



RESEARCH ARTICLE

10.1029/2021SW002892

Using Unsupervised and Supervised Machine Learning Methods to Correct Offset Anomalies in the GOES-16 Magnetometer Data

F. Inceoglu^{1,2,3}  and Paul T. M. Loto'aniu^{1,2}

¹Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, USA, ²National Centers for Environmental Information, National Oceanographic and Atmospheric Administration, Boulder, CO, USA, ³GFZ German Research Centre for Geosciences, Wissenschaftspark "Albert Einstein", Potsdam, Germany

Key Points:

- We show how machine learning can be used to correct contamination in magnetic field measurements taken from a geostationary satellite
- An unsupervised method was used to find four unique clusters in the time series data related to on-orbit thermal environment changes
- A supervised method was used to correct the clusters reducing the contamination from between 3–5 nT and 0–2 nT standard deviations

Correspondence to:

F. Inceoglu,
fadil.inceoglu@colorado.edu

Citation:

Inceoglu, F., & Loto'aniu, P. T. M. (2021). Using unsupervised and supervised machine learning methods to correct offset anomalies in the GOES-16 magnetometer data. *Space Weather*, 19, e2021SW002892. <https://doi.org/10.1029/2021SW002892>

Received 25 AUG 2021

Accepted 4 NOV 2021

Abstract This study uses supervised and unsupervised machine learning (ML) methods to correct unwanted offsets observed in the NOAA GOES-16 magnetometer data. All GOES satellites have an inboard and outboard magnetometer sensor mounted along a long boom. Post-launch testing of the GOES-16 magnetometers found that the inboard sensor suffers significant thermally induced magnetic contamination and currently only the outboard sensor is used in NOAA operations. The contamination varies both diurnally and seasonally making it very difficult to correct using basic statistical methods. For simplicity in explaining the offsets we are trying to correct, and methods used, we focus on correcting only one of the inboard vector components, the E-component (Earthward). We start by applying the unsupervised *k*-Shape method to the magnetic field vector E-component outboard minus inboard sensor time series, ΔE , resulting in four clusters that are closely related to the time of year and the solar β angle, which is a measure of the amount of time that a satellite is in direct sunlight. We then utilized LSTM networks as regressors to correct the offsets observed in GOES-16 inboard sensor E-component data. We trained our LSTMs using GOES-17 magnetometer data, which we show to exhibit much less variability compared with the GOES-16 data. The correction results reduced the offsets in the clusters from between 3–5 nT and 0–2 nT standard deviations. The combining of unsupervised and supervised ML methods is a powerful technique that can be applied to space-based instruments that produce time series data.

Plain Language Summary The Geostationary Operational Environmental Satellites (GOES) have two magnetometers, one is closer to the bus (inboard sensor) and another one mounted a few meters along the boom (outboard sensor). Post-launch testing of the GOES-16 magnetometers found that the inboard (IB) magnetometer suffers from contamination, which produces an offset in the IB data. Therefore, only the outboard (OB) sensor is used in NOAA operations. Having a good working second sensor on the satellite allows cross-sensor comparisons and validation both during post-launch testing and throughout the operational lifetime of the mission. We use unsupervised and supervised Machine Learning methods to correct the offset in the IB sensor data. We found that the offset is closely related to the time of year, more precisely the amount of time that the GOES satellite is in direct sunlight.

1. Introduction

The Geostationary Operational Environmental Satellite (GOES)-R series are the next generation weather satellites of the U.S. National Oceans and Atmospheric Administration (NOAA). Two of the four satellites in the series, GOES-16 and GOES-17, were launched in November 2016 and March 2018, respectively. The last two satellites are slated to be launched in February 16, 2022 and in April 2024. As with previous GOES, the GOES-R series supply NOAAs primary geostationary weather observations (NASA-GSFC, 2019). The GOES-R series have similar space weather payloads to previous GOES including magnetometers (MAGs) which measure the geomagnetic field and are the dataset of interest for this study. Products derived from GOES MAG observations are used by NOAA in forecasting, situational awareness, and development and validation of global space weather models (Singer et al., 1996; Loto'aniu et al., 2020). The science community also routinely uses this data for scientific research.

The GOES-16 and GOES-17 MAGs are identical, and a description of the instrument can be found in Loto'aniu et al. (2019). A drawing of the GOES-R series satellite is shown in Figure 1a. MAG consists of two

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

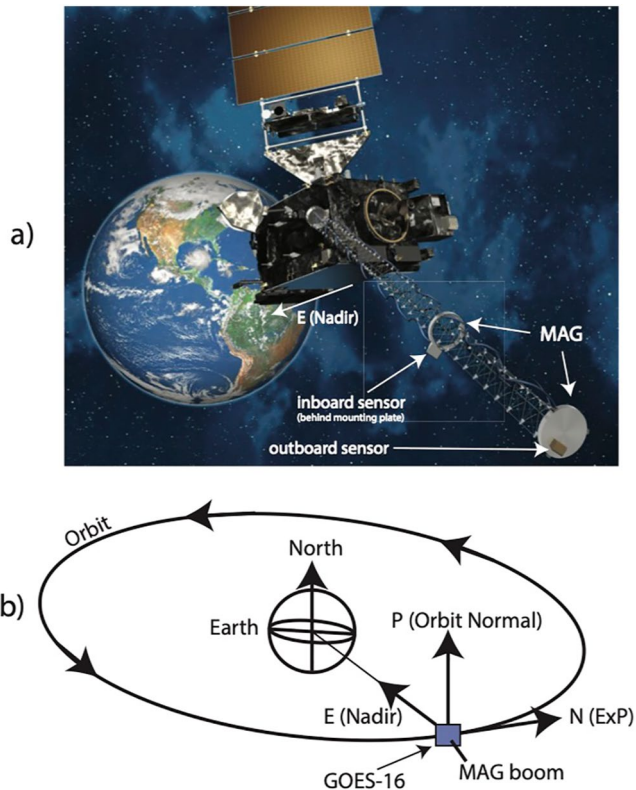


Figure 1. (a) The Geostationary Operational Environmental Satellite (GOES)-R spacecraft indicating the magnetometer (MAG) instrument locations along the boom (<http://www.goes-r.gov>). The MAG is comprised of an inboard (IB) and outboard (OB) magnetometer. (b) The geostationary orbit of GOES-16 in EPN-coordinates indicating the magnetometer (MAG) boom pointed away from Earth.

tri-axis fluxgate vector magnetometers mounted on a long boom with the inboard MAG sensor (IB) positioned 6.35 meters along the boom from the spacecraft and the outboard sensor (OB) at the end of the boom or 8.55 meters out from the edge of the spacecraft bus. The GOES-16 and GOES-17 satellites are, as of writeup, operational and located at geostationary orbit at 75.2° West longitude (called GOES-East) and 137.2° West longitude (GOES-West), respectively. The data used in this study were taken either from these two longitudes, or, in between longitudes where GOES satellites undergo testing or are placed into storage.

Post launch testing of the GOES-16 MAG revealed anomalies in the data, with the IB sensor data showing more significant issues (Loto'aniu et al., 2019). Similar post-launch tests of the GOES-17 MAG did not show the same level of contamination. This was in large part due to redesigns on GOES-17 as a result of lessons learned from issues observed with the GOES-16 MAG. Redesigns for the follow-on GOES-R MAGs (GOES-S, -T, and -U) concentrated on improving on-orbit thermal stability by adding extra thermal blankets, redesigning boom accommodation and refocusing on maintaining magnetic cleanliness best practices from hardware integration through to launch (see Loto'aniu et al. (2019) or further details).

Closer analysis of the GOES-16 IB sensor data offsets revealed that they were more sensitive to on-orbit thermal environment changes than expected. The source of the thermally induced contamination is not known. However, the GOES-16 geostationary orbit shown in Figure 1b is a particularly harsh thermal environment where temperatures can swing over 100°C between day and night. In addition to diurnal thermal environment variations, as Earth orbits, the Sun seasonal effects also result in instruments at geostationary orbit being exposed to varying thermal environment with time of year. Hence, any soft magnetic material(s) with magnetic properties that change with changing thermal conditions located close to the sensors will also produce a time varying contamination in the sensor data. For details of the anomalies observed in the GOES-16 MAG data and lessons learned, see Loto'aniu et al. (2019).

This study presents a demonstration of how machine learning (ML) methods can be used to correct contamination in the GOES-16 IB sensor time series data. Both unsupervised and supervised methods, which are two sub-classes of ML methods (Stinis, 2019), are utilized to correct the GOES-16 IB sensor vector magnetic field E-component (Earthward). The component direction is shown in Figure 1b (Also see, Section 3).

Unsupervised ML uses unannotated/untagged data to learn patterns and anomalies in the input data. The most widely used unsupervised ML method is clustering, which finds groups of data that have similar spatial and/or temporal behavior. We utilize the *k*-Shape clustering method, which was introduced in 2015 for time series data (Paparrizos & Gravano, 2015). Different from other clustering methods, *k*-Shape uses a different method to calculate the distances between time series as well as a different method to find centroids in each cluster. *k*-Shape uses normalized cross-correlation as a distance measure, which allows for shape comparison between time series because this distance measure is invariant to scaling and shifting. As a result, the *k*-Shape algorithm is domain independent and scalable for varying volumes of data and highly effective in clustering quality (Paparrizos & Gravano, 2015). The *k*-Shape method has been successfully applied in energy usage pattern recognition and energy load forecast (Jarábek et al., 2017; Yang et al., 2017). The *k*-Shape clustering method is used here to find patterns within the differences between the GOES-16 OB and IB sensor E-component time series that are signatures of contamination in the IB sensor data induced by thermal environment changes.

Supervised ML methods, on the other hand, use annotated data for a wide range of applications such as natural language processing (Greff et al., 2017), image convolution, super-resolution and denoting (Diaz Baso

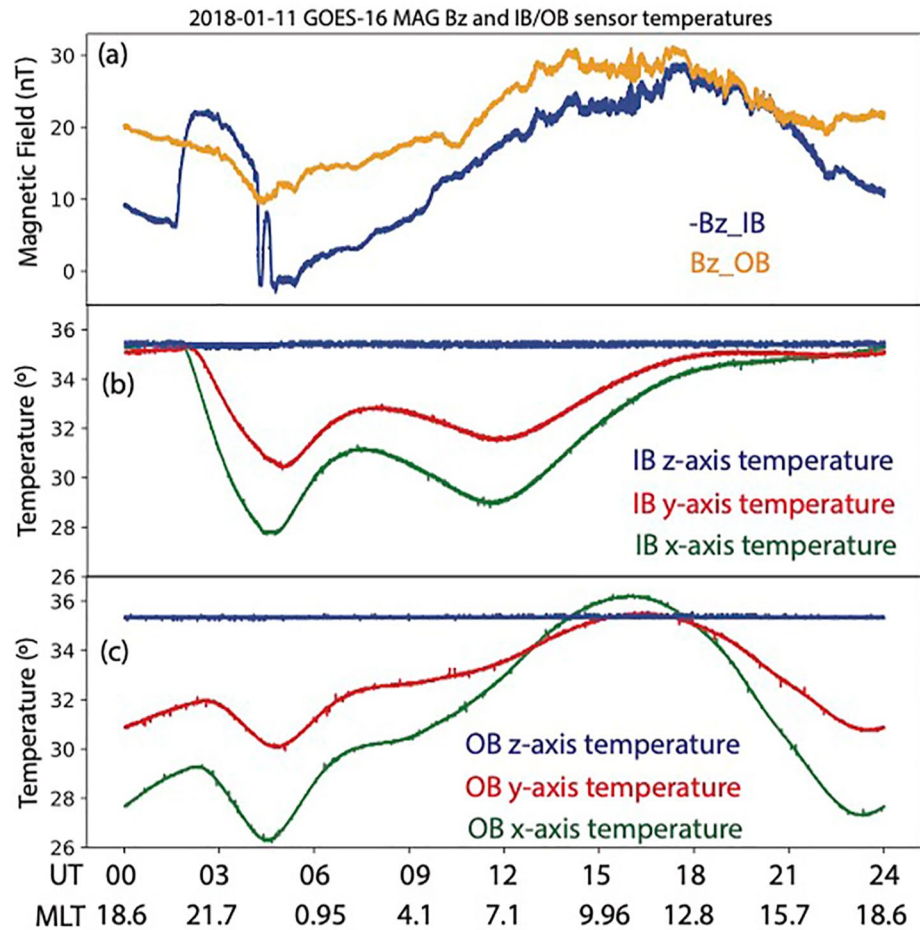


Figure 2. (a) Time series plot of GOES-16 MAG outboard (OB) and inboard (IB) sensor Bz magnetic field in the sensor frame for January 11, 2018. Corresponding temperature readings from all three components of the IB (b) and the OB (c) sensors. Note that the x-axis is given in both Universal Time (UT) and in Magnetic Local Time (MLT).

& Asensio Ramos, 2018; Díaz Baso et al., 2019), Earth observations (dos Santos, 2020; Jeppesen et al., 2019), and space weather prediction (Camporeale et al., 2020; Inceoglu et al., 2018). For sequenced data, such as time series, where the ordering of the data points is important, a modified version of recurrent neural networks named long short-term memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) is shown to be more effective and scalable (Greff et al., 2017) and we employ LSTM in this study. The LSTM networks are utilized for many problems from language modeling (Merity et al., 2017) to solar flare predictions (Liu et al., 2020). For each cluster identified, we build an LSTM and utilize the resultant networks as regressors to correct the GOES-16 IB E-component data.

In Section 2, we provide an example of the data issues we want to correct, while Section 3 describes the MAG data selected for the study. The analyses and results are explained in Section 4. Finally, we give our conclusions in Section 5.

2. Temperature Effects

An example of the GOES-16 high-resolution (10 Hz) MAG data is shown in Figure 2a. The panel shows OB and IB sensor frame magnetic field Bz-components, which point along the boom. The OB sensor is mounted on the outside of a baseplate at the end of the boom as shown in Figure 1a, while the IB sensor is mounted on the side of its baseplate facing the spacecraft. In this configuration, with the two sensors only a couple of meters apart, Bz_OB measurements should be virtually identical to $-Bz_{IB}$. However, as can be seen in the

top panel, they are not equal, and this observed difference in diurnal time series trends between the OB and IB data is an example of the issue we are trying to correct.

Figures 2b and 2c show the corresponding thermistor readings from the IB and OB sensors along each sensor X , Y , and Z axes. Each MAG sensor has a heater, and the heater duty cycle is triggered by the Z -axis temperature. As seen in Figures 2b and 2c, the Z -axis temperature remains practically constant throughout the day because the heater dead-band range is set to 0°C meaning that any drop from the operating temperature of the Z -axis (35°C in this example) turns the heater on. This is also true for the OB sensor. The constant Z -axis temperature readings provide no useful information to correct the IB/OB differences observed in Figure 2a.

The X -axis and Y -axis temperatures do not remain constant, indicating uneven heating of the sensor throughout the day by the heater and suggests that temperature gradients across the sensors are important factors. However, analysis found no clear correlations between temperature readings on three axes of the sensor and the IB/OB trend differences. Hence, thermistor readings from the GOES-16 MAG sensors are not good parameters to use for our correction algorithm. This could be because the thermistors are not located at optimal positions. However, solar angle changes throughout the year changes the illumination profile on the sensors, and we show below that this is a better parameter to use to account for thermal environment changes.

3. Data Selection

The MAG vector magnetic field data used for the machine learning analysis are in EPN coordinates, which is shown in Figure 1b. In the EPN frame, the P -axis is defined as normal to the orbit plane (Poleward), E is the Earthward (nadir) direction, and N (Normal) completes the right-handed system (Loto'aniu et al., 2019). We use EPN coordinates to provide consistency with the frame most commonly used for publicly available GOES MAG data.

Magnetic field data used from both IB and OB sensors spans January 2017 to November 2020 for GOES-16 and late March 2018 to November 2020 for GOES-17. Before any further analyses, we applied data selection criteria to remove low quality data for the purpose of this study. Our data selection criteria are; (a) data shall not have gaps more than 1 hr, (b) if the data gap is shorter than 1 hr, this gap cannot last more than 5 min continuously, meaning that it must consist of gaps shorter than 5 min. These two criteria are applied to the IB and OB magnetometer data as well as satellite location data in geographic latitude, longitude, and distance from Earth, which will be used to calculate the magnetic local times (MLT). For example, if a data set has 30 min total gap, which consists of 7 shorter gaps, each of which is shorter than 5 min, then the data set is accepted. However, if there are only 3 shorter gaps, durations of each is 10 min, then we disregard this data. After the selection criteria, we have magnetic field and satellite location data for 1,015 days for GOES-16 and 896 days of data for GOES-17, which span between November 29, 2017 and November 15, 2020, and between March 23, 2018 and November 18, 2020, respectively. NOAA currently uses the 1-min averaged MAG data in operations and therefore we also use the 1-min resolution data (~ 0.0167 Hz) in this study and linearly interpolate data gaps that are shorter than 5 min.

4. Analyses and Results

The GOES-16 and GOES-17 IB and OB MAG data in EPN coordinates spanning the data selection criteria period are shown in Figure 3. Note that around mid-October 2018, there is a shift in the GOES-17 magnetic field measurements, particularly in the E and N -components. During post-launch testing, GOES spacecraft undergo spin maneuvers to determine MAG instrument on-orbit calibration offsets. The observed shift in the GOES-17 magnetic field data corresponds to when the calibration values were applied to the magnetic field data products. Due to this shift, the GOES-17 MAG data from 2018 were not used in the LSTM networks.

The magnetic field observed at geostationary orbit shows both natural diurnal and seasonal variations. In the P -component, the field is stretched on the nightside lowering field strength and compressed on the day-side due to solar wind pressure resulting in stronger magnetic field on the dayside. Seasonally, geomagnetic

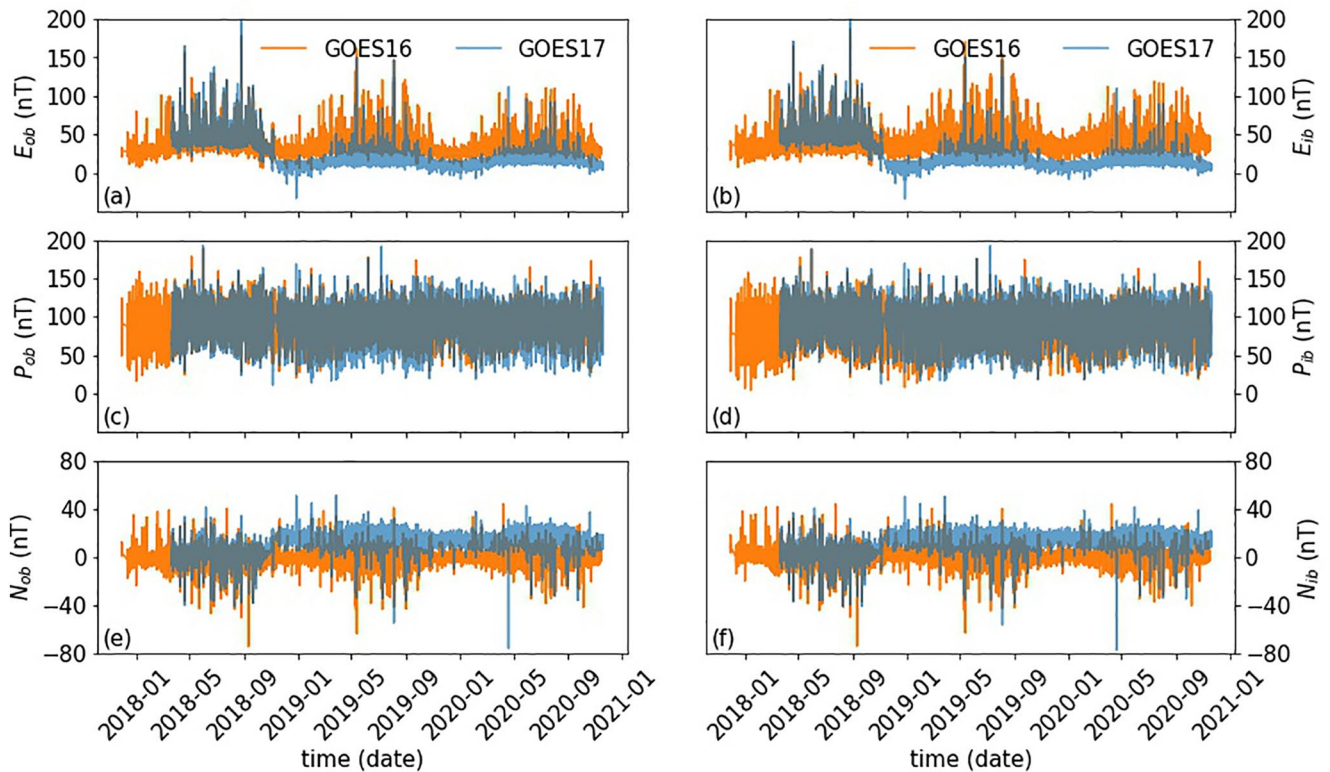


Figure 3. Magnetic field strengths in the E, P, and N coordinates measured by the outboard (left panel) and inboard (right panel) magnetometers mounted on the GOES-16 (orange) and GOES-17 (blue).

activity tends to enhance around equinox, where the Sun-Earth interaction is enhanced due to the equinoctial effect and the Russell-McPherron effect (Russell & McPherron, 1973). Around equinox, the position of Earth in heliographic latitude is also more favorable for interactions with high speed solar wind at high heliographic latitudes (the axial effect; Cliver et al., 2000). These changes in activity due to seasonal effects are clearly visible in the E and N-components of the magnetic field.

In order to justify using the GOES-17 MAG data to build the LSTM networks, we show in Figure 4 differences between OB and IB time series (E, P, and N coordinates) for all the GOES-16 and GOES-17 data used as a function of MLT. We compare the data from OB to IB for the reason that the two sensors should measure the same ambient magnetic field since they are located only a couple of meters apart along the boom of each spacecraft. Hence, OB-IB should be constant with time although not necessarily 0 nT because there could be a constant offset in one or both sensors. If there are significant issues such as magnetic contamination or thermal effects that vary with time on one or both sensors, then OB-IB should also vary with time. It is highly unlikely that time varying unwanted signals in each sensor data would be in lockstep such that OB-IB was constant with time. It is now common practice in space science missions to fly more than one identical magnetometer onboard in-part to be able to do this inter-sensor comparison.

In Figure 4, differences between OB and IB time series (ΔE , ΔP , and ΔN) for all the GOES-16 and GOES-17 data are shown as a function of MLT. The differences between the GOES-16 MAG OB and IB magnetic field measurements in E and P-components vary with amplitudes as high as 25 nT. The results for GOES-17, on the other hand, display much less diurnal variations, and the amplitude of the variations does not exceed around ~ 2 nT within each day. This is also confirmed by comparing the standard deviations in the datasets in Figures 4c, 4f, and 4i. The figure also justifies our decision to apply the correction first to the GOES-16 E-component data because it shows the largest variation in the OB-IB differences.

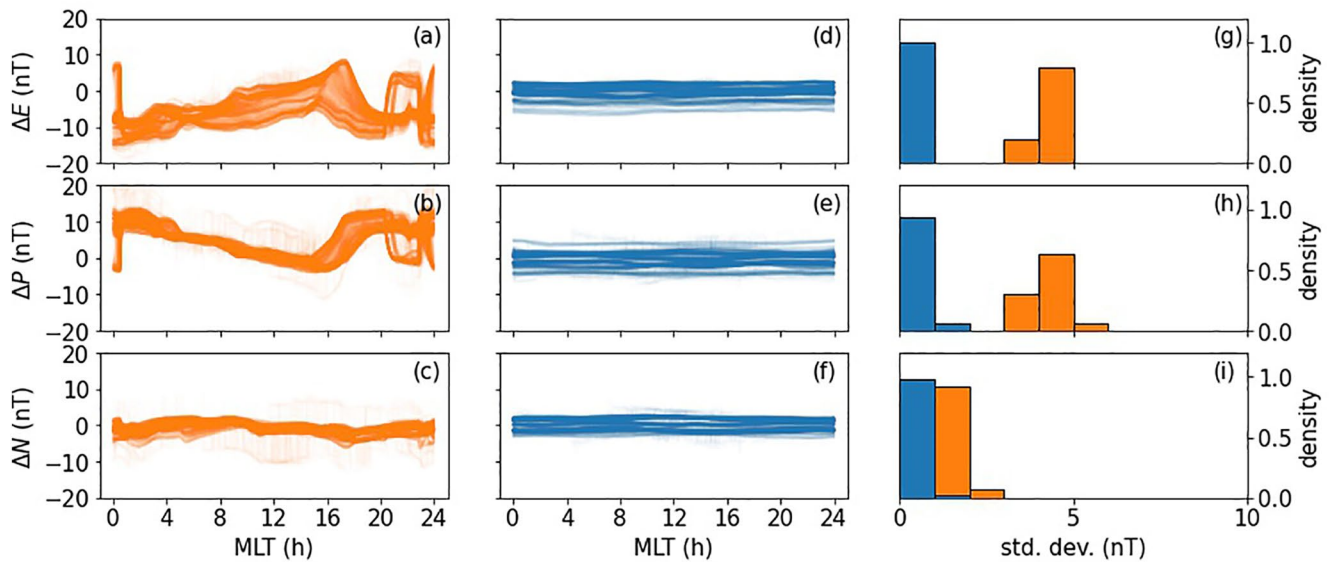


Figure 4. Differences between OB and IB time series for the E-component (ΔE), the P-component (ΔP), and the N-component (ΔN) for GOES-16 (a–c) and GOES-17 (d–f). Also shown are the corresponding distributions of standard deviations of the differences between OB and IB (g–i).

4.1. Clustering the Differences in the GOES-16 MAG Outboard and Inboard Sensor Data Using K-Shape

A closer inspection of the differences in the GOES-16 MAG OB and IB sensor data (OB-IB) hints at the presence of clusters, while the OB-IB differences from GOES-17 do not show such signs. We investigate the presence of any clustering behavior in the GOES-16 E-component OB-IB differences, ΔE , using *k*-Shape clustering (Paparrizos & Gravano, 2015). The *k*-Shape clustering method is an efficient and domain independent clustering method based on a scalable iterative refinement procedure. This method uses a normalized version of the cross-correlation measure as a distance measure to find the centroid of each cluster and update the members that belong to each cluster in each iteration. Additionally, using a normalized version of the cross-correlation measure as a distance measure enables comparison of the shapes of the time series and provides preservation of their shapes, as this method is invariant to scaling and shifting (Paparrizos & Gravano, 2015).

The number of clusters, *k*, is expected as an input for the *k*-Shape algorithm, much like *k*-means method that is another clustering algorithm based similarly on iterative refinement procedure (MacQueen, 1967). In estimating the number of clusters within the GOES-16 ΔE data, we input cluster numbers $k = 1, 2, 3, \dots, 10$

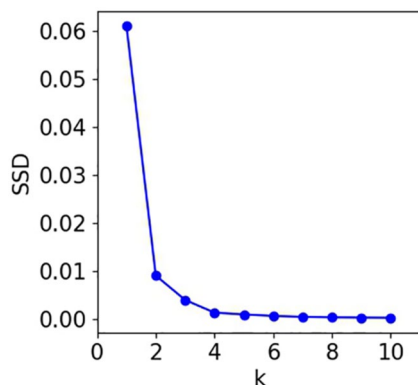


Figure 5. Sum of squared distances (SSD) calculated for each number of clusters, which ranges from 1 to 10.

and calculate the sum of squared distances (SSD) for each cluster, with results shown in Figure 5. The relationship between SSD and *k*, often called a scree-plot, is generally used as an indication of the optimum number of clusters in a data set (Paparrizos & Gravano, 2015). The results show that further decreases in SSD are very small after cluster number 4, meaning that the optimum number of clusters in the ΔE data is 4.

To evaluate the goodness of the *k*-Shape algorithm on clustering, we use the silhouette coefficient (Rousseeuw, 1987). The silhouette coefficient is used when the ground truth of the cluster labels are unknown. The silhouettes coefficient is calculated using the following equation:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

where $a(i)$ represents average dissimilarity of *i* to all other objects within cluster, whereas $b(i)$ denotes average dissimilarity of *i* to all other objects in the next nearest cluster (Rousseeuw, 1987), meaning that we need the

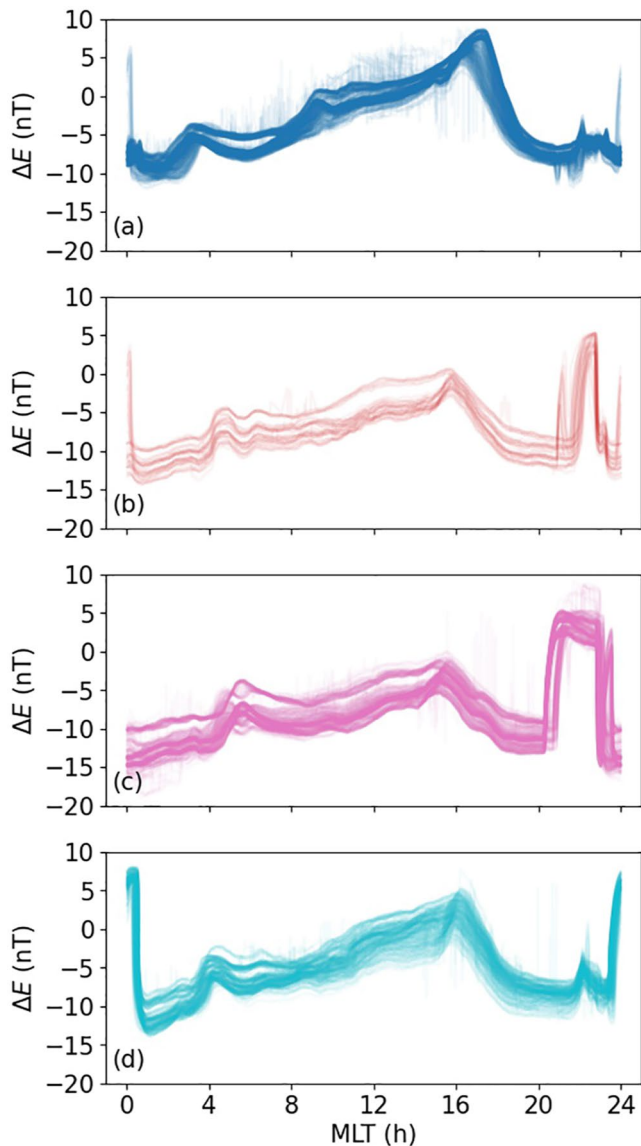


Figure 6. Four clusters, C1 (a), C2 (b), C3 (c), and C4 (d), detected in the GOES-16 OB-IB E-component magnetic field data.

labels obtained by a clustering method, and proximities between objects. The silhouette coefficient ranges from -1 to $+1$. Values close to -1 represent incorrect clustering, while values close to $+1$ show highly dense clustering. Additionally, we calculated the statistical significance of the silhouette coefficient by randomly generating four cluster labels and calculating the silhouette coefficients for each iteration. We iterate this process for 10,000 times. The silhouette coefficient for four clusters is calculated as 0.47 ($p \ll 0.01$).

Figure 6 visualizes the four clusters have varying temporal characteristics but similar maximum amplitude difference of ~ 20 nT. Cluster-1 (C1), shown in Figure 6a, consists of 413 members and shows an abrupt drop at the beginning of the day from about 6 nT toward -10 nT, then gradual increase to around 16 MLT to peak around 8 nT. The values then decrease more rapidly to ~ -10 nT until 20 MLT, after which the members display less variation except for the spike at the end of the day which is actually part of the same spike observed at the beginning of the day because $0 \text{ MLT} = 24 \text{ MLT}$. Cluster-2 (C2), shown in Figure 6b, consists of 44 members and has the same spike at the beginning/end of the day and shows similar gradual increase until 16 MLT before it decreases. However, there are two further peaks that occur between 20 and 24 MLT before the spike at the end of the day.

Cluster-3 (C3) has 326 members and has a temporal trend similar with C1 and C2 until 20 MLT. However, there is no peak at the beginning/end of the day. After 20 MLT, there is an abrupt increase in values from around -15 – -5 nT, and this level stays almost the same with a slight decrease until around 23 MLT, where it displays a sharp decrease down to around -15 nT (Figure 6c). Cluster-4 (C4) has 232 members (Figure 6d) and shows generally similar temporal variations to C1 with the exceptions that the spikes around 0–1 MLT and 23–24 MLT are made up of more members (thicker region) than C1. We discuss the meaning of the cluster temporal trends in the following section.

4.2. Effects of the GOES-16 Orbital Configuration

If the GOES-16 IB and OB magnetometer sensors were perfect instruments, the ΔE 's shown in Figure 6 would be zero, and therefore, no clusters would be found using k -shape analysis. However, analysis indicates an optimum number of four clusters in the ΔE data, and this suggests that the clusters may be seasonally dependent. This is confirmed in Figure 7, which shows the time of year over which each cluster occurs.

The C1 cluster covers the months from late April to early September (northern hemisphere spring and summer), while C2 covers late February and mid to late October. C3, on the other hand, covers late October to late February (around northern hemisphere winter), and C4 covers March to early April and September (equinox months). The importance of the clusters occurring at different times of the year is that solar illumination on the spacecraft varies as Earth orbits the sun resulting in varying thermal environment along the boom, including at the sensor locations.

The amount of direct sunlight a satellite experiences is defined by the solar beta angle, β . Figure 8 shows the distributions of the β angles for each cluster. GOES satellites encounter two periods during the year in which they are shadowed from sunlight by Earth around local midnight. Eclipses occur when $\beta \sim 0^\circ$ around the equinox months from about late February to mid-April, and from late August to mid-October. The C4 cluster β distribution is centered at $\sim 0^\circ$, and the spike in C4 around 0 or 24 MLT, shown in Figure 6, is probably the result of strong differential thermal conditions as GOES-16 moves in and out of Earth eclipse.

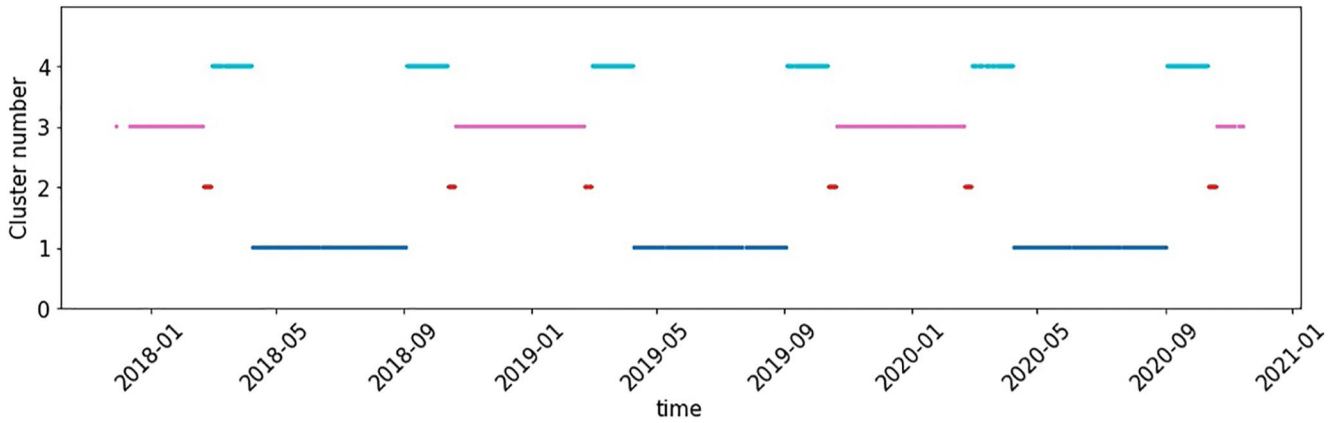


Figure 7. Temporal variation of the clusters. The color coding is the same with Figure 6.

Clusters C1 and C2 include small numbers of cluster members that overlap the equinox months where β is low, which probably explains the less dense spikes also around midnight MLT in Figures 6a and 6b.

The C1 cluster mostly occurs outside eclipse periods by Earth, with β positive and clustered around 23.45° . The 23.45° value is the obliquity of the ecliptic for Earth and therefore the maximum β for geostationary orbiting satellites. For C3, which includes the winter months, all of its cluster members occur outside eclipse

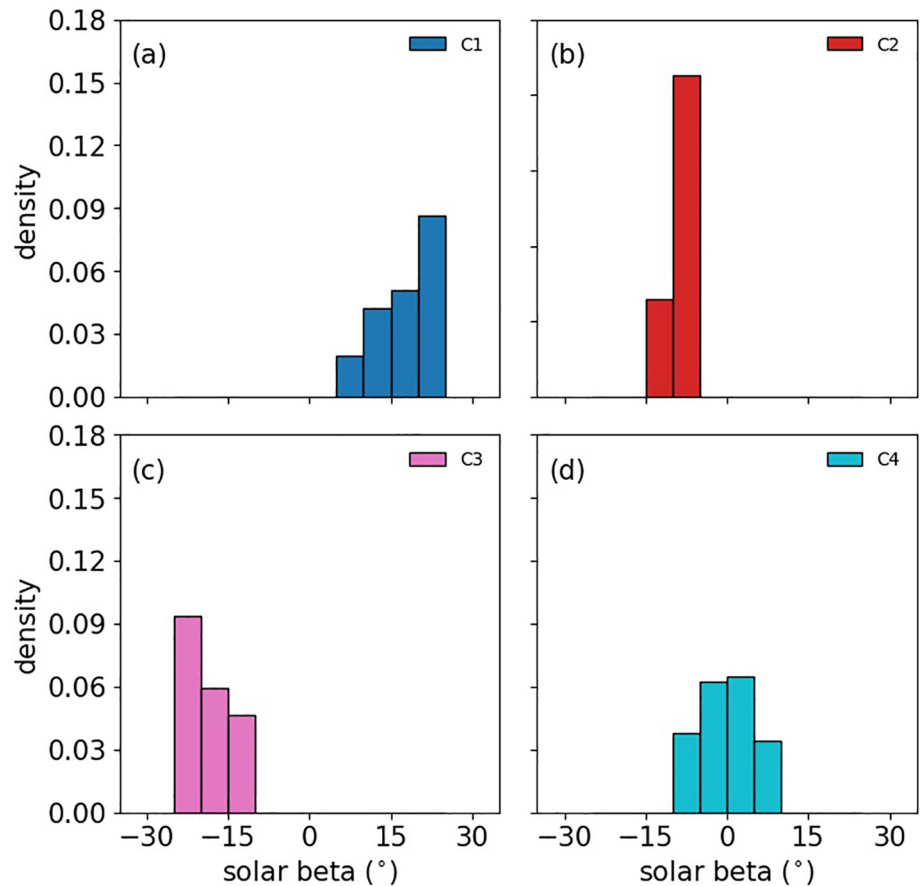


Figure 8. Distribution of the solar beta angles in C1 (a), C2 (b), C3 (c), and C4 (d). The color coding is the same with Figure 6.

season, and this explains the lack of spikes in the C3 cluster shown in Figure 6c at 0/24 MLT and the high absolute β clustered around the maximum -23.45° in Figure 8c.

Besides eclipse season where Earth shadows the spacecraft, from September to about mid-April each year the GOES-16 MAG sensors and boom can be completely or partially shadowed by the spacecraft bus (Loto'aniu et al., 2019). This time period mainly covers clusters C3 and C2 and explains why they both observed sudden jumps in ΔE starting around 21 MLT as seen in Figures 6b and 6c and lasting until about 22–23 MLT. The orientation of the boom relative to the spacecraft bus as shown in Figure 1 is the reason shadowing by the bus of the boom happens within the 20–24 local time sector. Hence, even though C3 is outside eclipse season as shown by its β distribution, local spacecraft shadowing can still affect the contamination on the IB sensor.

The cluster diurnal trends in Figure 6 between about 02 and 20 MLT show a similar pattern for all four clusters irrespective of time of year. This trend is likely the result of the changing temperature gradient across the boom and the sensors as the spacecraft orbits Earth toward the dayside across noon and back toward the nightside. ΔE approaches 0 nT for all clusters close to 16 MLT where the orbit is probably optimal to minimize the thermally induced effects. Afterward, ΔE goes negative again as solar illumination moves to the other side of the boom.

The results of the k -shape analysis found clusters in the ΔE time series and these clusters are associated with solar β angle. This justifies our use of β as an input to the LSTM network model development. However, we also need to include MLT because shadowing also depends on the spacecraft local time as the spacecraft bus can shadow the boom even outside eclipse season, and temperature gradients across the boom varies as the spacecraft orbits Earth.

4.3. Differences Between GOES-16 and GOES-17 MAG Sensor Measurements

In Figure 4, we showed that the MAG OB-IB sensor differences for GOES-17 showed less variability and were less sensitive to local time than corresponding GOES-16 MAG OB-IB differences. Here, we subtract the GOES-17 MAG OB E-component measurements from the GOES-16 MAG IB and OB observations in order show that for each cluster the GOES-16 IB sensor has greater unwanted variations. In order to take into account the differences in longitudes of the two satellites, a model magnetic field is used and subtracted from both satellites time series.

The model used was the well known TS04 model, and values were calculated using the GEOPACK IDL package (Tsyganenko, 2002a, 2002b; Tsyganenko & Sitnov, 2005; Tsyganenko et al., 2003) with input solar wind and geomagnetic indices taken from the NASA-CDAWeb. The model was calculated at 3-min resolution at both satellite locations and interpolated to the 1-min resolution of the measured OB and IB data. We then subtracted the model results from the GOES-16 OB and IB MAG measurements creating GOES-16_{OB}-GOES-16_{TS04} and GOES-16_{IB}-GOES-16_{TS04}, respectively. This was repeated for the GOES-17 OB sensor resulting in the time series GOES-17_{OB}-GOES-17_{TS04}, and then this time series was subtracted from the GOES-16 IB and OB results giving residuals for each E-component (ΔE) cluster, as shown in Figure 9.

It is clear from Figure 9 that the average variations (black lines) in ΔE for the GOES-16 OB clusters (left panel) show less variability than the GOES-16 IB ΔE clusters (right panels). The broadening or thickening of the color curves around the night time MLT hours is due to the fact that the magnetic field model is less accurate in this region. From the comparisons we conclude that the magnetic field measurements from the GOES-16 IB sensor is more affected than the OB sensor by the satellite orbital configuration, diurnal variations, and shadowing effects that were previously discussed.

4.4. Correcting GOES-16 IB MAG Data Using LSTM Networks

We used LSTM networks (Hochreiter & Schmidhuber, 1997) to correct the GOES-16 IB sensor E-component measurements. An LSTM node consists of four interactive parts; a memory cell, an input gate, an output gate, and a forget gate. An illustration of an LSTM node is shown in Figure 10. These gates are neural network layers with sigmoid activation functions that are followed by point-wise multiplications and they

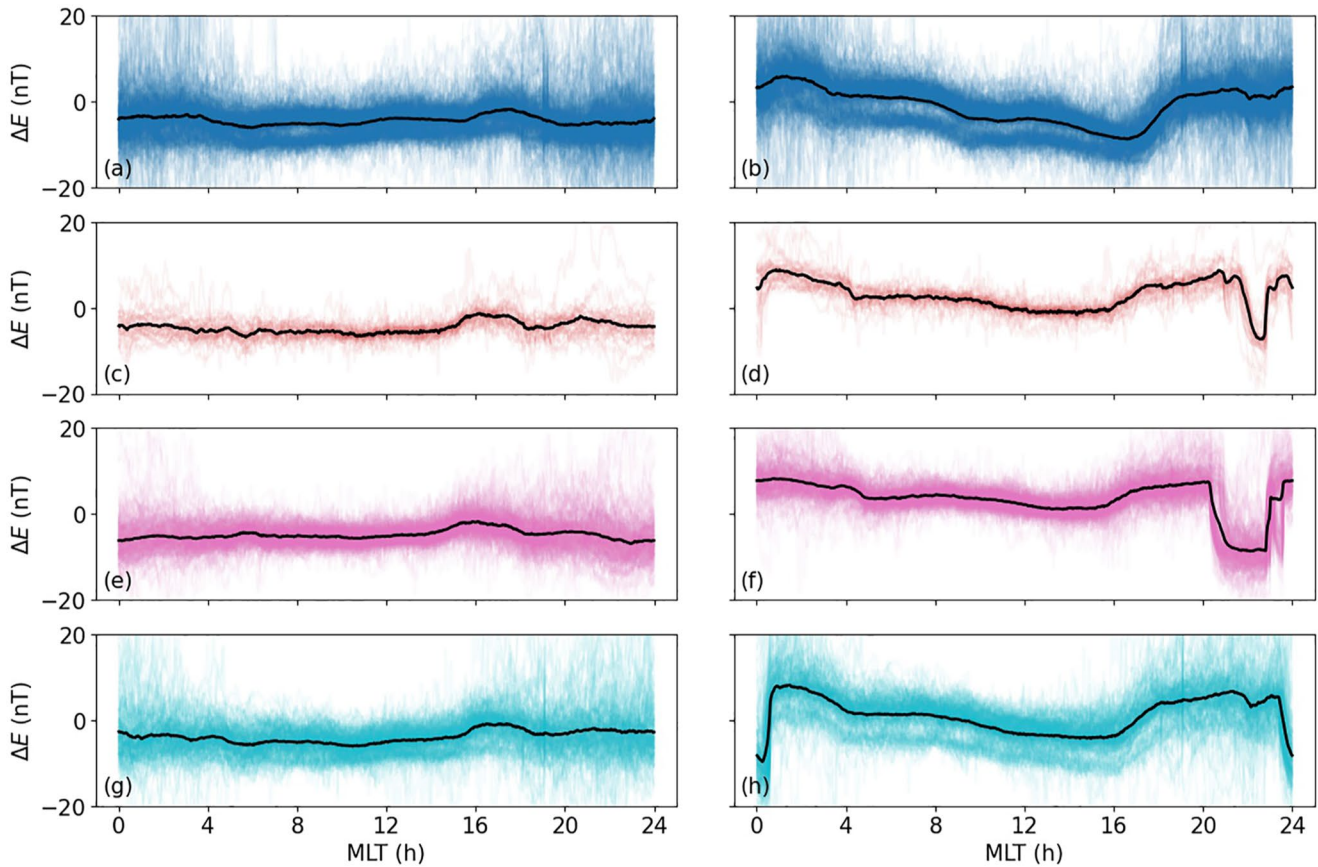


Figure 9. The left panel shows magnetic field E-component $[(GOES-16_{OB} - GOES-16_{TS04}) - (GOES-17_{OB} - GOES-17_{TS04})]$ for each cluster found in the GOES-16 data, while the right panel shows $[(GOES-16_{IB} - GOES-16_{TS04}) - (GOES-17_{OB} - GOES-17_{TS04})]$. The color coding is the same with Figure 6. The solid-black lines are the mean values in each cluster. Note that the y-axis in each plot is zoomed-in to make the variations more visible. This plot shows that the variability in the IB sensor (right panel) is more significant than the OB sensor (left panel) for all four clusters.

regulate the flow of information in an LSTM unit. The cell state, C_t , which is the core concept of the LSTMs, can learn which information to transfer forward or to forget using the gates. The cell state at time t (C_t), given by

$$C_t = C_{t-1} \odot f_t + \tilde{C}_t \odot i_t, \quad (2)$$

is continuously updated by the cell state at time $(t - 1)$ (C_{t-1}) and the candidate cell state at time t (\tilde{C}_t). In Equation 2, f_t and i_t are the forget and input gates, respectively. The forget gate regulates how much of the previous cell state (C_{t-1}) should be forwarded to the current cell state (C_t). The input gate, on the other hand, determines how much of the candidate cell state (\tilde{C}_t) is to be added to the current cell state. The forget and input gates can be written as

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f),$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i), \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c),$$

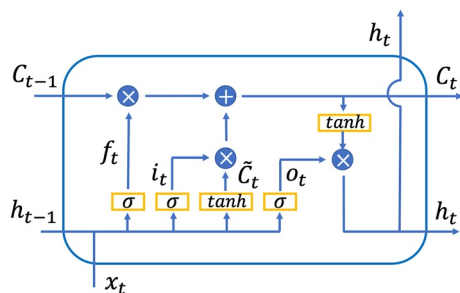


Figure 10. Illustration of an LSTM node; C_{t-1} and C_t denote the cell state, while h_{t-1} and h_t represent the output vector at time t and $(t - 1)$, respectively. The input, output, and forget gates are represented with i_t , o_t , and f_t , while x_t is the input vector. The point-wise multiplication and addition are represented with \times and $+$ in the blue circles. The sigmoid (σ) and tanh activation functions are shown in orange boxes.

Table 1
The Chosen Hyper-Parameters That Provide the Smallest $RMSE_{Test}$ Values Between the Measured GOES-16 OB E-Component Magnetic Field and Corrected GOES-16 IB E-Component Magnetic Field

Cluster	Training data (days)	Node	Epoch	Batch size
C1	160 (~60%)	10	10	64
C2	17 (~60%)	25	10	64
C3	110 (~60%)	10	10	64
C4	110 (~70%)	50	10	128

Note. The percentages in brackets show the length of the training period.

while the output gate (o_t), which controls how much of the cell state is to be forwarded to the output at time t (h_t), is defined as

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o), \quad (4)$$

$$h_t = o_t \odot \tanh(C_t),$$

where the sigmoid and hyperbolic tangent activation functions are given by

$$\sigma(z) = \frac{1}{1 + e^{-z}}, \quad (5)$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

In Equations 3 and 4, W and b represent the weights and biases in different gates, respectively. The input vector at time t is denoted as x_t , while the output vector at time $(t - 1)$ is represented as h_{t-1} . The point-wise multiplication is shown by \odot , whereas the matrix vector multiplication is indicated by \times .

The LSTM networks we used consist of an input, one hidden, and an output layer. We build one LSTM network for each cluster and utilize our networks as regressors. The LSTM networks were trained using GOES-17 OB data, MLT, solar β angle, and satellite longitudes as inputs, and the GOES-17 IB data as the output for the overlapping GOES-16 and GOES-17 observational time periods. As previously mentioned, GOES-17 MAG data from 2018 were excluded because it includes pre-calibrated MAG data (See, Figure 3). We then tested our algorithms using GOES-16 data.

For training of the LSTM networks, we employed an Adam optimizer (Kingma & Ba, 2014), and mean absolute error as our loss function as we are using our LSTM networks as regressors. Additionally, hyper-parameter optimization was done using a simple grid-search method. In this method, we used 60%, 70%, and 80% of the data for training periods containing 10, 25, and 50 nodes (neurons) in the LSTM network with 64, 128, and 256 data points as batch sizes, and 10, 50, and 100 as epoch numbers. Each configuration is then run 50 times to give a distribution of the root mean squared errors ($RMSE_{Test}$) for the test data. The $RMSE_{Test}$ is calculated between the measured GOES-16 OB E-component magnetic field and corrected GOES-16 IB E-component magnetic field because we expect the two datasets to be closer than the observed differences. The reason for using the $RMSE_{Test}$ is that we use our LSTM networks in a transfer learning fashion. Therefore, the most important thing is to choose the hyperparameters that would result in the smallest $RMSE_{Test}$, instead of RMSE calculated for the validation set, which is calculated between the measured GOES-17 OB E-component magnetic field and corrected GOES-17 IB E-component magnetic field.

As an example, we show results for hyper-parameter optimizations of the C1 cluster in Figure 11. For C1, epoch sizes 50 and 100 generally result in higher $RMSE_{Test}$ values, while epoch size 10, on the other hand, results in smaller numbers. The analysis shows that 160 days of training data (~60%) give better results. Smaller batch sizes, similar to epoch size, produces smaller $RMSE_{Test}$ values. As a result of the grid-search method for the hyper-parameter optimization, we chose 160 days of training data (~60%), 10 nodes, 10 epochs, and batch size of 64 to train our LSTM network for C1. Table 1 shows the hyper-parameters values chosen for all four clusters. For C2, these numbers are 17, 25, 10, and 64, respectively. As for C3, the minimum $RMSE_{Test}$ values are reached using 110 days of training data, 10 nodes, 10 epochs, and the batch size of 64, whereas 110 days of training data, 50 nodes, 10 epochs, and the batch size of 128 gave the smallest $RMSE_{Test}$ values for C4. Fluctuations in the $RMSE_{Test}$ values, even within one hyper-parameter setting, are quite large because of the stochastic nature of neural networks and therefore each cluster correction algorithm was run in a loop that ended when the $RMSE_{Test}$ reached its minimum.

Figure 12 shows examples of corrected GOES-16 IB E-component (IB_C or green lines) for four different days, which are chosen randomly from the four different clusters. They generally follow the same trend as the measured GOES-16 OB sensor E-component values (OB or blue lines), as well as being less offset from OB compared to the measured GOES-16 IB E-component values (IB_M or orange lines). On February 23, 2019

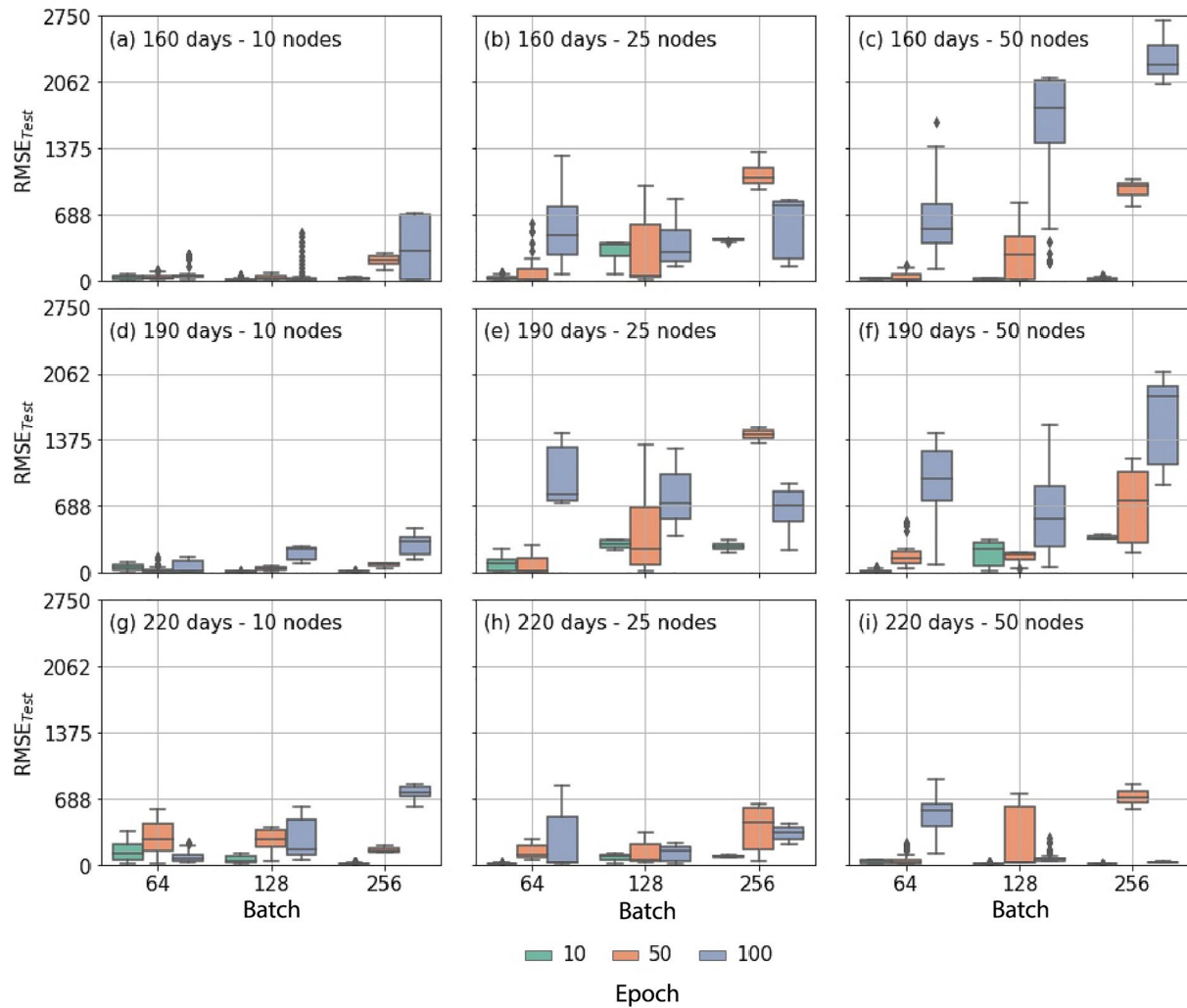


Figure 11. Calculated root mean squared errors (RMSE_{Test}) between the measured GOES-16 OB E-component magnetic field and corrected GOES-16 IB E-component magnetic field for different configurations of hyper-parameters for C1; length of the training data, number of nodes (neurons) in our LSTM network, batch size, and epoch numbers. For example, Figure 11a shows RMSE_{Test} values for 160 days of training data with an LSTM network of 10 nodes for Batch sizes of 64, 128, and 256, and for Epochs of 10 (green), 50 (orange), and 100 (purple).

(panel b), which is in the C2 cluster, the shadowing effect within 20–24 MLT is observed in IB_M but removed by the LSTM network in IB_C. The same is true on January 2, 2019 (panel c), which is in C3. On April 11 and March 2, 2019 (panels a and d), which are in the C1 and C4 clusters, there is a sharp drop in IB_M around local midnight (0/24 MLT) and on both days this drop is corrected using the LSTM networks.

However, the overall amount of correction varies with MLT for the days shown in Figure 12. The LSTM networks seem to almost completely remove sudden changes due to eclipse season (near local midnight) and spacecraft shadowing (20–24 MLT) effects, while trends throughout the day show varying offsets between IB_C and OB. There could be multiple explanations for this including that the GOES-16 OB sensor is also known to suffer some thermally induced effects, but to a much lesser extent than the IB sensor (Loto'aniu et al., 2019).

Looking more generally at the amount of correction after the LSTM networks were applied to the GOES-16 IB E-component data, we replotted ΔE as the difference between the GOES-16 OB E-component and the corrected GOES-16 IB E-component data, with results shown for each cluster in Figure 13. The ΔE for C1 varies between -4 and $+5$ nT with a standard deviation distribution between 0 and 2 nT, while the variances in the offsets between measured values are between 4 and 5 nT. ΔE values in C2 show the least

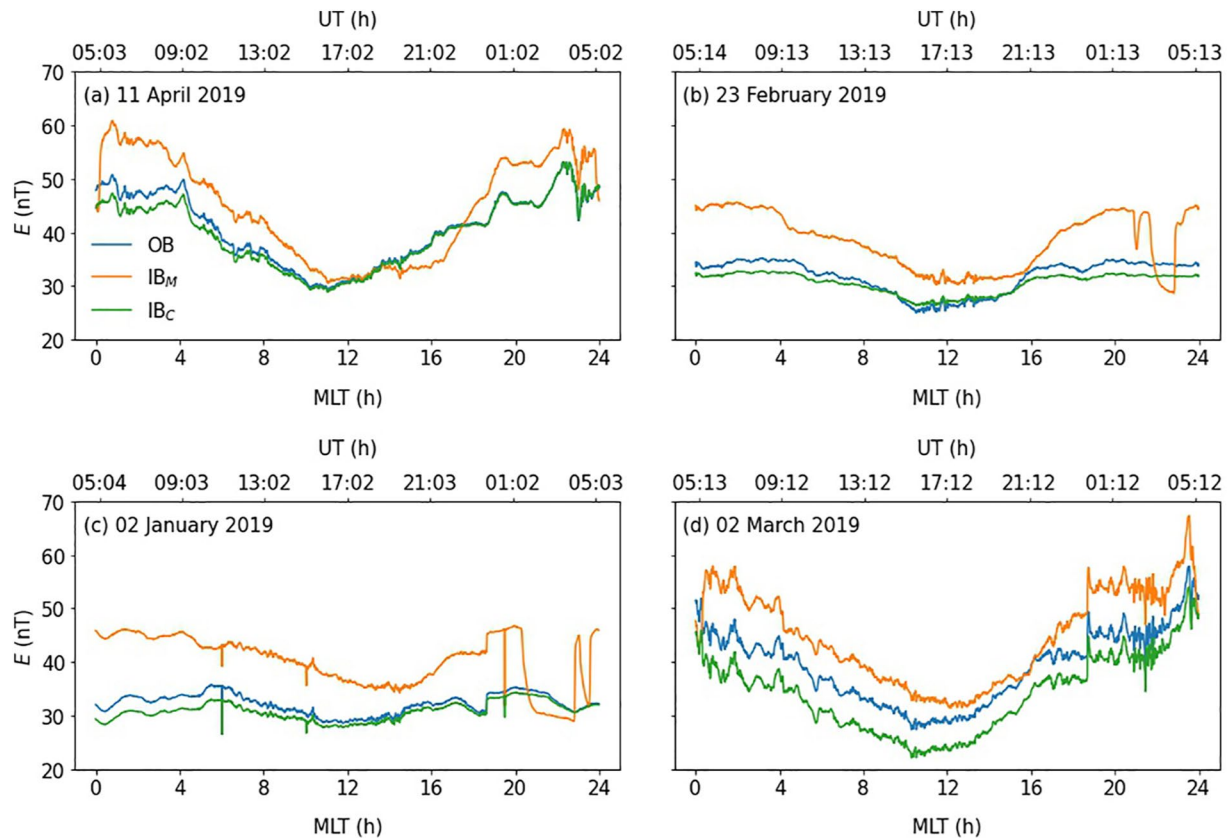


Figure 12. The corrected E-component IB (green), measured E-component IB (orange) and OB (blue) MAG from the GOES-16 satellite for C1 on the April 11, 2019 (a), for C2 on the February 23, 2019 (b), for C3 on the January 2, 2019 (c), and for C4 on the March 2, 2019 (d). Note that we provide both MLTs and UTs in the bottom and top x-axis of each panel.

improvement relative to measured values, with a corrected standard deviation as high as 5 nT. This is due to the fact that there are very few number of data sets in C2 and only 17 days of training data for the LSTM networks, which does not seem as effective in capturing the relationships between input variables.

In Figure 13, the ΔE values in C3 and C4 show similar results as those for C1 with slight differences. All three curves are fairly linear with slight negative gradients. However, C4 ΔE is offset from 0 nT and centered around 5 nT, while both C3 and C1 average closer to 0 nT. This indicates that during eclipse season, which is the C4 cluster time period, there is a few nT offset between the GOES-16 OB and IB E-component that is linear with MLT but not captured by the LSTM. Away from eclipse season, at any MLT the GOES-16 OB and IB E-component observations are closer on average. An example of this offset within the C4 cluster time period is shown in Figure 12d for March 2, 2019. As previously mentioned, there could be multiple explanation for this greater offset during eclipse season including thermal issues with the GOES-16 OB sensor, albeit to a much lesser extend than the IB sensor.

Two identically built and thermally stable magnetometers will still usually possess different factory temperature offset characteristics simply due to slight differences in hardware imperfections. However, the GOES-16 MAGs are not temperature corrected on-orbit because of the lack of correlation between the temperature curves and the magnetic field data as shown in Figure 2 example. The GOES-17 MAG data are also not temperature corrected. The observed differences between the GOES-16 OB and IB MAG measurements is likely partly due to the difference in temperature response between the sensors and partly due to some form of thermally induced magnetic contamination. The results presented do not separate these two effects. Nevertheless, the LSTM method identified that an additional offset remains, shown in Figure 13g, and also linearized this offset as a function of MLT allowing for a simple curve fit subtraction rather than dealing with the complex diurnal variations observed in the data before the LSTM method was applied.

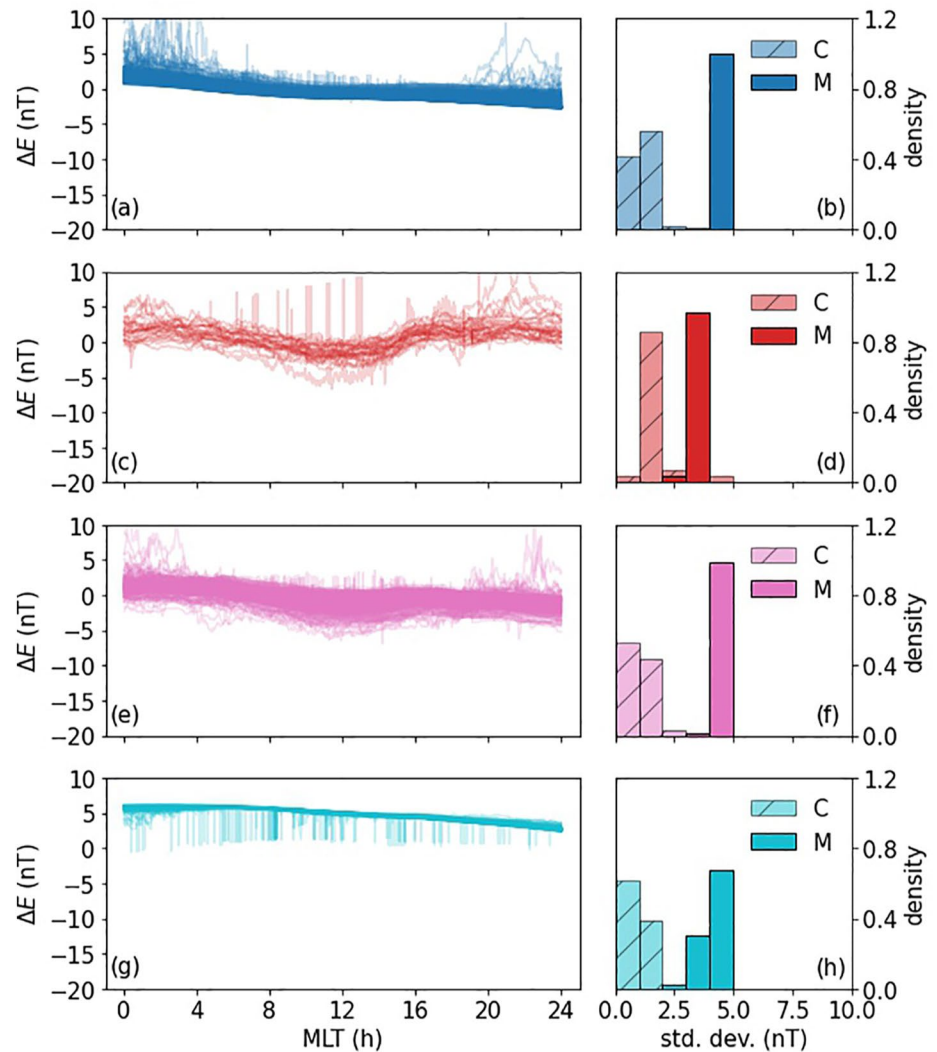


Figure 13. The left panel shows the difference between GOES-16 measured OB and corrected IB E-component magnetic field values for each cluster. On the right panel, we show the standard deviations calculated for the offsets between the measured OB and IB (darker colors) and also the standard variations in the differences between the measured OB and corrected IB (hatched bars). The color coding is the same with Figure 6.

5. Conclusions

This study shows how machine learning can be used to correct thermally induced magnetic contamination observed in the GOES-16 magnetometer IB sensor data. For simplification, we concentrated on correcting only one of the IB vector components, the E-component (Earthward). We started by comparing the differences between the GOES-16 magnetometer OB and IB sensor data in EPN coordinates. The OB-IB comparisons revealed variations reaching up to 25 nT on GOES-16, while this value is limited to about ~ 2 nT for GOES-17. Furthermore, GOES-16 OB-IB show variations in their overall levels with some flat-topped and peak-topped structures after about 20 MLT in the E and P-components, while the differences in the GOES-17 sensors does not exhibit such structures. The GOES-16 MAG has significantly more contamination issues than the GOES-17 MAG.

The *k*-Shape method was used to find possible clusters in the GOES-16 OB-IB E-component differences, ΔE . This method provides domain independent and efficient clustering of time series. The optimum number of clusters was found by comparing the sum of squared distances calculated for each cluster number ranging from 1 to 10 with results indicating that there are four clusters in the data. Further analyses show that the

4 clusters are closely related to the time of year and the solar β angle, which is a measure of the amount of time that a satellite is in direct sunlight. The ΔE trends for each cluster are consistent with thermally induced contamination due to Earth shadowing of the spacecraft during eclipse seasons as defined by β , and due to spacecraft shadowing of the boom.

In order to show that it is the IB sensor on GOES-16 that is the cause of the major variability in ΔE , we subtract the GOES-17 OB E-component time series from the GOES-16 OB and IB E-component time series, while taking into account the difference in the two satellite locations by subtracting the TS04 magnetic field model time series calculated at each of the satellite locations. The results clearly indicate that it is the GOES-16 IB sensor magnetic field measurements responsible for most of the shapes in each ΔE cluster.

Finally, we create LSTM networks for each cluster to minimize the ΔE variations in the clusters. LSTM networks are very effective when sequenced data are considered, such as time series data. Inputs to train the LSTM networks include the GOES-17 OB sensor E-component data, together with MLT, longitudes of the satellite, and solar β angles, while the output includes the GOES-17 IB data. The time periods used were all the periods where GOES-16 and GOES-17 had overlapping observations up to November 2020 from each cluster. We excluded data from 2018 as this period included pre-calibrated GOES-17 data. The hyper-parameters were optimized with a simple grid search algorithm. The trained LSTM algorithm for each cluster was then used to correct the GOES-16 IB E-component data. The results from the corrections indicate that C1, C3, and C4 showed the highest improvements where standard deviations reduced from a range between 3–5 nT and 0–2 nT. For C2, although there was improvement, the diurnal variation still persisted. This is because there are very limited training data available in this cluster.

We have showed in this study that unsupervised ML method, such as k -Shape, allows us to separate out time dependent offset trends in the differences between the GOES-16 OB and IB MAG measurements. In addition, using a supervised ML method like the LSTM networks enabled us to correct the effects from the diurnal and seasonal variations in the GOES-16 IB sensor E-component data. In future work, we plan to improve and extend this method to improve results for C2 and also to include corrections in the P-component. For the N-component, the differences between GOES-16 OB and IB (See ΔN in Figure 4) were not considered significant enough to apply the LSTM networks on the N-component data.

It should be noted that there are additional offsets and anomalies in the data that are not accounted for in this study. For example, GOES-17 undergoes a yaw-flip every 6 months and analysis of that data found an additional DC offset in the GOES-17 MAG instrument data. However, in the E-component, the offset correction was less than 0.1 nT (Loto'aniu et al., 2021) and well outside the resolution of our ML correction method. There is also contamination induced from the firing of arcjets on GOES-16 and GOES-17 (Califf, Early, et al., 2020; Califf, Loto'aniu, et al., 2020) and over the dates used in this study arcjet firing durations were about 90 min every four days. We decided that out of the total data length used, this small amount of data (~1.5%) could be ignored for the purposes of this study. The corrections to the GOES-16 IB sensor E-component data shown in this study are corrections to average unwanted seasonal and diurnal variations. Furthermore, the LSTM method applied to the k -Shape derived clusters tend to flatten out the shapes as a function of MLT but not necessarily always produce zero offsets. However, once flattened or linearized the remaining offsets are much simpler to correct. This proof-of-concept study showcases the strengths of the unsupervised and supervised ML methods in correcting space-based instrument time series data.

Acknowledgments

This work was supported by the NOAA Cooperative Agreement with CIRES-University of Colorado, NA17OAR4320101. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official National Oceanic and Atmospheric Administration, National Aeronautics and Space Administration, or other U.S. Government position, policy, or decision.

Data Availability Statement

Data used in this study are publicly available through the National Oceanic and Atmospheric Administration's National Centers for Environmental Information GOES-R data portal <https://ngdc.noaa.gov/stp/satellite/goes-r.html>.

References

- Califf, S., Early, D., Grotenhuis, M., Loto'aniu, T. M., & Kronenwetter, J. (2020). Correcting the arcjet thruster disturbance in goes-16 magnetometer data. *Space Weather*, 18(1), e2019SW002347. <https://doi.org/10.1029/2019SW002347>
- Califf, S., Loto'aniu, T. M., Early, D., & Grotenhuis, M. (2020). Arcjet thruster influence on local magnetic field measurements from a geostationary satellite. *Journal of Spacecraft and Rockets*, 57(1), 177–186. <https://doi.org/10.2514/1.A34546>

- Camporeale, E., Cash, M. D., Singer, H. J., Balch, C. C., Huang, Z., & Toth, G. (2020). A Gray-Box model for a probabilistic estimate of regional ground magnetic perturbations: Enhancing the NOAA operational geospace model with machine learning. *Journal of Geophysical Research*, *125*(11), e27684. <https://doi.org/10.1029/2019JA027684>
- Cliver, E., Kamide, Y., & Ling, A. (2000). Mountains versus valleys: Semiannual variation of geomagnetic activity. *Journal of Geophysical Research*, *105*, 2413–2424. <https://doi.org/10.1029/1999ja900439>
- Díaz Baso, C. J., & Asensio Ramos, A. (2018). Enhancing SDO/HMI images using deep learning. *Astronomy & Astrophysics*, *614*, A5. <https://doi.org/10.1051/0004-6361/201731344>
- Díaz Baso, C. J., de la Cruz Rodríguez, J., & Danilovic, S. (2019). Solar image denoising with convolutional neural networks. *Astronomy & Astrophysics*, *629*, A99. <https://doi.org/10.1051/0004-6361/201936069>
- dos Santos, R. S. (2020). Estimating spatio-temporal air temperature in london (uk) using machine learning and earth observation satellite data. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102066. Retrieved from <http://www.sciencedirect.com/science/article/pii/S030324341831033X>
- Greff, K., Srivastava, R. K., Koutnik, J., Steunebrink, B. R., & Schmidhuber, J. (2017). Lstm: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Inceoglu, F., Jeppesen, J. H., Kongstad, P., Hernández Marcano, N. J., Jacobsen, R. H., & Karoff, C. (2018). Using machine learning methods to forecast if solar flares will be associated with CMEs and SEPs. *The Astrophysical Journal*, *861*(2), 128. <https://doi.org/10.3847/1538-4357/aac81e>
- Jarábek, T., Laurinec, P., & Lucká, M. (2017). Energy load forecast using s2s deep neural networks with k-shape clustering. In *2017 IEEE 14th International Scientific Conference on Informatics* (pp. 140–145). <https://doi.org/10.1109/INFORMATICS.2017.8327236>
- Jeppesen, J. H., Jacobsen, R. H., Inceoglu, F., & Toftegaard, T. S. (2019). A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, *229*, 247–259. <https://doi.org/10.1016/j.rse.2019.03.039>
- Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv e-prints, arXiv:1412.6980.
- Liu, H., Liu, C., Wang, J. T. L., & Wang, H. (2020). Predicting coronal mass ejections using SDO/HMI vector magnetic data products and recurrent neural networks. *The Astrophysical Journal*, *890*(1), 12. <https://doi.org/10.3847/1538-4357/ab6850>
- Loto'aniu, P. T., Califf, S., Redmon, R. J., & Singer, H. J. (2020). Chapter 21 - Magnetic field observations from the goes-r series. In S. J. Goodman, T. J. Schmit, J. Daniels, & R. J. Redmon (Eds.), *The goes-r series* (pp. 251–259). Elsevier. Retrieved from <https://www.sciencedirect.com/science/article/pii/B9780128143278000214>
- Loto'aniu, T. M., Inceoglu, F., Grotenhuis, M., Rich, F., & Califf, S. (2021). *Peer stakeholder-product validation review (PS-PVR) for the full maturity of GOES-16 and GOES-17 MAG L1b product*. Boulder, Colorado: NASA-NOAA GOES-R Program Technical Review. Retrieved from https://www.noaasis.noaa.gov/GOES/PS_PVR_GOES17.html#parentHorizontalTab7
- Loto'aniu, T. M., Redmon, R. J., Califf, S., Singer, H. J., Rowland, W., Macintyre, S., et al. (2019). The GOES-16 Spacecraft Science Magnetometer. *Space Science Reviews*, *215*(4), 32. <https://doi.org/10.1007/s11214-019-0600-3>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press. Retrieved from <https://projecteuclid.org/euclid.bsm/1200512992>
- Merity, S., Shirish Keskar, N., & Socher, R. (2017). *Regularizing and optimizing LSTM language models*. arXiv e-prints, arXiv:1708.02182.
- NASA-GSFC. (2019). *Goes-r series data book*. NASA GSFC Technical Memorandum. CDRL PM-14.
- Paparrizos, J., & Gravano, L. (2015). K-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD international conference on management of data* (pp. 1855–1870). Association for Computing Machinery. <https://doi.org/10.1145/2723372.2737793>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. Retrieved from <https://www.sciencedirect.com/science/article/pii/0377042787901257>
- Russell, C. T., & McPherron, R. L. (1973). Semiannual variation of geomagnetic activity. *Journal of Geophysical Research*, *78*(1), 92–108. <https://doi.org/10.1029/ja078i001p00092>
- Singer, H. J., Matheson, L., Grubb, R., Newman, A., & Bouwer, S. D. (1996). Monitoring space weather with the GOES magnetometers. In E. R. Washwell (Ed.), *Goes-8 and beyond* (Vol. 2812, pp. 299–308). International Society for Optics and Engineering.
- Stinis, P. (2019). *Enforcing constraints for time series prediction in supervised, unsupervised and reinforcement learning*. arXiv e-prints, arXiv:1905.07501.
- Tsyganenko, N. A. (2002a). A model of the near magnetosphere with a dawn-dusk asymmetry 1. Mathematical structure. *Journal of Geophysical Research*, *107*(A8), 1179–1181. <https://doi.org/10.1029/2001JA000219>
- Tsyganenko, N. A. (2002b). A model of the near magnetosphere with a dawn-dusk asymmetry 2. Parameterization and fitting to observations. *Journal of Geophysical Research*, *107*(A8), 1176–1181. <https://doi.org/10.1029/2001JA000220>
- Tsyganenko, N. A., Singer, H. J., & Kasper, J. C. (2003). Storm-time distortion of the inner magnetosphere: How severe can it get? *Journal of Geophysical Research*, *108*(A5), 1209. <https://doi.org/10.1029/2002JA009808>
- Tsyganenko, N. A., & Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *Journal of Geophysical Research*, *110*(A3), A03208. <https://doi.org/10.1029/2004JA010798>
- Yang, J., Ning, C., Deb, C., Zhang, F., Cheong, D., Lee, S. E., et al. (2017). k-shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy and Buildings*, *146*, 27–37. <https://doi.org/10.1016/j.enbuild.2017.03.071>