

# Technical Report for Calibrated GRACE-FO Magnetometer Dataset

Kevin Styp-Rekowski  
styp-rekowski@tu-berlin.de  
TU Berlin  
Berlin, Berlin, Germany

Claudia Stolle  
Ingo Michaelis  
GFZ Potsdam  
Potsdam, Brandenburg, Germany

Odej Kao  
TU Berlin  
Berlin, Berlin, Germany

## 1 DATA ORIGIN

The data used for this dataset was published by Stolle et al. [1] and is available online <sup>1</sup>. The data is provided in NASA's CDF file format <sup>2</sup> which is a special file format for efficiently providing data together with descriptions and labels.

The used data is structured in different versions where for this publication the data from version 0201 was used, the main difference between the versions is the analytically calibrated result dataset, stored in 'ACAL\_CORR'. The version's subfolder is subdivided into the two satellites which are again subdivided into 7 folders, from which 'APEX', 'CHAOS7', and 'RAW' were mainly used and combined. This combined data includes the input data, meta-information like the QD-latitude and QD-longitude, as well as the target data.

## 2 PREPROCESSING STEPS

Before the proposed approach in our publication (to be published) was applied several steps of preprocessing have been applied. For any further action the data is aligned with the available Timestamp and can be considered as a matrix where each row represents a data point, identified by the timestamp, where the available information is represented by entries in the different columns that contain the features.

### 2.1 Preparation of interpolation functions

The linear 1-dimensional interpolation functions, which have been used in the proposed approach to calculate the time shift as well as applying the time shift, are based on all available data of a certain month without any restrictions. Normally, this means that there are data points in a resolution of 1Hz available. Therefore, one month of data has about 2.6 million data samples that are stored and used for the time shift application. For the application of the time shift, the time or index-based component of the interpolation functions are shifted according to the set or calculated time-shift and the result of the interpolation function is calculated. Thus, the shift in time is applied by the linear 1-dimensional interpolation functions according to the data.

### 2.2 Data Selection and Data Cleaning

As a first data cleaning step, all rows in the data are removed which contain any missing data in one of the columns. Only very small parts of the data are affected by this removal. Thus, a more sophisticated missing data handling is not needed for this dataset.

The data is then filtered for the calibration process for magnetic quiet times. Therefore, the Kp- and Dst-Indices are used. Only data with a Kp-Index  $k_p \leq 2$  and a Dst-Index  $|Dst| \leq 30$  are retained. As the CHAOS-7 Model does not model the magnetic disturbances, the training would otherwise not be able to reasonably model the underlying process. This is the reason for retaining the magnetic quiet data, such that the ground truth model is actually closely related to the remaining data.

For the magnetorquers, a saturation voltage has been detected which means that for these measurements the satellite is considered to be leaving its normal attitude. Thus, similar to [1], rows containing magnetorquer values equal or larger than 110 are filtered and removed from the available data. Similar to the missing data, these are only very few samples.

The data is not filtered for a certain QD-latitude interval as in [1] but instead these high latitudes are weighted lower as described in our publication (to be published).

Finally, features with a standard deviation of 0 are dropped before the training. This is done on a monthly basis as the models are trained month-wise. Features with a standard deviation of 0 are considered to be constant and thus considered to not add any information to the training. Dropping such features reduces the training time of the neural network.

### 2.3 Scaling

The additional housekeeping features which provide information about the payload and other systems onboard the satellite are scaled to lie in the same interval. For every month, for every feature from the selected dataset as described in section 2.2 is scaled to an interval of  $[-1, 1]$ , so all features are equally weighted and scaled to the same order of magnitude. This scaling is not applied for the raw measurements of the magnetometers as these are in general near to the target CHAOS-7 data in terms of order of magnitude. The same scaling is later applied when the calibrated dataset is generated after the training of the model. Thus, in the whole unselected dataset possibly values outside the given interval can be encountered which can be handled by the neural network.

## 3 DATA PUBLICATION STRUCTURE

Similar to the utilized data, the final calibrated dataset is published under the same link as the data origin within the 'ML' subfolder. Again, there is a subfolder for each of the two satellites where within the folders the CDF files can be found. The structure is the same as for the data origin where the analytically calibrated dataset could be found in the folder 'ACAL\_CORR' where the dataset calibrated with machine learning can be found in the folder 'ACAL\_CORR\_ML', containing the same variables as in the data origin, namely the

<sup>1</sup>[ftp://isdcftp.gfz-potsdam.de/grace-fo/MAGNETIC\\_FIELD/](ftp://isdcftp.gfz-potsdam.de/grace-fo/MAGNETIC_FIELD/)

<sup>2</sup><https://cdf.gsfc.nasa.gov/>

calibrated values of the magnetometer measurements in satellite frame as well as NEC, quaternions for the transformation between the frames, positional information, and 'B\_FLAG' which similarly to the data origin describes the data with a magnetorquer saturation as described in 2.2. Other data like QD-latitude and QD-longitude as well as the CHAOS-7 values can be derived from the data origin which follows a similar structure and can be matched via timestamps. The published data contains only the calibrated dataset utilizing machine learning to reduce data redundancy.

## 4 EVALUATION FIGURES

Explain additional evaluation and its structure for months as well as kp and non-kp selections

There is an additional evaluation folders, containing residual plots for all months under different conditions. The folder structure is organized by the satellites first (either GRACE-FO1 or GRACE-FO2), followed by folders with the naming structure "KP/DST-SELECTION\_YEARMONTH" where the Kp- / Dst-Selection is either "kpdst" or "all" referring to the data being filtered by low Kp- and Dst-Indices as explained in section 2.2 or respectively to all data without any magnetic quiet time filtering. The YEARMONTH placeholder is filled with the year and the month in a format like "201806" for the June of 2018 exemplary.

There are two kinds of plots included in the files: Map and Orbit plots. The map plots come with 3 files which are separated for

either all the current data, only the ascending orbits, or only the descending orbits, indicated by an ending of "\_all", "\_asc", or "\_desc" respectively. The plots show in the top row the x and y component residuals of the magnetometer measurements with the CHAOS-7 model ground truth, the lower row contains the z component and the color scale. The residuals are binned in bins of 5 and plotted with a contour plot over the earth's surface.

The orbit plots come with 9 files for each month, where there are 3 files for the 3 axes for each 3 of the data selection methods for either all orbits, only ascending orbits, or only descending orbits. The name indicates the membership of the file, e.g. "monthly\_fgm\_orbits\_all\_orbit\_residual\_x" contains the figure for all available orbits for the x-axis of the magnetometer measurement. The plots contain the residual between the calibrated data and the CHAOS-7 model ground truth, the left side shows the proposed ML calibration, while the right side shows the previous analytical calibration from [1]. Grey lines mark the 50th QD-latitude which is considered to not include any field-aligned currents (FAC) and was used for the analytical calibration.

## REFERENCES

- [1] Claudia Stolle, I Michaelis, C Xiong, Martin Rother, Th Usbeck, Yosuke Yamazaki, Jan Rauberg, and K Styp-Rekowski. 2021. Observing Earth's magnetic environment with the GRACE-FO mission. *Earth, Planets and Space* 73, 1 (2021), 1–21.