AAS 20-423

IMPACT OF SPACE WEATHER DRIVER FORECAST UNCERTAINTY ON DRAG AND ORBIT PREDICTION

Richard J. Licata[*], Piyush M. Mehta[†], and W. Kent Tobiska[‡]

Space Weather (SW) has a strong influence on satellite tracking, orbital decay, and collision avoidance in low Earth orbit (LEO). E.g., Satellite position Probability Density Functions (PDFs) essential for probability of collision, Pc, estimates are heavily dependent on drag. We have recently characterized the performance of SW driver forecast models used in operations by the US Air Force Space Command (AFSPC) with the High Accuracy Satellite Drag Model (HASDM). In this paper, we quantify the impact of SW driver forecast uncertainty on drag and orbit prediction.

INTRODUCTION

The Sun drives all of the variations in the near-Earth geospace environment. Largest variations in thermospheric mass density, up to several orders of magnitude, are caused by direct energy deposition from the Sun in the Extreme Ultraviolet (EUV) Spectrum. During solar maximum, temperature variations dominate; while during solar minimum, composition changes become more relevant. The variations caused by EUV heating have a relatively regular signature with periods of 27 days corresponding to solar rotation and 11 years corresponding to the solar cycle. Fluctuations on shorter scales can be caused by events such as solar flares.

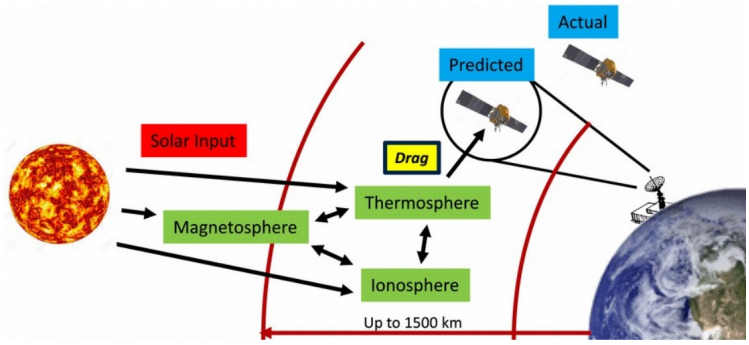

Figure 1. Coupling between Space Weather and Thermosphere/Drag and its impact on orbit prediction

[*]Graduate Research Assistant, Department of Mechanical and Aerospace Engineering, Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV 26506-6106.
[†]Assistant Professor, Department of Mechanical and Aerospace Engineering, Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV 26506-6106.
[‡]President / Chief Scientist, Space Environment Technologies, Pacific Palisades, California, USA.

A secondary source of energy input to the thermosphere is solar wind, which manifests itself as geomagnetic disturbances. They exhibit impulsive characteristics, from short periodic disturbances on the order of hours to longstanding, giant storms of several days' duration. Energy from the solar wind is transferred into the magnetosphere, which is eventually dissipated into the ionosphere/thermosphere through interaction with energetic plasma. The most dominant energy dissipation modes are Joule heating, ring currents, and auroral precipitation. During large geomagnetic storms, the thermospheric mass density can vary by up to three orders of magnitude and cause significant problems for space operations (including orbit tracking, prediction and collision) as illustrated in Figure 1. For example, during the March 1989 storm event, the NASA's Solar Maximum Mission (SMM) spacecraft was reported to have "dropped as if it hit a brick wall" due to the increased atmospheric drag. More than 1300 of the 8000 objects were lost and had to be reacquired as shown in Figure 2.[1] Because models and measurements used for predictions are not perfect, quantifying the confidence in the predictions is equally crucial.

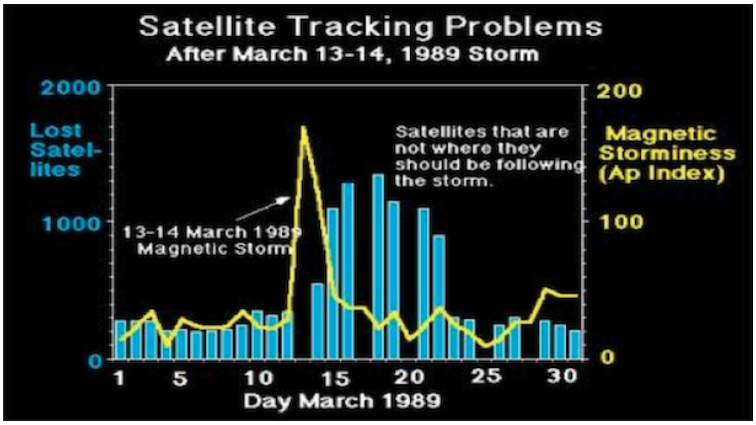

Figure 2. Number of satellites lost in connection to the March 1989 storm event.[1]

SW 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 the processes that cause energy flow and deposition in the geospace system. Forecasts of neutral mass density is a major uncertainty in operational orbit prediction and collision avoidance for objects in LEO. For the strongly driven system, accuracy of space weather driver forecasts is crucial for operations. The currently operational HASDM at AFSPC is driven by four (4) solar and two (2) geomagnetic proxies. Space Environment Technologies (SET) is contracted by the space command to provide forecasts for the drivers.[2,3]

Among a plethora of space weather indices and proxies, $F_{10.7}$ and ap are widely used to determine solar and geomagnetic activity, respectively. $F_{10.7}$ is a legacy proxy for solar EUV heating and has been recorded since 1947.[4] The 10.7 cm radio emission serves as an input for numerous models due to its correlation to thermospheric expansion and contraction. Centered ($F81c$) and backwards ($F54b$) averages for $F_{10.7}$ are often used in models as well to better determine the long-term state of the thermosphere.

In order to correlate the effect of geomagnetic storms and background activity to thermospheric mass density, Kp and ap are often used. The quasi-logarithmic planetary Kp index is used to assess atmospheric heating from disturbances within the magnetosphere on a global scale.[5] ap is the linearized version of Kp, and they both have a time resolution of three hours. Ap is the daily

average ap, which is utilized as well (e.g. MSIS).

Density forecasts used in orbit determination and Pc estimation have multiple sources of uncertainty. These include but are not limited to: SW driver uncertainty,[6,7] model uncertainty,[8,9] and subgrid uncertainty.[10,11] Each uncertainty has conditions where it dominates, but in the context of days, driver uncertainty is the major contributor.

In this paper, we will quantify the impact of SW driver forecast and model uncertainty on orbit prediction, crucial for conjunction assessment towards Space Situational Awareness and Space Traffic Management. We will do so for orbits at different altitudes in LEO and using various empirical and physics-based thermosphere models (TIE-GCM,[12] MSIS[13] and HASDM[14]).

METHODOLOGY

The algorithms used to forecast the aforementioned drivers are sub-optimal, and in the context of SSA, challenges arise when determining the reliability of the resulting density predictions. In previous work, the authors had used six years of SW driver forecasts provided by Space Environment Technologies (SET) to extract important statistics for the temporal evolution of the forecast error.[6] With this information, one can generate probabilistic samples of the driver forecasts. These probabilistic samples can subsequently be used in a Monte Carlo analysis with the goal to cover all likely possibilities in the evolution of the proxy or index of focus.

Density Models

The models used in this study are TIE-GCM, NRLMSISE00, and HASDM. TIE-GCM is a physics-based model that evolves species densities based on a set of SW inputs. Since it is a physics-based model, it is computationally expensive, compromising the feasibility of Monte Carlo analyses. To overcome this, a Long-Short Term Memory (LSTM) neural network model is used as a surrogate.[15] It is a reduced order model (ROM) trained on Principal Component Analysis (PCA) coefficients derived from TIE-GCM results. Another drawback to using TIE-GCM is the low altitude threshold the model provides. Since the densities produced from the model span both solar maximum and minimum conditions, the highest altitude output is 450 km. To use TIE-GCM at higher altitudes, the density data cubes were linearly extrapolated in log-scale.

NRLMSISE00, referred to here as MSIS, is an improvement over its predecessor, MSISE-90.[16] It is an empirical model which provides the advantage of prompt evaluation. As it pertains to this work, MSIS was evaluated using the GTD7 subroutine to generate 3-dimensional mass density data cubes every hour.

To incorporate HASDM, which is not a publicly available model, a HASDM machine learning model (HASDM-ML) was leveraged. The use of the HASDM-ML model is emblematic of an assimilated empirical model. The inputs for each model are found below in Table 1. For clarity, UTC refers to coordinated universal time, doy refers to the day of year, and the terms $Date$ and $Time$ refer to the year, day of year, and UTC in seconds.

Table 1. Inputs for the three models used in the study.

TIE-GCM ROM	$F_{10.7}$, Kp, UTC, doy
MSIS	$Date$, $Time$, $Position(LLA)$, $F_{10.7}$, $F81c$, Ap
HASDM-ML	$F_{10.7}$, $F54b$, ap, UTC, doy

Probabilistic Input Sampling

Since the bias and uncertainty in SET's driver forecast algorithms varied with different solar activity levels, three thresholds had been previously determined to generate four subpopulations of the forecasts. These forecasts were allocated based on the deterministic value one day from forecast epoch. For the geomagnetic forecasts (ap), the error analysis had been performed on the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) three day forecasts. The overall duration in this study is six days; therefore, the error analysis was still done on the last three days even though the forecasts are set to zero. This essentially accounts for the background ap within each condition. The ap forecasts are distributed based on the highest value in a given prediction period. The solar and geomagnetic activity levels and thresholds are displayed in Table 2.

Table 2. Bin thresholds for solar activity, $F_{10.7}$, and geomagnetic activity, ap.

Solar Activity	Low	$F_{10.7} \leq 75$
	Moderate	$75 < F_{10.7} \leq 150$
	Elevated	$150 < F_{10.7} \leq 190$
	High	$190 < F_{10.7}$
Geomagnetic Activity	Low	$ap \leq 10$
	Moderate	$10 < ap \leq 50$
	Active	$50 < ap$

Seven time periods between October 2012 and December 2019 were chosen to perform this analysis, which coincides with the availability of operational deterministic forecasts. The first four periods were chosen to cover each of the solar activity levels. For these cases, $F_{10.7}$ samples were generated based on the deterministic forecast and the temporal error statistics (mean and standard deviation at each point in time after forecast epoch). This is done 1,000 times for the Monte Carlo analysis, and the true ap variation is kept consistent between them. Figure 3 shows the true, deterministic, and probabilistic drivers for each of the four solar cases.

It is evident in the figure that many of the deterministic $F_{10.7}$ forecasts deviate significantly from the true variation in the six-day window. However, in each of these cases, the true variation is well-captured in the 3σ bounds. The right subplots show the true ap variation for the corresponding six-day period. The only forecast that does not exhibit any moderate geomagnetic activity is low solar activity. This is representative of solar minimum conditions.

In previous work, it had been observed that the ap forecast performance not only varied with geomagnetic activity level but also with solar activity level. Therefore, forecasts were distributed based on a combination of the two, resulting in twelve subpopulations for the temporal error statistics. Unfortunately, there were not enough forecasts in five of twelve subpopulations to extract meaningful statistics, so studies were only performed for the remaining seven conditions.

For this work in particular, only three conditions were chosen to study the effect of ap forecast uncertainty. These conditions are as follows: 1.) moderate solar / moderate geomagnetic activity, 2.) elevated solar / low geomagnetic activity, and 3.) high solar / moderate geomagnetic activity. For each of these conditions, $F_{10.7}$ will be kept at its true values consistently between the probabilistic ap samples. The true, deterministic and probabilistic drivers for these periods are displayed in Figure 4.

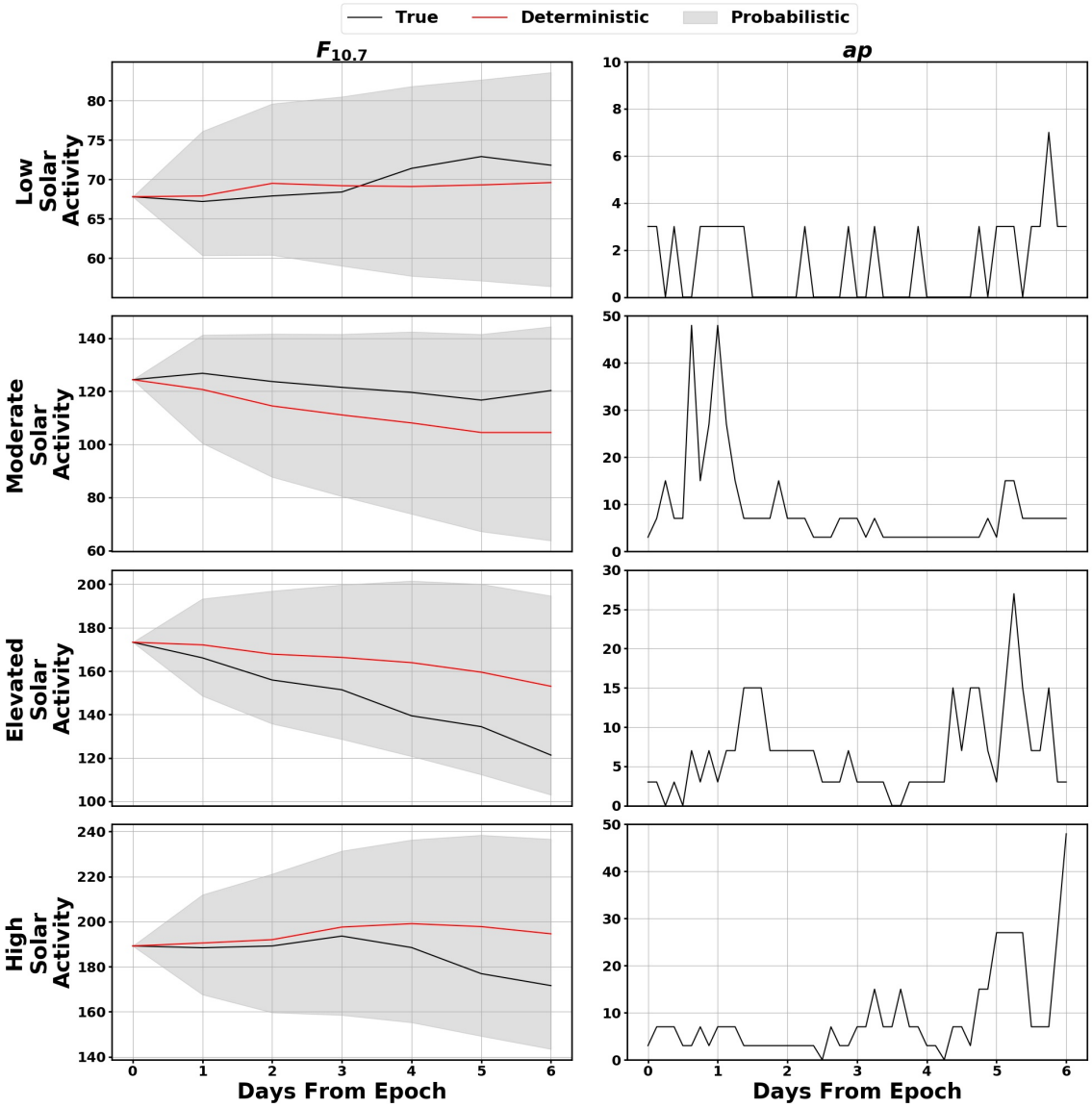

Figure 3. Space Weather inputs for the four solar cases. The shaded probabilistic region shows the 3σ bounds for the probabilistic samples.

Orbit Propagation

To generate three-dimensional density grids for each model and each set of inputs, a consistent method was chosen. At each hour in the six-day period, the models are evaluated for every set of probabilistic and true inputs. There are three altitudes of interest for this analysis (400 km, 600 km, and 800 km). A small theoretical satellite is propagated at each altitude accounting for two-body forces, J2, and atmospheric drag. Each probabilistic forecast and associated set of true inputs are fed into each model and evaluated on an hourly basis as previously mentioned. The propagator interpolated density at every location the satellite encounters, and the density data cube only changes when the time shifts from one hour to the next.

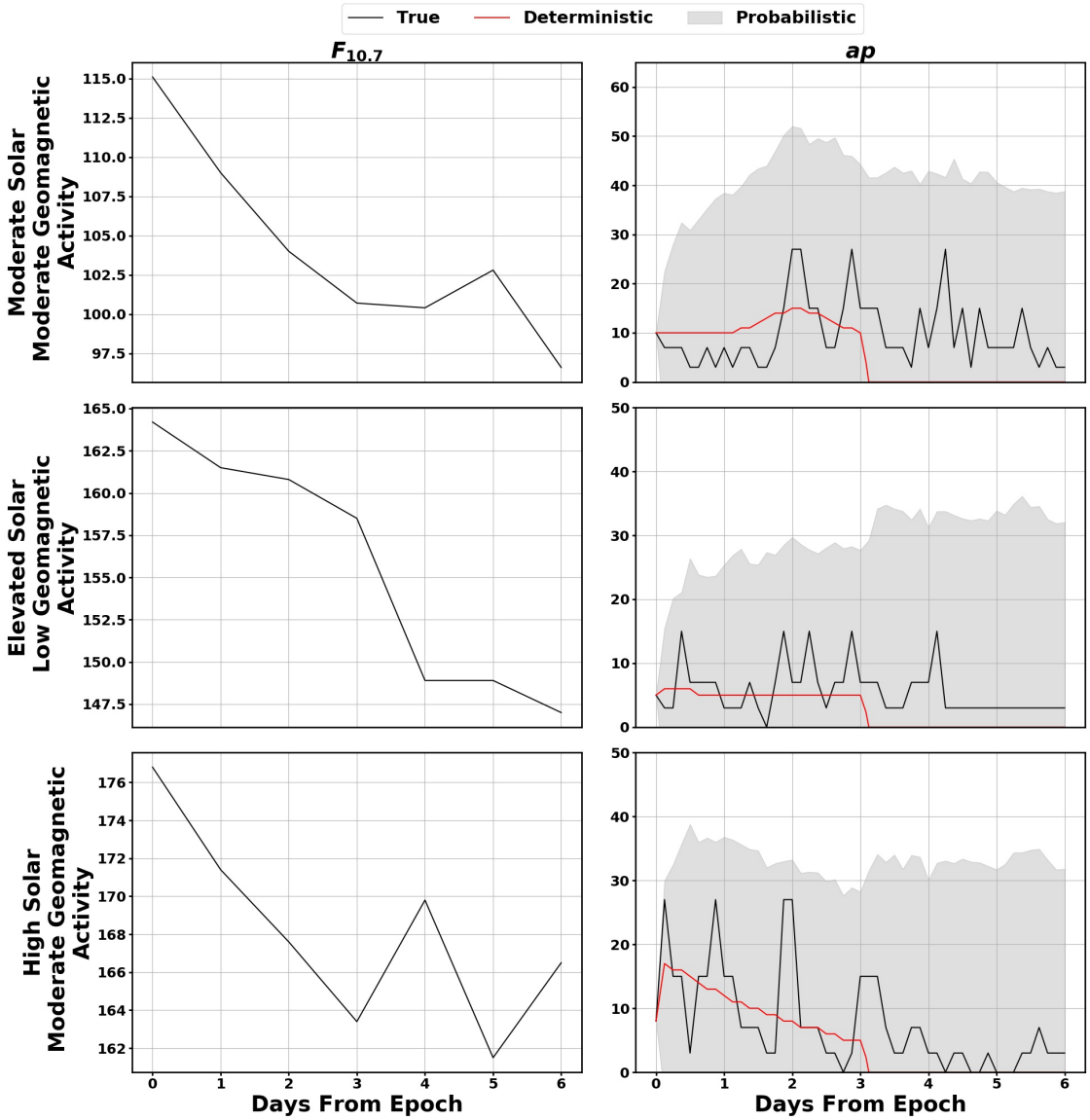

Figure 4. Space Weather inputs for the three geomagnetic cases. The shaded probabilistic region shows the 3σ bounds for the probabilistic samples.

Once the satellites are propagated through the density grids using the true and probabilistic inputs, the Earth-centered initial (ECI) positions at three and six days from epoch are saved. The position resulting from the HASDM-ML model using the true inputs is considered the true position, and a transformation matrix is derived from that information. Finally, in-track position errors with respect to the HASDM true position are computed for the probabilistic positions of each model. The in-track position errors are the focus of the study as they are the largest errors in the local reference frame.

RESULTS: DENSITY AND ORBIT PREDICTION WITH STATISTICAL SAMPLING

The results of this study show both the effects of driver uncertainty and model uncertainty. In-track position error histograms will be shown for three altitudes and all seven conditions. The standard deviation of each model will illustrate how sensitive each model is to the input while the mean position error will highlight the general over-or-under prediction of density for various conditions and altitudes.

At forecast epoch, the models are given the same set of inputs and there is no uncertainty associated with the inputs. These were fed to each model, and Figure 5 shows the average density profiles in semi-log plots. This was done for each of the seven SW conditions.

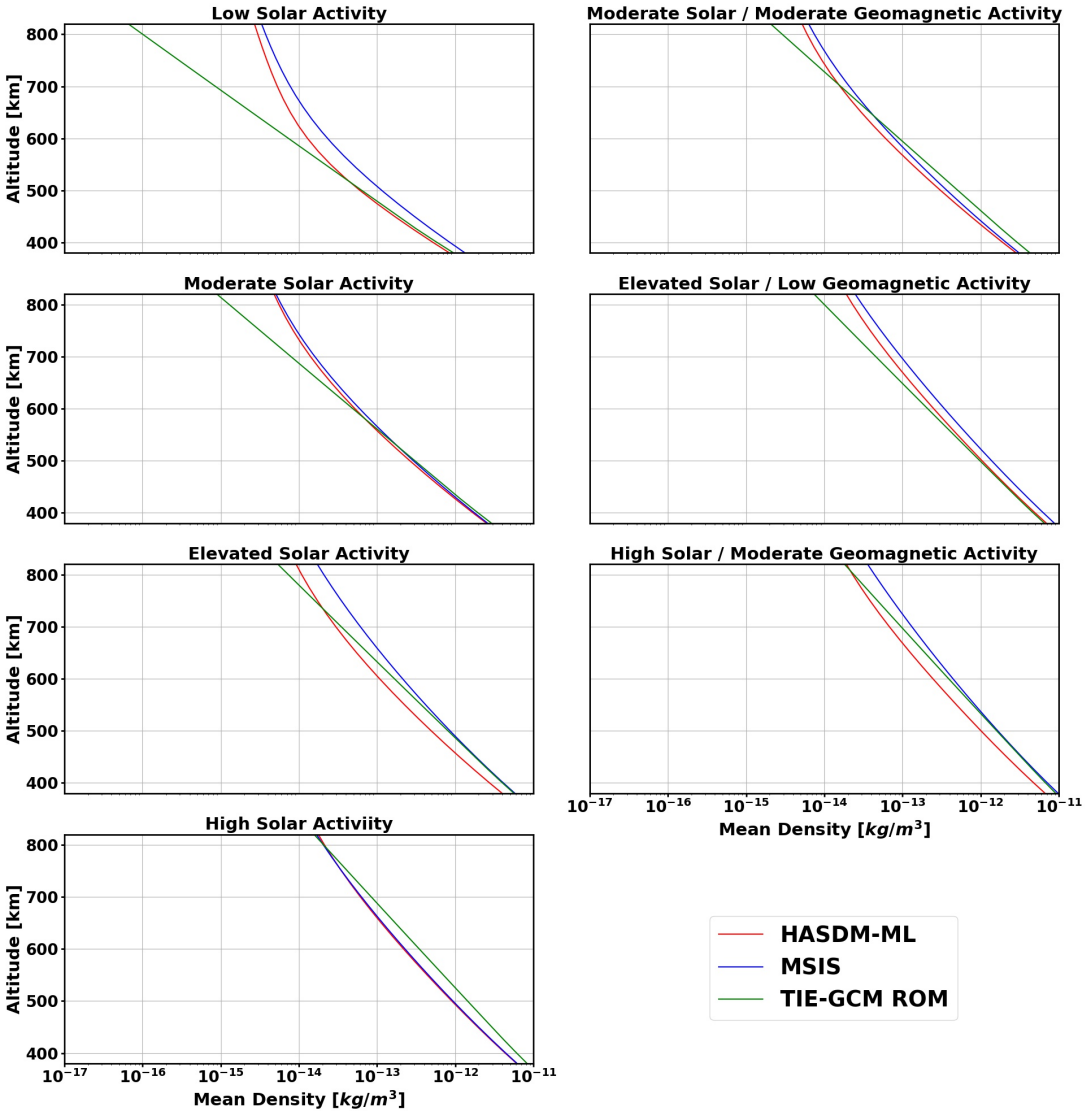

Figure 5. These are average density profiles for each model between 380 and 820 km. The left panels show the four solar conditions while the right panels show the three geomagnetic conditions. These are for inputs at forecast epoch (zero driver uncertainty).

In Figure 5, the linear extrapolation (in log-scale) for the TIE-GCM ROM densities is prominently displayed. There are linear trends for the other two models between 400 and 600 km, but beyond that, the linear extrapolation causes the density to be lower than it should be. This could be caused by the ROM only forecasting mass density and not number densities. The logarithmic extrapolation should be performed on the number densities, which would then be converted to mass density, for better results.[17,18] Data assimilation at solar minimum on the TIE-GCM ROM will be the focus of future work using the cited methodologies.[19,20]

For all of the proceeding figures, necessary context can be found in Figures 3-5 in order to see what the inputs are (both true and probabilistic) along with how the density profiles change for the different SW conditions. The first condition to be examined is low solar activity. The result for this condition are shown in Figure 6 and Table 3.

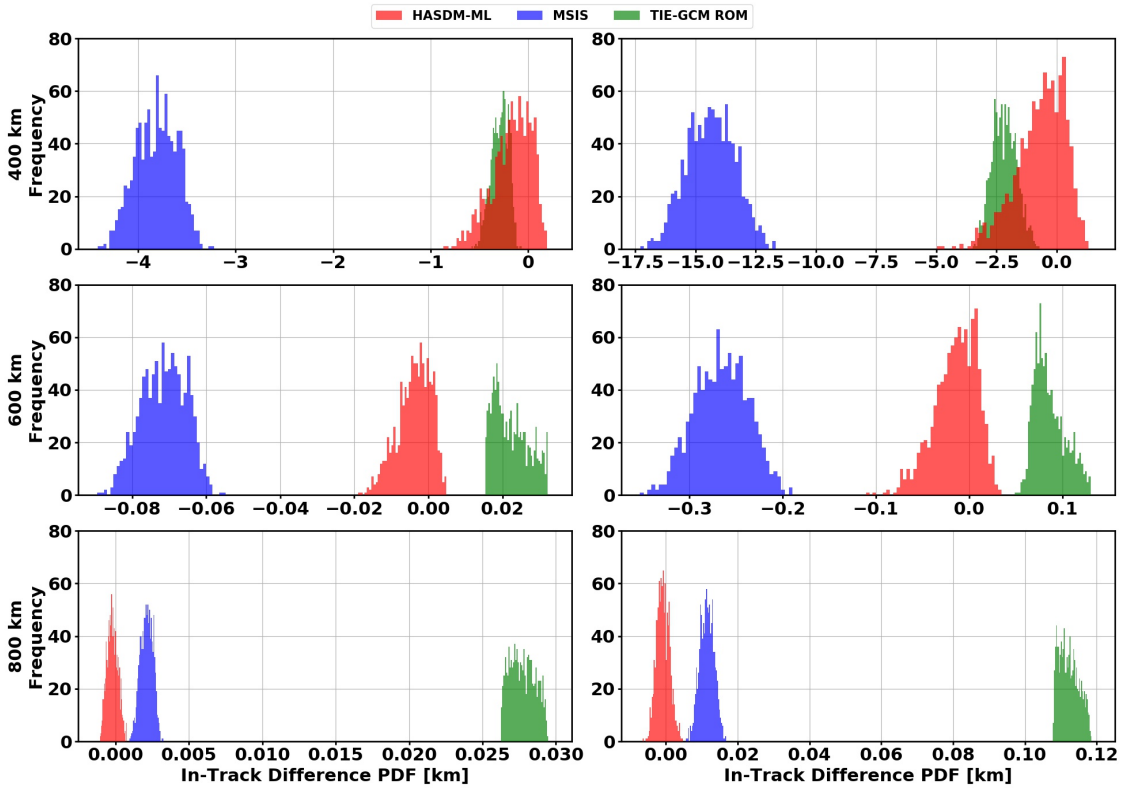

Figure 6. In-track position error distributions after three days (left) and six days (right). This is for low solar activity with $F_{10.7}$ being sampled.

At 400 km, MSIS is over-predicting density significantly. This results in more drag, loss of altitude, and increase in spacecraft velocity. Therefore, and this is true for all figures, higher density than the true case will cause the spacecraft to be ahead of the reference position (true inputs and HASDM-ML). That is what negative in-track difference represents. MSIS has a large bias at the lowest altitude during this solar minimum case.

The TIE-GCM ROM also has a slight tendency to over-predict density at 400 km, but that is reversed as altitude increases. Seen in the corresponding subplot in Figure 5, the logarithmic ex-

trapolation of mass density causes the densities to be much lower than the other two models with increasing altitude. This is the source of the right-shifting distribution of the TIE-GCM ROM in the second and third rows.

Table 3. Distribution statistics for low solar activity (Figure 6).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	-0.1449	0.1996	-0.5028	1.0126
	600 km	-0.0032	0.0045	-0.0109	0.0230
	800 km	-0.0002	0.0004	-0.0005	0.0018
MSIS	400 km	-3.7817	0.2085	-14.2209	0.9911
	600 km	-0.0707	0.0060	-0.2619	0.0288
	800 km	0.0022	0.0004	0.0117	0.0020
TIE-GCM ROM	400 km	-0.2787	0.0887	-2.1007	0.5023
	600 km	0.0228	0.0047	0.0881	0.0168
	800 km	0.0278	0.0008	0.1126	0.0027

Table 3 explicitly shows the bias and sensitivity of each model to the probabilistic inputs during this quiet solar condition. For both HASDM-ML and MSIS, the bias and sensitivity are reduced with increasing altitude as the overall densities become smaller and atmospheric drag is less significant. The sensitivity of TIE-GCM ROM follows this trend, but the bias has an abnormal trend due to the extrapolation. The major takeaways for this condition are the large bias exhibited by MSIS at the lower two altitudes, the difference in sign for the MSIS bias at 800 km, and the low sensitivity of TIE-GCM ROM to these low $F_{10.7}$ values. The results at moderate solar activity are displayed in Figure 7 and Table 4.

As solar activity level increases, the driver forecast uncertainty increases, as well.[6] This leads to a larger variance in $F_{10.7}$ between the probabilistic samples and therefore density and satellite position. In Figure 7, the distributions show this larger spread, as well as different relative performance. For this moderate solar activity condition, MSIS is more aligned with HASDM-ML than TIE-GCM ROM, particularly at higher altitudes.

Again, the shortcomings of the extrapolation are highlighted with the bottom two rows. The first row shows results from the actual TIE-GCM ROM output (no extrapolation), and the position error distribution is fairly similar to HASDM-ML and MSIS. There is also an abnormal distribution shape for HASDM-ML particularly earlier in the propagation (left subplots). All of the HASDM-ML distributions are skewed right, but it is unconventionally strong at the three-day mark.

The suspected cause of this is the two peaks in the ap variation for this period and how the non-optimized model reacts to them. After this abrupt storm ends, there is little geomagnetic activity, and the Gaussian shape of the HASDM-ML distribution begins to return at day six. This normalization of the distribution is more pronounced at lower altitudes, because atmospheric drag has a more pronounced effect. At 800 km, the storm disrupts the orbits in an unconventional manner, and the smaller drag effect is not able to dampen the abnormality in positions. However, even with this atypical model reaction, the true position is still captured (zero error within the distribution) at both points in the propagation at all altitudes.

At 400 km, all models actually capture the true position at both three and six days. At 600 km, only HASDM-ML and MSIS capture it, and at 800 km, MSIS drifts behind the true position. Table

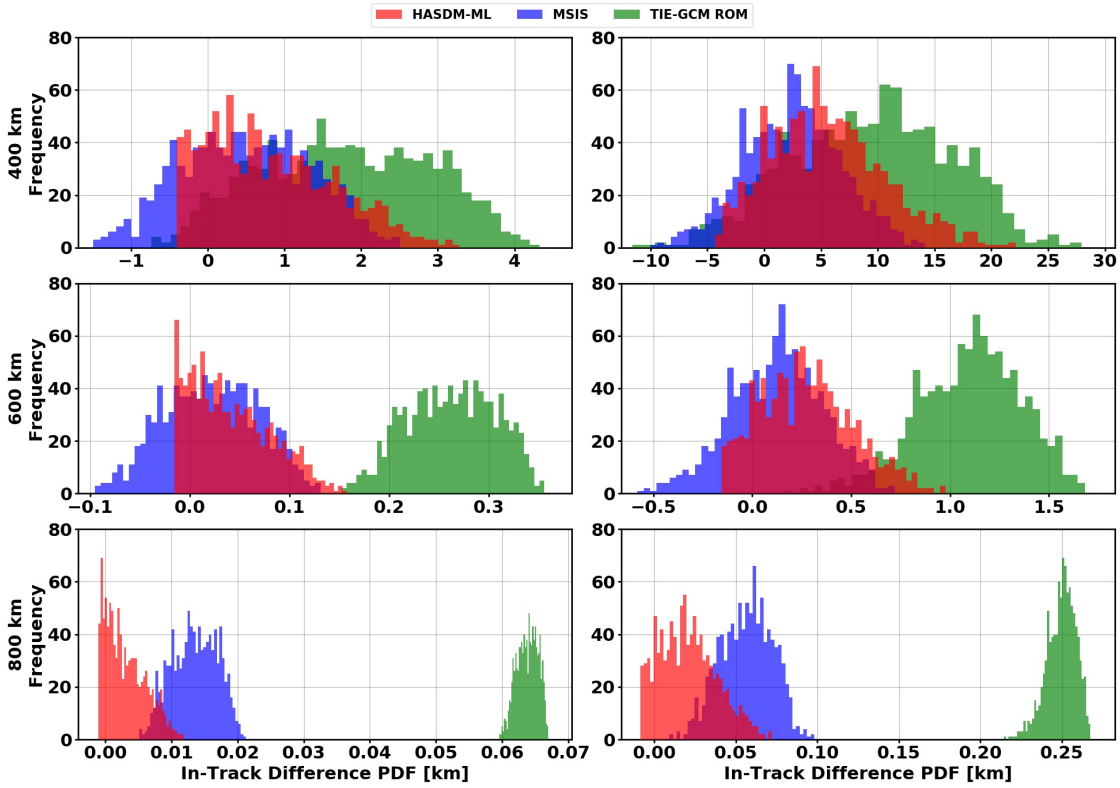

Figure 7. In-track position error distributions after three days (left) and six days (right). This is for moderate solar activity with $F_{10.7}$ being sampled.

Table 4. Distribution statistics for moderate solar activity (Figure 7).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	0.9128	0.8239	6.2897	5.0340
	600 km	0.0454	0.0402	0.3034	0.2338
	800 km	0.0034	0.0030	0.0227	0.0173
MSIS	400 km	0.5990	0.8334	3.0526	4.2574
	600 km	0.0273	0.0474	0.1574	0.2311
	800 km	0.0139	0.0034	0.0587	0.0159
TIE-GCM ROM	400 km	1.9182	1.0990	10.3535	6.9950
	600 km	0.2675	0.0443	1.1334	0.2643
	800 km	0.0640	0.0016	0.2510	0.0090

4 shows how similar the sensitivity of the models is to this set of probabilistic $F_{10.7}$ samples. The standard deviations for a given epoch and altitude are quite similar. The difference mean position error, or bias, is emphasized in this table. Next, the impact of $F_{10.7}$ uncertainty on the models for elevated solar activity is shown in both Figure 8 and Table 5.

At 400 km, within all model grids, the biases at elevated solar activity are significant. There is no overlap between any of the distributions at days 3 or 6. In fact, the bias only three days from

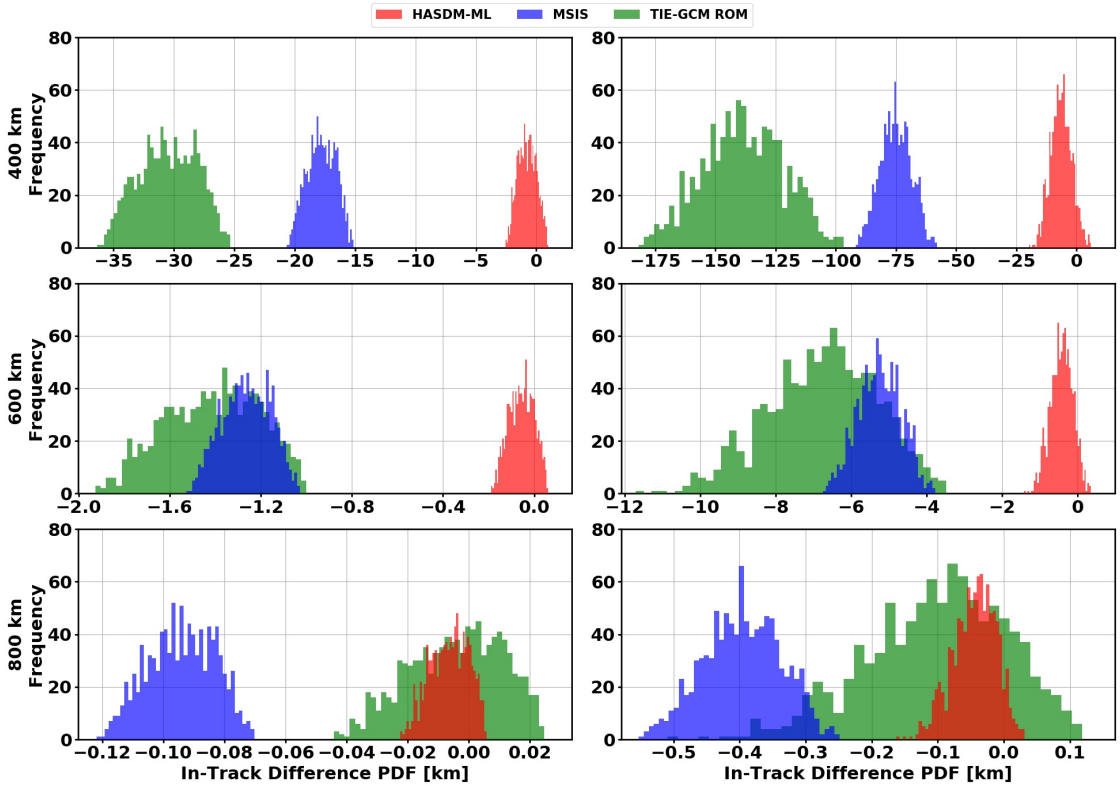

Figure 8. In-track position error distributions after three days (left) and six days (right). This is for elevated solar activity with $F_{10.7}$ being sampled.

forecast epoch for TIE-GCM ROM is $\sim$ 30 km and after six days, the bias grows to over 136 km. This is paramount, because in STM, a few meters of error can impact the decision-making process. This much bias is troubling, not to mention the almost 17 km standard deviation.

MSIS has a large, but less severe bias than TIE-GCM ROM. Both models are strongly overpredicting density, and its effect on orbit determination is critical. Even HASDM-ML has a measurable bias, but it is smaller than the other two models, as is the variance.

Table 5. Distribution statistics for elevated solar activity (Figure 8).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	-0.6724	0.7550	-5.7028	4.2967
	600 km	-0.0492	0.0545	-0.3762	0.2947
	800 km	-0.0057	0.0063	-0.0398	0.0325
MSIS	400 km	-17.6546	1.1699	-74.2236	6.4088
	600 km	-1.2510	0.1050	-5.1738	0.5705
	800 km	-0.0928	0.0109	-0.3878	0.0586
TIE-GCM ROM	400 km	-30.2316	2.4032	-136.5033	16.8159
	600 km	-1.3897	0.2069	-6.5461	1.4468
	800 km	-0.0025	0.0152	-0.0857	0.1062

At higher altitudes, the bias is greatly reduced, as seen in Table 5. The effect of density extrapolation for TIE-GCM ROM actually improves the performance with increasing altitudes, but that is attributed to the substantial overprediction within the model output. MSIS overpredicts density at all altitudes and both epochs, and its position error distribution does not overlap with HASDM-ML's distribution at any point. The last case for sampling $F_{10.7}$ is high solar activity, and the results are displayed in Figure 9 and Table 6.

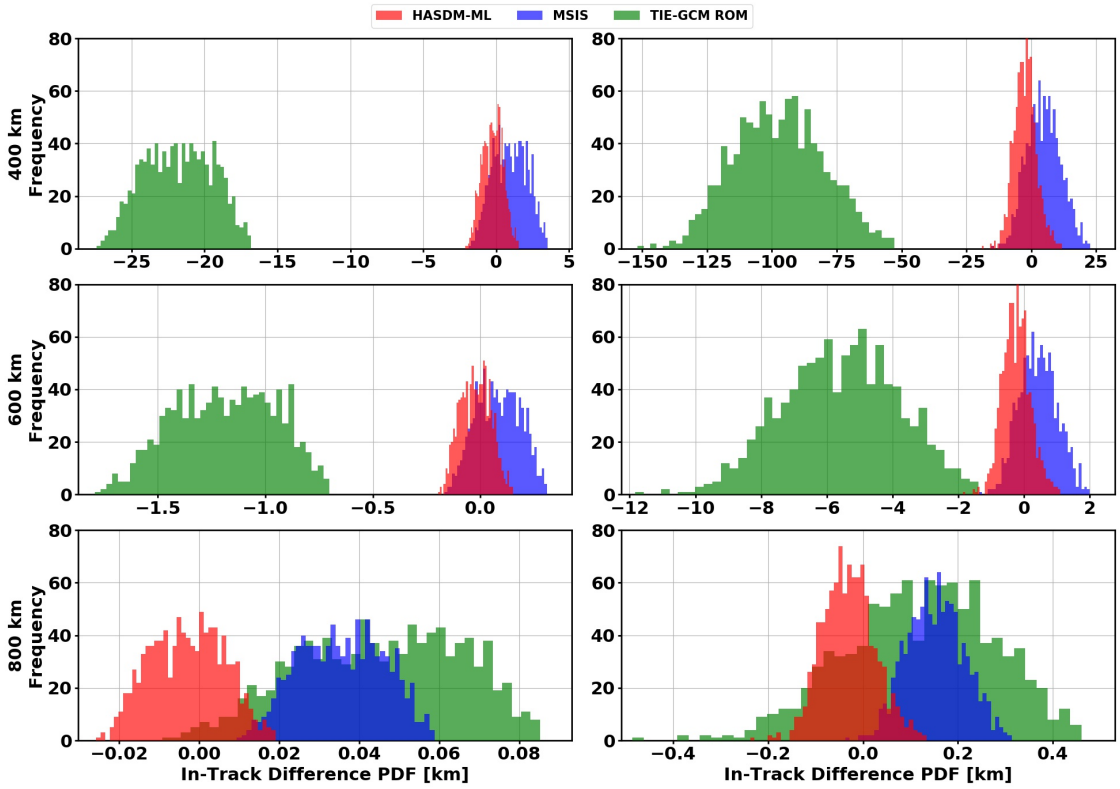

Figure 9. In-track position error distributions after three days (left) and six days (right). This is for high solar activity with $F_{10.7}$ being sampled.

For this particular case, many of the $F_{10.7}$ samples are in range of solar maximum level values. This is the activity level where the predictive $F_{10.7}$ algorithm struggles most and is most likely to overpredict $F_{10.7}$ in a deterministic forecast. However, as seen back in Figure 3, the true variation is well-captured by the 3σ bounds of the probabilistic samples. Therefore, the true position, at least for HASDM-ML, should be captured.

This is the case in both Figure 9 and Table 6, where the bias is consistently significantly smaller than the other two models. For this condition, MSIS underpredicts density for each altitude at both epochs. However, the bias is small and the distributions consistently look similar to that of HASDM-ML.

The similarity in HASDM-ML and MSIS distributions is confirmed in Table 6. At 400 km, the bias for TIE-GCM ROM is critically large again, and it proves to be extremely sensitive to the inputs at these magnitudes with a standard deviation of over 17 km. TIE-GCM ROM's high sensitivity to

Table 6. Distribution statistics for high solar activity (Figure 9).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	-0.1434	0.6995	-1.6310	4.2962
	600 km	-0.0132	0.0693	-0.1601	0.4236
	800 km	-0.0018	0.0092	-0.0211	0.0549
MSIS	400 km	1.0811	1.1546	5.7482	6.4553
	600 km	0.0955	0.1023	0.5128	0.5727
	800 km	0.0368	0.0106	0.1616	0.0598
TIE-GCM ROM	400 km	-21.5330	2.3395	-94.9865	17.4158
	600 km	-1.1635	0.2376	-5.3222	1.7692
	800 km	0.0473	0.0203	0.1431	0.1548

$F_{10.7}$ at all altitudes for this condition will result is vastly different Pc estimates relative to using either of the other two models. Similar to elevated solar activity, the effect of the extrapolation shows misleading good performance at 800 km. Back in Figure 5, the density profiles at forecast epoch emulate what is seen here in the position distributions for this case.

The remaining figures and tables are for the three geomagnetic conditions when $F_{10.7}$ is kept at the true values for the time period, and ap is sampled from the probabilistic distributions. The three cases were chosen to get a sense of the different combinations of solar and geomagnetic activity. The first case, which results are displayed in Figure 10 and Table 7, is moderate solar and moderate geomagnetic activity.

HASDM-ML and MSIS have similar sensitivities to these ap samples in this case. The bias in MSIS is much more pronounced, but the standard deviation, seen in the table, are nearly identical at the three-day mark and close at six days. TIE-GCM ROM is much more sensitive to the the probabilistic samples, which are interpolated to Kp. In relative terms, TIE-GCM ROM has a much larger standard deviation for 400 km and 600 km, but at 800 km, it is reduced to the level of the other two models. However, the extrapolation is still making the bias become more positive with increasing altitude.

Table 7. Distribution statistics for moderate solar / moderate geomagnetic activity (Figure 10).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	-0.6155	0.4697	-0.9924	1.4001
	600 km	-0.0270	0.0193	-0.0494	0.0558
	800 km	-0.0020	0.0014	-0.0037	0.0039
MSIS	400 km	-4.0117	0.4600	-16.7358	1.6527
	600 km	-0.2119	0.0199	-0.8674	0.0708
	800 km	-0.0048	0.0015	-0.0214	0.0054
TIE-GCM ROM	400 km	-8.9636	1.1085	-23.1105	6.0366
	600 km	-0.1519	0.0393	-0.1742	0.1868
	800 km	0.0402	0.0014	0.1715	0.0060

The TIE-GCM ROM distributions are fairly Gaussian in nature at three days, but after the last three days, it changes in shape more so than for the other models. After three days, there is no

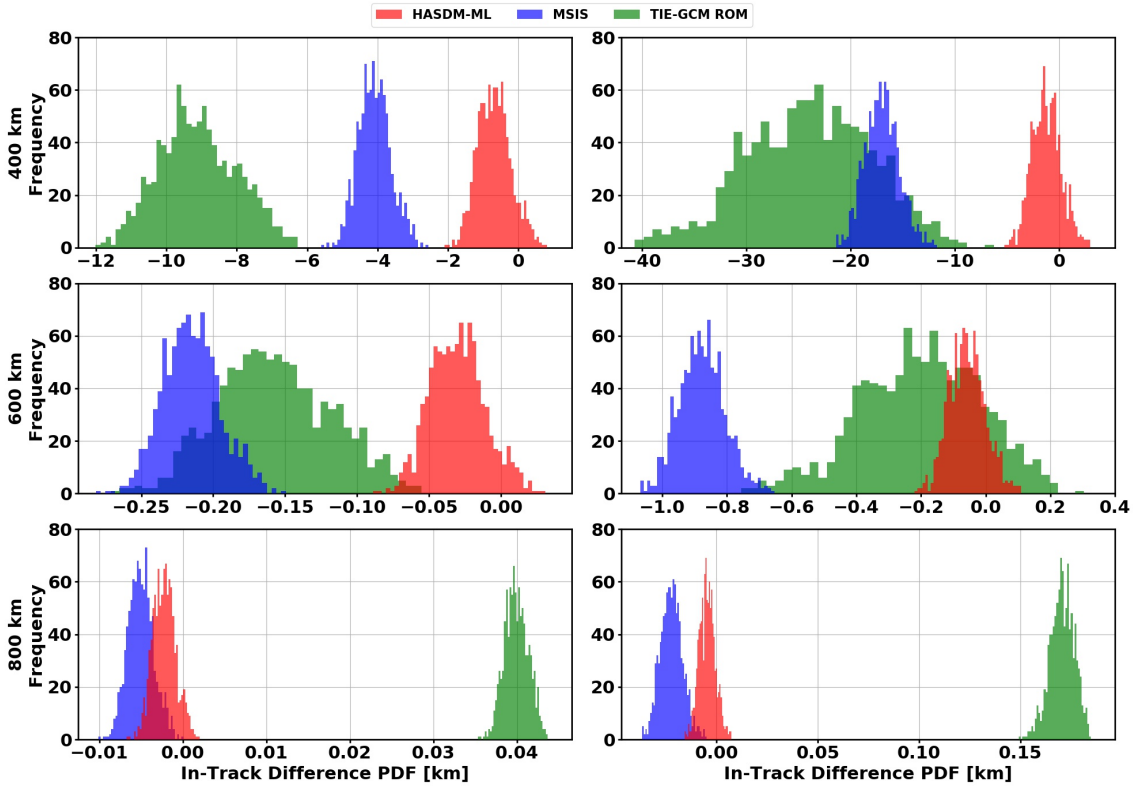

Figure 10. In-track position error distributions after three days (left) and six days (right). This is for moderate solar / moderate geomagnetic activity with ap being sampled.

deterministic forecast as ap gets set to zero. Therefore, all of the probabilistic samples in that time period are based in background activity for the specific condition. In Figure 10, it does not seem to alter the shape of the distributions for HASDM-ML and MSIS the way it does for TIE-GCM ROM.

When comparing Figure 10 to Figure 7, being in the same solar activity level, there is clearly more model sensitivity to sampling $F_{10.7}$ than ap sampling. However, there is much more spread in the biases of the models in this case when ap is sampled. The last key takeaway from this condition is the tendency for TIE-GCM and MSIS to overpredict density within the bounds of the model output.

The next geomagnetic condition to be analyzed is elevated solar and low geomagnetic activity. This represents quiet time during the rise or decline of solar maximum. The results are displayed in Figure 11 and Table 8.

The most distinct feature of Figure 11 is the spread of biases between models and the similarity in standard deviation. For all altitudes and both epochs, MSIS is overpredicting density. The MSIS bias at 400 km at the end of the propagation is almost 60 km. At both epochs, the MSIS bias is about three times the magnitude of TIE-GCM ROM, the model with the next largest bias. However, TIE-GCM ROM is underpredicting density in this case. All models have near-Gaussian distributions but skewed slightly left.

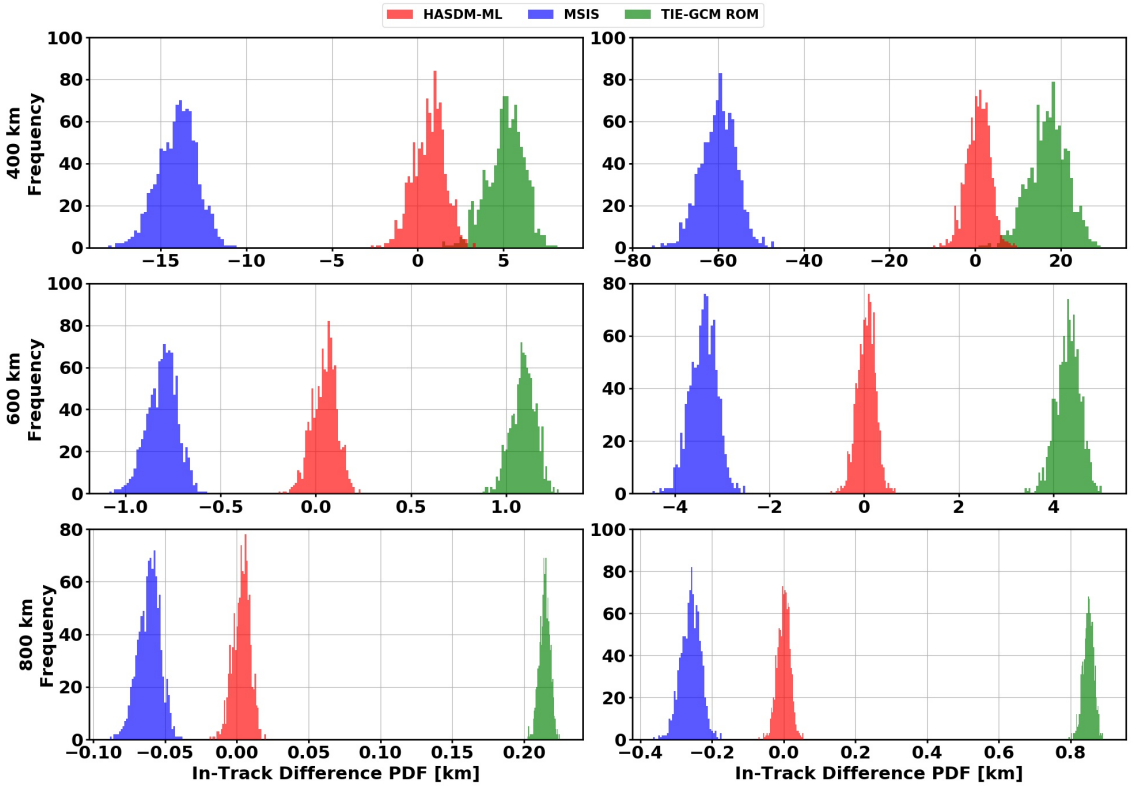

Figure 11. In-track position error distributions after three days (left) and six days (right). This is for elevated solar / low geomagnetic activity with ap being sampled.

At 600 km, the biases for both TIE-GCM ROM and MSIS are much more similar in magnitude, but now, TIE-GCM ROM has a larger absolute bias. The overall spread between the three models is much smaller, but that is a byproduct of the lesser effect of drag at this higher altitude. This is further exemplified by the spread of the models at 800 km. Relative to 400 km, the spread of the three models is two orders of magnitude lower at the highest altitude. Even still, the positions are distributed over a 1 km range after the propagation in the in-track direction which in terms of SSA and STM is a large area that these satellites may be.

Interestingly, in all cases for this condition, HASDM-ML has a lower standard deviation than MSIS. This is so striking, because MSIS takes daily average ap, or Ap, as its input for geomagnetic activity, where HASDM-ML takes each 3-hourly ap value. Even though the input to HASDM-ML is more dynamic, MSIS still has more sensitivity to geomagnetic activity, at least for this condition.

The last condition to be analyzed is high solar activity and moderate geomagnetic activity. Looking at Figure 4, the true $F_{10.7}$ used as an input for this period does not classify as high solar activity. However, the deterministic variation (not shown) has a value one day from forecast epoch above 190 sfu, and that causes ap to be sampled from the distribution based on high solar and moderate geomagnetic activity. These results are shown in both Figure 12 and 9.

As it was with the previous geomagnetic condition, the position error distributions has vastly different biases. For MSIS and TIE-GCM ROM (within its output bound grid), there is a strong

Table 8. Distribution statistics for elevated solar / low geomagnetic activity (Figure 11).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	0.8168	0.9127	1.2659	2.7447
	600 km	0.0566	0.0650	0.0820	0.1945
	800 km	0.0043	0.0059	0.0033	0.0178
MSIS	400 km	-13.8208	1.1483	-59.2142	4.1207
	600 km	-0.7935	0.0794	-3.3445	0.2837
	800 km	-0.0598	0.0078	-0.2547	0.0277
TIE-GCM ROM	400 km	5.2923	1.0633	17.8542	4.4484
	600 km	1.1019	0.0645	4.3543	0.2578
	800 km	0.2150	0.0037	10.8524	0.0150

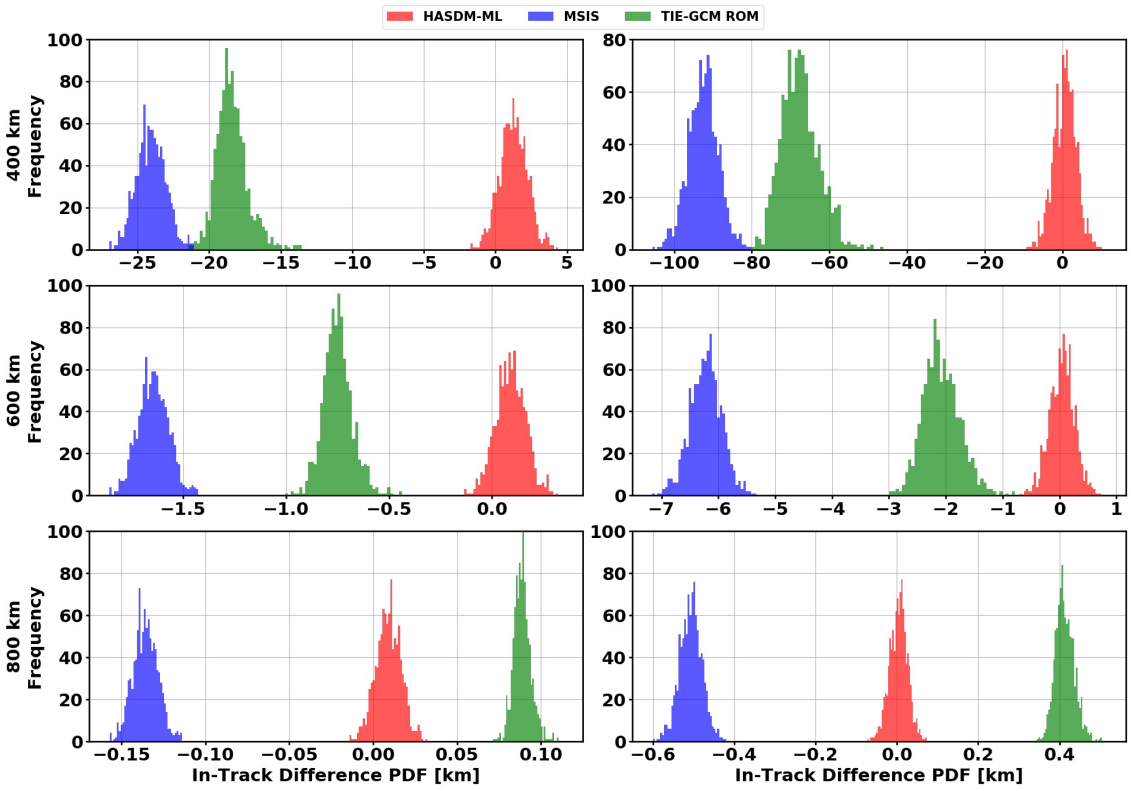

Figure 12. In-track position error distributions after three days (left) and six days (right). This is for high solar / moderate geomagnetic activity with ap being sampled.

tendency to overpredict density. This is the highest bias for MSIS out of any case, being almost 100 km. As previously stated, the values of $F_{10.7}$ for this case classify it as elevated solar activity, but ap is being sampled as if it is high solar activity. Therefore, it is more appropriate to compare these results to Figure 8 (elevated solar) than Figure 9 (high solar). At high solar activity conditions, the MSIS distribution is similar to HASDM-ML's, but MSIS' density predictions were lower than HASDM-ML. At elevated solar activity, MSIS was overpredicting density relative to HASDM-ML

had a similar bias to that of this geomagnetic case with a similar nominal $F_{10.7}$ variation.

This is another case whose results closely resemble the density profiles at forecast epoch. Referring back to Figure 5, it becomes evident that at these conditions, the models respond in a uniform manner across the forecast period. At the lowest altitude, MSIS and TIE-GCM ROM have similar density outputs which are larger than HASDM-ML. With increased altitude, the relative bias between MSIS and HASDM-ML stays the same, yet the extrapolation for TIE-GCM ROM causes its density output to change from being similar to MSIS to being more similar to HASDM-ML.

By looking at where the true position lies within the HASDM-ML distribution at both epochs, it is evident that the mean ap samples are lower than true for the first three days, resulting in smaller densities and probabilistic positions that are behind the true position. By the end of the forecast, the mean samples must be higher than true, because there are higher densities causing the satellites to drop and speed up, relative to true. At the end of the period, the mean HASDM-ML positions are closer to true than at the three-day mark. This also reiterates how important it is to know the range of positions the satellite may be at every point in the propagation. Assuming a more uniform uncertainty would result in a misrepresentation of possible satellite positions.

Table 9. Distribution statistics for high solar / moderate geomagnetic activity (Figure 12).

		3 Days		6 Days	
		μ (km)	σ (km)	μ (km)	σ (km)
HASDM-ML	400 km	1.4088	0.9696	1.1345	2.9469
	600 km	0.1023	0.0727	0.0798	0.2180
	800 km	0.0103	0.0075	0.0080	0.0222
MSIS	400 km	-23.917	1.0201	-91.8753	3.8552
	600 km	-1.6345	0.0744	-6.1746	0.2772
	800 km	-0.1346	0.0074	-0.5023	0.0274
TIE-GCM ROM	400 km	-18.2004	1.1054	-66.6659	4.9012
	600 km	-0.7348	0.0722	-2.0217	0.3246
	800 km	0.0898	0.0051	0.4173	0.0232

In Table 9, the exponential growth of the MSIS bias between the two epochs becomes more evident. The bias nearly quadrupled from three to six days from forecast epoch. There is a similar relative growth in the TIE-GCM ROM bias, but it is not at strong. It is also noted that HASDM-ML is the least sensitive to the input samples in nearly all cases.

CONCLUSIONS

These three density models, while each useful for different reasons and in certain applications, have strong underlying differences. First and foremost, TIE-GCM is a physics-based model where MSIS is empirical, and HASDM is an assimilative model. Therefore, TIE-GCM is does not have actual measurements embedded in it. It may better represent density in a physical sense, but there are conditions and areas where its accuracy is not as high. MSIS is derived from real measurements to tune the model's parameters and is quick to evaluate due to its form. HASDM is assumed to be the most accurate as it uses an empirical model along with real-time assimilation from calibration satellites. The main issue in using HASDM is that it is not publicly available as the other two are.

The benefit to using the two surrogate models, TIE-GCM ROM and HASDM-ML, is that they are both based in machine learning and can be evaluated for an entire week within a fraction of a

second. This makes a thorough analysis like this feasible. By comparing the model performance and their impact on orbit determination across different conditions, it becomes evident that there are significant biases and varying sensitivity for each model across the different SW inputs.

This analysis showed that sampling $F_{10.7}$ had a stronger impact on the resulting densities than sampling ap, even though it still highlighted the importance of sampling both drivers. MSIS proved to overpredict density stronger than it underpredicts density. It also overpredicted more frequently than not. In general, MSIS had similar input sensitivity to HASDM-ML, represented by the distributions' standard deviations. In the forecast epoch density profiles in Figure 5, MSIS and HASDM-ML have similar trends in density decay with altitude, even when there was a bias. In contrast, TIE-GCM ROM had to be extrapolated beyond 450 km, which resulted in mostly underpredicted densities. This proved to make its use unreliable above the model bounds. However, as previously noted, the proper method to extrapolate density for TIE-GCM is by extrapolating the number densities, not mass density. However, the ROM only outputs mass density.

For the final three cases, $F_{10.7}$ was kept at its true varitation and ap was sampled to see driver and model uncertainty based on geomagnetic activity. This showed fewer conditions with overlapping position error distributions. In fact, most cases had vastly different biases causing segmented distributions. TIE-GCM ROM proved to be the most sensitive to geomagnetic activity, and surprisingly, MSIS seemed slightly more sensitive to geomagnetic activity than HASDM-ML even though it takes Ap instead of ap as an input.

The difference in standard deviation, seen in the Figures 6-12 and Tables 3-9, will have a strong impact on the estimation of collision probability. The variability between the models is stark for many conditions and the choice of models is important in the decision making process for operators. In the future, we plan to use these methodologies and information to compute collision probability. There is some debate over how to compute Pc, but we will expand upon this work to incorporate these important calculations into the framework.

We plan to verify the input sensitivity of TIE-GCM and compare it to the results of TIE-GCM ROM. In addition, data assimilation will be performed on TIE-GCM ROM at solar minimum to study the possible improvements to the model and reduction of the large biases exhibited in many of the conditions. More work will be done with HASDM-ML, furthering its capability and making it publicly available for the research community.

ACKNOWLEDGMENT

This research was made possible by NASA West Virginia Space Grant Consortium, Training Grant #NNX15AI01H and NASA Agreement #80NSSC20M0055.

REFERENCES

[1] Space Weather Prediction Center, "Satellite Drag," 2019. https://www.swpc.noaa.gov/impacts/satellite-drag.

[2] W. K. Tobiska, S. D. Bouwer, and B. R. Bowman, "The development of new solar indices for use in thermospheric density modeling," *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 70, No. 5, 2008, pp. 803 – 819, https://doi.org/10.1016/j.jastp.2007.11.001.

[3] W. K. Tobiska, D. Knipp, W. J. Burke, D. Bouwer, J. Bailey, D. Odstrcil, M. P. Hagan, J. Gannon, and B. R. Bowman, "The Anemomilos prediction methodology for Dst," *Space Weather*, Vol. 11, No. 9, 2013, pp. 490–508, 10.1002/swe.20094.

[4] W. K. Tobiska, S. D. Bouwer, and B. R. Bowman, "The development of new solar indices for use in thermospheric density modeling," *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 70, No. 5, 2008, pp. 803 – 819, https://doi.org/10.1016/j.jastp.2007.11.001.

[5] W. McClain and D. Vallado, *Fundamentals of Astrodynamics and Applications*. Space Technology Library, Springer Netherlands, 2001.

[6] R. J. Licata, W. K. Tobiska, and P. M. Mehta, "Benchmarking Forecasting Models for Space Weather Drivers," *under review, Space Weather*, 2020, arXiv:2003.04743v1.

[7] C. D. Bussy-Virat, A. J. Ridley, and J. W. Getchius, "Effects of Uncertainties in the Atmospheric Density on the Probability of Collision Between Space Objects," *Space Weather*, Vol. 16, No. 5, 2018, pp. 519–537, 10.1029/2017SW001705.

[8] C. He, Y. Yang, B. Carter, E. Kerr, S. Wu, F. Deleflie, H. Cai, K. Zhang, L. Sagnières, and R. Norman, "Review and comparison of empirical thermospheric mass density models," *Progress in Aerospace Sciences*, Vol. 103, 2018, pp. 31 – 51, https://doi.org/10.1016/j.paerosci.2018.10.003.

[9] L. Sagnieres and I. Sharf, "Uncertainty Characterization of Atmospheric Density Models for Orbit Prediction of Space Debris," *7th European Conference on Space Debris*, 2017.

[10] S. L. Bruinsma and J. M. Forbes, "Medium- to large-scale density variability as observed by CHAMP,"

[11] F. Schiemenz, J. Utzmann, and H. Kayal, "Propagation of grid-scale density model uncertainty to orbital uncertainties," *Advances in Space Research*, Vol. 65, No. 1, 2020, pp. 407 – 418, https://doi.org/10.1016/j.asr.2019.10.013.

[12] L. Qian, A. Burns, B. Emery, B. Foster, G. Lu, A. Maute, A. Richmond, R. Roble, S. Solomon, and W. Wang, "The NCAR TIE-GCM: A community model of the coupled thermosphere/ionosphere system," *Geophysical Monograph Series*, Vol. 201, 01 2013, pp. 73–83, 10.1029/2012GM001297.

[13] A. E. Hedin, "Extension of the MSIS Thermosphere Model into the middle and lower atmosphere," *Journal of Geophysical Research: Space Physics*, Vol. 96, No. A2, 1991, pp. 1159–1172, 10.1029/90JA02125.

[14] M. Storz, B. Bowman, and J. Branson, *High Accuracy Satellite Drag Model (HASDM)*. 2005, 10.2514/6.2002-4886.

[15] R. J. Licata and P. M. Mehta, "Physics-informed Machine Learning for Probabilistic Space Weather Modeling and Forecasting: Thermosphere and Satellite Drag," 2019, 10.13140/RG.2.2.32538.18880.

[16] J. M. Picone, A. E. Hedin, D. P. Drob, and A. C. Aikin, "NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues," *Journal of Geophysical Research: Space Physics*, Vol. 107, No. A12, 2002, pp. SIA 15–1–SIA 15–16, 10.1029/2002JA009430.

[17] J. C. G. Walker, "Analytic Representation of Upper Atmosphere Densities Based on Jacchia's Static Diffusion Models," *Journal of the Atmospheric Sciences*, Vol. 22, 07 1965, pp. 462–463, 10.1175/1520-0469(1965)022¡0462:AROUAD¿2.0.CO;2.

[18] "Some Problems concerning the Terrestrial Atmosphere above about the 100 km Level," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 253, No. 1275, 1959, pp. 451–462.

[19] P. M. Mehta and R. Linares, "A New Transformative Framework for Data Assimilation and Calibration of Physical Ionosphere-Thermosphere Models," *Space Weather*, Vol. 16, No. 8, 2018, pp. 1086–1100, 10.1029/2018SW001875.

[20] P. M. Mehta and R. Linares, "Data-driven framework for real-time thermospheric density estimation," *Accepted, JGCD*, 2018.