Space Weather

RESEARCH ARTICLE
10.1029/2021SW002918

Key Points:
- We develop a nonlinear global model for exospheric temperature prediction called EXospheric TEMPeratures on a PoLyhedrAL gRid Machine Learned (EXTEMPLAR-ML)
- We leverage principal component analysis to improve our understanding of the EXTEMPLAR-ML temperature formulation
- EXTEMPLAR-ML shows increased accuracy relative to satellite observations across a majority of conditions, locations, and during geomagnetic storms

Supporting Information:
Supporting Information may be found in the online version of this article.

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

Citation:
Licata, R. J., Mehta, P. M., Weimer, D. R., & Tobiska, W. K. (2021). Improved Neutral Density Predictions Through Machine Learning Enabled Exospheric Temperature Model. Space Weather, 19, e2021SW002918. https://doi.org/10.1029/2021SW002918

Received 15 SEP 2021
Accepted 3 DEC 2021

Abstract
The community has leveraged satellite accelerometer data sets in previous years to estimate neutral mass density and exospheric temperatures. We utilize derived temperature data and optimize a nonlinear machine-learned (ML) regression model to improve upon the performance of the linear EXospheric TEMPeratures on a PoLyhedrAL gRid (EXTEMPLAR) model. The newly developed EXTEMPLAR-ML model allows for exospheric temperature predictions at any location with one model and provides performance improvements over its predecessor. We achieve reductions in mean absolute error of 2 K on an independent test set while providing similar error standard deviation values. Comparing the performance of both EXTEMPLAR models and the Naval Research Laboratory Mass Spectrometer and Incoherent Scatter radar Extended model (NRLMSISE-00) across different solar and geomagnetic activity levels shows that EXTEMPLAR-ML has the lowest mean absolute error across 80% of conditions tested. A study for spatial errors demonstrated that at all grid locations, EXEMPLAR-ML has the lowest mean absolute error for over 60% of the polyhedral grid cells on the test set. Like EXEMPLAR, our model's outputs can be utilized by NRLMSISE-00 (exclusively) to more closely match satellite accelerometer-derived densities. We conducted 10 case studies where we compare the accelerometer-derived temperature and density estimates from four satellites to NRLMSISE-00, EXEMPLAR, and EXEMPLAR-ML during major storm periods. These comparisons show that EXEMPLAR-ML generally has the best performance of the three models during storms. We use principal component analysis on EXEMPLAR-ML outputs to verify the physical response of the model to its drivers.

Plain Language Summary
Density in the upper atmosphere is highly variable and difficult to model. Empirical density models often rely on temperature profile predictions to determine species and mass densities. One of three key parameters in determining the temperature profiles is the asymptotic value at the top of the thermosphere called the exospheric temperature. By using temperatures derived from satellite acceleration measurements, we develop a machine-learned global temperature model called EXospheric TEMPeratures on a PoLyhedrAL gRid Machine Learned (EXTEMPLAR-ML). We achieve a 2 K reduction in mean absolute error on the independent test set relative to the model's predecessor. Additional analyses showed that EXEMPLAR-ML was more accurate than linear EXEMPLAR across a majority of conditions and grid locations. We also look at temperatures and densities along satellite orbits during 10 major geomagnetic storms from the 21st century. In this study, we see major improvements over a significant empirical model called NRLMSISE-00 and the linear predecessor to EXEMPLAR-ML-ML. We leveraged a mathematical decomposition tool on the model outputs to assess its internal formulation. This shows that EXEMPLAR-ML is most heavily driven by solar activity and the seasons.

1. Introduction
Thermospheric mass density modeling is vital to satellite tracking and orbit prediction, yet it remains a formidable task for researchers and operators. The thermosphere is highly driven by external forcing, such as space weather events, and complex internal dynamics. The primary external driver to the thermosphere is solar irradiance (Qian & Solomon, 2011). A majority of the solar irradiance energy input to the thermosphere can be captured with various solar indices and proxies (Bowman & Tobiska, 2006). However, these model drivers become less effective during solar minimum where other processes (e.g., composition changes) become increasingly relevant (Bowman et al., 2008; Mehta et al., 2019). Other space weather events, such as coronal mass ejections, can send mass and energy toward Earth. This mass and energy interact with the magnetosphere resulting in Joule heating and particle precipitation, which can cause large, sudden changes in density (Deng et al., 2013; Fedrizzi et al., 2012).
Nitric oxide (NO) is a cooling mechanism responsible for long-term cooling trends present during solar minimum (Kockarts, 1980) and short-term temperature decreases following large geomagnetic storms (Knipp et al., 2017; Mlynczak et al., 2003). Lei, Burns, et al. (2012) found that for the 2003 Halloween storm, temperature and density post-storm were appreciably lower than pre-storm levels although the specific cause of this is highly contested (Lei et al., 2021; Mikhailov & Perrone, 2020). Many empirical models do not model this phenomena well and predict higher density in the recovery phase of major storms relative to observations (Licata, Mehta, Tobiska, Bowman, & Pilinski, 2021; Oliveira & Zesta, 2019).

The Naval Research Laboratory Mass Spectrometer and Incoherent Scatter radar Extended model (NRLMSISE-00 but referred to in this paper as MSIS) is a commonly used empirical thermospheric density model (Picone et al., 2002). As with many models (e.g., DTM, Bruinsma, 2015 and JB2008, Bowman et al., 2008), MSIS heavily relies on temperature profiles to determine species densities and therefore mass density throughout the thermosphere. A key parameter in predicting the temperature profile is the exospheric temperature ($T_x$), which is the asymptotic value that the temperature profile approaches at the top of the thermosphere or thermopause (Bates, 1959; Jacchia, 1965). MSIS uses the Bates-Walker temperature profile (Walker, 1965).

The availability of accelerometer-derived density estimates from satellites—such as CHAllenging Minisatellite Payload (CHAMP), Gravity Recovery and Climate Experiment (GRACE), and Swarm—has been advantageous for model development and assessment (Bettadpur, 2012; Friis-Christensen et al., 2006; Lühr et al., 2002). Over the lifetime of satellites with onboard accelerometers, we accumulate measurements over an abundance of locations and space weather conditions. Researchers have used these measurements to derive density estimates by removing accelerations from other sources (Bruinsma & Biancale, 2003; Calabia & Jin, 2016; Doornbos, 2012; Liu et al., 2005; Mehta et al., 2017; Sutton, 2008). Weimer et al. (2016) used the density estimates from Mehta et al. (2017) to approximate exospheric temperatures by varying the temperature parameter in MSIS using the bisection method until the model density closely matched that of the satellite. Weng et al. (2017) followed this methodology and used Sutton’s CHAMP density estimates to create an exospheric temperature model.

Weimer et al. (2020) had used the derived exospheric temperatures to fit 1,620 linear models to make predictions on a polyhedral grid as a function of different space weather conditions over time. The model is called EXospheric TEMPeratures on a PoLyhedrAl gRid (EXEMPLAR). In this work, we develop an improved exospheric temperature model by using a single nonlinear artificial neural network (ANN) to make predictions at any location. This global model is called EXEMPLAR Machine Learned (EXEMPLAR-ML).

Principal component analysis (PCA), also referred to as Empirical Orthogonal Function (EOF) analysis or Proper Orthogonal Decomposition (POD), is used in this work to investigate the most dominant modes of variability in EXEMPLAR-ML to ensure that the temperature formulation is as expected. PCA has been used to analyze thermospheric density data sets previously and is often used in the development of reduced-order models (Gondelach & Linares, 2020; Mehta & Linares, 2017; Mehta et al., 2018). PCA has also been used to study satellite accelerometer data sets (Calabia & Jin, 2016; Lei, Matsuo, et al., 2012; Matsu & Forbes, 2010). Sutton et al. (2012) used PCA to produce basis functions that represented the variability of temperature parameters used in an empirical Jacchia family model to improve its nominal density formulation (Jacchia, 1970). Ruan et al. (2018) used CHAMP density estimates and a physics-based density model to develop an exospheric temperature model based in PCA. Machine learning (ML) models tend to be ambiguous in nature, so we utilize PCA on EXEMPLAR-ML outputs only to see if the model’s response to the inputs is physical without physical constraints being present in training.

In this work, we show that the extension of ML to develop an exospheric temperature model has the following key benefits over previous models:

- The ML model improves prediction accuracy over its linear predecessor in many conditions.
- While EXEMPLAR is composed of 1,620 individual regression models for specific locations, EXEMPLAR-ML is a single model capable of predicting exospheric temperature and any local time-latitude combination. That is particularly important when predicting temperature and density along an orbit at a high frequency.
- The most significant benefit is the nonlinear formulation of EXEMPLAR-ML. A linear model cannot capture nonlinear processes that are essential to represent during geomagnetic storms. We show in Sections 3.2
and 3.3 that EXEMPLAR-ML more closely matches satellite estimates for temperature and density during these periods.

The paper is organized as follows: we start by detailing the data and model development. Then, we discuss the methodology for temperature and density prediction using the model. We then compare errors between MSIS, EXEMPLAR, and EXEMPLAR-ML across different space weather conditions and at different locations. We investigate the dominant modes and PCA coefficients across an eight-year period. As a case study, we compare the temperature and density predictions of MSIS, EXEMPLAR, and EXEMPLAR-ML to CHAMP, GRACE-A, Swarm A, and Swarm B during 10 major geomagnetic storms.

2. Data and Methods

2.1. Data

Typically, MSIS predicts the exospheric temperature as a function of position and space weather drivers. Using this computed value, the model then calculates species densities as a function of altitude and therefore neutral mass density. The user can override the internally computed $T_\infty$ and MSIS will determine density based on the provided value. In the past, this has been leveraged in numerical schemes to match MSIS to satellite measurements in order to estimate $T_\infty$ (Forbes et al., 2009, 2011; Weimer et al., 2016). Weimer et al. (2020) used the Mehta et al. (2017) CHAMP and GRACE density estimates from 2002 to 2010 to perform a similar derivation of exospheric temperatures. They had also used Swarm A and Swarm B density estimates from Astafyeva et al. (2017). These two satellites, such as CHAMP and GRACE, had orbits with high inclination and were at 470 and 520 km in altitude, respectively (Friis-Christensen et al., 2006). Weimer et al. (2020) had varied $T_\infty$ in MSIS using a binary search method until the temperature was determined to 2 K. They also generated a geodesic polyhedral grid made of 1,620 cells. These cells do not have equal areas, but they were designed to all be within 10.6% of the mean area. The exospheric temperatures are binned to cells based on their geographic coordinates. The distribution of samples on the grid is shown in Figure 1. The number of samples in each bin ranges from 25,998 to 505,180. The distribution heavily favors high latitudes as the satellites have high inclinations.

The $F_{10.7}, S_{10.7}, M_{10},$ and $Y_{10}$ indices are representative of solar activity and are part of Space Environment Technologies' (SET) SOLAR2000 algorithm (Tobiska et al., 2000), which has been recently benchmarked by Licata, Tobiska, and Mehta (2020). $F_{10.7}$ is a legacy solar proxy measured since 1947 (Covington, 1948). It is measured by the Dominion Radio Astrophysical Observatory in Canada. The $S_{10.7}$ index is representative of the integrated 26–34 nm solar EUV emission measured by the Solar and Heliospheric Observatory research satellite's Solar Extreme-ultraviolet Monitor. $M_{10}$ is a proxy for the Mg II core-to-wing ratio and is measured by the Solar Backscatter Ultraviolet spectrometer on NOAA operational satellites. $Y_{10}$ is a hybrid index of 0.1–0.8 nm solar X-rays during solar maximum and Lyman-α emissions during solar minimum. Descriptions of these indices are thoroughly explained by Tobiska et al. (2008). Poynting flux values represent the electrodynamic energy that flows into the atmosphere and are calculated using an electrodynamics model (referred to as the W05 model) described by Weimer (2005a, 2005b), which requires IMF and solar wind velocity data from the Advanced Composition Explorer spacecraft.

2.2. Model Development

We had access to over 81 million exospheric temperature estimates from Weimer et al. (2020), the associated polyhedral grid locations, and different space weather indices/proxies as potential drivers. The temperature estimates spanned four satellite missions and approximately 16 years. During 2002–2010, there were abundant measurements with CHAMP and GRACE having a 10- and five-second cadence, respectively. From 2013 to 2018, there are fewer measurements as Swarm A and Swarm B only have measurements every 30 s. Furthermore, Weimer et al. (2020) only used Swarm B through June 2016. Figure 2 shows the training, validation, and test splits used for model development. The validation set uses one geomagnetically active year from the Swarm mission and part of solar minimum from CHAMP and GRACE to test the model’s capabilities on diverse conditions. The test set was chosen to be from 30 May 2005 to 2006. This period contains many storms and solar medium conditions. The training-validation-test split represents 74.50%-7.25%-18.25% of the total data.
The best linear model from previous work used $S_{10}$, $A\sqrt{M_{10}}$, Poynting flux totals ($S_N$ and $S_S$), a temperature perturbation term ($\Delta T$), day of year (doy), and universal time (UT). The cooling effect of Nitric Oxide emissions was simulated in the calculation of $\Delta T$. $S_N$ and $S_S$ represent the Poynting flux in the northern and southern hemispheres, respectively. Even though the JB2008 thermospheric density model (Bowman et al., 2008) uses $F_{10}$, $S_{10}$, $M_{10}$, and $Y_{10}$, Weimer et al. (2020) found that $S_{10}$ and $M_{10}$ had the highest correlation to temperature ($r = 0.97$). We found that the using all four JB2008 solar indices improved model accuracy. Therefore, we used $F_{10}$, $S_{10}$, $M_{10}$, $Y_{10}$, $S_N$, $S_S$, $\Delta T$, local solar time (LST), geodetic latitude, doy, and UT. For doy and UT, the model uses sine and cosine functions of the fractional doy and UT, generating four temporal inputs ($t_1$-$t_4$), see Equation 1. In an effort to prevent any local time discontinuities, transformations were also made to the local solar time input ($LST_1$ and $LST_2$), see Equation 2.

$$t_1 = \sin \left( \frac{2\pi \text{doy}}{365.25} \right) \quad t_2 = \cos \left( \frac{2\pi \text{doy}}{365.25} \right) \quad t_3 = \sin \left( \frac{2\pi \text{UT}}{24} \right) \quad t_4 = \cos \left( \frac{2\pi \text{UT}}{24} \right)$$

$$LST_1 = \sin \left( \frac{2\pi \text{LST}}{24} \right) \quad LST_2 = \cos \left( \frac{2\pi \text{LST}}{24} \right)$$

Upon having set up the data into inputs (described above) and labels (associated $\log_{10} T_\infty$), we leverage a tool called Keras Tuner (O’Malley et al., 2019). This allows us to provide a range of hyperparameters upon which the tuner searches to find the best architecture/model through a Bayesian optimization scheme to minimize the validation loss. The tuner settings are shown in Table 1. The tuner is provided 2 million random training samples.
and 500,000 validation samples. These 2 million and 500,000 samples are randomly selected from the complete training and validation sets, respectively.

The tuner settings and the necessary number of training and validation samples were determined through preliminary testing. With 150 architectures tested for a tuner with 2 million training and 500,000 validation samples, the tuner takes approximately 14 and a half hours to run being parallelized across four NVIDIA Quadro 5000 GPUs. Once complete, the tuner returns the 10 best models, which we evaluate on independent data to confirm model performance. The best architecture to come out of the tuner had 148,057 trainable parameters and is displayed in Table 2. This model was trained further using the full training set to 220 epochs where the validation loss was at its minimum value. While EXEMPLAR-ML is not restricted to prediction at the polyhedral grid locations, it

![Figure 2](image_url)

**Figure 2.** $S_{\|}$, $S_{\perp}$, and exospheric temperature across the training, validation, and test sets. Arrows show the span of the measurements used from each satellite.

| Parameter                      | Choices                                      |
|--------------------------------|----------------------------------------------|
| Number of Hidden Layers        | 1–10                                         |
| Neurons                        | min = 64, max = 1024, step = 4               |
| Activations                    | relu, sofplus, tanh, sigmoid, sofsign, selu, elu, linear |
| Dropout                        | min = 0.10, max = 0.60, step = 0.01         |
| Optimizer                      | RMSprop, Adam, Adadelta, Nadam              |
still contains the name EXTEMPLAR, because it is trained on temperatures that are binned to those locations.

EXTEMPLAR-ML is compared to EXTEMPLAR and MSIS across multiple space weather conditions and as a function of location. First, we time-match the valid $T_\infty$ predictions from each model between 2002 and 2018. The mean absolute error and error standard deviation are computed for each model across all samples. This is not the same methodology used by Weimer et al. (2020) to compute the model statistics, which is why they differ. Next, we look at different $F_{10}$ and $a_p$ ranges and compute the mean absolute error for each model within these space weather conditions across the test set. Test errors for each model are segmented into each EXTEMPLAR grid location to compare the mean absolute error and error standard deviation visually. This will show how the models’ performance compares in different regions (e.g., high latitudes, dayside, and nightside).

### 2.3. Model Interpretation With PCA

ML models can be difficult to interpret, which makes them ambiguous in nature. For EXTEMPLAR-ML, we want to validate that the model is driven by the same dominant processes one would expect in MSIS and linear EXTEMPLAR. To investigate this, we leverage PCA as an indirect interpretation technique. PCA is an eigen-decomposition technique that determines uncorrelated linear combinations of the data that maximize variance (Hotelling, 1933; Pearson, 1901). As mentioned in the Introduction, PCA is widely used in thermospheric density and exospheric temperature studies as both a modeling and analytical tool. We use PCA to get insight into EXTEMPLAR-ML, which requires predictions covering a vast array of conditions. To accomplish this, we evaluated the model at all 1,620 EXTEMPLAR grid locations from 2002 to 2009 at a three-hour cadence. These predictions provide the global evolution of exospheric temperatures spanning over half a solar cycle. We perform PCA on the spatiotemporal temperature maps to obtain the $U$, $\Sigma$, and $V$ matrices. PCA decomposes the data and separates spatial and temporal variations such that:

$$\mathbf{x}(s, t) = \sum_{i=1}^r \alpha_i(t) U_i(s)$$  \hspace{1cm} (3)

where $\mathbf{x} \in \mathbb{R}^r$ is the model output state (full 2D temperature maps), $r$ is the choice of order truncation, $\alpha_i$ are temporal coefficients, and $U_i$ are orthogonal modes or basis functions. The modes are the first $r$ columns of the left singular vector derived by performing PCA on an ensemble of model output solutions such that:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \ldots \mathbf{x}_m \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^T$$  \hspace{1cm} (4)

In Equation 4, $m$ represents the ensemble size (eight years). The temperature data is denoted by $\mathbf{X}$. $U$ is the left unitary matrix, and it is made of orthogonal vectors that represent the modes of variation. $\Sigma$ is a diagonal matrix consisting of the squares of the eigenvalues that correspond to the vectors in $U$. We can extract temporal coefficients by performing matrix multiplication between $\Sigma$ and $V^T$. Therefore, the signs of the modes and coefficients are important in the analysis phase.

### 2.4. Geomagnetic Storm Case Study

To assess the performance of EXTEMPLAR-ML during extreme events, we compare MSIS, EXTEMPLAR, and EXTEMPLAR-ML to CHAMP- and GRACE-A-derived temperature and density estimates during 10 storm periods. The storms are listed in Table 3 along with their statistics/attributes. Five of the storms are used by Bruinsma et al. (2021) to assess the capabilities of thermospheric density models. Three of the storms in Table 3 with available CHAMP and GRACE data are from 2005 and were classified by Knipp et al. (2013) as problem

| Neurons | Activation | Dropout rate |
|---------|------------|--------------|
| Layer 1 | 864        | sigmoid      | 0.18         |
| Layer 2 | 156        | relu         | 0.32         |
| Output  | 1          | linear       | 0.00         |

*Note. There are 14 inputs for Layer 1.*
Table 3
List of 10 Geomagnetic Storms Used to Compare MSIS, EXTEMPLAR v6, and EXTEMPLAR-ML.

| Start date       | Min $D_{st}$ [nT] | Max $a_r$ [2nT] | Storm type | Set | Satellites |
|------------------|-------------------|-----------------|------------|-----|------------|
| 29 October 2003  | −400              | 400             | CME        | Training | CH GR      |
| 19 November 2003 | −422              | 300             | CME        | Training | CH GR      |
| 11 February 2004 | −109              | 94              | HSS        | Training | CH GR      |
| 26 July 2004*    | −197              | 300             | CME        | Training | CH GR      |
| 14 May 2005      | −263              | 236             | CME        | Training | CH GR      |
| 23 August 2005   | −216              | 300             | CME        | Test    | CH GR      |
| 10 September 2005| −147              | 179             | CME        | Test    | CH GR      |
| 14 December 2006 | −146              | 236             | CME        | Test    | CH GR      |
| 16 March 2015    | −223              | 240             | CME        | Validation | SA SB |
| 21 June 2015     | −204              | 240             | CME        | Validation | SA SB |

Note: For storm type, CME = coronal mass ejection and HSS = high-speed stream. For satellites, CH = CHAMP, GR = GRACE, SA = Swarm A, and SB = Swarm B.

*Look only at a 48-hr period starting at 1200 UT on the start date.

The error distributions show that MSIS has a strong tendency to underpredict temperature on all three sets. Both EXEMPLAR models have similar characteristics for the training and validation sets but EXEMPLAR v6 has a more negative bias on the test set. A major takeaway from Figure 3 is that all three models struggle on the test set although it is only truly independent data for EXEMPLAR-ML. It is hypothesized that during this period, there is an issue with the data or the model drivers may not correspond to as much of the variation in the system as they do elsewhere. Table 4 shows the mean absolute error and standard deviation of the error for MSIS, EXEMPLAR v6, and EXEMPLAR-ML in Kelvin. The mean absolute error is also shown as a percentage. Table 5 shows the mean absolute percent error for these three models across different space weather conditions on the test data. These are shown as relative errors to remove the impact of the general temperature level for the conditions.

EXEMPLAR-ML has the lowest mean absolute error on all three sets, while having a lower error standard deviation on the training and validation sets. The difference between the two EXEMPLAR models in terms of the test error standard deviation is only 0.41 K. However, EXEMPLAR-ML has reduced the mean absolute error by 2 K on the test set. In Table 5, we see that both the linear and ML versions of EXEMPLAR are more accurate than MSIS, regardless of solar and geomagnetic activity level. Furthermore, EXEMPLAR-ML has the lowest mean absolute error for 16 of the 20 conditions investigated. The difference in mean absolute error between the two EXEMPLAR models never deviates by more than 0.6% though. We show the mean absolute error and error standard deviation for all models at each polyhedral grid cell in Figure 4. Note that only the test set is used for this comparison, and the number of samples in each cell ranges from 4,023 at the equator to 104,278 at the poles.

In Figure 4, the highest MSIS errors on the dayside are at nearly all latitudes while on the nightside, the highest errors favor high latitudes. There is a similar pattern for MSIS’ error standard deviation, though they are more

storms due to overproduction of nitric oxide that resulted in the inaccurate density predictions. The October 2003 storm was chosen as it was one of the most significant storms of the 21st century. The July 2004 storm was used for comparison between MSIS and EXEMPLAR by Weimer et al. (2020). The two storms in 2015 were chosen as they reside in the validation set and use Swarm measurements. We will compare the exospheric temperature predictions of each model to the satellite estimates in Section 3.2, and those values will be used in MSIS to predict density shown in Section 3.3.

3. Results

Once we obtained the model described in Section 2.2, we evaluated it on all training, validation, and test data. This is shown in Figure 3 alongside MSIS and the best linear EXEMPLAR model from Weimer et al. (2020) (EXTEMPLAR v6). The left panels show the original and predicted $T_e$ values in a scatterplot with a background contour, showing the absolute error relative to the satellite estimate. The right panels show the error distributions for training, validation, and test samples.

The $T_e$ scatterplots are fairly centered on the 1:1 line, which indicates a zero-error prediction. There is a skew toward underprediction at very high temperatures, particularly on the test set. However, some of these exospheric temperatures are not physical, due to some instances where an abnormally high temperature needs to be input to MSIS to obtain a match with the measured density. All models seem to have poor performance at temperatures above 1500 K on the test set. This subset of the data only makes up 0.03% of the test set, meaning the impact on overall performance is minor, and the scatter can be misleading. There is a similar artifact for the training set. Visually, it seems as if there is an abundance of high-temperature samples where the models perform poorly; however, there are only 778 samples above 2,500 K, making up only 0.001% of the training set. EXEMPLAR v6 seems to have a lower bias during these conditions but has much higher variance. EXEMPLAR-ML has the highest coefficient of determination on all three sets, though the EXEMPLAR v6 values are similar.

The error distributions show that MSIS has a strong tendency to underpredict temperature on all three sets. Both EXEMPLAR models have similar characteristics for the training and validation sets but EXEMPLAR v6 has a more negative bias on the test set. A major takeaway from Figure 3 is that all three models struggle on the test set although it is only truly independent data for EXEMPLAR-ML. It is hypothesized that during this period, there is either an issue with the data or the model drivers may not correspond to as much of the variation in the system as they do elsewhere. Table 4 shows the mean absolute error and standard deviation of the error for MSIS, EXEMPLAR v6, and EXEMPLAR-ML in Kelvin. The mean absolute error is also shown as a percentage. Table 5 shows the mean absolute percent error for these three models across different space weather conditions on the test data. These are shown as relative errors to remove the impact of the general temperature level for the conditions.

EXEMPLAR-ML has the lowest mean absolute error on all three sets, while having a lower error standard deviation on the training and validation sets. The difference between the two EXEMPLAR models in terms of the test error standard deviation is only 0.41 K. However, EXEMPLAR-ML has reduced the mean absolute error by 2 K on the test set. In Table 5, we see that both the linear and ML versions of EXEMPLAR are more accurate than MSIS, regardless of solar and geomagnetic activity level. Furthermore, EXEMPLAR-ML has the lowest mean absolute error for 16 of the 20 conditions investigated. The difference in mean absolute error between the two EXEMPLAR models never deviates by more than 0.6% though. We show the mean absolute error and error standard deviation for all models at each polyhedral grid cell in Figure 4. Note that only the test set is used for this comparison, and the number of samples in each cell ranges from 4,023 at the equator to 104,278 at the poles.

In Figure 4, the highest MSIS errors on the dayside are at nearly all latitudes while on the nightside, the highest errors favor high latitudes. There is a similar pattern for MSIS’ error standard deviation, though they are more
localized. The mean absolute error was lowest for EXEMPLAR v6 in 638/1,620 cells, and it was lowest for EXEMPLAR-ML in the remaining 982 cells. MSIS had the lowest error standard deviation in 24/1,620, EXEMPLAR v6 has the lowest error standard deviation in 1,190 cells, and EXEMPLAR-ML only had the lowest value for the remaining 406. Therefore, in the test set, EXEMPLAR-ML is generally more accurate as a function of location, but EXEMPLAR v6 is more precise.

3.1. Model Interpretation

With the extensive EXEMPLAR-ML prediction set described in Section 2.3, we can investigate the most dominant modes of variability and their associated PCA coefficients. This is shown in Figure 5. It is important to note...
that the data were not centered prior to performing PCA, so the first mode is
time representative of the mean temperature distribution over this period.
The first mode is representative of solar EUV heating denoted by the
diurnal temperature map and strong correlation with $S_{10}$. This mode accounts for over
97% of the system’s variance. Mode 2 represents a latitudinal Summer-Winter
variation. There is a linear progression of the mode with latitude and
$\alpha_2$ oscillated about zero with a period of 365 days. It has an inverse relationship
to $t_2$ described in Equation 1, and its amplitude is a function of the solar
activity. The second mode and coefficient are multiplied by $-1$ to align with
$t_2$. For Mode 3, the peaks are at high latitudes and the magnitude decreases
toward the equator.

$\alpha_3$ most strongly correlates with $\Delta T$ and the Poynting
flux totals ($\sim 0.60$ with $S_{10}$ and $S_{11}$). We suspect that this mode corresponds
to the effects of high-latitude heating from either Joule heating or electron
precipitation. These dominant modes are clear and would likely match that of
MSIS or EXTEMPLAR, confirming the model behavior. Again, this specific
analysis was only done on EXTEMPLAR-ML outputs for insights on the
model behavior. PCA was not used for model development.

### 3.2. Modeled Versus Observed Temperature Along Orbits

In an effort to compare model performance between other temperature models and the observations, we evaluate the MSIS (unmodified), EXEMPLAR v6, and EXEMPLAR-ML exospheric temperature values along satellite orbits for the 10 storms listed in Table 3. The mean absolute percent error relative to the satellite temperature estimates is shown in Table 6, and the predictions for the August 2005 storm are shown in Figure 6. Similar figures were generated for all 10 storms and
are provided in Supporting Information S1 in addition to a table of mean absolute error in Kelvin.

For all storms, both EXEMPLAR models perform better than MSIS. Concurrently, EXEMPLAR-ML has the
lowest error along 13 of the 20 satellite orbits out of these three models. EXEMPLAR-ML reduces $T_0$ prediction

![Image](image_url)

**Table 4**

| Model         | Mean absolute error | Standard deviation |
|---------------|---------------------|--------------------|
| Training set  |                     |                    |
| NRLMSISE-00   | 5.66%               | 47.10 K            |
| EXEMPLAR v6   | 3.62%               | 31.23 K            |
| EXEMPLAR-ML   | 3.16%               | 27.04 K            |
| Validation set|                     |                    |
| NRLMSISE-00   | 6.05%               | 47.76 K            |
| EXEMPLAR v6   | 3.85%               | 31.40 K            |
| EXEMPLAR-ML   | 3.70%               | 30.02 K            |
| Test set      |                     |                    |
| NRLMSISE-00   | 5.55%               | 43.37 K            |
| EXEMPLAR v6   | 3.74%               | 29.59 K            |
| EXEMPLAR-ML   | 3.47%               | 27.59 K            |

**Note.** Bold text represents the lowest error for the given condition.

**Table 5**

|                  | $F_{10} \leq 75$ | $75 < F_{10} \leq 90$ | $90 < F_{10} \leq 110$ | $F_{10} > 110$ | All $F_{10}$ |
|------------------|------------------|------------------------|------------------------|----------------|--------------|
| **NRLMSISE-00**  |                  |                        |                        |                |              |
| $a_p \leq 10$    | 5.64%            | 4.90%                  | 5.66%                  | 5.14%          | 5.23%        |
| $10 < a_p \leq 50$ | 7.13%            | 5.69%                  | 6.48%                  | 5.90%          | 6.28%        |
| $a_p > 50$       | 6.95%            | 7.16%                  | 9.26%                  | 8.04%          | 8.28%        |
| All $a_p$        | 6.00%            | 5.10%                  | 6.13%                  | 5.53%          | 5.55%        |
| **EXEMPLAR v6**  |                  |                        |                        |                |              |
| $a_p \leq 10$    | 3.44%            | 3.67%                  | 3.82%                  | 3.52%          | 3.62%        |
| $10 < a_p \leq 50$ | 3.83%            | 3.90%                  | 4.00%                  | 4.20%          | 3.92%        |
| $a_p > 50$       | 4.71%            | 6.23%                  | 6.52%                  | 5.20%          | 6.10%        |
| All $a_p$        | 3.55%            | 3.74%                  | 4.02%                  | 3.81%          | 3.74%        |
| **EXEMPLAR-ML**  |                  |                        |                        |                |              |
| $a_p \leq 10$    | 3.26%            | 3.20%                  | 3.56%                  | 3.29%          | 3.27%        |
| $10 < a_p \leq 50$ | 3.77%            | 3.80%                  | 4.01%                  | 4.27%          | 3.86%        |
| $a_p > 50$       | 4.55%            | 5.67%                  | 6.73%                  | 5.24%          | 6.03%        |
| All $a_p$        | 3.39%            | 3.36%                  | 3.88%                  | 3.68%          | 3.47%        |

**Note.** Bold text represents the lowest errors for that condition across the three models.
error up to 3% relative to EXTEMPLAR v6 and up to 9% relative to MSIS. Looking at the temperature predictions across these storms (Figure 6 and in Supporting Information S1), it is clear that EXTEMPLAR-ML most closely matches the general trends of the satellite estimates for a majority of each storm period. An important attribute of the model (as for EXTEMPLAR v6) is that it does not follow the outliers. In the data processing scheme, there are erroneous temperature predictions. While both EXTEMPLAR models are regressed on the satellite derived temperatures, it is evident that they pick up on physical trends and are not overfit to outliers. Other takeaways from the figures are that MSIS seems to follow the correct variation in $T_e$ but has a tendency to overpredict, and there are periods where EXTEMPLAR v6 deviates from the other temperature sources (e.g., other models and the satellite estimates). These trends are more evident in other storms found in Supporting Information S1, but this storm is shown as it is from the test set.
3.3. Modeled Versus Observed Density Along Orbits

We input the exospheric temperatures from Table 6 and Figure 6 into MSIS in order to obtain the associated mass density values along the satellite orbits. This is performed for the 10 storms, and the mean absolute percent errors are shown in Table 7. The predictions for the August 2005 storm are shown in Figure 7. The figures for the nine remaining storms and a table of mean absolute error in kg/m$^3$ are presented in Supporting Information S1.

The table shows improvements from both MSIS and EXEMPLAR v6 along 15 of the 20 orbits. For the March 2015 storm, EXEMPLAR v6 had the lowest errors along both Swarm orbits, but in terms of density, EXEMPLAR-ML had lower error. It is important to note that the difference between the two models is fairly negligible for some of the storms. However, the benefit of EXEMPLAR-ML is its global nature. We do not need to find the closest polyhedral grid cell and execute a specific model; EXEMPLAR-ML takes the location as an input. In Figure 7, all three models have similar predictions prior to the storm. During the storms, MSIS deviates

Figure 5. First three modes (left) and corresponding principal component analysis coefficients (right) for the exospheric temperatures between 1 January 2002 and 31 December 2009. The most highly correlated drivers are plotted against the coefficients for comparison with the Pearson correlation coefficient shown in the title (Schober et al., 2018).
Table 6
Mean Absolute Error of $T_e$ for MSIS, EXTEMPLAR v6 (Referred to as EXT v6), and EXTEMPLAR-ML (Referred to as EXT-ML) With Respect to CHAMP and GRACE-A or Swarm A and Swarm B Estimates for the Storms in Table 3

| Start date       | MSIS | EXT v6 | EXT-ML | MSIS | EXT v6 | EXT-ML |
|------------------|------|--------|--------|------|--------|--------|
| 29 October 2003  | 15.91% | 10.89% | 7.65%  | 13.03% | 9.98%  | 7.54%  |
| 19 November 2003 | 13.54% | 6.10%  | 4.96%  | 12.72% | 4.93%  | 4.03%  |
| 11 February 2004 | 5.05%  | 3.86%  | 3.92%  | 4.00%  | 3.43%  | 3.70%  |
| 26 July 2004     | 15.05% | 8.55%  | 5.71%  | 12.86% | 7.46%  | 5.42%  |
| 14 May 2005      | 10.79% | 6.89%  | 5.80%  | 9.54%  | 5.78%  | 5.42%  |
| 23 August 2005   | 9.65%  | 8.31%  | 8.06%  | 8.15%  | 6.39%  | 6.32%  |
| 10 September 2005| 8.61%  | 5.72%  | 6.42%  | 7.42%  | 5.10%  | 5.43%  |
| 14 December 2006 | 11.16% | 5.69%  | 5.98%  | 9.75%  | 4.69%  | 4.61%  |
| 16 March 2015    | 6.39%  | 3.65%  | 3.77%  | 5.79%  | 3.72%  | 3.77%  |
| 21 June 2015     | 11.95% | 5.46%  | 4.29%  | 11.89% | 5.71%  | 4.87%  |

*Note.* Bold text represents the lowest error across the three models for each satellite.

from the satellite estimates more significantly than the EXEMPLAR models. Through the entire period, EXEMPLAR v6 tends to result in higher density values than the other models.

4. Summary

In this work, we developed and optimized a global machine-learned nonlinear regression model to predict exospheric temperatures, given a set of space weather and temporal drivers. This model, called EXEMPLAR-ML, has fairly a similar performance across the training, validation, and test sets. This is an extension of a linear EXEMPLAR model developed by Weimer et al. (2020). EXEMPLAR v6 is a set of 1,620 linear regression models that predict exospheric temperature for their respective locations. While condensing that to a single model in EXEMPLAR-ML, we decreased the mean absolute error on all three sets relative to EXEMPLAR v6. While EXEMPLAR v6 had a lower error standard deviation on the test set, EXEMPLAR-ML’s value was within 0.5 K. It is difficult to comment on the generality of EXEMPLAR-ML, because Figure 3 and Table 5 show similar trends between the three models in terms of their relative performance on the three sets. We show that EXEMPLAR-ML had lower error in 16/20 solar and geomagnetic activity conditions investigated for the test set. When investigating error as a function of position within the test set, EXEMPLAR-ML had lower mean absolute errors for 61% of the polyhedral grid cells, while EXEMPLAR v6 had a lower error standard deviation for 73% of cells. Both of the EXEMPLAR models are developed exclusively for use with NRLMSISE-00.

The use of PCA on model outputs provided insight to the temperature formulation within the “black-box” ML model. The first mode represented the effects of solar EUV heating and accounted for 97.6% of the system's variance. Seasonal-latitudinal variations accounted for the next 0.76% of the variance and were still a function of solar activity. The last mode we looked at only accounted for 0.20% of the variance but described the effects of high-latitude heating caused by geomagnetic storms. These findings are consistent with what would be expected from performing the analysis with MSIS or EXEMPLAR v6. This analysis was only conducted on EXEMPLAR-ML outputs after training; PCA was not used for model development. The purpose was to show that even without physical constraints in the training process, the model's response to the inputs was in fact physical.

Table 7
Mean Absolute Error of Density for MSIS, MSIS + EXEMPLAR v6 (Referred to as EXT v6), and MSIS + EXEMPLAR-ML (Referred to as EXT-ML) With Respect to CHAMP and GRACE-A or Swarm A and Swarm B Estimates for the Storms in Table 3

| Start date       | MSIS | EXT v6 | EXT-ML | MSIS | EXT v6 | EXT-ML |
|------------------|------|--------|--------|------|--------|--------|
| 29 October 2003  | 41.62% | 24.44% | 17.49% | 43.75% | 26.69% | 19.91% |
| 19 November 2003 | 41.71% | 15.61% | 13.45% | 55.08% | 15.68% | 14.12% |
| 11 February 2004 | 14.51% | 10.94% | 10.99% | 15.66% | 13.73% | 14.55% |
| 26 July 2004     | 62.32% | 27.78% | 19.12% | 75.38% | 32.33% | 22.74% |
| 14 May 2005      | 29.30% | 16.69% | 14.02% | 42.05% | 23.16% | 18.68% |
| 23 August 2005   | 29.65% | 26.14% | 24.82% | 38.89% | 31.42% | 29.20% |
| 10 September 2005| 28.18% | 17.13% | 19.96% | 41.63% | 21.68% | 25.56% |
| 14 December 2006 | 36.54% | 15.59% | 16.12% | 57.79% | 21.50% | 20.12% |
| 16 March 2015    | 20.88% | 12.18% | 12.15% | 22.47% | 14.84% | 14.82% |
| 21 June 2015     | 64.25% | 22.34% | 17.83% | 82.49% | 29.52% | 25.08% |

*Note.* Bold text represents the lowest error across the three models for each satellite.
We performed a case study where EXTEMPLAR-ML, EXTEMPLAR v6, and MSIS predicted $T_\infty$ along CHAMP, GRACE-A, Swarm A, and Swarm B orbits during 10 major geomagnetic storms. These storms were diverse in their properties and spanned nearly 12 years. EXTEMPLAR-ML provided the most accurate predictions along 13/20 satellite orbits. When these temperatures were used for density prediction, the relative accuracy of EXTEMPLAR-ML became more pronounced. EXTEMPLAR-ML was the most accurate model for 15/20 orbits and had errors significantly lower than MSIS for nearly all storms. Figures were generated for all storms, but only one storm was displayed (both temperature and density). The remaining figures are provided in Supporting Information S1. These figures show the extent to which EXEMPLAR-ML follows the satellite temperature and density estimates while being robust to outliers that result from the processing of the accelerometer measurements. In the future, we plan to incorporate model uncertainty into EXEMPLAR-ML (Licata & Mehta, 2021; Licata, Mehta, & Tobiska, 2020, 2021). We plan to develop a newer model using temperatures derived with NRLMSIS 2.0 (Emmert et al., 2021). With the successive model, we also plan to use the exact observation locations to reduce errors associated with binning measurements to the polyhedral grid.

Figure 6. $T_\infty$ for MSIS, EXEMPLAR v6, and EXEMPLAR-ML plotted alongside CHAMP (left) and GRACE-A (right) for the geomagnetic storm in August 2005.
Data Availability Statement

CHAMP and GRACE density estimates from Mehta et al. (2017) can be found at http://tinyurl.com/densitysets. A data archive containing the supplemental graphs of neutral density predictions can be accessed online at https://doi.org/10.5281/zenodo.3525166. Also contained here are the adjustments to the NRLMSISE-00 model supplied by J. Emmert, the total Poynting flux into both Northern and Southern Hemispheres from the Weimer (2005a) model for years 2002–2017, the derived ΔT values, and EXEMPLAR model code with the required files.
Acknowledgments
This work was supported by NASA grant 80NSSC20K1362 to Virginia Tech under the Space Weather Operations 2 Research Program with subcontracts to WVU and SET. The authors would like to thank Douglas Drob for his insight into the MSIS model. The authors also appreciate the work of the anonymous reviewers for all of their time and effort in helping improve this manuscript.

References
Astafyeva, E., Zakharenkova, I., Haba, J. D., Doornbos, E., & van den Ijssel, J. (2017). Global ionospheric and thermospheric effects of the June 2015 geomagnetic disturbances: Multi-instrumental observations and modeling. Journal of Geophysical Research: Space Physics, 122(11), 11716–11742. https://doi.org/10.1002/2017JA024174
Bates, D. R. (1959). 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, 253(1275), 451–462. https://doi.org/10.1098/rspa.1959.0207
Bettsapir, S. (2012). Gravity recovery and climate experiment: Product specification document. GRACE 327-720, CSR-GR-03-02. Center for Space Research, The University of Texas.
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/AAA Astrodynamics Specialist Conference. AIAA 2008-6438. https://doi.org/10.2514/6.2008-6438
Bowman, B. R., & Tobiska, W. K. (2006). Improvements in modeling thermospheric densities using new EUV and FUV solar indices. In AIAA/ AAS Astrodynamics Specialist Conference. AAS 06-237.
Bruinsma, S. (2015). The DTM-2013 thermosphere model. Journal of Space Weather and Climate, 5, A1. https://doi.org/10.1051/ swsc/2015001
Bruinsma, S., & Biancale, R. (2003). Total densities derived from accelerometer data. Journal of Spacecraft and Rockets, 40(2), 230–236. https://doi.org/10.2514/2.3937
Bruinsma, S., Boniface, C., Sutton, E. K., & Fedrizzi, M. (2021). Thermosphere modeling capabilities assessment: Geomagnetic storms. Journal of Space Weather and Space Climate, 11, 12. https://doi.org/10.1051/swsc/2021002
Calabia, A., & Jin, S. (2016). New modes and mechanisms of thermospheric mass density variations from GRACE accelerometers. Journal of Geophysical Research: Space Physics, 121(11), 11,191–11,212. https://doi.org/10.1002/2016JA02594
Covington, A. E. (1948). Solar noise observations on 10.7 centimeters. Proceedings of the IRE, 36(4), 454–457. https://doi.org/10.1109/ JRPROC.1948.234598
Deng, Y., Fuller-Rowell, T. J., Ridley, A. J., Knipp, D., & Lopez, R. E. (2013). Theoretical study. Influence of different energy sources on the cusp neutral density enhancement. Journal of Geophysical Research: Space Physics, 118(5), 2340–2349. https://doi.org/10.1002/ jgra.50197
Doornbos, E. (2012). Producing density and crosswind data from satellite dynamics observations (pp. 91–126). Springer. https://doi. org/10.1007/978-3-642-25129-0_4
Enmert, J. T., Drob, D. P., Picone, J. M., Siskind, D. E., Jones, M., Jr., Mlynczak, M. G., et al. (2021). NRLMSIS 2.0: A whole-atmosphere empirical model of temperature and neutral species densities. Earth and Space Science, 8(3), e2020EA001321. https://doi.org/10.1029/2020EA001321
Fedrizzi, M., Fuller-Rowell, T. J., & Codrescu, M. V. (2012). Global Joule heating index derived from thermospheric density physics-based modeling and observations. Space Weather, 10(3), 530301. https://doi.org/10.1002/2011SW000724
Forbes, J. M., Bruinsma, S. L., Zhang, X., & Oberheide, J. (2009). Surface-escosphere coupling due to thermal tides. Geophysical Research Letters, 36(15), L15S12. https://doi.org/10.1029/2009GL038748
Forbes, J. M., Zhang, X., Bruinsma, S., & Oberheide, J. (2011). Sun-synchronous thermal tides in exosphere temperature from CHAMP and GRACE accelerometer measurements. Journal of Geophysical Research, 116(A11), A11309. https://doi.org/10.1029/2011ja016855
Fris-Cristensen, E., Lühr, H., & Halot, G. (2006). Swarm: A constellation to study the Earth's magnetic field. Earth, Planets and Space, 58, 351–358. https://doi.org/10.1186/BF035193
Gondelach, D. J., & Linares, R. (2020). Real-time thermospheric density estimation via two-line element data assimilation. Space Weather, 18(2), e2019SW002356. https://doi.org/10.1029/2019SW002356
Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441. https://doi.org/10.1037/h0071325
Jacchia, L. G. (1986). Static diffusion models of the upper atmosphere with empirical temperature profiles. Smithsonian Contributions to Astro- physics, 8, 215. https://doi.org/10.5479/si.00810231.8.9.213
Jacchia, L. G. (1970). New static models of the thermosphere and exosphere with empirical temperature profiles. SAO Special Report (p. 313).
Knipp, D., Kilcommons, L., Hunt, L., Mlynczak, M., Pilipenko, V., Bowman, B., et al. (2011). Thermospheric damping response to shear-enhanced geospace storms. Geophysical Research Letters, 40(7), 1263–1267. https://doi.org/10.1002/grl.50197
Knipp, D. J., Petree, D. V., Kilcommons, L. M., Isaacs, T. L., Cruz, A. A., Mlynczak, M. G., et al. (2017). Thermospheric nitric oxide response to shock-led storms. Space Weather, 15(2), 325–342. https://doi.org/10.1029/2016SW001567
Kockarts, G. (1980). Nitric oxide cooling in the terrestrial thermosphere. Geophysical Research Letters, 7(2), 137–140. https://doi.org/10.1029/ GL007i002p00137
Lee, J., Burns, A. G., Thayer, J. P., Wang, W., Mlynczak, M. G., Hunt, L. A., et al. (2012). Overcooling in the upper thermosphere during the recovery phase of the 2003 October storms. Journal of Geophysical Research, 117(A3), A03314. https://doi.org/10.1029/2011JA016994
Lee, J., Matsumo, T., Dou, X., Sutton, E., & Luan, X. (2012). Annual and semiannual variations of thermospheric density: EOF analysis of CHAMP and GRACE data. Journal of Geophysical Research, 117(A1), A01310. https://doi.org/10.1029/2011JA017324
Lee, J., Wang, W., Burns, A. G., Zhang, S.-R., & Dang, T. (2021). Comments on “Poststorm Thermospheric NO Overcooling?” by Mikhailov and Perrone (2020). Journal of Geophysical Research: Space Physics, 126(4), e2020JA027992. https://doi.org/10.1029/2020ja027992
Licata, R., & Mehta, P. (2021). Physics-informed Bayesian deep learning for space weather science and Operations. In Proceedings of the 18th Conference on Space Weather - 101st AMS Annual Meeting. https://doi.org/10.13140/RG.2.2.13248.76807
Licata, R., Mehta, P., & Tobiska, W. K. (2020). Data-driven HASDM density model using machine learning. In Proceedings of the AGU Fall Meeting 2020. https://doi.org/10.1029/EOSS.10505213.1
Licata, R., Mehta, P., & Tobiska, W. K. (2021). Impact of driver and model uncertainty on drag and orbit prediction. In Proceedings of the 31st AAS/AIAA Space Flight Mechanics Meeting.
Licata, R. J., Mehta, P. M., Tobiska, W. K., Bowman, B. R., & Pilinski, M. D. (2021). Qualitative and quantitative assessment of the SET HASDM database. Space Weather, 19, e2021SW002798. https://doi.org/10.1029/2021SW002798
Licata, R. J., Tobiska, W. K., & Mehta, P. M. (2020). Benchmarking forecasting models for space weather drivers. Space Weather, 18(10), e2020SW002496. https://doi.org/10.1029/2020SW002496
Liu, H., Lühr, H., Henize, V., & Köhler, W. (2005). Global distribution of the thermospheric total mass density derived from CHAMP. Journal of Geophysical Research, 110(A4), A04301. https://doi.org/10.1029/2004JA010741
Lühr, H., Gruenwald, L., & Forste, C. (2002). CHAMP reference systems, transformations and standards. Tech. rep. CH-GFZ-RE-002, GFZ-Potsdam.
Matsuo, T., & Forbes, J. M. (2010). Principal modes of thermospheric density variability: Empirical orthogonal function analysis of CHAMP 2001–2008 data. Journal of Geophysical Research, 115(A7), A07309. https://doi.org/10.1029/2009JA015109

Mehta, P. M., & Linares, R. (2017). A methodology for reduced order modeling and calibration of the upper atmosphere. Space Weather, 15(10), 1270–1287. https://doi.org/10.1002/2017sw001642

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(5), 569–588. https://doi.org/10.1002/2018SW001840

Mehta, P. M., Linares, R., & Sutton, E. K. (2019). Data-driven inference of thermosphere composition during solar minimum conditions. Space Weather, 17(9), 1364–1379. https://doi.org/10.1029/2019SW002264

Mehta, P. M., Walker, A. C., Sutton, E. K., & Godinez, H. C. (2017). New density estimates derived using accelerometers on board the CHAMP and GRACE satellites. Space Weather, 15(12), 1105–1116. https://doi.org/10.1002/2016SW001562

Mikhailov, A. V., & Perrone, L. (2020). Poststorm thermospheric NO overcooling? Journal of Geophysical Research: Space Physics, 125(11), e2019JA027122. https://doi.org/10.1029/2019JA027122

Mlynczak, M., Martin-Torres, F. J., Russell, J., Beaumont, K., Jacobson, S., Kozyra, J., et al. (2003). The natural thermostat of nitric oxide emission at 5.3 μm in the thermosphere observed during the solar storms of April 2002. Geophysical Research Letters, 30(21), 2100. https://doi.org/10.1029/2003GL017693

Oliveira, D. M., & Zesta, E. (2019). Satellite orbital drag during magnetic storms. Space Weather, 17(11), 1510–1533. https://doi.org/10.1029/2019SW002287

O’Malley, T., Barsztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Keras tuner. Retrieved from https://github.com/keras-team/keras-tuner

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559–572. https://doi.org/10.1080/14786410109462720

Picone, J. M., Hedin, A. E., Droh, D. P., & Akin, A. C. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. Journal of Geophysical Research, 107(A12), 4617. https://doi.org/10.1029/2002ja009430

Qian, L., & Solomon, S. C. (2011). Thermospheric density: An overview of temporal and spatial variations. Space Science Reviews, 168, 147–173. https://doi.org/10.1007/s11214-011-9810-z

Ruan, H., Lei, J., Dou, X., Liu, S., & Aa, E. (2018). An exospheric temperature model based on CHAMP observations and TIEGCM simulations. Space Weather, 16(2), 147–156. https://doi.org/10.1029/2017SW001759

Schober, P., Boer, C., & Schwarte, L. (2018). Correlation coefficients: Appropriate use and interpretation. Anesthesia & Analgesia, 126, 1763–1768. https://doi.org/10.1213/ane.0000000000002864

Sutton, E. K. (2008). Effects of solar disturbances on the thermosphere densities and winds from CHAMP and GRACE satellite accelerometer data. Ph.D. thesis. University of Colorado at Boulder.

Sutton, E. K., Cable, S. B., Lin, C. S., Qian, L., & Weimer, D. R. (2012). Thermospheric basis functions for improved dynamic calibration of semi-empirical models. Space Weather, 10(4), S10001. https://doi.org/10.1029/2012SW000827

Tobiska, W., 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

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

Walker, J. C. G. (1965). Analytic representation of upper atmosphere densities based on Jacchia’s static diffusion models. Journal of Atmospheric Sciences, 22(4), 462–463. https://doi.org/10.1175/1520-0469(1965)022<0462:ARUOUA>2.0.CO;2

Weimer, D. R. (2005a). Improved ionospheric electrodynamics models and application to calculating Joule heating rates. Journal of Geophysical Research, 110(A5), A05306. https://doi.org/10.1029/2004JA010884

Weimer, D. R. (2005b). Predicting surface geomagnetic variations using ionospheric electrodynamics models. Journal of Geophysical Research, 110(A12), A12S07. https://doi.org/10.1029/2005JA011270

Weimer, D. R., Mehta, P. M., Tobiska, W. K., Doornbos, E., Mlynczak, M. G., Droh, D. P., & Emmert, J. T. (2020). Improving neutral density predictions using exospheric temperatures calculated on a geodesic, polyhedral grid. Space Weather, 18(1), e2019SW002355. https://doi.org/10.1029/2019SW002355

Weimer, D. R., Sutton, E. K., Mlynczak, M. G., & Hunt, L. A. (2016). Inter calibration of neutral density measurements for mapping the thermosphere. Journal of Geophysical Research: Space Physics, 121(6), 5975–5990. https://doi.org/10.1002/2016JA022691

Weng, L., Lei, J., Sutton, E., Dou, X., & Fang, H. (2017). An exospheric temperature model from CHAMP thermospheric density. Space Weather, 15(2), 343–351. https://doi.org/10.1002/2016SW001577