12/31/01 00:10*	12/31/01 00:15	5	TP	–	–	–	–	–	5	–
01/10/02 20:45*	01/10/02 21:32*	01/10/02 21:31	−47	FN	–	–	–	–	–	−47	–
01/15/02 13:20*	01/15/02 11:17*	01/15/02 11:16	123	TP	–	–	–	–	–	123	–
–	01/27/02 15:22*	01/27/02 15:22	–	FP	–	–	–	–	–	–	–
02/20/02 07:30*	02/20/02 07:43*	02/20/02 07:43	−13	FN	–	–	–	–	–	−13	–
03/17/02 08:20*	03/17/02 07:49*	03/17/02 09:00	31	TP	–	–	–	–	–	31	–
03/18/02 13:00*	03/18/02 13:14*	03/18/02 13:15	−14	FN	–	–	–	–	–	−14	–
03/18/02 13:00*	03/20/02 15:37*	03/20/02 15:38	–	???	–	–	–	–	–	–	–
03/22/02 20:20*	03/22/02 14:51*	03/22/02 15:00	329	TP	–	–	–	–	–	329	–
04/17/02 15:30*	04/17/02 13:12*	04/17/02 13:30	138	TP	–	–	–	–	–	138	–
04/21/02 02:25	04/21/02 01:37*	04/21/02 01:40	48	TP	04/21/02 01:55*	04/21/02 01:57	04/21/02 02:00	−2	FN	18	–
05/22/02 17:55*	05/22/02 17:13*	05/22/02 17:15	42	TP	–	–	–	–	–	42	–
07/07/02 18:30*	07/07/02 16:00*	07/07/02 16:00	150	TP	–	–	–	–	–	150	–
07/16/02 17:50*	07/15/02 20:35*	07/16/02 02:00	1,275	TP	–	–	–	–	–	1,275	–
07/19/02 10:50*	EXTENDED	EXTENDED	–	FN	–	07/19/02 09:07	07/19/02 09:10	–	FP	–	–
07/22/02 06:55*	07/22/02 06:32*	07/22/02 06:35	23	TP	–	–	–	–	–	23	–
08/14/02 09:00*	08/14/02 09:33*	08/14/02 09:35	−33	FN	–	–	–	–	–	−33	–
08/22/02 04:55	08/22/02 04:21	08/22/02 04:30	34	TP	08/22/02 03:57*	08/22/02 03:40*	08/22/02 03:50	17	TP	17	–
08/24/02 01:40	08/24/02 01:39	08/24/02 01:39	1	TP	08/24/02 01:30*	08/24/02 01:31*	08/24/02 01:31	−1	FN	−1	–
–	09/06/02 15:57*	09/06/02 16:00	–	FP	–	–	–	–	–	–	–
09/07/02 04:40*	09/07/02 04:59*	09/07/02 04:59	−19	FN	–	–	–	–	–	−19	–
11/09/02 19:20*	11/09/02 18:31*	11/09/02 19:00	49	TP	–	–	–	–	–	49	–
–	12/20/02 00:07*	12/20/02 00:30	–	FP	–	–	–	–	–	–	–
–	05/27/03 23:30*	05/27/03 23:50	–	FP	–	–	–	–	–	–	–
05/28/03 23:35*	05/28/03 02:33*	05/28/03 04:00	1,262	TP	–	–	–	–	–	1,262	–
05/31/03 04:40*	05/31/03 05:26*	05/31/03 05:26	−46	FN	–	–	–	–	–	−46	–
06/18/03 20:50*	06/18/03 15:55*	06/18/03 17:00	295	TP	–	–	–	–	–	295	–
10/26/03 18:25*	10/26/03 18:11*	10/26/03 18:11	14	TP	–	10/26/03 18:15	10/26/03 18:15	–	FP	14	–
10/26/03 18:25*	EXTENDED	EXTENDED	–	–	–	10/27/03 03:15	10/27/03 03:15	–	FP	–	–
10/28/03 12:15	10/28/03 12:01	10/28/03 12:01	14	TP	10/28/03 11:50*	10/28/03 11:46*	10/28/03 11:47	4	TP	4	–
11/02/03 11:05*	11/02/03 00:56*	11/02/03 01:25	609	TP	11/02/03 17:40	11/02/03 17:53	11/02/03 17:53	−13	FN	609	–
11/02/03 11:05*	EXTENDED	EXTENDED	–	–	11/05/03 05:20	11/04/03 19:59	11/04/03 20:00	561	TP	–	–
–	11/20/03 09:09*	11/20/03 10:00	–	FP	–	–	–	–	–	–	–

Table B1
Continued

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
11/21/03 23:55*	11/21/03 20:11*	11/21/03 20:30	224	TP	–	–	–	–	–	224	–
12/02/03 15:05*	12/02/03 15:17*	12/02/03 15:17	−12	FN	–	–	–	–	–	−12	–
04/11/04 11:35*	04/11/04 11:10*	04/11/04 11:10	25	TP	–	–	–	–	–	25	–
07/25/04 18:55*	07/25/04 17:59*	07/25/04 18:30	56	TP	–	–	–	–	–	56	–
09/13/04 20:11*	09/13/04 21:44*	09/13/04 21:45	−93	FN	–	–	–	–	–	−93	–
09/19/04 19:25*	09/19/04 18:42*	09/19/04 19:00	43	TP	–	–	–	–	–	43	–
11/01/04 07:03	11/01/04 06:58*	11/01/04 07:00	5	TP	11/01/04 06:41*	–	–	–	FN	−17	–
11/07/04 19:10*	11/07/04 18:28*	11/07/04 18:29	42	TP	–	–	–	–	–	42	–
11/07/04 19:10*	EXTENDED	EXTENDED	–	–	11/10/04 03:20	11/10/04 03:33	11/10/04 03:33	−13	FN	–	–
–	01/15/05 08:55*	01/15/05 09:00	–	FP	–	–	–	–	–	–	–
01/16/05 02:10*	EXTENDED	EXTENDED	–	FN	–	–	–	–	–	–	–
01/16/05 02:10*	EXTENDED	EXTENDED	–	–	01/17/05 12:15	01/17/05 12:21	01/17/05 12:22	−6	FN	–	–
01/16/05 02:10*	EXTENDED	EXTENDED	–	–	01/20/05 07:01	01/20/05 18:45	01/20/05 18:45	−704	FN	–	–
05/14/05 05:25*	05/14/05 03:29*	05/14/05 03:50	116	TP	–	–	–	–	–	116	–
06/16/05 22:00	06/16/05 20:57*	06/16/05 21:15	63	TP	06/16/05 21:25*	06/16/05 21:15	06/16/05 21:30	10	TP	28	–
07/14/05 02:45*	07/14/05 01:43*	07/14/05 02:30	62	TP	–	–	–	–	–	62	–
NO ALERT	EXTENDED	EXTENDED	–	FN	–	–	–	–	–	–	–
NO ALERT	EXTENDED	EXTENDED	–	FN	–	–	–	–	–	–	–
07/27/05 23:00*	07/27/05 22:56*	07/27/05 23:05	4	TP	–	–	–	–	–	4	–
08/22/05 20:40*	08/22/05 19:38*	08/22/05 23:00	62	TP	–	–	–	–	–	62	–
09/08/05 02:15*	09/08/05 00:53*	09/08/05 02:00	82	TP	09/08/05 04:05	09/08/05 02:42	09/08/05 03:40	83	TP	82	–
09/14/05 01:00*	09/13/05 21:47*	09/13/05 21:55	193	TP	–	–	–	–	–	193	–
12/06/06 15:55*	12/06/06 09:45*	12/06/06 09:45	370	TP	12/07/06 01:15	12/07/06 00:09	12/07/06 00:15	66	TP	370	–
12/13/06 03:10	12/13/06 03:10	12/13/06 03:10	0	FN	12/13/06 03:00*	12/13/06 03:01*	12/13/06 03:01	−1	FN	−1	–
12/13/06 03:10	EXTENDED	EXTENDED	–	–	12/14/06 22:55	12/14/06 22:52	12/14/06 22:52	3	TP	–	–
08/14/10 12:30*	08/14/10 12:06*	08/14/10 12:07	24	TP	–	–	–	–	–	24	–
–	08/18/10 08:57*	08/18/10 09:25	–	FP	–	–	–	–	–	–	–
03/08/11 01:20*	03/07/11 22:18*	03/07/11 22:30	182	TP	–	–	–	–	–	182	–
03/21/11 19:50*	03/21/11 08:01*	03/21/11 08:30	709	TP	–	–	–	–	–	709	–
06/07/11 08:05	06/07/11 07:45	06/07/11 07:44	20	TP	06/07/11 07:35*	06/07/11 07:43*	06/07/11 07:42	−8	FN	−8	–
–	06/17/11 08:17*	06/17/11 08:20	–	FP	–	–	–	–	–	–	–
08/04/11 06:35	08/04/11 05:03	08/04/11 05:30	92	TP	08/04/11 05:10*	08/04/11 04:59*	08/04/11 04:59	11	TP	11	–
08/09/11 08:45	08/09/11 08:32	08/09/11 08:32	13	TP	08/09/11 08:40*	08/09/11 08:31*	08/09/11 08:31	9	TP	9	–
–	09/07/11 04:42*	09/07/11 05:00	–	FP	–	–	–	–	–	–	–
09/23/11 22:55*	09/23/11 04:48*	09/23/11 06:00	1,087	TP	–	–	–	–	–	1,087	–
10/23/11 15:05*	10/23/11 15:02*	10/23/11 14:55	3	TP	–	–	–	–	–	3	–
11/26/11 11:25*	11/26/11 11:19*	11/26/11 11:20	6	TP	–	–	–	–	–	6	–
01/23/12 05:30	01/23/12 05:01	01/23/12 05:00	29	TP	01/23/12 04:49*	01/23/12 04:51*	01/23/12 04:49	−2	FN	−2	–
01/27/12 19:05	01/27/12 18:29*	01/27/12 18:26	36	TP	01/27/12 19:00*	01/27/12 18:56	01/27/12 19:00	4	TP	31	–
–	03/05/12 15:17*	03/05/12 21:00	–	FP	–	–	–	–	–	–	–
03/07/12 05:10	03/07/12 00:19*	03/07/12 00:30	291	TP	03/07/12 04:05*	03/07/12 02:56	03/07/12 03:00	69	TP	226	–

Table B1
Continued

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
03/13/12 07:45*	03/13/12 08:00*	03/13/12 08:00	−15	FN	–	–	–	–	–	−15	–
03/13/12 18:10*	03/13/12 08:00*	03/13/12 08:00	610	TP	03/13/12 18:17	03/13/12 18:18	03/13/12 18:17	−1	FN	610	–
05/17/12 02:55	05/17/12 02:55*	05/17/12 02:55	0	FN	05/17/12 02:52*	05/17/12 02:56	05/17/12 02:55	−4	FN	−3	–
05/27/12 05:05*	05/26/12 23:46*	05/26/12 23:46	319	TP	–	–	–	–	–	319	–
06/16/12 19:55*	06/16/12 17:16*	06/16/12 17:15	159	TP	–	–	–	–	–	159	–
07/07/12 04:00*	07/07/12 02:55*	07/07/12 02:55	65	TP	–	–	–	–	–	65	–
07/09/12 01:30*	07/08/12 20:19*	07/08/12 20:30	311	TP	–	07/08/12 22:28	07/08/12 23:00	–	FP	311	–
07/12/12 18:35*	07/12/12 17:34*	07/12/12 17:33	61	TP	–	07/12/12 17:40	07/12/12 17:39	–	FP	61	–
07/17/12 17:15*	07/17/12 16:25*	07/17/12 16:30	50	TP	–	–	–	–	–	50	–
07/23/12 15:45*	07/23/12 11:01*	07/23/12 11:00	284	TP	–	–	–	–	–	284	–
07/24/12 07:20*	07/24/12 07:48*	07/24/12 07:20	−28	FN	–	–	–	–	–	−28	–
09/01/12 13:35*	09/01/12 11:51*	09/01/12 11:52	104	TP	–	–	–	–	–	104	–
09/28/12 03:00*	09/28/12 01:47*	09/28/12 02:00	73	TP	–	–	–	–	–	73	–
–	12/15/12 01:58*	12/15/12 01:57	–	FP	–	–	–	–	–	–	–
03/16/13 19:40*	03/16/13 16:47*	03/16/13 16:45	173	TP	–	–	–	–	–	173	–
04/11/13 10:55	04/11/13 08:57*	04/11/13 09:15	118	TP	04/11/13 09:40*	04/11/13 09:32	04/11/13 09:35	8	TP	43	–
05/15/13 13:35*	05/15/13 12:33*	05/15/13 12:32	62	TP	–	–	–	–	–	62	–
05/22/13 14:20*	05/22/13 13:52*	05/22/13 13:55	28	TP	05/22/13 14:55	05/22/13 14:04	05/22/13 14:05	51	TP	28	–
06/23/13 20:10*	06/23/13 19:29*	06/23/13 19:30	41	TP	–	–	–	–	–	41	–
09/30/13 05:05*	09/30/13 04:46*	09/30/13 05:00	19	TP	–	–	–	–	–	19	–
–	10/28/13 02:33*	10/28/13 02:35	–	FP	–	–	–	–	–	–	–
–	11/19/13 13:04*	11/19/13 13:05	–	FP	–	–	–	–	–	–	–
12/28/13 21:50*	12/28/13 21:18*	12/28/13 21:20	32	TP	–	–	–	–	–	32	–
01/06/14 09:15	01/06/14 08:37	01/06/14 08:37	38	TP	01/06/14 08:35*	01/06/14 08:36*	01/06/14 08:35	−1	FN	−1	–
01/06/14 09:15	EXTENDED	EXTENDED	–	–	01/07/14 20:15	01/07/14 20:40	01/07/14 20:15	−25	FN	–	–
02/20/14 08:55*	02/20/14 08:55*	02/20/14 08:55	0	FN	–	–	–	–	–	0	–
02/25/14 13:55*	02/25/14 06:29*	02/25/14 08:30	446	TP	–	02/25/14 12:57	02/25/14 13:15	–	FP	446	–
04/18/14 15:25*	04/18/14 14:00*	04/18/14 14:00	85	TP	–	04/18/14 14:12	04/18/14 14:12	–	FP	85	–
–	09/06/14 07:36*	09/06/14 07:36	–	FP	–	–	–	–	–	–	–
09/11/14 02:40*	09/10/14 21:05*	09/10/14 21:15	335	TP	09/11/14 04:25	09/10/14 23:22	09/10/14 23:30	303	TP	335	–
–	10/24/14 22:04*	10/24/14 23:00	–	FP	–	–	–	–	–	–	–
–	11/02/14 20:50*	11/02/14 20:50	–	FP	–	–	–	–	–	–	–
–	12/23/14 11:24*	12/23/14 11:24	–	FP	–	–	–	–	–	–	–
–	03/16/15 08:01*	03/16/15 08:00	–	FP	–	–	–	–	–	–	–
–	05/12/15 06:37*	05/12/15 06:37	–	FP	–	–	–	–	–	–	–
06/18/15 11:35*	06/18/15 09:07*	06/18/15 09:10	148	TP	–	–	–	–	–	148	–
06/21/15 20:35*	06/21/15 19:07*	06/21/15 19:30	88	TP	–	–	–	–	–	88	–
06/26/15 02:30*	06/25/15 23:41*	06/25/15 23:45	169	TP	–	–	–	–	–	169	–
10/29/15 05:50	10/29/15 03:47*	10/29/15 04:30	123	TP	10/29/15 04:35*	10/29/15 04:30	10/29/15 04:45	5	TP	48	–
01/02/16 04:30*	01/02/16 01:03*	01/02/16 01:03	207	TP	–	–	–	–	–	207	–
07/14/17 09:00*	07/14/17 05:30*	07/14/17 05:30	210	TP	–	–	–	–	–	210	–

Table B1
Continued

≥10 MeV					≥100 MeV					Event	
Alert[a]	Warning issue[b]	Warning begin[c]	Lead time[d]	Cat.[e]	Alert[f]	Warning issue[g]	Warning begin[h]	Lead time[i]	Cat.[j]	Lead time[k]	Cat.[l]
09/05/17 00:38*	09/05/17 00:30*	09/05/17 00:30	8	TP	–	–	–	–	–	8	–
09/05/17 00:38*	EXTENDED	EXTENDED	–	–	–	09/06/17 13:02	09/06/17 13:05	–	FP	–	–
09/10/17 16:45	09/10/17 16:32*	09/10/17 16:30	13	TP	09/10/17 16:25*	09/10/17 16:32*	09/10/17 16:30	−7	FN	−7	–

Note. Columns a-c list times for the ≥10 MeV proton Alert and Warning products, column d lists the ≥10 MeV proton Warning lead time and column e lists the corresponding ≥10 MeV proton event categorization as a true positive (TP), false negative (FN) or false positive (FP). Columns f-h list times for the ≥100 MeV proton Alert and Warning products, column i lists the ≥100 MeV Warning lead time and column j lists the corresponding ≥100 MeV event categorization. Columns k and l list the overall Event lead time and categorization based on the first energy range to cross threshold and the corresponding first Warning issued. Asterisks indicate which Alert and Warning issue times were used for the overall Event lead time.

Abbreviations: FN, number of false negatives; FP, number of false positives; SWPC, Space Weather Prediction Center; TP, number of true positives.

Data Availability Statement

All data used in this study is publicly available. GOES observational data is available from the NOAA NCEI web archive https://www.ngdc.noaa.gov/stp/satellite/goes/. Archived NOAA SWPC forecasts used in this paper are also publicly available from NOAA NCEI at https://www.ngdc.noaa.gov/stp/spaceweather.html. V2 sunspot number data is publicly available from the World Data Center SILSO, Royal Observatory of Belgium (http://www.sidc.be/silso/datafiles).

Acknowledgments
Author H. M. Bain carried out this work while supported by the Cooperative Agreement award NA17OAR4320101. The authors would like to thank Christopher Balch and Robert Rutledge (NOAA/SWPC) for their helpful discussion, and the referees for their thoughtful comments, which helped improve the paper.

References

Balch, C. C. (1999). SEC proton prediction model: Verification and analysis. *Radiation Measurements*, *30*(3), 231–250. https://doi.org/10.1016/S1350-4487(99)00052-9

Balch, C. C. (2008). Updated verification of the space weather prediction center's solar energetic particle prediction model. *Space Weather*, *6*(1), S01001. https://doi.org/10.1029/2007SW000337

Beck, P., Latocha, M., Rollet, S., & Stehno, G. (2005). TEPC reference measurements at aircraft altitudes during a solar storm. *Advances in Space Research*, *36*(9), 1627–1633. https://doi.org/10.1016/j.asr.2005.05.035

Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. (2012). Toward reliable benchmarking of solar flare forecasting methods. *The Astrophysical Journal*, *747*(2), L41. https://doi.org/10.1088/2041-8205/747/2/l41

Bröcker, J., & Smith, L. A. (2007). Increasing the reliability of reliability diagrams. *Weather and Forecasting*, *22*(3), 651–661. https://doi.org/10.1175/WAF993.1

Dierckxsens, M., Tziotziou, K., Dalla, S., Patsou, I., Marsh, M. S., Crosby, N. B., et al. (2015). Relationship between Solar Energetic Particles and properties of flares and CMEs: Statistical analysis of Solar Cycle 23 events. *Solar Physics*, *290*(3), 841–874. https://doi.org/10.1007/s11207-014-0641-4

Flueck, J. A. (1987). A study of some measures of forecast verification. In *10th Conf. on Probability and Statistics in the Atmospheric Sciences, Edmonton*.

Hanssen, A. W., Kuipers, W. J. A., & Instituut, K. N. M. (1965). *On the relationship between the frequency of rain and various meteorological parameters : (with reference to the problem of objective forecasting)*. s-Gravenhage: Staatsdrukkerij- en Uitgeverijbedrijf.

Heidke, P. (1926). Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler*, *8*, 301–349. https://doi.org/10.1080/20014422.1926.11881138

Hill, F. (2018). The Global Oscillation Network Group facility—An example of research to operations in space weather. *Space Weather*, *16*(10), 1488–1497. https://doi.org/10.1029/2018SW002001

Howard, R. A., Moses, J. D., Vourlidas, A., Newmark, J. S., Socker, D. G., Plunkett, S. P., et al. (2008). Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI). *Space Science Reviews*, *136*(1–4), 67–115. https://doi.org/10.1007/s11214-008-9341-4

Jolliffe, I. T., & Stephenson, D. R. (2012). *Forecast verification: A practitioner's guide in atmospheric science* (Second ed.). The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England: Wiley-Blackwell. https://doi.org/10.1002/9781119960003

Kahler, S. W., Cliver, E. W., & Ling, A. G. (2007). Validating the proton prediction system (PPS). *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(1–2), 43–49. https://doi.org/10.1016/j.jastp.2006.06.009

Kahler, S. W., White, S. M., & Ling, A. G. (2017). Forecasting E > 50-MeV proton events with the proton prediction system (PPS). *Journal of Space Weather and Space Climate*, *7*, A27. https://doi.org/10.1051/swsc/2017025

Kress, B. T., Rodriguez, J. V., & Onsager, T. G. (2020). Chapter 20 - The GOES-R Space Environment In Situ Suite (SEISS): Measurement of energetic particles in Geospace. In S. J. Goodman, T. J. Schmit, J. Daniels, & R. J. Redmon (Eds.), *The GOES-R Series* (pp. 243–250). Elsevier. https://doi.org/10.1016/B978-0-12-814327-8.00020-2

Laurenza, M., Cliver, E. W., Hewitt, J., Storini, M., Ling, A. G., Balch, C. C., & Kaiser, M. L. (2009). A technique for short-term warning of solar energetic particle events based on flare location, flare size, and evidence of particle escape. *Space Weather*, *7*(4), S04008. https://doi.org/10.1029/2007SW000379

Leka, K. D., Park, S.-H., Kusano, K., Andries, J., Barnes, G., Bingham, S., et al. (2019). A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. *The Astrophysical Journal Supplement Series*, *243*(2), 36. https://doi.org/10.3847/1538-4365/ab2e12

Lemen, J. R., Title, A. M., Akin, D. J., Boerner, P. F., Chou, C., Drake, J. F., et al. (2012). The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Physics*, *275*(1–2), 17–40. https://doi.org/10.1007/s11207-011-9776-8

Luhmann, J. G., Ledvina, S. A., Odstrcil, D., Owens, M. J., Zhao, X. P., Liu, Y., & Riley, P. (2010). Cone model-based SEP event calculations for applications to multipoint observations. *Advances in Space Research*, *46*(1), 1–21. https://doi.org/10.1016/j.asr.2010.03.011

Marsh, M. S., Dalla, S., Dierckxsens, M., Laitinen, T., & Crosby, N. B. (2015). SPARX: A modeling system for Solar Energetic Particle Radiation Space Weather forecasting. *Space Weather*, *13*(6), 386–394. https://doi.org/10.1002/2014SW001120

McIntosh, P. S. (1990). The classification of sunspot groups. *Solar Physics*, *125*(2), 251–267. https://doi.org/10.1007/BF00158405

Park, S.-H., Leka, K. D., Kusano, K., Andries, J., Barnes, G., Bingham, S., et al. (2020). A comparison of flare forecasting methods. IV. Evaluating consecutive-day forecasting patterns. *The Astrophysical Journal*, *890*(2), 124. https://doi.org/10.3847/1538-4357/ab65f0

Peirce, C. S. (1884). The numerical measure of the success of predictions. *Science*, *ns-4*(93), 453–454. https://doi.org/10.1126/science.ns-4.93.453-a

Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. (2012). The Solar Dynamics Observatory (SDO). *Solar Physics*, *275*(1–2), 3–15. https://doi.org/10.1007/s11207-011-9841-3

Posner, A. (2007). Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather*, *5*(5), S05001. https://doi.org/10.1029/2006SW000268

Richardson, I. G., Mays, M. L., & Thompson, B. J. (2018). Prediction of Solar Energetic Particle event peak proton intensity using a simple algorithm based on CME speed and direction and observations of associated solar phenomena. *Space Weather*, *16*(11), 1862–1881. https://doi.org/10.1029/2018SW002032

Rodriguez, J. V., Krosschell, J. C., & Green, J. C. (2014). Intercalibration of GOES 8–15 solar proton detectors. *Space Weather*, *12*(1), 92–109. https://doi.org/10.1002/2013SW000996

Sauer, H. H. (1989). SEL monitoring of the Earth's energetic particle radiation environment. In A. C. Rester & J. I. Trombka (Ed.), High-energy radiation background in space (Vol. 186, pp. 216–221). https://doi.org/10.1063/1.38171

Schou, J., Scherrer, P. H., Bush, R. I., Wachter, R., Couvidat, S., Rabello-Soares, M. C., & Tomczyk, S. (2012). Design and ground calibration of the Helioseismic and Magnetic Imager (HMI) instrument on the Solar Dynamics Observatory (SDO). *Solar Physics*, *275*(1–2), 229–259. https://doi.org/10.1007/s11207-011-9842-2

Schrijver, C., & Siscoe, G. (2010). *Heliophysics: Space Storms and Radiation: Causes and Effects*. https://doi.org/10.1017/CBO9781139194532

Schwadron, N. A., Townsend, L., Kozarev, K., Dayeh, M. A., Cucinotta, F., Desai, M., & Squier, R. K. (2010). Earth-Moon-Mars Radiation Environment Module framework. *Space Weather*, *8*(10), S00E02. https://doi.org/10.1029/2009SW000523

Seaton, D. B., Darnel, J. M., Hsu, V., & Hughes, J. M. (2020). Chapter 18 - GOES-R Series Solar Dynamics. In S. J. Goodman, T. J. Schmit, J. Daniels, & R. J. Redmon (Eds.), *The GOES-R Series* (pp. 219–232). Elsevier. https://doi.org/10.1016/B978-0-12-814327-8.00018-4

Sharpe, M. A., & Murray, S. A. (2017). Verification of space weather forecasts issued by the Met Office Space Weather Operations Centre. *Space Weather*, *15*(10), 1383–1395. https://doi.org/10.1002/2017SW001683

Smart, D. F., & Shea, M. A. (1992). Modeling the time-intensity profile of solar flare generated particle fluxes in the inner heliosphere. *Advances in Space Research*, *12*(2–3), 303–312. https://doi.org/10.1016/0273-1177(92)90120-M

Toriumi, S., & Wang, H. (2019). Flare-productive active regions. *Living Reviews in Solar Physics*, *16*(1), 3. https://doi.org/10.1007/s41116-019-0019-7

Vasudevan, G., Shing, L., Mathur, D., Edwards, C., Shaw, M., Seaton, D., & Darnel, J. (2019). Design and on-orbit calibration of the Solar Ultraviolet Imager (SUVI) on the GOES-R Series weather satellite. In Z. Sodnik, N. Karafolas, & B. Cugny (Eds.), International conference on space optics icso 2018 (Vol. 11180, pp. 2769–2778). SPIE. https://doi.org/10.1117/12.2536196

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences*. Elsevier Academic Press.