AAS 21-415

IMPACT OF DRIVER AND MODEL UNCERTAINTY ON DRAG AND ORBIT PREDICTION

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

The capability to accurately model and subsequently forecast thermospheric density is imperative for satellite orbit prediction and conjunction operations. Two of the major sources of uncertainty are Space Weather (SW) driver uncertainty and the underlying uncertainty within the density models, called model uncertainty. These uncertainties propagate into drag estimates and therefore satellite position estimates, which require confidence in decision making with regards to avoidance maneuvers. In this paper, we quantify the impact of SW driver forecast uncertainty and, for the first time, model uncertainty on drag and orbit prediction.

INTRODUCTION

With the rapid increase of satellite launches in recent years, Space Situational Awareness (SSA) has become a major focus in the effort to protect assets and humans in space. There are a plethora of existing atmospheric density models that are used to make forecasts employed by orbit determination algorithms. Two of the subsets of these models are physics-based and empirical. These types of models each have their strengths and weaknesses and both have their place in science and operations.

Physics-based models, such as TIE-GCM[1] and GITM,[2] solve continuity and other physical equations on a global grid in a dynamic fashion. For this reason, they are excellent to use for scientific investigation of short and long-term density trends. Meanwhile, these models are often computationally expensive and require a vast number of resources for parallelization. This can make uncertainty assessment infeasible in operations. Empirical models, like JB2008[3] or NRMSISE00,[4] use various measurements to solve parametric equations. They can provide local density values or multi-dimensional arrays to form global density grids. A key benefit to using an empirical model is the brief evaluation time. They are often used in data analysis and in orbit propagation. Assimilative models, most notably HASDM,[5] have a background density model that is corrected using real-time density estimates to reduce bias and achieve greater accuracy in forecasts.

All density models have multiple sources of uncertainty that weaken our confidence in their prediction. Two prominent sources of uncertainty are SW driver uncertainty and model uncertainty. In previous work, we quantified the effect of driver uncertainty on drag and orbit prediction.[6] Current capabilities to forecast SW indices and proxies (e.g. $F_{10.7}$ and ap) are imperfect and we had

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

quantified the effects on satellite position for certain conditions. Model uncertainty stems from our incomplete knowledge of how the thermosphere is impacted by the aforementioned drivers and how interactions within the thermosphere create density gradients and various phenomena. Model uncertainty is unique to each model and is not something that can be easily obtained with current models.

In this work, we will quantify the impact of both SW driver and model uncertainty on satellite positions using derivatives of TIE-GCM and HASDM. Similar to previous work, we will look at multiple time periods that span a diverse set of Space Weather conditions in order to compare the effects of each uncertainty source and observe the combined effect.

METHODOLOGY

There are two major components to conduct this analysis. First, a method to create probabilistic driver forecasts is required. Previously, the authors used six years of SW driver forecasts, provided by Space Environment Technologies (SET), to acquire temporal error statistics over multiple solar and geomagnetic conditions.[7] Using these statistics, along with the original driver forecasts, probabilistic samples can be generated for use in Monte Carlo analyses.

The second component is to extract uncertainty statistics from the density model itself (model uncertainty). Since no widely-used atmospheric density models provide uncertainty estimates, we leveraged machine learning techniques to generate a probabilistic model using Monte Carlo (MC) dropout.[8,9]

Probabilistic Input Sampling

There are seven time periods chosen for this analysis. The driver forecasts that were available span from October 2012 to December 2019. The first four cover the various solar activity levels, displayed in Table 1. For these conditions, $F_{10.7}$ is sampled with respect to the deterministic forecast while ap is kept at its true values. The last three cases are a combination of different solar and geomagnetic conditions, where ap is sampled relative to the deterministic forecast and the true $F_{10.7}$ variation is used.

Table 1. 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	$50 < ap$

For more information on the error statistics of the driver forecasts or on the generation of the probabilistic forecasts, refer to previous work.[6,7] Figure 1 shows the SW drivers used in the solar activity cases.

In each case, there is a deviation between the deterministic forecast and the true variation of $F_{10.7}$. In the low solar activity example, the forecast does not deviate much, but for the other three cases, a significant error accumulates. However, the true variation is well-contained within the 3σ bounds.

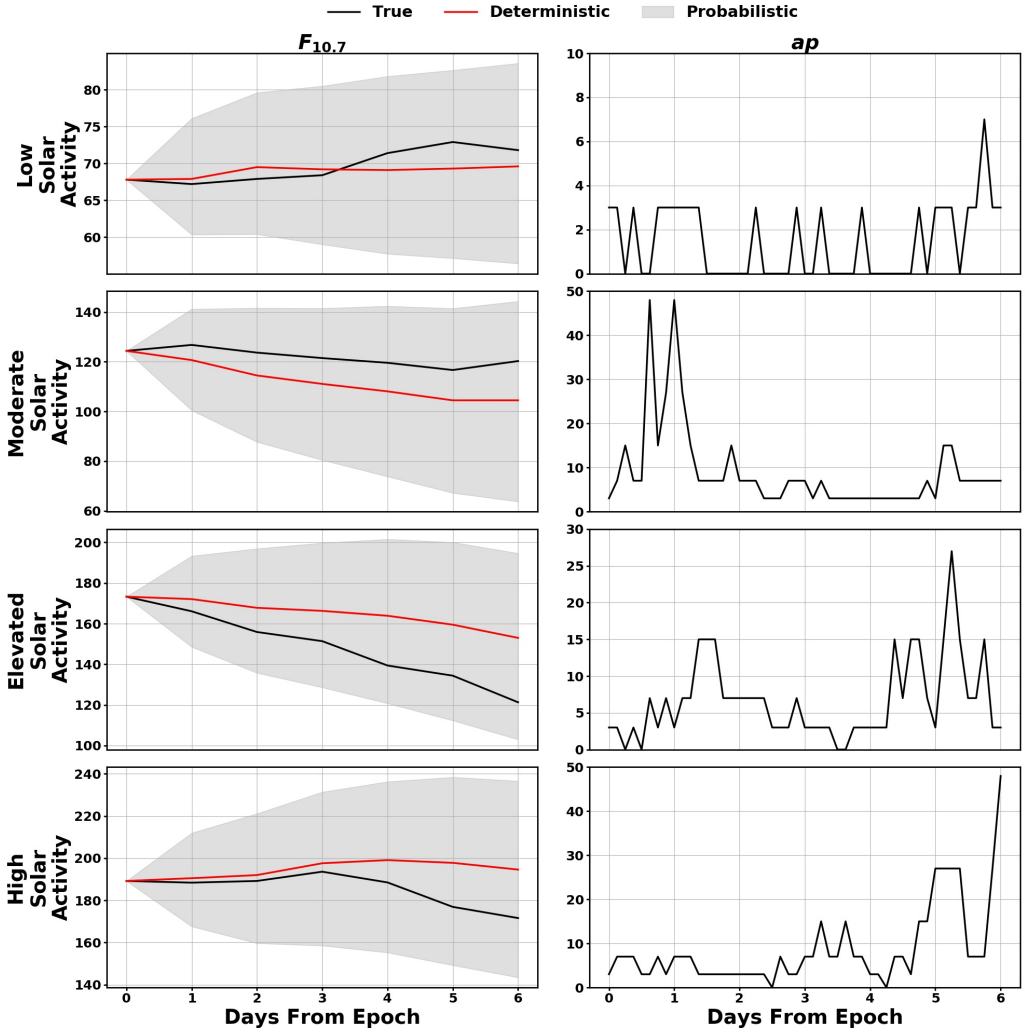

Figure 1. Space Weather inputs for the four solar cases. The shaded probabilistic region shows the 3σ bounds for the probabilistic samples.

Previously, we observed that the ap forecast errors had varying behavior that was dependent on both solar and geomagnetic activity. Therefore, ap forecast error is split into twelve distributions rather than three. The last three conditions used for this work are as follows: 1.) moderate solar / moderate geomagnetic activity, 2.) elevated solar / low geomagnetic activity, and 3.) high solar / moderate geomagnetic activity. The drivers for the three cases are shown in Figure 2.

The geomagnetic activity level is determined by the maximum value in the deterministic forecast. There is also an evident drop to zero in all three deterministic forecasts. The forecasts SET uses for ap come from NOAA's Space Weather Prediction Center (SWPC), and they only extend three days from forecast epoch. The error distributions were created with respect to a six-day window, so the

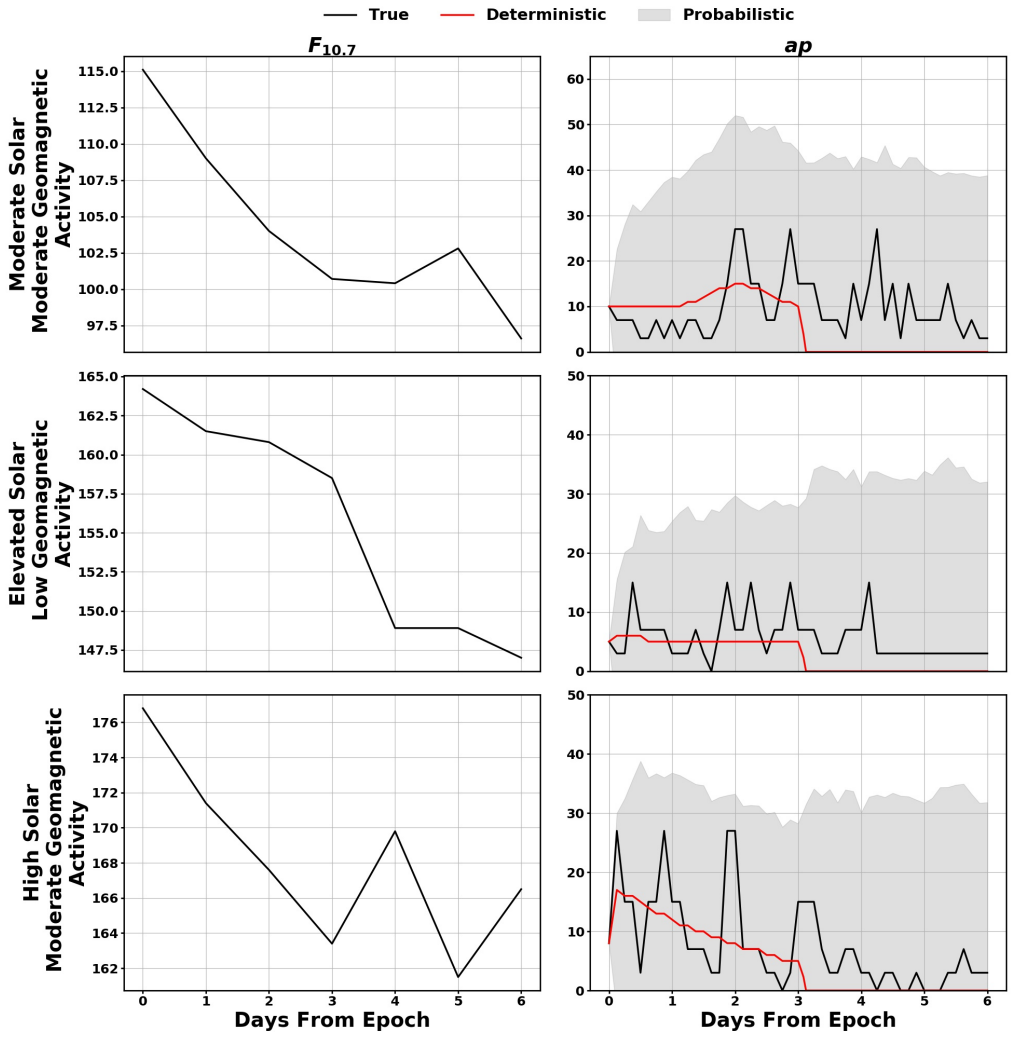

Figure 2. Space Weather inputs for the three geomagnetic cases. The shaded probabilistic region shows the 3σ bounds for the probabilistic samples.

last three days essentially sample background ap for that condition.

Monte Carlo Dropout

Dropout is a regularization tool commonly used in machine learning (ML) to prevent overfitting to the training data.[10] In a ML model, each upstream layer sends the output from all nodes to each node in the subsequent layer, where weights and biases are introduced to generate an output for each of that layer's nodes. When a model has too many parameters, it can overfit the function to the training data and provide poor performance when interpolating or extrapolating, similar to using a high-order polynomial to fit experimental data.

Dropout layers use Bernoulli distributions, one for each upstream node, with probability *P*. Every time a set of inputs is run through the model during training, each distribution is sampled, and if a sample is "true", that node's output is nullified. Each output from a given layer is then scaled-up based on the number of nullified outputs. Dropout is thought to make each node more independently sufficient and not reliant on the outputs of other nodes in the layer.[11]

This feature, by default, is only active during training making the resulting model deterministic. However, if it is manually set to always be on, the model's outputs are probabilistic, because different nodes are nullified each time a set of inputs are given to the model. With a sufficient number of runs, there is a distribution of model outputs for each input set. This is referred to as Monte Carlo dropout. The effect is that every prediction with the MC dropout represents a different functional representation of the model fit to the data; MC dropout is a Bayesian approximation of Gaussian Process.[8]

Density Models

The two models used in this study are TIE-GCM and HASDM. As previously mentioned, TIE-GCM is a physics-based model that evolves numerous species densities based on a given set of SW inputs. Due to its form, it is computationally expensive and does not provide uncertainty estimates. We recently developed a Long-Short Term Memory (LSTM) neural network model based on TIE-GCM outputs and used dropout as a surrogate for TIE-GCM that can provide uncertainty estimates. This model is referred to as TIE-GCM reduced-order Bayesian emulator (TIE-GCM ROBE).[12] In a similar previous paper, we used TIE-GCM ROM which is another LSTM with an identical structure, but it did not utilize dropout during training. This results in a slightly different model. When comparing the driver uncertainty distributions at 400 km in that paper to the results here, a minor difference is apparent.

HASDM is not a publicly available model; however, SET recently released a database containing twenty years of density data, called the SET HASDM Density Database.[13] We trained a neural network on this database, called HASDM-ML,[14] that is emblematic of an assimilated empirical model. This model also utilizes dropout to provide uncertainty estimates. The inputs for both models are displayed in Table 2. For clarity, UTC refers to coordinated universal time and doy refers to the day of year.

Table 2. Inputs for the two models used in the study.

TIE-GCM ROBE	$F_{10.7}$, Kp, UTC, doy
HASDM-ML	$F_{10.7}$, $F54b$, ap, UTC, doy

A benefit of both models is their ability to replicate their original models with less than 15% mean error and provide uncertainty estimates if desired. For this work, each model was evaluated 1,000 times for 145 time-steps (per case) in approximately three minutes without need for parallelization. This is an advantage of leveraging ML.

To visualize the uncertainty bounds that can be extracted from using MC dropout for these two models, refer to Figure 3. It shows the density predictions for a single location over a seven-day period using both models.

There is a notable difference in the 3σ bounds between the two models. First, the uncertainty grows with time in conjunction with the inputs for TIE-GCM ROBE while the uncertainty for

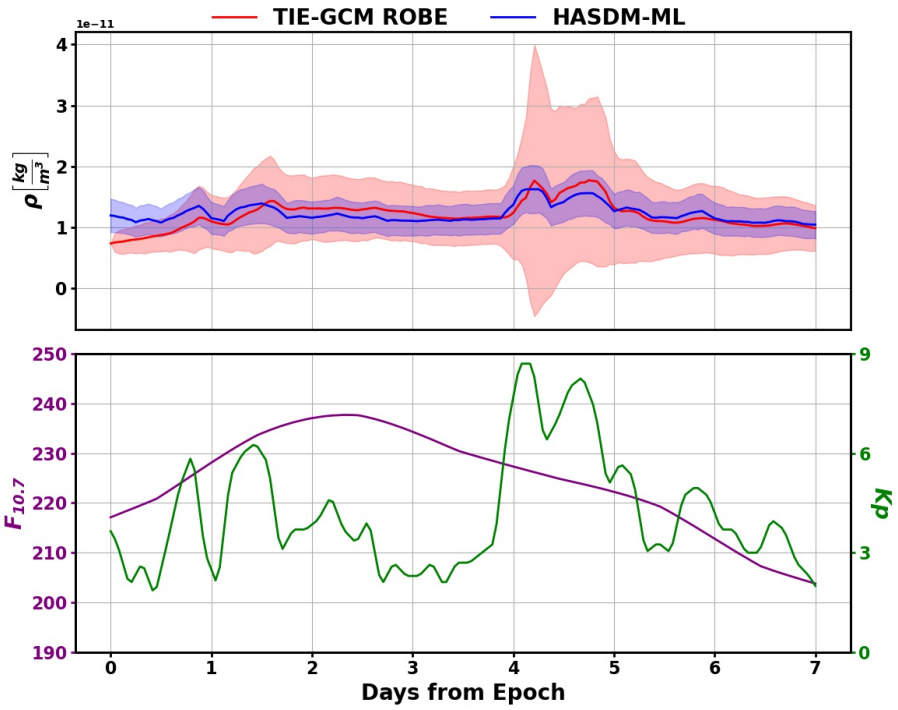

Figure 3. Local density prediction with uncertainty bounds for TIE-GCM ROBE and HASDM-ML.

HASDM-ML is simply a function of the inputs. This is the difference between the uncertainty in a dynamic (TIE-GCM ROBE) and a static (HASDM-ML) model. The uncertainty for TIE-GCM ROBE is also considerably larger during peak storm-time. This is likely due to the difference in training samples between the two models. HASDM-ML is trained on 17 years (nearly 50,000 samples) of the SET HASDM database, while TIE-GCM ROBE is only trained on one year (8,760 samples) of TIE-GCM outputs using simulated inputs.[15] The fewer number of training samples still provides enough information to train a model with sufficient performance on independent data, and further analysis is required to investigate the effect of the number of training samples on the resulting model uncertainty. Lastly, we have not yet performed a quantitative calibration of the uncertainty estimates.

Orbit Propagation

The two models provide a global density grid. The drivers were interpolated to an hourly cadence and run through each model. The density grid remains unchanged as a small theoretical satellite is propagated through it, and the grid updates each hour over the six-day period. The satellites all have identical initial conditions that provide an orbit which oscillates in altitude between ∼390 km and ∼410 km. The propagation algorithm accounts for two-body forces, J2, and atmospheric drag.

At both the three and six-day marks, the Earth-centered inertial (ECI) positions are saved. We use the position and velocity that result from HASDM-ML densities using the true inputs as the baseline position from which a transformation matrix is generated. In-track position differences are

computed with respect to the baseline for the three uncertainty conditions: driver uncertainty, model uncertainty, and the combined effect.

The positions resulting from the driver uncertainty cases use the models in deterministic form and the probabilistic forecasts. For model uncertainty, the true drivers are used, and the model is utilized with MC dropout. For the third case, the probabilistic forecasts are input to the model with MC dropout.

RESULTS: UNCERTAIN SOLAR ACTIVITY

The results of this study show both the effects of driver uncertainty and model uncertainty. In-track position error histograms will be shown for the three uncertainty cases 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 types of uncertainty.

For all of the proceeding figures, necessary context can be found in Figures 1-3 in order to see what the inputs are (both true and probabilistic) along with the difference in model uncertainty between HASDM-ML and TIE-GCM ROBE. The first condition to be examined is low solar activity. The result for this condition are shown in Figure 4 and Table 3.

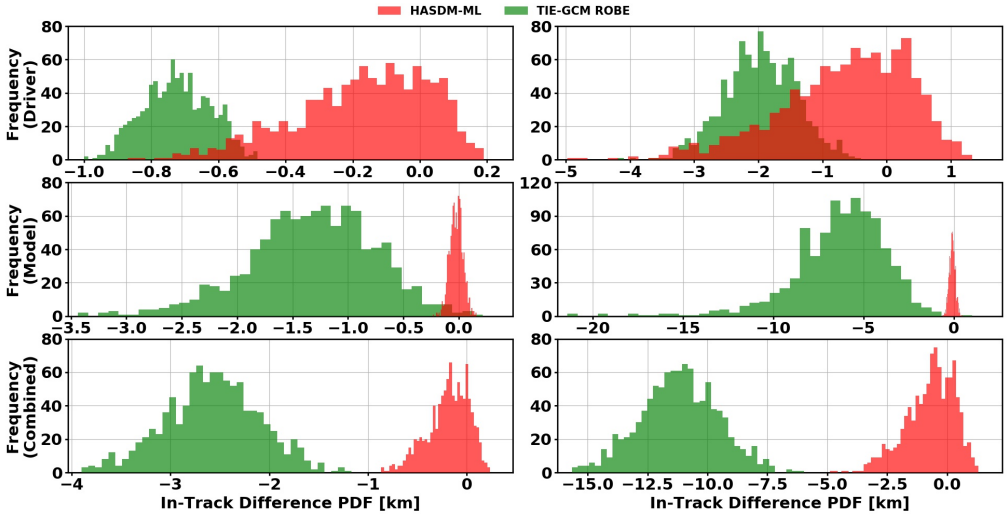

Figure 4. In-track position error distributions after three days (left) and six days (right). This is for low solar activity with $F_{10.7}$ being sampled.

When comparing the driver uncertainty distributions, TIE-GCM ROBE has a slight bias to over-predict density relative to the reference case. When the model produces higher densities, the spacecraft loses energy and therefore altitude. In order to conserve angular momentum, the spacecraft then speeds up which causes it to be ahead of the reference position. This results in a negative in-track difference based on how it was computed. This bias increases from three to six days, but the distributions of the two models have more overlap by the end of the period.

In the second row, large discrepancy in the distributions stand out. TIE-GCM ROBE positions are far more spread out than those from HASDM-ML at both epochs, and their mean positions are more spread out. In this case, all 1,000 runs for both models are done with a single set of inputs meaning the models treat them differently. TIE-GCM ROBE is overpredicting density and shows significantly more uncertainty.

Table 3. Distribution statistics for low solar activity (Figure 4).

		3 Days		6 Days	
	Uncertainty	μ **(km)**	σ **(km)**	μ **(km)**	σ **(km)**
HASDM-ML	Driver	-0.1451	0.1996	-0.5032	1.0126
	Model	-0.0121	0.0582	-0.0472	0.1668
	Combined	-0.1579	0.2083	-0.5593	1.0216
TIE-GCM ROBE	Driver	-0.7117	0.0990	-1.8799	0.5629
	Model	-1.2374	0.5732	-5.7487	2.6096
	Combined	-2.5060	0.4771	-10.9365	1.6232

Table 3 shows the mean and standard deviation of all the distributions in Figure 4. The previous notes on the qualitative analysis are upheld by these statistics of the first two uncertainty types. The combined effect of the two on the different models is quite interesting. For HASDM-ML, the mean in-track position for the combined uncertainties is approximately the sum of the individual uncertainties. However, the standard deviation is only slightly larger than the larger of the two individual sources. The mean position for the combined uncertainties for TIE-GCM ROBE is about twice that of the largest individual mean value. Another unique statistic for this model is that the standard deviation is lower for the combined uncertainties is smaller than that of the larger single source. The results at moderate solar activity are displayed in Figure 5 and Table 4.

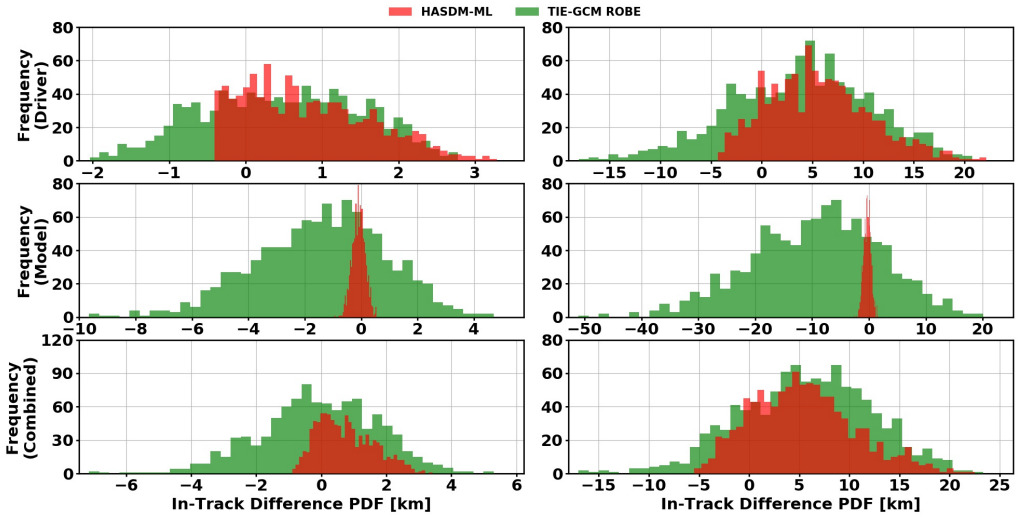

Figure 5. In-track position error distributions after three days (left) and six days (right). This is for moderate solar activity with $F_{10.7}$ being sampled.

There is a peculiar trend with the driver uncertainty case for HASDM-ML at the three-day mark. The distribution is skewed right and there is a drop-off on the left side of this distribution. As

discussed in the previous paper, this is believed to be a combination of the strong storm early in the forecast and the model not being optimized which can cause an abnormal response from the model. By the end of the time period, when there was virtually no geomagnetic activity after the storm, the distribution returns to a Gaussian form.

The two models have similar levels of position uncertainty in response to the probabilistic $F_{10.7}$ variations. This similarity does not carry over to the model uncertainty distributions. Again, TIE-GCM ROBE carries much more uncertainty than HASDM-ML. After six days, the position spread is approximately twice as large as it is for driver uncertainty. The effect on HASDM-ML is the opposite, where the spread is much larger for driver uncertainty than it is for model uncertainty.

Table 4. Distribution statistics for moderate solar activity (Figure 5).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	0.9128	0.8239	6.2897	5.0340
	Model	-0.0484	0.2282	-0.1897	0.6322
	Combined	0.8581	0.8585	6.0889	5.1177
TIE-GCM ROBE	Driver	0.5931	1.0303	4.7901	6.8775
	Model	-1.0831	2.3141	-7.5959	11.4058
	Combined	0.2497	1.8126	6.6931	6.5555

As with the previous condition, the mean in-track positions for HASDM-ML and the combined uncertainties is approximately the sum of the two individual uncertainty sources. The standard deviation follows the same trend as it had for low solar activity, as well. The combined uncertainty mean position for TIE-GCM ROBE is closer to the reference position than either of the uncertainties at three days, but it is the most positive of the three uncertainty cases at six days. The last notable statistic in Table 4 is standard deviation at six days. The combined uncertainties provided a smaller standard deviation than either of the individual uncertainties. This may be caused by an unlikely offsetting effect of the two different uncertainty sources, similar to out of phase sine waves. Next, the impact of $F_{10.7}$ uncertainty on the models for elevated solar activity is shown in both Figure 6 and Table 5.

For the elevated solar activity scenario, TIE-GCM ROBE is visibly more sensitive to the probabilistic driver forecasts than HASDM-ML. There is also a considerable bias, particularly at six days. This was noted in the previous study. The model significantly over-predicts density for this set of inputs. In regards to model uncertainty, HASDM-ML has a reduced position spread for both epochs, while TIE-GCM ROBE shows less impact from model uncertainty than it does for driver uncertainty. This is the first condition where that is true.

Looking at Table 5, the previous trends for HASDM-ML are sustained. For TIE-GCM ROBE, combined mean positions fall between those of the individual uncertainties where the standard deviations are the largest for the combined case. The last case for sampling $F_{10.7}$ is high solar activity, and the results are displayed in Figure 7 and Table 6.

The driver uncertainty distributions for both models are similar to the previous condition with regards to bias and spread. Again, the effect of driver uncertainty is greater than that of model uncertainty on HASDM-ML while the opposite is true for TIE-GCM ROBE. However, at six days the position distribution spread is extremely similar for TIE-GCM ROBE with regards to the uncertainty sources.

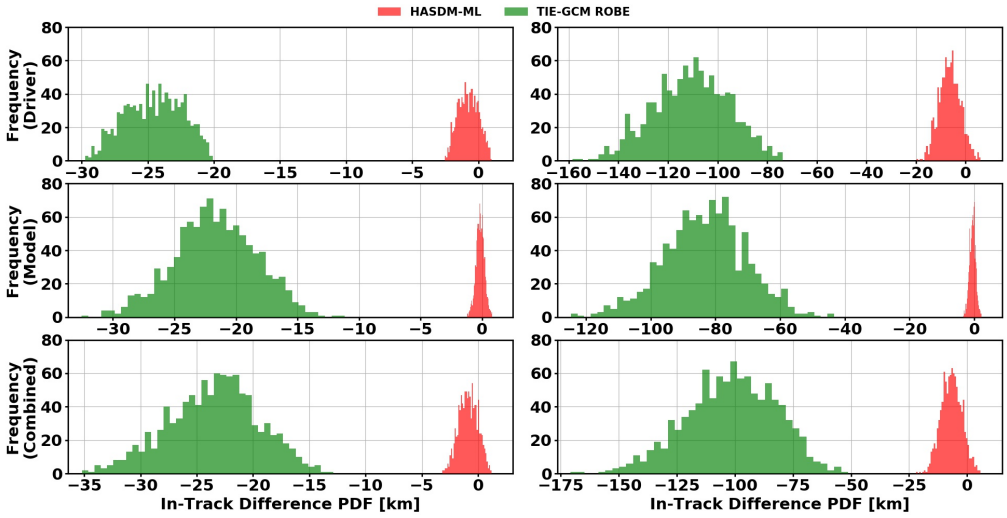

Figure 6. In-track position error distributions after three days (left) and six days (right). This is for elevated solar activity with $F_{10.7}$ being sampled.

Table 5. Distribution statistics for elevated solar activity (Figure 6).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	-0.6724	0.7550	-5.7028	4.2967
	Model	-0.0963	0.3402	-0.3442	0.9447
	Combined	-0.7602	0.8329	-6.0373	4.4323
TIE-GCM ROBE	Driver	-24.3160	2.1598	-108.5845	15.2076
	Model	-21.1565	3.3462	-81.2311	12.6726
	Combined	-22.9693	4.0129	-98.7382	19.3081

Table 6. Distribution statistics for high solar activity (Figure 7).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	-0.1434	0.6995	-1.6310	4.2962
	Model	-0.1389	0.5607	-0.5969	1.6432
	Combined	-0.2798	0.9018	-2.1862	4.6235
TIE-GCM ROBE	Driver	-19.5515	2.3039	-77.7999	15.6343
	Model	-17.7222	4.0500	-64.0621	15.8708
	Combined	-18.4270	4.7575	-70.2415	22.1433

A notable statistic for this high solar activity condition is the similarity in the standard deviation for TIE-GCM ROBE between the driver and model uncertainty cases. This was pointed out in the qualitative analysis, but the raw information reinforces how close they are.

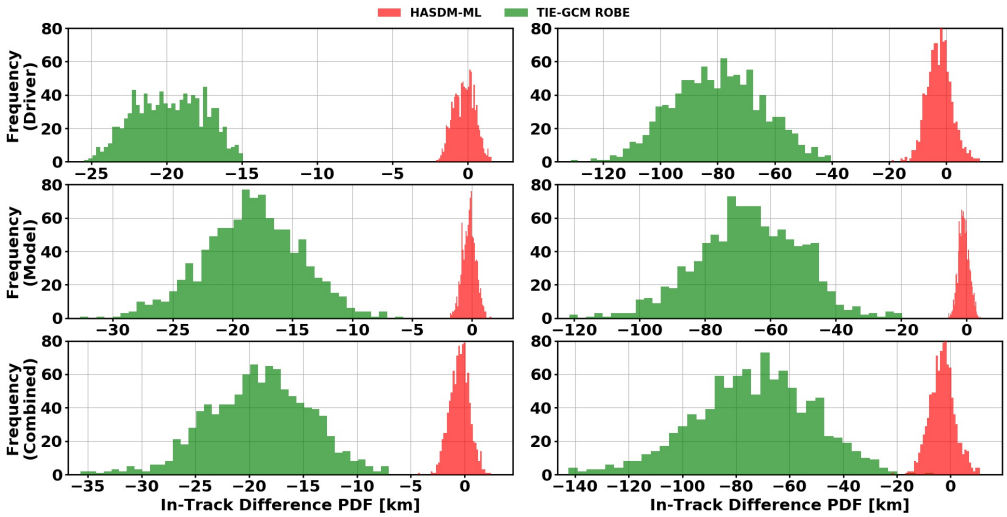

Figure 7. In-track position error distributions after three days (left) and six days (right). This is for high solar activity with $F_{10.7}$ being sampled.

RESULTS: UNCERTAIN GEOMAGNETIC ACTIVITY

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 8 and Table 7, is moderate solar and moderate geomagnetic activity.

Keeping with the trend, TIE-GCM ROBE is over-predicting density in the driver uncertainty case. Although, the spread of the model distributions are less diverse, hinting that there may be a discrepancy between the models' treatment of $F_{10.7}$ and ap samples. Again, there is significantly more model uncertainty in TIE-GCM ROBE than there is for HASDM-ML.

Table 7. Distribution statistics for moderate solar / moderate geomagnetic activity (Figure 8).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	-0.6155	0.4697	-0.9924	1.4001
	Model	-0.0412	0.1987	-0.1615	0.5511
	Combined	-0.6675	0.5087	-1.1819	1.5061
TIE-GCM ROBE	Driver	-10.0313	0.8634	-36.6811	3.5702
	Model	-3.0209	1.0749	-3.4655	3.9822
	Combined	-9.6028	2.2021	-32.9720	9.1021

Table 7 shows similar tendencies for the models in the way the uncertainty statistics compare to each other. There is still an additive nature to HASDM-ML's mean positions and similar uncertainties for its driver and combined uncertainties' standard deviation values. It is important to note that the standard deviation in TIE-GCM ROBE's two uncertainty cases are more similar than they have

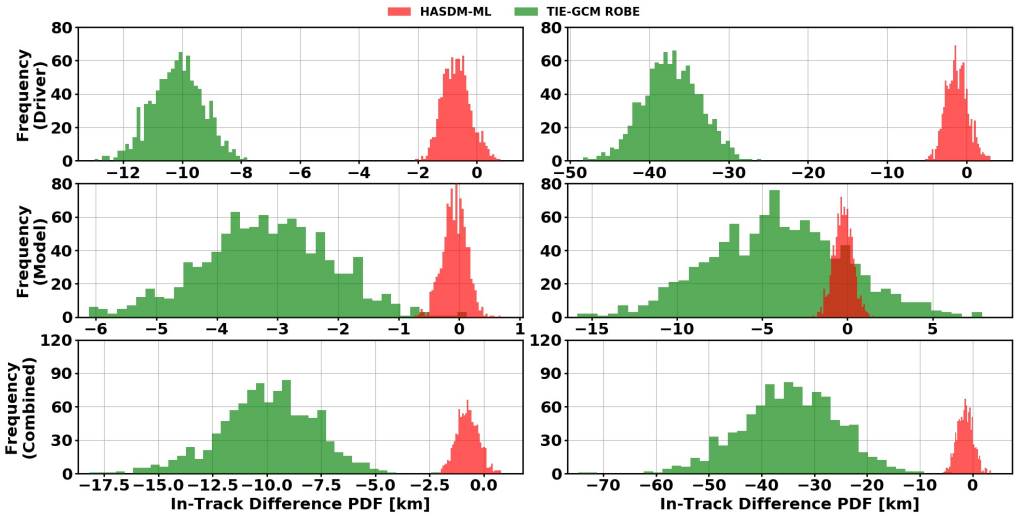

Figure 8. In-track position error distributions after three days (left) and six days (right). This is for moderate solar / moderate geomagnetic activity with ap being sampled.

previously shown to be. 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 9 and Table 8.

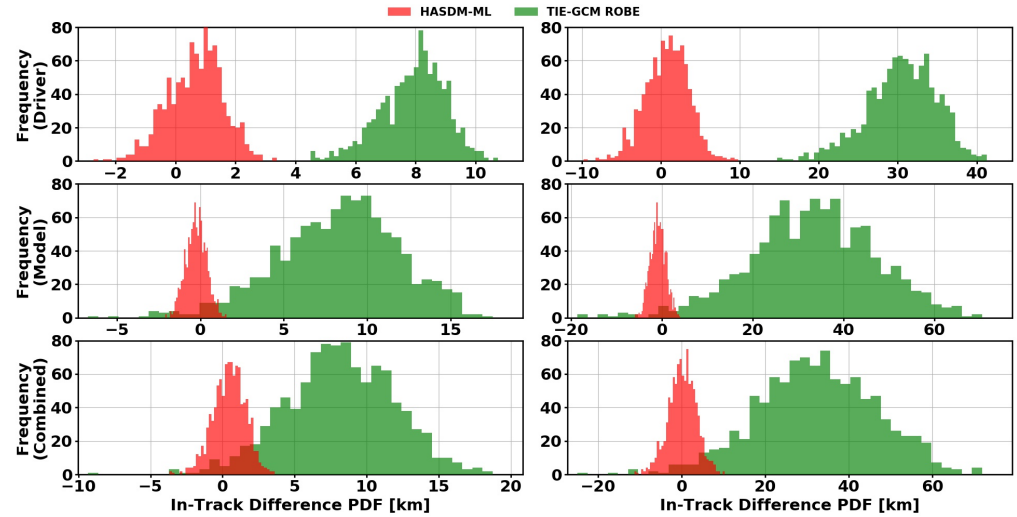

Figure 9. In-track position error distributions after three days (left) and six days (right). This is for elevated solar / low geomagnetic activity with ap being sampled.

The first characteristic to note in Figure 9 is that TIE-GCM ROBE is underpredicting density in this condition. In addition, the sensitivity of both models to the probabilistic ap forecasts is comparable. The mean positions for TIE-GCM ROBE are similar in all three uncertainty cases for both epochs.

Table 8. Distribution statistics for elevated solar / low geomagnetic activity (Figure 9).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	0.8168	0.9127	1.2659	2.7447
	Model	-0.1543	0.6081	-0.6066	1.6949
	Combined	0.6489	1.1155	0.6695	3.3309
TIE-GCM ROBE	Driver	8.0766	1.0719	31.1664	4.4166
	Model	8.8591	3.6687	35.3401	13.9023
	Combined	8.8474	3.7708	34.6877	14.2979

Table 8 confirms the similarities in TIE-GCM ROBE's mean positions, and it exposes a unique characteristic in its standard deviations. The combined standard deviation is slightly larger than that of the model uncertainty case at the three and six day marks. This is the first condition where that holds for this particular model. This is also the condition with the largest discrepancy in the standard deviation for HASDM-ML between the driver and combined uncertainty cases.

The last condition to be analyzed is high solar activity and moderate geomagnetic activity. Looking at Figure 2, 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 the 190 sfu threshold, and that causes ap to be sampled from the distribution based on high solar and moderate geomagnetic activity. It is performed this way for consistency. These results are shown in both Figure 10 and 9.

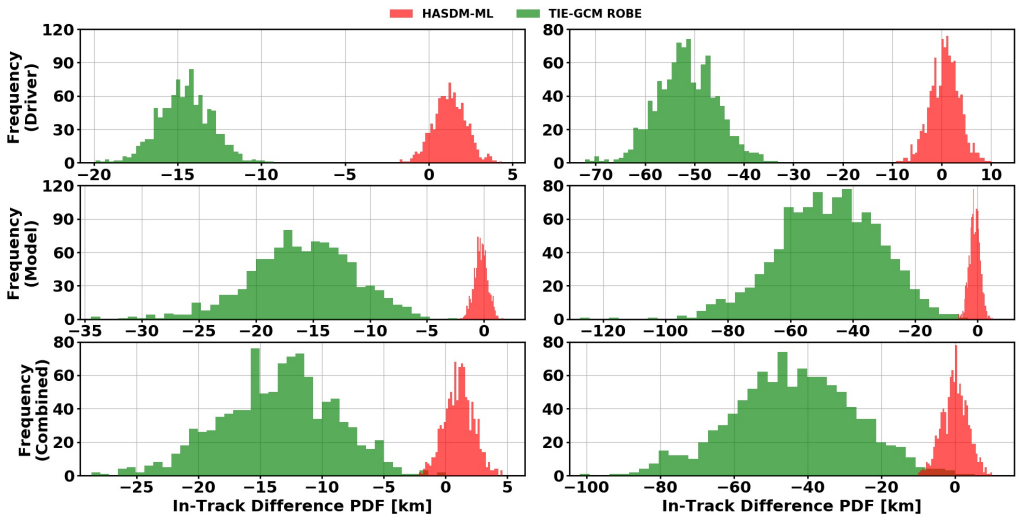

Figure 10. In-track position error distributions after three days (left) and six days (right). This is for high solar / moderate geomagnetic activity with ap being sampled.

This condition results in similar distributions using probabilistic ap forecasts, but there is a prominent bias for TIE-GCM ROBE signifying density overprediction. When comparing this distribution's spread to the highest two levels of solar activity (having similar inputs), we are seeing less sensitivity when sampling ap. There is a similar bias for TIE-GCM ROBE when considering either uncertainty source for this condition, and the values are lower than the higher two solar conditions.

Table 9. Distribution statistics for high solar / moderate geomagnetic activity (Figure 10).

		3 Days		6 Days	
	Uncertainty	**μ (km)**	**σ (km)**	**μ (km)**	**σ (km)**
HASDM-ML	Driver	1.4088	0.9696	1.1345	2.9469
	Model	-0.1615	0.5855	-0.6220	1.6739
	Combined	1.2652	1.1656	0.5210	3.4577
TIE-GCM ROBE	Driver	-14.2832	1.5868	-50.6131	6.0373
	Model	-15.1977	4.5907	-44.8212	16.4871
	Combined	-12.8815	4.6060	-41.3057	16.6883

Finally, Table 9 supports these observations. The standard deviation for driver uncertainty using TIE-GCM ROBE is notably lower than for elevated and high solar activity. This observation is also true for the mean. Despite that, the standard deviations for model uncertainty in this condition are higher than the two being compared for the same model.

CONCLUSIONS

These two density models, HASDM-ML and TIE-GCM ROBE, have evident underlying differences. First and foremost, TIE-GCM ROBE (as a recurrent neural network) is dynamic in nature. It is concerned with the evolution of the system with respect to time and the SW inputs. It was developed this way intentionally, being a derivative of TIE-GCM. HASDM-ML, on the other hand, is based on HASDM data, and HASDM is a static model. Therefore, HASDM-ML is a regression model and solely relates the SW inputs at the current epoch to a global density grid. Both surrogate models provide predictions with acceptable error, relative to their original modes, and can be evaluated in little time on standard desktop computers.

There are stark differences between the performance of these two models with regard to each uncertainty source. HASDM-ML has relatively low variance when looking at driver uncertainty, model uncertainty, or the combined effect of the two. An interesting observation from the HASDM-ML statistics is that the combined mean and variance could be approximated from the two individual uncertainty statistics. The combined mean was roughly the sum of the two mean positions, and the standard deviation was always slightly larger than that of the driver uncertainty. The standard deviation for driver uncertainty was greater than the model uncertainty for all seven conditions, and by six days, it was nearly an order of magnitude larger.

The TIE-GCM ROBE statistics were less consistent and diverged from the HASDM-ML observations. There was not an additive property in the mean positions for any of the conditions. Rather, the combined mean position did not show any consistency in how it compared to the two individual trends. It was between the two values in 9/14 cases (7 conditions and two epochs), but it was either more positive or negative than either single source in remaining five cases.

In nearly every scenario, the effect of model uncertainty on TIE-GCM ROBE was greater than driver uncertainty. The lone deviation from this trend came at the six-day mark of the elevated

solar activity condition. The standard deviation for model uncertainty was only 83% of the value for driver uncertainty. Another unique result came from the moderate solar activity condition. The TIE-GCM ROBE six-day standard deviation was smaller for the combined uncertainties than both of the single uncertainty cases. This is an unlikely but possible result of combining probabilistic inputs and a probabilistic model.

The results of this study highlight the importance of understanding the levels of uncertainty that come from various sources in density prediction and therefore orbit determination. These vary between models and also are a function of Space Weather. Solar and geomagnetic activity impact the level of uncertainty in driver forecasts, but they can also impact the confidence in a model's prediction. Neither driver or model uncertainty have shown to be definitively more important than the other upon studying these results. Conversely, the conclusion is that we need to further investigate uncertainty on a model-to-model basis.

FUTURE WORK

We plan to leverage data assimilation to de-bias TIE-GCM ROBE. In addition, both models need to be optimized, with the help of tools like AutoKeras or Keras Tuner,[16,17] in order to better grasp their sensitivity to inputs and their prediction capabilities. We also intend to investigate the impact of training samples on the model uncertainty bounds for both HASDM-ML and TIE-GCM ROBE. Going forward, we will also investigate custom cost functions in training the deep learning models for consistency and reliability of uncertainties.[9]

ACKNOWLEDGMENT

This research was made possible by NASA West Virginia Space Grant Consortium, Training Grant #NNX15AI01H. SET and WVU gratefully acknowledge support from the NASA SBIR contract #80NSSC20C0292 for Machine learning Enabled Thermosphere Advanced by HASDM (META-HASDM).

REFERENCES

[1] 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.

[2] A. Ridley, Y. Deng, and G. Tóth, "The global ionosphere–thermosphere model," *Journal of Atmospheric and Solar-Terrestrial Physics*, Vol. 68, No. 8, 2006, pp. 839 – 864, https://doi.org/10.1016/j.jastp.2006.01.008.

[3] B. Bowman, W. K. Tobiska, F. Marcos, C. Huang, C. Lin, and W. Burke, "A New Empirical Thermospheric Density Model JB2008 Using New Solar and Geomagnetic Indices," 2012, 10.2514/6.2008-6438.

[4] 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.

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

[6] R. Licata, P. Mehta, and W. K. Tobiska, "Impact of Space Weather Driver Forecast Uncertainty on Drag and Orbit Prediction," 08 2020.

[7] R. J. Licata, W. K. Tobiska, and P. M. Mehta, "Benchmarking Forecasting Models for Space Weather Drivers," *Space Weather*, Vol. 18, No. 10, 2020, https://doi.org/10.1029/2020SW002496.

[8] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," 2016.

[9] G. J. Anderson, J. A. Gaffney, B. K. Spears, P.-T. Bremer, R. Anirudh, and J. J. Thiagarajan, "Meaningful uncertainties from deep neural network surrogates of large-scale numerical simulations," 2020.

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, Vol. 15, 06 2014, pp. 1929–1958.

[11] R. Alake, "Understanding and Implementing Dropout in TensorFlow and Keras," https://towardsdatascience.com/understanding-and-implementing-dropout-in-tensorflow-and-keras-a8a3a02c1bfa, 05 2020.

[12] R. Licata and P. Mehta, "Physics-informed Bayesian Deep Learning for Space Weather Science and Operations," 01 2021, 10.13140/RG.2.2.13248.76807.

[13] W. K. Tobiska, B. R. Bowman, D. Bouwer, A. Cruz, K. Wahl, M. D. Pilinski, P. M. Mehta, and R. J. Licata, "The SET HASDM density database," *Earth and Space Science Open Archive*, 2020, p. 5, 10.1002/essoar.10504902.1.

[14] R. Licata, P. Mehta, and W. K. Tobiska, "Data-Driven HASDM Density Model using Machine Learning," 12 2020, 10.1002/essoar.10505213.1.

[15] P. M. Mehta, R. Linares, and E. K. Sutton, "A Quasi-Physical Dynamic Reduced Order Model for Thermospheric Mass Density via Hermitian Space-Dynamic Mode Decomposition," *Space Weather*, Vol. 16, No. 5, 2018, pp. 569–588, 10.1029/2018SW001840.

[16] H. Jin, Q. Song, and X. Hu, "Auto-Keras: An Efficient Neural Architecture Search System," 2019.

[17] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi, *et al.*, "Keras Tuner," https://github.com/keras-team/keras-tuner, 2019.