Space Weather

RESEARCH ARTICLE
10.1029/2020SW002496

Benchmarking Forecasting Models for Space Weather Drivers

Richard J. Licata[1], **W. Kent Tobiska**[2], **and Piyush M. Mehta**[1]

[1]Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, WV, USA, [2]Space Environment Technologies, Pacific Palisades, CA, USA

Abstract Space weather indices are commonly used to drive operational forecasts of various geospace systems, including the thermosphere for mass density and satellite drag. The drivers serve as proxies for various processes that cause energy flow and deposition in the geospace system. Forecasts of neutral mass density are a major uncertainty in operational orbit prediction and collision avoidance for objects in low Earth orbit (LEO). For the strongly driven system, accuracy of space weather driver forecasts is crucial for operations. The High Accuracy Satellite Drag Model (HASDM) currently employed by the U.S. Air Force in an operational environment is driven by four solar and two geomagnetic proxies. Space Environment Technologies (SET) is contracted by the space command to provide forecasts for the drivers. This work performs a comprehensive assessment for the performance of the driver forecast models. The goal is to provide a benchmark for future improvements of the forecast models. Using an archived data set spanning 6 years and 15,000 forecasts across Solar Cycle 24, we quantify the temporal statistics of the model performance.

Key Points:
- Four solar ($F_{10.7}$, $S_{10.7}$, $M_{10.7}$, and $Y_{10.7}$) and two geomagnetic (*ap* and *Dst*) driver indices are used as inputs by the operational HASDM system
- Temporal statistics using 6 years of historical data for driver forecasts are developed
- Baseline for future developments within the community are provided

Correspondence to:
R. J. Licata,
rjlicata@mix.wvu.edu

Citation:
Licata, R. J., Tobiska, W. K., & Mehta, P. M. (2020). Benchmarking forecasting models for space weather drivers. *Space Weather*, *18*, e2020SW002496. https://doi.org/10.1029/2020SW002496

Received 6 MAR 2020
Accepted 29 SEP 2020
Accepted article online 2 OCT 2020

1. Introduction

Accurately quantifying mass density in the thermosphere remains a predicament for the community. The difficulty stems from the highly dynamic nature of the thermosphere, an environment driven by a number of factors ranging from solar extreme ultraviolet (EUV) and geomagnetic heating to gravity waves in the lower atmosphere. Emmert (2015) provides a thorough overview of the physical drivers and their effects on thermospheric density. Current capabilities limit our ability to predict satellites' trajectories with precision in an operational setting. During large solar and geomagnetic storms, operators struggle to locate many resident space objects, let alone have the means to predict their orbits (Berger et al., 2020). Many resources in the United States and abroad are devoted to tracking satellites and determining their orbits in order to protect humans and other assets in space.

HASDM (Storz et al., 2005) is an assimilative empirical model that uses a large batch of calibration satellites to make corrections to a density nowcast from the Jacchia-Bowman 2008 (JB2008) model (Bowman et al., 2008, 2012). The resulting density data cube is then propagated forward in time using driver forecasts supplied to JB2008. The solar forecasts are deterministic in nature, not stochastic. The data-driven geomagnetic forecasts for *Dst* are also deterministic, while the *ap* forecasts are a hybrid of human forecaster and geophysical model methods.

The JB2008 model provides neutral density in the thermosphere using global exospheric temperature equations driven by four solar indices/proxies to represent different solar heating sources (Bowman et al., 2008; Tobiska et al., 2008). From ISO 21348 (2007), an index is a measured indicator of level of activity, while a proxy is a surrogate for other physical processes. The four solar indices and proxies are all reported in solar flux units (10^{-22}W m^{-2} Hz^{-1}) which are denoted as sfu.

The $F_{10.7}$ proxy has a strong correlation to solar EUV irradiance which has led to its longtime use as a surrogate for solar EUV energy. However, $F_{10.7}$ has no physical relation to solar EUV irradiances. The $S_{10.7}$, $M_{10.7}$, and $Y_{10.7}$ indices and proxies were derived from actual solar irradiance measurements and scaled to $F_{10.7}$ magnitudes in the original JB2008. This has also allowed an ease of comparison between these disparate time series. $S_{10.7}$ is an index indicative of activity of the integrated 26- to 34-nm bandpass solar chromospheric EUV emission, which penetrates to the middle thermosphere and is absorbed by atomic oxygen. The

$M_{10.7}$ proxy is used as a surrogate for far ultraviolet (FUV) photospheric 160-nm Schumann-Runge Continuum emissions, which penetrate to the lower thermosphere and cause molecular oxygen dissociation. The fourth solar index is $Y_{10.7}$ which is a composite of X_{b10} and Lyman-alpha. This is a hybrid index of solar coronal 0.1- to 0.8-nm X-ray emissions and 121.6-nm Lyman-alpha, both of which penetrate to the mesosphere and participate in water chemistry. In order to forecast the solar indices/proxies, the Space Environment Technologies (SET) uses a linear predictive algorithm that captures persistence and recurrence (Tobiska, Bowman, & Bouwer, 2008). Specifically, this is the "TS_FCAST" subroutine in Interactive Data Language (IDL).

To capture the impact of geomagnetic activity, the model uses a synthesis of *ap* and *Dst* indices. The *ap* index is a measure of global geomagnetic activity derived from 12 observatories that fall between 48°N and 63°S in latitude (McClain & Vallado, 2001). The utilization of *ap* during quiet geomagnetic conditions results in low density errors, but *Dst* proves to be a more effective driver during storm times (Bowman et al., 2008). *Dst* is an index that represents the strength of the storm time ring current in the inner magnetosphere (Tobiska, Bowman, & Bouwer, 2008). Its forecast is generated using SET's *Anemomilos* algorithm, which provides a forecast with maximum prediction window of 6 days (Tobiska et al., 2013) using a data-driven deterministic algorithm. For further details on all of the JB2008 drivers, see Tobiska, Bowman, and Bouwer (2008) and ISO 14222 (2013).

In contrast to the other indices, the 3-hourly *ap* forecasts are actually interpolated values from the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center's (SWPC) *Kp* forecasts (https://www.swpc.noaa.gov/products/planetary-k-index). Additionally, they are generated from an ensemble of individual human forecasters' predictions informed by model output (University of Michigan's Geospace Model since 2017) (Haiducek et al., 2017; Singer, 2013; Steenburgh et al., 2014). This forecast only extends 3 days, so the value is set to zero for the last 3 days of each forecast window (Days 4–6). Even though SWPC only recently switched to using the Geospace Model, this data represent the official NOAA SWPC forecast and we use it as such.

Errors in these space weather driver forecasts cause errors in the resulting densities, therefore impairing satellite conjunction analyses. Bussy-Virat et al. (2018) recently performed a study to show the effects of driver uncertainty on the probability of collision between two space objects. In order to achieve this, the authors performed an analysis on 2 years of $F_{10.7}$ and *ap* forecast errors.

We expand upon the work of Bussy-Virat et al. (2018) by using (i) all solar and geomagnetic drivers that are used in operations, (ii) a large historical data set covering a period of 6 years, (iii) an extended forecast window of up to 6 days, and (iv) the initial driver values to characterize model performance as a function of the solar and geomagnetic activity. This expansion is performed to get a more complete picture of the legacy drivers and specifically analyze the performance of the driver forecasts that are directly fed to JB2008 and subsequently HASDM.

The following section introduces the techniques and thresholds that bin solar and geomagnetic drivers. The motivation for coarse binning is to understand the forecasts for general levels of solar and geomagnetic conditions. The basis of the need for further evaluation is the community effort led by the OSTP SWORM and SWAP (2019) activity to create a benchmark for upper atmosphere expansion. Our resulting work here presented for the first time shows the error in the existing baseline for operational thermospheric density driver specification and prediction used by the USAF HASDM system. Their output densities are then used for deriving the NORAD TLE sets. Below, the resulting uncertainty in the automated, operational forecast drivers is presented. Proceeding the results of the statistical analysis, we provide a very brief discussion on the impact of these uncertainties on orbit prediction and operations. A more comprehensive study was recently conducted by the authors (Licata et al., 2020).

2. Methodology

The proprietary SET algorithms automatically produce files every 3 hr generating updated 6-day forecasts for solar and geomagnetic indices and proxies. The forecasts are used with exclusive, restricted access by the USAF customer. This study is the first time that the metrics of the forecasts have been evaluated and made public. These forecasts have a temporal resolution of 3 hr. In addition, they archive the observed values for

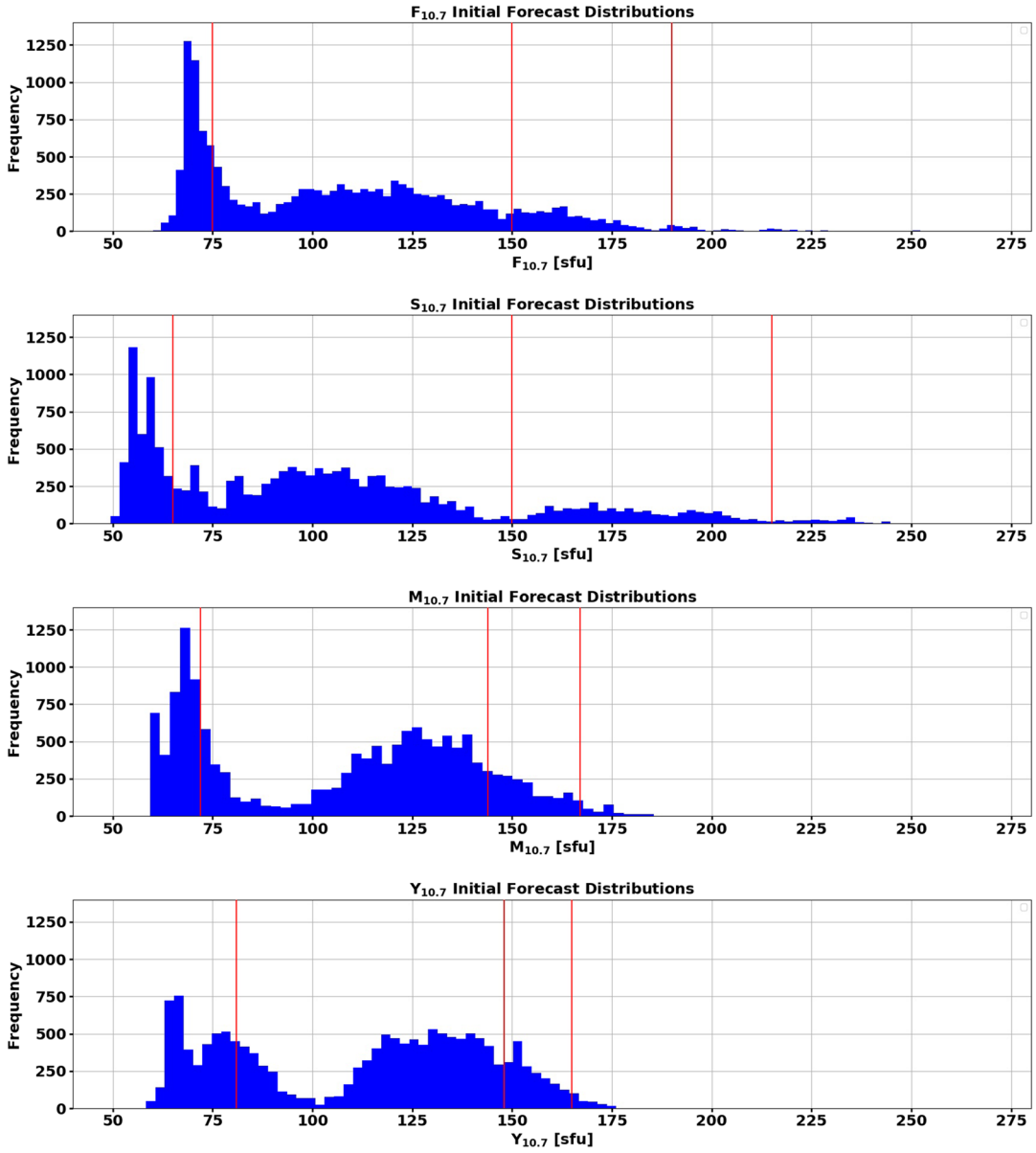

Figure 1. Distributions of initially forecasted values for each solar index with partitions shown in red.

each time step. To conduct this analysis, forecasts from October 2012 through the end of 2018 were used with the exception of a small number of missing/corrupted forecasts. In total, there were over 15,000 files to leverage for this study.

In order to effectively examine the solar and geomagnetic indices in comparable terms, a consistent approach had to be determined. To provide the clearest possible representation for all indices, different methods are used for solar indices/proxies and geomagnetic indices but kept consistent within each of the domains. Each index was split into separate subpopulations depending on the initial forecasted value. Populations that ended up with fewer than 100 forecasts are not shown, because there is insufficient data to draw statistical conclusions.

2.1. Solar Indices

The task of generating statistical results for the four solar indices investigated ($F_{10.7}$, $S_{10.7}$, $M_{10.7}$, and $Y_{10.7}$) was relatively straightforward. The forecasts are generated using SET's *SOLAR2000* algorithm (Tobiska, Bouwer, & Bowman, 2008; Tobiska et al., 2000). The thresholds to assess activity level for $F_{10.7}$ and *ap* have

Table 1
Activity Level Thresholds and Units for the Four Solar Indices

Solar index	Level	Value
$F_{10.7}$ (sfu)	Low	$F_{10.7} \leq 75$
	Moderate	$75 < F_{10.7} \leq 150$
	Elevated	$150 < F_{10.7} \leq 190$
	High	$F_{10.7} > 190$
$S_{10.7}$ (sfu)	Low	$S_{10.7} \leq 65$
	Moderate	$65 < S_{10.7} \leq 150$
	Elevated	$150 < S_{10.7} \leq 215$
	High	$S_{10.7} > 215$
$M_{10.7}$ (sfu)	Low	$M_{10.7} \leq 72$
	Moderate	$72 < M_{10.7} \leq 144$
	Elevated	$144 < M10.7 \leq 167$
	High	$M_{10.7} > 167$
$Y_{10.7}$ (sfu)	Low	$Y_{10.7} \leq 81$
	Moderate	$81 < Y_{10.7} \leq 148$
	Elevated	$148 < Y_{10.7} \leq 165$
	High	$Y_{10.7} > 165$

been described by Licata et al. (2019) and Mehta (2013) and are combined here with a supplementary statistical analysis for the remaining solar indices and proxies. Our objective in setting thresholds is to group data by general solar activity levels. Figure 1 depicts how the solar indices are distributed based on the initially forecasted value (1 day from forecast epoch), and Table 1 describes the solar activity levels.

Using these partitions on the 15,000+ forecasts resulted in a distinct number of individual $F_{10.7}$ forecasts for each activity level. These were used to classify the remaining solar indices and proxies, with the absence of a natural partition, or lull, in the distribution. A natural partition for $S_{10.7}$ can be seen at 150 sfu. This was chosen for that particular threshold as it did not greatly disrupt the number of forecasts in the adjacent activity levels since the goal was to have a similar number of forecasts across all solar indices and proxies for a given activity level. Peaks in the Figure 1 distribution data are a result of the natural distributions of solar activity estimated in a 3-hr cadence. Reading from right to left in the figures (high to low solar activity), the decline of Solar Cycle 24 from 2012 to 2018 is clearly portrayed and is the source of the predominantly bimodal distributions.

Figure 1 shows how the forecasts are distributed and that all activity levels have sufficient data to perform the following analyses. Note that the shapes of the distributions within each activity level are not indicative of the distributions of the forecast errors within them. The four levels of solar activity are defined in Table 1.

With each index's/proxy's forecast appropriately divided on its initial forecasted value, uncertainty distributions could be generated with respect to time from the forecast epoch. The uncertainty for the solar indices is defined as the error with respect to the issued (actual archival) value, normalized by the issued value. It is important to note that all errors shown (for both solar and geomagnetic indices) have a consistent sign convention. Positive percentages represent a forecasted value that was *more positive* than the issued (actual) value. For the solar indices and proxies, the error in solar flux units is also provided. All of the solar indices are updated daily, so there are 24 distributions for each (4 magnitude based and 6 temporal partitions).

2.2. Geomagnetic Indices

The analysis of the two geomagnetic indices, *ap* and *Dst*, was more intricate. Not only are the uncertainties functions of their magnitudes and time from epoch, they vary with solar activity level. To analyze *ap*, three geomagnetic activity levels were chosen: low, moderate, and active. In analyzing *Dst*, six geomagnetic activity levels were chosen and are consistent with the NOAA G-scale as operationally applied by SET. To allocate the geomagnetic forecasts, the largest value in the forecast for *ap* and the most negative value for *Dst* are the controlling factors. Figure 2 shows how the two geomagnetic indices are distributed based on these characteristics.

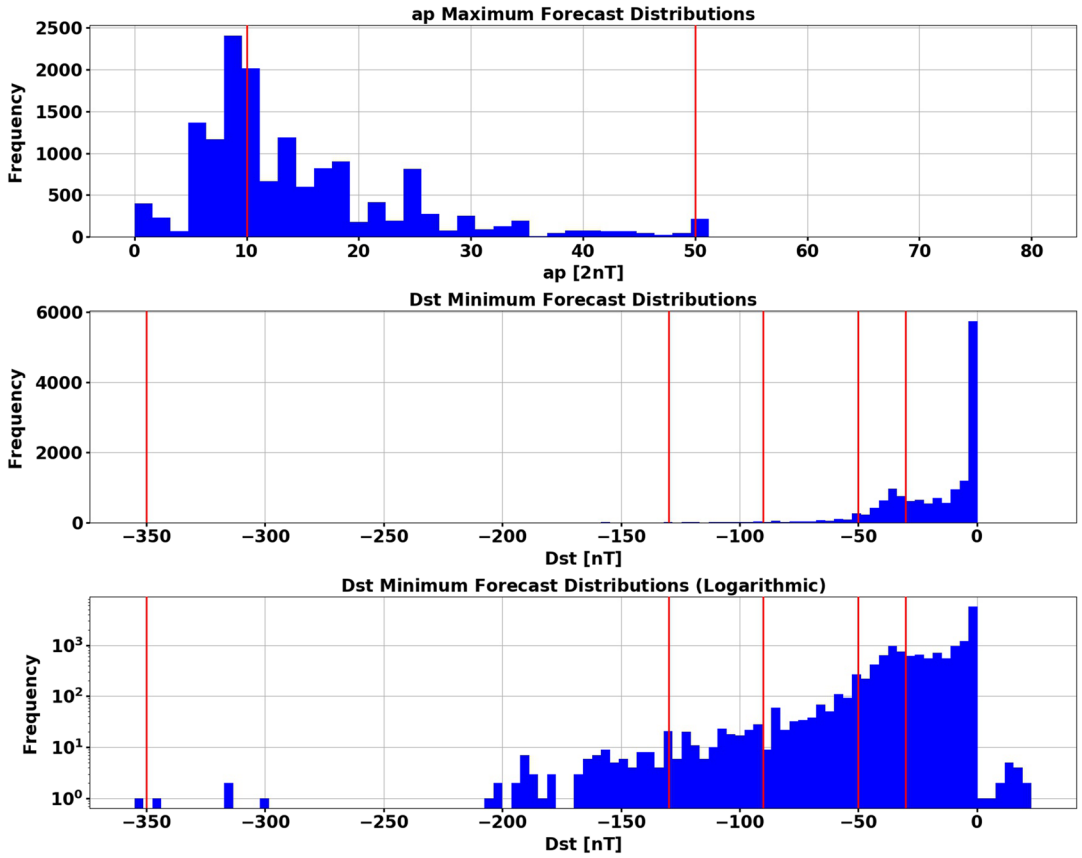

Figure 2. Distributions of initially forecasted values for the two geomagnetic indices with partitions shown in red. The *Dst* distribution is shown a second time with the frequency on a logarithmic scale for improved reading.

The *ap* distribution shows strong decay in forecast frequency with increasing *ap* values. There is a noticeable number of forecasts with a maximum value of 50 2nT which get classified as moderate geomagnetic activity, even though the histogram shows the bar in the next activity level. The *Dst* distribution had such a significant amount of minimum forecasted values at or near zero that the distribution is also shown in a log scale. This increases the visibility of areas with fewer forecasts. As previously noted, these distributions are not indicative of the forecast error distributions. Table 2 explicitly states the thresholds for *ap* and *Dst*.

In addition to the geomagnetic conditions, the forecast is classified by the initial forecasted $F_{10.7}$ value. Since the distributions have a finer temporal resolution and a solar dependency, there are 576 and 1,152

Table 2
Activity Level Thresholds and Units for Geomagnetic Activity, ap and Dst

Index	Level	Value
ap (2nT)	Low	$ap \leq 10$
	Moderate	$10 < ap \leq 50$
	Active	$ap > 50$
Dst (nT)	G0	$Dst \geq -30$
	G1	$-30 > Dst \geq -50$
	G2	$-50 > Dst \geq -90$
	G3	$-90 > Dst \geq -130$
	G4	$-130 > Dst \geq -350$
	G5	$Dst \leq -350$

Table 3
Distribution Statistics $F_{10.7}$ Error Distributions (Figure 3)

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	−0.2685	−0.7472	−0.6672	−0.3721	−0.0674	0.2428
	σ	3.6985	4.7031	5.5001	6.1683	6.7677	7.2050
	EBM	0.1126	0.1432	0.1675	0.1878	0.2061	0.2194
Moderate solar	μ	−0.8251	−0.8095	−0.9639	−1.1450	−1.1456	−1.1679
	σ	12.0854	14.9853	17.8425	20.2973	21.9353	23.3389
	EBM	0.2489	0.3086	0.3674	0.4180	0.4517	0.4806
Elevated solar	μ	5.7270	7.2425	9.0385	10.3829	11.0017	10.9559
	σ	18.3328	22.1021	25.2942	27.0774	27.5279	26.9074
	EBM	0.8572	1.0335	1.1827	1.2661	1.2872	1.2582
High solar	μ	15.7448	19.5749	24.2444	26.6674	26.0230	23.9778
	σ	20.2227	24.7236	27.6092	30.8869	33.9069	35.9795
	EBM	2.5639	3.1345	3.5003	3.9159	4.2988	4.5616

distributions for *ap* and *Dst*, respectively. The geomagnetic indices could be classified by other metrics instead of $F_{10.7}$, such as average thermospheric density, but $F_{10.7}$ was used since that was the most readily available additional metric.

It becomes difficult to generate a standard percent error normalized by the issued value, because the issued value can be small or even zero. Therefore, another method had to be chosen to provide a similar comparison. Instead of normalizing errors by the issued value, they are normalized by the absolute value of the long-term mean value of the index. Therefore, an error of −200% for *Dst* signifies an error twice the magnitude of the long-term mean *Dst*, and the prediction was more negative than the issued value. The long-term mean values for *ap* and *Dst* are 9.2 2nT and −8.8 nT, respectively. The time period to determine the mean values spans from October 2012 to the end of 2018, the same time period of the forecasts available for this study. Results are displayed in both normalized and absolute terms for the geomagnetic indices, as well.

The statistics provided in the proceeding section are the mean, standard deviation, and the error bound for the population mean (EBM). These are generated only for the forecast errors in the proxy's/index's units. Equations 1 and 2 show how the errors are computed in both absolute terms and in percentage form.

$$\text{Error} = forecast - issued \tag{1}$$

$$\text{Percent Error} = 100\% \cdot \frac{forecast - issued}{issued} \tag{2}$$

Table 4
Distribution Statistics $S_{10.7}$ Error Distributions (Figure 4)

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	0.2206	−0.0343	−0.1044	−0.0985	−0.0660	−0.0501
	σ	5.8378	5.7474	5.8682	6.0709	6.1153	6.0475
	EBM	0.1797	0.1769	0.1806	0.1869	0.1883	0.1862
Moderate solar	μ	0.1111	0.2190	0.2974	0.3426	0.2977	0.1862
	σ	10.0715	11.7059	13.5434	14.9844	16.1132	17.1467
	EBM	0.2102	0.2443	0.2827	0.3127	0.3363	0.3579
Elevated solar	μ	−1.1311	−1.7989	−2.2361	−2.6910	−2.8148	−2.7593
	σ	16.5592	19.2050	21.9271	24.1118	25.8017	27.2658
	EBM	0.7120	0.8257	0.9428	1.0367	1.1094	1.1723
High solar	μ	16.2040	23.1628	28.3943	31.4085	31.6158	30.1409
	σ	35.3267	39.6577	39.6471	38.7875	38.5307	38.6613
	EBM	4.4060	4.9458	4.9445	4.8373	4.8052	4.8215

Table 5
Distribution Statistics $M_{10.7}$ Error Distributions (Figure 5)

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	−0.9582	−1.1063	−1.2667	−1.4203	−1.5303	−1.6171
	σ	5.1317	5.6375	6.2250	6.4693	6.8538	7.2024
	EBM	0.1562	0.1716	0.1895	0.1969	0.2086	0.2192
Moderate solar	μ	−0.5138	−0.7424	−0.9976	−1.1619	−1.2370	−1.3348
	σ	8.9027	11.2745	13.4329	15.1962	16.4221	17.4041
	EBM	0.1835	0.2324	0.2768	0.3132	0.3384	0.3587
Elevated solar	μ	3.7282	5.6301	7.3375	8.5258	9.0192	8.9334
	σ	10.4528	13.4784	16.1156	17.9805	18.9175	19.0770
	EBM	0.4861	0.6269	0.7495	0.8363	0.8798	0.8872
High solar	μ	4.6517	6.3966	8.8328	10.9500	12.3435	12.9784
	σ	9.2079	12.7892	16.4932	19.2241	20.9130	21.5820
	EBM	1.1849	1.6457	2.1223	2.4738	2.6911	2.7772

As previously mentioned, if computing the statistics in normalized form (Equation 2) for the geomagnetic indices, the *issued* term in the denominator is replaced with the absolute value of the long-term mean value of the index. In order to account for the sample mean not perfectly representing the population mean, the 95% confidence EBM is provided which can be used to determine the 95% confidence interval for the population mean. This is shown in Equations 3 and 4. $CI_{95\%}, \bar{x}, \sigma$, and n represent the 95% confidence bounds, the sample mean, the standard deviation, and the number of samples, respectively.

$$EBM = Z_{95\%}\frac{\sigma}{\sqrt{n}} \tag{3}$$

$$CI_{95\%} = \bar{x} \pm EBM \tag{4}$$

The Z value that corresponds to 95% confidence is 1.9600, and in the proceeding tables, the EBM values are given with respect to the standard deviation in the respective units, not the normalized form. Other EBMs and confidence intervals can be easily computed using the corresponding Z value and the values in Tables 3–8.

3. Results

In the resulting uncertainty figures, the mean and standard deviation of forecast error (as a function of time from forecast epoch) are presented for each activity level. This way, biases can be identified and the

Table 6
Distribution Statistics $Y_{10.7}$ Error Distributions (Figure 6)

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	−0.4270	−0.2282	−0.0684	0.0902	0.2697	0.4650
	σ	4.5093	4.9373	5.4416	5.9977	6.4255	6.8160
	EBM	0.1372	0.1502	0.1656	0.1825	0.1955	0.2074
Moderate solar	μ	−0.8277	−1.3317	−1.7651	−2.1734	−2.5561	−2.9688
	σ	8.5137	10.1811	11.8475	13.0170	13.8737	14.8937
	EBM	0.1753	0.2096	0.2439	0.2680	0.2856	0.3066
Elevated solar	μ	2.1089	1.7243	2.0679	2.5656	2.7714	2.8423
	σ	7.7173	9.3066	10.6086	11.4955	11.8996	11.9965
	EBM	0.3593	0.4333	0.4939	0.5352	0.5541	0.5586
High solar	μ	5.2729	4.2075	4.5855	5.2131	5.6561	6.0140
	σ	8.5806	10.8097	11.5436	11.3048	11.0190	10.4416
	EBM	1.1497	1.4483	1.5466	1.5146	1.4761	1.3990

Table 7
Distribution Statistics for ap Error Distributions (Figure 8) in Units of 2nT

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	0.6782	0.4987	−1.2448	−6.4479	−7.0534	−6.6291
	σ	6.4053	6.8368	10.2313	10.6877	11.0401	9.4683
Low geomagnetic	*EBM*	0.2902	0.3098	0.4636	0.4843	0.5003	0.4290
Low solar	μ	2.1330	2.1653	3.0192	−7.1221	−6.8853	−7.5399
	σ	12.6072	12.7522	10.9242	10.6639	10.9207	11.9126
Moderate geomagnetic	*EBM*	0.5160	0.5220	0.4471	0.4365	0.4470	0.4876
Moderate solar	μ	−0.0492	−0.5465	−1.7532	−6.9294	−7.0114	−6.7315
	σ	6.8140	8.0992	12.1259	12.1145	11.2090	9.9162
Low geomagnetic	*EBM*	0.2258	0.2684	0.4019	0.4015	0.3715	0.3287
Moderate solar	μ	1.3877	1.3850	0.9151	−10.0225	−10.0000	−10.2871
	σ	16.7665	16.5156	14.9577	14.6091	14.9821	15.6026
Moderate geomagnetic	*EBM*	0.4444	0.4378	0.3965	0.3872	0.3971	0.4136
Elevated solar	μ	−0.2166	−0.5019	−1.3707	−6.4573	−5.6573	−5.0204
	σ	7.2675	7.6940	7.9520	9.0765	8.6062	6.2746
Low geomagnetic	*EBM*	0.5084	0.5382	0.5563	0.6350	0.6021	0.4389
Elevated solar	μ	2.9701	2.7038	1.2755	−7.7028	−7.6295	−8.0547
	σ	10.9766	12.0786	11.9995	10.0915	10.2440	10.7734
Moderate geomagnetic	*EBM*	0.6911	0.7605	0.7555	0.6354	0.6450	0.6783
High solar	μ	8.1667	5.2540	0.5079	−7.9921	−9.0397	−10.3968
	σ	15.2611	12.7671	8.9307	7.3996	8.1156	10.3549
Moderate geomagnetic	*EBM*	2.6647	2.2293	1.5594	1.2920	1.4171	1.8081

Note. Days 1–3 represent the error statistics for the actual forecasts, where Days 4–6 simply show background error that is a result of setting the forecast to zero.

algorithm's temporal uncertainty can be determined. Figure 3 shows the performance of the $F_{10.7}$ forecast algorithm, and Table 3 shows the statistics in sfu.

At low and moderate levels of solar activity, the $F_{10.7}$ algorithm is fairly unbiased. It is not until elevated and high solar activity that a bias accumulates, showing a tendency of overforecasting the proxy. The evolution of the error's standard deviation has an expected growth with time from epoch for all activity levels, showing the uncertainty of the forecast increasing with time. The algorithm performs well when the first forecasted $F_{10.7}$ value is below 150 sfu, which accounted for approximately 87% of the forecasts. This analysis points to needed improvements in $F_{10.7}$ prediction for periods of elevated and high solar activity. For moderate solar activity, the normalized error has a slightly positive bias where the bias is slightly negative when looking at the actual mean errors (to the right). This is caused by the range of this activity level (75 to 150 sfu). This shows that the algorithm is likely overforecasting $F_{10.7}$ toward the higher end of the activity level and underforecasting at the lower end. This would cause the normalized mean to rise relative to the actual mean errors. This is confirmed by the subplots for low and elevated solar activity where the algorithm is underforecasting and overforecasting, respectively. This analysis on the discrepancy between the normalized and actual mean errors is applicable to the remaining solar indices and proxies.

Figure 4 and Table 4 provide the algorithm performance for $S_{10.7}$. There is little bias through low, moderate, and elevated activity levels (over 98% of forecasts) displaying strong overall performance. The uncertainty at these activity levels is similar to $F_{10.7}$, but the performance at high solar activity is not as stable. For high solar activity, there is a dominant tendency to overforecast in addition to a large uncertainty. This is a by-product of the forecasting method. The uncertainty also does not consistently grow with time. Thus, $S_{10.7}$ prediction needs more attention for high solar activity periods.

The $F_{10.7}$ and $S_{10.7}$ algorithms are both vulnerable to high solar activity, but the comprehensive effectiveness is visible. The limitation during high activity is due to the volatility of the Sun during solar maximum, that is, the inability to accurately forecast flares and the lack of information from the solar East limb and solar

Table 8
Distribution Statistics for Dst Error Distributions (Figure 10) in Units of nT

Condition	Statistics	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days
Low solar	μ	1.1077	1.1851	1.8191	1.9479	4.7670	4.9067
	σ	14.6264	15.3676	16.0768	16.6027	15.2969	15.5632
G0	*EBM*	0.5654	0.5940	0.6214	0.6418	0.5913	0.6016
Moderate solar	μ	8.8130	9.1455	9.5665	9.8593	10.1396	10.2912
	σ	18.2244	18.3000	18.9761	19.0591	19.2045	19.1569
G0	*EBM*	0.4094	0.4111	0.4262	0.4281	0.4314	0.4303
Elevated solar	μ	9.6871	8.8767	8.0980	8.0500	8.9239	9.4235
	σ	14.8961	14.4229	14.4208	14.5799	15.5498	16.2407
G0	*EBM*	0.7751	0.7504	0.7503	0.7586	0.8091	0.8450
High solar	μ	15.9664	15.9076	13.1176	9.8571	6.0840	6.0420
	σ	16.9190	17.6202	18.2560	20.0190	19.3606	18.7587
G0	*EBM*	3.0399	3.1659	3.2801	3.5969	3.4786	3.3704
Low solar	μ	−24.6131	−24.1075	−23.7094	−23.1334	−0.0849	−0.0027
	σ	16.6175	16.1616	15.9475	15.9702	14.9486	14.9425
G1	*EBM*	0.8390	0.8160	0.8052	0.8063	0.7547	0.7544
Moderate solar	μ	−1.2407	−0.5186	0.1395	1.1314	8.9733	9.5651
	σ	27.9324	29.1335	26.8674	26.7461	21.9025	21.3862
G1	*EBM*	1.8669	1.9471	1.7957	1.7876	1.4639	1.4294
Elevated solar	μ	5.0405	1.6185	5.6069	8.3584	7.9191	7.9075
	σ	23.1907	21.5794	18.6853	16.0778	12.3314	15.6321
G1	*EBM*	3.4558	3.2157	2.7844	2.3959	1.8376	2.3294
Moderate solar	μ	−2.1743	0.0804	2.8874	7.8525	15.9303	16.6944
	σ	29.8673	28.2948	26.2478	23.7456	18.4655	17.0010
G2	*EBM*	3.0311	2.8715	2.6638	2.4098	1.8740	1.7253
Moderate solar	μ	−19.3704	−13.3611	−2.8704	5.1481	8.4444	12.3148
	σ	47.9640	45.4857	28.6925	20.9562	23.4494	16.9666
G3	*EBM*	9.0461	8.5787	5.4114	3.9524	4.4226	3.1999

farside active region's growth. The algorithms for the remaining indices prove to be more robust to solar activity. The $M_{10.7}$ performance is presented in Figure 5 and Table 5.

For $M_{10.7}$, there is a minimal bias of $\pm 2\%$ for the lower two activity levels, but the right subplots show that there is a slight tendency to underpredict. At elevated and high solar activity, the bias is accumulating with time and increases in intensity. Across all levels, the uncertainty starts below 4% and grows steadily with time. An interesting characteristic that contrasts the prior two indices is the lower uncertainty at high solar activity. The difference in performance is not drastic relative to the other conditions. Therefore, improvement in $M_{10.7}$ is needed for elevated and high solar activity periods.

To conclude the analysis of the solar indices, Figure 6 and Table 6 both show the performance of the $Y_{10.7}$ algorithm. Relative to the previous three indices, the $Y_{10.7}$ algorithm is considerably robust to activity levels and has less overall uncertainty. In the first two activity levels, the bias is less than $\pm 1\%$ for nearly the entire prediction window. The uncertainty grows with time for all activity levels, but its magnitude is less significant than the other indices. The bias never exceeds 5% and the uncertainty 12%.

As previously stated, the geomagnetic indices were more difficult to analyze due to an increase in dependencies and a finer time resolution. Each geomagnetic index has its own set of activity levels but are both also based on $F_{10.7}$ thresholds. The performance of the *ap* forecasts is shown in Figures 7 and 8 along with Table 7.

Unlike the solar indices, there are multiple conditions with insufficient data to conduct the analysis. The most distinct difference in the *ap* forecast performance, relative to the other indices, is the discontinuity at

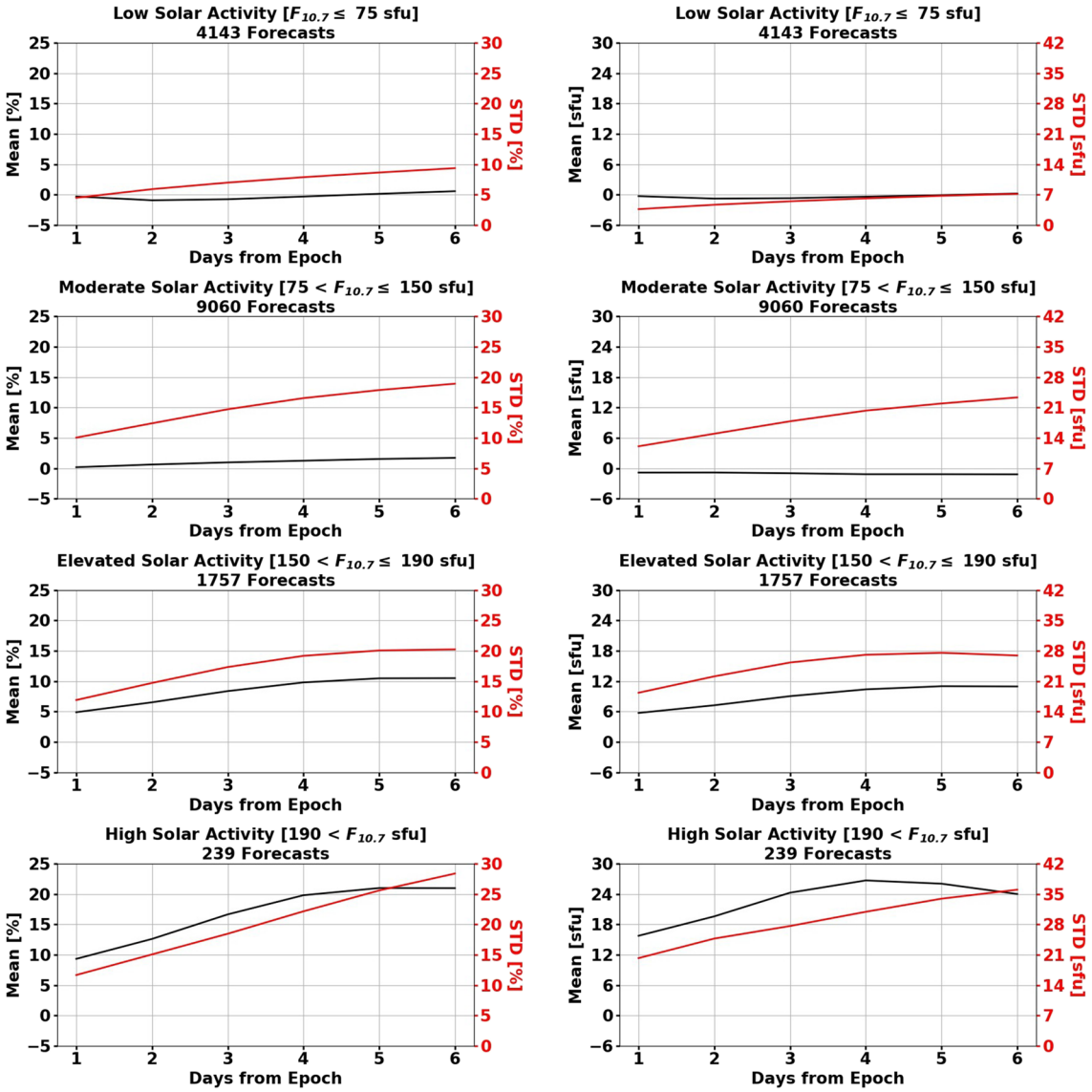

Figure 3. $F_{10.7}$ algorithm performance across four levels of solar activity with normalized error shown on the left and absolute error shown on the right.

the 3-day mark. Mentioned in section 1, the forecasts only have a 3-day prediction window. The forecasts are created by the ensemble of Space Weather forecasters at NOAA SWPC with the aid of the Geospace model.

Figures 7 and 8 show uncertainty results for a 6-day prediction window to be consistent with the other indices, even though SET sets every *ap* value to zero after 3 days. There are still interesting results in the latter 3 days of the forecasts across the different conditions. For example, the magnitude of underprediction (when *ap* is set to zero) is different for each condition as is the volatility of *ap*, shown by the standard deviation. Even so, the most important aspect of two figures is the first 3 days when forecasts are provided, and this figure presents a benchmark on the prediction accuracy by NOAA SWPC for *ap*.

During low geomagnetic activity (across all solar activity levels), there is no significant bias detected. With moderate geomagnetic activity, there is a general overprediction that decreases over the 3-day provided forecast. It shows a possible path for prediction improvement by relying on persistence when *ap* is high at the start of the forecasts. Another key determination is shown by the rightmost panels where there is only a single forecast that has a value greater than 50 2nT. This reflects the difficulty in quantifying the intensity of a storm, even with the aid of a physics-based model.

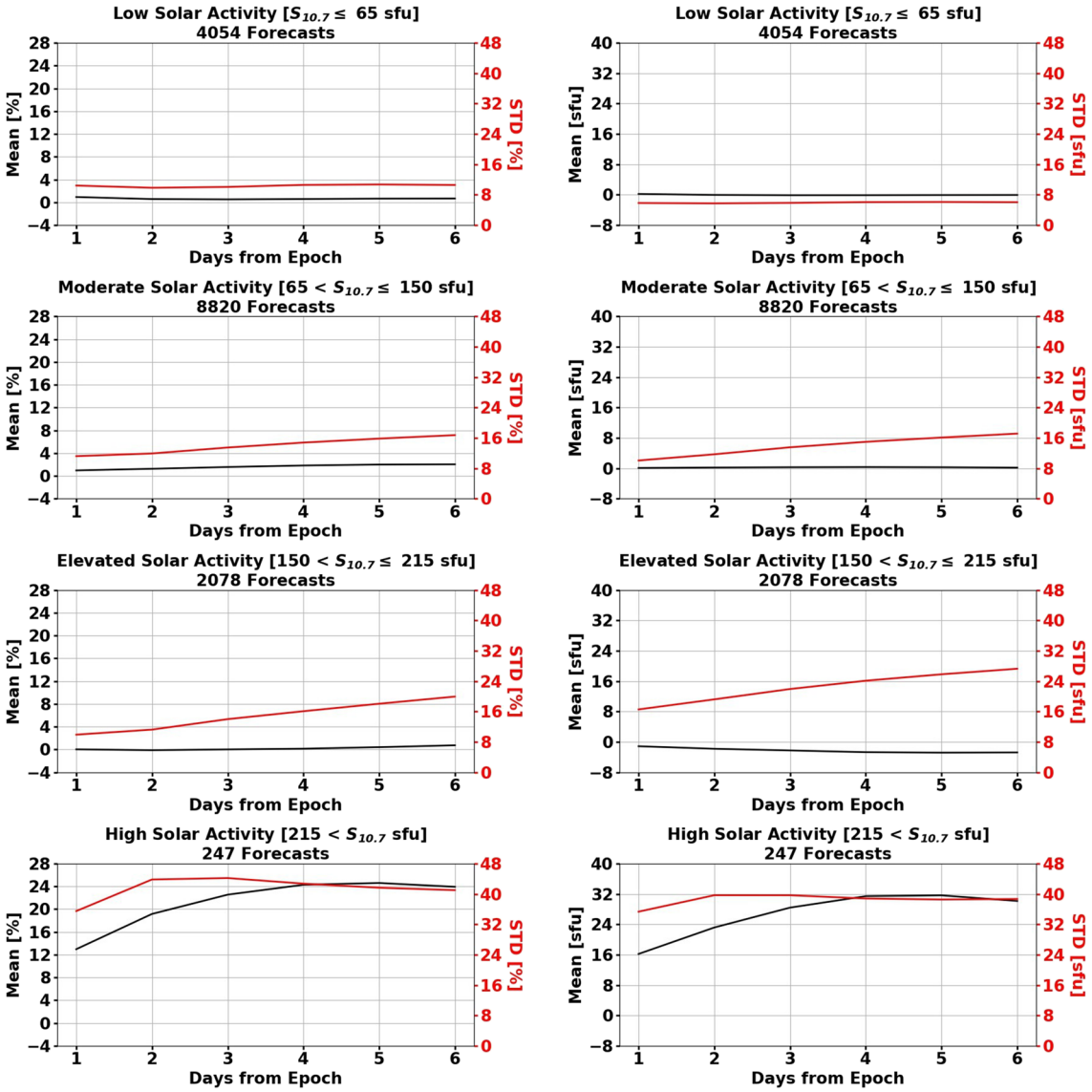

Figure 4. $S_{10.7}$ algorithm performance across four levels of solar activity with normalized error shown on the left and absolute error shown on the right.

The last algorithm analyzed is SET's *Anemomilos* for *Dst* forecasts, shown in Figures 9 and 10 along with Table 8. The G5 row is not shown since there was only a single forecast where a G5 storm was expected. There are only 9 of 24 conditions with enough forecasts to perform the analysis, but the remaining results provide insight to the strengths and weaknesses of the algorithm.

In the top left subplot (when conditions are quiet), the forecasts remain relatively unbiased, and the uncertainty slowly increases with time. Figure 10 shows a general tendency to predict *Dst* to be more positive for nearly all G0 and G1 conditions, with the exception of G1 low solar activity conditions. In this case, the algorithm has a strong bias to expect *Dst* to be ~23 nT more negative than the issued values over the first 4 days of the forecast. Following the strong inclination after Day 4, the algorithm tends to neutralize the bias. This is interpreted as accurate prediction of *Dst* recovery to quiet conditions but overprediction of the initial magnitude at the onset of the storm.

The bias for G1–G3 moderate solar activity conditions shows a strong temporal dependency transitioning from underprediction to overprediction in each case. G2 moderate solar activity is a case with a peculiar trend of the uncertainty decaying with time from epoch. The algorithm tends to miss the magnitude and start of events and then achieves recovery to background after the main phase of the storm too quickly. This

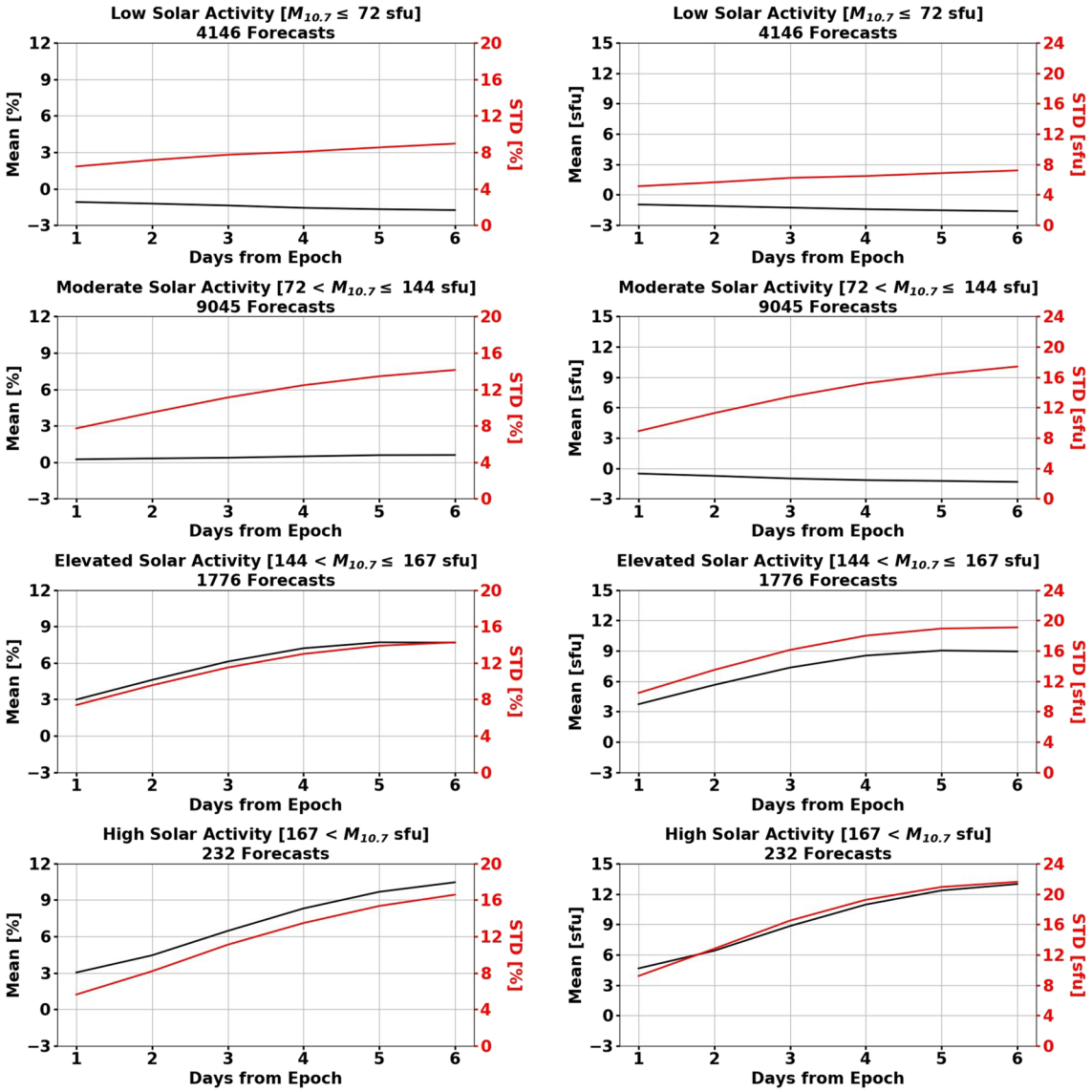

Figure 5. $M_{10.7}$ algorithm performance across four levels of solar activity with normalized error shown on the left and absolute error shown on the right.

is also the case for G3 moderate solar activity, which shows even more pronounced error in this direction. This prediction error points to a need for improvement in understanding the arrival timing, magnitude, and duration of events. A source of the *Dst* overprediction in G0–G3 conditions is that *Anemomilos* does not model (ignores) high-speed streams (HSS).

4. Operational Impacts

As seen in Figures 3–10 and Tables 3–8, the driver forecasting models and algorithms are imperfect. The operational driver forecasts do not provide quantified uncertainty, requiring statistical analyses to understand its effects on orbit propagation. A study was performed by Licata et al. (2019) to investigate this, and examples of the results are shown in Figure 11.

The probabilistic $F_{10.7}$ forecasts in Figure 11 were generated using data in the current study. There was a constraint of the maximum change in the driver ($dF_{10.7}$) from one time step to the next. This limiting factor was chosen through further statistical analyses. In general, all forecasts were within the 95% confidence level, but it is clear that there is a strong tendency to overestimate the $F_{10.7}$ predicted magnitudes. Each driver forecast was additionally input to a quasi-physical model of the mass density built using recurrent neural

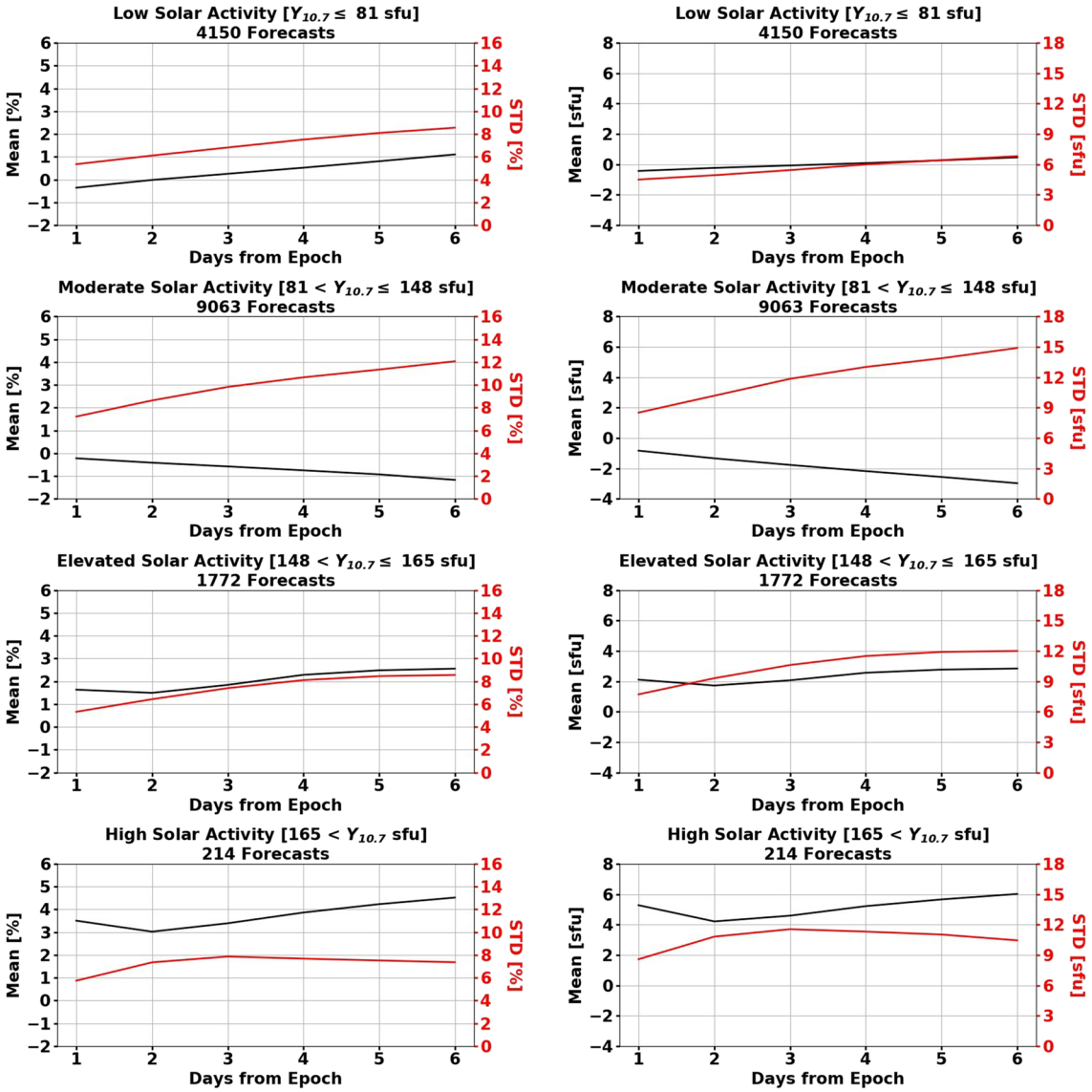

Figure 6. $Y_{10.7}$ algorithm performance across four levels of solar activity with normalized error shown on the left and absolute error shown on the right.

network to forecast a resulting 3-D density grid that would be used in orbit propagation to understand the effects of prediction error on densities (Licata & Mehta, 2019, 2020; Mehta et al., 2018). The satellite position distributions after 6 days (relative to the position with the true drivers) give light to the need for probabilistic approaches in determining satellites' orbits. In this fairly quiet case, there was an ~25.5-km position error with deterministic approaches. Probabilistic forecasting allows for the true position to be captured through the analysis. This is seen in both the radial and in-track position difference distributions with zero error being covered in the tails. In this case, the mean probabilistic in-track position (~22 km) was more accurate than the deterministic position.

In addition to driver uncertainty, the choice of density model is important in estimating collision probability. Each model will have different sensitivities to the driver variations in addition to underlying biases. Therefore, it becomes important to study the impacts of multiple uncertainty sources (Licata et al., 2020).

5. Conclusions

The analysis of the SET algorithms used by the JB2008 and HASDM models provided clear performance baselines for the current state-of-the-art of operational, automated density model driver forecasts. This work

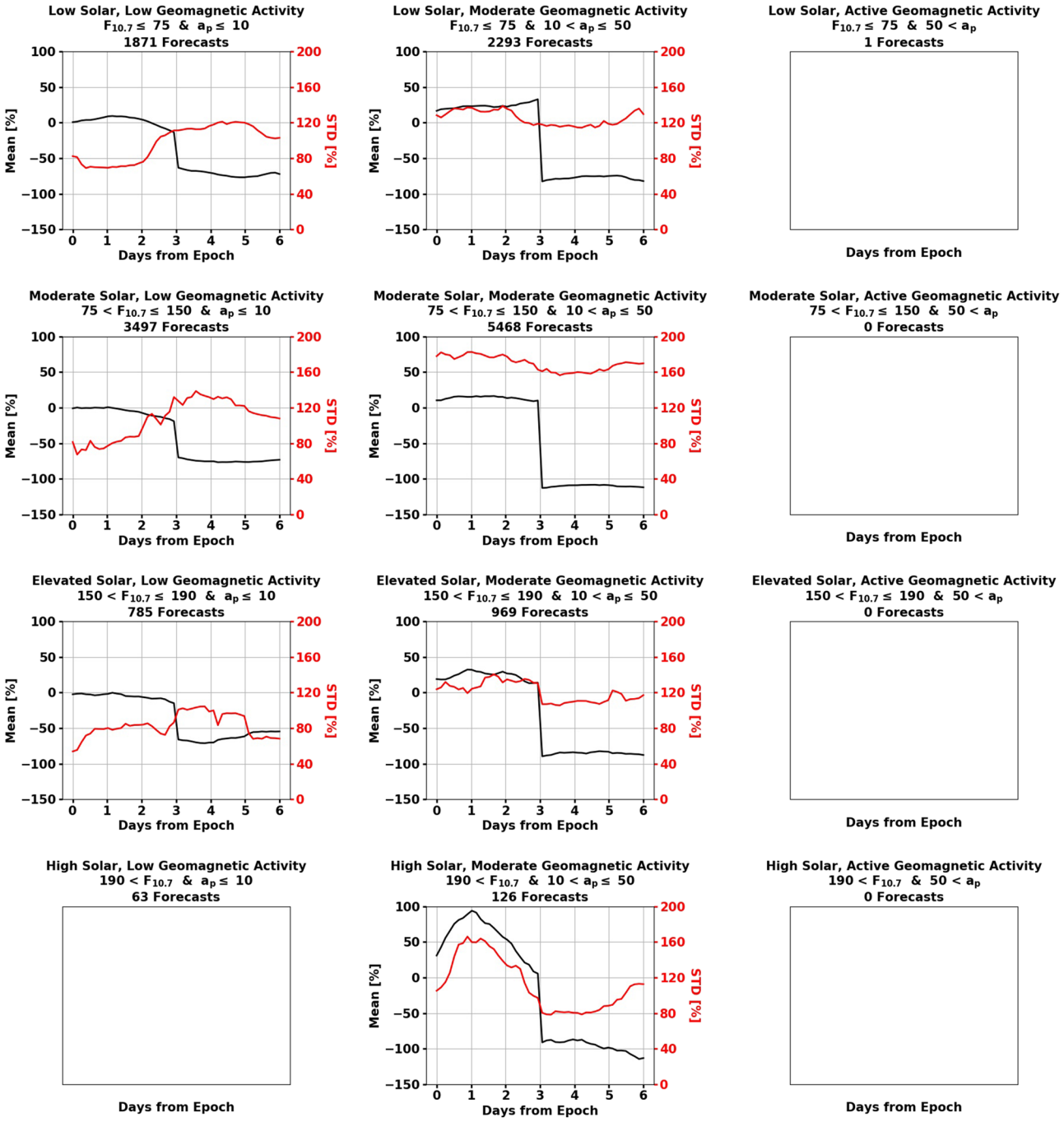

Figure 7. *ap* forecast uncertainty for the 12 solar and geomagnetic conditions in normalized form. The normalizing value is the long-term mean: 9.2 2nT.

showed the strengths of these predictive algorithms while also showing conditions where improvements can be made. In general, the forecasting capability for solar indices at low and moderate activity levels has comparably low uncertainty and virtually no bias. This performance is degraded to an extent at elevated and especially high activity levels, where the Sun is more volatile, and the evolution of flaring active regions is still poorly predicted.

The best performing algorithm is for $Y_{10.7}$ whose forecasting method is the most complex of the four solar indices investigated. The algorithm for $M_{10.7}$ also has low uncertainty and low bias at the two lower solar activity levels. The forecasts for $F_{10.7}$ and $S_{10.7}$ prove to be more uncertain and with generally higher biases.

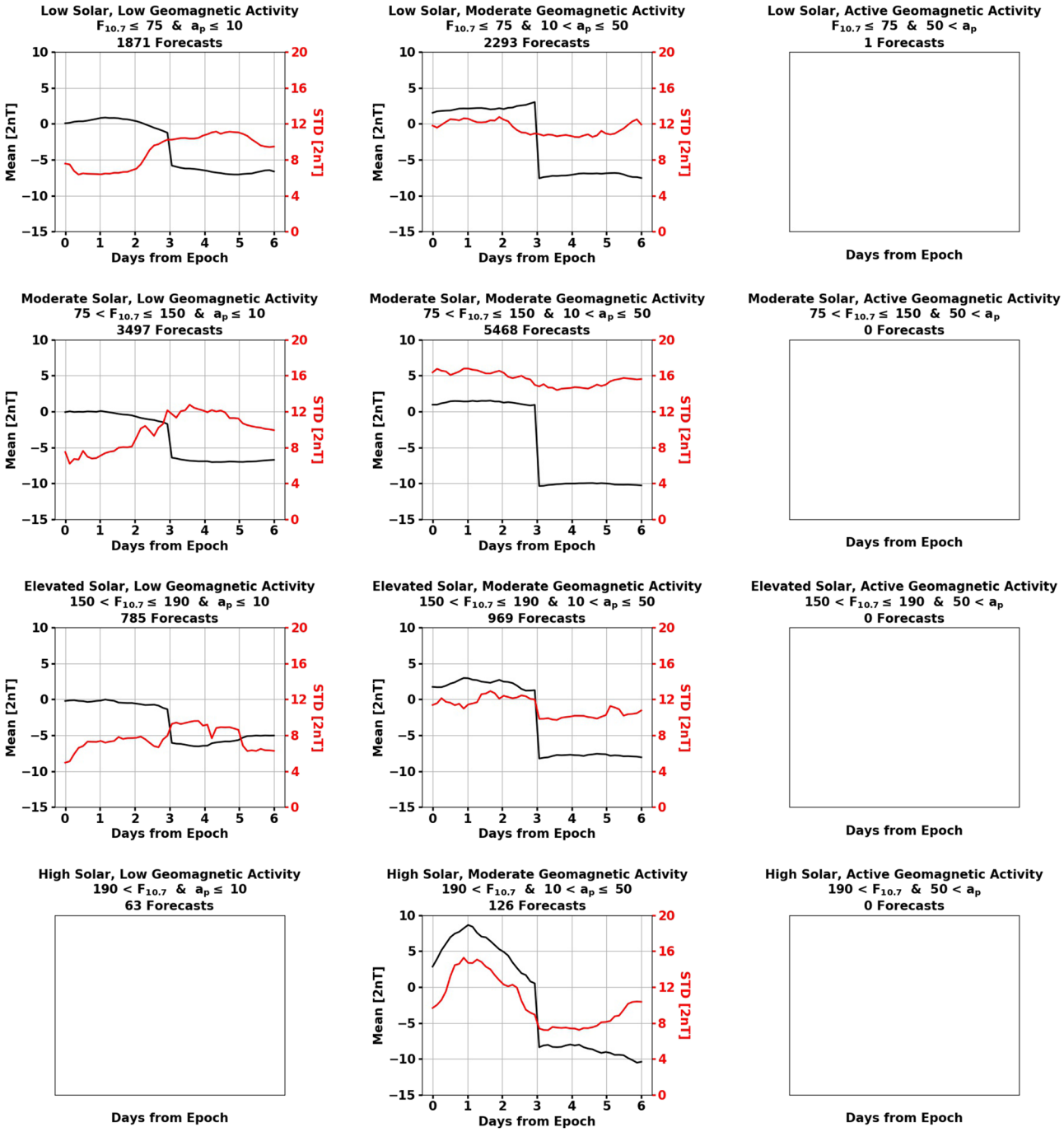

Figure 8. *ap* forecast uncertainty for the 12 solar and geomagnetic conditions in absolute terms.

Both indices had strong tendencies to overpredict at high solar activity. The index that delivers the greatest energy input to the atmosphere is $S_{10.7}$, so reducing the error in this driver would significantly improve density forecasting overall.

The geomagnetic indices, *ap* and *Dst*, proved to be difficult to predict even using two diverse methods. The forecasts for *ap* are determined by a team of forecasters with the aid of a model, and there was still a low probability of detection for geomagnetic storms. In most conditions, however, there was little or no bias in the predictions. The 3-day prediction window also ended up being a limitation, and results from a full 6-day forecast would be intriguing. The *Dst* algorithm performed well during G0 (or quiet) conditions. The standard deviation of error stayed steady around 150% in these cases, but that is only relative to the long-term mean

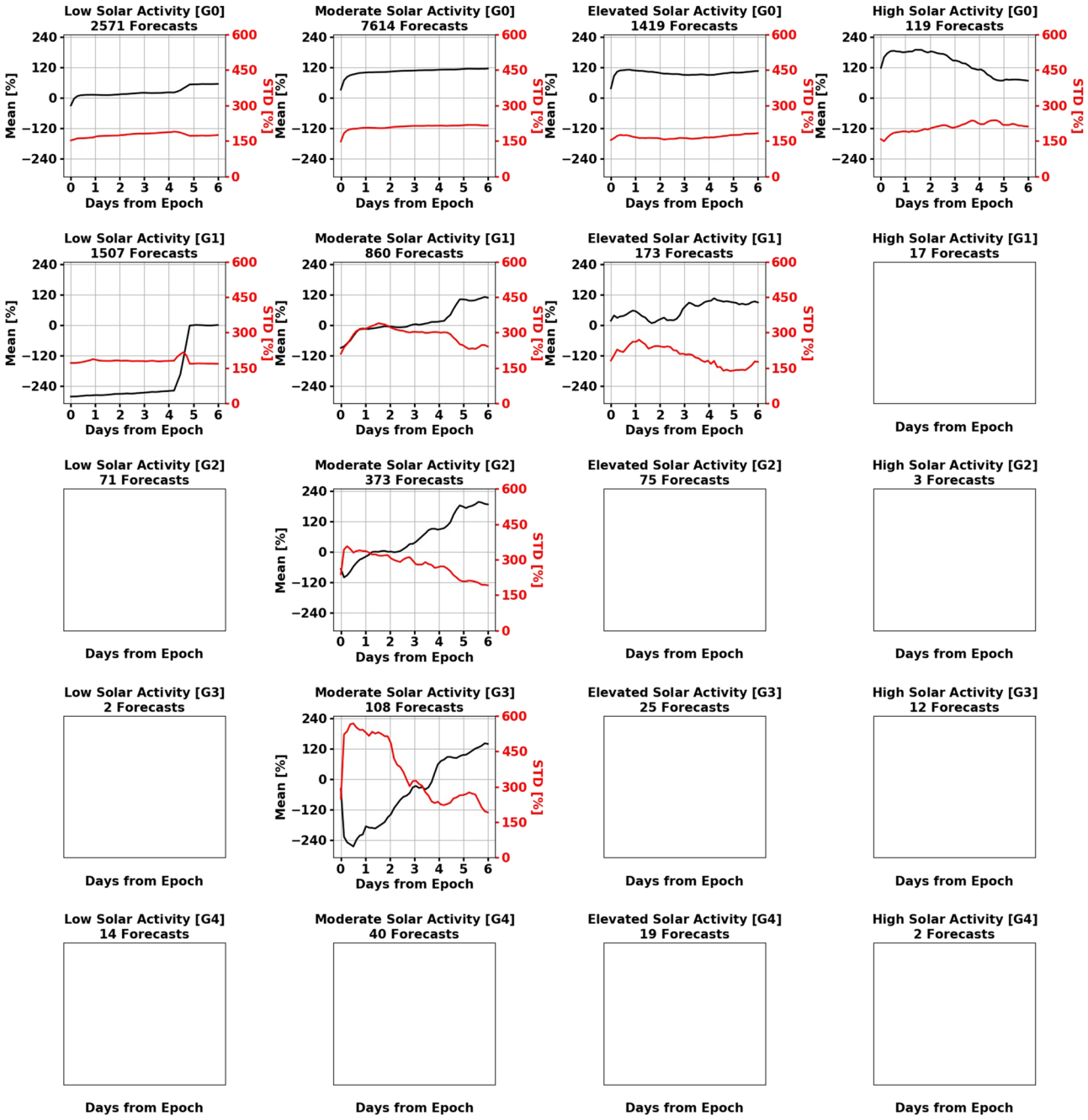

Figure 9. *Dst* forecast uncertainty for the combined solar and geomagnetic conditions in normalized form. The normalizing value is the absolute value of the long-term mean: 8.8 nT.

value, not the issued values. About 150% for *Dst* corresponds to ~13 nT. The algorithm showed poorer trends with increased geomagnetic activity. The increased uncertainty is attributed to the lack of HSS prediction and an inability to accurately and consistently forecast CME arrival time and magnitude.

A major limitation in this study was the lack of forecasts under certain conditions. This was particularly problematic for the geomagnetic indices, and using the most extreme index value to bin forecasts was used to offset this limitation. Even with this technique, a large percentage of conditions had insufficient data to perform the uncertainty analysis. In the future, we hope to include additional forecasts to the analysis to update the results in order to cover more conditions, particularly during the rise of Solar Cycle 25.

For clarification, these six indices and proxies do not represent or capture all of the energy deposited into the atmosphere, and therefore, perfect forecasting of these drivers will not necessarily result in perfect den-

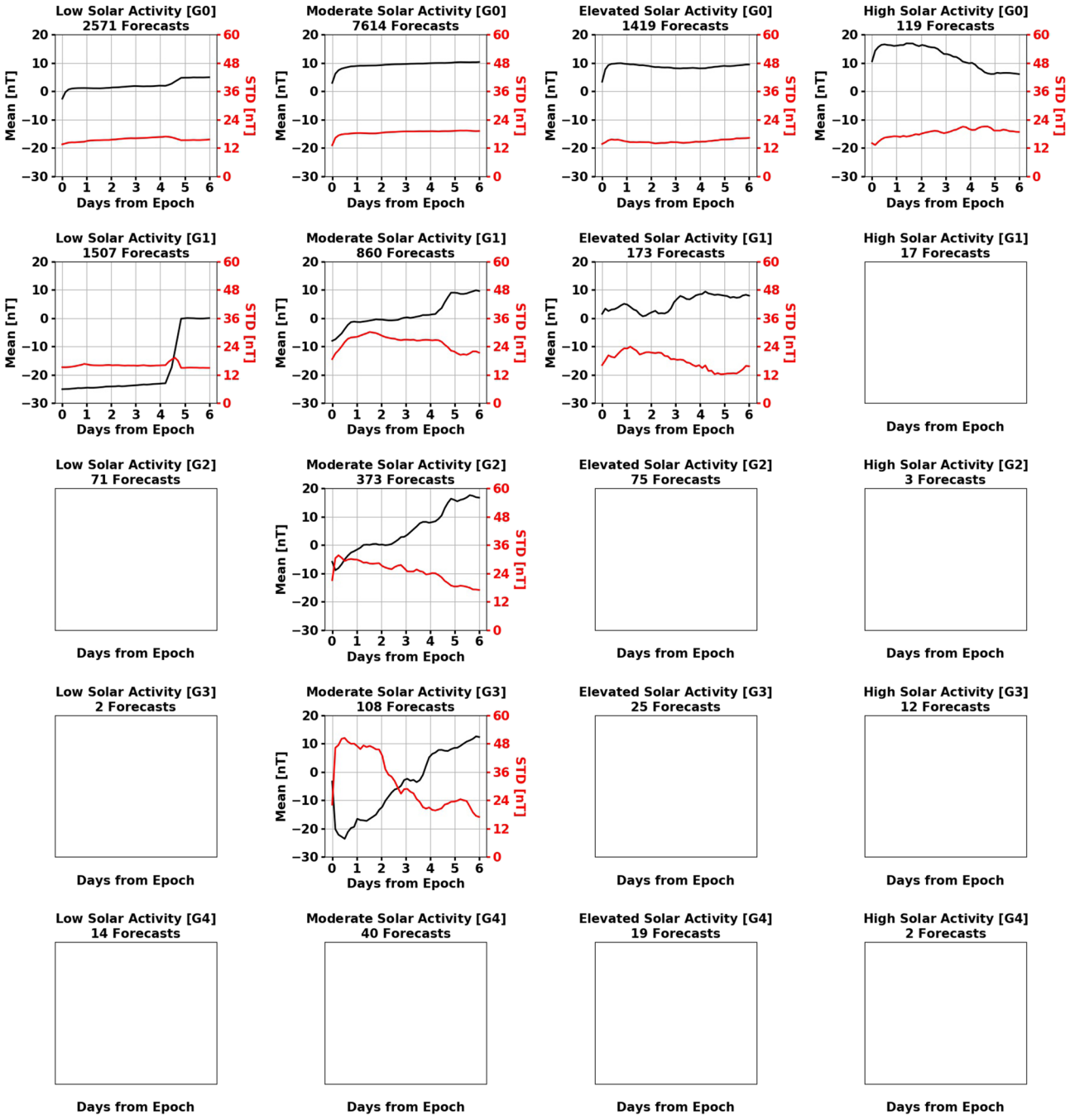

Figure 10. *Dst* forecast uncertainty for the combined solar and geomagnetic conditions in absolute terms.

sities. However, this work is intended to provide the community with a benchmark for future algorithm and model development. In comparison with these current capabilities, researchers and operators will be able to accurately identify improvements to forecasts of density and develop more precise satellite trajectories.

Data Availability Statement

SET proprietary data for this research are not made publicly available since they reside on operational servers run for the sole benefit of the USAF. Data are provided courtesy of Space Environment Technologies, 2019. These data have been provided to West Virginia University with license to use for scientific research.

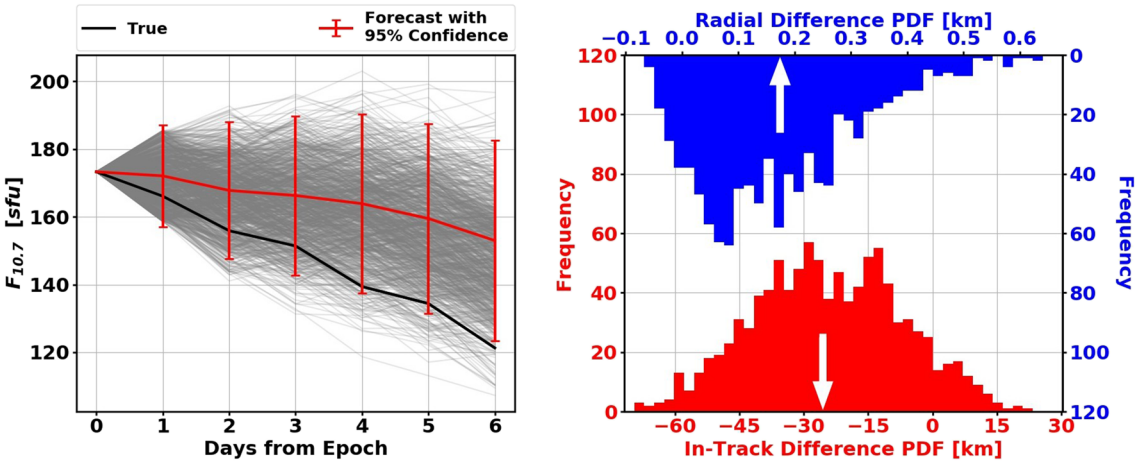

Figure 11. (left) Deterministic and probabilistic $F_{10.7}$ forecasts in addition to the true variation during the time period. (right) Satellite position distributions relative to the true position after encountering 6 days of probabilistic densities resulting from the corresponding $F_{10.7}$ fluctuations. White arrows represent position using deterministic $F_{10.7}$ values.

Acknowledgments
This work was made possible by NASA West Virginia Space Grant Consortium, Training Grant NNX15AI01H and NASA Established Program to Stimulate Competitive Research Grant 80NSSC19M0054. SET and WVU gratefully acknowledge support from the NASA SBIR Contract 80NSSC20C0292 for Machine learning Enabled Thermosphere Advanced by HASDM (META-HASDM). The authors would like to acknowledge Bruce Bowman, Dave Bouwer, and Alfredo Cruz of Space Environment Technologies for access to forecasts and relevant documentation to perform this study in addition to providing important insights into the extraction and use of operational data records.

References

Berger, T. E., Holzinger, M. J., Sutton, E. K., & Thayer, J. P. (2020). Flying through uncertainty. *Space Weather*, *18*, e2019SW002373. https://doi.org/10.1029/2019SW002373

Bowman, B., Tobiska, W. K., Marcos, F., Huang, C., Lin, C., & Burke, W. (2008). A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices. In *AIAA/AAS Astrodynamics Specialist Conference*. AIAA 2008-6438.

Bowman, B., Tobiska, W. K., Marcos, F., Huang, C., Lin, C., & Burke, W. (2012). A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices. https://doi.org/10.2514/6.2008-6438

Bussy-Virat, C. D., Ridley, A. J., & Getchius, J. W. (2018). Effects of uncertainties in the atmospheric density on the probability of collision between space objects. *Space Weather*, *16*, 519–537. https://doi.org/10.1029/2017SW001705

Emmert, J. T. (2015). Thermospheric mass density: A review. *Advances in Space Research*, *56*, 773–824. https://doi.org/10.1016/j.asr.2015.05.038

Haiducek, J. D., Welling, D. T., Ganushkina, N. Y., Morley, S. K., & Ozturk, D. S. (2017). SWMF global magnetosphere simulations of january 2005: Geomagnetic indices and cross-polar cap potential. *Space Weather*, *15*, 1567–1587. https://doi.org/10.1002/2017SW001695

ISO 14222 (2013). Space environment (natural and artificial)—Earth upper atmosphere. Geneva, CH: International Organization for Standardization.

ISO 21348 (2007). Space environment (natural and artificial)—Process for determining solar irradiances. Geneva, CH: International Organization for Standardization.

Licata, R. J., & Mehta, P. M. (2019). Physics-informed machine learning for probabilistic space weather modeling and forecasting: Thermosphere and satellite drag. https://doi.org/10.13140/RG.2.2.32538.18880

Licata, R. J., & Mehta, P. M. (2020). Physics-informed Machine learning with autoencoders and LSTM for probabilistic space weather modeling and forecasting. https://doi.org/10.13140/RG.2.2.17039.74401

Licata, R. J., Mehta, P. M., & Kay, C. (2019). Data-driven framework for space weather modeling with uncertainty treatment towards space situational awareness and space traffic management. In *Astrodynamics Specialist Conference*. AAS 19-603.

Licata, R., Mehta, P., & Tobiska, W. K. (2020). Impact of space weather driver forecast uncertainty on drag and orbit prediction, *Astrodynamics specialist conference*: AAS 20-423. Lake Tahoe, California.

McClain, W. D., & Vallado, D. A. (2001). *Fundamentals of astrodynamics and applications*, Space technology library (pp. 556–557). Springer Netherlands.

Mehta, P. M. (2013). Thermospheric density and satellite drag modeling (Ph.D. Thesis), University of Kansas.

Mehta, P. M., Linares, R., & Sutton, E. K. (2018). A quasi-physical dynamic reduced order model for thermospheric mass density via Hermitian space-dynamic mode decomposition. *Space Weather*, *16*, 569–588. https://doi.org/10.1029/2018SW001840

SWAP (2019). National space weather strategy and action plan: Space Weather Operations, Research, and Mitigation Working Group. https://www.whitehouse.gov/wp-content/uploads/2019/03/National-Space-Weather-Strategy-and-Action-Plan-2019.pdf

Singer, H. (2013). Report on the selection of geospace model(s) for transition to operations at NOAAs space weather prediction center (SWPC): Space Weather Prediction Center.

Steenburgh, R. A., Biesecker, D. A., & Millward, G. H. (2014). From predicting solar activity to forecasting space weather: Practical examples of research-to-operations and operations-to-research. *Solar Physics*. *289*, 675–690. https://doi.org/10.1007/s11207-013-0308-6

Storz, M., Bowman, B., & Branson, J. (2005). High accuracy satellite drag model (HASDM). https://doi.org/10.2514/6.2002-4886

Tobiska, W. K., Bouwer, S. D., & Bowman, B. R. (2008). The development of new solar indices for use in thermospheric density modeling. *Journal of Atmospheric and Solar-Terrestrial Physics*, *70*(5), 803–819. https://doi.org/10.1016/j.jastp.2007.11.001

Tobiska, W. K., Bowman, B., & Bouwer, S. D. (2008). Solar and geomagnetic indices for the JB2008 thermosphere density model. chap. 4, COSPAR CIRA Draft.

Tobiska, W. K., Knipp, D., Burke, W. J., Bouwer, D., Bailey, J., Odstrcil, D., et al. (2013). The Anemomilos prediction methodology for Dst. *Space Weather*, *11*, 490–508. https://doi.org/10.1002/swe.20094

Tobiska, W. K., Woods, T., Eparvier, F., Viereck, R., Floyd, L., Bouwer, D., et al. (2000). The SOLAR2000 empirical solar irradiance model and forecast tool. *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(14), 1233–1250. https://doi.org/10.1016/S1364-6826(00)00070-5