Physics-informed Neural Networks for the Improvement of Platform Magnetometer Measurements

Kevin Styp-Rekowski[1,2,3], **Ingo Michaelis**[2], **Monika Korte**[2], **Claudia Stolle**[4]

[1]Technical University of Berlin, Berlin, Germany
[2]GFZ Germany Research Center for Geosciences, Potsdam, Germany
[3]University of Rostock, Rostock, Germany
[4]Leibniz Institute for Atmospheric Physics at the University of Rostock, Kühlungsborn, Germany

Key Points:

- A calibration procedure of platform magnetometers from non-dedicated satellite missions is presented.
- A new calibration method using physics-informed neural networks with the Biot-Savart law is introduced.
- Datasets of space-based magnetic field measurements for the GOCE and GRACE-FO satellite missions are made publicly available.

Corresponding author: Kevin Styp-Rekowski, `styp@gfz-potsdam.de`

Abstract
Space-based measurements of the Earth's magnetic field with a good spatiotemporal coverage are needed to understand the complex system of our surrounding geomagnetic field. High-precision magnetic field satellite missions form the backbone for research, but they are limited in their coverage. Many satellites carry so-called platform magnetometers that are part of their attitude and orbit control systems. These can be re-calibrated by considering different behaviors of the satellite system, hence reducing their relatively high initial noise originating from their rough calibration. These platform magnetometer data obtained from non-dedicated satellite missions complement the high-precision data by additional coverage in space, time, and magnetic local times. In this work, we present an extension to our previous Machine Learning approach for the automatic in-situ calibration of platform magnetometers. We introduce a new physics-informed layer incorporating the Biot-Savart formula for dipoles that can efficiently correct artificial disturbances due to electric current-induced magnetic fields evoked by the satellite itself. We demonstrate how magnetic dipoles can be co-estimated in a neural network for the calibration of platform magnetometers and thus enhance the Machine Learning-based approach to follow known physical principles. Here we describe the derivation and assessment of re-calibrated datasets for two satellite missions, GOCE and GRACE-FO, which are made publicly available. We achieved a mean residual of about 7 nT and 4 nT for low- and mid-latitudes, respectively.

Plain Language Summary

This study revolves around enhancing our understanding of Earth's magnetic field by leveraging satellite data. While certain satellites provide highly detailed magnetic field information, their coverage is limited in geographical and temporal scope. Many satellites carry basic magnetic sensors as part of their control systems, but these sensors are initially rather inaccurate. We developed a computational method that combines machine learning and physics to refine these sensor readings. Our approach specifically addresses and corrects errors stemming from the satellite's own magnetic interference. We applied and tested this method on data from two specific satellites, namely GOCE and GRACE-FO. The improved magnetic field data resulting from our method is made publicly accessible, offering a more accurate and reliable dataset for researchers studying Earth's magnetic field.

1 Introduction

Platform magnetometers, commonly installed on various satellites in low Earth orbit, offer a promising means to expand the spatial and temporal coverage of Earth's magnetic field measurements from space. However, these instruments, initially not dedicated for geoscience applications, require calibration to ensure the scientific accuracy and usability of the collected data. To achieve this, gathering information about the satellite is essential to correct for artificial disturbances caused by other payload systems and other influencing properties associated with the satellite.

To quantify the Earth's magnetic field, several high-precision satellite missions have been operated. From 1999 to 2004, magnetic field data from the Ørsted mission (Neubert et al., 2001) are available. From 2000 to 2010, the CHAMP (Reigber et al., 2002) satellite mission was in orbit, followed by a gap from 2010 to 2013 where no high-precision mission measured the magnetic field. Since 2013, the Swarm constellation (Friis-Christensen et al., 2006; Olsen et al., 2013) provides again high-precision measurements. Recently, there have been ongoing efforts to complement the high-precision missions with additional data from non-high-precision or platform magnetometers with an analytical approach to the calibration. In this way, data from the GRACE, CryoSat-2, DMSP, GRACE-FO, Swarm-Echo, AMPERE, and GOCE missions have been calibrated and made pub-

licly available (Olsen, 2021; Olsen et al., 2020; Alken et al., 2020; Broadfoot et al., 2022; Anderson et al., 2000; Stolle, Michaelis, et al., 2021; Michaelis et al., 2022). Although having slightly higher noise these data complement dedicated magnetic field mission data well for understanding the Earth's magnetic field variations. They enlarge the spatiotemporal distribution, e.g., providing enhanced coverage of the altitudes or local times of measured phenomena mainly of magnetospheric or ionospheric origin. Section 2 provides a brief overview of the geomagnetic field. Subsequent work has shown the analytical potential of these additional data (e.g., Stolle, Olsen, et al. (2021); Xiong et al. (2021); Park et al. (2020)).

In earlier works, we demonstrated the effectiveness of Machine Learning-based methods for the calibration of platform magnetometers (Styp-Rekowski et al., 2021, 2022b; Bader et al., 2022). Leveraging Machine Learning (ML) techniques, we can adapt the magnetometer signal to compensate for artificial disturbances originating from the payload of the satellite. Our proposed non-linear regression approach automates the identification of relevant features and their interactions, broadening the range of inputs that can be utilized. This reduces the analytical work required for the calibration of platform magnetometers, resulting in faster, more precise, and easily accessible magnetic datasets derived from non-dedicated satellite missions. These calibrated datasets are made publicly available, promoting broader scientific access and utilization (Styp-Rekowski et al., 2022a, 2023).

In this work, we propose an extension for the known approach by incorporating the physical Biot-Savart law into a neural network (NN), which results in a physics-informed neural network (PINN). This improves the modeling and correction of the impact of electric current-induced artificial magnetic fields on the satellite's magnetic measurements, as the PINN is more constrained to follow first-principle physical laws. In addition, the B-field estimates of the Average Magnetic field and Polar current System (AMPS) model (Laundal et al., 2018) are combined with the B-field estimates of the CHAOS-7 model (Finlay et al., 2020), improving the reference model of the calibration, especially for the polar regions. This extended approach is applied to the GOCE (Floberghagen et al., 2011; Drinkwater et al., 2003) and GRACE-FO (Kornfeld et al., 2019) satellite missions and their respective measurements. In the future, the proposed approach can be applied to a wider variety of satellites to improve the accuracy of their platform magnetometer measurements. We hope to enable other satellite operators to calibrate their magnetic instruments, improve the quality of their data, and make additional data available to the scientific community.

In classic, on-ground calibration, a Helmholtz cage is used to determine the response of the magnetic field instrument to different applied magnetic fields considering different satellite states and the response of the instrument under calibration (Friis-Christensen et al., 2006). Recently, Springmann et al. (2010) described the satellite noise signals by dipoles, in terms of location, orientation, and strength, by employing multiple magnetometers in a research facility on-ground and a least-squares minimization. In this work, we present an approach to determine a similar estimation of dipoles in-flight for single magnetometers on board a satellite through data assimilation. Due to the availability of a large set of electric current strengths and measured magnetic fields, the dipoles are estimated as part of a larger optimization problem. The satellites in this study carry their magnetometers at a single position, which makes the localization more inaccurate. However, this configuration offers the advantage of a large amount of data, encompassing various activation strengths. Consequently, numerous data points linking electric currents with their induced magnetic fields are available for analysis.

Physics-informed neural networks (PINN) represent a powerful combination of traditional physics-based modeling and ML-based techniques (Cuomo et al., 2022). These networks are designed to incorporate physical principles into their architecture, allowing them to leverage data-driven insights and fundamental laws of physics. This inno-

vative method has been applied in numerous fields, from fluid dynamics and materials science to geophysics and beyond, making it a promising tool for complex physical systems (Yuan et al., 2020).

We evaluate our approach on multiple levels: First, a residual analysis comparing the calibration quality to our reference model is conducted, followed by a feasibility study assessing the ability of the trained PINN models to calibrate out-of-sample data of subsequent months. This shows how the calibration could be adapted for a near real-time application. We also show how disturbance sources can be extracted and analyzed utilizing the proposed physics-informed approach, followed by an analysis of magnetic phenomena, namely field-aligned currents (FAC) and magnetic storms, using the calibrated magnetometer data.

The remainder of the paper is structured as follows. First, Section 2 gives an introduction to Earth's magnetic field, followed by Section 3 with an overview of the data used within the presented methodology. Section 4 introduces the proposed approach, presenting the improvements to the ML-based calibration. The results of the proposed approach are described in Section 5, while section 6 summarizes our findings.

2 Background

This chapter provides a brief overview of Earth's magnetic field including its sources, structure, and phenomena.

The geomagnetic field originates from Earth's molten, electrically conducting outer core, primarily composed of iron and nickel (Lowrie, 2023). The heat in the core drives electric currents through thermal convection, and influenced by the rotation of the Earth, results in the geodynamo mechanism. At Earth's surface, this self-sustaining process establishes a dipole-dominated magnetic field. The field surrounding Earth is further affected by external sources in the ionosphere, magnetosphere, and solar wind (Prölss, 2012).

Interactions between Earth's magnetic field and the interplanetary magnetic field (IMF), which is carried along with the solar wind, lead to various geomagnetic phenomena, such as the formation of the magnetosphere the magnetospheric ring current, and field-aligned currents (FACs) in polar regions, while the charged particles of the upper atmosphere in the presence of the magnetic field form ionospheric currents, such as the electrojet at the equator and the solar quiet daily magnetic variation (Sq), (Olsen & Stolle, 2012). Geomagnetic indices like the Kp (Matzka et al., 2021), Dst (Sugiura, 1964), or Hp30-indices (Yamazaki et al., 2022) are indicators of the magnetic activity of these interactions manifested as short-lived disturbances (of minutes to days) of the geomagnetic field.

Various data including ground-based observatories, air and marine magnetic campaigns, and space-based measurements contribute to empirical modeling of Earth's magnetic field and its variations. While ground observatories offer continuous long-term observations, satellites provide global coverage, including remote areas and oceans. High-precision data from missions like Ørsted, CHAMP, and Swarm have largely improved geomagnetic field models. An overview on current capabilities to describe the geomagnetic field are given in Stolle et al. (2017, and references therein).

3 Data

The newly derived calibration method we propose is applied to two satellite missions, namely the GOCE mission (from 2009 to 2013) and the GRACE-FO mission (from 2018 and ongoing).

3.1 Satellite Data

The data collection process is very similar to previous works by Styp-Rekowski et al. (2022b) for the GOCE satellite and by Styp-Rekowski et al. (2021) for the GRACE-FO satellite duo. To summarize, all available data for the satellites are used, which means any publicly available measurements taken onboard the satellites. Therefore, an interpolation onto common timestamps is needed, which was set to be the timestamps of the position and attitude determination of the satellites. There has already been extensive preprocessing of features in the form of One-hot encoding for textual features, removal of non-informative features, addition of external features, scaling of the features to an interval of $[-1, 1]$, automatic filling of missing values, removal of outliers, and magnetic quiet time filtering. For more details, please refer to the previous publications.

The GOCE mission has been completed in 2013. Therefore, we have a completed dataset with no new measurements. It sums up to about 6.4 million data points with 984 features used within our calibration model. For the GRACE-FO mission, data have been collected since June 2018 at a rate of 1Hz for the magnetic measurements. Together with this publication, the datasets until July 2023 will be published, which means about 162 million data points with about 71 features.

The previous data collection process has been partly changed and extended. For both satellite missions, the magnetic quiet time filtering based on the *Kp* (Matzka et al., 2021) and *Dst* (Sugiura, 1964) indices has been replaced by a new filtering based on the Hp30 (Yamazaki et al., 2022) and *Dst* indices. The increased resolution of a 30-minute resolution, as compared to the previous 3-hour resolution, provides enhanced control over the selection of filtered data points. As the Hp30 has been set up to contain the same magnitudes as the *Kp* index, filtering for data points with Hp30 ≤ 2 has been applied. The Dst-based filtering was changed to utilize the rate of change with $|d(Dst)/dt| < 4nT/hr$ to filter the data points.

In addition, previous data have been flagged if the magnetorquer control currents were applied at their maximum value, which was interpreted as the satellite being out of attitude. This has been extended to mark the whole time span of such occurrences in what is described as a center-of-mass calibration maneuver, steering the magnetorquers with the maximum and minimum control currents alternatingly (Cossavella et al., 2022). Such identified data is ignored during the training but added in the final calibrated dataset with a corresponding flag.

In preparation for the PINN, it is necessary to identify the features that represent electric currents onboard the satellite. For the GOCE satellite, these features have been determined using a publicly available feature description list, which specifies the physical units of the measurements, such as 'A' and 'mA'. In the case of the GRACE-FO satellite mission, features were identified based on their names, specifically by including the keyword "current" in their names. These features are subsequently standardized to a common unit of Ampère and undergo no additional scaling during the feature preprocessing stage.

3.2 Reference Model

As a reference model for the calibration, the CHAOS-7.15 model has been used, which is based on ground observatories as well as space-based measurements of the magnetic field, e.g., from nearly ten years worth of Swarm data (Finlay et al., 2020). The model has been evaluated at each satellite position to be calibrated and then rotated into the satellite frame. This reference model will be combined with the AMPS model by combining their respective B-field estimates into the finally used combined reference model.

Laundal et al. (2018) introduced the AMPS model that describes the large-scale structure of the current system in the polar regions, also known as FACs. The model is based on several space weather parameters that are needed as input and can be used to calculate the present magnetic field at a certain position and time, whereby the magnetic field is based only on these external phenomena that are not part of the CHAOS-7 model. Thus, this model is well-suited to be used in conjunction with the CHAOS-7 model to enhance the quality of the used reference model. Recently, Kloss et al. (2023) have presented a similar idea to co-estimate the core field with the ionospheric field while creating models of the geomagnetic field, effectively combining the AMPS with the CHAOS-7 model during the creation of a new model. Here, we combine the AMPS and CHAOS-7 model by combining the magnetic field estimates.

For the introduction of the AMPS model, additional space weather properties are needed for its parametrization. Therefore, B_y, B_z of the interplanetary magnetic field (IMF), solar wind speed V_{sw}, the solar flux index $F10.7$, and the dipole tilt angle of the Earth *tilt* have been added to the set of features that are only used for calculating the reference model values.

Here, 1-min data available for the space weather parameters, provided by the National Aeronautics and Space Administration (NASA)(Papitashvili & King, 2020), have been used where missing data are interpolated as the gaps are small and changes are expected continuously. As recommended by Laundal et al. (2018), the space weather parameters are smoothed with a rolling 20-minute window mean before input into the AMPS model. In addition, as the model is parameterized for the polar regions, only values polewards of 40° quasi-dipole latitude (QDLat) have been used, otherwise set to 0. Within the published datasets, the magnetic fields of the AMPS model are provided separately from the CHAOS-7 data so the calibrated data can be investigated with either of the reference models. For the combination of the CHAOS-7 and the AMPS model, the evaluated B field estimates are added and subsequently used as the reference model.

A larger error of the calibration results is expected at high latitudes than at mid and low latitudes due to increased ionospheric disturbances. The discrepancy can be partly mitigated since

the AMPS model contains the mean large-scale structures, whereas the satellite likewise measures large-scale and small-scale features of the FACs. Thus, the inclusion of the AMPS model improves the calibration process as the values provided by the combined model cover the mean variation at high latitudes.

4 Physics-informed Calibration

Traditionally, platform magnetometer data have received a careful calibration where features were preselected and new features created based on the experience of the domain expert (Olsen et al., 2020; Michaelis et al., 2022). These calibrations are well comprehensible and are excellent in extrapolation or generalization tasks. Similar approaches have been applied to a variety of satellites, either calibrating the instrument directly or co-estimating the instrument calibration with a model calculation (Alken et al., 2020). Recently, a calibration utilizing ML tools has been proposed with several advantages, like the possibility to use all available measurements on the satellite as the ML approach can select the relevant features for the calibration itself (Styp-Rekowski et al., 2021). In addition, non-linear interactions between measurements and the reference model are automatically found, thus yielding the possibility for an advanced calibration achieving lower remaining noise in the produced dataset.

4.1 Methods and Concepts

As presented in Styp-Rekowski et al. (2022b), there have been shortcomings in the ML approach. First, the calibration underestimates the FACs in the calibrated datasets, rendering the datasets less valuable for studying geomagnetic phenomena in high-latitude regions. Second, the previous ML model is only analyzable by techniques like Shapley Additive Explanations (SHAP) (Lundberg & Lee, 2017), allowing a view into the inner workings of a black box. In addition, the generalizability of the previous models, e.g., when studying month-to-month models, could be improved. Within this work, we propose an extended approach to tackle these shortcomings.

The modeling of FAC in high-latitude regions has been found to be associated with the distinct characteristics of the specific satellite mission's orbit, i.e., GOCE. The combination of the sun-synchronous polar orbit and magnetic local times (MLT) of about 6 and 18 LT for the descending and ascending orbit, respectively, the tilt angle of the Earth towards the sun, and the Earth's rotation in combination lead to the phenomenon of the satellite flying through a larger area of shade around the south pole for the period around June solstice. The spatially limited shading of the satellite leads to a certain uniqueness of the housekeeping parameters in this area. On the other hand, the FACs also appear in this area and are not part of the CHAOS-7 reference model. This fact leads to the previously purely statistically driven NN to correlate the unique property of the housekeeping data with the unique signal property of the FACs in this region, thus optimizing for this correlation, which means that the housekeeping data are 'misused' to correct for the FAC signals present in the measured data.

Therefore, two extensive further developments of the previous approach have been developed to tackle this problem. First, as described in Section 3.2, the AMPS model by Laundal et al. (2018) is incorporated into our reference model, allowing us to include the average large-scale features of the FAC regions in our calibration for a better calibration result. Secondly, we will incorporate known laws from physics into the NN optimization, ensuring that the calibration will be physically more correct instead of purely statistics-driven.

4.2 Physics-informed Component

One main contributor to the artificial disturbances relevant to the calibration of platform magnetometers is known to be electric currents flowing within the electronics system of a satellite. The disturbance mainly originates in the induced magnetic fields caused by a flowing electric current, which influences the measurements of the magnetometers.

For this well-known phenomenon, the Biot-Savart Law gives a 3-dimensional representation of Ampère's Law. As the magnetometer measurements originate from a 3d fluxgate magnetometer, the Biot-Savart law is the best fit to describe these processes in the satellite that couple the electric currents with induced disturbances in the form of magnetic fields. In a closed and compact system like a satellite, every circuit should be considered and modeled as closed. Therefore, the search focuses on approximations for dipoles of planar magnetic field-inducing coils. The Biot-Savart law for magnetic dipoles is given in its vector form in Equation (1),

$$\overrightarrow{B}(\overrightarrow{r}, \overrightarrow{m}) \;=\; \frac{\mu_0}{4\pi}\left(\frac{3\hat{r}(\hat{r} * \overrightarrow{m}) - \overrightarrow{m}}{|\overrightarrow{r}|^3}\right) \tag{1}$$

with μ_0 as the permeability of free space,

$$\mu_0 \;=\; 4\pi 10^{-7}\frac{Vs}{Am} \tag{2}$$

$\overrightarrow{r}$ as the position at which the magnetic field is calculated, $\hat{r}$ as the unit vector in the direction of $\overrightarrow{r}$ as defined by

$$\hat{r} = \frac{\overrightarrow{r}}{|\overrightarrow{r}|} \tag{3}$$

and the magnetic dipole moment $\overrightarrow{m}$, used in a similar way like by Springmann et al. (2010) and defined by Jackson (1999). The magnetic dipole moment $\overrightarrow{m}$, as defined in Equation (4),

$$\overrightarrow{m} = IN\overrightarrow{a} \tag{4}$$

is the product of the electric current I, the area orthogonal vector $\overrightarrow{a}$ which is perpendicular to the inner area of the coil and has the magnitude of the area itself, and the number of coil windings N. When substituting Equation (3) into Equation (1) and some reformatting, we arrive at our finally used formula in Equation (5).

$$\overrightarrow{B}(\overrightarrow{r}, \overrightarrow{m}) = \frac{\mu_0}{4\pi}\left(\frac{3\overrightarrow{r}(\overrightarrow{m} * \overrightarrow{r})}{|\overrightarrow{r}|^5} - \frac{\overrightarrow{m}}{|\overrightarrow{r}|^3}\right) \tag{5}$$

This formula gives the induced magnetic field of a magnetic dipole generated by a coil. There are some assumptions when using this formula to approximate the dipoles of inducing electric currents within the satellite system. The main assumption is that the complex induced magnetic fields onboard the satellite system are expected to be approximated by one dipole per sub-system. For this simple form of the equation, the center of the coil is set to be at the origin while the position of the induced magnetic field is calculated at the position $\overrightarrow{r}$, which in our case corresponds to the position of the magnetometer relative to the center of the coil. As in the calibration scenario of the satellites, the only given parameter in this equation is the electric current I, which is part of the housekeeping datasets delivered together with the magnetometer data by the satellite operators, the 3-dimensional position $\overrightarrow{r}$ and the area vector $\overrightarrow{a}$ as well as the coil windings parameter N need to be estimated by the optimization function. This will be done separately for every available electric current in the housekeeping data. We omit the winding parameter N in our estimation and set it to 1, with the goal of fewer parameters to estimate. If a coil has more than one winding, the optimization would directly factor this property into the area vector since it acts as a factor for the dipole moment term. In addition, another assumption is to use the permeability of free space, which is a good estimator within a satellite system.

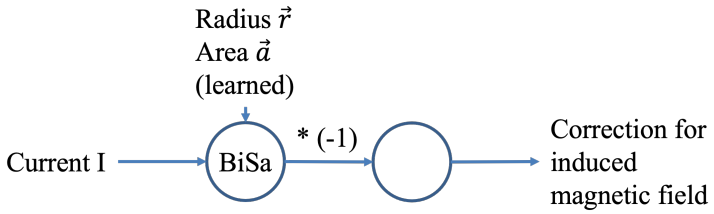

Figure 1. BiSa (Biot-Savart) layer with the electric current I as the input, learnable parameters $\overrightarrow{r}$ and $\overrightarrow{a}$ and the correction for the induced magnetic field.

For the satellite calibration within a PINN, this means that the input to the Biot-Savart layer is the electric current of a particular subsystem of the satellite, as can be seen in Figure 1 which shows the Biot-Savart layer schematically. This means that for every electric current present in the data, a separate Biot-Savart layer is instantiated with

random vectors $\overrightarrow{r}$ and $\overrightarrow{a}$. After calculating the result of Equation (5), which corresponds to the induced magnetic field, the field is inverted to correct for the artificial disturbance introduced by the electric current being present. During the gradient-based optimization of the PINN, the learnable parameters $\overrightarrow{r}$ and $\overrightarrow{a}$ will be changed to match the measured behavior of the satellite as a system. If an electric current present in our dataset is not relevant for the measurements of the magnetometer at hand, the $\overrightarrow{a}$ will become very small, and the position of the coil $\overrightarrow{r}$ will become very large relative to the magnetometer, thus neglecting the influence of this electric current.

This means that the calibration becomes physically more constrained as the producing source for the induced magnetic fields is restricted to follow the Biot-Savart law. In a feed-forward neural network, an arbitrary non-linear function with no constraints would estimate the induced magnetic field of an electric current. With the PINN methodology, this function is limited to follow the Biot-Savart law, e.g., a non-present electric current of 0A cannot induce any magnetic field.

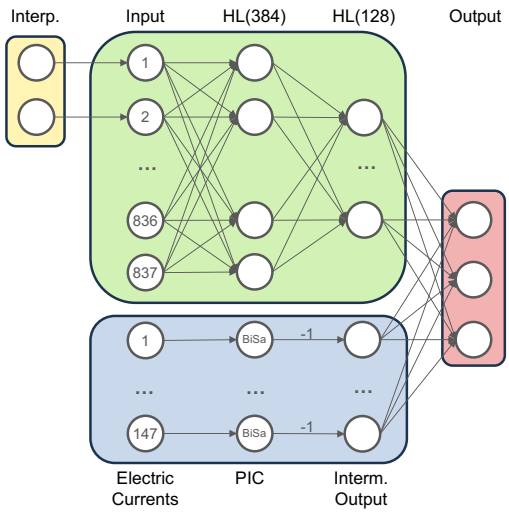

Figure 2. Architecture of the PINN calibration model for the GOCE satellite: In yellow is the time shift interpolation, in green is the FFNN, in blue is the PIC, and in red is the 3-dimensional output. The final architecture was found through a hyperparameter optimization.

Figure 2 shows an illustration of the architecture of the calibration model for the GOCE satellite. The architecture for the GRACE-FO is similar, with different dimensions for the input. The number and size of the hidden layers were determined with a hyperparameter optimization. The electric currents identified within the feature set are excluded from the feed-forward neural network (FFNN), illustrated in green, and instead form the new physics-informed component (PIC), shown in blue. Within the PIC, the electric currents are translated into induced magnetic fields by the BiSa neurons and then inverted to be corrected within the last concatenation step. Finally, the 3-dimensional output layer, representing the X-, Y-, and Z-axes, illustrated in red, combines the intermediate outputs from the FFNN and the PIC and thus generates the final PINN calibration model.

During the network training, the gradient of the model predictions is compared to our reference model, from which the weights and biases of the network are adjusted according to the gradient. For the PINN, this means that the location, encoded in $\overrightarrow{r}$, and the magnetic moment, encoded in $\overrightarrow{a}$, are adjusted by translating the magnetic field-related gradient into a position-related gradient, effectively adapting the relative position to fit

the data. In addition, the learned parameters $\overrightarrow{r}$ and $\overrightarrow{a}$ are accessible in the trained model, giving the possibility for further analysis, as will be shown in Section 5.

With the GOCE mission providing data at 16s intervals and the GRACE-FO mission providing data at 1s intervals, the model training needs to be structured. Therefore, for the GOCE mission, a global model spanning the whole mission duration (2009 to 2013), and for the GRACE-FO mission, yearly models as defined by the calendar are trained. This ensures good coverage of different behaviors within the satellite as a system while, on the other hand, not exceeding computational limits in terms of memory and computation time. Afterward, the global or yearly models, respectively, are trained on a per-month basis with a much lower learning rate of $1{*}10^{-5}$. Monthly data ensure enough data points for this finetuning step while enabling the models to adjust for small perturbations in the calibration over time. In addition, L1-regularization has been added with a regularization parameter of $1{*}10^{-3}$ (Ng, 2004). This penalizes high weights during the training of the neural network and makes it more sparse to focus on the important input features.

5 Evaluation

The proposed approach is evaluated on two satellite missions, namely the GOCE and GRACE-FO satellite missions, consisting of one and two satellites, respectively. The following subchapters show different aspects of our evaluation for either one or both missions. As different housekeeping data are available, the utilized model for the two satellite missions looks slightly different. For the GOCE satellite, as seen in Figure 2, 837 of the total 984 features are input features into the FFNN part of the PINN, while 147 features represent electric currents onboard the satellite and are fed into the PIC of the PINN. For the GRACE-FO satellite, there are nine features representing currents that are fed into the PIC of the PINN, while the other features are used in the FFNN part of the PINN. These are about 71 features, depending on the filtering applied on a yearly basis. As described in Styp-Rekowski et al. (2021) and re-evaluated in this study, the magnetometer readings are shifted in time by 0.67 s for the GRACE-FO1 and GRACE-FO2 satellites, while the magnetorquers are shifted by -0.33 s.

5.1 Residual Evaluation

In the following analysis, we restrict the data to low- and mid-latitudes and apply the filtering provided by the B_FLAG of the dataset to restrict the evaluation to non-erroneous data as well as a Hp30 ≤ 2 and $|d(Dst)/dt| < 4nT/hr$ filtering for magnetic quiet times.

Satellite	Mean absolute error	Standard deviation
GOCE	6.56	9.66
GRACE-FO1	3.57	5.13
GRACE-FO2	3.82	5.23

Table 1. Mean absolute error and standard deviation of the calibrated data for the different satellites, spanning the whole calibration time range.

The residual is calculated on a per-point basis between the calibrated data and the reference model estimates. Looking at the residual for the GOCE mission, calculated on a per-month basis and averaged over all months, a mean absolute error (MAE) of about 6.56 nT with a standard deviation (SD) of 9.66 nT was achieved, as can be seen in Table 1. For the GRACE-FO1 satellite, an MAE of about 3.57 nT with an SD of 5.13 nT, and for the GRACE-FO2, an MAE of 3.82 nT with an SD of 5.23 nT was achieved. This

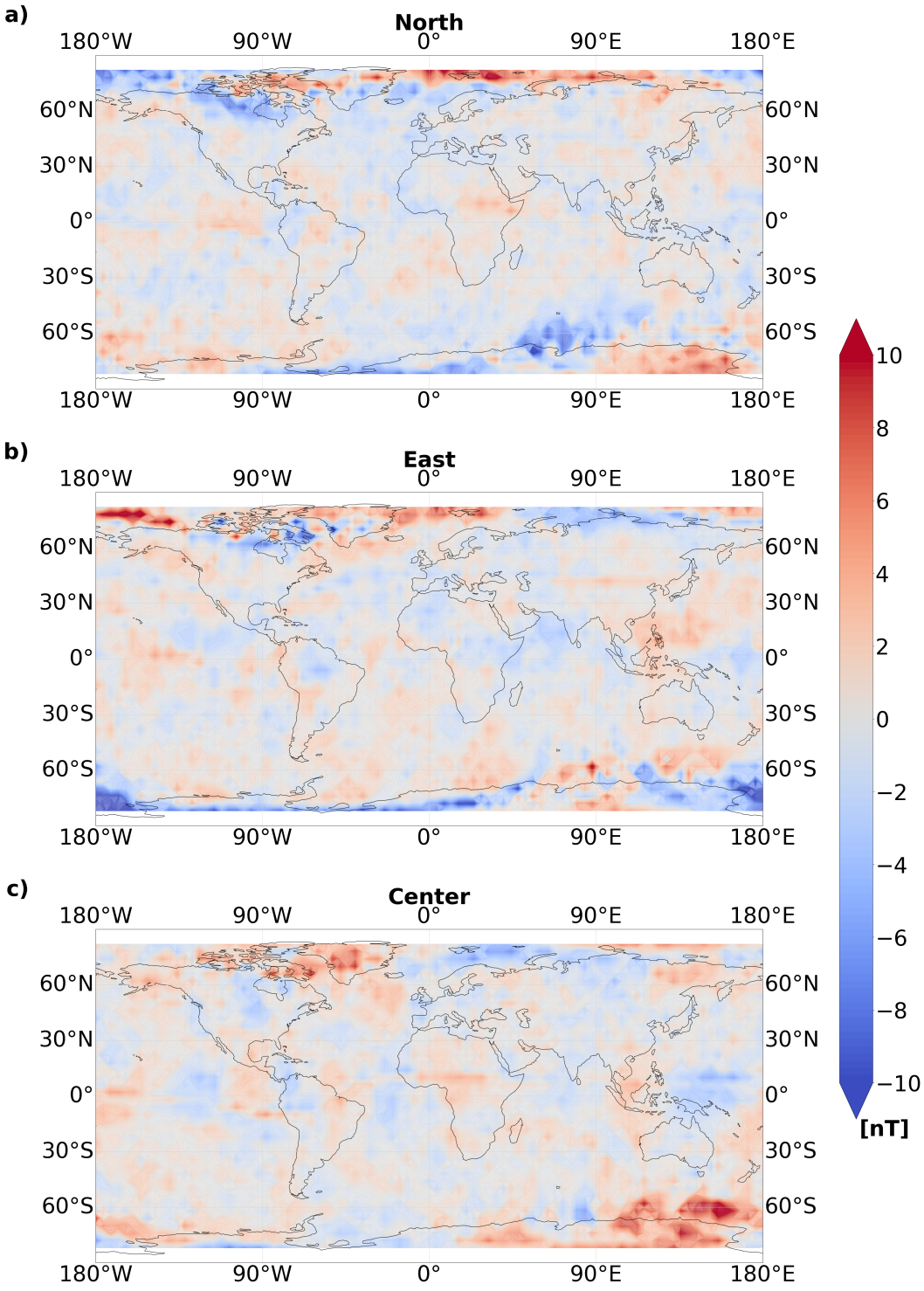

Figure 3. Map of the binned and averaged residuals between the PINN calibration and the combined reference model in the NEC-frame on a scale of 10 nT with GOCE data for the year 2013, panel a) shows the North component, panel b) the East component, and panel c) the Center component. A grey color indicates a residual of close to 0. Note the color scale maximum and minimum values contain saturated values.

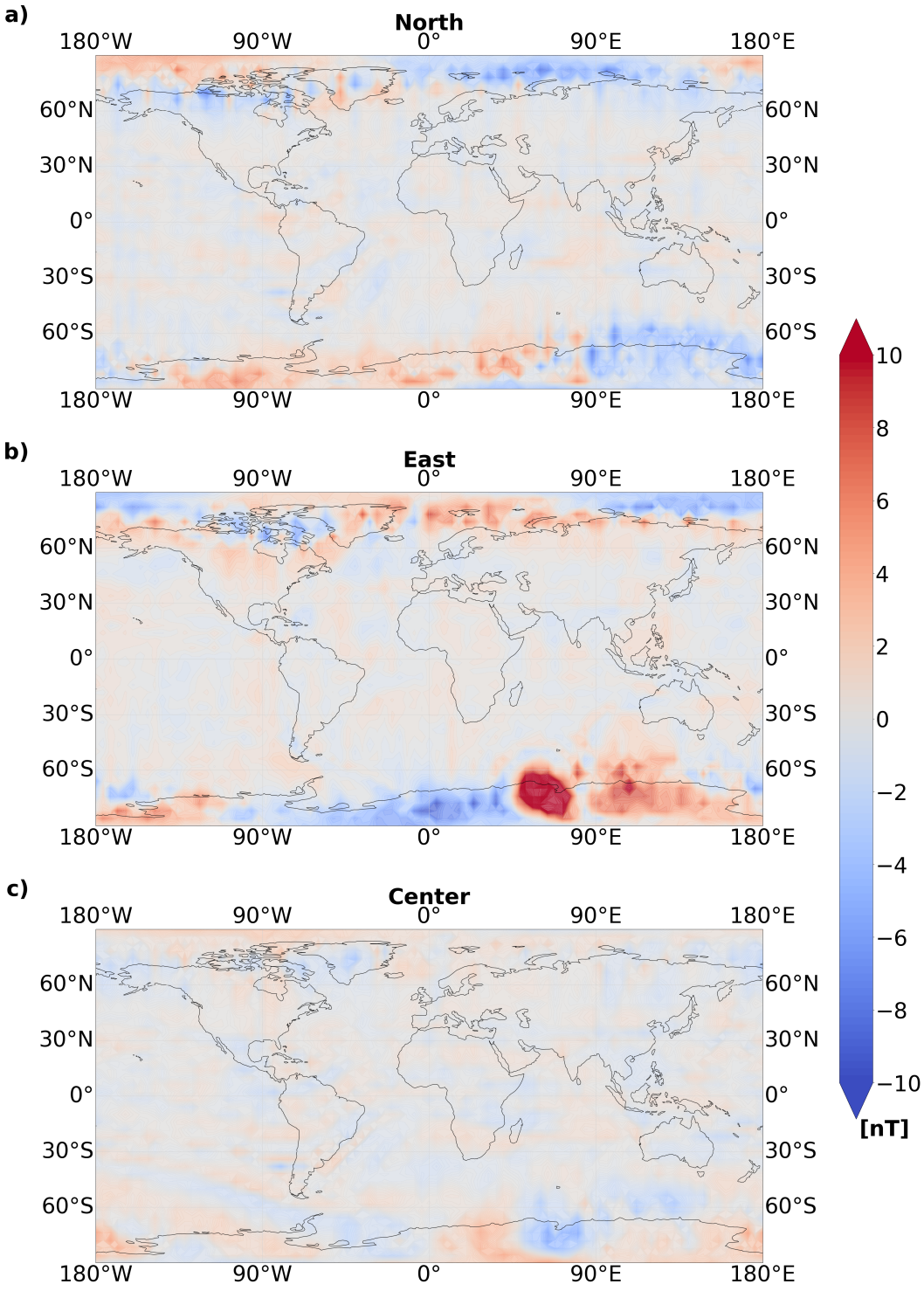

Figure 4. Map of the binned and averaged residuals between the PINN calibration and the combined reference model in the NEC-frame on a scale of 10 nT with GRACE-FO1 data for the year 2019, panel a) shows the North component, panel b) the East component, and panel c) the Center component. A grey color indicates a residual of close to 0. Note the color scale maximum and minimum values contain saturated values.

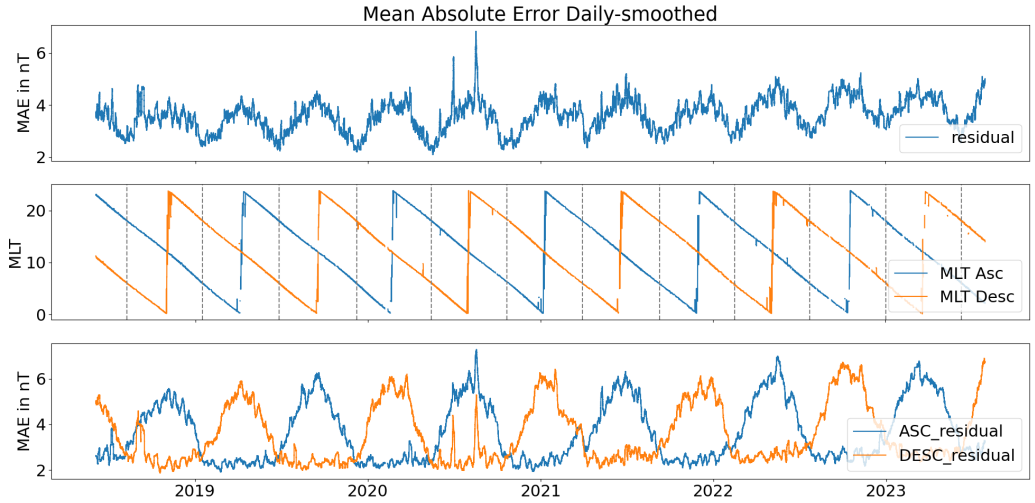

Figure 5. First row: Residual over whole calibration period for GRACE-FO1 with 86400s (daily) smoothing applied. Second row: MLT changes for the ascending and descending orbit of the mission. Dawn-dusk orbits marked with grey dotted lines for times of low residual. Third row: Same as the first row, with a separation for ascending descending orbits.

lies well within the margin to enable a scientific application of the proposed calibration data. In the previously calibrated dataset of the GOCE satellite (Styp-Rekowski et al., 2022b), an MAE of 6.47 nT, of the GRACE-FO1 satellite (Styp-Rekowski et al., 2021), an MAE of about 2.96 nT and of the GRACE-FO2 satellite an MAE of 3.51 nT was achieved. When comparing the same period as for the previous calibration of the GRACE-FO mission and GOCE mission, the new calibration has a MAE that is 0.4 nT larger than previously, which is due to the fact that the proposed calibration method is constraining the model more in the usage of the additional features. A comparison with data from the Swarm satellites is conducted in Section 5.6.

Exemplary, for the GOCE and GRACE-FO1 satellite, the residual data for the years 2013 and 2019, respectively, have been plotted on a map of the Earth where the residual has been averaged in bins of 4° latitude by 4° longitude for the contour plot in the NEC-frame. Figure 3 and Figure 4 show the result of this evaluation. With the scale given on the right and the three components North, East, and Center, it can be seen that the residual has no apparent local distribution and is overall close to 0. The high-latitude areas show that the satellites measure actual data of the polar current sheets, which in average slightly differs from the given AMPS model used within the reference model.

By now, over five years of mission data have been acquired for the GRACE-FO satellite mission. This allows us to examine the behavior of the residual over a long period, spanning multiple seasons, multiple passages of the same MLTs for the satellite mission, and different levels of the solar cycle. The mean absolute residual over the whole calibrated data of the GRACE-FO1 satellite has been plotted in Figure 5, with daily smoothing applied. The data have been selected for the low- and mid-latitude range with a filter for the B_FLAG of the calibrated dataset applied and selected for magnetic quiet times using the previously defined filter. This gives an overview of how well the calibration performs compared to our reference model over time. In general, there is a reappearing seasonality with a periodicity of about 11 months in the first panel, visible in a low residual of about 2.5 nT every 5.5 months. The second row of the plot shows how the GRACE-FO1 satellite drifts through different MLTs throughout the mission. There

is a relationship between the residual and the MLT of the mission, as the satellite repeats its MLT drift every 11 months, more precisely 320 days. The drops in the residual correlate with the dawn-dusk orbits visited by the satellite mission, where orbits with an MLT of 18 and 6 are marked with grey dotted lines.

Dawn, dusk and the local night time in between are the time when the least influence of magnetic dayside phenomena is present. Therefore, we see the lowest residuals for dawn, dusk and night side orbits. This result is also due to the fact that the CHAOS-7 combined with AMPS reference model, does include only averaged ionospheric currents and does not capture fast, small scale variations. The third panel shows the ascending and descending residuals separately, e.g., the residual peaks are high around times when the orbit goes through a noon MLT while maintaining low residuals on the nightside. Thus, a 2-3 nT residual can be maintained if only nighttime orbits are selected. Still, daytime data should be included in our calibration and evaluation since dayside phenomena are of interest to the geoscientific community. This could, e.g., include dayside Sq variations caused by electric currents in the ionosphere which are not modelled by the CHAOS-7 model (Finlay et al., 2020), effects of the equatorial electron jet (EEJ), or a more complex behavior of the satellite system not modeled by the calibration.

5.2 Predictive Calibration

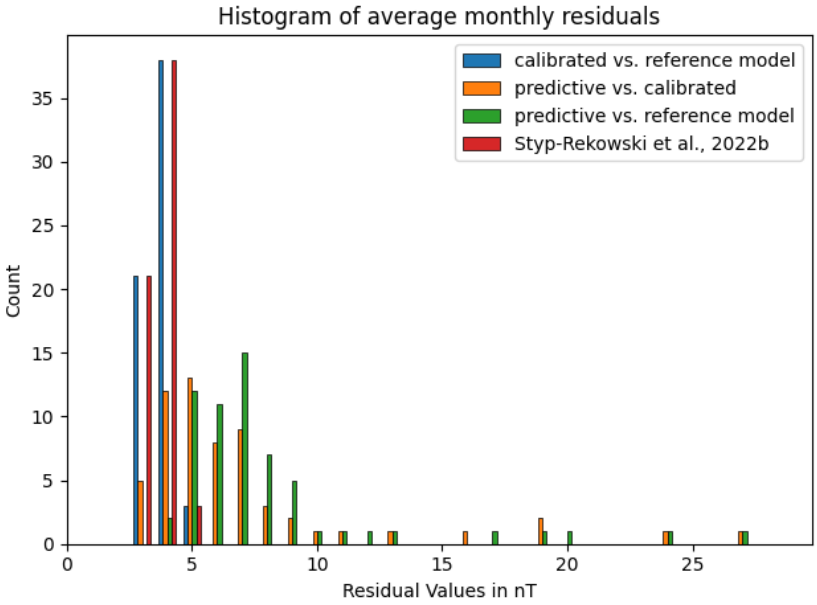

Figure 6. Histogram plot of monthly average residuals for predictive calibration models evaluated on subsequent monthly data for the whole mission data of the GRACE-FO1 satellite. Additionally, the direct calibration is included and compared against the reference model as well as the predictive calibration. Finally, the results are compared to the previous calibration by Styp-Rekowski et al. (2022b).

Additionally, a feasibility evaluation was conducted utilizing the different monthly models of the GRACE-FO1 satellite mission. For every month of the satellite mission, the calibration model that was trained on the data of the previous month has been used to calibrate the data of the current month. This is done to show how the calibrated models perform on data of unseen months, which also carry a slight shift in MLT and yearly

	X	Y	Z
MTQ1	18.43	-2.39	-1.31
MTQ2	1.57	15.33	0.68
MTQ3	0.63	-0.99	18.54

Table 2. Maximum magnetic dipole moment (in Am^2) as calculated by the maximum applied electric currents of the three magnetorquers (MTQ) and derived by the learned parameters from the PINN model.

seasonality. Furthermore, this is a building block towards near real-time usage of the proposed calibration methods as this enables the use of precomputed models to quickly assess newly measured satellite data without the need for immediate training. The model coefficients of the reference models change on a larger time-scale of several months to years, while the availability of their input parameters lies in the range of hours to days, enabling a near real-time usage in the future.

In Figure 6 a histogram of average monthly residuals is shown for different configurations. First, the calibrated dataset is compared with the reference model, where most values lie within a range of 3 to 5 nT, given in blue. This is comparable to the previous method by Styp-Rekowski et al. (2022b), given in red. In addition, the predictive calibration is compared directly with the calibrated data in orange, as well as with the reference model, given in green. Comparing the predictive calibration with the reference model, the residual values lie within a range of about 4 nT to 10 nT, also reflected by a median value of about 6.7 nT. There are some outliers that mainly originate from the unseen behavior of the supporting features within the calibration model. Generally, the residual values are larger compared to the direct monthly calibration but remain reasonable within one order of magnitude. Comparing the predictive with the direct calibration, a generally lower residual can be seen as the calibration follow the same restrictions of the models.

For the potential continuous processing of GRACE-FO satellite data, two strategies could be deployed. A rolling yearly model could be pre-trained and applied to newly arriving monthly data, or the current yearly model could be extended by new monthly data and then fine-tuned for the latest monthly data.

5.3 Predicted Dipole Locations

As stated previously, the usage of the Biot-Savart formula within the PIC of the neural network model enables the extraction of the learned disturbance locations $\overrightarrow{r}$ and their dipole moment $\overrightarrow{m}$ if the electric current is combined with the area orthogonal vector. Therefore, the maximum control currents for the magnetorquers have been identified over the whole time series, which are about 0.109 Ampère for every magnetorquer. With these values, the induced magnetic field for the magnetometer position can be calculated. As the Biot-Savart formula, viewed as an equations system, is underdetermined, a variety of possible $\overrightarrow{r}$ and $\overrightarrow{a}$ solutions are possible to achieve the same induced magnetic field at only one point within the satellite. Indeed, infinitely many solutions exist. Therefore, the induced magnetic field and the known magnetorquer positions are taken together with the maximum current and inserted into Equation (5), which results in the dipole moment generated by the maximum magnetorquer control currents. With three given variables, the equation becomes a solvable linear equations system. Thus, the dipole moment can be extracted.

The results of this analysis can be seen in Figure 7. Here, a 3D model of the GRACE-FO satellite has been rendered to visualize the results. The front of the panel, defined

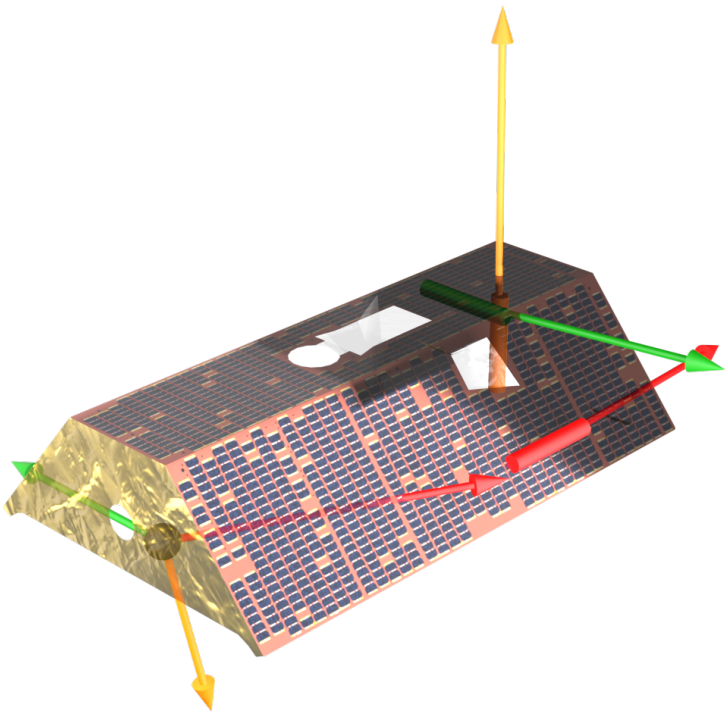

Figure 7. Model of the GRACE-FO satellite in decimeters(dm) with the magnetometer position in front(left) of the satellite given as a black sphere. The induced magnetic field in nT is shown at the magnetometer position in the form of vectors. At the back of the satellite, the magnetic dipole moment at maximum control currents for the magnetorquers is given in Am^2 for the same magnetorquer colors, while the magnetorquers are depicted as bars with the same colors as their respective magnetometer axis. The dipole moment vectors approximately align with the X, Y, and Z axes of the satellite and are nearly orthogonally aligned to each other $\pm 1^\circ$.

by the direction of the laser instrument, can be seen to the left. Also, the magnetometer is located in front of the satellite, depicted by a black sphere. The induced magnetic field for the three magnetorquers is shown at the magnetometer position. Then, for the approximate magnetorquer positions, as derived from construction drawings, the magnetic moment is calculated for the maximum magnetorquer control currents. These can be seen in the back of the satellite. The values for the magnetorquers are also given in Table 2. The table is diagonally dominated, which shows that the magnetorquer momenta are mainly aligned with the X, Y, and Z-axes, while the X and Y axes show some minor mixing. In addition, the angles between the different momenta measure 89.3°, 90.4°, and 89.2° respectively, meaning that the momenta are approximately orthogonal.

These results closely align with the expectations for the satellite: The three magnetorquers are built in an orthogonal fashion to enable the control of the attitude of the satellite. The positions and their momenta also match with the alignment axes of the built-in magnetorquer bars. This means that the PINN can reliably find and assign the position and strength of the sources of artificial satellite disturbances caused by dipoles. In particular, introducing PINNs can open the black box that NNs represent, allowing insight into the patterns learned from the data.

5.4 FAC Analysis

To show the viability for analyzing geomagnetic phenomena, auroral FACs are investigated. Therefore, Figure 8 and Figure 9 show the summarized FAC in dependence of the MLT and QDLat of the GOCE and GRACE-FO1 satellites, respectively. The FACs have been derived from the calibrated magnetometer data and summarized by an aggregation function into bins of 2°.

For the GOCE satellite, the FAC values have been aggregated by the median, while for the GRACE-FO satellite, the mean could be used as it contains many more measurements in the dataset. For both satellites, the bow-shaped Region 1 and 2 currents become visible. This confirms that the calibration process expectedly preserves natural signals, and the shortcomings of the previous approach by Styp-Rekowski et al. (2022b) could be overcome by including the AMPS model and introducing physical laws into the NN. Figure 8 contains a comparison of the newly calibrated GOCE data with the previously published dataset. The FACs in the Northern Hemisphere were less pronounced, and for the Southern Hemisphere, no bow-like shapes were visible, as can be seen in the upper half of the figure. With the new calibration method, this has changed drastically, and the expected shape is present in the bottom half of the figure which is especially visible within the Southern Hemisphere. For the GRACE-FO satellites, the result has been similar to previous studies.

A comparison between the GRACE-FO-derived FACs and Swarm A- and B-derived FACs has been conducted. Therefore, a slice of the MLT was taken during June 2019, ranging from MLTs of 5.5 to 6.5, representing the dawn. This choice has been made as strong currents are present during Northern Hemisphere summer for this MLT range, as seen in Figure 9.

The data have been low-pass filtered with a 20s cutoff similar to Xiong et al. (2021), representing large-scale FACs (>150km) for the GRACE-FO mission. For every degree of QDLat, the values have been summarized by a boxplot and visualized in Figure 10. In addition, the altitude, MLT, and magnetic indices are given as mean values in dependence on QDLat. Both GRACE-FO satellites are in good agreement with each other and capture similar structures as the Swarm satellites. Remaining differences can be attributed to the MLT of the satellites and the noise level of the calibrated GRACE-FO platform magnetometer data. Thus, the usage of calibrated GRACE-FO magnetic data to support measurements by the Swarm satellite is encouraged for investigations of magnetic phenomena on a global scale.

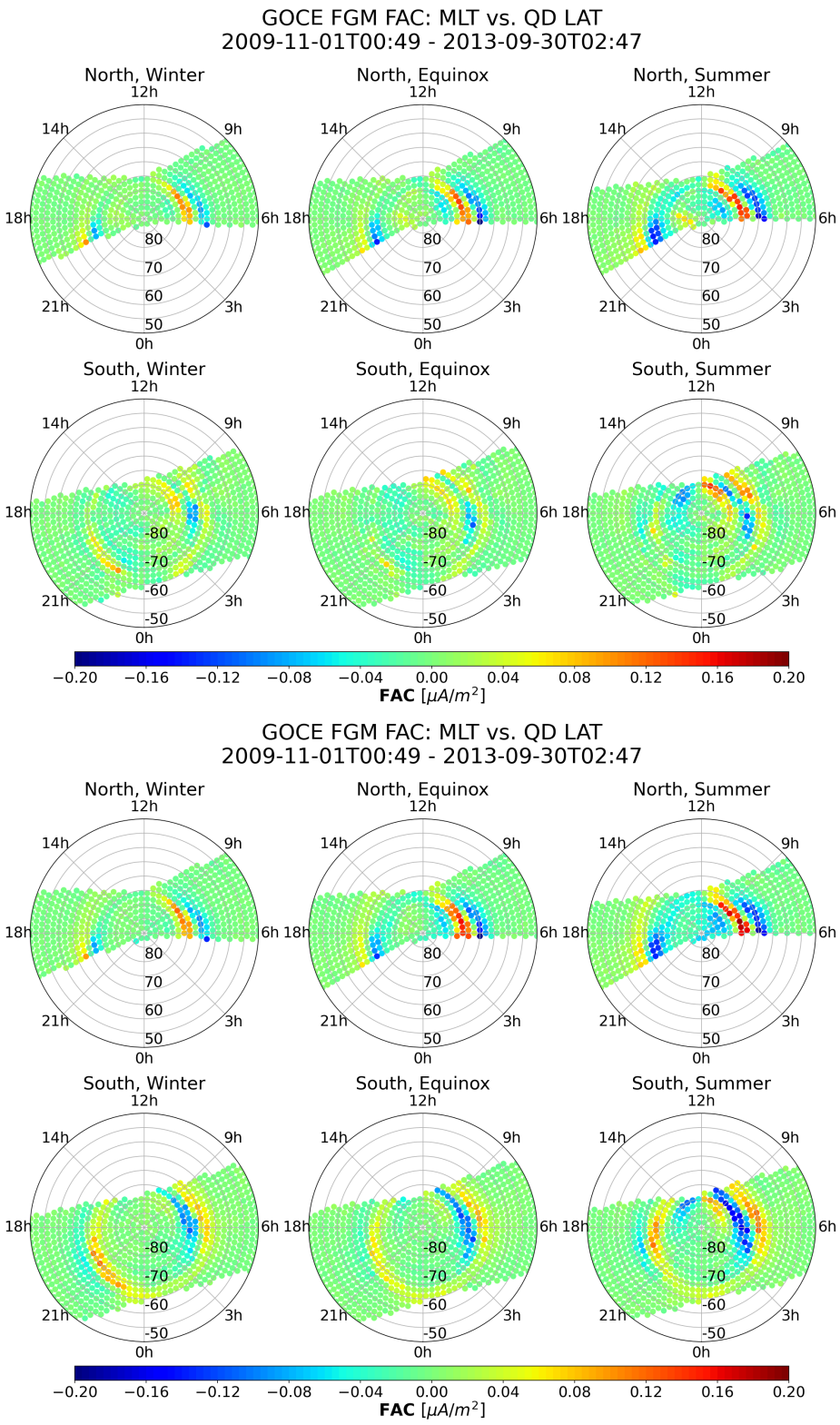

Figure 8. Field-aligned currents as derived from calibrated data of the GOCE satellite mission. Summarized median by MLT and QDLat for the Northern and Southern Hemispheres. At the top is the plot as derived from the previous approach (Styp-Rekowski et al., 2022b) and at the bottom is the newly derived plot with enhanced FACs for the Southern Hemisphere.

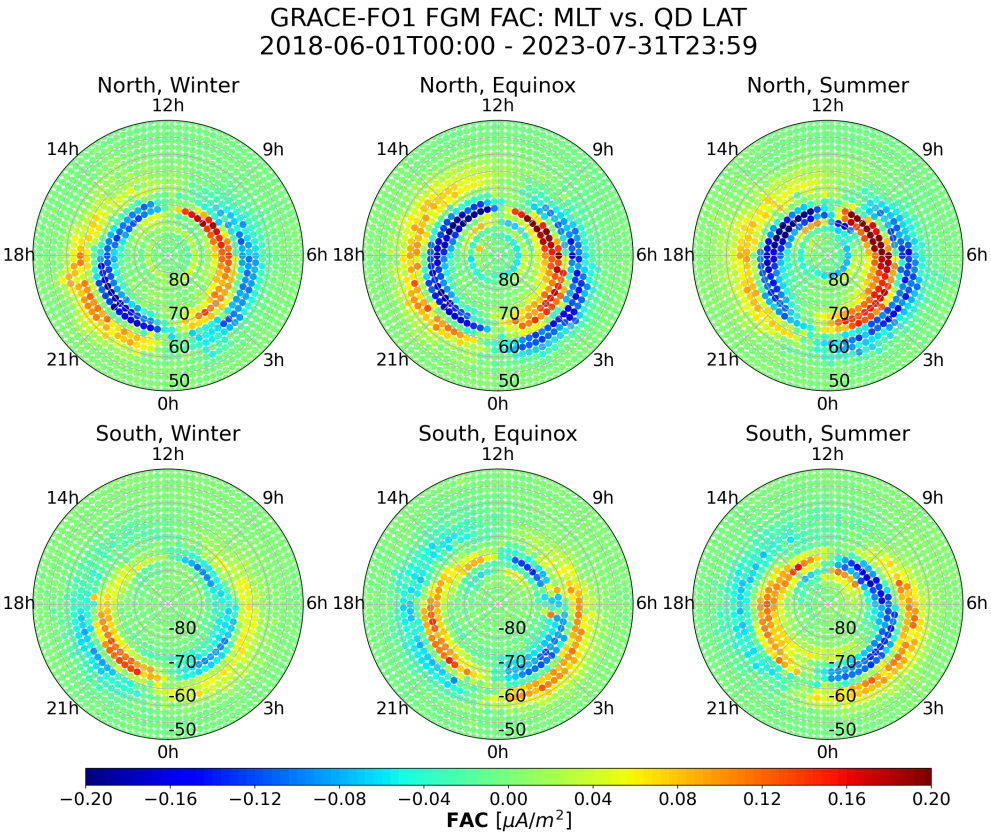

Figure 9. Field-aligned currents as derived from calibrated data of the GRACE-FO1 satellite. Summarized mean by MLT and seasons for the Northern and Southern Hemispheres.

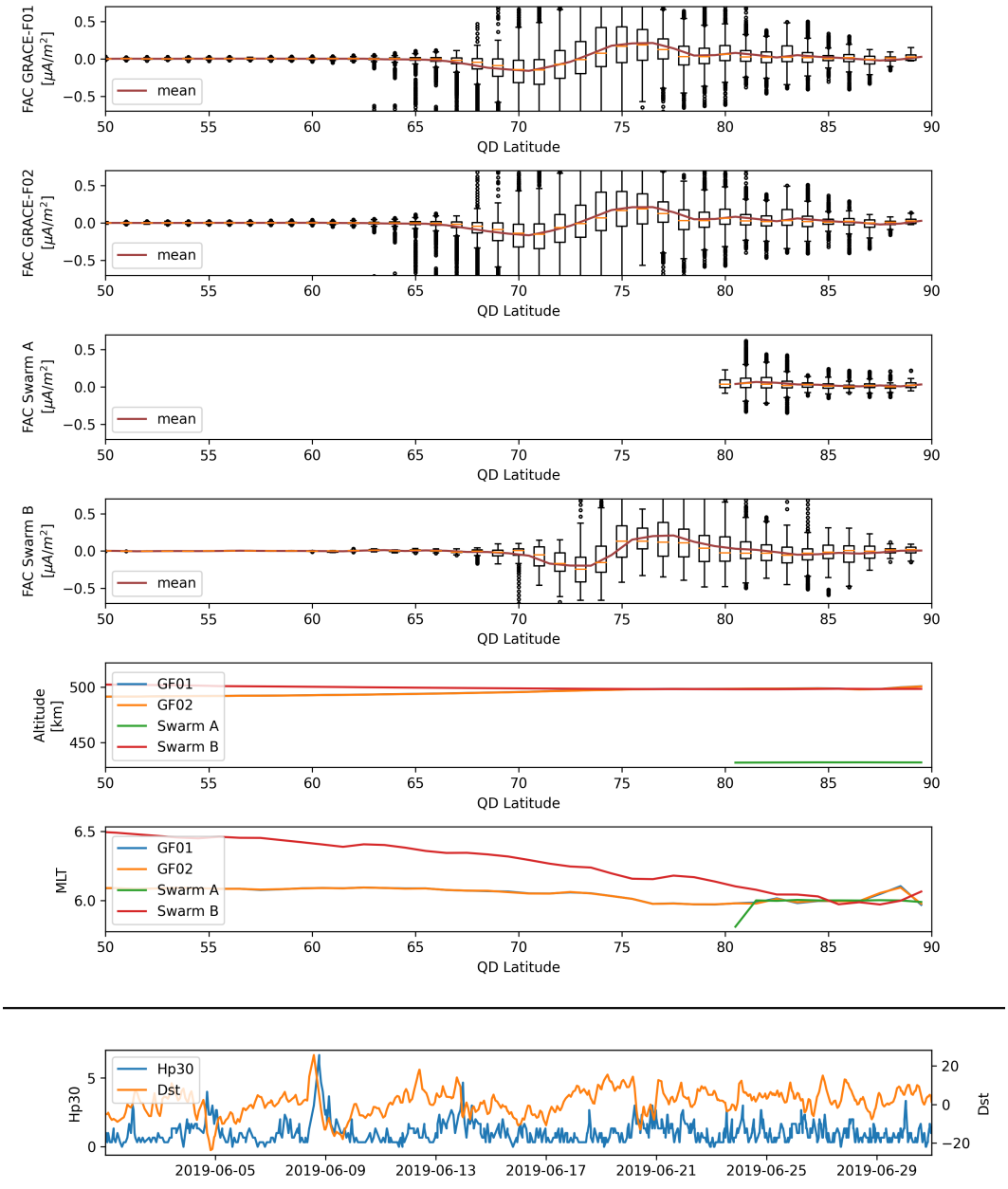

Figure 10. Boxplot summary for FAC data derived from GRACE-FO1, GRACE-FO2, Swarm-A, and Swarm-B missions. Data are selected from June 2019 for an MLT between 5.5 LT and 6.5 LT and quasi-dipole latitude between 50° and 90°, representing dawn. In addition, the altitude and MLT are given as mean values in dependence on quasi-dipole latitude. The magnetic indices are given in dependence of the time in the last panel.

5.5 Multi-mission orbit constellations

As an example of the application of calibrated GRACE-FO data, a recent geomagnetic storms has been investigated. The analyzed storm occurred on the 4th of November 2021, shown in Figure 11 for four days around the high geomagnetic activity. The distribution of FAC values in dependence of QDLat and MLT for the Northern and Southern Hemisphere is shown. The Swarm-A and Swarm-B measurements are shown in squares, while the GRACE-FO1 measurements are given in circles. The coloring contains an alpha value. Thus, if a circle is prominently visible on a square, the GRACE-FO measurements disagree with the Swarm measurements, while for an agreement, the circle visually merges with the square. The figure is separated into three rows, where for every row a different Hp30 selection has been applied to visualize differences in the geomagnetic activity, ranging from low activity with $Hp30 <= 2.0$, medium activity with $2.0 < Hp30 <= 4.0$, to high activity with $4.0 < Hp30$.

The MLTs of the satellite mission strongly vary and show the strength of additional data from non-dedicated missions. The global coverage of MLTs is strongly increased by using GRACE-FO data that has MLTs with a difference of about 3.7 and 3.8 hours on average compared to the Swarm A and B satellites for this time period, respectively. The extension of the auroral oval during storm time becomes visible as the covered area within the plot is larger because of the increased spatiotemporal coverage, enabling a global picture during magnetic storms. The idea of improving the global coverage of geomagnetic field measurements through non-dedicated satellite missions becomes evident here.

5.6 Evaluation against Swarm data

As the GRACE-FO mission has been operating since 2018 and the Swarm mission has been in orbit since 2013, a comparison between the calibrated GRACE-FO data and the data provided by the Swarm mission is possible, which is not the case for GOCE, since GOCE did not operate simultaneously with Swarm.

Therefore, the residuals between the Swarm and GRACE-FO1 data against their respective CHAOS-7 model prediction for low- and mid-latitudes during geomagnetic quiet times have been compared. Figure 12 shows a histogram of the North, East, and Center components of the vector magnetic field measurements for both missions. The data have been filtered with their respective flags and the resulting histogram is normalized. The Swarm data used in this study were downloaded in October 2023 from the VirES platform (Smith et al., 2023).

With Swarm as the high-precision mission achieving a steeper Gaussian distribution, it can be seen how a significant intersection of calibrated platform magnetometer data achieves similarly low residuals. The best result was achieved for the North component. This highlights the potential of platform magnetometer data to accompany high-precision missions with additional data of only modestly higher noise. Still, it needs to be emphasized that the calibration of platform magnetometer data would not be possible without a high-precision mission in space to act as a reference point because non-dedicated satellites mostly do not carry absolute magnetometers.

In addition, all conjunctions between the GRACE-FO1 satellite and the Swarm A satellite between June 2018 and July 2023 have been analyzed. Again, filtering for geomagnetic quiet times and flags has been applied. For both missions, for every data point where the distance between the two satellites was below 400 km, their respective residual with the CHAOS-7 model has been computed and the difference between the calculated residuals has been used for the conjunctions. The resulting conjunctions are binned by QDLat and MLT and aggregated by the mean, as shown in Figure 13. Overall, the conjunctions carry a low residual for low- and mid-latitude while having areas of larger residual around the poles. No apparent correlation is visible between QDLat or MLT for

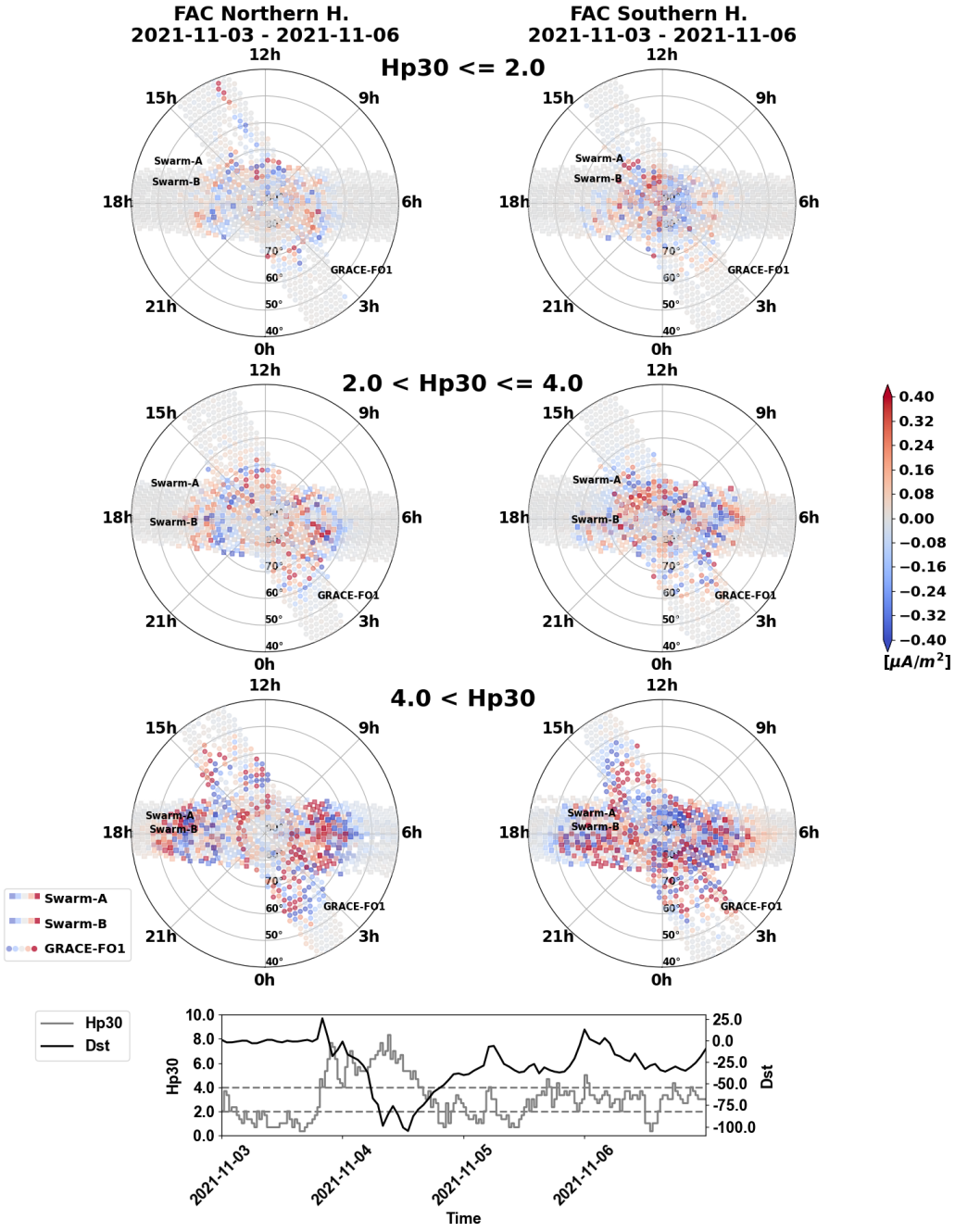

Figure 11. The mean FAC values for the Swarm-A, Swarm-B, and GRACE-FO satellite missions around the magnetic storm of the 4th of November, 2021, shown for four days, in dependence of magnetic local time and quasi-dipole latitude. The plot is divided into three rows, depending on the Hp30 index. Additionally, the Hp30 and *Dst* indices for this time frame are given.

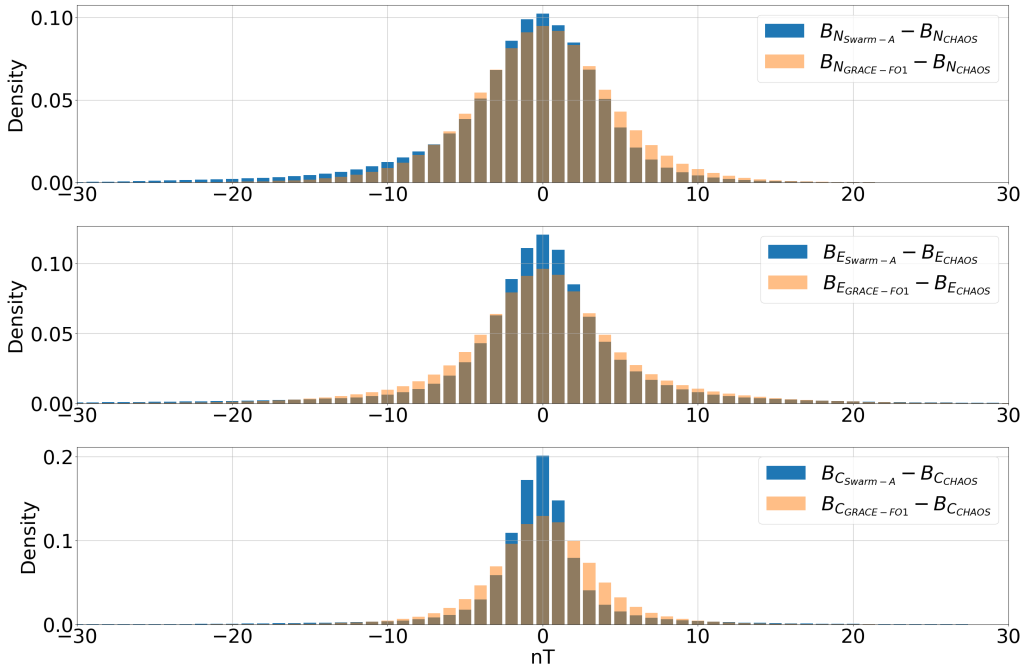

Figure 12. Residual distribution comparison of the Swarm (blue) and GRACE-FO1 (orange) calibrated data compared to the CHAOS7 reference model for the whole period from June 2018 to July 2023 within a histogram plot with bin sizes of 1 nT for the magnetic North (top), East (middle) and Center (bottom) component. Note the different vertical scales.

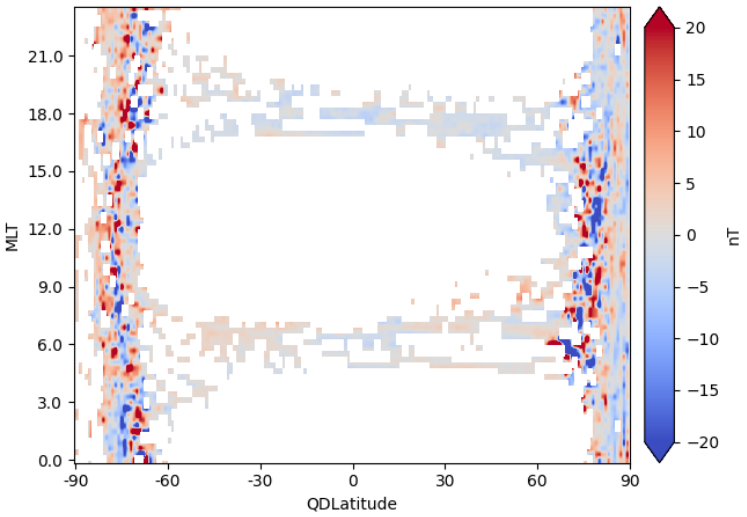

Figure 13. Conjunctions between the Swarm A and GRACE-FO1 satellites from June 2018 to July 2023. Data are selected by flags and geomagnetic quietness. The heatmap compares the residual to their respective CHAOS-7 model and shows the aggregated mean by quasi-dipole latitude and magnetic local time binning.

the conjunctions. Overall, this shows that the calibrated platform magnetometer data are in good agreement with the current high-precision mission in orbit.

6 Conclusion

This work introduced a major extension to the previous ML approach (Styp-Rekowski et al., 2022b) to calibrating platform magnetometers carried by non-dedicated satellites. By introducing the physical Biot-Savart law into the NN, the new PINN is able to correctly handle and identify magnetic dipoles acting within the satellite system. Additionally, the AMPS model was added to our reference model to anticipate large-scale auroral current system disturbance, increasing the calibration quality, particularly in the polar regions. When applied to the two satellite missions, GOCE and GRACE-FO1 together with GRACE-FO2, mean absolute residual values of 6.56 nT, 3.57 nT, and 3.82 nT could be obtained, respectively. Compared to the previous approach, the residuals of the proposed methodology lie in a similar range while overcoming the identified shortcomings. These results enable the application of the calibrated data to analyze geomagnetic phenomena, as was shown exemplarily for FACs and geomagnetic storms. By its nature, this approach is mostly automated, so that it is straightforward to apply it to the calibration of magnetometer data from other non-dedicated satellites in the future. The dataset of the two missions calibrated alongside this work is available (Styp-Rekowski et al., 2022a, 2023).

Acronyms

AMPS Average Magnetic field and Polar current System
CHAMP CHAllenging Minisatellite Payload
ETL Extract, transform, and load process
FAC Field-aligned currents
FFNN Feed-forward neural network
GOCE Gravity and steady-state Ocean Circulation Explorer
GRACE Gravity Recovery And Climate Experiment
GRACE-FO Gravity Recovery And Climate Experiment Follow-On
IMF Interplanetary Magnetic Field
MAE Mean absolute error
ML Machine Learning
MLT Magnetic local time
MTQ Magnetorquer
NEC North-East-Center frame
NN Neural network
PIC Physics-informed component
PINN Physics-informed neural network
QDLat Quasi-dipole latitude
SD Standard deviation

Open Research Section

Data used in this study are publicly available from the European Space Agency (ESA) for the GOCE satellite (`https://earth.esa.int/eogateway/missions/goce/data`) and from the German Research Center for Geosciences (GFZ) for the GRACE-FO satellites (Michaelis et al., 2021). The Swarm data were accessed through the viresclient (Smith et al., 2023). The different indices and supplementary data were available from the NASA for the B_y and B_z of the IMF, the solar wind speed V_{sw}, the Dst-index, the $F10.7$-index (Papitashvili & King, 2020); the Hp30-index is provided by the GFZ (Yamazaki et al.,

2022). The reference models used in this publication can be accessed through their respective publications for the AMPS model (Laundal et al., 2018) and the CHAOS-7 model (Finlay et al., 2020).

The generated data from this publication for the calibrated geomagnetic field measurements, as well as their respective CHAOS-7 estimates and the derived FACs, can be found under version 301 Styp-Rekowski et al. (2022a) for the GOCE satellite and version 302 Styp-Rekowski et al. (2023) for the GRACE-FO satellites.

Acknowledgments

We thank Guram Kervalishvili, Heinz-Peter Brunke, Jan Rauberg, and Martin Rother for the vivid and fruitful discussions. This work is supported through HEIBRIDS - Helmholtz Einstein International Berlin Research School in Data Science under contract no. HIDSS-0001. This study has been partly supported by Swarm DISC activities funded by ESA under contract no. 4000109587/13/I-NB.

References

Alken, P., Olsen, N., & Finlay, C. C. (2020). Co-estimation of geomagnetic field and in-orbit fluxgate magnetometer calibration parameters. *Earth, Planets and Space*, *72*, 1–32.

Anderson, B. J., Takahashi, K., & Toth, B. A. (2000). Sensing global Birkeland currents with Iridium® engineering magnetometer data. *Geophysical Research Letters*, *27*(24), 4045–4048.

Bader, J., Styp-Rekowski, K., Doehler, L., Becker, S., & Kao, O. (2022). Macaw: The machine learning magnetometer calibration workflow. In *2022 ieee international conference on data mining workshops (icdmw)* (pp. 1095–1101).

Broadfoot, R. M., Miles, D. M., Holley, W., & Howarth, A. D. (2022). In situ calibration of the Swarm-Echo magnetometers. *Geoscientific instrumentation, methods and data systems*, *11*(2), 323–333.

Cossavella, F., Herman, J., Hoffmann, L., Fischer, D., Save, H., Schlepp, B., & Usbeck, T. (2022). Attitude Control on GRACE Follow-On: Experiences from the First Years in Orbit. In *Space operations: Beyond boundaries to human endeavours* (pp. 493–517). Springer.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, *92*(3), 88.

Drinkwater, M., Floberghagen, R., Haagmans, R., Muzi, D., & Popescu, A. (2003). Vii: Closing session: Goce: Esa's first earth explorer core mission. *Space science reviews*, *108*(1), 419–432.

Finlay, C. C., Kloss, C., Olsen, N., Hammer, M. D., Tøffner-Clausen, L., Grayver, A., & Kuvshinov, A. (2020). The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. *Earth, Planets and Space*, *72*(1), 1–31. doi: 10.1186/s40623-020-01252-9

Floberghagen, R., Fehringer, M., Lamarre, D., Muzi, D., Frommknecht, B., Steiger, C., ... Da Costa, A. (2011). Mission design, operation and exploitation of the gravity field and steady-state ocean circulation explorer mission. *Journal of Geodesy*, *85*(11), 749–758.

Friis-Christensen, E., Lühr, H., & Hulot, G. (2006, 04). Swarm: A constellation to study the Earth's magnetic field. *Earth, planets and space*, *58*, 351-358. doi: 10.1186/BF03351933

Jackson, J. D. (1999). *Classical electrodynamics* (3rd ed. ed.). New York, NY: Wiley. Retrieved from `http://cdsweb.cern.ch/record/490457`

Kloss, C., Finlay, C. C., Laundal, K. M., & Olsen, N. (2023). Polar ionospheric

currents and high temporal resolution geomagnetic field models. *Geophysical Journal International*, *235*(2), 1736–1760.

Kornfeld, R. P., Arnold, B. W., Gross, M. A., Dahya, N. T., Klipstein, W. M., Gath, P. F., & Bettadpur, S. (2019). GRACE-FO: the gravity recovery and climate experiment follow-on mission. *Journal of spacecraft and rockets*, *56*(3), 931–951.

Laundal, K. M., Finlay, C. C., Olsen, N., & Reistad, J. P. (2018). Solar Wind and Seasonal Influence on Ionospheric Currents From Swarm and CHAMP Measurements. *Journal of Geophysical Research: Space Physics*, *123*(5), 4402-4429. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JA025387` doi: https://doi.org/10.1029/2018JA025387

Lowrie, W. (2023). *The Earth's Magnetic Field*. Oxford University Press.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Matzka, J., Stolle, C., Yamazaki, Y., Bronkalla, O., & Morschhauser, A. (2021). The geomagnetic Kp index and derived indices of geomagnetic activity. *Space weather*, *19*(5), e2020SW002641.

Michaelis, I., Stolle, C., & Rother, M. (2021). *GRACE-FO calibrated and characterized magnetometer data.* GFZ Data Services. doi: https://doi.org/10.5880/GFZ.2.3.2021.002

Michaelis, I., Styp-Rekowski, K., Rauberg, J., Stolle, C., & Korte, M. (2022). Geomagnetic data from the GOCE satellite mission. *Earth, Planets and Space*, *74*(1), 1–16.

Neubert, T., Mandea, M., Hulot, G., Von Frese, R., Primdahl, F., Jørgensen, J. L., ... Risbo, T. (2001). Ørsted satellite captures high-precision geomagnetic field data. *Eos, Transactions American Geophysical Union*, *82*(7), 81–88.

Ng, A. Y. (2004). Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on machine learning* (p. 78).

Olsen, N. (2021). Magnetometer data from the GRACE satellite duo. *Earth, Planets and Space*, *73*(1), 1–20.

Olsen, N., Albini, G., Bouffard, J., Parrinello, T., & Tøffner-Clausen, L. (2020). Magnetic observations from CryoSat-2: calibration and processing of satellite platform magnetometer data. *Earth, Planets and Space*, *72*(1), 1–18.

Olsen, N., Friis-Christensen, E., Floberghagen, R., Alken, P., Beggan, C. D., Chulliat, A., ... others (2013). The Swarm satellite constellation application and research facility (SCARF) and Swarm data products. *Earth, Planets and Space*, *65*, 1189–1200.

Olsen, N., & Stolle, C. (2012). Satellite geomagnetism. *Annual Review of Earth and Planetary Sciences*, *40*, 441–465.

Papitashvili, N. E., & King, J. H. (2020). Omni 1-min data. *Dataset]. NASA Space Physics Data Facility. https://doi. org/10.48322/45bb-8792*. doi: https://doi.org/10.48322/45bb-8792

Park, J., Stolle, C., Yamazaki, Y., Rauberg, J., Michaelis, I., & Olsen, N. (2020). Diagnosing low-/mid-latitude ionospheric currents using platform magnetometers: CryoSat-2 and GRACE-FO. *Earth, Planets and Space*, *72*, 1–18.

Prölss, G. (2012). *Physics of the Earth's space environment: an introduction.* Springer Science & Business Media.

Reigber, C., Lühr, H., & Schwintzer, P. (2002). CHAMP mission status. *Advances in space research*, *30*(2), 129–134.

Smith, A. R. A., Pačes, M., & Santillan, D. (2023, July). *ESA-VirES/VirES-Python-Client.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8176069` doi: 10.5281/zenodo.8176069

Springmann, J., Cutler, J., & Bahcivan, H. (2010). Magnetic sensor calibration and residual dipole characterization for application to nanosatellites. In *Aiaa/aas*

astrodynamics specialist conference (p. 7518).

Stolle, C., Michaelis, I., Xiong, C., Rother, M., Usbeck, T., Yamazaki, Y., ... Styp-Rekowski, K. (2021). Observing Earth's magnetic environment with the GRACE-FO mission. *Earth, Planets and Space*, *73*(1), 1–21.

Stolle, C., Olsen, N., Anderson, B., Doornbos, E., & Kuvshinov, A. (2021). Special issue "Characterization of the geomagnetic field and its dynamic environment using data from space-based magnetometers". *Earth, Planets and Space*, *73*(1), 1–4.

Stolle, C., Olsen, N., Richmond, A. D., & Opgenoorth, H. J. (2017). Topical Volume on Earth's Magnetic Field—Understanding Geomagnetic Sources from the Earth's Interior and Its Environment. *Space Science Reviews*, *206*, 1–3.

Styp-Rekowski, K., Michaelis, I., Korte, M., & Stolle, C. (2023). *GRACE-FO ML-calibrated magnetic field data.* GFZ Data Services. (V. 0302) doi: https://doi.org/10.5880/GFZ.2.3.2023.001

Styp-Rekowski, K., Michaelis, I., Stolle, C., Baerenzung, J., Korte, M., & Kao, O. (2022a). *GOCE ML-calibrated magnetic field data.* GFZ Data Services. (V. 0301) doi: https://doi.org/10.5880/GFZ.2.3.2022.002

Styp-Rekowski, K., Michaelis, I., Stolle, C., Baerenzung, J., Korte, M., & Kao, O. (2022b). Machine learning-based calibration of the GOCE satellite platform magnetometers. *Earth, Planets and Space*, *74*(1), 1–23.

Styp-Rekowski, K., Stolle, C., Michaelis, I., & Kao, O. (2021). Calibration of the grace-fo satellite platform magnetometers and co-estimation of intrinsic time shift in data. In *2021 ieee international conference on big data (big data)* (pp. 5283–5290).

Sugiura, M. (1964). Hourly values of equatorial Dst for the IGY. *Ann. Int. Geophys.*, *35*(9).

Xiong, C., Stolle, C., Michaelis, I., Lühr, H., Zhou, Y., Wang, H., ... Rauberg, J. (2021). Correlation analysis of field-aligned currents from the magnetic measurements of GRACE follow-on mission. *Earth, Planets and Space*, *73*, 1–13.

Yamazaki, Y., Matzka, J., Stolle, C., Kervalishvili, G., Rauberg, J., Bronkalla, O., ... Jackson, D. R. (2022). Geomagnetic Activity Index Hpo. *Geophysical Research Letters*, *49*(10), e2022GL098860. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022GL098860` (e2022GL098860 2022GL098860) doi: https://doi.org/10.1029/2022GL098860

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., ... others (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716.