Universität Rostock
Traditio et Innovatio

MACHINE LEARNING CALIBRATION OF SATELLITE PLATFORM MAGNETOMETER DATA

Kumulative Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Rostock

https://doi.org/10.18453/rosdok_id00004838

vorgelegt von Kevin Marcel Styp-Rekowski
aus Berlin
Rostock, 19.02.2024

Gutachter:
Prof. Dr. Claudia Stolle, University of Rostock, Leibniz Institute of Atmospheric Physics
Prof. Dr. Odej Kao, Technical University Berlin, Distributed and Operating Systems
PD Dr. Alexander Grayver, University of Cologne, Institute of Geophysics and Meteorology

Jahr der Einreichung: 2024
Jahr der Verteidigung: 2024

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank all my mentors, colleagues, friends, supporters, and my family. In particular, I would like to thank my supervisor, Prof. Claudia Stolle, for the countless invaluable meetings, guidance, experiences, and patience with me. I wish to extend my thanks to my second supervisor, Prof. Odej Kao, who provided me with his advice, infected me with his passion, and helped me to keep perspective. In addition, I want to thank my reviewers for reviewing my work, especially PD Dr. Alexander Grayver as the third reviewer.

Working with two institutions on an interdisciplinary project during a pandemic has been a valuable experience in my life. I therefore would like to thank all the members from the GFZ section 2.3 for geomagnetism. A special thanks goes to Ingo for his always open ear, the countless hours of fruitful discussions and result analysis, as well as for increasing my knowledge about the Earth's magnetic field and satellites immensely. Dear Alex, Anna, Rosa, Franziska, and Mareike, I would like to thank you all for your support in all organizational and technical matters concerning my work. Of course, I equally want to thank my colleagues from the TU Berlin DOS group for all the support and for pushing each other to new heights. A special thanks goes to Florian, with whom I wrote my master's thesis and who paved the way for me as a PhD Candidate. Dear Jana and Britta, I would like to thank you for your support in all organizational and technical matters.

I would like to thank HEIBRiDS for giving me this unique opportunity to do research on this thesis. The numerous seminars, summer schools, other meetings, and networking with fellow students have helped me to develop personally. Special thanks to Eirini, Nina, and Sandra.

I want to thank the European Space Agency for enabling this research within an environment of openness and support through the publication of data or the organization of community events. Many thanks to the Swarm community for welcoming me and my research with open arms. Dear co-authors, a very special thanks to you because without your cooperation, your dedication, and your hard work, I would not be here.

I would like to say a big thank you to my family. I would especially like to thank my parents for their years of support, love, and dedication, and also Patrick, Marietta, and Mathilda from the bottom of my heart. Also, I would like to thank my friends who have helped me stay grounded with their support, happy memories, and encouragement.

I would like to thank my wife Carolin, who always has my back, whom I love unconditionally, and who has endured the last few years of this intense time with me. Despite the eventful time as a PhD candidate, a lot has also happened in my private life: We got married, looked for a house together, and, of course, you gave birth to our daughter Emilia. Thank you both from the bottom of my heart.

ABSTRACT

The Earth's magnetic field is subject to constant evolution, the trend and changes of the magnetic field, the impact on climate, different time-varying and small-scale phenomena, space weather and hazard prediction, the magnetosphere-ionosphere coupling, or the continuation of monitoring are current research subjects.

With its complex nature and coupling with external phenomena, analysis heavily relies on measurement data from which the empirical models are created. Space-based measurements are vital for understanding Earth's magnetic field, providing comprehensive global coverage and spanning various local times and altitudes. Despite these advantages, the accessibility of high-precision data remains limited, with one high-precision mission currently in orbit, the Swarm satellites. Non-dedicated satellites, designed for purposes other than measuring the magnetic field, frequently carry platform magnetometers for navigational tasks. Increasing the data availability through the calibration of platform magnetometers from non-dedicated satellites is investigated in this work. However, the inherent rough calibration of these magnetometers makes a more refined calibration and characterization approach necessary for scientific exploration.

This dissertation focuses on applying Machine Learning for the post-launch calibration of platform magnetometers, introducing a novel methodology to enhance the quality and reliability of the datasets that are produced. The publication-based chapters provide a detailed illustration of how the calibration was achieved. First, an analytical calibration for the GOCE satellite mission is demonstrated, which achieves good results by thorough analysis of the input data and shows applications of the dataset for magnetospheric phenomenon analysis. Subsequently, the ML-based calibration is introduced, which is applied to the GOCE and GRACE-FO satellite missions and achieves residuals of well below 10 nanoTesla, the unit of magnetic flux density, enabling scientific application. Next, the proposed calibration methodology is extended by incorporating the first-principle physical law of Biot-Savart into the neural network, incorporating physics-based information into the calibration to improve the physical validity. The evaluation highlights the achievements in the calibration of platform magnetometers, showcasing the efficacy of the Machine Learning-based approach in improving the accuracy and usability of geomagnetic data while confirming the plausibility through the detection of geophysical phenomena.

By leveraging Machine Learning techniques, this research aims to automate and improve the calibration process, addressing the inherent challenges given by the rough calibration of platform magnetometers. The quality of the research is verified with two use cases, the GOCE and GRACE-FO satellite missions.

In conclusion, this thesis contributes to increasing the data availability for geomagnetic field analysis and lays a foundation for future application to other missions. By introducing a novel calibration methodology using Machine Learning, this research drives the field toward a more sophisticated and nuanced understanding of Earth's magnetic field dynamics by monitoring the Earth with a true swarm of satellites.

ZUSAMMENFASSUNG

Das Erdmagnetfeld unterliegt einer ständigen Entwicklung, der Trend und die Veränderungen des Magnetfelds, die Auswirkungen auf das Klima, verschiedene zeitlich veränderliche und kleinräumige Phänomene, die Vorhersage des Weltraumwetters und -gefahren, die Kopplung von Magnetosphäre und Ionosphäre oder die Fortsetzung der Überwachung sind aktuelle Forschungsthemen.

Die Analyse des komplexen Magnetfelds und seiner Wechselwirkungen mit externen Phänomenen basiert stark auf Messdaten, die zur Erstellung empirischer Modelle dienen. Weltraumbasierte Messungen sind entscheidend, da sie globale Abdeckung bieten, aber der Zugang zu hochpräzisen Daten bleibt begrenzt. Diese Arbeit untersucht die Kalibrierung von Plattformmagnetometern auf nicht spezialisierten Satelliten, um die Datenverfügbarkeit zu verbessern, was jedoch einen verfeinerten Ansatz erfordert, um wissenschaftliche Forschung zu ermöglichen.

Diese Dissertation befasst sich mit der Anwendung von maschinellem Lernen für die Kalibrierung von Plattformmagnetometern nach dem Start und stellt eine neuartige Methodik zur Verbesserung der Qualität und Zuverlässigkeit der erzeugten Datensätze vor. In den auf Publikationen basierenden Kapiteln wird detailliert dargestellt, wie die Kalibrierung erreicht wurde. Zunächst wird eine analytische Kalibrierung für die GOCE-Satellitenmission demonstriert, die durch eine gründliche Analyse der Eingabedaten gute Ergebnisse erzielte und Anwendungen des Datensatzes für die Analyse magnetosphärischer Phänomene aufzeigt. Anschließend wird die ML-basierte Kalibrierung vorgestellt, die auf die GOCE- und GRACE-FO-Satellitenmissionen angewendet wird und Residuen von deutlich unter 10 nanoTesla, der Einheit der magnetischen Flussdichte, erzielt, was wissenschaftliche Anwendungen ermöglicht. Als nächstes wird die vorgeschlagene Kalibrierungsmethode erweitert, indem das grundlegende physikalische Gesetz von Biot-Savart in das neuronale Netz integriert wird, wodurch die physikalische Gültigkeit der Kalibrierung verbessert wird. Die Bewertung hebt die Erfolge bei der Kalibrierung von Plattformmagnetometern hervor und zeigt die Wirksamkeit des auf maschinellem Lernen basierenden Ansatzes bei der Verbesserung der Genauigkeit und Nutzbarkeit geomagnetischer Daten, während die Plausibilität durch die Erkennung geophysikalischer Phänomene bestätigt wird.

Durch den Einsatz von Techniken des maschinellen Lernens zielt diese Forschung darauf ab, den Kalibrierungsprozess zu automatisieren und zu verbessern, um die mit der groben Kalibrierung von Plattformmagnetometern verbundenen Herausforderungen zu bewältigen. Die Qualität der Forschung wird anhand von zwei Anwendungsfällen, den Satellitenmissionen GOCE und GRACE-FO, verifiziert.

Zusammenfassend lässt sich sagen, dass diese Arbeit dazu beiträgt, die Datenverfügbarkeit für die Analyse des geomagnetischen Feldes zu erhöhen und eine Grundlage für die künftige Anwendung bei anderen Missionen schafft. Durch die Einführung einer neuartigen Kalibrierungsmethode unter Verwendung von maschinellem Lernen treibt diese Forschung das Feld in Richtung eines differenzierteren und nuancierteren Verständnisses der Dynamik des Erdmagnetfeldes voran, indem die Erde mit einem echten Satellitenschwarm überwacht werden soll.

Contents

List of Figures

Abbreviations

AMPS - Average Magnetic field and Polar current System
AOCS - Attitude and Orbit Control System
ASM - Absolute scalar magnetometer
CDF - Common data format
CHAMP - CHAllenging Minisatellite Payload
CHAOS - CHAmp Ørsted SAC-C magnetic field model
DOI - Digital object identifier
ELU - Exponential Linear Unit
FAC - Field-aligned currents
GFZ - Helmholtz Centre Potsdam
GOCE - Gravity and steady-state Ocean Circulation Explorer
GRACE-FO - Gravity Recovery And Climate Experiment Follow-On
IGRF - International Geomagnetic Reference Field
IMF - Interplanetary Magnetic Field
JPL - Jet Propulsion Laboratory
LEO - Low Earth orbit
LRI - Laser Ranging Interferometer
MAE - Mean absolute error
ML - Machine Learning
MSE - Mean squared error
NASA - National Aeronautics and Space Administration
NEC - North-East-Center frame
NN - Neural network
PINN - Physics-informed neural network
RELU - Rectified Linear Unit
VFM - Vector field magnetometer

1 MOTIVATION

The Earth's magnetic field is constantly changing and remains a current research topic. The geomagnetic field has been studied for a long time and has been used in applications like navigation for hundreds of years. Protection from solar wind and radiation provided by the geomagnetic field is essential for life on our planet. Furthermore, understanding the dynamic nature of the Earth's magnetic field is vital not only for scientific exploration but also for practical applications such as space weather forecasting. In our continuous pursuit of increased insight, empirical models provide the backbone for analysis.

For empirical models, the data is the main source of information. Magnetic flux density data are needed for empirical modeling of the Earth's magnetic field. These stem from ground observatory data, combined with aerial campaigns and high-precision satellite data. The available data has led to a well-established understanding of the Earth's magnetic field and resulted in the creation of magnetic field models.

Space-based measurements play a significant role in modeling the Earth's magnetic field and geomagnetic phenomena as they, contrary to ground stations, provide a global coverage that is equally distributed. This global perspective enables a comprehensive understanding of the complex dynamics of the Earth's magnetic field, contributing valuable insights into geophysical processes. High-precision magnetic satellite data have thus a huge importance in generating sophisticated models of the Earth's magnetic field.

However, these high-precision satellite missions are limited in their coverage in terms of space, time, and different magnetic local times. In recent years, there have been efforts to increase the available data by utilizing other non-dedicated missions in space [56, 4, 54]. These non-dedicated satellites have a different mission goal than measuring the Earth's magnetic field. Nevertheless, many satellites in low Earth orbit (LEO) carry platform magnetometers onboard as part of their attitude and orbit control system (AOCS), used for navigation and satellite operations. A high noise in the measurement data is acceptable for their navigational tasks, thus they are placed within or at the body of the satellite system. The main idea is to reduce the noise of these data by calibration and characterization and thus increase their quality, which results in magnetic datasets accompanying the measurements of high-precision missions. To achieve this, their measurement data must be corrected for artificial disturbances introduced through their placement and mounting. These calibrated datasets also contain the opportunity to improve the data availability in the past, where no magnetic mission was available. This helps bridge the gap of high-precision measurements that occurred between 2010 and 2013.

In the past, this has been achieved using an analytical calibration where the platform magnetometer features are identified and adjusted for scaling, offset, and misalignment in their mutual angles within the calibration. Then, hand-selected features known to introduce the largest artificial disturbances are used in additional terms related to the characterization of their disturbance factor. Within their methodology, a preselection of features is conducted. Similar approaches have been applied to GRACE [44], CryoSat-2 [47], DMSP [1], GRACE-FO [55], Swarm-Echo [9], AMPERE [4], and GOCE [38]. For this, a sophisticated magnetic field model is required as it serves as the reference model needed within the calibration methodology because platform magnetometers deliver a relative vector measurement that needs to be calibrated with the absolute values of the reference model. Hereby, the aforementioned magnetic field models like the IGRF model [2], the CHAOS-7 model [17], the Mag.num model [51], or the Kalmag model [26] are needed. The empiric CHAOS-7 model is currently the most-used model in LEO and is constantly developed further [17]. This approach requires manual analytical work to determine the relevant features and their relationship to the magnetometer measurements before applying the calibration.

In this work, a new methodology is presented that utilizes neural networks that are trained on the available data of the satellite with a geomagnetic field model as the reference model to adjust the measurements such that the initial residual is reduced compared to the geomagnetic field model. Here, Machine Learning (ML) is used for fast, precise, and readily accessible automated calibration of magnetic datasets from non-dedicated missions. The proposed non-linear regression not only automatically identifies relevant features and their crosstalk but also adjusts the values of raw measurements, such as scaling, offsetting, or rotating, while characterizing corrections for external influences on sensor measurements, like temperature dependence. Particularly, handling multi-satellite magnetic data contributes to a substantial increase in geomagnetic datasets, offering valuable insights into Earth's magnetic field.

Challenges Since many satellites are build-wise different from each other, individual analysis and planning for every satellite are not possible in a timely manner. Satellites are among the best-monitored systems available, which means that there are many features available from which only a fraction is relevant to the magnetic properties of the satellite. Still, it is not known in advance which features may be relevant for the calibration, so finding them is an exponentially complex task. In addition, a lot of cross-talk between relevant features is present, meaning that feature combinations play a major role, which becomes exponentially hard to detect for analytic methods. Handling sensor data introduces additional challenges as they originate from a variety of instruments, measuring in different physical units and at different time intervals. The data include misreadings, missing data, misaligned and non-plausible timestamps, rendering data preprocessing a major task within this work. Also, the evaluation of existing models is strongly dependent on the determination of the satellite's position in terms of longitude, latitude, and altitude, as well as the exact timestamp for the position. One of the challenges includes, e.g., the lack of ground truth in high-latitude regions where existing magnetic field models do not include the so-called field-aligned currents (FAC), which occur highly fluctuating.

Research Objectives The goal of this work is to automatically learn the magnetic behavior of the satellite system post-launch to create an ML-based calibration model. The model is able to calibrate the instrument and correct for the artificial disturbances originating from the satellite itself and thus extract a calibrated dataset of the magnetometer signal to measure Earth's magnetic field in the background. This objective can be subdivided into 3 research objectives:

- RO-1: Show feasibility of ML-based calibration of platform magnetometers to improve the sensor readings.
- RO-2: Incorporate physics-based information into ML calibration of sensors to improve physical validity.
- RO-3: Extract information about the most important parts of the calibration and gain explainable insights.

Key contributions To answer these research objectives, the key contributions of this thesis are:

- The data have been investigated and analyzed, leading to the development of different preparation steps to streamline the given data from the satellite and other supplementary sources. A sophisticated data preprocessing pipeline that is generally applicable to platform magnetometer datasets is developed and implemented to prepare, convert, clean, and filter the data for subsequent calibration.
- I show that ML methods are able to achieve an automated post-launch calibration of satellite platform magnetometers. A successful calibration for two satellite missions, GRACE-FO and GOCE, is presented, overcoming different challenges like a partial lack of ground truth or cross-talk between the features that need to be considered. In addition, an extensive evaluation verifies the plausibility and quality of the calibration.
- An extension utilizing physics-informed neural networks has been introduced that incorporates the Biot-Savart formula for dipoles into the ML-based calibration model. Utilizing known laws from physics increased the physical plausibility of the calibration and enabled new insights into the interplay of the electronic components onboard the satellite.

- A feedback loop of current insights has been established. Thus, results and new information from the calibration have had an impact on the further development of the proposed approach. This includes information about the most influential properties onboard the satellite influencing the calibration or the position and strength of the dipoles.

Outline In the following, different parts of the described approach will be discussed in more detail. The Earth's magnetic field will be introduced in Section 2. Section 3 presents the three satellite missions used in this work and describes platform magnetometers in more detail. Section 4 gives an overview of ML in general and its application for the calibration of platform magnetometers. Section 4.2 summarizes the work conducted during this dissertation, while Sections 5 to 7 contain the publication-based chapters and detailed studies. Finally, Section 8 summarizes this work and points to future work directions.

2 The Earth's Magnetic Field

This chapter will provide a brief overview of Earth's magnetic field. First, the source, structure, and phenomena of the geomagnetic field will be described. Afterward, the underlying modeling data that led to the insights and understanding of the geomagnetic field will be described.

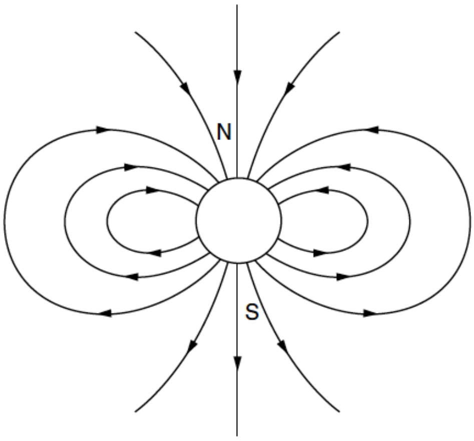


Figure 1: Schematic of dipole magnetic field lines in a vacuum. Source: [29]

The Earth's magnetic field evolves from Earth's molten, electrically conducting core, composed mainly of Iron and Nickel [29]. The heat in the core drives electric currents through thermal convection. In this process, the rotation of the Earth plays a crucial role, leading to the generation of electric currents through the geodynamo mechanism. The geodynamo involves the self-sustaining process wherein the convective motion of molten iron generates electric currents, and the resulting magnetic field further influences the motion of the conductive fluid. This self-sustaining process ultimately establishes the Earth's dipole-dominated magnetic field. These electric currents, combined with the Earth's rotation, contribute to establishing a dipole magnetic field. While these convection processes are very dynamic, the dominant lower-order terms lead to a dipole dominated magnetic field at the surface level of the Earth. Figure 1 contains a schematic visualization of the dipolar geomagnetic field, taken from Kelley [29], which shows an idealized view of a dipole magnetic field without considering further details.

While this is true for the driving processes of Earth's magnetic field without its surrounding environment, the actual magnetic field looks different. In our solar system, the Earth is embedded within the interplanetary magnetic field (IMF) originating from the Sun [49]. Driven by the solar wind, the IMF evolves from the Sun and forms the heliosphere up to the heliopause. The solar wind consists of a continuous stream of charged particles emitted by the Sun. The IMF interacts with the Earth's magnetic field at different levels, which leads to significant changes in the simplified dipole structure shown in Figure 1, while on the surface of the Earth the largest proportion of deviations from a simple dipole still originate from the core. The external magnetic field components change the structure further.

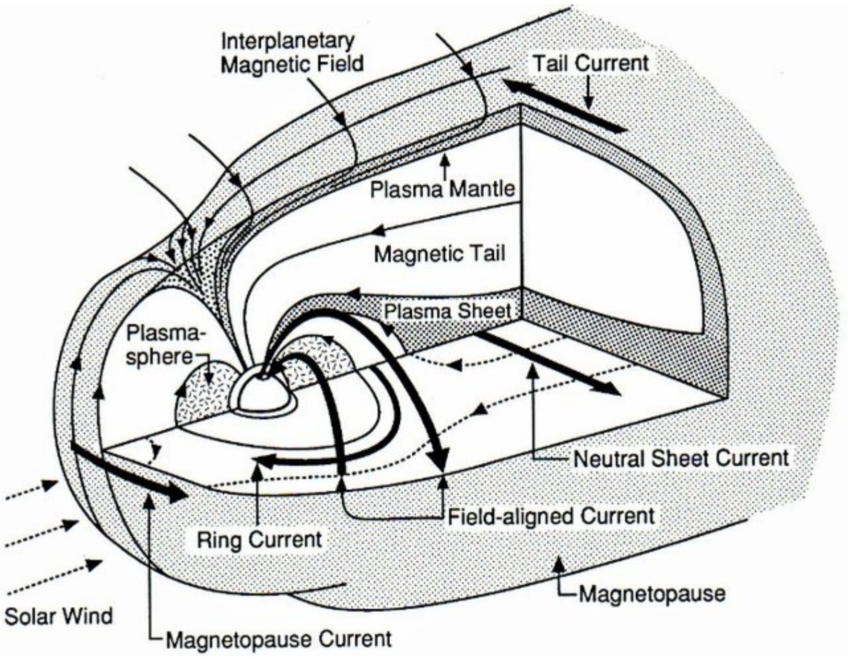


Figure 2: Structure of the magnetosphere. Source: [31]

Figure 2 shows the structure of the magnetosphere with phenomena evolving from the interaction of the geomagnetic field with the IMF and the solar wind, as provided by Kivelson and Russell [31]. Many geomagnetic phenomena at a smaller scale evolve from the interactions of Earth's magnetic field with its environment and the interaction with the IMF and the solar wind. On a large scale, the magnetic field is tailed in the direction of the side facing away from the Sun, with the charged particles from the Sun influencing different phenomena of the geomagnetic field. Among these phenomena are the ring current and electrojet at the equator, as well as FACs found in polar regions. These phenomena are still subject to research where additional data and the increased coverage can help in their analysis and understanding of the coupling between these.

Data from many sources is available for modeling the geomagnetic field, e.g., ground observatories, air and marine campaigns, and space-based measurements [46]. The positioning and distribution of geostationary observatories are constrained by their fixed locations on the surface of the Earth, leading to an uneven distribution across the globe that is weighed toward the Northern Hemisphere and dominated by locations on the mainland. These observatories provide consistent, long-term observations of the Earth's magnetic field and, with their static position, are especially well-suited to measure variations over time. In general, stationary sensors can detect varying fields well, while moving sensors, such as satellites, are better positioned in terms of altitude and are well-suited for detecting static fields and variations over time.

On the other side, space-based measurements onboard LEO satellites provide very good coverage and distribution of their measurements. This enables modeling efforts for the geomagnetic field as satellite data covers remote areas and oceans well. The utilization of high-precision data from past missions such as Ørsted and CHAMP, along with the ongoing high-precision mission Swarm, has led to an improvement of the geomagnetic field models. The mysteries of the Earth's magnetic field can be explained with an ever-growing resolution [23]. This resulted in highly sophisticated empirical models like the IGRF [2], Mag.num [51], Kalmag [26], or CHAOS-7 model [17], amongst others. The CHAOS-7 model is the currently most-used model in LEO and consists of three main components, namely the core, crustal, and large-scale magnetospheric fields [22]. Among others, the trend and changes of the magnetic field, the impact on climate, different time-varying and small-scale phenomena, space weather and hazard prediction, or the magnetosphere-ionosphere coupling are subjects to ongoing research. This work aims to provide additional space-based measurements by leveraging the possibility of calibrating platform magnetometers from non-dedicated missions to accompany the data of high-precision satellites.

3 SATELLITE MISSIONS TO SENSE THE EARTH'S MAGNETIC FIELD

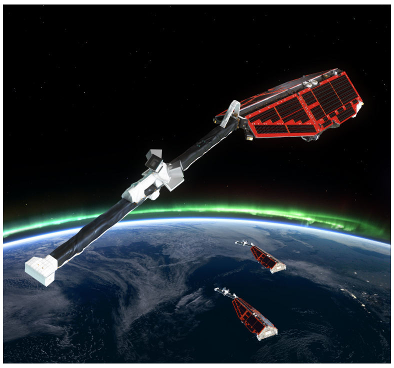


Figure 3: The Swarm satellite mission consists of three satellites. Source: [1]

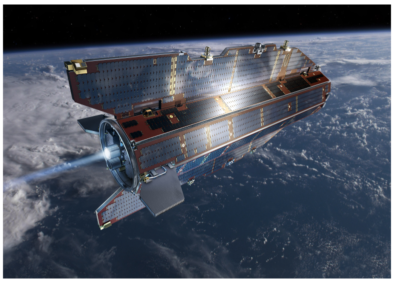


Figure 4: The GOCE satellite mission consists of a single satellite. Source: [2]

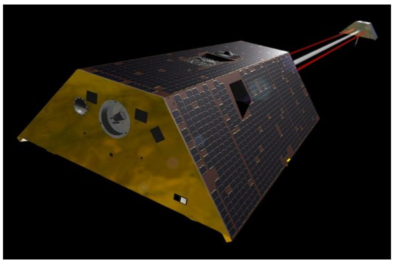


Figure 5: The GRACE-FO satellite mission consists of two satellites following each other. Source: [3]

This chapter will give an overview of the three main satellite missions used in this dissertation. Figures 3 to 5 contain an illustration of the three satellite missions in orbit. First, the current high-precision magnetic satellite mission Swarm will be presented. Then, the two non-dedicated satellite missions, GOCE and GRACE-FO, will be introduced, from which the platform magnetometer measurements have been calibrated.

3.1 SWARM

By now, there is a history of high-precision satellite missions dedicated to measuring the geomagnetic field. Between 1999 and 2004, magnetic field data from the Ørsted mission [42] are accessible. The CHAMP satellite mission [50] orbited from 2000 to 2010, followed by a gap from 2010 to 2013, during which no high-precision mission measured the magnetic field. Since 2013, the Swarm constellation [19, 45] has been delivering high-precision measurements, illustrated in Figure 3. The Swarm mission is the current high-precision magnetic satellite mission, in orbit since the 22nd of November 2013 [45]. A high-precision measurement is defined by a co-measurement of the absolute and vector magnetic flux density with very low noise. All these missions had specially designed architectures and star trackers for attitude estimation to acquire highly accurate magnetic measurements. Several properties of the Swarm mission make it a successful mission that has provided many insights into the geomagnetic field.

The Swarm mission consists of three satellites flying in a specific constellation. Swarm A(lpha) and C(harlie) fly side-by-side with a separation of about 1.4° longitude at the

---

[1] `https://www.esa.int/ESA_Multimedia/Images/2012/03/Swarm`

[2] `https://www.esa.int/ESA_Multimedia/Images/2008/03/GOCE2`

[3] `https://www.eoportal.org/api/cms/documents/163813/6584772/GRACE-FO.jpg/ad2ce666-d3e7-8c9f-3069-f48f0599bab4?t=1666999711989`

equator with an altitude of approximately 462km in a near-polar orbit with about 87° inclination. Swarm B(ravo) is flying at a higher altitude of approximately 511km with an inclination of about 88°. This constellation yields several advantages, like the ability to measure along-track differences as two satellites fly side-by-side, with additional coverage of altitude, space, and time by the Swarm-B satellite flying in a different orbit. The configuration of the constellation was temporarily altered through maneuvers undertaken to acquire insights from specific constellations, e.g., Swarm A and B following each other with a separation of only 2 seconds on the same track. The mission was initially designed for four years of operation and had its tenth anniversary in space on the 22nd of November 2023, and it is expected to continue.

The design and architecture of the satellites are targeted to acquire clean and high-precision magnetic measurements. Apart from other instruments onboard the Swarm mission, like the electric field instrument, Langmuir probes, and others, the main interest lies in the magnetic instruments of this mission. The body of the satellite has been built so that sources of magnetic disturbance are placed far away from the measuring instruments, and general care is taken for magnetic cleanliness, e.g., avoiding conducting loops in the wiring of the satellite. The magnetometer instruments are placed on a boom that is about 4m long, far away from the satellite body. On the tip of the boom, the absolute scalar magnetometer (ASM) instrument is placed, which is an optically pumped Helium proton magnetometer providing an accuracy of < 0.3 nT [24]. The ASM is essential for the precision of the Swarm satellites as it gives an absolute measurement of the magnetic field needed to calibrate the vector field magnetometer (VFM) measurement, a fluxgate magnetometer. While the ASM produces a magnitude measurement of the magnetic flux density, the VFM measures the magnetic field with three directional orthogonal components as a vector. The magnitude of the vector readings is then adjusted with the values from the ASM readings. The X, Y, and Z measurements of the VFM magnetometer measure along the direction of flight, laterally to the direction of flight, and vertically to the direction of flight, respectively. This gives the possibility to gain insights into the shape of the magnetic field and the spatial distribution of various geomagnetic phenomena. The VFM is placed on a so-called optical bench located at the center of the boom and measures with an accuracy of < 0.5 nT 1σ-accuracy [20]. The optical bench is a platform that hosts the mounting points of the VFM and the star tracker cameras used for attitude determination. This unique architecture enables the highest precision utilizing the VFM measurements with good attitude determination for the vector components of the VFM located on the same stiff platform.

The satellites themselves went through on-ground calibration and characterization of the different instruments and their interaction with the behavior of the satellite system. For this, the assembled satellites were placed in a large Helmholtz coil where the external field, i.e., the geomagnetic field, had been removed. Then, an extensive study of the magnetic behavior of the satellite during different activations was conducted. This includes determining scale factors, offsets, and Euler angles of the magnetometer itself but also activations of the magnetorquers, solar array currents, battery currents, and many more activations. This calibration was later adjusted after launch with slight corrections. With such preparation before launch, the magnetic understanding of the Swarm satellites is well established, which allows for the highest accuracy. Such a calibration is only possible on-ground and with a significant time slot dedicated to it. This is different for non-dedicated missions where the magnetic measurements did not have a high priority and where such a calibration is impossible to conduct after launch. Nevertheless, the ASM of Swarm C stopped working shortly after the launch. In addition, during the operation of the mission, an influential factor of the sun in combination with the zenith angle towards the sun was found, and a correction was introduced to counteract this. Nonetheless, the Swarm mission is considered a huge success among the community and has provided numerous insights into the geomagnetic field.

3.2 GOCE

The Gravity field and steady-state Ocean Circulation Explorer (GOCE) [18, 14] satellite mission was launched on the 17th of March 2009 and ended on the 11th of November 2013. Figure 4 contains a visualization of the mission in space. The geodetic mission flew at an altitude of about 268 km, which is relatively low for LEO satellites. The satellite flies in a near-polar, sun-synchronous orbit with an inclination of 96.7° and a repeat cycle of 61 days, ensuring uniform coverage of the entire Earth within this period. The primary mission goal was to measure the Earth's static gravity field with high accuracy, leading to the design decision of the relatively low altitude.

The orbit altitude has huge implications for the mission. Because of the higher air density and the resulting drag, the satellite was equipped with fins to keep the satellite stable. In addition, the fuel usage of the thrusters is increased compared to satellites at a higher altitude to keep the satellite in its orbit. Increased air drag and need for propulsion are also reflected in a relatively short mission length of about four years. The main instrument of the mission is a large accelerometer located at the center of the satellite. A very steady flight of the satellite and a precise position and attitude estimation were needed to achieve the mission goal.

The satellite's significance within the geomagnetic community stems from several advantages of the satellite mission's design. The satellite carries three platform magnetometers manufactured by Billingsley Aerospace & Defense of type *TFM100S* onboard, which have been used to gather measurements of the Earth's magnetic field. The good position determination and the very stable flight ensure accurate location and orientation assignments of the measurements. In addition, the low altitude is interesting for studies of Earth's crustal magnetic field as other high-precision satellite missions like Swarm, or previously CHAMP and Ørsted, fly at higher altitudes. The GOCE mission operated between 2009 and 2013, an important time range for space-based measurement of the Earth's magnetic field. The CHAMP mission ended on the 19th of September 2010, while the Swarm mission launched on the 22nd of November 2013. No high-precision space-based measurements of the Earth's magnetic field exist for this period [1, 32]. Especially for this period, the magnetic measurement data of the GOCE mission can play a crucial role in bridging this gap.

3.3 GRACE-FO

The Gravity Recovery and Climate Experiment-Follow-On (GRACE-FO) [33] satellite mission was launched on the 22nd of May 2018 and is expected to be ongoing for several years from today [34]. The geodetic mission consists of two satellites in a string-of-pearl constellation, where the satellites follow each other at a distance of about 220 km, as can be seen in Figure 5. Hereby, the first satellite is turned by 180° to face the following second satellite for the main instrument of the mission to be targeted at each other. The satellites fly in a near-polar orbit with 89° and a repeat cycle of 30 days at an altitude of about 482 km.

The GRACE-FO mission carries a Laser Ranging Interferometer (LRI) with which laser beams are sent from one satellite to the other. By analysis of the traveling time of the laser, conclusions about the gravity field and mass changes of the Earth can be made. In addition, accelerometers and GPS antennas support the LRI in positioning and detection.

Although the primary goal of the mission is geodetic, each GRACE-FO satellite carries two fluxgate magnetometers onboard for attitude control, of which only one is turned on, as the other is on standby as a redundancy if the first instrument experiences a failure. Utilizing the same foundation as the Swarm and CHAMP satellites, a certain level of magnetic cleanliness in the architecture is preserved, which is beneficial for the quality of the results. In addition, the special constellation is also interesting for space science, as the second satellite revisits the same position as the first satellite with a delay of about 29 seconds. As an ongoing mission, it benefits from continuous data availability, a high data rate of 1 Hz, and resolution, collectively contributing to its scientific significance. With the Swarm satellite operating since 2013, the whole mission period of the GRACE-FO mission overlaps with the Swarm satellite. This fact makes it possible to analyze conjunctions where the orbits of the two satellites are close together.

3.4 Platform Magnetometers

Figure 6: The Billingsley TFM100S fluxgate magnetometer is used as a platform magnetometer onboard the GOCE and GRACE-FO satellite missions. Source: [4]

As emphasized, at the core of this work are the magnetometric sensors of the presented non-dedicated satellite missions GOCE and GRACE-FO. These magnetometers are mounted on the satellite platform and are, therefore, also referred to as platform magnetometers. Both satellite missions use magnetometer sensors provided by Billingsley Aerospace & Defense of type TFM100S, depicted in Figure 6. The specifications can be found in the data sheet[5], of which the key elements are the measurement range with $\pm 100\ \mu$T and the root mean square noise level of the instrument with 100 pT. This means that both sensors provide a theoretical accuracy of well below 1 nT. They are part of the AOCS and play a crucial role in determining the orientation of the satellite by comparing the measurement of the sensors with the known structure of Earth's magnetic field, e.g., the location of the North Pole. In the GRACE-FO satellites, the platform magnetometers are placed at the front of the satellite close to the LRI, as far away as possible from the magnetorquer bars placed in the back of the satellite. In the GOCE satellite, there are three platform magnetometers placed side-by-side in the center of the body. With their usage as platform magnetometers, several challenges arise when attempting to calibrate these magnetometers for scientific purposes.

Fluxgate Magnetometer These sensors are fluxgate magnetometers that deliver a vector measurement with three orthogonal magnetic field components while having no absolute measurement of the magnetic field amplitude. Figure 7 contains a 3D visualization of the general design of a fluxgate magnetometer by Auster et al. [5]. The principal idea of a magnetometer is to convert the magnetic flux density signal into an electric current to quantify the magnetic signal [48]. The core component of a fluxgate magnetometer is a ferromagnetic core, typically made of a highly permeable material, here pictured by

---

[4] `https://magnetometer.com/products/fluxgate-magnetometers/tfm100s/`

[5] `https://magnetometer.com/wp-content/uploads/TFM100S-Spec-Sheet-February-2008.pdf`, accessed: 06.01.2024

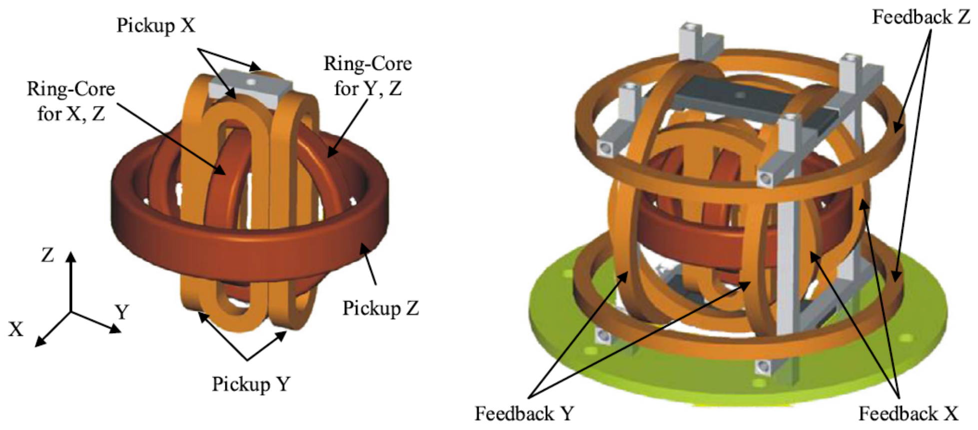


Figure 7: Working principle of a fluxgate magnetometer. Source: [5]

the ring cores. The system consisting of pickup coil and ring core measures the magnetic field. An opposing field is applied with the feedback coils so that the pickup coils measure a zero field. The current flowing through the feedback coils then correlates with the measured magnetic field. This is also called coil compensated fluxgate magnetometer, as the external field is compensated for the measurement technique to be applied.

Due to their working principle, fluxgate magnetometers need calibration as they provide a relative measurement of the magnetic field components in the form of an electric current signal with arbitrary amplitude. Both non-dedicated satellite missions do not carry an additional ASM that could provide an absolute measurement of the magnetic field amplitude. Therefore, the calibration of the fluxgate magnetometers relies on an external reference that is, in our case, provided by the utilized reference model. The satellite engineers already applied an initial calibration that is used for navigational tasks but is insufficient for the scientific usage of this data, rendering a more precise calibration necessary. The reference model, mainly consisting of the CHAOS-7 model, is heavily carried by the Swarm mission, which can provide absolute measurements of the magnetic field. The reference model is then evaluated for the positions of the non-dedicated satellite missions to provide a magnitude estimation of the measured magnetic field components. To sum up, while the platform magnetometer data can accompany high-precision data with additional coverage, their calibration would not be possible without a high-precision satellite mission in space that delivers the anchor points for the calibration.

The electric current signal of the fluxgate magnetometers is then converted through an analog-to-digital converter, where they are transformed into data. On a sensor level, the number of bits used to save the magnetometer readings is an important factor stemming from the satellite's operation. This can become a major limiting factor. For GOCE, the magnetometer readings are saved with a bit precision of 3 nT, and for GRACE-FO, with a bit precision of 1 nT. Another important factor is the rate of the measurements, which is dependent on the sensor itself but also on other properties of the satellite. When analyzing magnetometer readings taken onboard satellites, the position and attitude of the satellite are crucial for interpreting the readings. Therefore, the usability of the measurements also strongly depends on other components like the star camera or accelerometers, which are combined within the AOCS systems of the GOCE and GRACE-FO satellites to provide positional data with high accuracy. This also means that when these sensors face difficulties, resulting in missing or unusable position determination, the magnetometer readings will become unusable as a consequence of this. In addition, the time resolution of the position determination dictates the use of the magnetometer measurements as they act as the anchor point when creating a dataset of magnetic measurements of the satellite.

Sensor Measurement In a larger picture, various external factors play a role in the measurements of platform magnetometers. As they are mounted within the satellite body, they are exposed to artificial disturbances and influences originating from the satellite. Many influencing mechanisms can play a role, the most prominent of which are artificial magnetic fields, temperature-dependent behavior, and current-induced magnetic fields. These stem from the working principle of fluxgate magnetometers. The temperature influences the materials and the coils and thus has an impact on the measurements from the magnetometers. A calibration should consider the temperature of the magnetometer and adjust the measurements accordingly. The magnetometer itself does not provide a temperature measurement, hence an approximation needs to be found, often relying on a temperature measurement from a nearby subsystem. Another major factor is induced magnetic fields from electric currents flowing onboard the satellite. Following Ampère's Law, stronger currents produce stronger magnetic fields and thus potentially influence the magnetometer sensor readings more, which depends on the relative position and orientation between the inducing source and the sensor. Still, this means that the strongest currents have the highest potential to influence the magnetometer readings. On satellites, these are the solar arrays together with the battery currents, as well as magnetorquers inducing artificial magnetic fields or reaction wheels used to control the attitude of the satellite.

In addition, there can be cross-talk or overlapping between neighboring fluxgate magnetometers, which originates in their working principle of alternating applied magnetic fields. This can, e.g., happen in the GOCE satellite with its special positioning of the three magnetometers placed side-by-side with each other and measuring simultaneously in close proximity. Depending on the intensity and local position, there can be an influence of these neighboring magnetometers on each other that needs to be accounted for when calibrating magnetometer measurements.

Acquired Measurements The platform magnetometers from the GOCE satellite mission measure at an interval of 16 s. Over the whole mission, about 6.4 million data points with 984 features have been collected. For the GRACE-FO mission, the data is measured at a rate of 1 Hz, and as an ongoing mission, about 2.6 million data points are collected per month and satellite. Until the end of 2023, about 174.2 million data points with 97 features were accumulated. The data is provided by the Helmholtz Centre Potsdam (GFZ) and the Jet Propulsion Laboratory (JPL).

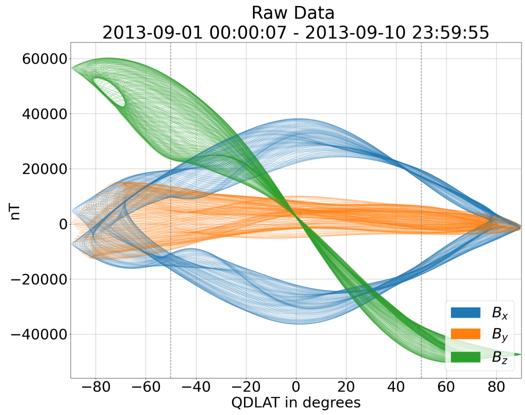


Figure 8: Raw magnetometer data as measured by the GOCE satellite in September 2013.

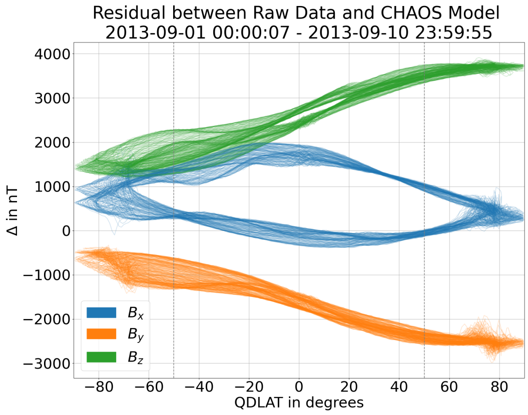


Figure 9: Raw Measurement data of the GOCE satellite with CHAOS-7 model values subtracted. Please note the different y-scale.

In the context of satellite observations and geomagnetic measurements, distinct frames play a crucial role in accurately representing and interpreting data. The satellite frame serves as a reference tied to the orientation of the spacecraft, aiding in the analysis of instrument readings and ensuring proper alignment. On the other hand, the North East Center (NEC) frame is an Earth-centred and Earth-fixed frame used commonly by magnetic field models and global models [16]. Data can be transformed between different frames by rotation, therefore the estimations of the geomagnetic field models are transformed to the satellite frame because the influences on the instruments and the modeling of the calibration of the platform magnetometer happen in the satellite frame so the calibration should be performed in the same frame. In addition, when evaluating geomagnetic phenomena, the quasi-dipole latitude is commonly used, defined as the latitude projection between the magnetic North and South Pole, which differ from the geographic poles [15]. Many geomagnetic phenomena are rather visible in dependence on the quasi-dipole latitudes defined by the geomagnetic field than in a geographic representation. These frames contribute to a comprehensive framework for studying and interpreting satellite-based geomagnetic measurements.

As an example of the measured data, Figure 8 shows the raw measurements of the GOCE satellite mission for a period of ten days in September 2013 in dependence on the quasi-dipole latitude. The three axes components of the measurement are shown with B_x aligned with the flying direction, B_y sidewards of the flying direction, and B_z as the horizontal direction. It can be seen that for similar quasi-dipole latitudes, similar measurement values are measured by looking at the saturation of the drawn, transparent lines. As the satellite flies in a polar orbit and the B_x values are aligned with the flight direction, a separation of the blue lines is visible that stems from the ascending and descending orbits of the satellite flying either parallel or anti-parallel with the magnetic field lines of the Earth's dipole dominated magnetic field, respectively. On the right side, Figure 9 shows the residual between the raw measurements and the CHAOS-7 magnetic field model without any further calibration, note the change of the y-axis scale compared to the previous plot. It can be seen that the delta of the measurements is much smaller, but an adjustment is still needed for the measurements to receive a smaller residual. This gives the calibration task a visual representation. Furthermore, the high latitude area, below -60° and above 60° quasi-dipole latitude, shows a disturbed behavior. This originates from FACs not represented within the CHAOS-7 model but measured by the satellite as they are part of the natural phenomena within Earth's magnetic field.

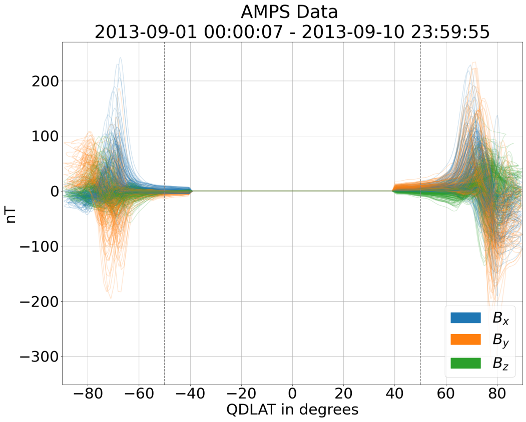


Figure 10: AMPS model data as evaluated for the GOCE satellite positions. The AMPS model is only valid above 40° quasi-dipole latitude.

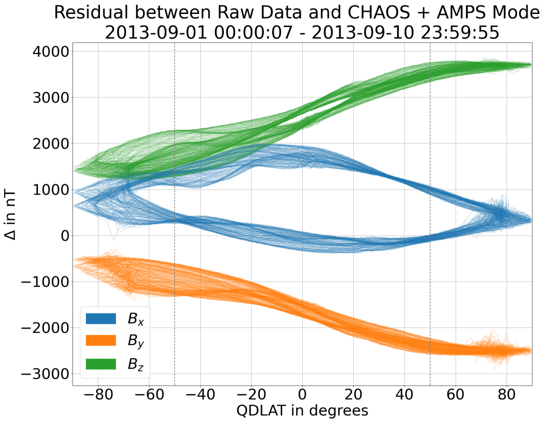


Figure 11: Raw Measurement data of the GOCE satellite with combined CHAOS-7 and AMPS model values subtracted. Please note the different y-scale.

Considering the average magnetic field and polar current system (AMPS) model that provides the average large-scale features of the FACs, the evaluation for the GOCE satellite positions can be seen in Figure 10. The model is only valid for high latitudes above 40° quasi-dipole latitude. In general, it can be seen that the magnitudes of the FACs are much smaller than the magnitudes of the geomagnetic field modeled by the CHAOS-7 model. Nevertheless, the global structure follows the disturbances visible in Figure 9. Combining the AMPS model estimates with the CHAOS-7 model estimates and subtracting them from the measured data leads to a picture like Figure 11. It can be seen that the disturbances in the high-latitude regions are significantly reduced, but the small-scale measured data remains noisy in this area. This is one of the challenges and describes the given situation of the measured data.

4 Application of Machine Learning in Satellite Magnetometer Calibration

In this section, an overview of the basics of ML is given, highlighting its relevance in the upcoming chapters, where the details of its application are explained. Following this, there's a summary of the work carried out in this thesis.

4.1 Relevant Fundamentals of Machine Learning

In recent years, the field of Machine Learning (ML) has gained significant attention [10]. With an ever-growing amount and rate of data collected in various fields, the need for efficient techniques to analyze this data and find patterns has grown alongside [60]. Hereby, the input data to the ML algorithms is as important as the algorithms and mathematical concepts used. This section provides a general overview of the most important concepts.

4.1.1 Classification and Regression

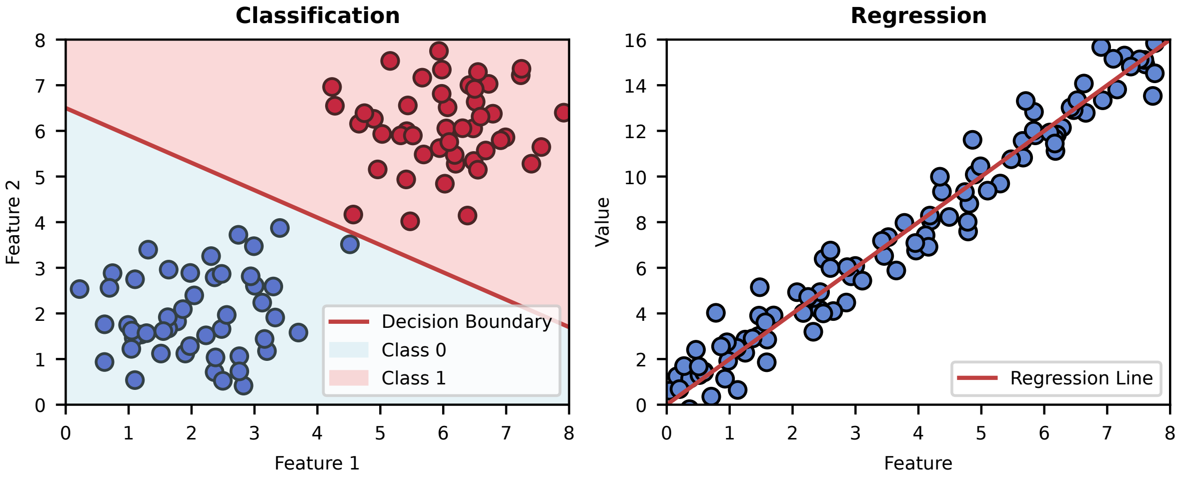


Figure 12: Comparison of classification and regression. Source: [41]

In general, there are two main tasks for ML applied to data: Classification and regression [36]. As visualized in Figure 12, classification attempts to achieve a categorization based on the information given, e.g., assigning a class to data points. Such categories can be binary, e.g., true and false or normal and abnormal, but a larger number of categories is also possible. On the other hand, regressions predict or estimate a numerical output variable, e.g., a regression line based on the available data points. This output variable is referred to as the dependent variable that is calculated using the input of the ML technique in the form of the independent variables [35]. In this work, I will focus on the regression as the expected output is continuous.

4.1.2 Supervised and Unsupervised Methods

Depending on the available data, one of the most important decisions is whether to use supervised or unsupervised methods in ML, which divides the available ML techniques into two categories. In supervised methods, the expected output or outcome is known for a given input set. Therefore, these techniques focus on enabling an ML model to estimate the expected outcome from the given inputs using a variety of algorithms, while afterward applying that model to further input data. Examples are linear or logistic regression, trees, forests, support vector machines, or neural networks of different kinds [11]. In contrast, for unsupervised methods, only the input is given without a clear expectation for the output. Here, the task is to find patterns within the data and present these as the outcome of the ML training for further usage or analysis. Examples include clustering techniques, Autoencoders learning to represent data structures at hand, or recommender systems based on the choices of other users [3].

The boundaries between the two techniques are fluid, e.g., with semi-supervised learning, some workflows operate in between these two categories. Techniques include deriving labels from some (partial) prior information, e.g., a small subset of labels or a grouping based on specific properties of the data, then applying supervised learning. In this work, the focus lies on supervised methods as a geomagnetic field model can be used as a reference model to generate the expected outputs, with a special focus on neural networks.

4.1.3 Training Process

The term of training a neural network evolves from the optimization process for the internal parameters of different supervised ML algorithms. When training an ML model, different data pairs with input x and expected output y are presented to the model. A predicted output $\hat{y}$ is made by a neural network that acts as a function f on the input x as can be seen in Equation (1).

$$\hat{y} = f(x) \tag{1}$$

The predicted output is compared to the expected output for every input pair, where the discrepancy between the predicted output and the expected output is defined as the error, with many possibilities to calculate this error. Among the most prominent is the mean absolute error (MAE), defined in Equation (2), and the mean squared error (MSE), defined in Equation (3).

$$MAE = \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{2}$$

$$MSE = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{3}$$

These errors are used to adjust the internal parameters of the ML model by a gradient-based optimization. The gradient is derived from the error of the presented samples and sets the direction for the change in the model parameters so that it will predict the presented values with a smaller error after adjustment. Depending on the ML algorithm, a different paradigm is used to adjust the internal parameters to incorporate the knowledge gained from the presented samples. This process is repeated with all available data points multiple times until convergence is achieved, resulting in the "training" of the ML model.

Multiple data pairs are predicted simultaneously in so-called batches to speed up the process, yielding computational benefits. When predicting the output in batches, there are as many errors calculated as there are data pairs in the batch. These errors are averaged to form the gradient of the batch, after which the internal model parameters are adjusted. Due to this averaging, the gradient summarizes multiple data points and is therefore more stable, e.g., against outliers. Additionally, this enhances the trustworthiness of the gradient, resulting in a higher possible learning rate. The learning rate is a hyperparameter in the training of ML models, as it adjusts the amount of adjustment towards the current gradient. As each gradient only represents a small subset of the data, the gradient is not directly applied to adjust the internal model parameters but instead multiplied by the learning rate before application, whereby the learning rate lies in the interval of $[0, 1]$.

4.1.4 Overfitting and Regularization

To check the performance of the model, the model can be evaluated on data pairs, producing output estimates, and the error towards the expected output can be analyzed. However, the model would be biased towards these data pairs if they were included in the training process. Therefore, before the training starts, the common practice includes splitting the dataset into a training dataset and a test dataset, with usual orders of 70%/30% up to 90%/10%, respectively [25]. With such a split, the model can be trained on the training dataset and evaluated on the test dataset whose data samples have not been seen by the model beforehand. This ensures a meaningful evaluation of the model's generalization capability on out-of-sample predictions.

With this split, a comparison between the train and test error can be conducted. Ideally, the model performs as well on the test dataset as the training dataset. Thus, the underlying distribution or pattern of the training dataset is well represented by the model, and the model shows the capability to generalize these results on out-of-sample data, which should follow the same distribution.

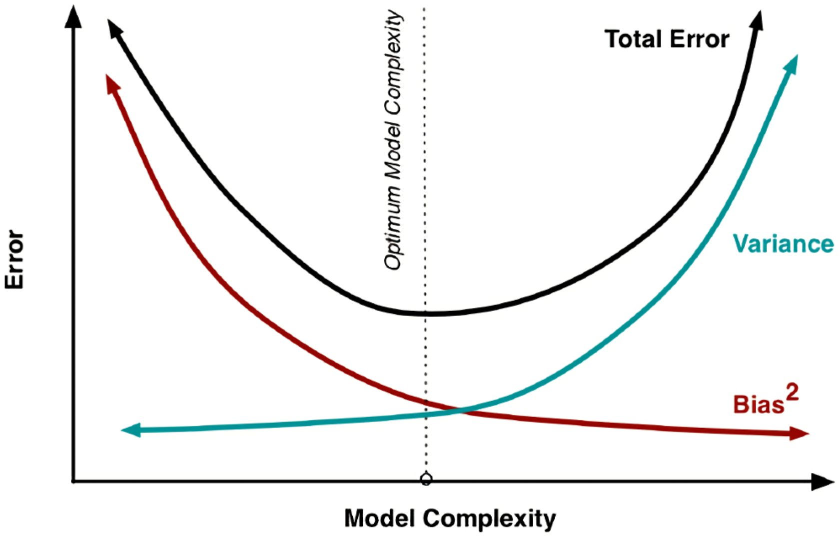


Figure 13: Illustration of the bias-variance tradeoff. Source: [41]

If the training error is significantly lower than the testing error, a phenomenon called overfitting can occur [21]. In that case, the model is capturing noise and detailed fluctuations of the data rather than the underlying patterns. This can, e.g., happen if the model was trained with too many iterations of the training data, thus specializing in the fine-granular patterns of the training data instead of the general data distribution of the training data. Another reason may be a model that is too large or too complex and can adapt to many small structures in the data. This is also known as the bias-variance tradeoff [7]: With increased model complexity, the model has a high variance, enabling it to adapt to the dataset very well until the adaption is too close and overfitting occurs, as can be seen in Figure 13 towards the right side where the total error increased. With a too simplistic model complexity, the model has a high bias, which means that a larger error is present because the model is too simple to explain the underlying distribution of the data, as can be seen towards the left side. Finding the correct model complexity is thus a task of reducing the bias-variance tradeoff, which consequently reduces the total error. This means that the model complexity has to be chosen so that the model complexity is neither too large nor too small for the task at hand.

Another way to handle overfitting is the use of regularization. Regularization is a technique that shifts the task of reducing model complexity to the optimization process of training the model. Here, a second goal of the optimization is introduced by extending the error term with another error term related to the model complexity. Examples of this are L1- or L2-regularization [43]. Here, the error term is extended by a secondary term that takes the sum of the weights w or the sum of the squared weights into account, respectively. This term is then weighted with a parameter λ that regulates the amount of reduction of the weights in the network. Equation (4) shows how the MSE from Equation (3) would be extended with a L2-regularization.

$$MSE_{reg} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + \lambda * \sum_{j=1}^{m} w_j^2 \tag{4}$$

With this term in the optimization process, the optimization not only reduces the error of the model for the prediction task but also reduces the weights within the model, thus reducing the complexity of the model by enforcing the elimination of unnecessary weights.

4.1.5 Machine Learning Methods

The framework, as described previously, is mostly the same for various algorithms and ML techniques. The model can consist of different architectures and statistical principles, but the learning process on data samples is generally similar. For this work, the focus will lie on feed-forward neural networks, which have primarily been used.

Neural Networks A feed-forward neural network (NN) consists of multiple neurons where each neuron's output y is defined in Equation (5),

$$y = a(w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \ldots + w_m * x_m + b) \tag{5}$$

where x_1 to x_m are different features of a data point, which get multiplied by their respective weight w_1 to w_m, with a bias b added finally. The sum of these products is then fed into an activation function a, explained in the next paragraph. NNs consist of many such neurons, which will be assigned different weights. Multiple neurons in parallel, which means that they work on the same input x_1 to x_n, are called a layer of neurons. These layers of neurons can also be stacked, meaning another layer takes the outputs y_1 to y_p of the first layer as the inputs for the neurons in the next layer. Each neuron of a layer can also be seen as a feature, as it is a new combination of the inputs that acts like a machine-crafted feature.

The initial formula is very similar to a linear regression formula, and by stacking these neurons, there would be no non-linearity added to the system, limiting its expressiveness in the processes it could explain. Therefore, the sum of the products is fed into the non-linear activation function. There are many activation functions available that act in different ways. One of these is the Rectified Linear Unit (RELU) [39], defined in Equation (6).

$$r(x) = max(0, x) \tag{6}$$

Negative values become 0, while positive neuron calculations will get forwarded into the next layer. This enhances the expressiveness of the NN and gives the ability to model non-linearity. The Exponential Linear Unit (ELU) [12] used in this work has a similar function shape to the RELU but overcomes some shortcomings. Usually, all intermediate layers, also called hidden layers, before the final output layer use activation functions in their output. This is often not done for the final output layer, as a linear combination of the previous layer is desired.

The construct of a number of neurons, a number of layers, and the last layer of neurons, which constitutes the output, is called the NN or the model. In the beginning, all the weights within this system are initialized randomly. For the model to represent the desired function, e.g., mapping the satellite data input to the calibrated measurement, the weights need to be adjusted in a meaningful way, and this is called the training of the NN. Contrary to algorithms like the Least Squares approach, NNs are trained with the Backpropagation algorithm [52]. Therefore, each available data point in the training dataset is forwarded through the NN, and the current result of this calculation is compared to the expected result for this data point, also called the ground truth or the reference model. This error is then used as the gradient to adjust the weights and backpropagated through the NN to adjust the weights in the different layers so that the prediction would come closer to the expected result if the calculation were repeated. To speed this process up and not rely on single data points, this error is calculated on groups of data points, also called a batch, for which one common gradient is calculated, and then an adjustment to the network is made. The adjustment towards the calculated gradient can be controlled with the learning rate, which usually lies in the interval of [0,1] and is multiplied with the calculated gradients before they are applied to the weights. One iteration through all batches of available data points and adjustments to the NN is called an epoch. The NN will be trained for multiple epochs, and optimizers like the Adam optimizer [30] speed up the process by modifying the gradients with historical information. After this training process, the weights of the NN do not change anymore, and it is assumed that the NN represents the statistics of the data. Afterward, it can predict similar data to the one it was trained with.

4.1.6 Data Science

The underlying data used for the training of the model is equally as important as the model architecture. As the models are statistics-driven, they will learn what is present in the data. Thus, the data are equally as important as the used algorithms [40]. The combination of data processing and model training is commonly referred to as Data Science [13]. This holistic approach has recently gained more traction as including the data into the process with its form, content, and data preprocessing has a huge influence on the performance and architecture of the model. There are many subfields of Data Science, some of which will be explained here.

First of all, the data needs to be cleaned. Frequently, the initial dataset needs to undergo so-called cleaning. Unexpected data records are common when measuring data, e.g., from sensors. This can include many different aspects, such as outliers, missing data, data gaps, or wrongly measured data. In the preprocessing of the data, there are a variety of techniques to handle such data and thus improve the model's performance subsequently.

Outliers in the data can occur in several different forms. When looking at sensor data, these can be unrealistic readings with, e.g., extremely high or unexpected numbers. Such data points interfere with the training process as the model computes a prediction based on these values. A large error is the result, as the reading was faulty and did not originate from the data distribution. This large error is then translated into a large gradient that backpropagates to the model's parameters and thus interferes with the model's performance by adjusting the internal parameters based on faulty data. To deal with outliers, outliers can be corrected or filtered out of the data so they are excluded from the training process and marked during the prediction process, e.g., with the usage of flags.

Missing data can occur in two forms: When looking at a time series of data, there could be a complete gap in any recording. Such data is not recoverable, and only probabilistic remodeling techniques exist, if at all. In the use case of satellite measurements, this is considered unrecoverable data. The other case involves data where only parts of the data are missing. In a system of sensors, this could be a subsystem not delivering measurements for a certain period of the data records. Such data can be filled or reconstructed by filling techniques. One approach includes filling such missing data with the mean value of the corresponding feature, which acts as an estimate for the missing value and thus renders these data points useful again because they can be included in the training data. Other methods include reconstruction techniques based on other available features that allow drawing conclusions and reapplying patterns, thus reconstructing missing features based on available features. For magnetometer calibration, this is especially applicable if auxiliary features contain gaps. Missing magnetometer measurements or positional data mean that these data points cannot be restored.

For the magnetometer calibration, data from a variety of subsystems onboard the satellites needs to be merged. Additional external sources like indices and space weather parameters are also included in this. This poses a challenge as the systems have different measurement intervals, different time stamps, and so on. For the calibration of platform magnetometers, the attitude and position determination are crucial, therefore the timestamps defined by these are used as the common ground to interpolate other data onto. Thus, different sources of data can be merged and used as a common dataset for the calibration.

Another important step in Data Science is the preprocessing of data. This includes, e.g., scaling the data to a common interval. Raw data, as measured by sensors, has a variety of measuring intervals, which has an implication on the model training. The intervals can be vastly different, ranging from values behind the decimal point up to orders of magnitude larger. In the calculation of the error and the subsequent gradient, larger value ranges will receive a higher weighting within the adjustments of neural networks. Therefore, all available features will be scaled to a common interval so the training of the neural network can be conducted with an equal weighting of the features.

Another essential step in the data preprocessing is preparing the data for the model training. The model learns the statistical distribution of the data that is presented during the model training. Therefore, addressing selection bias, where the training dataset is not a representative sample of the overall population, is crucial for the learning ability of the model. Selection bias occurs when certain characteristics in the data distribution are overrepresented or underrepresented in the dataset, potentially leading to inaccurate or biased model predictions. If the training dataset does not accurately represent the underlying problem, the model may not perform well on new, unseen data. As an example, in the calibration of platform magnetometers, filtering for magnetic quiet conditions is applied as the model is supposed to be trained on the underlying mechanisms within the satellite apart from external geomagnetical signals during magnetic active conditions. This stems from the fact that the CHAOS-7 reference model does not contain short-lived and fluctuating signals during active times, which are measured by the satellite's magnetometer. Thus, the data needs to represent the task to learn the adjustments needed for the satellite as a system.

For the generalization ability of neural networks, it is crucial that the training dataset contains the same data distribution as the whole application set of the model. In classical analytical calibrations, a linear regression algorithm is applied that will naturally generalize well by linearly expanding the predictions for out-of-sample distributions like in high-latitude regions. Therefore, the data is limited to a low- and mid-latitude range as these ranges do not contain FACs that are not fully modeled in the reference model to ensure a stable training of the model. Contrary to this, for the calibration of platform magnetometers using neural networks, all quasi-dipole latitudes need to be included, as the high-latitude regions contain out-of-scope values when compared to the low- and mid-latitude regions. Therefore, data from all quasi-dipole latitudes needs to be included in the training, which results in the challenge of calibrating the platform magnetometer data with natural signals left in the measurements.

Finally, the subsequent goal of this work is to publish the calibrated dataset for use by the community. Several steps must be considered for publication. For the publication, the

FAIR (Findability, Accessibility, Interoperability, and Reuse of digital assets) principles are followed. In our case, this means that our published data uses digital object identifiers (DOI) and the common data format (CDF) file standard provided by the National Aeronautics and Space Administration (NASA) while being publicly accessible. In addition, to render the dataset as useful as possible, datasets that are as complete as possible are published. This means that any filtering or selection of data is only applied during the model training, while for the generation of the final dataset, all data is used. In order to do this, flagging is used to give meta information about which data was included in the training and which data points have been filtered for a particular reason during the training of the calibration model. Therefore, the final result contains a dataset that, apart from some data gaps contained in the satellite's measurements, contains a temporally coherent series of measurements of the satellite's calibrated magnetic values that can be used for further analyzing the geomagnetic field.

4.1.7 Physics-informed Neural Networks

Recently, there has been an increasing focus on a specific subset of ML known as Physics-informed neural networks (PINN) [61]. As the application of neural networks and ML has expanded in various domains, data scientists have encountered challenges where the results of the trained models may violate physical laws in different domains. There are many problems at hand where solely data-based models have limitations. In response to this, PINNs have emerged to incorporate essential physics-based rules into the model results, setting new standards for accuracy while ensuring agreement with the underlying physical principles.

PINNs represent a powerful hybrid approach of ML and physical principles, designed in such a way as to integrate domain knowledge into the learning process of ML models. There are various ways to achieve this, e.g., adjusting the training process, changing the model architecture, or further processing the output. By bridging the gap between traditional physics-based models and modern ML techniques, PINNs have been shown to enhance the accuracy and interpretability of the results [27].

In the field of geosciences, ML and PINN have seen a rise in usage as well. Smirnov et al. [53] have introduced the NET model to predict the electron density of the topside ionosphere. Here, the results of the ML methodology are used in a physically governed equation system to form the result of the approach. Zhelavskaya et al. [62] used a Kalman filter to combine the outputs of a physics-based model with the outputs of an NN-based model to retrieve a good estimation for predicting global plasma density. This dissertation adds a contribution to the automatic calibration of platform magnetometers in the field of geomagnetism by utlizing PINNs to calibrate the platform magnetometers onboard non-dedicated satellites.

4.2 METHODOLOGICAL APPLICATION TO PLATFORM MAGNETOMETERS

This chapter shall give a short summary of the dissertation. First, the general aim of the work will be given, followed by a chronological summary of the works conducted alongside the dissertation that have been published.

I calibrate and characterize platform magnetometer data of different satellite missions utilizing ML techniques. Therefore, satellite systems carrying platform magnetometers flying in LEO are looked at, mainly the GOCE and GRACE-FO satellite missions. The platform magnetometers measure the Earth's magnetic field and are mounted only roughly calibrated for navigation purposes on many satellites. For scientific usage of this magnetic data, the artificial disturbances introduced by the satellite's payload systems need to be removed from the recorded data. For this, supporting data from a variety of sensors mounted on the satellite can be used, e.g., information about the activation of other subsystems. This includes the activation of magnetorquers, battery currents, solar array currents, or thruster activation.

By applying ML techniques, the measured magnetometer signal is adapted for artificial disturbances from other satellite payload systems. The proposed non-linear regression can automatically identify relevant features as well as their crosstalk, allowing for a wider range of available inputs. This approach reduces analytical work for the calibration of platform magnetometers, thus leading to faster, more precise, and easily accessible magnetic datasets from non-dedicated missions. In fact, when mentioning calibration, the calibration and characterization is meant within this work. The calibration itself is adjusting the values of the raw measurement, e.g., by scaling, offsetting, or rotating, while the characterization describes the correction for external influences on the measurements of the sensor, e.g., correction in dependence of temperature. Since a neural network constitutes a non-linear regression, calibration and characterization are performed in the same model.

Combining all of the data into a common mapping procedure is challenging: it comprises different sampling rates, signal amplitudes, noise levels, and latency. Combining data of different sampling rates, signal amplitude, noise levels, and latency in one mapping procedure requires special care for data handling and inter-calibration to achieve an unbiased result uniformly valid over the globe. Hereby, the challenges described in the introduction are addressed and overcome.

After the initial results for GRACE-FO, a post-launch calibration of the platform magnetometers was conducted utilizing ML methods which come with some inherent beneficial properties for solving the present challenges [57]. Here, I proposed a feed-forward neural network architecture for a non-linear regression that is able to identify relevant features automatically, thus taking over the task of Feature Selection. By using as much information as possible about the satellite's payload systems, the ML model can generate higher-order features and identify crosstalk between the magnetometers. I showed that ML models are able to model the magnetic behavior of non-dedicated satellites post-launch. The GRACE-FO data contains an inherent time shift in certain parts of the dataset. For a given timestamp, not all available features correspond precisely to that timestamp. To address that some features must be adjusted along the time-axis to align with the specific timestamp, I suggest the use of a data-generating interpolation neuron, capable of co-estimating and applying this time shift during the training of the neural network. The evaluation has shown promising results, outperforming analytical calibrations.

Next, the GOCE satellite mission has been investigated, which is presented in Section 5 [38]. The GOCE mission, together with the acquired data, is presented in this publication. The difference to the previously calibrated GRACE-FO missions lies in a different architecture of the satellite, together with a lower precision and lower measurement rate of the magnetometer readings. On the other hand, for the GOCE satellite mission, many more features are provided that can be analyzed for the calibration. In this study, the used features have been selected by domain experts and the calibration has been conducted using a linear regression on hand-crafted features. The evaluation showed that this calibration is stable over time and produces satisfying results that are able to be used in a scientific setting, as was shown for two magnetospheric phenomena.

Next, I further developed and applied the ML-based non-linear regression method on the GOCE satellite, which is presented in Section 6 [58]. The proposed approach effectively addresses challenges in the dataset, like the lack of ground truth, Feature Selection, and Feature Combination. An extensive preprocessing pipeline was developed to handle the huge variety in the measured data of the GOCE mission which includes outlier handling, missing data, data gaps, or flagging of incorrect data. The application to a different satellite highlights the advantages of the automatic approach in its applicability. The results show how the ML approach surpasses existing methods and demonstrates a reduction in the residuals compared to the reference model CHAOS-7. In addition, its practical application possibilities are showcased, which include the detection of geomagnetic storms or a derivation of a lithospheric field model solely from calibrated GOCE calibrated platform magnetometer data.

Finally, in Section 7, an extension to the previous ML approach is presented that is applied to both satellite missions, GOCE and GRACE-FO [59]. In this work, I combine the advantages of the analytical bottom-up approach with the ML-based top-down approach by incorporating the first-principle physical law of Biot-Savart into the neural network, creating a PINN. This constraints the previously purely statistics-driven model with physical laws, thus ensuring a more stable calibration. In addition, the CHAOS-7 reference model is combined with the AMPS model, improving the reference model of the calibration, especially for the polar regions. This improved on the challenge of a lack of ground truth for the polar regions. After the application of this approach, two new calibrated magnetic datasets for the two satellite missions are published that show how they could overcome the shortcomings of the previous approach for certain constellations. In the evaluation, the remaining residual has been analyzed, as well as the application in the analysis of geomagnetic phenomena utilizing calibrated platform magnetometer data. For example, it could be shown that the majority of residual values of the calibrated GRACE-FO dataset are in the same range as the high-precision satellite mission Swarm. Finally, using a PINN enables the analysis of the learned magnetic behavior of the satellite within the model by investigating the learned, internal parameters of the Biot-Savart neuron incorporated into the PINN.

PUBLICATION-BASED CHAPTERS

Sections 5 to 7 contain the publication-based chapters.

Michaelis *et al. Earth, Planets and Space* *(2022) 74:135*
https://doi.org/10.1186/s40623-022-01691-6

Earth, Planets and Space

TECHNICAL REPORT **Open Access**

Geomagnetic data from the GOCE satellite mission

I. Michaelis[1*], K. Styp-Rekowski[1,2], J. Rauberg[1], C. Stolle[3] and M. Korte[1]

Abstract
The Gravity field and steady-state Ocean Circulation Explorer (GOCE) is part of ESA's Earth Explorer Program. The satellite carries magnetometers that control the activity of magnetorquers for navigation of the satellite, but are not dedicated as science instruments. However, intrinsic steady states of the instruments can be corrected by alignment and calibration, and artificial perturbations, e.g. from currents, can be removed by their characterisation correlated to housekeeping data. The leftover field then shows the natural evolution and variability of the Earth's magnetic field. This article describes the pre-processing of input data as well as calibration and characterisation steps performed on GOCE magnetic data, using a high-precision magnetic field model as reference. For geomagnetic quiet times, the standard deviation of the residual is below 13 nT with a median residual of (11.7, 9.6, 10.4) nT for the three magnetic field components (x, y, z). For validation of the calibration and characterisation performance, we selected a geomagnetic storm event in March 2013. GOCE magnetic field data show good agreement with results from a ground magnetic observation network. The GOCE mission overlaps with the dedicated magnetic field satellite mission CHAMP for a short time at the beginning of 2010, but does not overlap with the Swarm mission or any other mission flying at low altitude and carrying high-precision magnetometers. We expect calibrated GOCE magnetic field data to be useful for lithospheric modelling and filling the gap between the dedicated geomagnetic missions CHAMP and Swarm.

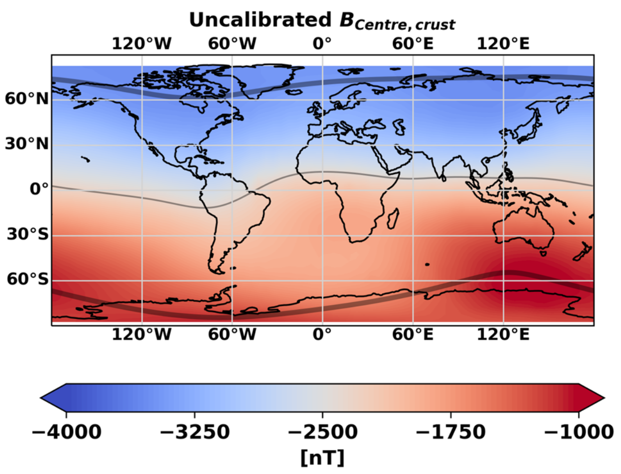
Keywords: Earth's magnetic field, Geomagnetism, Ionospheric currents, Magnetospheric ring current, Satellite-based magnetometers, Platform magnetometers, GOCE

---

*Correspondence: ingo.michaelis@gfz-potsdam.de

[1] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
Full list of author information is available at the end of the article

Graphic Abstract



Introduction

In the last two decades, low Earth orbiting (LEO) satellites have been available for accurate measurement of the geomagnetic field using dedicated instruments, e.g. missions like CHAMP (CHAMP 2019) and Swarm (Olsen et al. 2013). However, there is a temporal gap of about 3 years between these dedicated missions.

In addition, single missions can only provide limited coverage in local time at a given time. Enhancement of simultaneous local time coverage is given by multi-mission constellations. To this aim, magnetometer data from missions like CryoSat-2 (Olsen et al. 2020), GRACE (Olsen 2021), and GRACE-FO (Stolle et al. 2021) has been characterised and calibrated and made publicly available. Some of those missions can fill the gap between the high-level missions CHAMP and Swarm from 2010 to 2013, e.g. CryoSat-2 and GRACE, others can fill the gap in magnetic local time (MLT) distribution, such as GRACE-FO. An overview of scientific and platform magnetometer (PlatMag) missions is shown in Fig. 1. Stolle et al. (2021) have shown that large-scale field-aligned currents can be derived from GRACE-FO, as well as equatorial ring currents. The standard deviation of the residuals of those datasets compared to high-level geomagnetic models like CHAOS-7 (Finlay et al. 2020) have been reduced to values well below 10 nT for geomagnetic quiet times, depending on the mission. This report introduces a calibrated magnetometer data set from the Gravity field

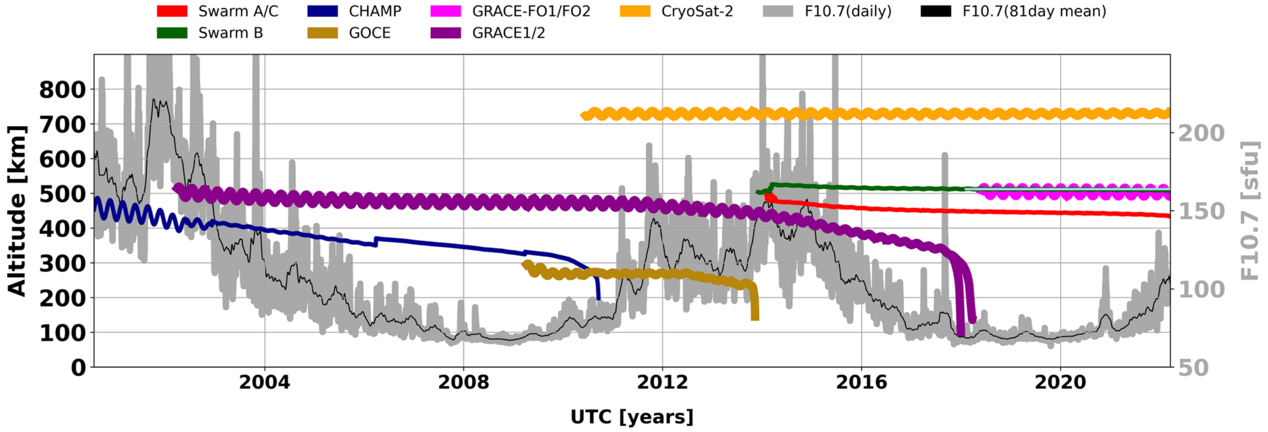


Fig. 1 Overview of the two satellite missions dedicated to geomagnetic measurements CHAMP (blue line) and Swarm (red and green lines) and a selection of missions carrying platform magnetometers at their respective altitudes. Also shown is the F10.7 solar irradiation index as an indication of solar activity (grey with mean as black solid line, right axis)

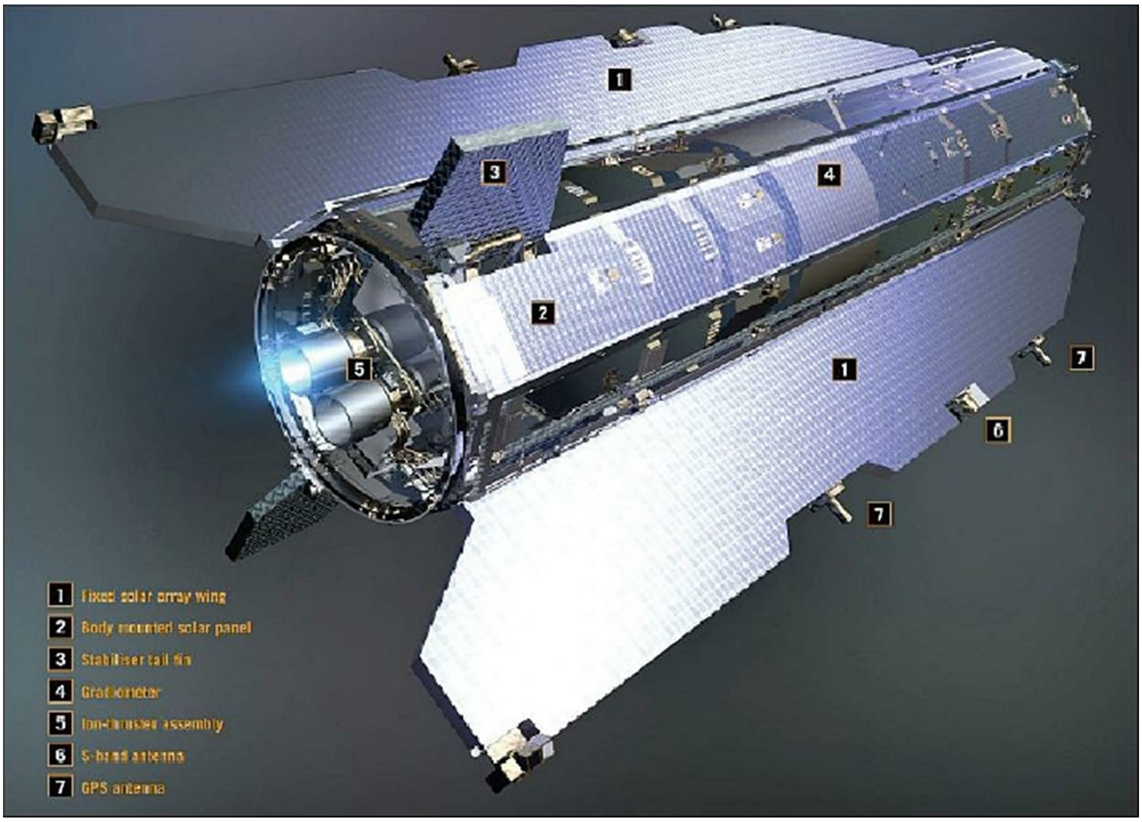


Fig. 2 Schematic view of the GOCE satellite. (Credits: ESA)

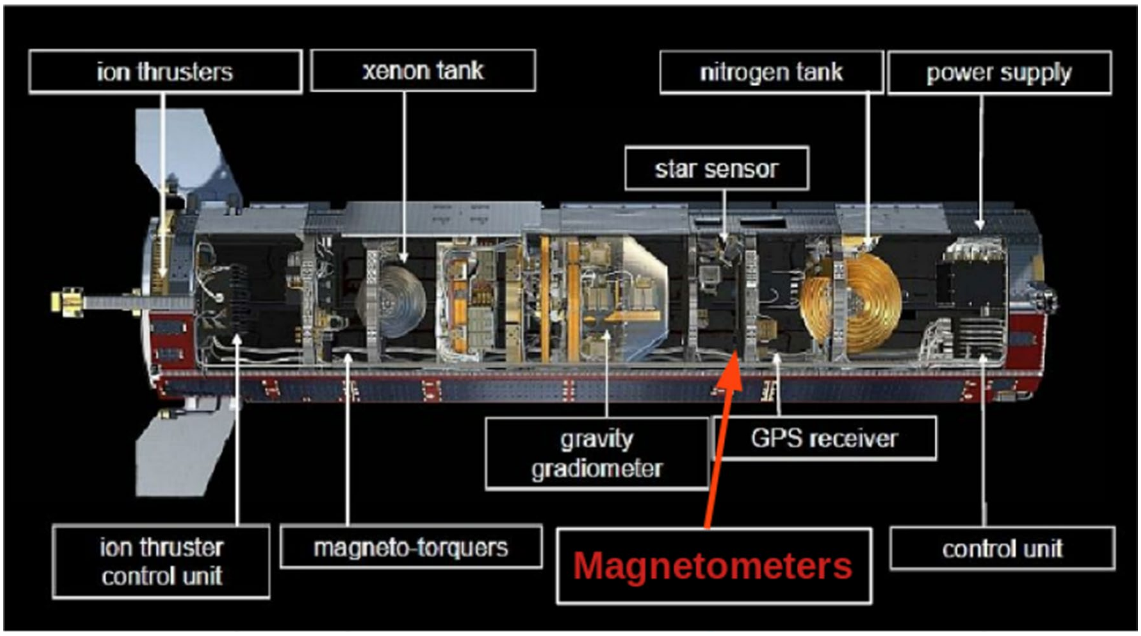


Fig. 3 Location of instruments at the satellite body. (Credits: ESA)

and steady-state Ocean Circulation Explorer (GOCE) mission, following a similar calibration and characterisation procedure of GRACE-FO (Stolle et al. 2021).

The GOCE mission has been operated by ESA. The primary objective of GOCE (Floberghagen et al. 2008, 2011; GOCE Flight Control Team 2014) was to obtain precise global and high-resolution models for both the static and the time-variable components of the Earth's gravity field and geoid. GOCE was successfully launched on 17 March 2009 and completed its mission on 11 November 2013. It was flying on a near-circular polar dawn–dusk orbit with an inclination of 96.7 °C and at a mean altitude of about 262 km, (https://www.esa.int/Applications/Observing_the_Earth/FutureEO/GOCE/Facts_and_figures). A sketch of the satellite is shown in Fig. 2 and a summary on the satellite's orbits and body is available at (https://www.esa.int/Enabling_Support/Operations/GOCE). The GOCE satellite carried three magnetometers as part of its attitude and orbit control system mounted side-by-side displaced by 80 mm. The attitude was mainly controlled by ion thrusters to achieve a drag-free flight, and in addition magnetorquers are used. For magnetorquer activation, the magnetic background field at each time and location of the satellite needs to be measured by magnetometers.

This article describes the original data, methods, and procedures of data processing, characterisation of disturbances, and calibration of instrument-intrinsic parameters that are necessary to obtain scientifically useful magnetic field data from the GOCE platform magnetometers. We show the performance of the calibration and characterisation procedure by comparison to the CHAOS-7 field model, the illustration of Field Aligned Currents (FAC), and a comparison of the time series characterising a geomagnetic storm to the commonly used Dst index that is obtained from ground data. The processed magnetometer data described in this article are available at (Michaelis and Korte 2022), for November 01, 2009 to September 30, 2013. The data published with this article is version 0205.

Data sets and data pre-processing

Instruments

As part of the Drag-free Attitude and Orbit Control System (DFACS), the GOCE satellite carries three active 3-axis fluxgate magnetometers, called MGM. The calibration and characterisation effort is part of Swarm DISC (Swarm DISC 2022). The PlatMag consortium within Swarm DISC decided to call magnetometer instrument reference frames MAG. Hence MGM will be further called MAG. Figure 3 shows the locations of the magnetometers onboard the satellite. The magnetometers are manufactured by Billingsley Aerospace &Defence and are of type TFM100S (Billingsley 2020). The measurement range is ±100 μT, the root mean square noise level of the instrument is 100 pT and the resolution of the digitisation is 3.05185 nT/bit, (Kolkmeier et al. 2008). Hence, the instrument noise is below the digitisation level. The data are sampled at 1/16 Hz. The MAG data have been pre-calibrated achieving biases of less than 500 nT.

Magnetometer calibration further relies on attitude data derived from the Electrostatic Gravity Gradiometer (EGG), which is GOCE's main instrument, and three star cameras (STR) that are mounted on the shaded side of the satellite, shown in Fig. 2. The strongest magnetic disturbance is expected from the magnetorquers (MTQ), although they are located as far away as possible from the magnetometers; see the overview of instrument location in Fig. 3. Since measurements of the magnetorquer currents are available, an almost full correction for them can be expected.

GOCE's whole telemetry of the satellite, including, e.g. magnetometer, magnetorquer currents, attitude, solar array currents, battery currents, and magnetometer temperatures, is publicly available at https://earth.esa.int/eogateway/missions/goce, European Space Agency (2009). The telemetry datasets used for this article are listed in Table 1. GOCE L1b and L2 data are provided in zip files that contain ESA's Earth Explorer Format (EEF) files for each L1b product. An overview of used products with given names, source, unit, and time resolution is listed in Table 1. Data stored as telemetry are given in zip files that contain ESA's Earth Explorer header and data in ASCII. Time values are always handled as defined in the EEF. The dataset with the highest cadence and quality is the attitude information since it relies on the main instrument of the mission. An interpolation of attitude data may add numerical noise. Therefore, it makes sense to use timestamps from the attitude dataset as reference for creating a series of timestamps. The timestamps are selected from the attitude dataset that are closest to MAG dataset timestamps. This subset of input data was used to linearly interpolate all other data, that is position, magnetometer, magnetorquer, currents and other housekeeping (HK) data listed in Table 1. If the interpolation distance is larger than 16 seconds a flag has been set that indicates a data gap. For each timestep, the predictions of the high-level geomagnetic field model CHAOS-7 including core, crustal and external contributions have been calculated, following Finlay et al. (2020). For the selection of the low-latitude range ($|QDLAT| < 50°$), we also calculate the quasi-dipole latitude (QDLAT) and magnetic local time (MLT) (Richmond 1995; Emmert et al. 2010) for each record. For selection of the geomagnetic quiet days, we use the geomagnetic Kp index ($Kp \leq 3$) (Matzka et al. 2021) and the geomagnetic equatorial Dst index ($|Dst| \leq 30\,\mathrm{nT}$) (Nose et al. 2015).

Coordinate frames

The Satellite Physical Coordinate Frame (SC_O_p), called SC in the following, is defined in Kolkmeier et al. (2008). The three MAGs are aligned with the principal axis of the satellite. The rotation of a vector in SC to MAG reference frame is given in Eq. (1):

$$\overline{\mathrm{MAG}_i} = \underline{R_{\mathrm{SC2MAG}}}\,\overline{\mathrm{SC}}, \tag{1}$$

with

$$\underline{R_{\mathrm{SC2MAG}}} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{2}$$

That means negative $\mathbf{MAG_{i,x}}$ is aligned with the flight-direction, $\mathbf{MAG_{i,z}}$ points to the Earth and $\mathbf{MAG_{i,y}}$ completes the orthogonal coordinate system.

The Gradiometer Reference Frame (GRF) is the coordinate system in which the measurements of GOCE's main instrument, the Electrostatic Gravity Gradiometer (EGG), are given. These are the gravity tensor and the combined EGG and STR attitude of the satellite with respect to the International Celestial Reference Frame (ICRF). GOCE provides a high quality attitude product, EGG_IAQ_1i (Frommknecht et al. 2011), which is the combination of the Electrostatic Gravity Gradiometer (EGG) and the star cameras. Fixed reference frames for all instruments are expected to be stable with respect to each other. Missing static rotations between reference frames will be corrected by Euler angle estimation during calibration.

Scientific evaluation of the data will be done in the Earth-fixed North–East–Centre (NEC) reference frame, which is also the frame for predictions of the CHAOS-7 reference model. The calibration and characterisation procedure has to be done in the same reference frame for measurements and model data. Calibration parameters are instrument intrinsic and depend on the instrument reference frame. Characterisations of local disturbances are systematic in a local satellite reference frame. That leads to the decision to apply calibration and characterisation in the MAG reference frame.

For rotation of CHAOS-7 predictions, $\mathbf{B_{model,NEC}}$, from NEC to MAG reference frame a chain of rotations is needed. The first is the rotation from NEC to International Terrestrial Reference Frame (ITRF) depending on the latitude and longitude of the satellite location. We use Seeber (2003, page 23) to define a North–East–Zenith reference frame. By changing the sign of the z-direction (3rd row) we get a North–East–Centre reference frame, Eq. (3):

$$\underline{R_{\mathrm{ITRF2NEC}}} = \begin{pmatrix} -sin(\Phi)\cdot cos(\Lambda) & -\,sin(\Phi)\cdot sin(\Lambda) & cos(\Phi) \\ -sin(\Lambda) & cos(\Lambda) & 0 \\ -cos(\Phi)\cdot cos(\Lambda) & -\,cos(\Phi)\cdot sin(\Lambda) & -\,sin(\Phi) \end{pmatrix}$$
$$\text{with latitude } \Phi \text{ and longitude } \Lambda. \tag{3}$$

The second is a time-dependent rotation from ITRF to ICRF, taking into account Earth's nutation and precession. $\underline{R_{\mathrm{ITRF2ICRF}}}$ is calculated by application of the SOFA library function iauC2t06a (IAU SOFA Board 2019) and using Earth rotation parameters that are derived from the International Earth Rotation and Reference Systems service (IERS 2020).

The rotation from ICRF to GRF frame is given by quaternions available in the EGG_GGT_li product. GRF and SC reference frames are nominally parallel (Kolkmeier et al. 2008), we can set the quaternions given in EGG_GGT_li product to derive the rotation from ICRF to SC, q_{ICRF2SC}.

Rotations can be combined very stably using quaternion algebra. Hence, we need to convert the direction cosine representation of $\underline{R_{\mathrm{NEC2ITRF}}}$, $R_{\mathrm{ITRF2ICRF}}$ and $\underline{R_{\mathrm{SC2MAG}}}$ to a quaternion representation q_{NEC2ITRF}, $q_{\mathrm{ITRF2ICRF}}$ and q_{SC2MAG} following (Wertz 1978, page 415). In summary, the complete rotation from the NEC to the MAG frame is given as:

$$\mathbf{q_{NEC2MAG}} = \mathbf{q_{NEC2ITRF}} \cdot \mathbf{q_{ITRF2ICRF}} \cdot \mathbf{q_{ICRF2SC}} \cdot \mathbf{q_{SC2MAG}}, \tag{4}$$

$$\mathbf{B_{NEC}} \xrightarrow{\mathbf{q_{NEC2ITRF}}} \mathbf{B_{ITRF}} \xrightarrow{\mathbf{q_{ITRF2ICRF}}} \mathbf{B_{ICRF}} \xrightarrow{\mathbf{q_{ICRF2SC}}} \mathbf{B_{SC}} \xrightarrow{\mathbf{q_{SC2MAG}}} \mathbf{B_{MAG}}. \tag{5}$$

CHAOS-7 predictions are finally rotated from NEC to the MAG frame applying the rotation quaternion in Eq. (4) following (Wertz 1978, page 759):

$$\mathbf{B_{model,MAG}} = \mathbf{q^{-1}_{NEC2MAG}} \cdot \mathbf{B_{model,NEC}} \cdot \mathbf{q_{NEC2MAG}}. \tag{6}$$

For rotation of calibrated and characterised MAG data, Eq. (6) has to be applied in inverse order on $\mathbf{B_{MAG}}$.

Pre-processing

The three equal fluxgate magnetometers on the GOCE satellite are mounted perfectly aligned side-by-side with a distance of 80 mm. For that reason one would expect them to give the same results at the same times. However, when looking at the residuals to CHAOS-7 of the individual components from different magnetometers, respectively, some large steps are visible. We found no correlation with activity of GOCE instruments or major

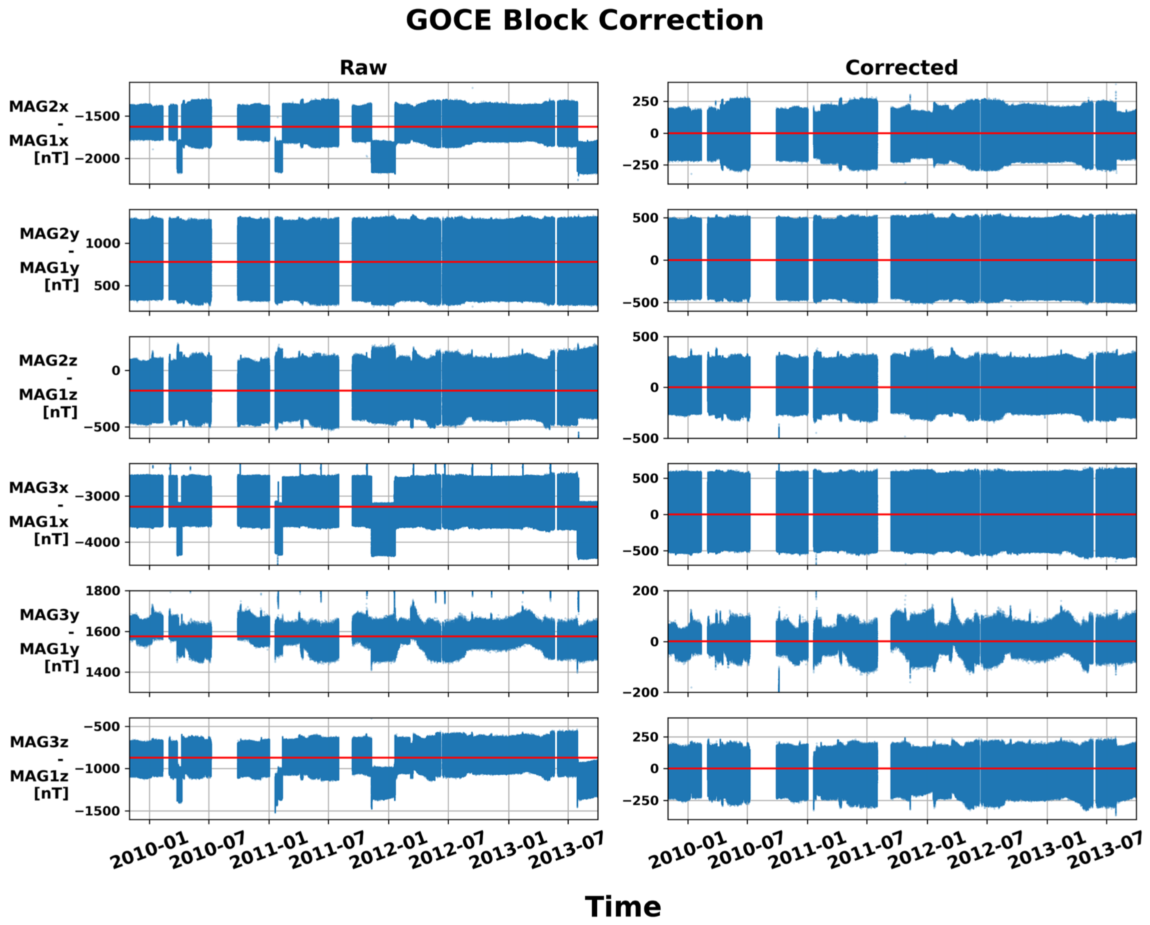


Fig. 4 Overview of block correction for the whole mission. Shown are the differences between magnetometers 2 and 1, and 3 and 1 for the *x*, *y*, and *z* components from top to bottom. Without block correction (left) and after applied block correction (right)

events. We had to correct those events by hand before applying the calibration, and call this step block correction in the following. For each component of MAG2 and MAG3 we subtracted the corresponding component of MAG1. We identified timestamps of the beginning of each block correction by using a higher resolution figure of Fig. 4. The first block has been set as reference for all components of MAG2 and MAG3. For all further blocks the offset of MAG2 and MAG3 has been corrected to reach the same mean value as the first block. At the end the mean value of all blocks has been removed from MAG2 and MAG3. A table containing the timestamps of each event and the corresponding correction values is given as supplementary material in Additional file 1. After the block correction has been applied the residuals between the magnetometers look similar, as can be seen in Fig. 4. Since there will be no relevant scientific output from three calibrated magnetometers very close to each other we decided to combine the three magnetometers into one single instrument by using the mean value, Eq. (7):

$$\mathbf{B}_{\mathbf{MAG}} = \frac{\sum_{i=1}^{3} \mathbf{B}_{\mathbf{MAGi}}}{3}. \tag{7}$$

By combination of the three instruments, we reduce the noise level of the input data and fill small gaps in single magnetometer records.

Calibration and characterisation

Since the magnetometers of GOCE are used for the Drag-free Attitude and Orbit Control System (DFACS) they have been calibrated on-ground to fulfil the specification for DFACS which has biases of less than 500 nT. The pre-calibrated dataset is provided in the AUX_NOM_1B product. Previous studies, like Stolle et al. (2021) for GRACE-FO and Olsen et al. (2020) for CryoSat-2 showed that adding more internal features like

Table 1 Input data used for calibration and characterisation, including product name, variable name, unit, and temporal resolution

	Description	Product	Variable	Unit	Cadence
E	Magnetic field	AUX_NOM		nT	16 s
			MGM1_X_out_1i		
			MGM1_Y_out_1i		
			MGM1_Z_out_1i		
			MGM2_X_out_1i		
			MGM2_Y_out_1i		
			MGM2_Z_out_1i		
			MGM3_X_out_1i		
			MGM3_Y_out_1i		
			MGM3_Z_out_1i		
A_{MTQ}	Magnetorquer currents	Telemetry		A	1 s
	mtr1_current		CAT20044		
	mtr2_current		CAT20045		
	mtr2_current		CAT20046		
POS	Satellite position in ITRF			km	1 s
	PSO_PKI and PSO_PRD	PSO_2G	X,Y,Z		
q	EGG_IAQ	EGG_NOM_1B			1 s
	ICRF to GRF	EGG_IAQ_1i	q1,q2,q3,q4		
T_{MAG}	Magnetometer temperature			degC	32 s
	MGM_HTR_T1		THT00004		
	MGM_HTR_T2		THT00012		
	MGM_HTR_T3		THT00068		
A_{BAT}	Battery currents			A	16 s
	BAT_CHARGE_PWR		PHD95002		
	BAT_PROVIDED_PWR		PHD95021		
	BAT_CHARGE_CUR_N		PHT10040		
	BAT_DISCH_CUR_N		PHT10060		
A_{SA}	Solar array current			A	32 s
	THT10000		SA W+Z T N		
	THT10001		SA W-Z T N		
HK	Housekeeping data	Telemetry			
	CDE_A_Status		MHT00000		16 s
	PCUx_INPUT_CUR		PHD94003	A	16 s
	PCU1_INPUT_CUR		PHD94001	A	16 s
	PCU2_INPUT_CUR		PHD94002	A	16 s
	PCU3_INPUT_CUR		PHD94003	A	16 s
	PCU4_INPUT_CUR		PHD94004	A	16 s
	PCU5_INPUT_CUR		PHD94005	A	16 s
	PCU6_INPUT_CUR		PHD94006	A	16 s
	PCU1_REG1_CUR		PHT11960	A	16 s
	PCU1_REG2_CUR		PHT11980	A	16 s
	PCU2_REG1_CUR		PHT12100	A	16 s
	PCU2_REG2_CUR		PHT12120	A	16 s
	PCU2_REG3_CUR		PHT12140	A	16 s
	PCU3_REG3_CUR		PHT12320	A	16 s
	PCU4_REG1_CUR		PHT12420	A	16 s
	PCU4_REG2_CUR		PHT12440	A	16 s
	PCU4_REG3_CUR		PHT12460	A	16 s
	PCU3_REG1_CUR		PHT12280	A	16 s

Table 1 (continued)

Description	Product	Variable	Unit	Cadence
PCU3_REG2_CUR		PHT12300	A	16 s
PCU5_REG1_CUR		PHT12560	A	16 s
PCU5_REG2_CUR		PHT12580	A	16 s
PCU5_REG3_CUR		PHT12600	A	16 s
SA O+Z-X TEMP		THT10002	degC	32 s
SA C+Z-X TEMP		THT10003	degC	32 s
SA O-Z+X TEMP		THT10004	degC	32 s
SA C-Z+X TEMP		THT10005	degC	32 s

currents, and temperatures to the data processing, that may cause perturbations, can lead to much better calibrated datasets. We follow the same approach as Stolle et al. (2021) but adapt it to conditions and limitations of the GOCE satellite, e.g. availability of currents and temperature measurements. Calibration and characterisation will be applied on a subset only, to avoid that natural variations are interpreted as disturbances, but remain part of the data after the calibration procedure. Therefore, we use only geomagnetic quiet times when natural variations should not be measured by the satellite, thus allowing for a post-launch calibration of the satellite system itself. Concretely, we use only data with $|\mathrm{QDLAT}| < 50°$, $Kp \leq 3, |\mathrm{Dst}| \leq 30\,\mathrm{nT}$ and $B\_\mathrm{Flag} = 0$. B_Flag is a quality flag that gives non-zero values if the data gap for interpolation of input data is larger than 16 s. Since the resolution of the magnetometer data is only 16 s, we decided to use monthly data for the estimation of calibration and characterisation parameters. That avoids rapid fluctuation in estimated parameters, but still gives a long-term trend of parameter evolution with time to cope with system changes and deterioration.

Parameters for vector calibration

The previously combined magnetometer data B_{MAG} act as the raw magnetic field vector for calibration, in MAG frame further named $\mathbf{E} = (E_1, E_2, E_3)^T$ in nT. The calibration estimates the nine instrument-intrinsic parameters scale factors $\mathbf{s} = (s_1, s_2, s_3)^T$, offsets $\mathbf{b} = (b_1, b_2, b_3)^T$ and misalignment angles of the coil windings $\mathbf{u} = (u_1, u_2, u_3)^T$. Additionally, misalignment between static reference frames may occur, e.g. due to slight rotation during mounting of instruments. This misalignment is estimated in a vector of Euler (1-2-3) angles $\mathbf{e} = (e_1, e_2, e_3)^T$, following Wertz (1978, page 764), or in a direction cosine rotation matrix, $\underline{R_A}$, which includes the three external parameters. Euler (1-2-3) represents three rotations about the first, second and third axis, in this order. The parameters are used to describe

$$\mathbf{B_{cal}} = \underline{R_A}\,\underline{P}^{-1}\underline{S}^{-1}(\mathbf{E} - \mathbf{b}) = \underline{A}(\mathbf{E} - \mathbf{b}) = \underline{A}\mathbf{E} - \mathbf{b_A}, \tag{8}$$

where $\underline{R_A}$ is the direction cosine matrix representation of the Euler (1-2-3) angles $\mathbf{e}$, $\underline{P}^{-1}$ is the misalignment angle lower triangular matrix

$$\underline{P}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{sin(u_1)}{cos(u_1)} & \frac{1}{cos(u_1)} & 0 \\ -\frac{sin(u_1)sin(u_3)+cos(u_1)sin(u_2)}{w cos(u_1)} & -\frac{sin(u_3)}{w cos(u_1)} & 1/w \end{pmatrix}$$
$$\text{with: } w = \sqrt{1 - sin^2(u_2) - sin^2(u_3)}, \tag{9}$$

and $\underline{S}^{-1}$ is the diagonal matrix including the inverse of the scale factor

$$\underline{S}^{-1} = \begin{pmatrix} 1/s_1 & 0 & 0 \\ 0 & 1/s_2 & 0 \\ 0 & 0 & 1/s_3 \end{pmatrix}. \tag{10}$$

Equation (8) is valid for fluxgate magnetometers treated as linear instruments. Brauer et al. (1997) showed that Eq. (8) needs to be extended for non-linear effects of 2nd ($\underline{\xi}$) and 3rd ($\underline{\nu}$) order by 2nd ($\mathbf{E}_\xi$) and 3rd ($\mathbf{E}_\nu$) order data:

$$\mathbf{B_{cal}} = \underline{A}\mathbf{E} - \mathbf{b_A} + \underline{\xi}\mathbf{E}_\xi + \underline{\nu}\mathbf{E}_\nu, \tag{11}$$

with non-linearity parameters of 2nd order:

$$\underline{\xi} = \begin{pmatrix} \xi^1_{11} & \xi^1_{22} & \xi^1_{33} & \xi^1_{12} & \xi^1_{13} & \xi^1_{23} \\ \xi^2_{11} & \xi^2_{22} & \xi^2_{33} & \xi^2_{12} & \xi^2_{13} & \xi^2_{23} \\ \xi^3_{11} & \xi^3_{22} & \xi^3_{33} & \xi^3_{12} & \xi^3_{13} & \xi^3_{23} \end{pmatrix}, \tag{12}$$

non-linearity parameters of 3rd order:

$$\underline{\nu} = \begin{pmatrix} \nu^1_{111} & \nu^1_{222} & \nu^1_{333} & \nu^1_{112} & \nu^1_{113} & \nu^1_{223} & \nu^1_{122} & \nu^1_{133} & \nu^1_{233} & \nu^1_{123} \\ \nu^2_{111} & \nu^2_{222} & \nu^2_{333} & \nu^2_{112} & \nu^2_{113} & \nu^2_{223} & \nu^2_{122} & \nu^2_{133} & \nu^2_{233} & \nu^2_{123} \\ \nu^3_{111} & \nu^3_{222} & \nu^3_{333} & \nu^3_{112} & \nu^3_{113} & \nu^3_{223} & \nu^3_{122} & \nu^3_{133} & \nu^3_{233} & \nu^3_{123} \end{pmatrix}, \tag{13}$$

and modulated data vectors of 2nd and 3rd order:

Table 2 Estimated calibration and characterisation parameters including units and dimensionality

Parameter	Description	Unit	Dimension
$\mathbf{s}$	Scale factors	$\frac{\mathrm{nT}}{\mathrm{nT}}$	3
$\mathbf{b}$	Offsets	nT	3
$\mathbf{u}$	Misalignment angles	rad	3
$\mathbf{e}$	Euler (123) angles	rad	3
$\underline{\xi}$	2nd order non-linearity	$\frac{1}{nT}$	3 × 6
$\underline{\nu}$	3rd order non-linearity	$\frac{1}{\mathrm{nT}^2}$	3 × 10
$\underline{bt}$	Temperature dependency of offsets b	$\frac{\mathrm{nT}}{^\circ\mathrm{C}}$	3 × 3
$\underline{st}$	Temperature dependency of scale factors s	$\frac{\mathrm{nT}}{\mathrm{nT}^\circ\mathrm{C}}$	3 × 3
$\underline{bat}$	Battery current scale factor	$\frac{\mathrm{nT}}{\mathrm{mA}}$	3 × 4
$\underline{sa}$	Solar array current scale factor	$\frac{\mathrm{nT}}{\mathrm{mA}}$	3 × 2
$\underline{M}$	Magnetorquer current scale factor	$\frac{\mathrm{nT}}{\mathrm{mA}}$	3 × 3
$\underline{hk}$	Housekeeping data scale factor		3 × 25

Parameters for characterisation

Characterisation consists of the identification and, if possible, correction of artificial magnetic perturbations contained in the raw magnetic data. By simple correlation analysis combined with knowledge from former satellite missions like CHAMP, Swarm and GRACE-FO, we identified the magnetorquer currents, $\mathbf{A_{MTQ}}$, the magnetometer heater temperatures, $\mathbf{T_{MAG}}$, the battery currents, $\mathbf{A_{BAT}}$, the solar array panel currents, $\mathbf{A_{SA}}$, and a set of housekeeping currents, and temperatures $\mathbf{A_{HK}}$, to affect the GOCE magnetometer data. We also consider an effect from the correlation between the magnetometer temperature and magnetic field residuals, $\mathbf{E_{st}} = \mathbf{E} \cdot (\mathbf{T_{MAG}} - T_0)$, where T_0 is the monthly median of $\mathbf{T_{MAG}}$.

The characterisation equation is a combination of all identified disturbances:

$$\begin{aligned} \mathbf{E}_\xi &= (E_1^2,\, E_2^2,\, E_3^2,\, E_1E_2,\, E_1E_3,\, E_2E_3)^T, \\ \mathbf{E}_\nu &= (E_1^3,\, E_2^3,\, E_3^3,\, E_1^2E_2,\, E_1^2E_3,\, E_2^2E_3,\, E_1E_2^2,\, E_1E_3^2,\, E_2E_3^2,\, E_1E_2E_3)^T. \end{aligned} \tag{14}$$

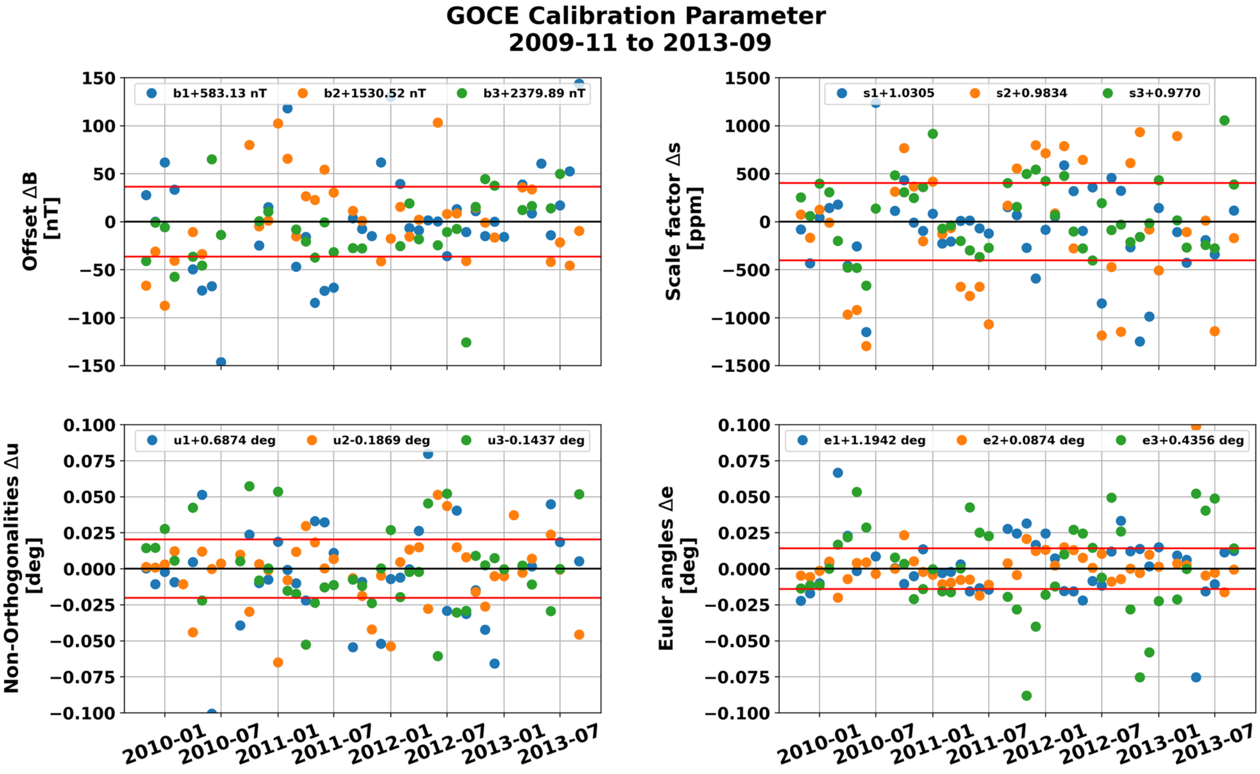


Fig. 5 Time series of instrument-intrinsic calibration parameters offset (top-left), scale factors (top right), non-orthogonalities (bottom-left) and Euler angles (bottom-right) with respect to their median value. Red lines indicate average mean absolute deviation

$$\begin{aligned}\mathbf{B_{char}} =&\underline{M} \cdot \mathbf{A_{MTQ}} + \underline{bat} \cdot \mathbf{A_{BAT}} \\ &+ \underline{sa} \cdot \mathbf{A_{SA}} + \underline{hk} \cdot \mathbf{A_{HK}} \\ &+ \underline{bt} \cdot (\mathbf{T_{MAG}} - T_0) + \underline{st} \cdot \mathbf{E_{st}},\end{aligned} \tag{15}$$

with magnetorquer current scale factor ($\underline{M}$), battery current scale factor ($\underline{bat}$), solar array current scale factor ($\underline{sa}$), housekeeping data scale factor ($\underline{hk}$), temperature dependency of offsets b ($\underline{bt}$) and temperature dependency of scale factors s ($\underline{st}$). Input data used in Eq. (11) and (15) are listed in Tables 1 and 2, respectively. All input parameters and calibrated magnetic observation products are provided in CDF format, in the same format as for GRACE-FO (Michaelis et al. 2021).

Parameter estimation

An ordinary least squares linear regression has been applied to estimate the parameters $\mathbf{m_{cal}}$ and $\mathbf{m_{char}}$ to minimise for S:

$$S = |(\mathbf{B_{cal}}(\mathbf{m_{cal}}, \mathbf{E}) + \mathbf{B_{char}}(\mathbf{m_{char}}, \mathbf{d_{char}})) - \mathbf{B_{model,MAG}}|^2, \tag{16}$$

with the calibrated magnetic field vector $\mathbf{B_{cal}}$ using instrument-intrinsic calibration parameters, $\mathbf{m_{cal}} = (\mathbf{b}, \mathbf{s}, \mathbf{u}, \mathbf{e}, \underline{\xi}, \underline{\nu})$ that have been applied on the raw magnetic field vector $\mathbf{E}$, as given in Eq. (11). For estimation of the characterised magnetic field vector $\mathbf{B_{char}}$ parameters describing the impact on the housekeeping data $\mathbf{m_{char}} = (\underline{M}, \underline{bat}, \underline{sa}, \underline{bt}, \underline{st}, \underline{hk})$ have been applied to the housekeeping data $\mathbf{d_{char}} = (\mathbf{A_{MTQ}}, \mathbf{A_{BAT}}, \mathbf{A_{SA}}, \mathbf{A_{HK}}, \mathbf{T_{MAG}}, \mathbf{E_{st}})$, as given in Eq. (15). $\mathbf{B_{model,MAG}}$ is the CHAOS-7 magnetic field estimations for the core, crustal and large-scale magnetospheric field rotated into the instrument MAG frame as described by Eq. (6).

From previous satellite missions like GRACE-FO it is known that additional time shifts between instrument measurements may occur. We repeated the calibration and characterisation procedure for a range of time shifts within an interval of ±2 s in steps of 0.1 s on the most quiet data set, which was in December 2009. Best calibration results (minimum of the absolute values of residual to CHAOS-7) have been determined with a shift of 0.4 s for MAG data.

Results and discussion

In this section, we discuss the final GOCE data set and some potential applications. We assess the residuals to CHAOS-7 predictions of all vector components and compare the lithospheric field measured from the GOCE data to the lithospheric field contribution included in CHAOS-7. Moreover, we calculate auroral field-aligned currents (FAC) and compare magnetospheric ring currents measured by GOCE with ground-based estimations like the Geomagnetic Equatorial Disturbance Storm Time Index (Dst).

Assessment of the final data set

To assess the temporal robustness of the calibration, time series of calibration parameters are shown in Fig. 5 for offsets, scale factors, non-orthogonalities and Euler angles. Red lines show the average mean absolute deviation of the parameters. The parameters show no long-term trends over the mission duration. Comparisons with previously published studies gave similar order results for the mean absolute deviation of the parameter time series for CryoSat-2 (Olsen et al. 2020). However, in detail GOCE shows much higher variations in each of the parameters. That might be caused by higher air pressure at GOCE's low altitude which is compensated for by near-continuous operation of the drag-free attitude and orbit control system.

Residuals for the calibrated magnetic field vector have been calculated with respect to CHAOS-7 predictions for geomagnetic quiet conditions and low latitudes, i.e. $|\mathrm{QDLAT}| < 50^o$, $Kp <= 3$, and $|\mathrm{Dst}| <= 30\,\mathrm{nT}$. Table 3 shows the mean and standard deviation of these residuals

Table 3 Mean and standard deviation of residuals to CHAOS-7 for GOCE for geomagnetic quiet times and for a single quiet day, 2009-12-01

Parameter	Whole period						Single day					
	Mean [nT]			Std [nT]			Mean [nT]			Std [nT]		
	x	y	z	x	y	z	x	y	z	x	y	z
ΔB_{MAG}	0.0	− 0.1	− 0.0	116.7	276.3	115.9	1.1	− 1.6	0.4	8.3	6.1	5.6
ΔB_{NEC}	− 2.8	− 0.1	− 0.3	135.5	271.5	106.3	0.1	1.0	0.4	8.3	6.4	5.6
ΔB_{RAW1}	− 592.4	− 1618.6	− 2318.3	763.1	554.2	623.1	− 549.1	− 1587.2	− 2269.3	752.9	495.4	594.3
ΔB_{RAW2}	− 597.3	− 1613.6	− 2311.7	796.1	743.2	695.9	− 543.2	− 1580.9	− 2273.3	792.8	700.4	664.5
ΔB_{RAW3}	− 589.3	− 1620.6	− 2314.0	721.1	565.4	560.2	− 531.7	− 1595.6	− 2285.6	712.0	502.3	517.1

$\mathbf{B_{MAG}}$ and $\mathbf{B_{NEC}}$ represent residuals for calibrated data and $\mathbf{B_{RAW}}$ for data before calibration

Table 4 Standard deviation of residuals to CHAOS-7 for GOCE for all months in the mission period

Month	ΔB_{MAG}			ΔB_{NEC}			ΔB_{RAW1}			Used		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	Data	$\overline{K_p}$	$\overline{Dst}$
	[nT]	[nT]	[nT]	[nT]	[nT]	[nT]	[nT]	[nT]	[nT]	[%]		[nT]
2009-11-01	8.8	6.4	5.5	8.7	6.4	5.5	737.1	468.5	603.0	54.0	0.62	−2.0
2009-12-01	8.9	6.4	5.9	8.9	6.4	5.9	736.4	473.9	607.1	51.7	0.46	4.0
2010-01-01	9.2	6.9	6.4	9.2	6.9	6.4	739.0	470.7	600.7	51.6	0.63	−2.0
2010-02-01	8.9	7.6	5.6	8.9	7.6	5.6	737.1	472.3	601.1	18.1	1.11	−8.0
2010-03-01	80.7	276.7	59.1	34.5	284.0	68.7	754.9	551.5	611.0	50.6	1.06	−5.0
2010-04-01	13.7	9.4	40.8	13.5	9.8	40.8	735.8	489.0	615.2	42.1	1.08	−12.0
2010-05-01	13.1	11.5	15.5	13.1	11.5	15.5	732.2	485.2	613.8	40.5	1.12	−7.0
2010-06-01	18.9	169.8	31.2	37.8	160.6	54.2	730.6	512.6	614.0	47.6	1.41	−9.0
2010-07-01	30.7	44.0	28.1	30.6	44.1	28.1	721.9	474.0	617.0	8.5	1.49	−12.0
2010-08-01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	NaN
2010-09-01	14.9	17.3	15.9	13.6	18.3	15.9	750.8	473.8	595.0	5.9	1.13	−12.0
2010-10-01	14.9	9.2	40.3	14.8	9.4	40.3	746.5	470.8	593.7	48.2	1.05	−9.0
2010-11-01	9.1	7.1	5.9	9.1	7.1	5.9	745.6	470.4	607.9	52.4	1.03	−8.0
2010-12-01	9.8	7.7	6.1	9.8	7.8	6.1	743.2	476.3	611.3	50.8	0.8	−7.0
2011-01-01	15.3	10.0	28.5	15.8	9.2	28.5	758.8	467.7	592.8	20.8	0.92	−2.0
2011-02-01	9.6	7.4	12.0	9.5	7.4	12.0	739.0	476.4	614.9	46.6	1.02	−9.0
2011-03-01	9.4	7.5	6.1	9.3	7.6	6.1	734.9	479.4	606.3	44.4	1.07	−5.0
2011-04-01	10.5	9.1	8.8	10.5	9.2	8.7	755.5	481.7	603.0	41.9	1.13	−5.0
2011-05-01	11.9	10.0	12.0	11.9	10.0	12.0	737.5	477.9	604.9	46.4	1.29	−7.0
2011-06-01	13.7	12.4	16.8	13.8	12.4	16.8	741.9	486.0	606.1	45.5	1.52	−9.0
2011-07-01	13.4	11.7	15.6	13.5	11.7	15.6	737.5	485.5	606.2	47.3	1.59	−9.0
2011-08-01	3.5	2.0	5.4	3.6	1.9	5.4	780.0	394.5	525.7	0.1	1.71	−15.0
2011-09-01	9.2	8.9	6.4	9.0	9.0	6.4	750.1	478.4	591.5	20.3	1.09	−14.0
2011-10-01	9.8	8.6	6.6	9.7	8.7	6.5	753.9	476.8	603.6	44.7	1.09	−11.0
2011-11-01	31.4	202.6	53.2	45.3	200.5	50.9	763.4	516.9	612.1	49.0	0.9	−9.0
2011-12-01	10.5	9.2	8.9	10.4	9.3	8.9	762.6	467.5	597.6	54.8	0.92	−3.0
2012-01-01	13.2	9.8	30.8	13.1	10.1	30.7	773.6	473.3	621.9	43.6	1.25	−3.0
2012-02-01	9.8	8.2	6.3	9.8	8.3	6.3	736.1	478.8	603.1	45.7	1.46	−9.0
2012-03-01	10.4	10.9	7.7	10.4	10.9	7.7	748.9	476.1	594.1	24.6	1.42	−14.0
2012-04-01	10.7	8.6	7.1	10.6	8.7	7.1	745.7	476.3	600.2	42.7	1.35	−12.0
2012-05-01	12.6	12.8	14.2	12.6	12.8	14.2	747.8	482.0	603.1	48.9	1.27	−5.0
2012-06-01	34.7	156.2	42.2	25.2	160.8	29.5	759.5	513.1	607.7	33.6	1.3	−5.0
2012-07-01	16.4	12.9	26.1	16.3	12.9	26.2	736.9	486.5	601.5	38.1	1.72	−9.0
2012-08-01	11.6	9.6	8.9	11.5	9.7	8.9	739.2	475.7	605.9	49.3	1.4	−4.0
2012-09-01	12.0	10.6	8.6	12.1	10.6	8.6	766.6	483.3	605.6	45.4	1.22	−2.0
2012-10-01	10.0	9.0	6.8	9.9	9.1	6.8	757.5	486.0	611.7	38.9	0.93	−7.0
2012-11-01	10.8	9.0	7.3	10.8	9.0	7.3	767.9	484.0	620.4	45.4	0.98	−6.0
2012-12-01	10.1	8.5	6.9	10.1	8.5	6.9	749.7	477.0	619.9	55.6	0.78	8.0
2013-01-01	33.9	180.6	49.4	37.3	183.7	32.7	754.9	512.4	617.2	50.2	1.01	0.0
2013-02-01	757.7	1736.0	757.6	888.5	1697.6	689.6	1117.1	1815.9	971.8	44.0	1.3	−6.0
2013-03-01	10.0	8.6	6.6	9.9	8.7	6.6	758.8	477.1	604.0	39.8	1.18	−6.0
2013-04-01	10.6	9.1	8.0	10.6	9.1	8.0	755.2	485.3	598.0	53.4	1.06	−4.0
2013-05-01	208.8	646.2	129.0	192.9	652.6	120.1	813.0	806.5	616.8	26.4	1.25	−5.0
2013-06-01	14.1	12.7	16.8	14.1	12.7	16.8	756.1	493.5	604.6	39.1	1.34	−11.0
2013-07-01	15.9	13.2	32.4	15.9	13.2	32.4	759.2	492.1	603.9	41.8	1.27	−9.0
2013-08-01	107.8	130.6	121.6	108.4	131.1	120.6	808.4	504.6	595.9	45.1	1.34	−9.0
2013-09-01	10.8	9.6	7.5	10.6	9.7	7.5	772.9	477.4	586.6	51.7	1.16	−3.0

$\mathbf{B}_{\mathbf{MAG}}$ and $\mathbf{B}_{\mathbf{NEC}}$ represent residuals for calibrated data and $\mathbf{B}_{\mathbf{RAW1}}$ for MAG$_1$ data before calibration. The amount of data used for calibration and the averages of the two geomagnetic activity indices Kp and Dst are also given

Table 5 Magnetic impact of calibration and characterisation, respectively, for each parameter given in Eq. (15) and the non-linear parameters in Eq. (11)

Parameter	**Std [nT]**			**Min [nT]**			**Max [nT]**		
	x	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
ΔB_{ξ}	67.8	141.2	67.5	− 13448.9	− 27276.5	− 12631.9	12828.2	35776.8	21688.7
ΔB_{ν}	48.0	82.4	42.8	− 11447.7	− 20148.4	− 10435.0	3830.4	10720.9	9133.3
ΔB_{MTQ}	56.6	33.6	29.3	− 298.2	− 705.7	− 390.5	451.4	704.7	234.8
ΔB_{BAT}	33.4	93.3	48.4	− 634.0	− 725.3	− 744.5	430.9	1225.6	1022.1
ΔB_{SA}	123.6	156.1	185.4	− 885.4	− 573.2	− 784.2	814.0	1630.7	974.0
ΔB_{HK}	212.1	271.1	484.0	− 1049.1	− 1939.5	− 2453.0	2985.5	1502.7	2469.2
ΔB_{BT}	12.4	4.4	43.9	− 93.4	− 128.3	− 329.4	213.3	57.6	289.0
ΔB_{ST}	7.3	7.2	7.4	− 387.7	− 550.9	− 347.5	307.4	340.0	400.9
$\Delta B_{cal,NEC}$	135.5	271.5	106.3	− 28906.3	− 25670.8	− 10717.4	10169.4	21387.6	32442.4

Results are given in the MAG reference frame

for the whole mission period, and for the most quiet day in the most quiet month. The mean values are close to zero which means that the calibration removed the offsets correctly. For very quiet conditions, Kp < 1, the standard deviation can be reduced to values below 8 nT. The calibration has been applied on monthly data. Results for the standard deviation of residuals with respect to the CHAOS-7 model are given for each month in Table 4 for calibrated magnetometer data in MAG and NEC as well as for raw data of magnetometer MAG_1 as representative example. The last three columns give the percentage of data used for the specific month, the mean Kp value and mean Dst value from within data selection for the calibration. Standard deviations vary strongly from month to month. For the majority of months the standard deviation is reduced to the level of very quiet conditions. However, some months deviate strongly from the quiet days. For some of those extreme months, a correlation with missing data or higher geomagnetic conditions seems to exist. However, we cannot state a general correlation of high residuals with high activity. In general, the values for mean and standard deviation have been significantly reduced by the calibration to values between 7 and 13 nT, and are similar to residuals for GRACE-FO given by Stolle et al. (2021) and for CryoSat-2 by Olsen et al. (2020), which varied between 3 nT and 10 nT (GRACE-FO) and 4 nT and 15 nT (CryoSat-2).

The estimation of impact for non-intrinsic instrument parameters is shown in Table 5. The impact has been estimated by residual calculation between using all estimated parameters and using all but one parameter and setting this one parameter to a neutral value. As an example, to estimate the impact of $\mathbf{\Delta B_{SA}}$, first all estimated parameters are applied to Equation 15 to compute $\mathbf{B_{char}}$. Then, the same approach is repeated with $\underline{sa}$ being set to zero and calculating $\mathbf{B_{char,zerosa}}$. The difference between $\mathbf{B_{char}}$ and $\mathbf{B_{char,zerosa}}$ is the impact of parameter $\underline{sa}$, called $\mathbf{\Delta B_{SA}}$. The results indicate that $\underline{hk}$ and $\underline{sa}$ have the largest impact. On other missions, e.g. GRACE-FO (Stolle et al. 2021), an even larger standard deviation of impact from solar panels than for the other parameters was found. The influence might be smaller on GOCE due to design and orbit characteristics of the GOCE satellite. The solar arrays are mounted such that they are always on the bright side with the GOCE dusk–dawn orbit, so that currents induced by the solar arrays are more or less constant and do not vary much.

Figure 6a provides global maps of the residuals between the processed data and CHAOS-7 predictions for December 2009 with the mean of the residuals summarised in bins of size of 5° geocentric latitude and 5° geocentric longitude. The three columns represent the B_N, B_E and B_C components of the NEC frame, respectively. The first row displays residuals to the core, the crustal and the large-scale magnetospheric field predictions of CHAOS-7. The second row shows residuals to only the core and the large-scale magnetospheric field predictions, i.e. in particular the lithospheric field is now included in the data. The third row shows the crustal field prediction from CHAOS-7. The grey lines indicate 0° and ±70° magnetic latitude (QDLAT). Figure 6b gives distribution of geomagnetic and solar indices and magnetic local time of the data set of this month, which was geomagnetically quiet. Auroral electrojet and field-aligned currents at high latitudes produce the largest deviations as they are measured by the satellite but not included in the CHAOS-7 model. Since the data are collected at a dawn–dusk orbit, no significant low and mid latitude ionospheric disturbances are expected, nor significant effects from magnetospheric currents during the quiet

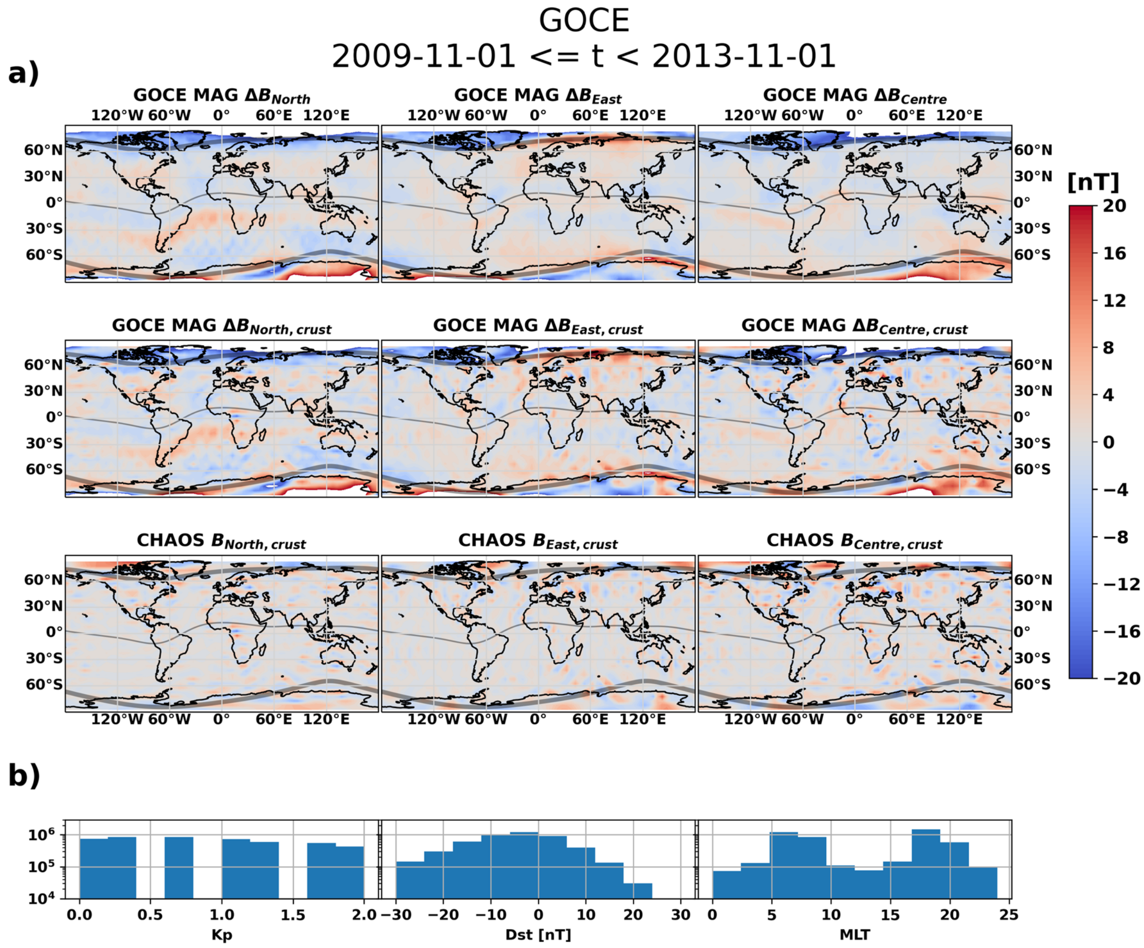


Fig. 6 Top panel of **a** shows magnetic residuals to CHAOS-7 (core, crustal and large-scale magnetospheric field). Middle panel of **a**: magnetic residuals to CHAOS-7 (core and large-scale magnetospheric field). Bottom panel of **a**: crustal field from CHAOS-7 model. The columns show the three NEC components North, East and Centre. **b** Shows the distribution of geomagnetic and solar activity indices and magnetic local time for data selection used in **a**

times. Still, there are systematic deviations that follow the geomagnetic equator in all components, and these are already known from GRACE-FO carrying the same type of magnetometers. However, besides the prominent disturbance at the geomagnetic equator there are large areas with absolute residuals below 4 nT as indicated by greyish colours. The comparison of second and third row of Fig. 6a also shows that the calibrated GOCE data can reproduce the large-scale crustal anomalies quite well. For example, the Bangui and Kursk anomaly in central Africa and Russia, respectively, are clearly seen. Still, a systematic artificial field with low amplitude along the geomagnetic equator is visible.

Large-scale field-aligned currents

Field-aligned currents (FAC) are not part of the CHAOS7 model and should remain in the measured data after calibration and characterisation. Since platform magnetometers have a higher noise level than science magnetometers, we expect only large-scale auroral field-aligned currents to be visible. Figure 7 shows results for FACs derived from GOCE MAG for the whole period of the mission on the Northern (top) and Southern (bottom) hemisphere, selected for the northward (left) and southward (right) z-component of interplanetary magnetic field (IMF). FACs have been put in bins of 2° using the median as the aggregation function. Region 1 and 2 currents are prominently visible, similar to results from the PlatMag feasibility study for Swarm and GOCE https://www.esa.int/Enabling_Support/Preparing_for_the_Future/Discovery_and_Preparation/ESA_s_unexpected_fleet_of_space_weather_monitors and in Lühr et al. (2016).

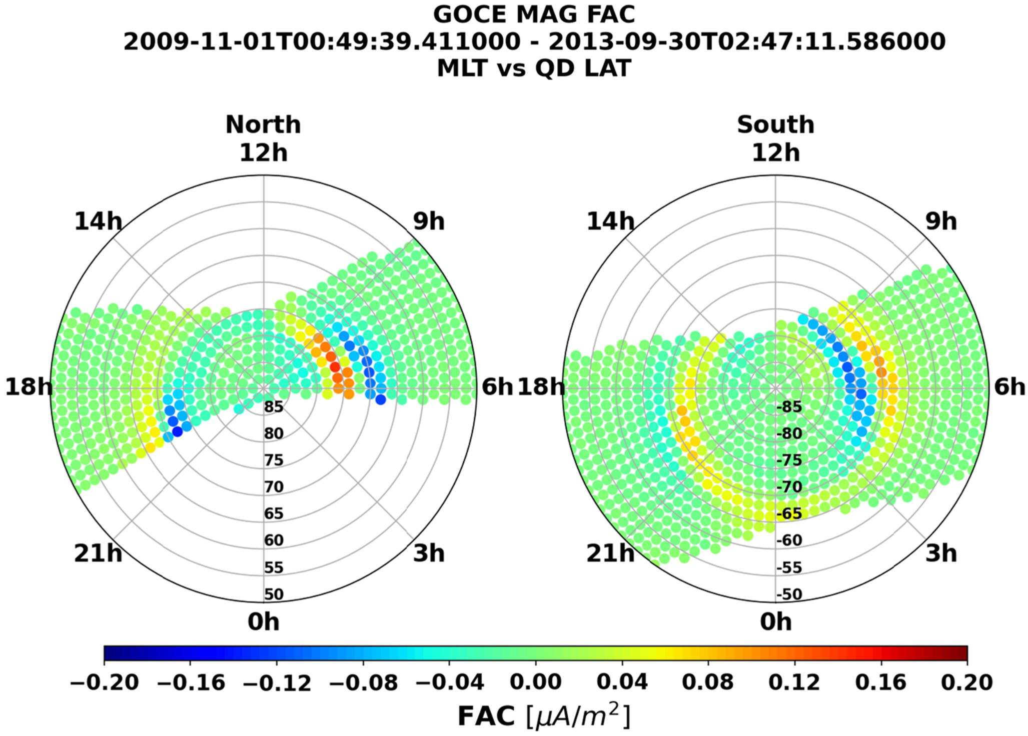


Fig. 7 Quasi-dipole latitude (QDLAT) versus magnetic local time (MLT) large-scale field-aligned currents for the whole mission duration. The left panel shows the northern hemisphere and the right panel the southern hemisphere

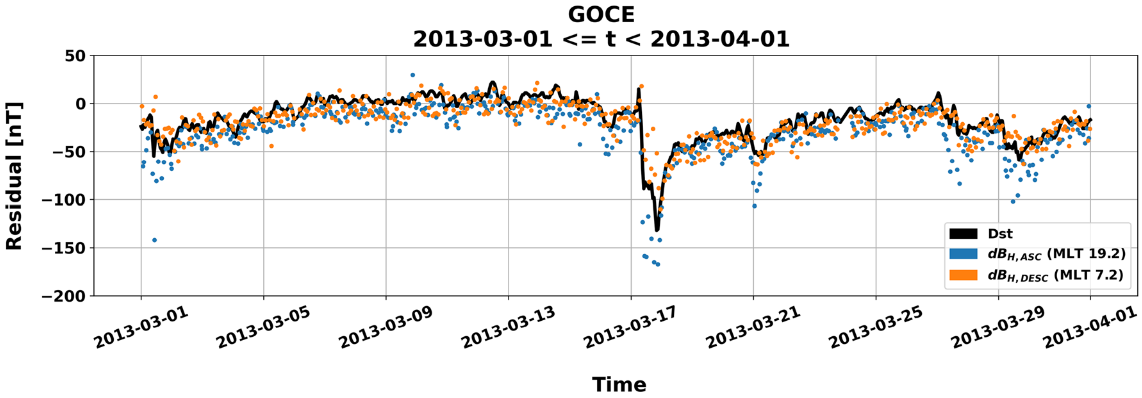


Fig. 8 Time series of residuals of calibrated GOCE magnetic data to the core and crustal field of CHAOS-7 around the magnetic storm in March 2013. Ascending (ASC) nodes are plotted in blue, descending (DESC) nodes in orange. The Dst index is also plotted in black

The magnetic effect of the magnetospheric ring current during the March 17, 2013 storm

A geomagnetic storm with values of Dst < -130 nT occurred on March 17, 2013 (Fig. 8). The circles represent medians of residuals of the horizontal component of the magnetic field ($\sqrt{B_N^2 + B_E^2}$) within ± 10° geomagnetic latitude and projected to 0° geomagnetic latitude for each low-latitude orbital segment for ascending (blue) and descending (orange) orbits. The residuals are calculated with respect to the CHAOS-7 core and crustal field predictions. The large-scale magnetospheric field was not subtracted, and signatures from magnetospheric currents (including their induced counterparts in the Earth) remain included in the data. The ascending and descending orbit data generally agree well with each other and with the Dst index, despite the different retrieval technique for magnetospheric signatures in ground and satellite data. It is known from earlier studies that ground-based derived ring current signatures show systematic differences to those derived in space and that in particular the Dst index does not have the correct magnetospheric baseline (Maus and Lühr 2005; Olsen et al. 2005; Lühr et al. 2017; Pick et al. 2019). The ring current signal obtained from LEO satellites is generally lower than from ground, which is also reflected in an offset between the Dst index and the satellite derived residuals. In detail the ring current at ascending (MLT 6) nodes shows systematic weaker residual than for descending (MLT 18) nodes. That agrees well with dawn–dusk asymmetries found in studies from Newell and Gjerloev (2012) for Super MAG Ring current and Love and Gannon (2009) for Dst.

Conclusions

The GOCE mission carries three vector magnetometers for attitude and orbit control. We applied a calibration and characterisation procedure that significantly reduces perturbations produced artificially by the satellite itself. The calibrated data from non-dedicated magnetometers in LEO can be used to fill gaps between dedicated magnetic field missions and in the MLT distribution. However, since non-dedicated missions do not carry an absolute magnetometer as a reference, a high-level geomagnetic model based on dedicated missions is still needed for the calibration. Although calibrated platform magnetometer data cannot reach residuals below 1 nT when compared to high-level geomagnetic models as dedicated mission data from, e.g. CHAMP and Swarm do, we have shown that they contain information about lithospheric and magnetospheric field signatures and field-aligned currents. With standard deviations of residuals between 7 nT and 13 nT for quiet times, our GOCE results are of similar order to those of CryoSat-2 and GRACE-FO calibrated magnetometer data (Olsen et al. 2020; Stolle et al. 2021). For a mission not dedicated to magnetic field research and not carrying scientific magnetometers, residuals in this order of magnitude are acceptable. The calibrated GOCE data are freely available and may be used for studying different magnetic field sources and the near-Earth space environment.

Abbreviations

CHAMP: CHAllenging Minisatellite Payload; CHAOS: CHAmp Ørsted SAC-C magnetic field model; CDF: Common data format; DFACS: Drag-free attitude and orbit control system; Dst: Geomagnetic Equatorial Disturbance Storm Time Index; ESA: European Space Agency; FAC: Field-aligned currents; GFZ: Helmholtz Centre Potsdam, German Research Centre for Geosciences; GOCE: Gravity field and steady-state Ocean Circulation Explorer; GRACE-FO: Gravity Recovery and Climate Experiment Follow-On; GRF: Gradiometer Reference Frame; HK: Housekeeping; ICRF: International Celestial Reference Frame; IGRF: International Geomagnetic Reference Field IGRF-13; ISDC: Information System and Data Center at GFZ; ITRF: International terrestrial reference frame; L1b: GOCE level 1b data; LEO: Low Earth Orbit; MAG: Magnetometer; MLAT: Modified apex latitude; MLT: Magnetic local time; MTQ: Magnetorquer; NEC: North, East, Centre coordinate system; PlatMag: Platform magnetometer; QDLAT: Quasi-dipole latitude; SC: Spacecraft physical reference frame; STR: Star cameras (Star TRackers).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40623-022-01691-6.

Additional file 1. Pre-processed block correction of magnetometers.

Acknowledgements

The European Space Agency (ESA) is gratefully acknowledged for providing the GOCE data. Special thanks to Björn Frommknecht (ESA) for pre-selection of housekeeping data. Kp is provided by GFZ, the Dst and AE indices by the Geomagnetic World Data Centre Kyoto, and F10.7 by the Dominion Radio Astrophysical Observatory and Natural Resources Canada. We thank Martin Rother (GFZ) for fruitful discussions.

Author contributions

IM and CS defined the study. IM pre-processed and calibrated the data. JR derived FACs. CS, MK, IM, KSR analysed and interpreted the results. IM wrote the manuscript. All authors read and approved the final manuscript.

Funding

This study has been partly supported by Swarm DISC activities funded by ESA under contract no. 4000109587/13/I-NB. KSR is supported through HEIBRIDS-Helmholtz Einstein International Berlin Research School in Data Science under contract no. HIDSS-0001.

Availability of data and materials

The data generated and analysed in this paper are available at (Michaelis and Korte 2022) ftp://isdcftp.gfz-potsdam.de/platmag/MAGNETIC_FIELD/GOCE/Analytical/v0205/.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. [2] Technical University of Berlin, Electrical Engineering and Computer Science, Ernst-Reuter-Platz 7, 10587 Berlin, Germany. [3] Leibniz-Institut für Atmosphärenphysik e.V. an der Universität Rostock, Schloßstraße 6, 18225 Kühlungsborn, Germany.

Received: 8 April 2022 Accepted: 13 August 2022

References

Billingsley (2020) Billingsley TFM100SH Magnetometer. https://magnetometer.com/products/fluxgate-magnetometers/tfm100s,https://magnetometer.com/wp-content/uploads/magnetometer-comaparison.pdf, https://magnetometer.com/wp-content/uploads/TFM100S-Spec-Sheet-February-2008.pdf

Brauer P, Merayo JMG, Nielsen OV, Primdahl F, Petersen JR (1997) Transverse field effect in fluxgate sensors. Sens Actuators A Phys 59:70–74. https://doi.org/10.1016/s0924-4247(97)01416-7 (**ISSN 0924-4247**)

CHAMP: overview of final ME products and format description (2019) CHAMP Satellite Mission. GFZ German Research Centre for Geosciences, Scientific Technical Report STR - Data ; 19/10), (2019). https://doi.org/10.2312/GFZ.b103-19104

Emmert JT, Richmond AD, Drob DP (2010) A computationally compact representation of Magnetic Apex and Quasi-Dipole coordinates with smooth base vectors. J Geophys Res 115:A08322. https://doi.org/10.1029/2010JA015326
European Space Agency (2009) GOCE Level 1 Data Collection. Version 1. https://doi.org/10.5270/esa-8sfucze
Finlay C, Kloss C, Olsen NMH, Tøffner-Clausen L, Grayver A, Kuvshinov A (2020) The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. Earth Planets Space 72:156. https://doi.org/10.1186/s40623-020-01252-9
Floberghagen R, Drinkwater M Haagmans R, Kern M (2008) GOCE's Measurements of the Gravity Field and Beyond. ESA Bulletin 133. https://www.esa.int/esapub/bulletin/bulletin133/bul133d_floberghagen.pdf
Floberghagen R, Fehringer M, Lamarre D, Muzi d, Frommknecht b, Steiger C, Piñeiro J, and 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):749–75. https://doi.org/10.1007/s00190-011-0498-3
Frommknecht B, Lamarre D, Meloni M, Bigazzi A, Floberghagen R (2011) GOCE level 1b data processing. J Geod 85:759–775. https://doi.org/10.1007/s00190-011-0497-4
GOCE Flight Control Team (2014) GOCE End-of-mission operations report. Technical report GO-RP-ESC-FS-6268, Issue 1, Revision 0. https://earth.esa.int/documents/10174/85857/2014-GOCE-Flight-Control-Team.pdf
IAU SOFA Board (2019) IAU SOFA Software Collection. http://www.iausofa.org
IERS (2020) International Earth rotation and Reference systems Service (IERS) Earth Orientation Center, Bulletin B. ftp://hpiers.obspm.fr/eop-pc/bul/bulb_new
Kolkmeier A, Präger G, Möller P, Strandberg T, Kempkens K, Stark J, Gessler L, Hienerwadel K (2008) GOCE-DFAC Interface Control Document, Tech. Rep. GO-IC-ASG-0005_12, EADS Astrium
Love JJ, Gannon JL (2009) Revised D_{st} and the epicycles of magnetic disturbance: 1958–2007. Ann Geophys 27(8):3101–3131. https://doi.org/10.5194/angeo-27-3101-2009
Lühr H, Kervalishvili G, Rauberg J, Michaelis I, Stolle C (2016) Advanced ionospheric current estimates by means of the Swarm constellation mission: A selection of representative results. In: Proceedings, (Special publications / European Space Agency; 740), ESA living planet symposium, Prague, Czech Republic 2016. https://doi.org/10.48440/2.3.2022.001
Lühr H, Xiong C, Olsen N, Le G (2017) Near-earth magnetic field effects of large-scale magnetospheric currents. Space Sci Rev 206:521–545. https://doi.org/10.1007/s11214-016-0267-y
Matzka J, Bronkalla O, Tornow K, Elger K, Stolle C (2021) Geomagnetic Kp index. V. 1.0. GFZ Data Services. https://doi.org/10.5880/Kp.0001
Maus S, Lühr H (2005) Signature of the quiet-time magnetospheric magnetic field and its electromagnetic induction in the rotating. Earth Geophys J Int 162:755–763. https://doi.org/10.1111/j.1365-246X.2005.02691.x
Michaelis I, Stolle C, Rother M (2021) GRACE-FO calibrated and characterized magnetometer data. GFZ Data Services. https://doi.org/10.5880/GFZ.2.3.2021.002
Michaelis I, Korte M (2022) GOCE calibrated and characterised magnetometer data. GFZ Data Services. https://doi.org/10.5880/GFZ.2.3.2022.001
Newell PT, Gjerloev JW (2012) SuperMAG-based partial ring current indices. J Geophys Res 117:A05215. https://doi.org/10.1029/2012JA017586
Nose M, Iyemori T, Sugiura M, Kamei T (2015). Geomagnetic Dst index. https://doi.org/10.17593/14515-74000
Olsen N, Sabaka T, Lowes F (2005) New parameterization of external and induced fields in geomagnetic field modeling, and a candidate model for IGRF 2005. Earth Planets Space 57(12):1141–1149. https://doi.org/10.1186/BF03351897
Olsen N, Friis-Christensen E, Floberghagen R, Alken P, Beggan CD, Chulliat A, Doornbos E, Da Encarnação JT, Hamilton B, Hulot G, Van Den Ijssel J, Kuvshinov A, Lesur V, Lühr H, Macmillan S, Maus S, Noja M, Olsen PEH, Park J, Plank G, Püthe C, Rauberg J, Ritter P, Rother M, Sabaka TJ, Schachtschneider R, Sirol O, Stolle C, Thébault E, Thomson AWP, Tøffner-Clausen L, Velímský J, Vigneron P, Visser PN (2013) The Swarm Satellite Constellation Application and Research Facility (SCARF) and Swarm data products. Earth Planets Space 65:1189–1200. https://doi.org/10.5047/eps.2013.07.001
Olsen N (2021) Magnetometer Data of the GRACE Satellite Duo. Earth Planets Space 73:62. https://doi.org/10.1186/s40623-021-01373-9
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 Space 72:48. https://doi.org/10.1186/s40623-020-01171-9 (**ISSN 1880-5981**)
Pick L, Korte M, Thomas Y, Krivova N, Wu C-J (2019) Evolution of large-scale magnetic fields from near-Earth space during the last 11 solar cycles. J Geophys Res Space Phys. https://doi.org/10.1029/2018JA026185,2527-2540. https://doi.org/10.1029/2018JA026185
Richmond A (1995) Ionospheric electrodynamics using magnetic apex coordinates. J Geomagn Geoelectr 47:191–212. https://doi.org/10.5636/jgg.47.191
Seeber G (ed) (2003) Satellite Geodesy, 2nd, completely revised and extended. Walter de Gruyter, Berlin, New York. ISBN 3-11-017549-5
Stolle C, Michaelis I, Xiong C, Rother M, Usbeck T, Yamazaki Y, Rauberg J, Styp-Rekowski KM (2021) Observing Earth's magnetic environment with the GRACE-FO mission. Earth Planets Space 73:51. https://doi.org/10.1186/s40623-021-01364-w
Swarm DISC (2022) Swarm DISC (Swarm Data, Innovation, and Science Cluster). https://earth.esa.int/eogateway/activities/swarm-disc
Wertz JR (ed) (1978) Spacecraft Attitude Determination and Control, vol 73. Springer, Netherlands. ISBN 978-90-277-1204-2. https://doi.org/10.1007/978-94-009-9907-7

Publisher's Note

Styp-Rekowski *et al. Earth, Planets and Space* (2022) 74:138
https://doi.org/10.1186/s40623-022-01695-2

Earth, Planets and Space

FULL PAPER

Open Access

Machine learning-based calibration of the GOCE satellite platform magnetometers

Kevin Styp-Rekowski[1*], Ingo Michaelis[2], Claudia Stolle[3], Julien Baerenzung[2], Monika Korte[2] and Odej Kao[1]

Abstract

Additional datasets from space-based observations of the Earth's magnetic field are of high value to space physics and geomagnetism. The use of platform magnetometers from non-dedicated satellites has recently successfully provided additional spatial and temporal coverage of the magnetic field. The Gravity and steady-state Ocean Circulation Explorer (GOCE) mission was launched in March 2009 and ended in November 2013 with the purpose of measuring the Earth's gravity field. It also carried three platform magnetometers onboard. Careful calibration of the platform magnetometers can remove artificial disturbances caused by other satellite payload systems, improving the quality of the measurements. In this work, a machine learning-based approach is presented that uses neural networks to achieve a calibration that can incorporate a variety of collected information about the satellite system. The evaluation has shown that the approach is able to significantly reduce the calibration residual with a mean absolute residual of about 6.47nT for low- and mid-latitudes. In addition, the calibrated platform magnetometer data can be used for reconstructing the lithospheric field, due to the low altitude of the mission, and also observing other magnetic phenomena such as geomagnetic storms. Furthermore, the inclusion of the calibrated platform magnetometer data also allows improvement of geomagnetic field models. The calibrated dataset is published alongside this work.

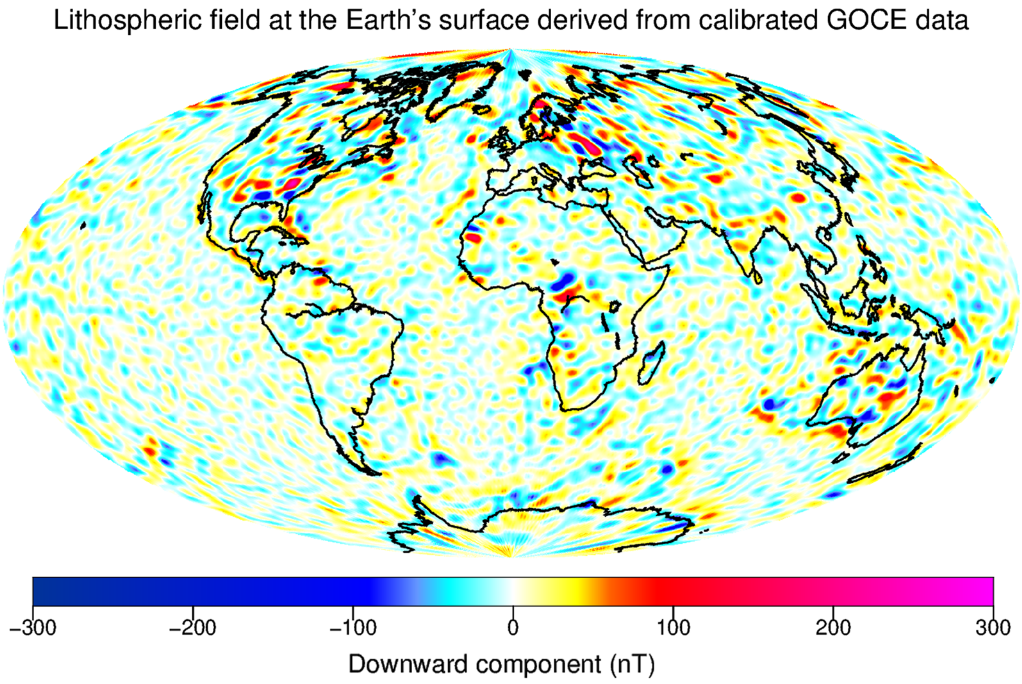
Keywords: Machine learning, Calibration, Platform magnetometer, GOCE satellite, Magnetic field model

---

*Correspondence: styp-rekowski@tu-berlin.de

[1] Distributed and Operating Systems, Technical University of Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
Full list of author information is available at the end of the article

Graphical Abstract



Introduction

Space-based observations of the Earth's magnetic field from low Earth orbit are of high value for space physics and geomagnetism due to their potential global coverage. High-precision magnetic satellite missions carrying a magnetometer package to provide absolute vector data of the geomagnetic field have revolutionized our knowledge of its distribution and variations (see, e.g., Olsen and Stolle (2012)). The most prominent missions with decade-long continuous time series are the CHAllenging Minisatellite Payload (CHAMP) (Reigber et al. 2002) and Swarm missions (Olsen et al. 2013). The measurements of these missions have led to high-quality models, describing different components of the geomagnetic field, e.g., the core, the lithospheric, and the large-scale magnetospheric field, e.g., (Finlay et al. 2020; Baerenzung et al. 2020).

In addition to high-precision magnetic mission data, the spatiotemporal coverage of magnetic measurements can be enhanced by applying observations of so-called platform magnetometers which are mounted on many satellites primarily for attitude control. Usually, they do not provide absolute data, have coarse sampling rates of one to several seconds, and are mounted with low attention to magnetic cleanliness. However, based on high-quality magnetic field models and information on the satellite attitude and on-board operations and control describing possible artificial magnetic signals, measurements of platform magnetometers can be carefully calibrated. Previously, platform magnetometer data from the GRACE, GRACE-FO, Cryosat-2, and DMSP missions have been calibrated and made publicly available (Olsen 2021; Stolle et al. 2021a; Olsen et al. 2020; Alken et al. 2020). A summary of these calibrations and examples of scientific applications in both geomagnetism and space science is given by Stolle et al. (2021b). Especially between the end of the CHAMP mission in September 2010 and the launch of the Swarm mission in November 2013, these data have had the potential to enhance magnetic field models.

In this work, we present a machine learning (ML) algorithm to calibrate platform magnetometer data of the Gravity and steady-state Ocean Circulation Explorer (GOCE) (Floberghagen et al. 2011; Drinkwater et al. 2003). The satellite was launched in March 2009 and ended in November 2013, i.e., being another satellite with the potential of bridging the previously mentioned gap of high-precision measurements. With its low altitude of about 255 km, the magnetic part of the GOCE mission may be especially interesting in detecting the lithospheric field. The GOCE satellite flies in a polar orbit of 98° inclination and the mission follows a sun-synchronous dawn–dusk orbit at local times of about 6 and 18 LT for the descending and ascending orbits, respectively. The spacecraft carries three 3-axis magnetometers of the type Billingsley TFM100-S measuring the Earth's magnetic field at a rate of 16 s (Billingsley 2020). More details on the mission and especially its magnetometer package is given by Michaelis et al. (2022).

All above-mentioned data from platform magnetometers including the dataset of GOCE by Michaelis et al. (2022) have applied analytical approaches solving a least-square problem. The purpose of this work is to develop and present a ML technique to calibrate magnetometer data with the aim to provide a method that does not need preselection of parameters that describe potential artificial magnetic disturbances. As an example, parameters like magnetorquer activations, battery currents, or sensor temperatures are known to contribute to such disturbances. Rather, all available parameters are fed into the process and the ML algorithm itself identifies relevant properties for the calibration. In addition, timing issues of the satellite clock or non-linear relationships between parameters are automatically identified. By that, we aim at providing a calibration tool for platform magnetometers that is easily applicable to other missions as well. A similar calibration has been presented earlier for the GRACE-FO satellite mission (Styp-Rekowski et al. 2021).

In the following, the Chapter ”Datasets and preprocessing” describes the datasets used and the application of different preprocessing steps. Afterward, the Chapter ”Machine learning-based calibration” describes the ML approach and its application to the data. The assessment of the results as well as examples of geophysical applications are presented in the Chapter ”Results and discussion”. Finally, our findings are concluded in the Chapter ”Conclusion”.

Datasets and preprocessing

The GOCE satellite produces data in an interval of 16 s. In a normal month containing 30 days, this means a total of 162000 data points. With an inclination of 98° degrees, the GOCE mission has a 61-day orbit periodicity, meaning that after this timespan it has covered the whole earth in a uniform pattern.

Dataset

The data used in this work consist of the same input dataset as described in Michaelis et al. (2022). The magnetometer measurements, the supporting housekeeping data, the available telemetry data, and the CHAOS7 data were interpolated onto the same timestamps because of different subsystems of the satellite measuring at different timestamps and time intervals. Contrary to Michaelis et al. (2022), all available data were used instead of selecting a subset. The measurements are recorded in the satellite frame and are provided by the European Space Agency (ESA). Therefore, calibration is performed in the same frame in which the instruments are mounted on the satellite since the available information affects the magnetometer measurements in this way. The CHAOS7 reference model was therefore also rotated into the satellite frame.

The available information was collected, interpolated, and finally merged into our final dataset. With available position and attitude information provided by the combined product of the Electrostatic Gravity Gradiometer and Star Trackers of the GOCE mission, the rotations were calculated such that the final calibration dataset of this study will be available in the satellite frame as well as the Earth-fixed North–East–Center (NEC) frame.

In this work, the measured properties or attributes by the satellite are referred to as features, which is a common term in ML. Features of the dataset used include the magnetometer readings, the magnetorquer activations, the solar array as well as battery and other available currents, temperatures, thruster activations, and also the available telemetry data of the satellite which includes a multitude of properties like status variables, flags, and others. They also carry a variety of physical units that are not considered further. Overall, there are 975 of the available 2233 features taken into account for the calibration after the application of the following preprocessing steps. The final list of used features can be found in the 'feature_list.csv' file published together with the dataset. The ML approach does not differentiate these features and will inherently select the most relevant features for the calibration. The magnetometer readings before the proposed calibration originate from the L1b product and have been pre-calibrated to a bias below 500nT.

Feature preprocessing

As previously mentioned, a multitude of features is collected to automatically identify relevant features for the calibration of the platform magnetometers. The more information can be collected and included about the satellite as a system, the better the calibration can potentially become. Before the ML approach is applied, the available data need to be preprocessed. While each new line in the data represents a new data record with an assigned unique timestamp, each of the columns or recorded measurements associated with this timestamp are referred to as features of the measurement for the rest of the work. In the first step of the analysis, these features will be converted, added, or filtered with a variety of goals.

One-hot encoding

As all available data of the GOCE mission are used, a lot of telemetry data are included as well which is partly delivered as textual information, e.g., certain states of systems being encoded with the literals 'AVAILABLE' or 'NOT_AVAILABLE'. As some features have more than 2 possible literals, a sophisticated method is needed to convert this textual information into numerical information which is usable by statistical models for the calibration. For each of the possible literals of a feature available, another new feature is generated which encodes a 1 if the feature equals this literal and 0 otherwise, this is also called one-hot encoding or dummy coding. Thus, if a certain literal that was recorded in a feature has an influence on the calibration, the calibration can utilize the associated feature. In addition, with this technique every resulting new feature has the same euclidean distance contrary to assigning number to every different literal available within the feature value range.

Removed features

Several features are explicitly not used for the calibration to perform well. These include timestamps, magnetic local time (MLT), and positional data like the latitude, longitude, or radius as these have the potential to encode positional information which needs to be avoided to not misguide the model calibration. When training against the reference model, the objective is not to learn an efficient mapping of position data to reference model values, but to model the magnetic conditions of the satellite system itself under different system activations. For the same reason, features like recorded measurements of the star trackers are removed as they inherit a potential to encode positional information about the satellite.

Additional features

The solar activity, given by the F10.7 index, an 81-day moving average of it as well as the day of the year were added as additional features (Tapping 2013). These were added to correct for potential influences from the angle of the satellite towards the sun or dependencies on the solar activity. Likewise, the 3-1-3-Euler angles were calculated using the quaternions and added as additional features to give the calibration model the possibility to adjust for inaccuracies in their estimation.

Missing data

Data gaps occur several times throughout the duration of the mission. These gaps are not part of the calibration because the data are unavailable. In addition, some positional data are either missing or measured incorrectly. For a minority of the data, there is no information about the position of the satellite, these data points have been removed. This was especially the case near data gaps like in July 2010. Gaps are caused either by missing data from the magnetometers or by lack of essential information such as position or attitude Also, data where the interpolation distance during the data gathering process described in Michaelis et al. (2022) was larger than 16 s are removed from the calibration process and flagged in the error flag "B_FLAG" as this data is considered similar to missing data because the data has to be considered uncertain.

NaN-filling

If only small shares of the additional housekeeping data are missing, a filling strategy is used to make these data usable again. For each feature, the share of missing data is calculated. If the share of missing data within a feature is smaller than 20% of the available data, the information situation is considered to be well enough to take these features into account as the features deliver potentially beneficial information in at least 80% of the cases which constitutes the majority. For these features, the missing share is filled up with the mean of the present values. The data points are flagged with an introduced "NaN-Flag" which indicates that data have been substituted with non-original measurements, 1 indicating that at least one feature value has been filled, 0 indicating no manipulation of the data.

Magnetic quiet time filtering

As the satellite is calibrated post-launch, the data need to be filtered for calibration to ensure that only the magnetic system of the satellite itself is modeled. Therefore, natural disturbances need to be removed before calibration, as they shall remain in the calibrated measurements of the satellite. The Kp and Dst indices are good indicators for magnetic activity. The data are filtered for values of Kp$\leq$2 (Matzka et al. 2021) and Dst$\leq$30 (Nose et al. 2015) which is considered to contain magnetic quiet times without omitting too much data for the training. Thus, the calibration shall model the underlying process of the satellite. Data filtered out like that is flagged with the "KP_Dst_Flag", 1 indicating a too high magnetic activity, 0 indicating low magnetic activity.

Outlier removal

Despite careful selection of the data and filtering for missing positional information as well as magnetic active times, some data points deviate strongly from their expected value. It is suspected that this behavior originates in a bad attitude estimation of the satellite but the underlying reasons can be manifold. To identify such data, the raw measurements of the magnetometers are compared to the reference model without any further calibration. The raw data follow the rough pre-calibration of the platform magnetometers as used during the mission. From the remaining residuals, the mean $\bar{x}_j$ is calculated as well as its standard deviation (STD). Then, it is checked whether each data point falls within the range of three standard deviations around the mean:

$$\text{with } \bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{i,j} \tag{1}$$

$$|x_{i,j}| \leq |\bar{x}_j| + 3 * \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{i,j} - \bar{x}_j)^2} \tag{2}$$

for all data points i and all magnetometer measurements j. This means at least 99.7% of data lies within the defined range if the data are Gaussian distributed which holds true for the GOCE dataset. If at least one of the magnetometer measurements deviates stronger from the mean, the whole orbit is considered to be an outlier and removed from the training data of the calibration, because analysis has shown that the whole orbit of the satellite behaves unusual when outliers are detected within the orbit. In addition, this data is flagged within the error flag ”B_Flag” and is part of the published dataset.

CHAOS-7 model

As the reference model for the calibration, the CHAOS-7.8 model has been used (Finlay et al. 2020). Utilizing a variety of magnetic measurements, including ground station observations as well as satellite data from current and previous high-precision magnetic satellite missions, the CHAOS-7 model is able to precisely model the core, the lithospheric, and the external field. This reference model is especially needed as the platform magnetometers onboard the GOCE satellite mission do not have an absolute component, but rather measure the Earth's magnetic field relatively. Thereby, similar to Michaelis et al. (2022) the reference model has been evaluated at each position of the satellite and was then rotated into the satellite frame for calibration purposes.

Machine learning-based calibration

The proposed method utilizes neural networks (NNs) to train a ML-based model for the calibration and characterization of the platform magnetometer measurements. In the proposed approach, the calibration and characterization are accomplished in one method and thus referred to as calibration in the remainder of this work. Contrary to previous work done on the calibration of platform magnetometers like for the GRACE, GRACE-FO, or Cryosat-2 satellites, there are several differences in the approach used within this work. In the proposed approach, the features are not selected manually based on experience about their relevance on improving the calibration result, rather as much information as possible about the satellite as a magnetic system is collected and presented to the calibration model which is able to choose the relevant features. Also, no interactions between the features need to be hand-crafted, e.g., polynomial combinations of the magnetometer axes, since the calibration method used has inherent non-linear modeling capabilities. Another difference lies in the usage of data from all available latitudes, this also includes the high-latitude area containing field-aligned currents (FACs).

Utilizing a large number of collected features as described in the data preprocessing step, renders linear models like linear regression unsuitable as the underlying statistical problem is over-parameterized and thus the calibrated model overfits the data. Overfitting means that the model performs very well on the seen data, but generalizes poorly to unseen data. NNs can overcome this problem as the number of neurons is limited and thus, the number of used features gets limited. In addition, the batch-wise learning of the stochastic gradient descent algorithm prevents the use of single uncorrelated features. However, NNs need a large number of data points for the gradient-based approach to converge to an optimum, representing a good calibration of the magnetometer measurements. This is also referred to as training the NN as the weights representing the model are adjusted within this process. This is contradictory to the slowly deteriorating instruments onboard the satellite system for which many temporally separated models would be better capable of representing the changing behavior of the instruments. Therefore, the approach shown in Fig. 1 was developed. After collecting, filtering,

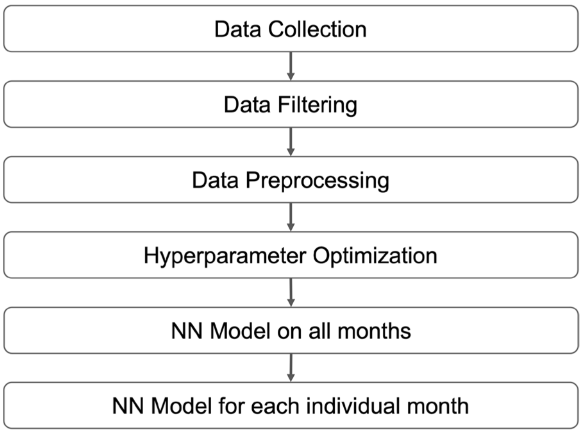


Fig. 1 Overview of the calibration process

and preprocessing the data as described in the previous section, first a Hyperparameter Optimization (HPO) on a randomly selected subset of the dataset is performed to determine the NN architecture and training parameters (see section below), then a NN model on all available data from the life span of the mission from November 2009 until September 2013 is trained using the found architecture, representing a global model of the calibration of the magnetometers. Afterward, a more fine-granularly trained model is trained for each month individually based on the global model, this is done with a much smaller learning rate when adjusting the weights of the NN. By using this approach, there is enough to train a global model which was presented with all variations of the data's statistics while still maintaining multiple models representing the differences between individual months, thus reflecting the change in the instruments throughout the duration of the mission. The fact that the GOCE mission flies in a sun-fixed orbit with stable MLTs supports the training of a global model as no changes in the behavior of the parameters are to be expected.

Before starting the training process, the STD of every feature as well as the auto-correlation among the features have been calculated. Features having an STD of 0 (depending on the computer precision) have been removed, meaning that these contain only constant values with no additional information for the calibration process. In addition, features having a correlation of 1 (depending on the computer precision) with any other feature have been removed as these represent the exact same information twice and are thus treated as duplicates of the same feature and removed. The STD-filtering removed 442 features, while the following correlation-filtering removed 780 additional features, as the telemetry dataset contains a lot of duplicated information. The reduced dataset dimension also helps to reduce the training time needed for the ML approach. In addition, the features are normalized with a min–max normalization to an interval of [− 1, 1], which is a technique by which the gradients converge to an optimal solution more smoothly when training NNs. This also ensures that each feature and the variations of the values within the feature are weighted equally during the training of the model. As mentioned earlier, there are 975 of the 2233 features taken into account during the calibration which includes the magnetometer measurements, the housekeeping data, and the telemetry data. Also, the applied filtering methods led to a reduction from about 6.4 million data points to about 4.8 million data points.

As mentioned earlier, NNs are not very well able to generalize to unseen values of features. Therefore, the training of the NN cannot be limited to the low- and mid-latitude region where the reference model is known to constitute the ground truth correctly as there are no magnetic phenomena in this region that are not modeled by the CHAOS-7 model. On the other hand, in the high-latitude regions the values measured by the magnetometers lie beyond the interval of previously seen values in the low- and mid-latitude regions, thus the NN would not be able to generalize well on these. Therefore, the high-latitude region has been taken into account for the training of the NN as well. Similar to previous work done with the ML-based calibration of the GRACE-FO satellite, a weighting was applied for the samples depending on the quasi-dipole latitude (QDLAT) they are originating from and based on the trustability of the reference model for these QDLATs. The low- and mid-latitude region samples have been weighted the highest, while the regions from 50° to 60° QDLAT have been weighted slightly lower and the high latitude regions beyond 60° QDLAT have been weighted the lowest as described in the following. The FACs appearing in the high-latitude regions were the main motivation for applying this weighting. The weights are calculated using the number of samples S_i of the three zones indexed by i in a relationship maintaining manner so that they must satisfy the following equation:

$$w_1 * S_1 + w_2 * S_2 + w_3 * S_3 = S_1 + S_2 + S_3. \tag{3}$$

Because this system of equations is undersatisfied with three weights, an additional ratio is given, set at $1 : \frac{1}{4} : \frac{1}{160}$, based on a rough estimate on the influence of FACs in the areas, so that samples from the second zone are weighted $\frac{1}{4}$ compared to the first zone, and samples from the third zone are weighted $\frac{1}{160}$ compared to the first zone. The reference model is mostly accurate for these regions as well, still, the samples containing FACs shall not be included with full weight as that would imply the model to calibrate these alongside the magnetometer calibration to meet the constraints imposed by the reference model, effectively setting the nulling of FACs as the goal of the modeling process. Thus, the final weights for the whole training dataset were set to about 1.68, 0.42, and 0.01 for the weights w_1, w_2, and w_3, respectively. When calculated for each month separately, they vary slightly from these values. The applied gradient generally is the result of the loss function that sums the equally weighted losses of batches, so only reducing the weights would lead to smaller gradients. A proportionally equally large gradient is achieved through rescaling of the weights.

Machine learning approach

Feed-forward NNs are general function approximators that are trained using the stochastic gradient descent algorithm (Hornik et al. 1989). This means that NNs can approximate any given function given enough complexity of the NN. The calibration of platform magnetometers can also be formulated as a function

$$f(x) = \hat{y}, \tag{4}$$

where x corresponds to our input including the 9 measurements of the 3 magnetometers as well as the previously processed housekeeping and telemetry data, f corresponds to the NN as the calibration function on our inputs, and $\hat{y}$ corresponds to the calibrated measurement by the platform magnetometers for the given input x. The function, or NN, is then trained using the mean squared error (MSE) which is the squared error between the predicted output $\hat{y}$ and the expected output derived from the CHAOS-7 reference model y:

$$\epsilon = (y - \hat{y})^2. \tag{5}$$

Utilizing this error ϵ, the gradients to adjust the weights of the NN are calculated and backpropagated through the different layers of the NN, thus optimizing for a low residual between the prediction of the calibration function and the expected output, reassembling a good calibration function. As there exists only one definitive position of the satellite for each measurement, there exists only one evaluation of the reference model for a given measurement. Therefore, the calibration function treats all three magnetometers as the input and outputs only one calibrated measurement. A detailed introduction to how NNs work can be found in Appendix A.

In addition, a lot of configurations for the training of the NN can be altered. For calculating the gradients during the training, a batch of data points is considered and a gradient for the whole batch is calculated for proceeding, the size of this batch is referred to as the batch size. The rate at which the calculated gradient is taken into account to adjust the weights of the NN is called the learning rate. While training the NN a step-decaying learning rate function has been chosen which reduces the learning rate after certain training progress. Also, there are a variety of activation functions to choose from as well as an arbitrary number of possible NN architectures, resulting from the number of neurons and layers used. Therefore, the architecture, as well as the configuration of the training parameters, needs to be explored which is done in the HPO.

Hyperparameter optimization

For the HPO, a variety of training and architecture configurations have been stochastically tested using the Bayesian Optimization algorithm (Snoek et al. 2012). The Bayesian Optimization algorithm is a Gaussian process that evaluates the trained NN with a certain parameterization and thus models the parametrization space connected with the target function as a Gaussian process. The search for an optimal set of parameters is then guided towards parameterizations of the NN which lead to an optimum of the target function, the smallest calibration residual, respectively. For the HPO, a variety of parameters and parameter ranges have been investigated, including the number and sizes of neuron layers within the NN architecture, the kind of activation functions used, the batch size, the number of training epochs, the learning rate, and the parametrization of the step-decaying learning rate function.

Since the model must be evaluated for each parametrization tried, a randomly chosen subset of the mission data, representing about 16% of the whole data or roughly 750000 data points, has been chosen for the HPO. This was mainly done because of performance reason as it reduces the training time of each evaluation, but still represents enough data for the NN to converge to an optimum and the data to reflect well the distribution of the entire data set. In addition, this HPO data are divided into two parts, 80% or about 600000 data points of the HPO data being used for the training while 20% of the HPO data are being used to test the trained NN. This is a common technique in ML since 80% of the HPO data is a good representation of the distribution of the data and enough data points for training the ML model while the remaining 20% of the HPO data are still a significant part of the data to test whether the model also performs well on unseen data. The error values of the trained NN will be compared between the training and test dataset to detect possible overfitting as well as the performance of the model. The score of the evaluation is then set to the MSE of the test part of the dataset.

The HPO resulted in the following parametrization of the NN: The learning rate was set to 0.01, the batch size to 1500, the number of epochs to 1200, the activation function as the Exponential Linear Unit (ELU) (Clevert et al. 2015), and the decaying function halving the learning rate every 90 epochs.

For the architecture of the NN, similar results have been suggested with a triangular-shaped architecture, consisting of a large first layer, a medium-sized second layer, and the output layer consisting of 3 neurons. Analysis has shown that with increasing layer sizes, the residual becomes smaller, but the further improvement was caused by reduced FAC amplitudes. Hence, a tradeoff between a large architecture for a small residual and a small architecture that retains the FACs had to be found. Therefore, as a compromise, the architecture was set to consist of the first layer with 384 neurons, the second layer with 128, followed by the output layer with 3 neurons.

Fine-tuning and timeshift analysis

After using the found parametrization for the architecture and training of the NN, the global model was trained on all available data of the mission. Again, a split of 80% of the about 4.8 million data points for the training and 20% of the data points for testing has been used. Similar to before, this is a common step in ML to evaluate the model's quality by evaluating the model on unseen test data. In a second step, this global model was fine-tuned with the monthly data of the different months using following NN parameters: 200 epochs and a small learning rate of 0.00001, allowing only for small specific adjustments. Thus, a different calibration is found for each month which is derived from the same initial global calibration.

In addition to the fine-tuning of the model parameters, a possible time shift in parts of the data has been searched for, similar to the time shift found in the analytical calibration of the platform magnetometer of the GRACE-FO or GOCE mission (Stolle et al. 2021a; Michaelis et al. 2022). Therefore, in parallel to the fine-tuning of the global model, for every month the interpolation neuron presented by Styp-Rekowski et al. (2021) has been used to search for a possible time shift. This interpolation neuron automatically finds a time shift which reduces the overall residual of the calibration. The results were averaged over all months and are presented in Table 1. The overall mean absolute error (MAE) which is defined as the average absolute residual over all samples as well as the MSE defined as the squared residual over all samples are very close when comparing the training and test scores which is a good sign of the model fitting the data distribution well. For the magnetometers, a mean shift of −0.51s, and for the magnetorquers, a mean shift of −3.61s have been found. But, the STD for both shifts found is very high, with 1.79s and 2.45s, respectively, rendering these time shifts very inconsistent across all months. As the time shifts cannot consistently be found during the fine-tuning of independent monthly data, they are considered to represent local minima, and thus they are not incorporated into the final solution, leaving the data as is. Also, the time shift of 0.4 s as found by Michaelis et al. (2022) has been evaluated with no observable improvements.

Final architecture

To sum up, the final approach uses the parametrization found in the HPO and applies no time shift. The carefully preprocessed data are thus input into the NN with 3 layers of 384 neurons, 128, and 3 neurons as shown in Fig. 2. As depicted in Fig. 1, the data are first collected, filtered, and preprocessed to then train the global model for 1200 epochs as described in the HPO chapter. Finally, the model is fine-tuned month-wise without applying a time shift with a lower learning rate of 0.0001 for 200 epochs. These trained monthly models are then used to generate the calibrated dataset of the platform magnetometers of the GOCE satellite.

Results and discussion

After producing the calibrated dataset, the resulting measurements of the Earth's magnetic field have been analyzed. First, the remaining residual of the calibrated data compared to a global magnetic field model is evaluated, followed by different magnetic phenomena like the FACs and magnetic storm behavior. Finally, the calibrated data have been used to support the modeling of the magnetic field of the Earth.

Residual to reference model

Table 2 shows the residuals of the calibrated dataset against the reference model CHAOS-7. Unless otherwise

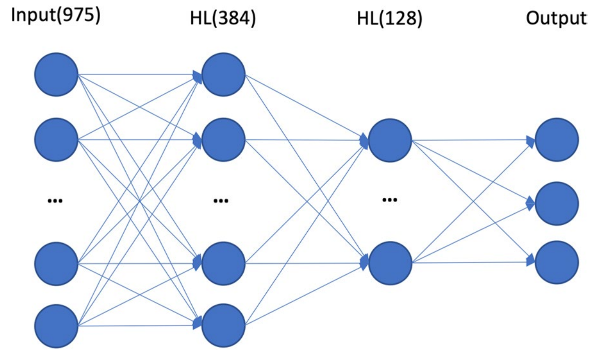


Fig. 2 Neural network architecture, with the inputs on the left side, two hidden layers (HL) and the 3-dimensional output standing for the 3-axis calibrated measurement

Table 1 Summarized residual and offset data over all available months for low- and mid-latitudes. Errors and standard deviation given in nT, while the offsets for the magnetometers (MAG) and magnetorquers (MTQ) are given in seconds

	MAE		MSE		StdDev		MAG	MTQ
	Train	Test	Train	Test	Train	Test		
Mean	13.42	13.55	854.95	882.26	28.32	28.57	− 0.51	− 3.61
Standard Dev.	3.57	3.71	464.58	552.58	7.34	8.20	1.79	2.45

Table 2 Error values in nT after fine-tuning training of the neural network for every month, on the left for the whole latitude region, on the right for low- and mid-latitudes

Month	90-QDLAT		50-QDLAT	
	MAE	STD	MAE	STD
	Train | Test	Train | Test		
200911	8.98 | 8.89	20.08 | 19.40	4.63	5.81
200912	8.88 | 8.85	19.31 | 19.07	4.61	5.78
201001	9.96 | 9.94	22.29 | 22.27	4.68	5.94
201002	10.48 | 10.38	23.50 | 23.72	4.67	6.08
201003	11.12 | 11.11	24.29 | 24.47	5.37	10.10
201004	12.05 | 12.21	26.51 | 26.75	5.37	6.88
201005	16.38 | 16.60	33.34 | 34.56	8.38	10.36
201006	17.14 | 16.93	35.20 | 34.08	8.54	10.50
201007	20.01 | 20.67	39.74 | 41.52	9.62	13.17
201009	9.94 | 10.53	20.53 | 22.34	5.56	9.60
201010	10.45 | 10.44	23.38 | 23.02	5.00	6.37
201011	10.52 | 10.45	23.24 | 22.84	4.86	6.15
201012	10.36 | 10.45	22.86 | 23.16	4.87	6.14
201101	11.02 | 10.97	24.16 | 23.76	5.38	9.95
201102	10.91 | 10.99	24.38 | 25.60	4.95	6.24
201103	10.96 | 10.99	25.20 | 25.24	4.85	6.14
201104	12.76 | 12.80	29.19 | 29.39	5.54	7.12
201105	15.58 | 15.37	34.06 | 33.09	7.04	8.90
201106	19.29 | 19.22	40.96 | 40.99	9.12	11.12
201107	18.70 | 19.09	40.11 | 41.03	8.74	10.71
201109	11.05 | 10.92	23.64 | 22.72	5.20	6.73
201110	10.82 | 11.02	23.34 | 23.78	5.15	6.63
201111	13.15 | 12.86	29.03 | 28.23	6.41	9.16
201112	11.35 | 11.41	24.51 | 24.87	5.75	7.49
201201	13.66 | 13.67	30.82 | 31.41	5.73	7.38
201202	11.66 | 11.54	25.82 | 25.59	5.09	6.46
201203	14.09 | 14.03	32.18 | 31.75	5.30	6.78
201204	14.35 | 14.45	33.20 | 33.20	5.48	7.07
201205	17.27 | 17.44	37.74 | 38.03	7.69	9.68
201206	18.38 | 18.21	39.55 | 38.63	9.24	11.57
201207	21.84 | 22.38	49.22 | 49.70	10.85	15.88
201208	15.88 | 16.09	36.26 | 36.56	6.22	8.00
201209	12.52 | 12.57	28.39 | 28.27	5.68	7.95
201210	11.29 | 11.13	24.42 | 23.38	5.48	7.09
201211	11.98 | 11.98	26.56 | 26.80	5.51	7.49
201212	11.50 | 11.57	26.13 | 26.00	5.42	6.91
201301	10.72 | 10.92	22.81 | 23.40	5.49	7.14
201302	24.03 | 23.90	53.19 | 52.30	14.99	43.09
201303	10.87 | 10.85	23.34 | 23.26	5.29	6.84
201304	12.47 | 12.43	27.60 | 27.10	5.68	7.29
201305	18.18 | 18.67	39.63 | 40.27	8.09	11.03
201306	18.69 | 18.82	39.99 | 39.49	9.20	11.33
201307	18.74 | 18.90	40.56 | 40.99	8.73	10.88
201308	15.16 | 15.67	35.63 | 37.21	6.01	8.71
201309	11.97 | 11.78	26.19 | 25.13	5.70	7.51

noted the residuals to the CHAOS-7 predictions of core, lithosphere and large-scale magnetospheric field are considered in the following sections. Each row shows the average residual for every month with the same data selection as described in the previous chapter. The left side shows the training results for the whole latitude range of the Earth, while the right side shows the low- and mid-latitudes between the 50° QDLAT. The residual is given in mean absolute error (MAE), mean squared error (MSE), and STD. As mentioned earlier, for the training the data have been randomly split into 80% of the data representing the training data, and the remaining 20% representing the test data. For the training results on the left side, the residual is always given for the training and test data.

It can be seen that the generalization error, meaning the gap between the training and test dataset is very small, indicating that the model performs similarly on unseen data. This means that the model was able to adapt to the statistics of the data in a way that holds true for unseen data. Naturally, it can be observed that the residual for the low- and mid-latitudes is smaller than for the whole latitude range as the highly fluctuating FACs are measured by the satellite in this range while not being included in the reference model, thus resulting in a higher error. For the low- and mid-latitude, a consistent residual of below 10 nT can be observed which enables scientific studies. A trend is visible, that around June solstice, in the months of May, June, and July the residuals with a range of 8-10 nT seem to be higher than for the majority of the other months with 4-6 nT, which seems to be related to a higher STD and will be discussed later. The February of 2013 appears to be an outlier.

Residual maps

In Fig. 3, the residuals of the calibrated platform magnetometer measurements against the CHAOS-7 reference model are plotted as a function of latitude and longitude for the North, East and Center B-field components on global maps. The data shown are annual averages for 2012, which have been determined by averaging all data from 2012 in bins of 4° latitude by 4° longitude. The averages have been assigned a color according to the scale in the bottom right, plotted as a contour plot.

It can be seen that the calibration performs well, especially in low- and mid-latitudes where grey is the dominating color, meaning that the mean remaining residual is close to 0. In high latitudes, the shapes of the high residuals clearly resemble the known patterns of FACs, which are not included in the reference model, thus indicating that this signal remains in the calibrated satellite measurements as desired. In the Center component, there is an anomaly visible along the magnetic equator. This remaining residual is rather small in amplitude and is suspected to be of artificial origin, as it was also observed in the calibration of the platform magnetometers of GRACE-FO and Cryosat-2, which use the same type of instrument. For the GOCE satellite, Michaelis et al. (2022) found the same anomaly which remained visible although a multitude of features have been used with the proposed approach.

Comparison of orbit residual

In Figures 4, 5 and 6 the residual of the calibration against the reference model CHAOS-7 is plotted as a function of the QDLAT. Each figure shows exemplary either the ascending or descending orbits of one axis of the platform magnetometer measurements for the whole year of 2012, separated by months starting on top with January to December on the bottom. Each drawn line represents the calibrated values of one corresponding half-orbit within that month, while the mean measured value for every QDLAT is depicted in black. It can be seen that the calibration performs especially well in the low- and mid-latitude region while in the high-latitude regions the remaining higher residuals mainly are due to the FACs that are not modeled in the CHAOS-7 model but measured by the satellite. For the southern hemisphere, it can be observed that the FACs are being correlated with features by the NN model as they are lower in amplitude and in their mean value than expected. Currently, it is subject of further research to detect the features which were used by the NN model to correct for the FACs as this occurs only in the southern hemisphere.

The higher STD in the northern hemisphere summer months is evident in a larger spread of individual orbit residuals (colored areas in the plots) for the June solstice in the months of May, June, and July, present in all three components and both orbital directions, supporting the assumption that the measurements contain higher noise as there is no systematic pattern detectable. The origin of the higher deviation is unknown. Although the complete set of housekeeping and telemetry data was used in the calibration process, that gives a multitude of information about the satellite system, there was no feature able to correct this remaining error. Thus this is subject to further investigation. Overall, the calibration works well as the residual is generally low as well as the STD, visible as the colored areas in the figures.

The remaining, artificial residual around the magnetic equator in the Center component mentioned above is found here too in both the ascending and descending orbits, and with a sinusoidal shape around the zero crossing of this magnetic component. For the East component, there is an increasing trend visible

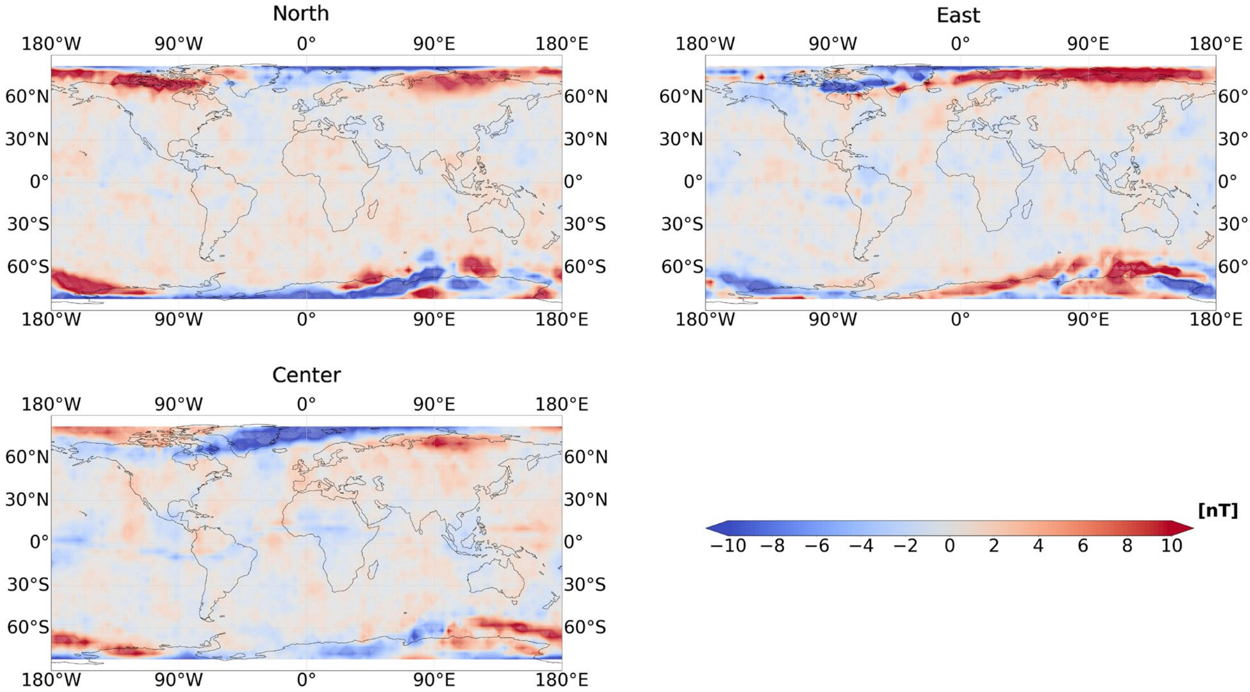


Fig. 3 Binned and averaged residual map between the calibrated platform magnetometers and the CHAOS-7 reference model for the year 2012 in the NEC Frame, North component in top left, East in top right, Center in bottom left. The color scale for all panels is given in the bottom right, and saturation of color is achieved beyond -10 or 10 nT, respectively. Residuals are low while FAC structures remain visible

around the equatorial region where the variation of the residual increases, especially for the descending orbits. This increased variation has been observed to increase throughout the duration of the mission from 2009 until 2013. It is suspected to be correlated to the solar cycle which had its minimum in December 2008 and then increased up to its maximum in April 2014.

Comparison of analytical and ML method

The calibrated dataset presented in this study is compared to the analytically calibrated dataset from Michaelis et al. (2022), in the following referred to as the Ana approach. Therefore, the Common Data Format (CDF) product files have been utilized to compare the residuals as well as differences between the calibrations. In addition, the proposed dataset was used to similarly compare against a magnetic storm event.

Residual distribution

First, the residual distribution of the ML approach with the CHAOS-7 reference model is compared to the residual distribution of the Ana approach with the CHAOS-7 reference model in Fig. 7, this was done on the December 2009 data that represent the most magnetically quiet month. The three panels show the North, East and Center component residuals of the Ana distribution in blue and the residuals of the ML distribution in orange. It can be seen that all distributions approximately follow a Gaussian distribution curve where the ML approach has a smaller standard deviation, manifesting in a steeper Gaussian distribution. This is especially the case for the North and East component where the lower overall residual of the ML approach appears to originate. The Center component looks very similar in both approaches. In addition, all distributions are centered near zero. For the East component there appears to be a small shift to the left which is suspected to be originated in the different treatment of the FACs in the southern hemisphere when comparing both approaches.

Figure 8 shows the difference between the ML and Ana calibration as a function of the calibrated measurements of the Ana approach, separated for the three North, East, and Center components, again for the most magnetically quiet December 2009, please note the different scales. The two calibrations mainly differ in the FAC regions that lie in the high-latitude regions, this means up to a value of 10000nT for the North component, the whole value range for the East component, and below about -40000nT and above about 40000nT for the Center component. We note that the difference is larger in the

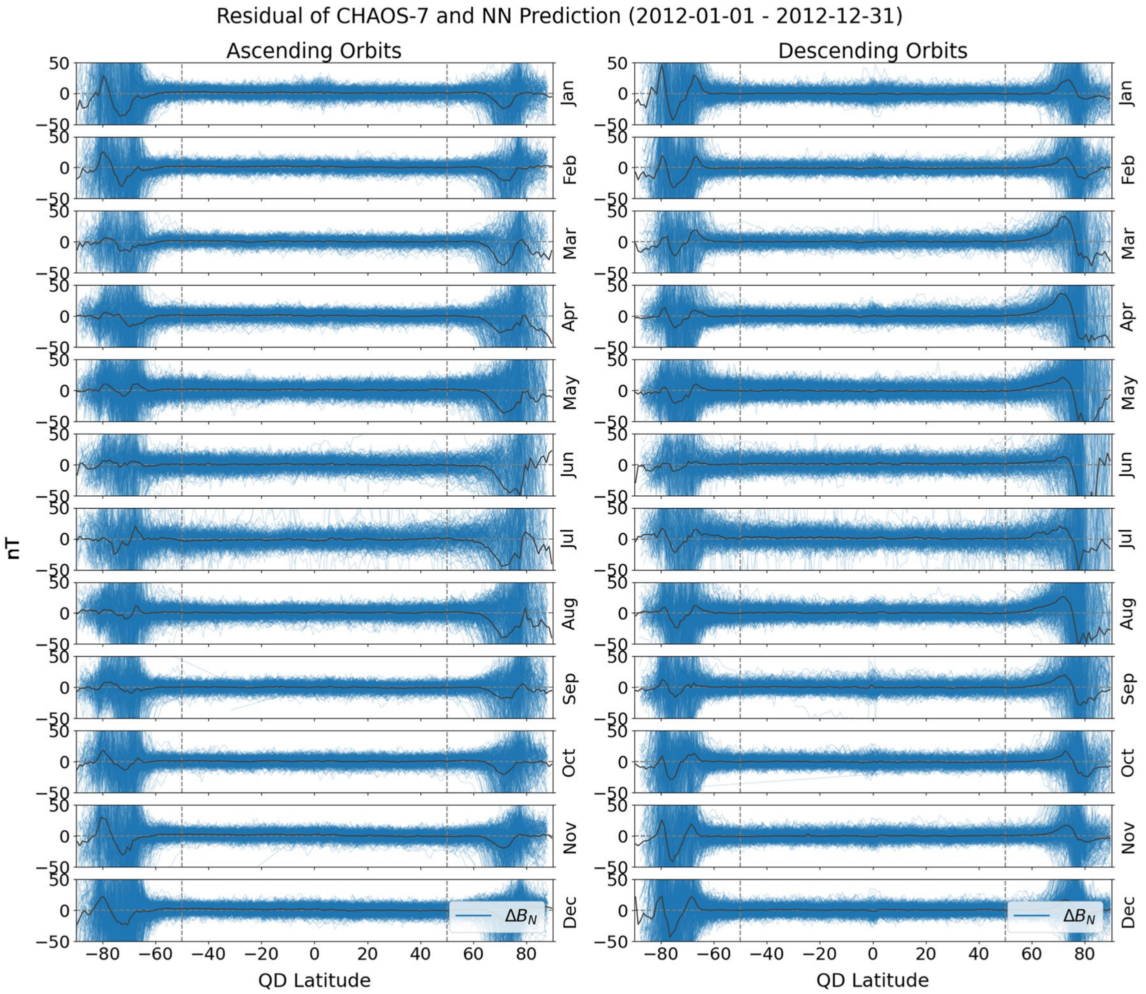


Fig. 4 Residual between the calibrated platform magnetometers and the CHAOS-7 reference model as a function of the quasi-dipole latitude (QDLAT) for every month of the year 2012 of the North component of the calibrated measurement in nT. The orbits are split into the ascending dusk-orbits on the left, and the descending dawn-orbits on the right, with the mean residual depicted in black

southern hemisphere (negative Center component), which is related to underestimation of FAC amplitudes in the ML method in this hemisphere described above that is currently under investigation. In addition, the regression curves show that there is no systematic difference in dependence of the value ranges between the two calibrations as the slope of the function is approximately zero. The shift indicates a difference between the two calibrations that lies mostly in the bit resolution of the measurement, which is about 3.05nT, being defined by the the last bit of the magnetometer measurement.

Comparison to Dst index

Figure 9 shows a magnetic storm in March 2013 as characterized by the Dst index that is obtained from the data of four low latitude ground magnetic observatories (Nose et al. 2015). The deviation of the horizontal component dB_H of the calibrated measurements from the CHAOS-7 core and lithospheric field model is plotted in orange and blue for the ascending and descending orbits, respectively. As the GOCE satellite flies in dusk–dawn orbits, the ascending measurements correspond to an MLT of about 18 while the descending measurements correspond to an MLT of about 6. It can be seen that the measured power follows very well the Dst index, thus showing the capability of the GOCE satellite to measure magnetic storm events. In addition, the ascending measurements are deviating stronger from the origin and the Dst index which is an expected phenomenon for the dusk side measurements, also known as the dawn–dusk asymmetry described by Anderson et al. (Anderson et al. 2005). The dusk side of the magnetic field is showing stronger

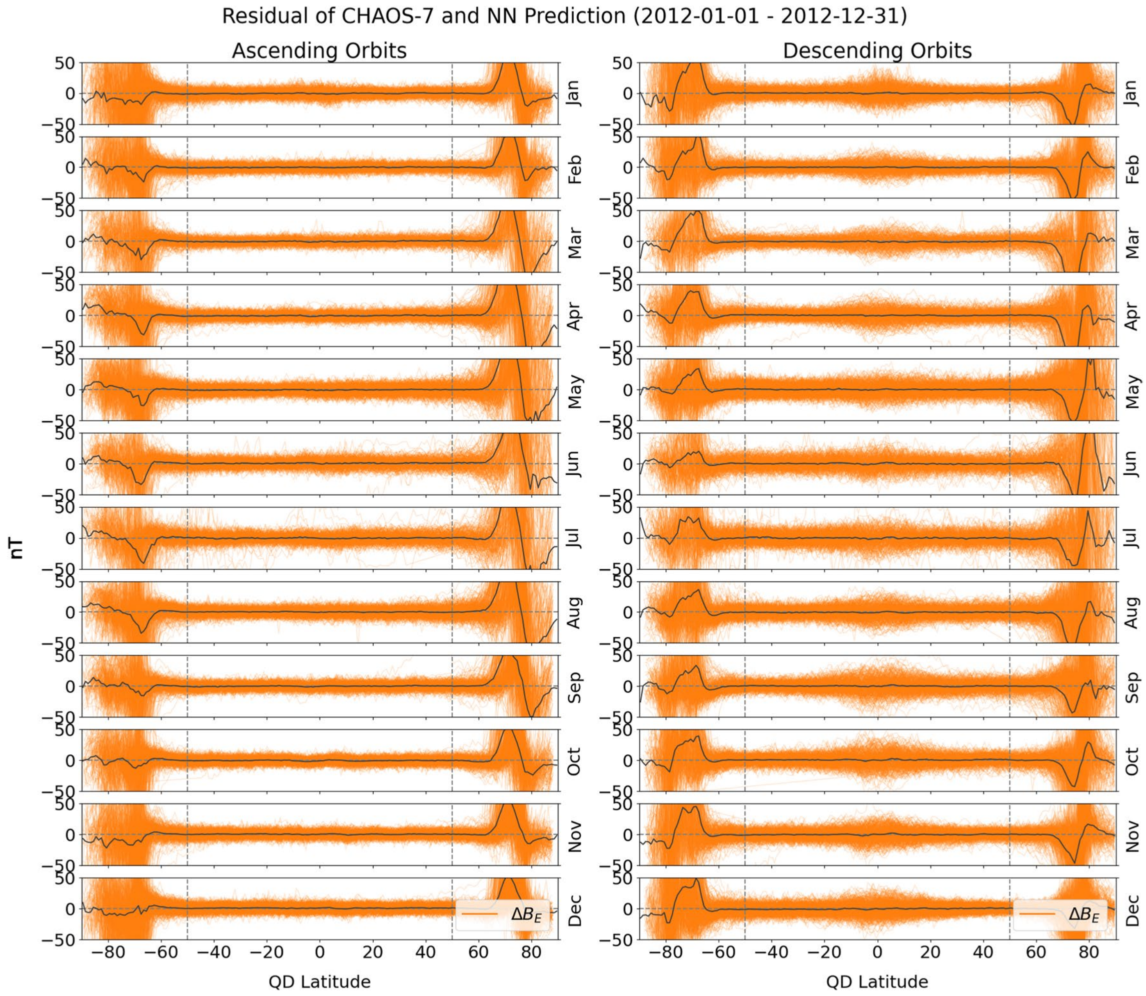


Fig. 5 Residual between the calibrated platform magnetometers and the CHAOS-7 reference model as a function of the quasi-dipole latitude (QDLAT) for every month of the year 2012 of the East component of the calibrated measurement in nT. The orbits are split into the ascending dusk-orbits on the left, and the descending dawn-orbits on the right, with the mean residual depicted in black

deviations during a magnetic storm. Similar behavior was observed for other storms during the lifetime of the GOCE mission. Similar results were obtained by Michaelis et al. (2022), confirming the quality of the calibration obtained, with even clearer separation by dusk and dawn.

Lithospheric field analysis

With an altitude of about 255km, the GOCE satellite has a rather low altitude compared to other magnetic satellite missions. Therefore, an analysis of the measured lithospheric field has been conducted.

As the data have a timely resolution of 16 s, the whole mission period has been taken into account for extracting the lithospheric field. The lithospheric field is known to not change within a time frame of about 4 years. Thébault et al (2021) recently proposed a high-resolution lithospheric field model which is depicted in the middle panel of Fig. 10. The top panel shows the lithospheric field as given by the ML-calibrated measurements of the GOCE satellite, after subtraction of the CHAOS-7 core and large-scale magnetospheric field. In general, a good agreement can be seen between the measured data of the GOCE satellite and the lithospheric field model. Anomalies like the Bangui in Africa, or the Kursk anomaly in Russia are detectable, as well as the striped ocean bottom structures in the Atlantic. The third panel shows the difference between the high-resolution model and the reconstructed lithospheric field from the GOCE measurements which, despite the small scale, shows nearly no saturation and has pale colors. This suggests that by incorporating GOCE data into lithospheric

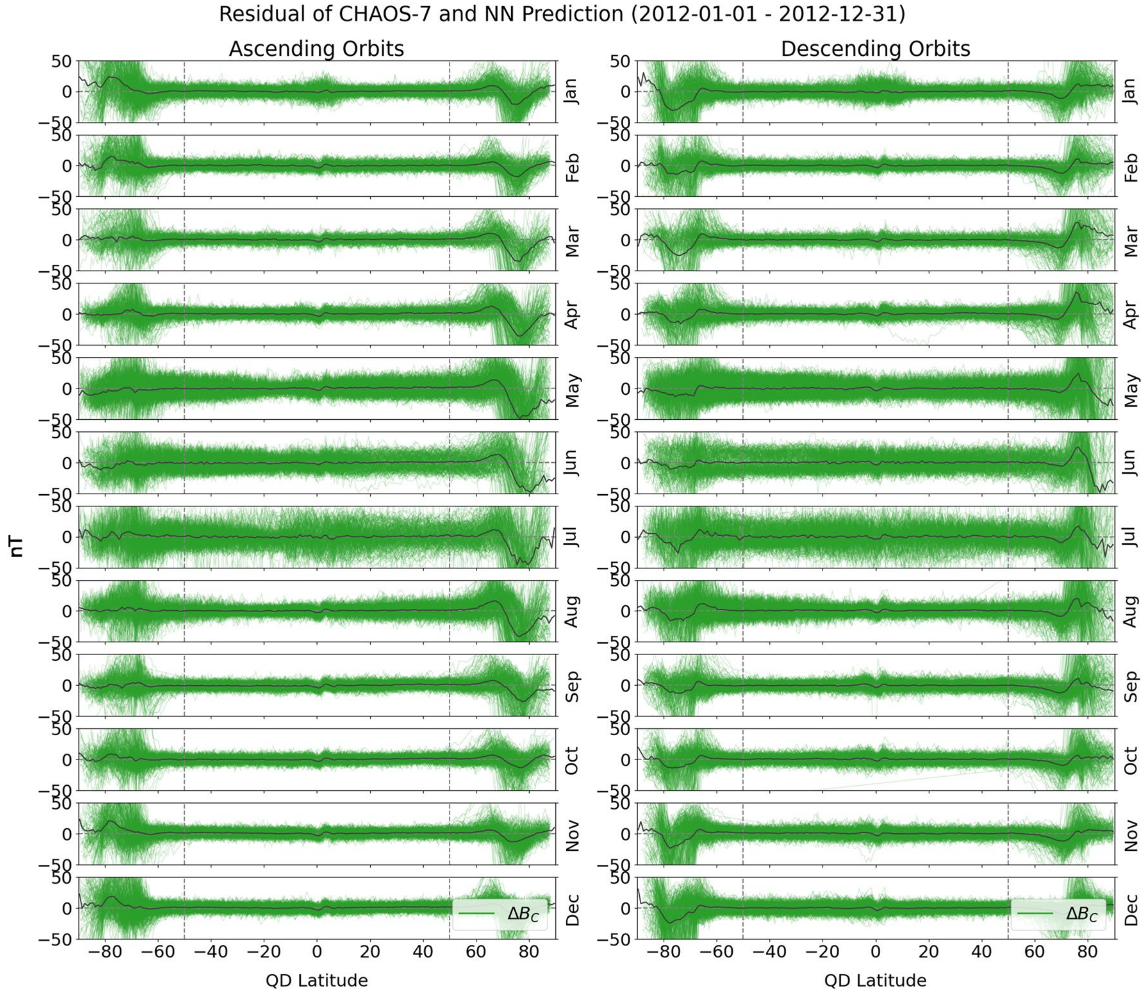


Fig. 6 Residual between the calibrated platform magnetometers and the CHAOS-7 reference model as a function of the quasi-dipole latitude (QDLAT) for every month of the year 2012 of the Center component of the calibrated measurement in nT. The orbits are split into the ascending dusk-orbits on the left, and the descending dawn-orbits on the right, with the mean residual depicted in black

field models, there is a potential improvement due to the additional amount of data points, filling the gap between CHAMP and Swarm.

Comparison to CHAMP

As the CHAMP mission ended in September 2010, there is a period of time during which both satellite missions flew simultaneously. Therefore, conjunctions have been calculated where the two missions have been close to each other to compare the retrieved measurements.

Figure 11 shows the found conjunctions between the 23rd of January 2010 and the 5th of February 2010. During this interval, the two satellite missions have been co-rotating and the magnetic environment was quiet as indicated by the first panel. The distance between the two missions was below 1000km and the relevant QDLATs are shown in the 3rd panel, the dawn and dusk orbit measurements are depicted in red and blue, respectively. The three bottom panels show the residual between the measurement of the GOCE and the CHAMP satellite for the three components of the magnetic field in the NEC frame. Most data points have an absolute difference of less than 20nT, while there is no systematic difference apparent.

Table 3 puts the results of the conjunctions in perspective. For the three components, the total count of measurements during the conjunction is shown, separately for the dawn and dusk orbit. The number of data points that lie above or below the 10nT mark is given in the fourth and fifth column. It can be seen that overall 80% of the

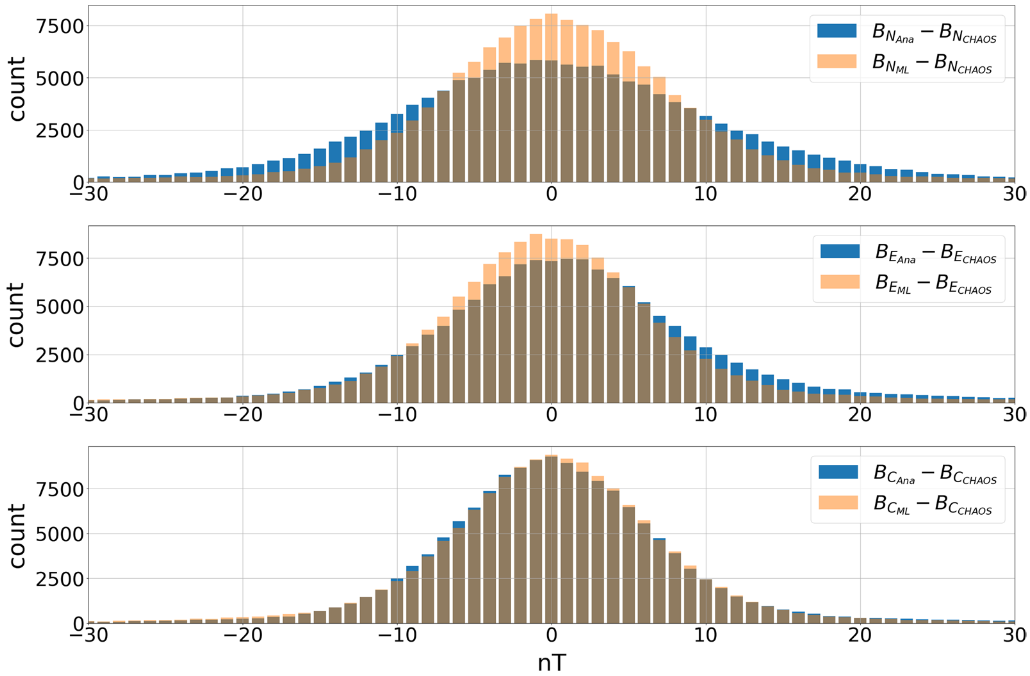


Fig. 7 Residual distribution comparison of the analytical (blue) and ML (orange) based calibrations compared to the CHAOS7 reference model for the December 2009 within a histogram plot with bin sizes of 1 nT for the magnetic North (top), East (middle) and Center (bottom) component

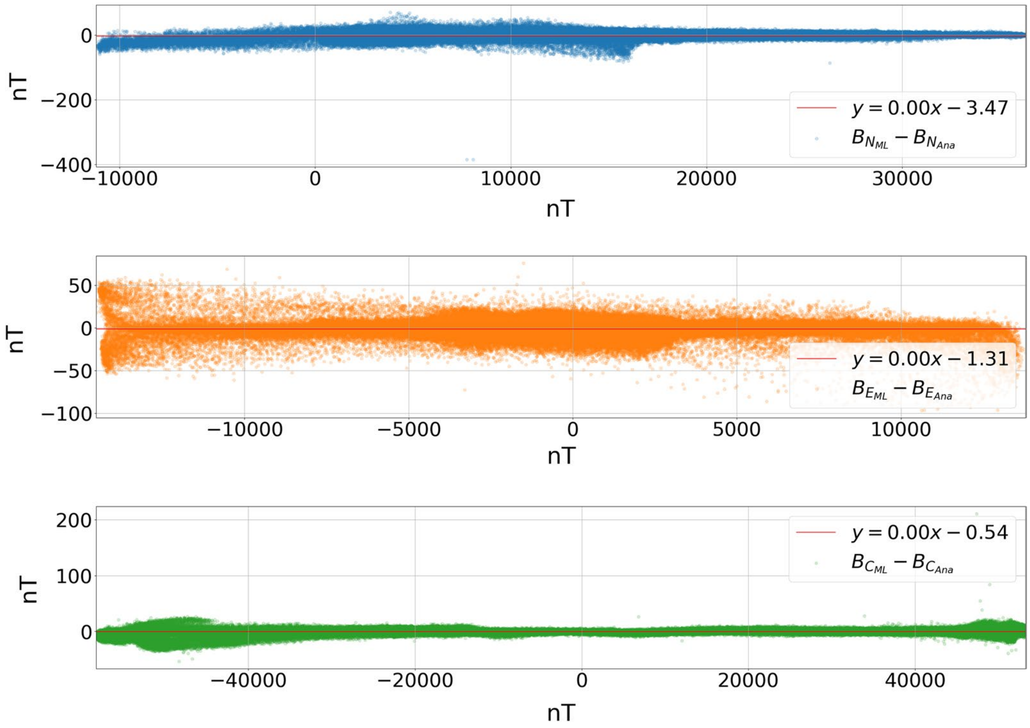


Fig. 8 Difference between ML and Ana calibrations as a function of the calibrated measured magnetic flux density by the analytical approach for the December of 2009 for the North (top), East (middle) and Center (bottom) component. The regression curves are given in red. Please note the different scales of the y-axis

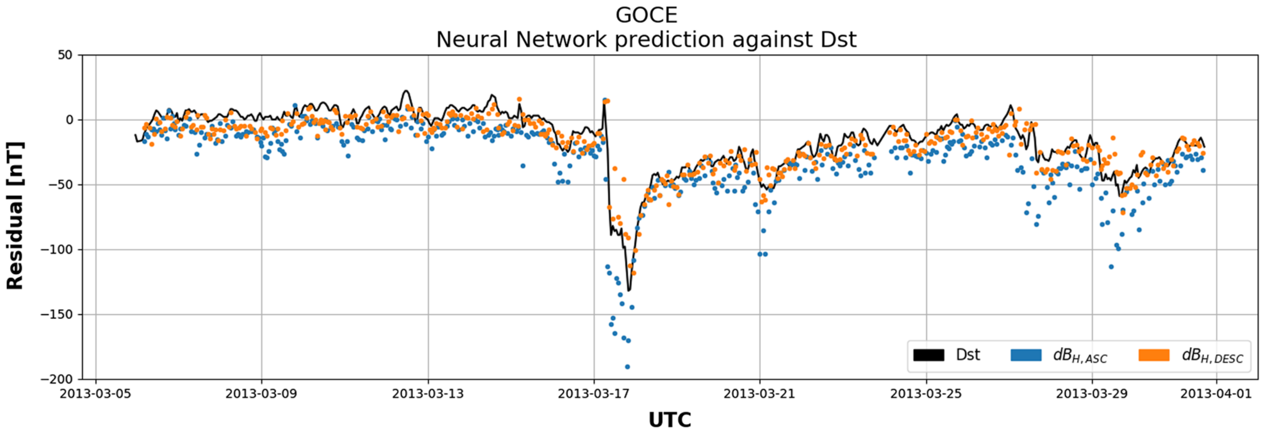


Fig. 9 Magnetic storm in March 2013 indicated by Dst index. The variations of the residual values of the calibrated GOCE platform magnetometer to the CHAOS-7 core and lithospheric field for the ascending, $dB_{H,ASC}$ (blue) and descending, $dB_{H,DESC}$ (orange) orbits follow the index line well, with some known offsets

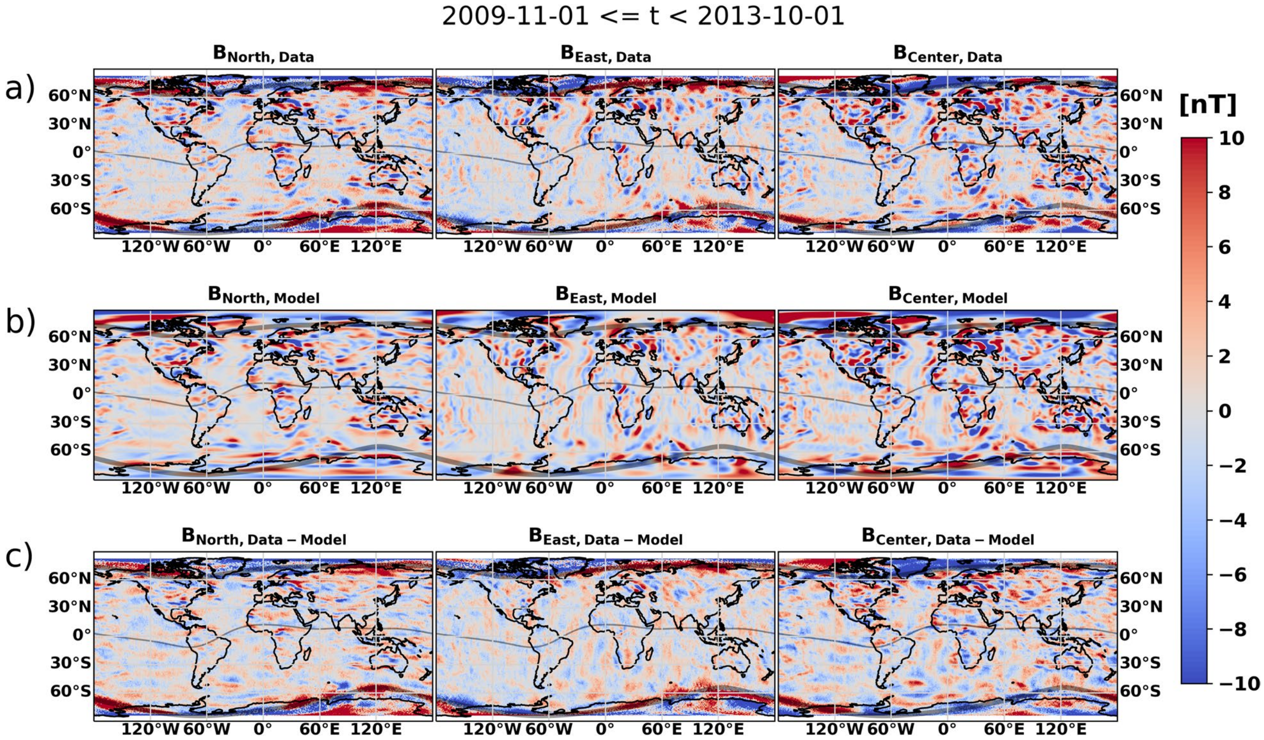


Fig. 10 Lithospheric field residual of GOCE data after subtraction of CHAOS-7 core and large-scale magnetospheric field components (top) and high-resolution lithospheric field model calculated from Thébault et al (2021) (middle). The bottom panel shows the difference between GOCE lithospheric field data and the high-resolution lithospheric model

measurements during the conjunction have a low residual compared with the CHAMP mission, although this varies depending on the component. This is an encouraging result as this result was achieved using a platform magnetometer.

Impact analysis

An interpretation analysis of the NN and the impact of the features has been conducted. Therefore, SHapley Additive exPlanations (SHAP) (Lundberg and Lee 2017) have been used. This method is able to compute the

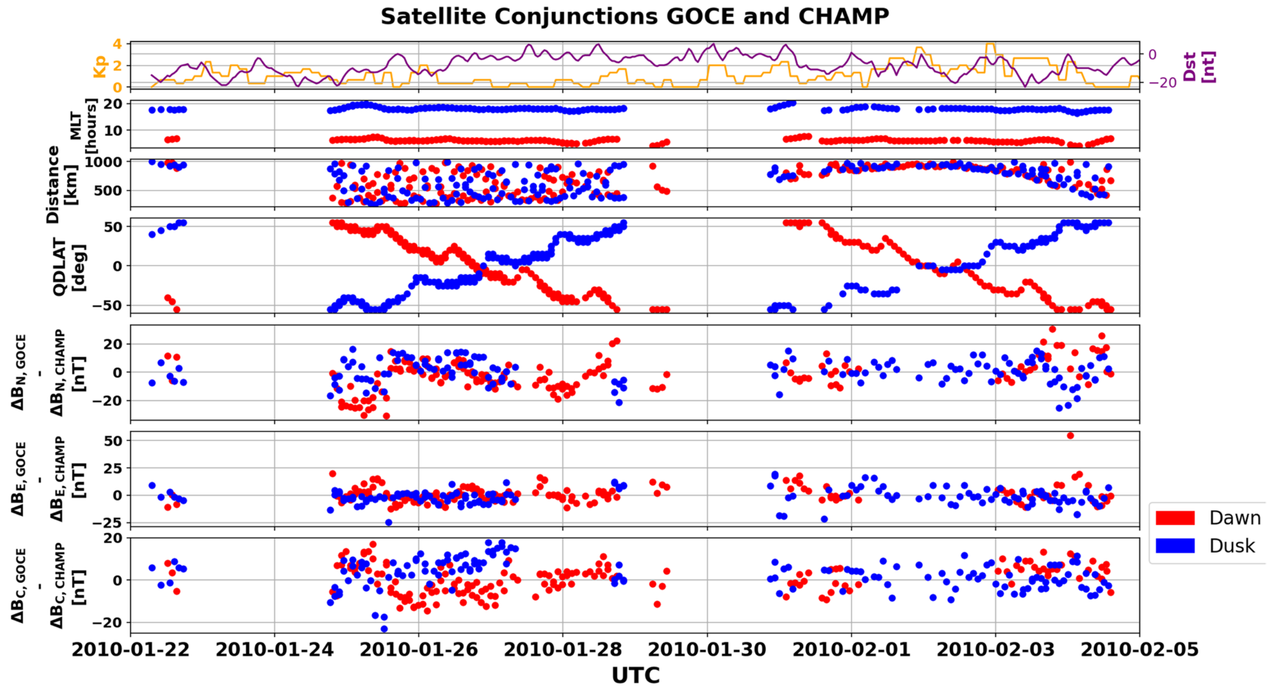


Fig. 11 Conjunctions for co-rotating GOCE and CHAMP satellites for the period of 2010-01-22 to 2010-02-05 are shown. The first panel shows the geomagnetic environment, indicators Kp and Dst show the geomagnetic quiet time. The second panel shows the MLT with descending dawn orbits in red and ascending dusk orbits in blue given by quasi-dipole latitude (QDLAT) in the 4th panel. The 3rd panel shows the distance between the GOCE and CHAMP satellites. The three bottom panels show residuals for GOCE with respect to the CHAOS7 reference model subtracted by residuals for CHAMP with respect to the CHAOS7 model for the North, East and Center components from top to bottom in nT

Table 3 Statistics of conjunction between the GOCE and CHAMP satellites corresponding to Fig. 11. The columns show different statistics for the dawn and dusk orbits for each NEC component for the residual between the satellites for distances within 1000 km, QDLAT <60° and split for below and above a residual of 10 nT

Component	Orbit	Total count	Count >= 10nT	Count < 10nT	Percentage < 10nT
North	Dawn	155	52	103	66.45
North	Dusk	148	41	107	72.30
East	Dawn	155	25	130	83.87
East	Dusk	148	20	128	86.49
Center	Dawn	155	24	131	84.52
Center	Dusk	148	23	125	84.46

contributions of each feature to the prediction of the NN. To achieve this, December 2009 is chosen as the example because it is the most magnetically quiet month. 600 data points are used to compute a mean prediction which is used as the background by the SHAP method. These data points are determined as 600 centroids of K-Means clustering, so a good representation of the monthly data is achieved. This step defines the mean predictions of the NN, which is needed so that the SHAP method can calculate the contributions of the different features to deviations from this mean expected result. In a second step, 600 randomly chosen data points and their prediction have been compared multiple times with the average result. For each data point, this step is repeated many times with masking some features absent, thus an estimate about the contribution of each feature can be given. As the calibration was done in satellite frame, this analysis has also been conducted in satellite frame.

Figure 12 shows the top 40 features used by the ML algorithm, sorted depending on their impact or contribution on the final prediction for the calibrated value. The list of features was split in two parts, depending on the scale of their contribution, please note the different scales. The most important feature can be found on the top left, while the 40th most important feature can be found in the bottom right. The order is based on the average impact on the model output, which consists of the added contributions to the X, Y, and Z components, depicted in different colors. The ESA provides an online sheet containing the feature abbreviations and their descriptions (ESA 2019). On the left, the magnetometer measurements can be found, e.g., the first three features are the Z-component of the three different magnetometers, which are accompanied by the same features found in the Telemetry data, AMT00104 to AMT00304. This confirms the assumption that the magnetometers have the largest influence on the magnetometer calibration. On the right, the most important housekeeping or telemetry features can be seen which are: temperatures (THT), Xenon Tank Heaters (XST), one Euler angle, some currents (PHD), and the magnetorquers (ATT). These are interesting findings as these generally correspond to the expectations for the calibration of platform magnetometers as used in the analytical method, but were found automatically by the ML approach. There are some differences in which features exactly are relevant when comparing both approaches, e.g., the Xenon Tank Heaters or different temperatures which were not used in the analytical approach. These have not been searched for manually in the previous analytical approach and show the ability of the ML approach to discover relevant features.

Integration and improvement in Kalmag model

To evaluate the dataset in the context of geomagnetic field modeling, it was assimilated by the Kalman filter algorithm used to build the Kalmag model (Baerenzung et al. 2020). This model is composed of 7 sources, a core field, a lithospheric field, an induced/residual ionospheric field, a remote, a close, and a fluctuating magnetospheric field, and a source associated with FACs. Each source is expanded in spherical harmonics and their dynamical evolution is controlled by scale-dependent autoregressive processes. For more detail about the Kalman filter algorithm and the spatiotemporal characterization of the different sources see Baerenzung et al. (2020).

The Kalmag model spanning the last 22 years was constructed through the assimilation of CHAMP, Swarm, and secular variation data derived from ground-based observatory measurements. It therefore is subject to the lack of observations taken by low orbiting satellites between 2010.7 and 2013.9. This new GOCE dataset could therefore have a great potential to fill this data gap. In order to evaluate this potential, three models were built, two including GOCE data and one without. The first model, referred as model C, is the one partially serving the construction of Kalmag. It spans the [2000.5; 2014.0] time period and is derived from CHAMP and observatory secular variation data. Its solution in 2009.8 was considered as a prior information for the second model (model G), which was built through the assimilation of GOCE and secular variation observations. The last model, namely model GL, is similar to model G except that the mean lithospheric field was a priori set to zero and its prior covariance initialized. Its purpose is to evaluate how the lithospheric field can be recovered from GOCE data alone. All comparisons between the different models are performed for 2014.0.

The results of this evaluation for the internal field, i.e., the sum of the core and the lithospheric field, and for the secular variation, are presented in terms of energy spectra at the Earth's surface in 2014.0 in Fig. 13. Spectra of the mean (solid lines), the standard deviation (dashed lines) and the difference with the Kalmag mean model (circles) are displayed for model G with thick lines and for model C with thin lines. For the main field (left panel), the spectra associated with mean solutions cannot be distinguished between the two models. This is not the case for the spectra of the difference with Kalmag. The mean solution of model G is globally closer to the Kalmag solution than the mean solution of model C. Since Kalmag is more accurate than both model G and C, due to the fact that it also derives from Swarm data taken before and after 2014.0, this result demonstrates that the assimilation of GOCE data helps to better resolve the main field. However, predicted uncertainties of model G, given by the spectra of the standard deviation, are slightly underestimated when compared to effective errors, as approximated by the spectra of the difference with Kalmag. This is an indication that some source contributing to the observations is not perfectly modeled. Ionospheric currents, which are still generating some magnetic signal at dawn and dusk (the orbit of the satellite), might be this source since the Kalman filter algorithm was not calibrated to account for them.

The impact of GOCE data to recover the secular variation is clearly positive as it can be observed on the right panel of Fig. 13. Not only the spatial resolution of model G is higher, but its level of error is globally lower than for model C. Furthermore, predicted and effective errors are consistent with one another. Contrary to the main field, no signal is apparently contaminating the secular variation. Note that performing the same analysis at previous epochs leads to similar results.

(See figure on next page.)
Fig. 12 The figure shows the top 40 features used by the ML algorithm, sorted by their average impact on the final prediction for the calibrated value, as was calculated by the SHAP method. This was evaluated for the X, Y, and Z components of the measurements in the satellite frame where each feature's contribution consists of the contribution to each component, distinguished by color. Note the different scale of the features on the left and right

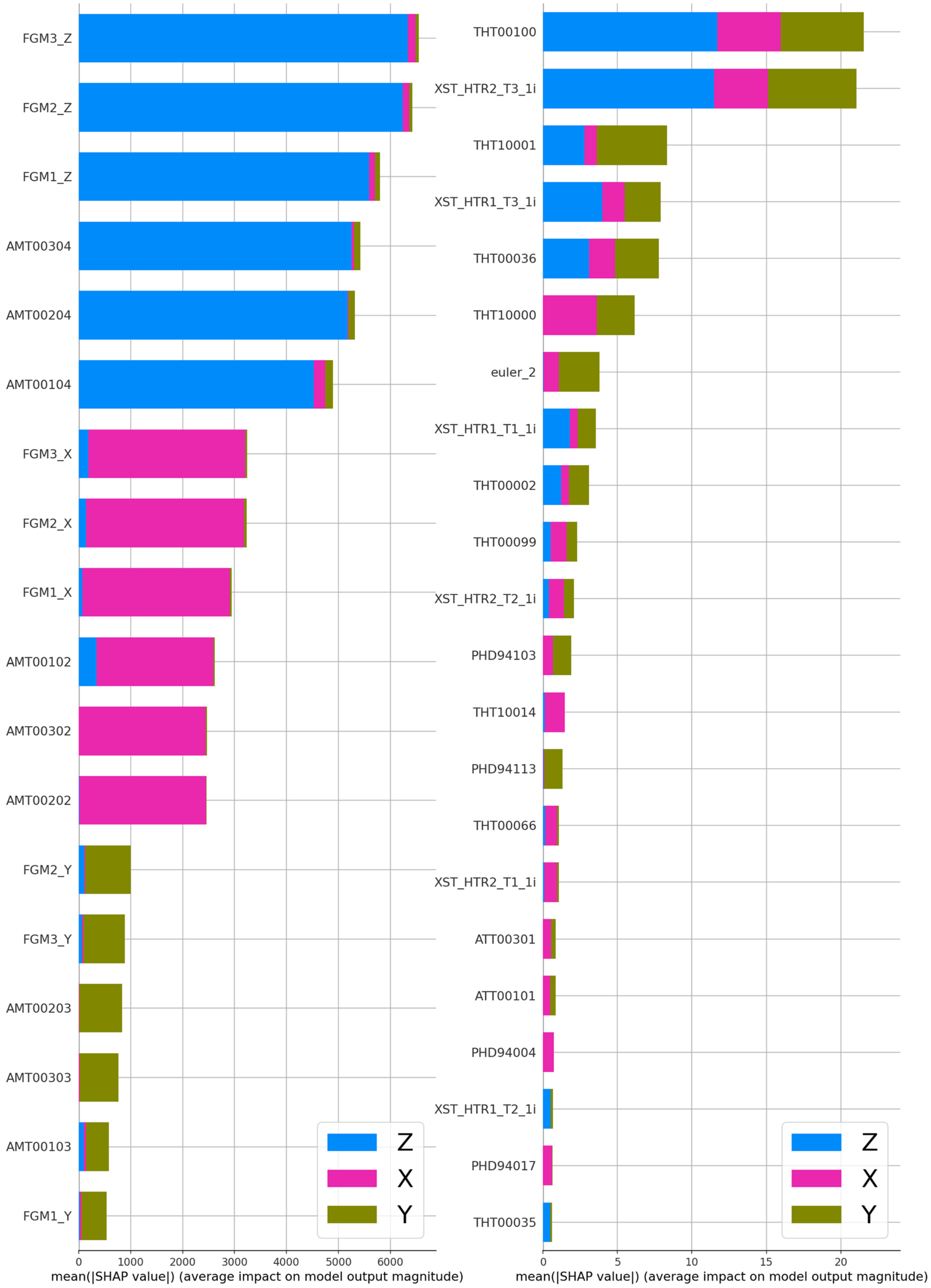


Fig. 12 (See legend on previous page.)

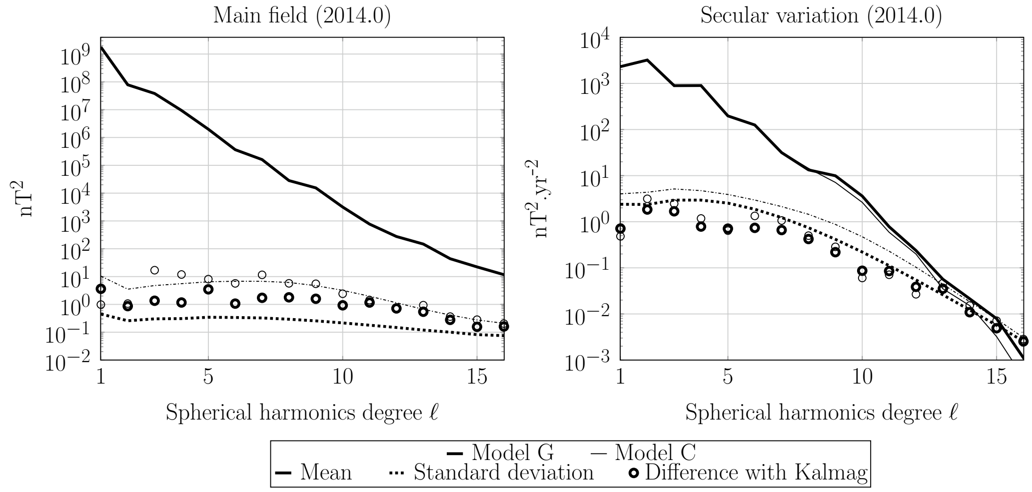


Fig. 13 Main field (left) and secular variation (right) energy spectra at the Earth's surface in 2014.0. Spectra of the mean (solid lines), standard deviation (dashed lines) and difference with the Kalmag model (circles) for model G (thick lines and symbols) and model C (thin lines and symbols)

The signal associated with the lithospheric field can also be well extracted from the dataset as illustrated in Fig. 14. Comparisons between the downward component of the field at the Earth's surface for model GL (top) and for model C (bottom) highlight their proximity. Most structures which can be recovered with CHAMP data are present within the GL solution. However, as for the core field, predicted uncertainties are also slightly underestimated (not shown).

Through this evaluation one can conclude that such a dataset is useful for the geomagnetic modeling community to cope with the lack of observations between the CHAMP and Swarm eras. Nevertheless, efforts in modeling the dayside ionospheric field are likely to be required in order to take full advantage of these new measurements.

Conclusion

To sum up, we could show that the ML-based calibration of the platform magnetometers onboard the GOCE mission yields promising results. With careful data collection and selection, as well as sophisticated data preprocessing it is possible to significantly reduce the remaining residual of the platform magnetometer measurements compared to the reference model CHAOS-7. Our evaluation has shown that on average a residual of 6.47nT for low- and mid-latitudes could be achieved which leads to a dataset that can help in studying the Earth's magnetic field. Some potential applications were shown like measurements of the lithospheric field, as well as additional information during geomagnetic storms. During a conjunction with the CHAMP satellite, it could be shown that the achieved calibration is in good agreement with other magnetic measurements. Finally, the enhancement of an existing magnetic field model could be shown for the time span of the gap of high-precision magnetic missions between 2010 and 2013. With the ML approach we are able to provide a calibration which is easily applicable to other satellite missions with some knowledge about the underlying Data Science techniques, whereas the previous analytical approach needs deep Domain knowledge to be applied. In the future, we hope that the provided dataset can support geoscientists by offering additional data to better cover the magnetic field in time, altitude, position, and MLT. The data of this study are published on the ISDC-Server of the GFZ at (Styp-Rekowski et al. 2022) ftp://isdcftp.gfz-potsdam.de/platmag/MAGNETIC_FIELD/GOCE/ML/v0204/ under the version 0204. In the future, the used features of the calibration as well as the correlation between features and the southpolar electrojet will be analyzed to investigate how to improve the calibration further.

Appendix A: Neural network introduction

A feed-forward neural network (NN) consists of multiple neurons where each neuron's output y is defined as follows:

$$y = a(w_1 * x_1 + w_2 * x_2 + w_3 * x_3 + \cdots + w_n * x_n + b)$$

where x_1 to x_n are different features of a data point which get multiplied by their respective weight w_1 to w_n, with a bias b being added finally. The sum of these products is then fed into an activation function a which is explained in the next paragraph. NNs consist of many such neurons which will be assigned different weights. Multiple neurons in parallel, which means that they work on the same input x_1 to x_n, are called a layer of neurons. These layers

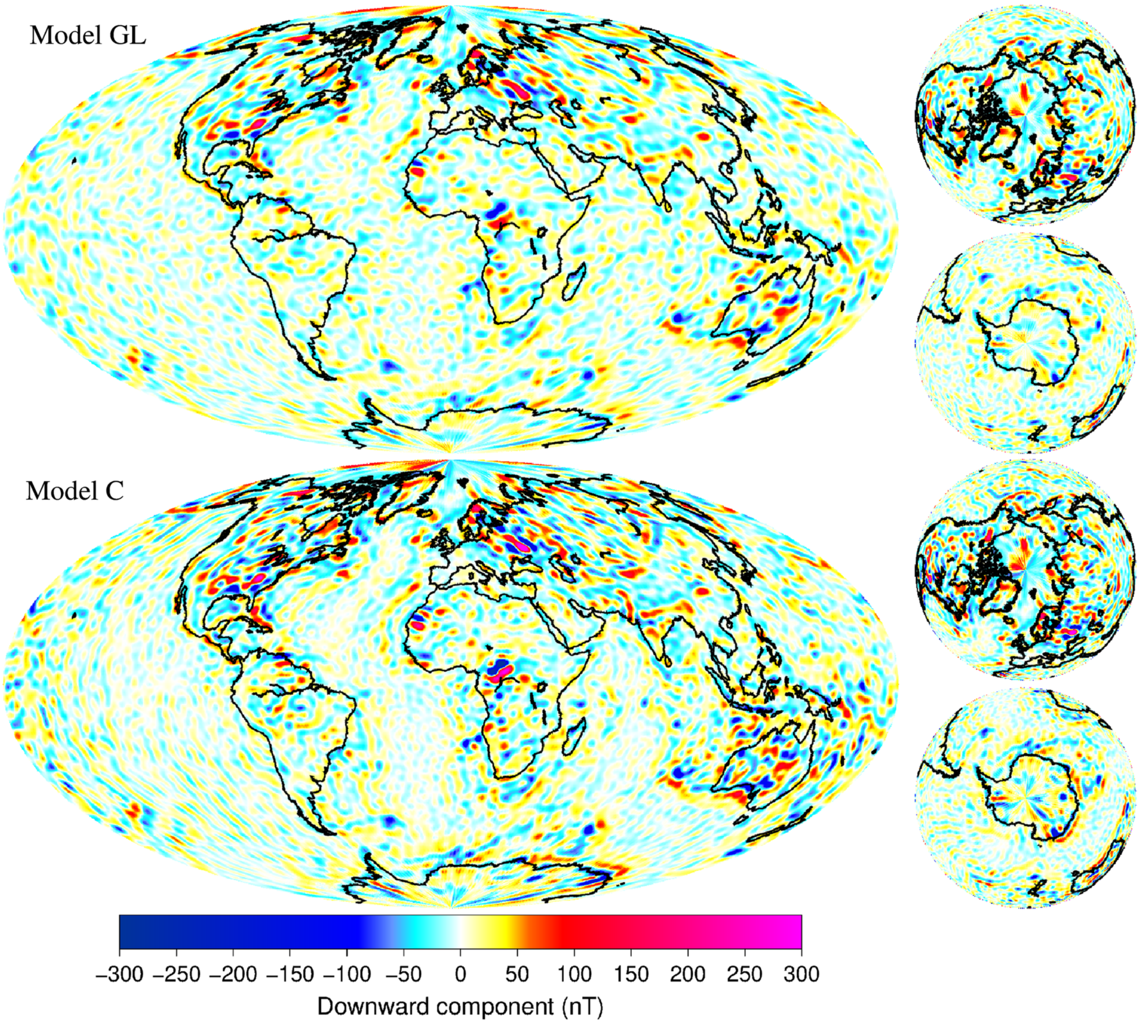


Fig. 14 Downward component of the lithospheric field at the Earth's surface expanded up to spherical harmonics degree $\ell = 100$. Top: model GL derived from GOCE data alone. Bottom: model C derived from CHAMP data

of neurons can also be stacked, meaning another layer takes the outputs y_1 to y_m of the first layer as the inputs for the neurons in the next layer. Each neuron of a layer can also be seen as a feature, as it is a new combination of the inputs that acts like a machine-crafted feature.

The initial formula is very similar to a linear regression formula and by stacking these neurons there would be no non-linearity added to the system, which would limit its expressiveness in the kind of processes it could explain. Therefore, the sum of the products is fed into the non-linear activation function. There are many activation functions available, that act in different ways, one of these is the Rectified Linear Unit (RELU) (Nair and Hinton 2010), defined as follows:

$$f(x) = max(0, x)$$

Negative values become 0, while positive neuron calculations will get forwarded into the next layer. This enhances the expressiveness of the NN and gives the ability to model non-linearity. The Exponential Linear Unit (ELU) (Clevert et al. 2015) used in this work has a similar function shape like the RELU but overcomes some shortcomings. Normally, all intermediate layers, also called hidden layers, before the final output layer use activation functions in their output. For the final output layer, this is oftentimes not done as a linear combination of the previous layer is desired.

The construct of number of neurons and number of layers, as well as the last layer of neurons which constitutes the output is then called the NN or the model. In the beginning, all the weights within this system are initialized randomly. For the model to represent the desired function, e.g., mapping the satellite data input to the calibrated measurement, the weights need to be adjusted in a meaningful way, this is called the training of the NN. Contrary to algorithms like the least squares approach, NNs are trained with the backpropagation algorithm (Rumelhart et al. 1986). Therefore, each available data point in the training dataset is forwarded through the NN and the current result of this calculation is compared to the expected result for this data point which is also called the ground truth or the reference model. This error is then used as the gradient to adjust the weights and backpropagated through the NN to adjust the weights in the different layers in such a way that the prediction would come closer to the expected result if the calculation was repeated. To speed this process up and not rely on single data points, this error is calculated on groups of data points, also called a batch, for which one common gradient is calculated and then an adjustment to the network is made. The amount of adjustment towards the calculated gradient can be controlled with the learning rate, which normally lies in the interval of [0,1] and is multiplied with the calculated gradients before they are applied to the weights. One iteration through all batches of available data points and adjustments to the NN is called an epoch. The NN will be trained for multiple epochs and there are optimizers like the Adam optimizer (Kingma and Ba 2014) speeding up the process by modifying the gradients with historic information. After this training process, the weights of the NN do not change anymore and it is assumed that the NN represents the statistics of the data. Afterward, it can be used to perform predictions on similar data as it was trained with.

Abbreviations

CHAMP: CHAllenging Minisatellite Payload; ELU: Exponential linear unit; FAC: Field-aligned currents; GOCE: Gravity and steady-state Ocean Circulation Explorer; GRACE: Gravity Recovery And Climate Experiment; GRACE-FO: Gravity Recovery And Climate Experiment Follow-On; HPO: Hyperparameter optimization; MAE: Mean absolute error; ML: Machine learning; MLT: Magnetic Local Time; MSE: Mean squared error; NEC: North–East–Center frame; NN: Neural network; QDLAT: Quasi-dipole latitude; STD: Standard deviation.

Acknowledgements

We thank Jan Rauberg for his help in interpreting and evaluating the results. We also thank the Geomagnetic World Data Centre Kyoto for the Dst index, and the Astrophysical Observatory and Natural Resources Canada for the F10.7 index.

Author Contributions

KS preprocessed and calibrated the data, and wrote the manuscript. KS, IM, and CS designed the study. JB integrated the data. KS, IM, CS, MK, and OK evaluated and reviewed the results of the study. All authors read and approved the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. 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.

Availability of data and materials

The datasets generated and analyzed during the current study are available under (Styp-Rekowski et al. 2022) ftp://isdcftp.gfz-potsdam.de/platmag/MAGNETIC_FIELD/GOCE/ML/v0204/.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Distributed and Operating Systems, Technical University of Berlin, Ernst-Reuter-Platz 7, 10587 Berlin, Germany. [2]GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Telegrafenberg, 14473 Potsdam, Germany. [3]Leibniz Institute of Atmospheric Physics at the University of Rostock, Schloßstraße 6, 18225 Kühlungsborn, Germany.

Received: 6 May 2022 Accepted: 24 August 2022

References

Alken P, Olsen N, Finlay CC (2020) Co-estimation of geomagnetic field and in-orbit fluxgate magnetometer calibration parameters. Earth, Planets and Space 72:1–32. https://doi.org/10.1186/s40623-020-01163-9

Anderson B, Ohtani S-I, Korth H, Ukhorskiy A (2005) Storm time dawn-dusk asymmetry of the large-scale Birkeland currents. J Geo Res, 110(A12), A12220. https://doi.org/10.1029/2005JA011246

Baerenzung J, Holschneider M, Wicht J, Lesur V, Sanchez S (2020) The Kalmag model as a candidate for IGRF-13. Earth Planets Space 72:1–13

Billingsley Billingsley TFM100S Magnetometer. Billingsley Aerospace Defense. https://magnetometer.com/wp-content/uploads/TFM100S-Spec-Sheet-February-2008.pdf

Clevert D-A, Unterthiner T, Hochreiter S (2015) Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289

Drinkwater M, Floberghagen R, Haagmans R, Muzi D, Popescu A (2003) VII: closing session: GOCE: ESA's first earth explorer core mission. Space Sci Rev 108:419–432

European Space Agency (2019) GOCE telemetry data collection. Version 1.0. GOCE telemetry packets description. https://doi.org/10.5270/esa-7nc8pjp

Finlay CC, Kloss C, Olsen N, Hammer MD, Tøffner-Clausen L, Grayver A, Kuvshinov A (2020) The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. Earth Planets Space 72:156. https://doi.org/10.1186/s40623-020-01252-9

Floberghagen R, Fehringer M, Lamarre D, Muzi D, Frommknecht B, Steiger C, Piñeiro J, Da Costa A (2011) Mission design, operation and exploitation of the gravity field and steady-state ocean circulation explorer mission. J Geo 85:749–758. https://doi.org/10.1007/s00190-011-0498-3

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neu Net 2:359–366

Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980

Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. Adv in Neu Inf Pro Sys 30:4765-4774

Matzka J, Bronkalla O, Tornow K, Elger K, Stolle C (2021) Geomagnetic Kp index. GFZ GRCG. https://doi.org/10.5880/Kp.0001

Michaelis I, Styp-Rekowski K, Rauberg J, Stolle C, Korte M (2022) Preprint) Geomagnetic data from the GOCE satellite mission. ESpace Science Open Archive. https://doi.org/10.1002/essoar.10511006.1

Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on international conference on machine learning. Omnipress, Haifa, Israel, pp 807-814

Nose M, Sugiura M, Kamei T, Iyemori T, Koyama Y (2015) Dst Index. WDC for Geomagnetism, Kyoto. https://doi.org/10.17593/14515-74000

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

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 Space 72:1–18

Olsen N, Friis-Christensen E, Floberghagen R, Alken P, Beggan CD, Chulliat A, Doornbos E, Da Encarnação JT, Hamilton B, Hulot G, Van Den Ijssel J, Kuvshinov A, Lesur V, Lühr H, Macmillan S, Maus S, Noja M, Olsen PEH, Park J, Plank G, Püthe C, Rauberg J, Ritter P, Rother M, Sabaka TJ, Schachtschneider R, Sirol O, Stolle C, Thébault E, Thomson AWP, Tøffner-Clausen L, Velímský J, Vigneron P, Visser PN (2013) The Swarm satellite constellation application and research facility (SCARF) and Swarm data products. Earth Planets Space 65:1189–1200. https://doi.org/10.5047/eps.2013.07.001

Olsen N, Stolle C (2012) Satellite geomagnetism. Ann Rev of Ear and Pla Sci 40:441–465

Reigber C, Lühr H, Schwintzer P (2002) CHAMP mission status. Adv in Spa Res 30:129–134

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536

Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. Adv in Neu Inf Proc Sys 25:2951-2959

Stolle C, Michaelis I, Xiong C, Rother M, Usbeck T, Yamazaki Y, Rauberg J, Styp-Rekowski K (2021) Observing Earth's magnetic environment with the GRACE-FO mission. Earth Planets Space 73:84–104. https://doi.org/10.1186/s40623-021-01364-w

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 Space 73:51–54. https://doi.org/10.1186/s40623-021-01409-0

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). IEEE, pp 5283-5290. https://doi.org/10.1109/BigData52589.2021.9671977

Styp-Rekowski K, Michaelis I, Stolle C, Baerenzung J, Korte M, Kao O (2022) GOCE ML-calibrated magnetic field data. V. 0204. GFZ Data Services. https://doi.org/10.5880/GFZ.2.3.2022.002

Tapping K (2013) The 10.7 cm solar radio flux (F10. 7). Space Weather 11:394–406

Thébault E, Hulot G, Langlais B, Vigneron P (2021) A spherical harmonic model of Earth's lithospheric magnetic field up to degree 1050. Geo Res Let 48:e2021GL095147. https://doi.org/10.1029/2021GL095147

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 publicly 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 innovative 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,

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

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

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

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

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

is the product of the electric current I, the area orthogonal vector $\vec{a}$ 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).

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

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 $\vec{r}$, 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 $\vec{r}$ and the area vector $\vec{a}$ 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.

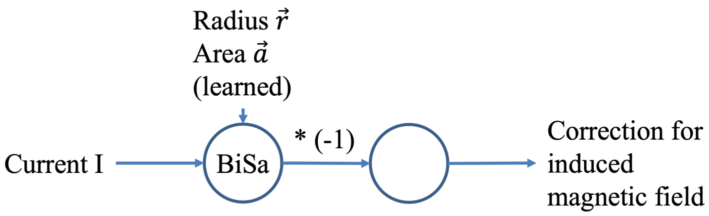


Figure 1. BiSa (Biot-Savart) layer with the electric current I as the input, learnable parameters $\vec{r}$ and $\vec{a}$ and the correction for the induced magnetic field.

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.

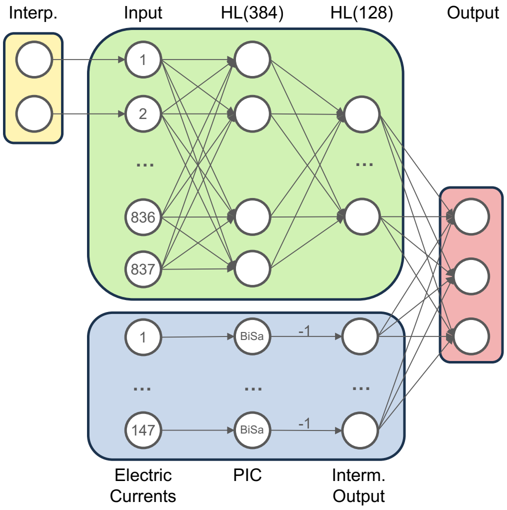


Figure 2. Architecture of the PINN calibration model for the GOCE satellite: In yellow is the time shift interpolation, in green is the FFNN, in blue is the PIC, and in red is the 3-dimensional output. The final architecture was found through a hyperparameter optimization.

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

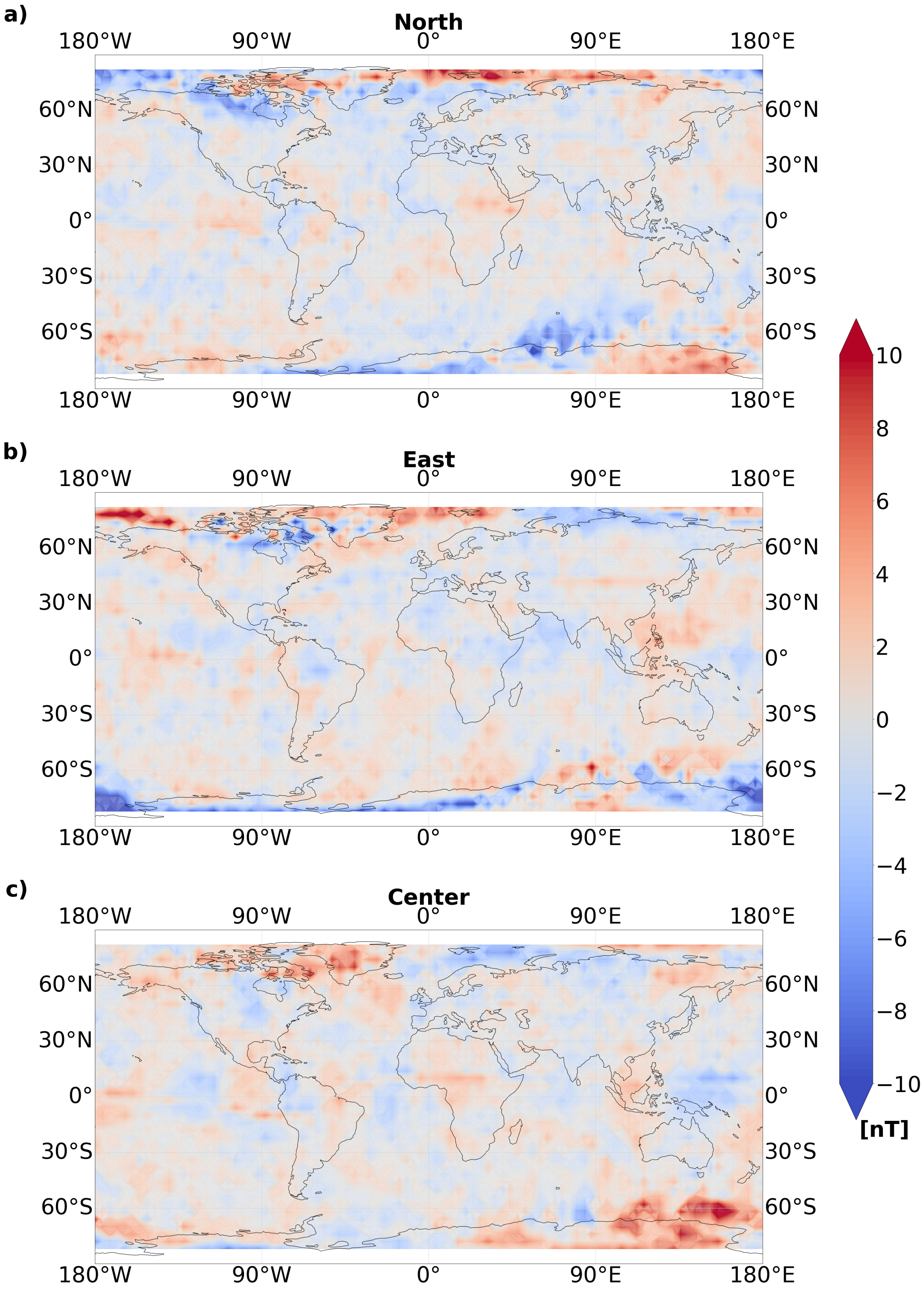


Figure 3. Map of the binned and averaged residuals between the PINN calibration and the combined reference model in the NEC-frame on a scale of 10 nT with GOCE data for the year 2013, panel a) shows the North component, panel b) the East component, and panel c) the Center component. A grey color indicates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

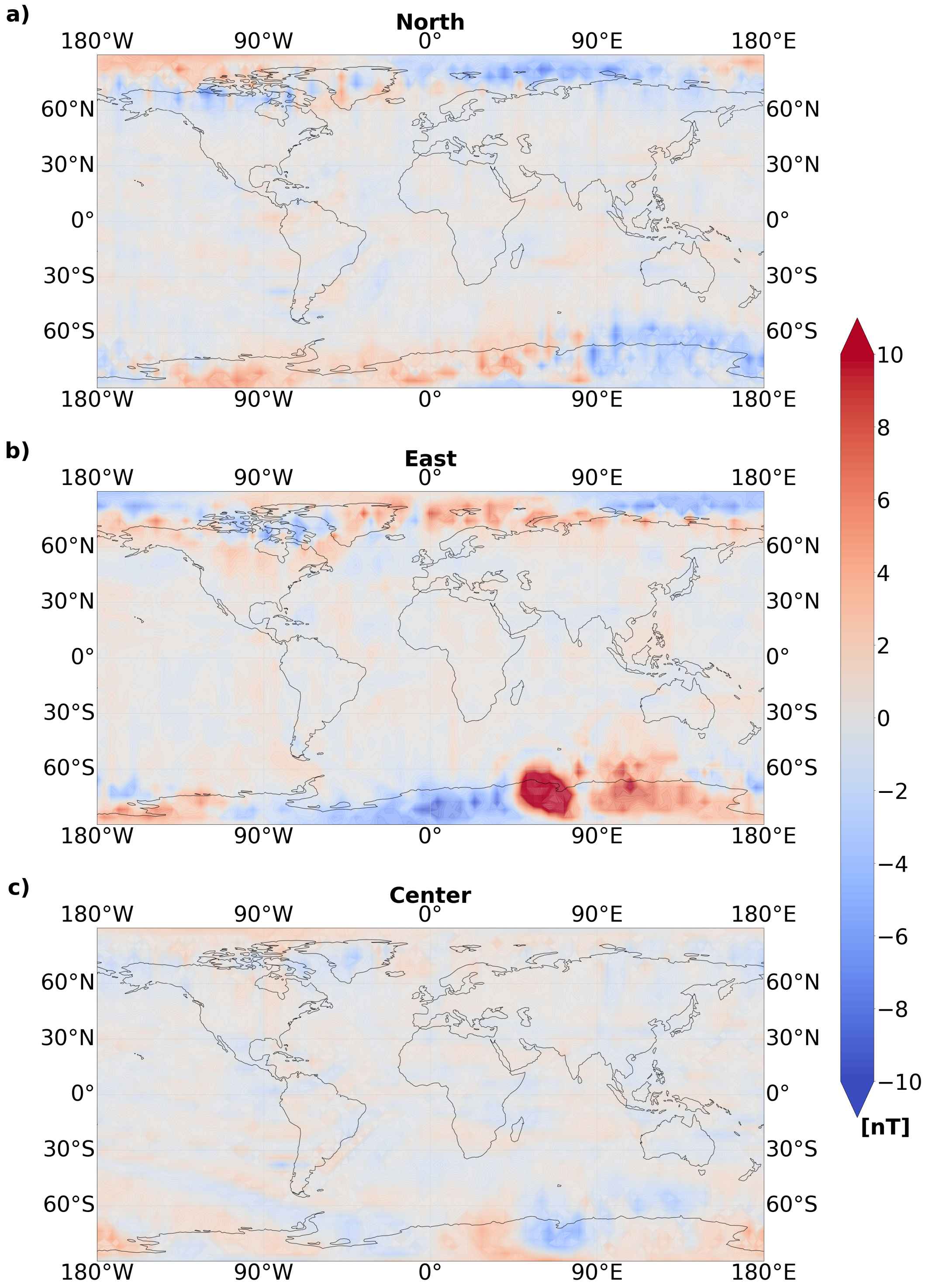


Figure 4. Map of the binned and averaged residuals between the PINN calibration and the combined reference model in the NEC-frame on a scale of 10 nT with GRACE-FO1 data for the year 2019, panel a) shows the North component, panel b) the East component, and panel c) the Center component. A grey color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

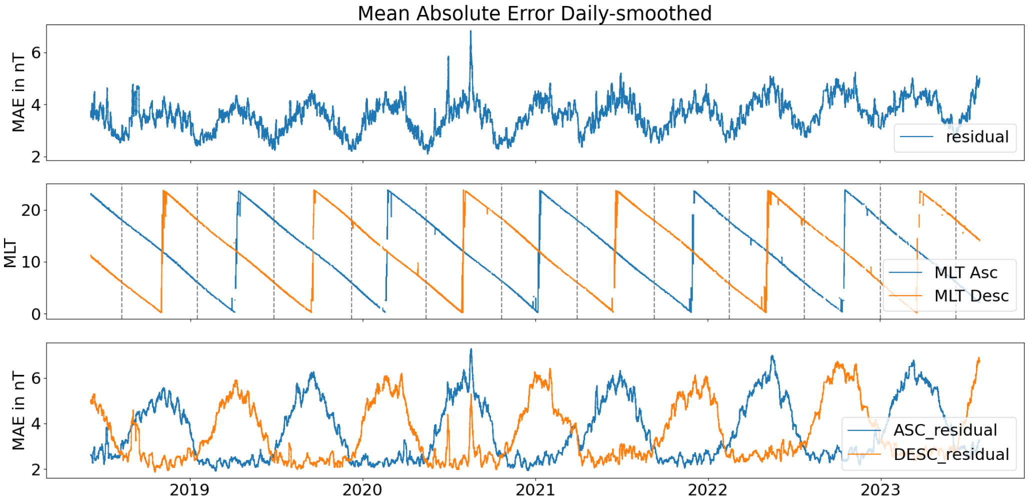


Figure 5. First row: Residual over whole calibration period for GRACE-FO1 with 86400s (daily) smoothing applied. Second row: MLT changes for the ascending and descending orbit of the mission. Dawn-dusk orbits marked with grey dotted lines for times of low residual. Third row: Same as the first row, with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

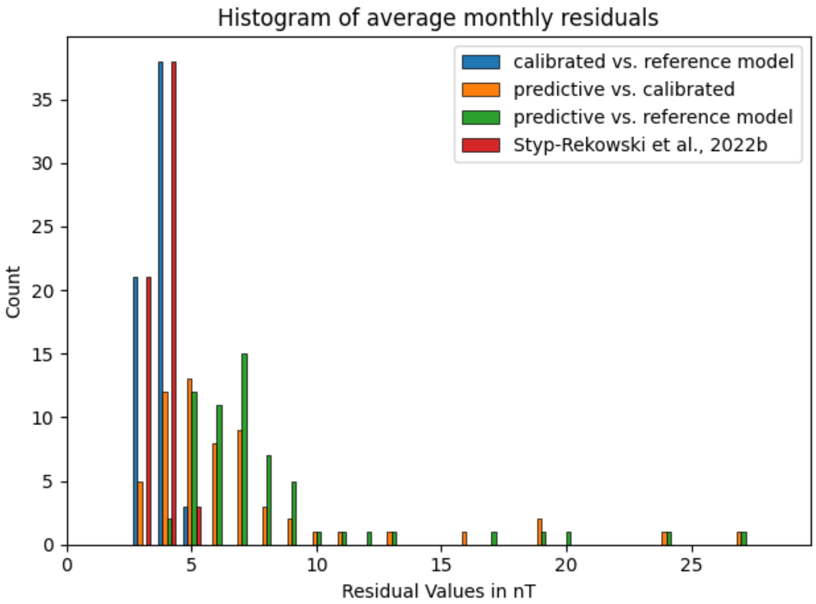


Figure 6. Histogram plot of monthly average residuals for predictive calibration models evaluated on subsequent monthly data for the whole mission data of the GRACE-FO1 satellite. Additionally, the direct calibration is included and compared against the reference model as well as the predictive calibration. 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 $\vec{r}$ and their dipole moment $\vec{m}$ 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 $\vec{r}$ and $\vec{a}$ 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

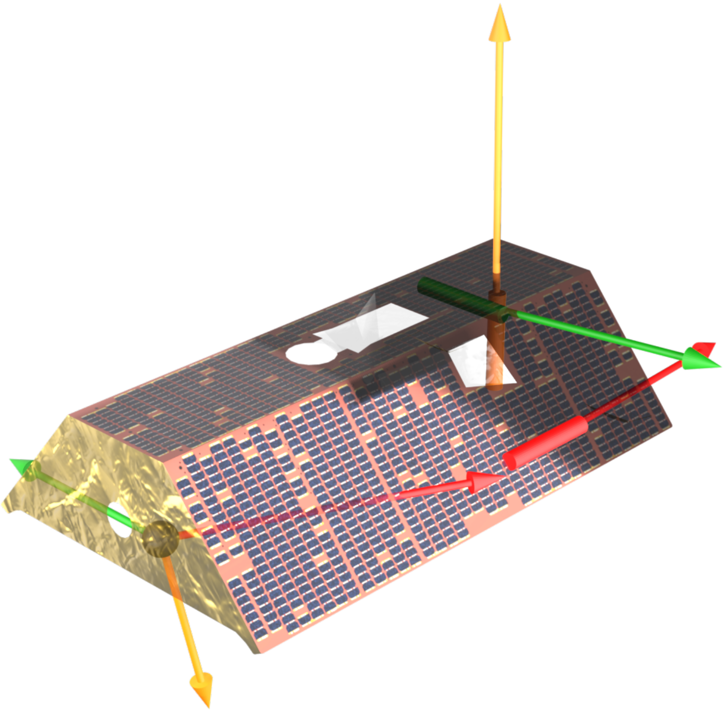


Figure 7. Model of the GRACE-FO satellite in decimeters(dm) with the magnetometer position in front(left) of the satellite given as a black sphere. The induced magnetic field in nT is shown at the magnetometer position in the form of vectors. At the back of the satellite, the magnetic dipole moment at 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 ±1°.

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.

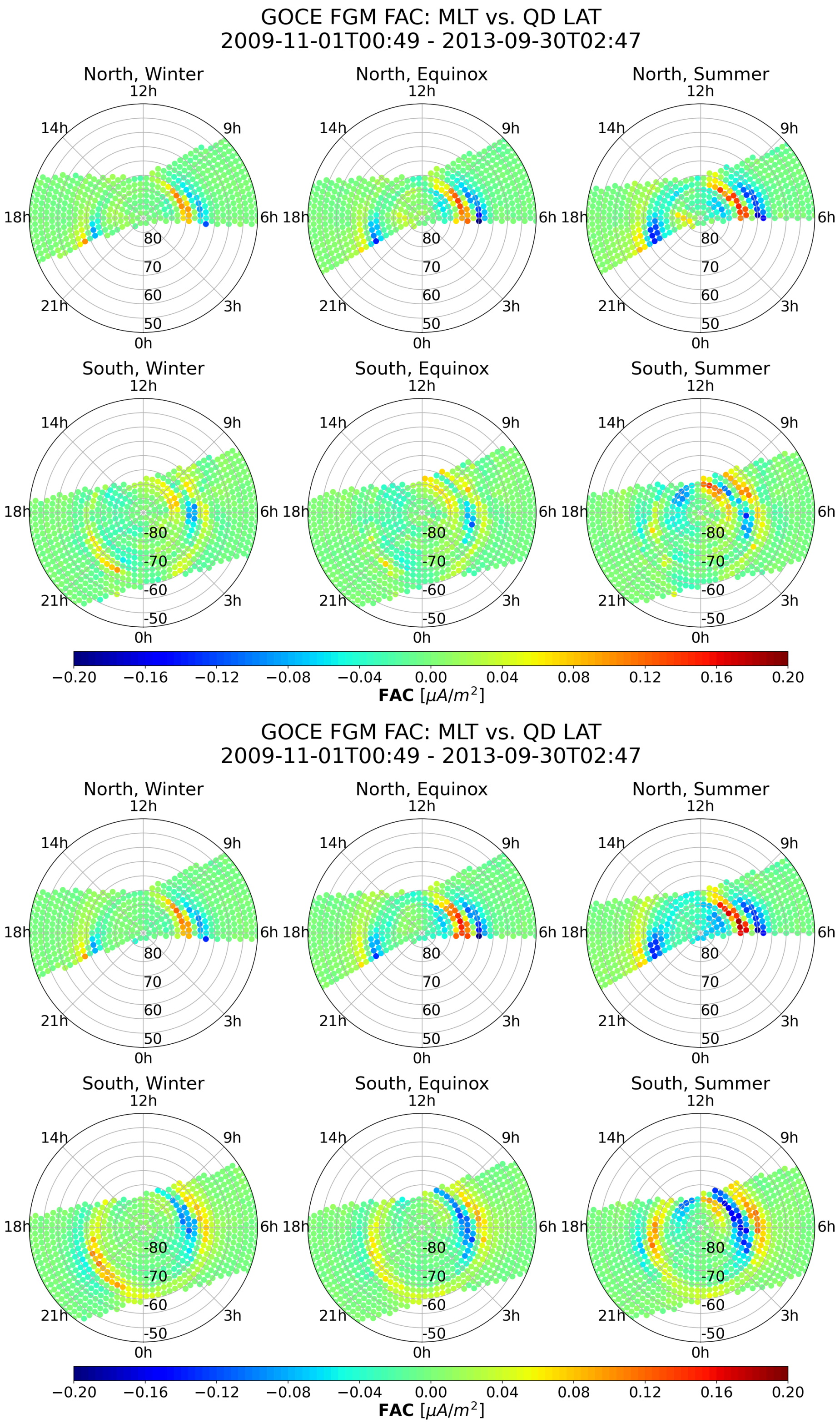


Figure 8. Field-aligned currents as derived from calibrated data of the GOCE satellite mission. Summarized median by MLT and QDLat for the Northern and Southern Hemispheres. At the top is the plot as derived from the previous approach (Styp-Rekowski et al., 2022b) and at the bottom is the newly derived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

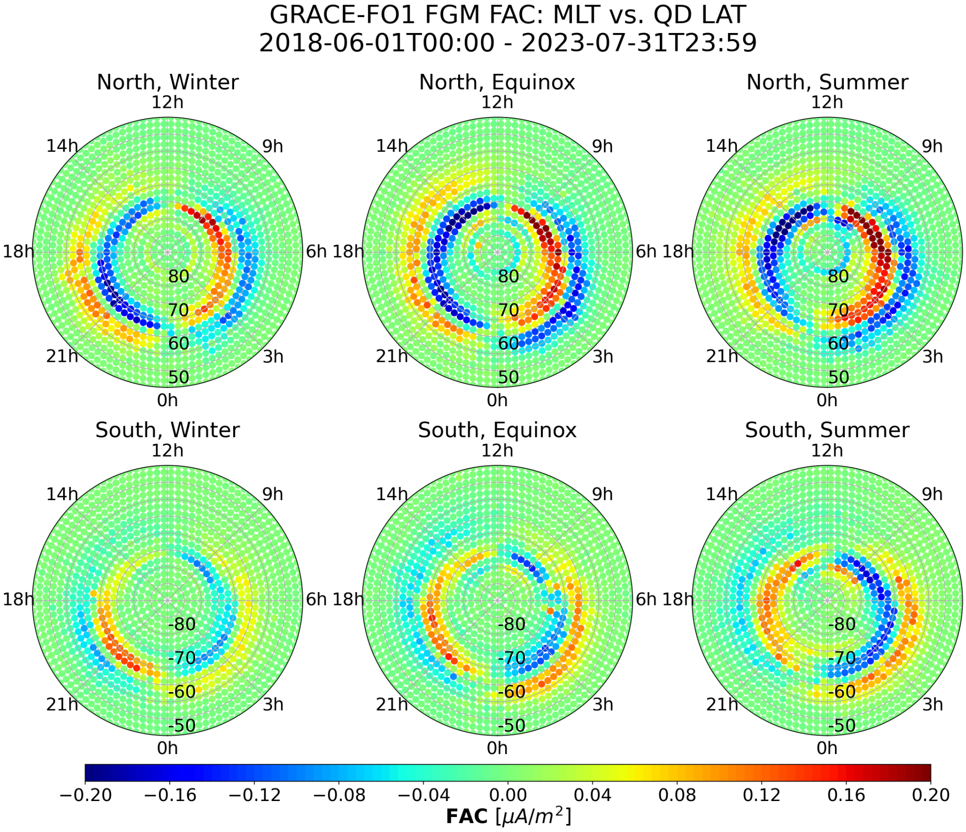


Figure 9. Field-aligned currents as derived from calibrated data of the GRACE-FO1 satellite. Summarized mean by MLT and seasons for the Northern and Southern Hemispheres.

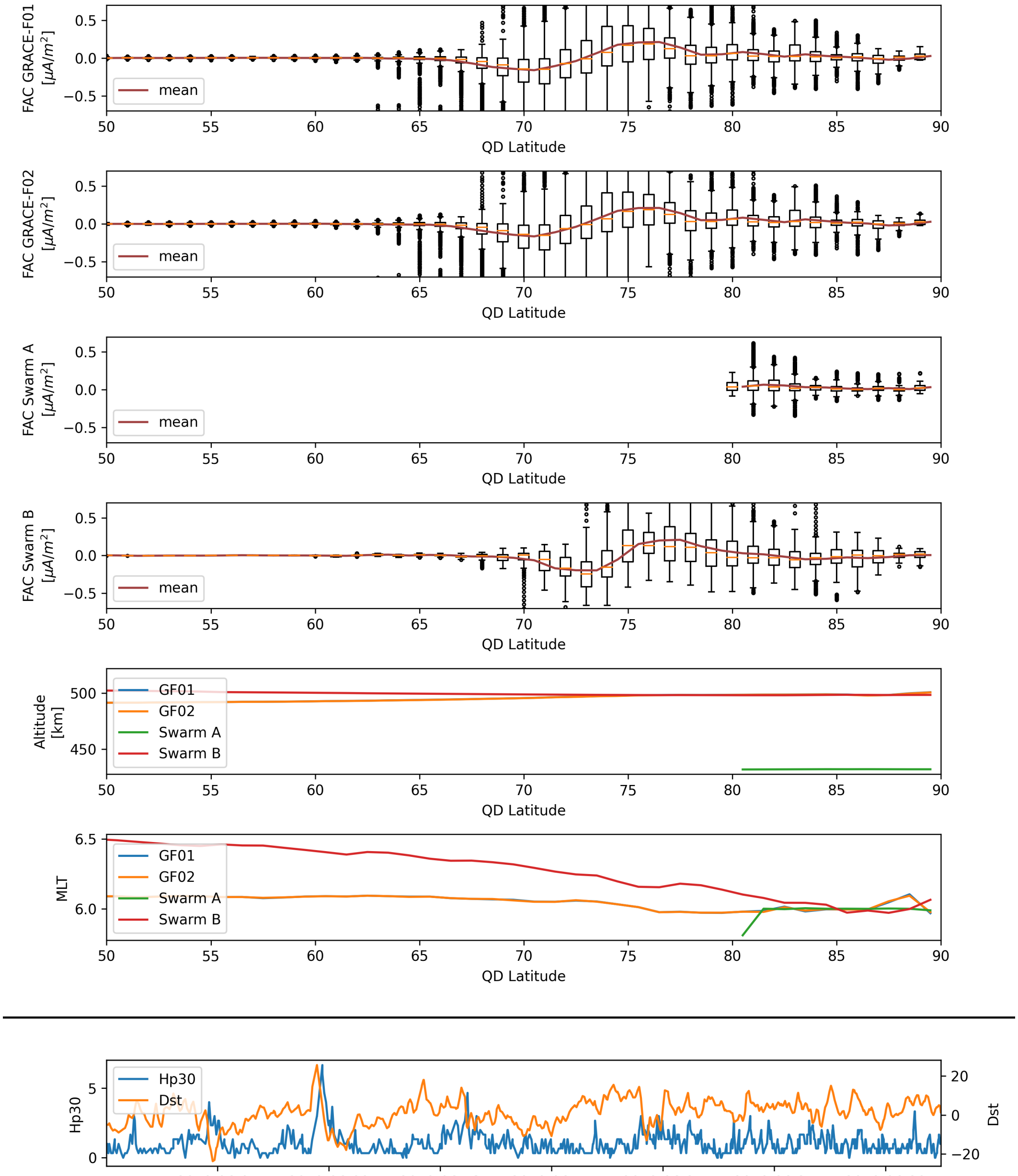


Figure 10. Boxplot summary for FAC data derived from GRACE-FO1, GRACE-FO2, Swarm-A, and Swarm-B missions. Data are selected from June 2019 for an MLT between 5.5 LT and 6.5 LT and quasi-dipole latitude between 50° and 90°, representing dawn. In addition, the altitude and MLT are given as mean values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

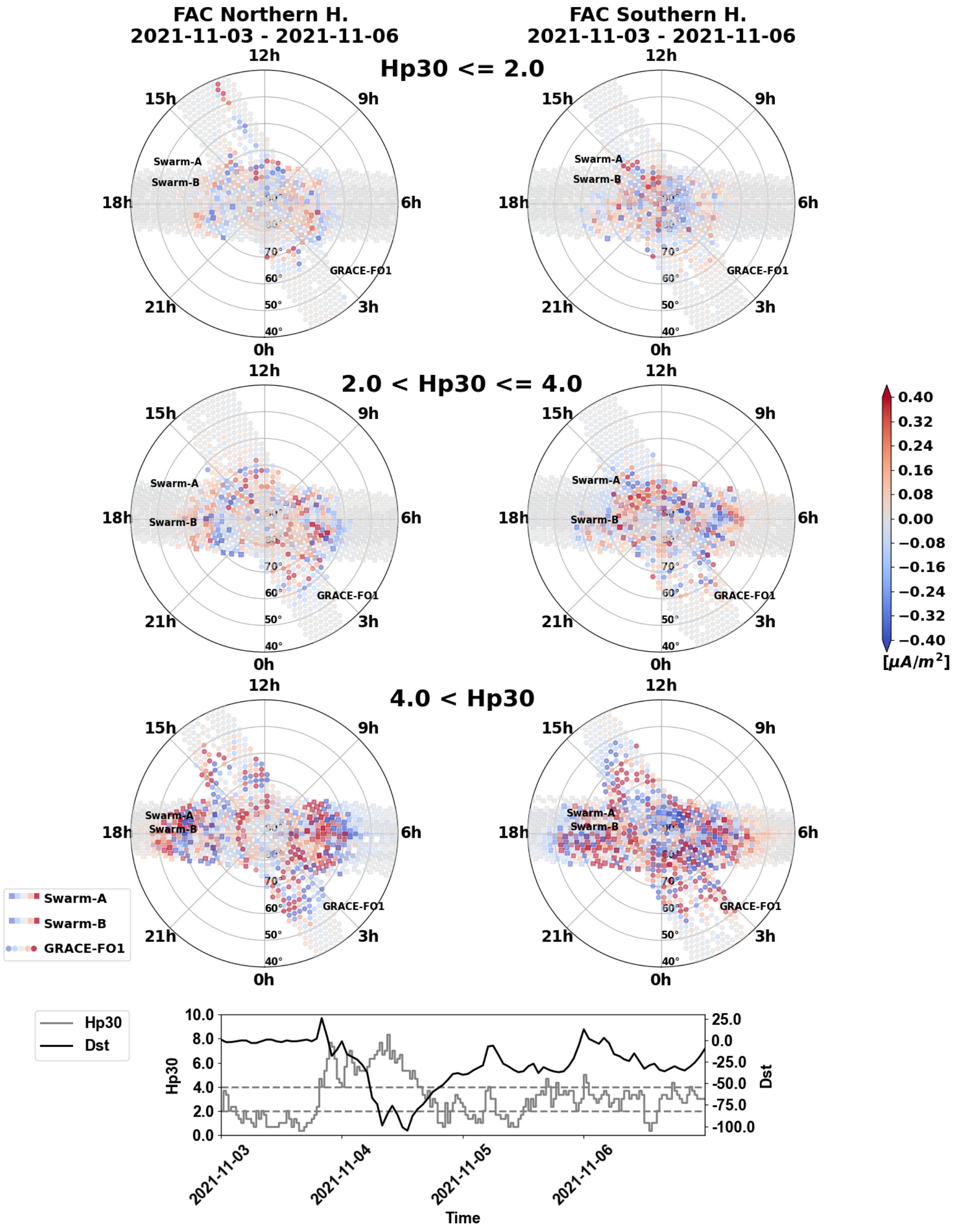


Figure 11. The mean FAC values for the Swarm-A, Swarm-B, and GRACE-FO satellite missions around the magnetic storm of the 4th of November, 2021, shown for four days, in dependence of magnetic local time and quasi-dipole latitude. The plot is divided into three rows, depending on the Hp30 index. Additionally,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

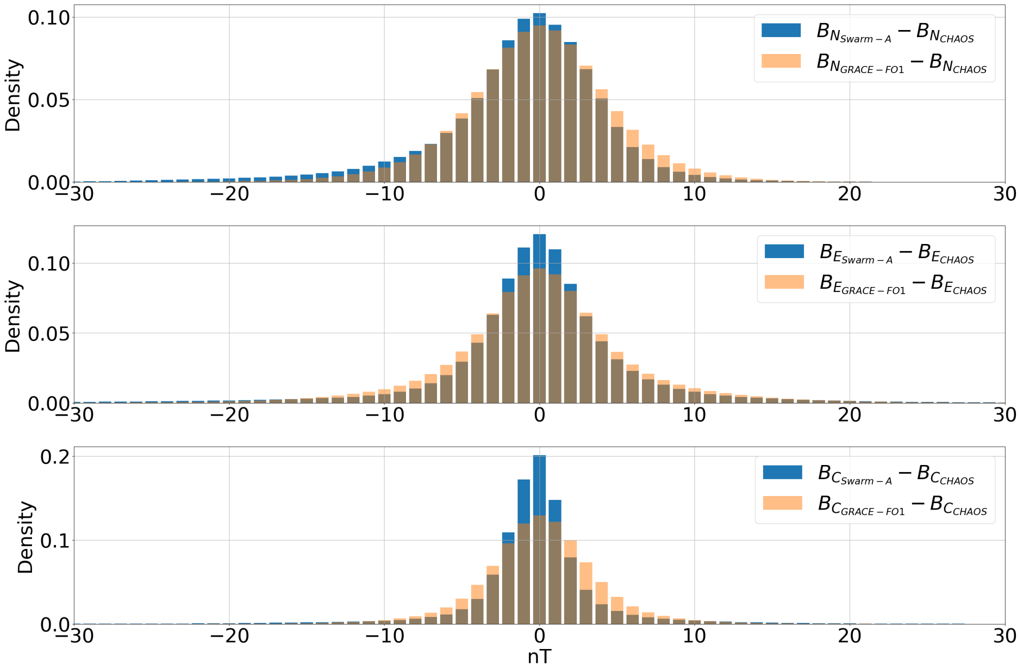


Figure 12. Residual distribution comparison of the Swarm (blue) and GRACE-FO1 (orange) calibrated data compared to the CHAOS7 reference model for the whole period from June 2018 to July 2023 within a histogram plot with bin sizes of 1 nT for the magnetic North (top), East (middle) and Center (bottom)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

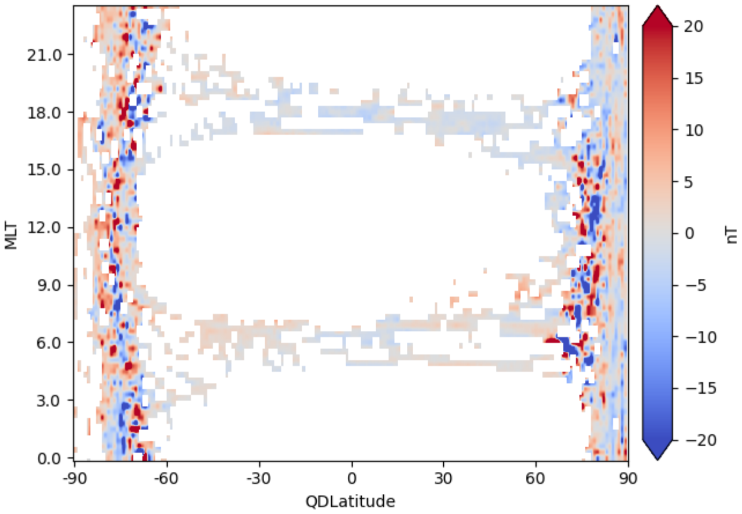


Figure 13. Conjunctions between the Swarm A and GRACE-FO1 satellites from June 2018 to July 2023. Data are selected by flags and geomagnetic quietness. The heatmap compares the residual to their respective CHAOS-7 model and shows the aggregated mean by quasi-dipole latitude and magnetic local time binning.

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.

8 SUMMARY AND PROPOSED FUTURE INVESTIGATIONS

In summary, these studies collectively advance the field of magnetometer calibration for satellite missions. Specifically, a sophisticated method for the automatic ML calibration of satellite platform magnetometer data has been shown.

In Section 5, an analytical approach for the calibration and characterization of platform magnetometers is presented with the GOCE mission as the use case. Here, the relevant features have been hand-selected with expert knowledge. A linear regression has been applied to correct the scaling, offset, and misalignment of the magnetometer measurements. In addition, the largest known artificial disturbances have been characterized and corrections applied. This led to a reduction of the noise present in the measured data, and the utility of calibrated data from non-dedicated magnetometers for filling gaps in dedicated magnetic field missions has been shown.

In Section 6, the methodology for an ML-based calibration for the GOCE mission is introduced, achieving a substantial reduction in residual compared to the CHAOS-7 reference model, with potential applications in studying the Earth's magnetic field and enhancing magnetic field models. Here, the data preprocessing pipeline has been established that is needed for the calibration. The GOCE satellite, with about 970 features monitored, acts as a use case for a satellite system with a huge variety of monitored information. The developed ML-based calibration methodology showed to be able to handle large data volumes and model the underlying calibration and characterization of the GOCE platform magnetometers. In addition, applications in the analysis of geophysical phenomena were shown. The extensive evaluation analyzed the remaining residual of the calibration and identified the most influential features.

A new approach introducing PINNs to the calibration of platform magnetometers is introduced in Section 7, demonstrating its efficacy in handling magnetic dipoles within the satellite system and achieving improved calibration quality, particularly in polar regions. Therefore, the Biot-Savart formula for dipoles has been successfully merged with an ML model, combining the advantages of physical and empirical models. In addition, the AMPS model has been combined with the CHAOS-7 model to form a combined reference model. This enhances the calibration performance in high latitudes. The application of this method to GOCE and GRACE-FO missions yields satisfying residual values enabling robust analysis of geomagnetic phenomena. In addition, this approach yields increased interpretability of the trained model by analyzing the learned locations and strengths of the dipoles present on the satellite, which makes it possible to draw conclusions about the satellite.

Regarding the research objectives posed in Section 1, Section 5 lays the foundation, and Sections 6 and 7 show the feasibility of the ML-based calibration (RO-1). The incorporation of physics-based information into the ML calibration is shown in Section 7 (RO-2). Finally, Sections 5 to 7 contribute to the extraction of information and explainable insights of the satellites (RO-3).

This dissertation establishes the foundation for several follow-up studies, with the findings and datasets from this thesis serving as a basis for future extensions.

First, the established method can be applied to other platform magnetometer-carrying satellites. As the proposed method leverages the advantages of ML, the approach can be applied to satellites previously calibrated using analytical methods like GRACE, CryoSat-2, or others. In addition, other satellites carrying platform magnetometers have been identified like TUBIN [6], Swarm-Echo, or EnMAP [28, 37], which are promising in their orbital and mission constellation.

Another way to establish a new direction would be to investigate additional modeling techniques. Therefore, a look at time-dependent features can be taken to model a better understanding of influences through time. Other than that, more physical phenomena can be modeled within the PINN, e.g., the influence of the temperature or thermoelectric currents [8].

The presented methodology is not limited to non-dedicated LEO satellites but can, in principle, be transferred onto other satellites, platforms, and orbits to calibrate the respective platform magnetometers. This could be applied to satellites in other orbits, e.g., higher altitudes than LEO satellites or satellites in an elliptical orbit. The requirement to use this approach is a period of time within the satellite's mission life span where a reference model is available so a calibration model can be created.

Finally, the usage of the datasets within geomagnetic field models is clearly one of the next steps founded on this work. This has, to some extent, been shown for the GOCE mission data and the modeling of the lithospheric field but could be done for a larger amount of satellites and longer time periods, combining the advantages of additional measurements and coverage provided by non-dedicated satellites.

To sum up, within this dissertation a novel ML-based approach to automatically calibrate platform magnetometers has been shown that overcomes the shortcomings and challenges present in earlier approaches. The robustness and wide range of potential applications for geomagnetism are demonstrated by the calibrated datasets of additional space-based magnetic measurements.

AUTHOR CONTRIBUTIONS

The first presented paper in Section 5 publishes the analytical approach of calibrating satellite-based magnetic field measurements which has been contrasted to the new Machine Learning approach, introduced and published by me in subsequent papers. The discussion of the results of this paper has been largely driven by the results achieved by the Machine Learning approach I was working on at this time. Thus, I importantly and actively contributed to the methodology of the study as well as to its manuscript, analysis, and interpretation.

I led the conceptualization and design of the study presented in Section 6, at which I was supported by Claudia and Ingo. The processing and calibration of the data, the production of the figures, and writing the manuscript was performed by me. I evaluated and reviewed the study, while the co-authors helped in the reviewing. The evaluation of the data with a lithospheric field model was performed by Julien.

I led the conceptualization and design of the study presented in Section 7, at which I was supported by Ingo, Claudia, and Monika. This manuscript is currently submitted to the JGR: Machine Learning and Computation journal. I implemented the methodology, processed the data, conducted the validation and formal analysis, produced the figures and visualizations, and wrote the manuscript. For the evaluation of the results, I collaborated with Ingo and Claudia.

REFERENCES

[1] Patrick Alken, Nils Olsen, and Christopher C Finlay. Co-estimation of geomagnetic field and in-orbit fluxgate magnetometer calibration parameters. *Earth, Planets and Space*, 72:1–32, 2020.

[2] Patrick Alken, Erwan Thébault, Ciarán D Beggan, Hagay Amit, J Aubert, J Baerenzung, TN Bondar, WJ Brown, S Califf, A Chambodut, et al. International geomagnetic reference field: the thirteenth generation. *Earth, Planets and Space*, 73(1):1–25, 2021.

[3] Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J Aljaaf. A systematic review on supervised and unsupervised machine learning algorithms for data science. *Supervised and unsupervised learning for data science*, pages 3–21, 2020.

[4] Brian J Anderson, Regupathi Angappan, Ankit Barik, Sarah K Vines, Sabine Stanley, Pietro N Bernasconi, Haje Korth, and Robin J Barnes. Iridium communications satellite constellation data for study of Earth's magnetic field. *Geochemistry, Geophysics, Geosystems*, 22(8):e2020GC009515, 2021.

[5] HU Auster, KH Glassmeier, W Magnes, O Aydogar, W Baumjohann, D Constantinescu, D Fischer, KH Fornacon, E Georgescu, P Harvey, et al. The THEMIS fluxgate magnetometer. *Space science reviews*, 141:235–264, 2008. doi: https://doi.org/10.1007/s11214-008-9365-9.

[6] Merlin F Barschke, Julian Bartholomäus, Karsten Gordon, Marc Lehmann, and Klaus Brieß. The TUBIN nanosatellite mission for wildfire detection in thermal infrared. *CEAS Space Journal*, 9:183–194, 2017.

[7] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

[8] Peter Brauer, Jose MG Merayo, Hauke Thamm, Lars Tøffner-Clausen, Manfred Amann, Gauthier Hulot, Pierre Vogel, Pierre Vigneron, and Rune Floberghagen. Magnetic perturbations from thermoelectric currents in Swarm thermal blankets. In *EGU General Assembly Conference Abstracts*, page 14561, 2018.

[9] Robert M Broadfoot, David M Miles, Warren Holley, and Andrew D Howarth. In situ calibration of the Swarm-Echo magnetometers. *Geoscientific instrumentation, methods and data systems*, 11(2):323–333, 2022.

[10] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002, 2019.

[11] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168, 2006.

[12] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv preprint arXiv:1511.07289*, 2015.

[13] David Donoho. 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4):745–766, 2017.

[14] MR Drinkwater, R Floberghagen, R Haagmans, D Muzi, and A Popescu. VII: Closing session: GOCE: ESA's first earth explorer core mission. *Space science reviews*, 108 (1):419–432, 2003.

[15] JT Emmert, AD Richmond, and DP Drob. A computationally compact representation of magnetic-apex and quasi-dipole coordinates with smooth base vectors. *Journal of Geophysical Research: Space Physics*, 115(A8), 2010.

[16] Christopher C Finlay. Models of the main geomagnetic field based on multi-satellite magnetic data and gradients—techniques and latest results from the swarm mission. *Ionospheric Multi-Spacecraft Analysis Tools: Approaches for Deriving Ionospheric Parameters*, pages 255–284, 2020.

[17] Christopher C Finlay, Clemens Kloss, Nils Olsen, Magnus D Hammer, Lars Tøffner-Clausen, Alexander Grayver, and Alexey Kuvshinov. The CHAOS-7 geomagnetic field model and observed changes in the South Atlantic Anomaly. *Earth, Planets and Space*, 72(1):1–31, 2020.

[18] Rune Floberghagen, Michael Fehringer, Daniel Lamarre, Danilo Muzi, Björn Frommknecht, Christoph Steiger, Juan Piñeiro, and Andrea Da Costa. Mission design, operation and exploitation of the gravity field and steady-state ocean circulation explorer mission. *Journal of Geodesy*, 85(11):749–758, 2011.

[19] Eigil Friis-Christensen, Hermann Lühr, and Gauthier Hulot. Swarm: A constellation to study the Earth's magnetic field. *Earth, planets and space*, 58:351–358, 04 2006. doi: 10.1186/BF03351933.

[20] Eigil Friis-Christensen, Hermann Lühr, D Knudsen, and R Haagmans. Swarm–an earth observation mission investigating geospace. *Advances in Space Research*, 41(1): 210–216, 2008.

[21] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[22] G Hulot, TJ Sabaka, and Nils Olsen. The present field. *Treatise of geophysics*, 5, 2007.

[23] Gauthier Hulot, André Balogh, Ulrich R Christensen, Catherine G Constable, Mioara Mandea, and Nils Olsen. The Earth's Magnetic Field in the Space Age: An Introduction to Terrestrial Magnetism. *Space Science Reviews*, 155(1-4):1, 2010.

[24] T Jager, Jean-Michel Léger, F Bertrand, I Fratter, and JC Lalaurie. SWARM Absolute Scalar Magnetometer accuracy: analyses and measurement results. *SENSORS, 2010 IEEE*, pages 2392–2395, 2010.

[25] V Roshan Joseph. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4):531–538, 2022.

[26] Baerenzung Julien, Holschneider Matthias, Jan Saynisch-Wagner, and Maik Thomas. Kalmag: a high spatio-temporal model of the geomagnetic field. *Earth, Planets and Space*, 74(1):139, 2022.

[27] Anuj Karpatne, William Watkins, Jordan Read, and Vipin Kumar. Physics-guided neural networks (pgnn): An application in lake temperature modeling. *arXiv preprint arXiv:1710.11431*, 2, 2017.

[28] Hermann Kaufmann, Karl Segl, Luis Guanter, Stefan Hofer, K-P Foerster, Timo Stuffler, Andreas Mueller, Rudolf Richter, Heike Bach, Patrick Hostert, et al. Environmental mapping and analysis program (EnMAP)-Recent advances and status. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 4, pages IV–109. IEEE, 2008.

[29] Michael C Kelley. *The Earth's ionosphere: Plasma physics and electrodynamics.* Academic press, 2009.

[30] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[31] Margaret Galland Kivelson and Christopher T Russell. *Introduction to space physics.* Cambridge university press, 1995.

[32] Clemens Kloss, Christopher C Finlay, and Nils Olsen. Co-estimating geomagnetic field and calibration parameters: modeling Earth's magnetic field with platform magnetometer data. *Earth, Planets and Space*, 73:1–21, 2021.

[33] Richard P Kornfeld, Bradford W Arnold, Michael A Gross, Neil T Dahya, William M Klipstein, Peter F Gath, and Srinivas Bettadpur. GRACE-FO: the gravity recovery and climate experiment follow-on mission. *Journal of spacecraft and rockets*, 56(3): 931–951, 2019.

[34] Felix Landerer, Frank Flechtner, Himanshu Save, Christopher McCullough, Christoph Dahle, Srinivas Bettadpur, Robert Gaston, and Krzysztof Snopek. Extending Global Mass Change Satellite Observations Into the Third Decade With GRACE-FO: Science Mission Status and Plans. GRACE/GRACE-FO Science Team Meeting 2022, 2022.

[35] J Scott Long and Jeremy Freese. *Regression models for categorical dependent variables using Stata*, volume 7. Stata press, 2006.

[36] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386, 2020.

[37] Thomas Meschede. *Design automation and performance analysis of modular reconfigurable attitude control systems.* PhD thesis, Dissertation, Berlin, Technische Universität Berlin, 2018, 2019.

[38] Ingo Michaelis, Kevin Styp-Rekowski, Jan Rauberg, Claudia Stolle, and Monika Korte. Geomagnetic data from the GOCE satellite mission. *Earth, Planets and Space*, 74:135, 2022.

[39] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[40] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. Deep learning applications and challenges in big data analytics. *Journal of big data*, 2(1):1–21, 2015.

[41] Brady Neal, Sarthak Mittal, Aristide Baratin, Vinayak Tantia, Matthew Scicluna, Simon Lacoste-Julien, and Ioannis Mitliagkas. A modern take on the bias-variance tradeoff in neural networks. *arXiv preprint arXiv:1810.08591*, 2018.

[42] Torsten Neubert, M Mandea, G Hulot, R Von Frese, Fritz Primdahl, John Leif Jørgensen, E Friis-Christensen, Peter Stauning, N Olsen, and T Risbo. Ørsted satellite captures high-precision geomagnetic field data. *Eos, Transactions American Geophysical Union*, 82(7):81–88, 2001.

[43] Andrew Y Ng. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78, 2004.

[44] Nils Olsen. Magnetometer data from the GRACE satellite duo. *Earth, Planets and Space*, 73(1):1–20, 2021.

[45] Nils Olsen, Eigil Friis-Christensen, Rune Floberghagen, Patrick Alken, Ciaran D Beggan, Arnaud Chulliat, Eelco Doornbos, João Teixeira Da Encarnação, Brian Hamilton, Gauthier Hulot, et al. The Swarm satellite constellation application and research facility (SCARF) and Swarm data products. *Earth, Planets and Space*, 65:1189–1200, 2013.

[46] Nils Olsen, Hermann Lühr, Christopher C Finlay, Terence J Sabaka, Ingo Michaelis, Jan Rauberg, and Lars Tøffner-Clausen. The CHAOS-4 geomagnetic field model. *Geophysical Journal International*, 197(2):815–827, 2014.

[47] Nils Olsen, Giuseppe Albini, Jerome Bouffard, Tommaso Parrinello, and Lars Tøffner-Clausen. Magnetic observations from CryoSat-2: calibration and processing of satellite platform magnetometer data. *Earth, Planets and Space*, 72(1):1–18, 2020.

[48] Fritz Primdahl. The fluxgate magnetometer. *Journal of Physics E: Scientific Instruments*, 12(4):241, 1979.

[49] Gerd Prölss. *Physics of the Earth's space environment: an introduction.* Springer Science & Business Media, 2012.

[50] Ch Reigber, Hermann Lühr, and P Schwintzer. CHAMP mission status. *Advances in space research*, 30(2):129–134, 2002.

[51] Martin Rother, Monika Korte, Achim Morschhauser, Foteini Vervelidou, Jürgen Matzka, and Claudia Stolle. The Mag. num core field model as a parent for IGRF-13, and the recent evolution of the South Atlantic Anomaly. *Earth, Planets and Space*, 73:1–17, 2021.

[52] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[53] Artem Smirnov, Yuri Shprits, Fabricio Prol, Hermann Lühr, Max Berrendorf, Irina Zhelavskaya, and Chao Xiong. A novel neural network model of Earth's topside ionosphere. *Scientific Reports*, 13(1):1303, 2023.

[54] C Stolle, J Baerenzung, EA Kronberg, J Kusche, H Liu, and H Shimizu. Special issue "DynamicEarth: Earth's interior, surface, ocean, atmosphere, and near space interactions". *Earth, Planets and Space*, 75(1):137, 2023.

[55] Claudia Stolle, Ingo Michaelis, C Xiong, Martin Rother, Th Usbeck, Yosuke Yamazaki, Jan Rauberg, and K Styp-Rekowski. Observing Earth's magnetic environment with the GRACE-FO mission. *Earth, Planets and Space*, 73:51, 2021.

[56] Claudia Stolle, Nils Olsen, B Anderson, Eelco Doornbos, and Alexey Kuvshinov. 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, 2021.

[57] Kevin Styp-Rekowski, Claudia Stolle, Ingo Michaelis, and Odej Kao. 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)*, pages 5283–5290. IEEE, 2021.

[58] Kevin Styp-Rekowski, Ingo Michaelis, Claudia Stolle, Julien Baerenzung, Monika Korte, and Odej Kao. Machine learning-based calibration of the GOCE satellite platform magnetometers. *Earth, Planets and Space*, 74:138, 2022.

[59] Kevin Styp-Rekowski, Ingo Michaelis, Monika Korte, and Claudia Stolle. Physics-informed Neural Networks for the Improvement of Platform Magnetometer Measurements. *Authorea*, January 2024. doi: 10.22541/au.170602061.18680921/v1. URL `https://doi.org/10.22541/au.170602061.18680921/v1`.

[60] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Elsevier, 2006.

[61] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems. *ACM Computing Surveys*, 55(4):1–37, 2022.

[62] Irina S Zhelavskaya, Nikita A Aseev, and Yuri Y Shprits. A combined neural network- and physics-based approach for modeling plasmasphere dynamics. *Journal of Geophysical Research: Space Physics*, 126(3):e2020JA028077, 2021.

Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig angefertigt und ohne fremde Hilfe verfasst habe. Dazu habe ich keine außer den von mir angegebenen Hilfsmitteln und Quellen verwendet und die den benutzten Werken inhaltlich und wörtlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Rostock

(Abgabedatum) (Vollständige Unterschrift)