# An Operational Geomagnetic Baseline Derivation Method for Magnetic Observatories Located in Mid-Latitudes

Veronika Haberle, Aurélie Marchaudon, Aude Chambodut, Pierre-louis Blelly

Space Weather

RESEARCH ARTICLE
10.1029/2024SW004048

# An Operational Geomagnetic Baseline Derivation Method for Magnetic Observatories Located in Mid-Latitudes

**Veronika Haberle[1,2,3,4], Aurélie Marchaudon[1], Aude Chambodut[2,3], and Pierre-Louis Blelly[1]**

[1]Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse, CNRS, CNES, Toulouse, France, [2]Université de Strasbourg, CNRS, Strasbourg, France, [3]Université de Strasbourg, CNRS, Ecole et Observatoire des Sciences de La Terre, Strasbourg, France, [4]Conrad Observatory, GeoSphere, Trafelberg, Austria

**Key Points:**

- A geomagnetic baseline method is proposed that can be deployed in operational (near-) real time environments
- An algorithm that identifies magnetic disturbances within geomagnetic field measurements is introduced
- Possible replacements for magnetic quiet variations during storm and disturbance periods are elaborated

**Supporting Information:**

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

**Correspondence to:**

A. Marchaudon,
aurelie.marchaudon@irap.omp.eu

**Citation:**

Haberle, V., Marchaudon, A., Chambodut, A., & Blelly, P.-L. (2024). An operational geomagnetic baseline derivation method for magnetic observatories located in mid-latitudes. *Space Weather*, *22*, e2024SW004048. https://doi.org/10.1029/2024SW004048

Received 24 JUN 2024
Accepted 12 DEC 2024

**Author Contributions:**

**Conceptualization:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut, Pierre-Louis Blelly
**Data curation:** Aude Chambodut
**Formal analysis:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut, Pierre-Louis Blelly
**Funding acquisition:** Aurélie Marchaudon, Pierre-Louis Blelly
**Methodology:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut, Pierre-Louis Blelly
**Project administration:** Aurélie Marchaudon
**Software:** Veronika Haberle, Pierre-Louis Blelly



**Abstract** Ground magnetic field measurements are an important tool to determine the strength of space weather events in the terrestrial environment on local and global scales. For that purpose, geomagnetic baselines play a vital role as they describe typical quiet variations within geomagnetic data which allows the successive isolation of magnetic storm contributions. This work introduces an operational baseline derivation method to accurately assess then replace the amplitude of space weather events within high-quality ground magnetic field measurements from magnetic observatories located in mid-latitudes. A two-step approach first identifies storm and disturbance intervals within the magnetic signal. Quiet variations, consistent with pre- and post-disturbance periods, are then used to replace the signal during the identified intervals. The final baseline is validated through comparisons with existing methods and through demonstration during moderate and strong space weather events at 13 globally distributed observatories, demonstrating its ability to track quiet variations accurately while maintaining them during disturbances. This supports the application of the introduced baseline for geomagnetic field description and new magnetic index derivation for space weather event characterization with high spatio-temporal resolution. As the method is deployable in near real-time, it is suitable for operational environments.

**Plain Language Summary** This research article is about a new method to improve the measurements of space weather effects on Earth's magnetic field. Space weather can cause significant disturbances in Earth's magnetic field, which are measureable from ground magnetic observatories. The new method helps remove these disturbances from the ground magnetic field measurements by describing the quiet magnetic field variations. This is important for understanding space weather and its effects. The introduced method works in two steps. First, it identifies periods of space weather activity in the measurements. Then, it replaces those periods with data from calmer times before or after the event. We show that this method works well for observatories in mid-latitude regions and can be deployed in near real-time making it useable in operational environments.

## 1. Introduction

The Earth's magnetic field, or geomagnetic field, is well known to follow variations and perturbations of solar origin. External perturbations coming from the Sun's surface, traveling through the solar wind and interacting with the Earth's magnetosphere and ionosphere, are the cause of intense and variable electric currents which modify the geomagnetic field and can cause geomagnetic storms (Gonzalez et al., 1994). The tracking and forecasting of these strong disturbances in the Earth's spatial environment has become a major challenge for space weather, requiring monitoring with improved spatio-temporal resolution.

Magnetic ground observatories (or stations) have been deployed as early as the beginning of the nineteenth century to monitor magnetic variations and their network has ever since been expanding to cover accessible latitudes and longitudes at the Earth's surface. Very early on, interest in generating proxies of magnetic perturbations has arisen, the aim being to characterize and access the strength of these perturbations but also to identify quiet periods allowing to isolate the internal contribution of the geomagnetic field. Bartels et al. (1939) were first to suggest such a proxy, called K-indices. Still widely used nowadays, these indices record the strength of the external magnetic perturbations at each observatory accounting for its ground location over a 3-hr period and use a logarithmic scale for describing the levels of activity. These K-indices were then combined over different networks of stations to give rise to magnetic indices describing Earth's magnetic activity state on a planetary scale.

**Supervision:** Aurélie Marchaudon, Aude Chambodut, Pierre-Louis Blelly
**Validation:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut, Pierre-Louis Blelly
**Visualization:** Veronika Haberle, Aurélie Marchaudon
**Writing – original draft:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut
**Writing – review & editing:** Veronika Haberle, Aurélie Marchaudon, Aude Chambodut

The first global index, the Kp index, was introduced by Bartels (1949) with its associated network of stations being geographically unevenly distributed and sparse for historical reasons. Then, Mayaud (1967) tried to overcome this drawback by developing the am index which is based on a more widely spread network. He also built the aa index, based on only 2 antipodal stations, but capable of going back in time as early as 1868 (Mayaud, 1972). Despite their broad scope of application, these indices are not perfectly suited to follow active solar periods when rapid temporal and regional variations are of interest. Nevertheless, Matzka et al. (2021) reviewed the properties of the Kp index and introduced the near real-time production of the nowcast Kp index. To overcome the main limitations of the K-indices, Yamazaki et al. (2022) introduced the Hpo indices which have a higher temporal resolution of up to 15 min and do not have a maximum class such that stronger storms can be classified. A few years later, further local indices were designed to record magnetic perturbations associated with specific current systems of ionospheric and/or magnetospheric origin, especially during magnetic storms. Examples include the family of AE indices aiming at recording auroral electrojets (Davis & Sugiura, 1966), the PC indices aiming at characterizing the North and South polar cap potentials (Troshichev & Janzhura, 2012), the Dst and the family of ASY/SYM indices (Iyemori, 1990; Sugiura, 1964; Sugiura & Kamei, 1991) describing the ring and magnetopause currents.

Whether describing the general state of the geomagnetic field or tracking the intensity of a dedicated current system, all indices have in common that they characterize the corresponding irregular variations within magnetic field measurements. As such, the contribution from these sources of interest have to be isolated within the geomagnetic field signal. Generally, this is achieved by removing the geomagnetic baseline. Thus, after this baseline subtraction, the residuals are supposed to contain solely information on the magnetic disturbances from external origin. Bartels et al. (1939) defined the geomagnetic baseline as

> "…a smooth curve (a regular daily variation) to be expected for that element on a magnetically quiet day, according to the season, the sunspot-cycle, and, in some cases, the phase of the Moon"

which is assumed to encompass the solar daily variation, the lunar daily variation and the after-effects of disturbances like recovery phases of the ring current enhancement. At the time, trained observers identified these quiet curves on analogue magnetograms by hand. This rather subjective definition was later made more concrete into a set of 7 practical rules by Mayaud (1980) to support the objectivity of the baseline determination. However, up to this date there is no qualitative ground truth on the shape of geomagnetic baselines and the general guidelines from the past are still widely employed to justify baseline derivations. The consensus that is forming nowadays is that a geomagnetic baseline includes the secular variation, solar cycle induced variations and the solar quiet variations (Gjerloev, 2012; Kamp, 2013).

Each of the introduced index families uses different methods to determine a suitable baseline and each of the baseline methods produces different baselines, that lead to differences in the disturbance fields (i.e., the residuals after removal of the dedicated baseline method). For mid-latitude observatories, the algorithm from Sucksdorff et al. (1991), referred to as the FMI method (the abbreviation coming from Finnish Meteorological Institute), is often deployed to produce the baselines used to generate the local K and thus K-derived magnetic activity indices such as Kp indices. It is based on a regression over a sliding window centered on the 24hr UT day, whereby the solar quiet variations are determined by using a fifth order harmonics on hourly means. More recently, Gjerloev (2012) introduced a baseline derivation method for SuperMAG indices, which involves rotating magnetometer measurements into a specific local magnetic coordinate system. This method incorporates long-term trends and diurnal variations by removing the mode (the value with the highest occurrence rate) for each day from the measurements, and then applying cubic convolution interpolation and weighted smoothing procedures to derive the diurnal variations and long-term trends, respectively.

Detailed investigations of the actually contained sources in these baselines are widely absent. Some methods use techniques of statistical averaging to describe quiet variations. When applied to mid-latitudes, this leads to an averaging of the Sq current footprints fully neglecting its intrinsic day-to-day (D2D) variability which amplitude can be in the order of weak to moderate magnetic disturbances. For example, Gjerloev (2012) shows that the baseline subtracted spectrum does not include the 24hr harmonics and thus argues that the solar quiet variations are included within the baseline. Additionally, many of the currently available baseline methodologies are not ready to be deployed in operational, near-real time environments for which the need grows as space weather now- and forecasting are getting more relevant.

In this study, our aim is to calculate a new baseline for mid-latitude observatories which allows deriving new generations of planetary magnetic indices with an improved spatio-temporal resolution while accurately following geomagnetic activity during storm events, such as for example, in Chambodut et al. (2015), while aiming at being applicable in near-real time. To check for consistency, we will compare our new derived baseline with the other baselines used for mid-latitude observatories and introduced above, even if ultimately, we do not expect them to be the very same. For such a purpose, Haberle et al. (2022) introduced a geomagnetic baseline derivation methodology for which quiet source contributions have been extensively studied and discussed. The derivation method is based on signal filtering and will be shortly recalled in Section 2.1. Its main interest lies in the fact that no a priori information is needed to derive it. This new baseline, named thereafter the filter baseline, follows the magnetic field variations tightly, in particular the D2D variability, which is a desirable characteristic during geomagnetically quiet conditions. However, the filter baseline is sensitive to disturbances. Specifically, disturbance and storm contributions within the filter baseline are still present during non-quiet periods, causing an underestimation of the strength of these perturbations in the residuals. Thus, we acknowledge that the filter baseline is not directly applicable for all time periods. The current paper addresses this issue and proposes a methodology to remove storm and disturbance contributions within the filter baseline. The Chambon-la-Forêt (CLF) magnetic observatory is used as a proof of concept and to illustrate the methodology. The first step is to identify disturbances within geomagnetic field signals, described in Section 2.2, and the second is to replace the filter baseline during the identified intervals with valid quiet variations, described in Section 2.3. In Section 3 we compare the introduced baseline algorithm to other methodologies. In Section 4, the algorithm is applied to different mid-latitude magnetic observatories to support its global application. In Section 5, the qualities and limitations of this method are discussed, in particular the fact that the filter baseline and the introduced methodology are solely reliant on magnetic field measurements and thus are suitable for (near-) real time applications.

## 2. Adapting the Filter Baseline During Disturbances

### 2.1. The Filter Baseline

We briefly recall the definition of the filter baseline from Haberle et al. (2022). For details please relate to their work directly. The goal of the geomagnetic baseline is to isolate contributions from quiet sources within geomagnetic field measurements. The geomagnetic field is a superposition of various sources. Assuming that sources operate on dedicated frequency bands, temporal filtering is applied to each of the geomagnetic horizontal components $X$ and $Y$ to extract the quiet sources. In mid-latitudes these are mainly the secular variation and the solar quiet current systems. The secular variation is extracted by applying a low-pass filter with cut-off period of 36 hr, noted as $x_{>24}, y_{>24}$. The solar quiet current systems are extracted by a set of four band-pass filters with passing periods of 36hr–24 hr; 24hr–12 hr; 12hr–8 h and 8–6 hr. The finite impulse response filters have a window-size of 3 days and make use of the Hamming-window. The corresponding filter responses are notated with subscripts, that is, $x_{24}$, $x_{12}$, $x_8$, $x_6$ and $y_{24}$, $y_{12}$, $y_8$, $y_6$. The sum of the four (sub-) diurnal filters results in the daily variations $x_D$, $y_D$ as a band-pass filter between 36hr and 6 hr. The combination of long-term filter and diurnal filter output defines the filter baseline, that is,

$$x_{FB} = x_{>24} + x_D \quad \text{for } X, \text{ and} \tag{1}$$

$$y_{FB} = y_{>24} + y_D \quad \text{for } Y \tag{2}$$

As is shown in Haberle et al. (2022) the filter baseline is well suited to follow typical variations during geomagnetically quiet periods and captures quiet sources accurately, including the intrinsic day-to-day variability of the Sq currents. However during periods of geomagnetic disturbances, the filter baseline follows the storm-induced portion too closely. In order to remove these unwanted contributions, the strategy is to substitute the filter baseline during disturbance periods with representative quiet variations. This is achieved in two steps.

1. Identification of disturbance periods, that is, periods during which the filter baseline does not follow quiet variations, see Section 2.2
2. Replacement of identified disturbance periods with representative quiet variations, see Section 2.3.

Keeping in mind that our goal is to deploy the final baseline in an operational, possible near-real time context, we pose the requirement that the identification of disturbance intervals and the corresponding replacement baseline are achieved without further manual intervention.

The magnetic field data used in this study comes from the Real-time Magnetic Observatory Network (INTERMAGNET) (see intermagnet.github.io). The advantage of using this data repository is that geomagnetic field measurements are of high quality, having a resolution of 0.1 nT and come in 1 minute resolution. The coordinate system in which the data is provided is the local geographic coordinate system which *X*-axis points toward North, *Y*-axis toward East and the *Z*-axis vertically Down, referred to as NED system. Here we concentrate at the horizontal component of the geomagnetic field, that is, the *X* and Y components of the magnetic field vector as these are typically used in magnetic index derivations. We focus on stations in mid-latitudes, in the range of absolute eccentric dipole magnetic latitudes from 20° to 60° at which auroral and equatorial electrojets play a secondary role. In the subsequent subchapters we deploy the magnetic observatory Chambon-la-Forêt as a proof of concept and to showcase the methodology. Thresholds, when used, are defined by station.

### 2.2. Detecting Disturbances Within Geomagnetic Field Measurements

It is common to make use of the prominent deflection of the horizontal magnetic field component to identify geomagnetic storms locally, see for example, Bailey and Leonhardt (2016), or globally by using magnetic indices like Dst or SYM-H, see for example, Echer et al. (2011) or Walach and Grocott (2019). These studies are limited to strong events only and do not consider weak to moderate disturbances. For the geomagnetic baseline, however, we are rather interested in a consistent deviation from quiet variations. Haberle et al. (2022) found that storm and disturbance contributions are present in each of the filter response. Though unwanted in the final baseline, we can leverage this circumstance to identify such deviations, that is, disturbed periods. For illustration we show in the following a period between 9th to 28th of December 2002 that includes a moderate event with minimum SYM-H of −49 nT during an initial quiet period, followed by a strong event with minimum SYM-H of −90 nT and a rather disturbed period (see the upper panel of Figure 3). Similar to Hutchinson et al. (2011) and Walach and Grocott (2019), we use the SYM-H index to identify the duration of the storm (from the start of the initial phase to the end of the recovery phase). The minimum SYM-H marks the end of the main phase, and the end of the recovery phase is taken as the crossing of −15 nT. The onset times of the initial phase is selected manually for the two storms. The full duration for the moderate event is from 2002 to 12–14 11:44 to 2002-12-15 20:29 and for the strong event from 2002 to 12–18 18:26 to 2002-12-22 10:30.

#### 2.2.1. Identification of Strong Magnetic Storms

Haberle et al. (2022) concluded that, in addition to storm signatures, the main source contributing to the long-term filter responses $x_{>24}$, $y_{>24}$ is the secular variation that induces the observed smooth trend on top of recurrent seasonal and 27-day oscillations. To isolate and highlight the storm signatures from the general trend and recurrent variations, we remove the moving average over the past 27-day to obtain the de-trended long-term filter responses. A window of 27-day guarantees the smoothing of the 27-day variation and the removal of seasonal and longer-term variations. We introduce the 27-day de-trended long-term horizontal intensity $\bar{h}_{>24}$, defined as

$$\bar{h}_{>24}(t) = \sqrt{x_{>24}(t)^2 + y_{>24}(t)^2} - \frac{1}{27d}\sum_{i=t-27d}^{t}\sqrt{x_{>24}(i)^2 + y_{>24}(i)^2}, \tag{3}$$

The horizontal intensity enables us to compare it to the SYM-H index. Indeed, considering the entire period between 1991 and 2019 in min resolution for CLF, $\bar{h}_{>24}$ and SYM-H are linearly correlated with a Pearson coefficient of 0.79, indicating strong linear correlation. This indicates that, when geomagnetic activity increases, the long-term filter values not only increase too, but increase linearly with the SYM-H index. Thus, for any given time-step *t*, we can derive the linear relationship

$$\bar{h}_{>24}(t) \approx a(t) \times \text{SYM-H} + b(t), \tag{4}$$

where the linear fit coefficients a(t) and b(t) are based on $\bar{h}_{>24}(t)$ and SYM-H values from the past 11 years to ensure that any solar cycle and internal field variations are taken into account.

Owing to threshold definitions for SYM-H, we can leverage (Equation 4) to derive a threshold $h_L(t)$ for $\bar{h}_{>24}(t)$. The threshold $h_L$ serves as discrimination for strong enough geomagnetic disturbances that is closely related to the method of using a SYM-H threshold through the found linear correlation. The only question remaining is which SYM-H value is best to be used for defining the threshold $h_L$. In the literature, it is common to define storms by their minimum Dst or SYM-H, for example, the conditions Dst $< -50$ nT or SYM - H $< -50$ nT are commonly used to define strong storms, see for example, Echer et al. (2011); Walach and Grocott (2019). This threshold is also consistent with statistical patterns of CLF magnetic variations binned by level of activity characterized by the SYM-H index for which the quiet pattern starts to deviate for SYM-H level comprised between $-40$ and $-50$ nT (see Figure S1 in Supporting Information S1 for details), this is thus supporting the choice of $-50$ nT as SYM-H threshold in the linear relationship (Equation 4).

For each time-step $t$, we fit the linear relationship between SYM-H and $\bar{h}_{>24}$ based on the past 11 years and using the determined threshold $-50$ nT, we derive $h_L$. Once $h_L$ is known, the identification of strong geomagnetic disturbances is enabled as soon as the long-term filter values are available. The linear fits and thresholds $h_L$ for CLF and further observatories can be found in the supplementary material (see Table S1 in Supporting Information S1 for CLF, as well as other magnetic observatories).

Each time-step $t$ that satisfies $\bar{h}_{>24}(t) < h_L(t)$ is then marked as a disturbance for both of the components, that is, both $x_{FB}(t)$ and $y_{FB}(t)$ need to be replaced. In the upper panel of Figure 1, $\bar{h}_{>24}$ and $h_L$ are illustrated. All values beneath $h_L$ are marked as a disturbance $t_D$, indicated in orange. The lower two panels indicate the resulting disturbance intervals $t_D$ on the components $X$ and $Y$ (also in orange) at CLF. It is clear that the onset and parts of the main phase are missed. To overcome this issue, the most recent previous local maximum $t_{\max}$ is identified marking the start-time of the event (indicated by the yellow triangle). Looking for the onset in this manner is in agreement with typical definition of storms, where the local maximum before a Dst or SYM-H minimum is used as start-time. As the shape of the long-term filter response is a wave, such a maximum can always be found, even if the storm is not accompanied by a clear Sudden Storm Commencement (SSC). The time-interval $t_O$ (in yellow) up until $t_{\max}$ is appended to $t_D$. The algorithm defined in this manner produces the disturbance intervals $t_O + t_D$ as illustrated in the lower panels of Figure 1 for the $X$ (left) and the $Y$ (right) components. Please note, that the intervals for both components are the same. This method is intended to identify strong events for which the likelihood is high, that the filter baseline needs to be replaced for both components.

### 2.2.2. Identification of Weak to Moderate Magnetic Disturbances

Even for weak to moderate events, directly removing our filter baseline from magnetic measurements may lead to underestimation of the disturbance's strength and the disturbance itself will still be partly present within the corresponding residuals which are defined as.

$$resX = X - x_{FB} \tag{5}$$

$$resY = Y - y_{FB}. \tag{6}$$

The featured elevated amplitudes will deviate from a general quiet level. Assuming that a significant increase in amplitude marks disturbance-time and that occurrences of increased amplitudes are less frequent than the general background/quiet level we can treat disturbance amplitudes like outliers. With the methodology of sigma-clipping (Lehmann, 2006), outliers can be characterized by their associated $i\sigma$th confidence interval with $i = 1, 2, 3\ldots$ In this way, upper and lower boundaries can be defined as $b_{U/L} = \mu \pm i\sigma$, where $\mu$ is the mean and $\sigma$ the standard deviation of the data.

To account for possible variabilities within the background level of the residuals, we use the running mean $\mu$ and running standard deviation $\sigma$ over a sliding window of the past 6 days. Six days will account for any long-term effects from solar cycle phases and seasons, but also account for periods which are very disturbed due to a strong event, increasing background levels for several days even after the end of such an event. Vennerstrom et al. (2016) investigated more than 2000 geomagnetic storms for which the longest duration was 93 hr Haines et al. (2019) investigated a similar set of storms and found the longest duration to be 75 hr. Using a sliding window of 6 days thus allows to include the full storm while ensuring sufficient non-storm data points for the statistics. Still, disturbances are associated with a larger standard deviation which holds the risk that beginnings of and

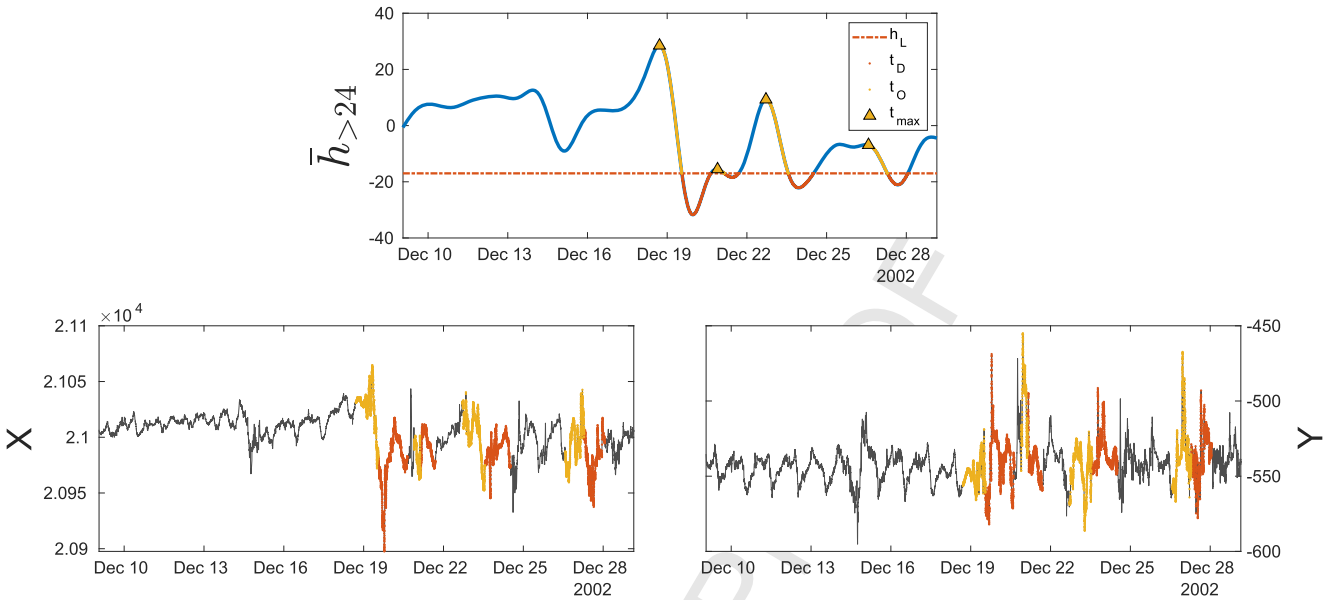


**Figure 1.** Storm detection with the long-term filter responses: Illustration of the algorithm to detect strong storms with the de-trended horizontal intensity $\bar{h}_{>24}$ from Section 2.2.1 in the upper panel (orange intervals correspond to variation below $h_L$ threshold and yellow intervals corresponds to the appended interval from the previous maximum $t_{max}$). The lower panels show the corresponding identified intervals on the $X$ (left) and the $Y$ (right) components. All vertical axes are in nanotesla.

subsequent disturbances are hidden and remain unnoticed. The nature of the residuals is to oscillate from negative to positive values crossing the zero-line in-between. Defining upper and lower boundaries enable the identification of single outliers, but misses values that occur when the residuals fluctuate from a positive to a negative outlier and vice-versa. As we are interested in full disturbances only, full time intervals within the filter baseline need to be replaced and not isolated minutes. To overcome these limitations from the sigma-clipping, we introduce a 2-steps algorithm. As disturbances can manifest themselves differently in $X$ and Y components this algorithm is applied to each of them individually.

In step 1 we determine an appropriate boundary that identifies outliers due to disturbances and in step 2 we combine the singular outliers to identify actual disturbance intervals. Supporting illustrations for these steps are found in Figure 2, showing the $X$ component on the left and the $Y$ component on the right.

- **Step 1: Determining quiet levels**

  The sigma-clipping enables the definition of quiet level boundaries. As pointed out, the $1\sigma$ interval is too constrained and disturbances may be missed due to the featured boundary broadening. Therefore we will adapt the quiet level boundary by combining a loose and a tight boundaries, $b_L$, $b_T$.

  – **Step 1.1: Definition of loose and tight boundaries**

  To allow for a more relaxed border, we apply the sigma-clipping with $3\sigma$ on each of the residuals res$X$ and res$Y$, that is, the boundaries are defined by the moving average $\mu_1$ and moving standard deviation $\sigma_1$ over the past 6 days: $b_1 = \mu_1 \pm 3\sigma_1$. This boundary $b_1$ is derived on, and thus, contains influences of very strong outliers due to the high confidence interval of 3. We re-calculate the moving average $\mu_2$ and moving standard deviation $\sigma_2$ only on the $3\sigma$ constrained values that is, for $-b_1 < \text{res} < b_1$, providing us with the loose boundary $b_L = \mu_2 + \sigma_2$.

  For the tight boundary, we remove once more the outliers from the data, that is, $-b_L < \text{res} < b_L$ and generate the weighted, moving mean $\hat{\mu}$ and weighted standard deviation $\hat{\sigma}$. The weight function is half a cosine which, for $t = t_0$, is 0 at $t = t_0 - 6d$ and 1 at $t = t_0$. This ensures that disturbances that are farther in the past influence the tight boundary less than disturbances that are closer in time. The tight boundary then writes as $b_T = \hat{\mu} \pm 3\hat{\sigma}$. Both boundaries ($b_L$ in yellow and $b_T$ in purple) are indicated in panels (a) in Figure 2.

  – **Step 1.2 Combination of $b_L$ and $b_T$**

  By default, we assume quiet time and apply the loose boundary $b_L$. As soon as a value above $b_L$ is detected, the tight boundary is applied for the successive 12 hr. The resulting quiet level boundary $b$ is

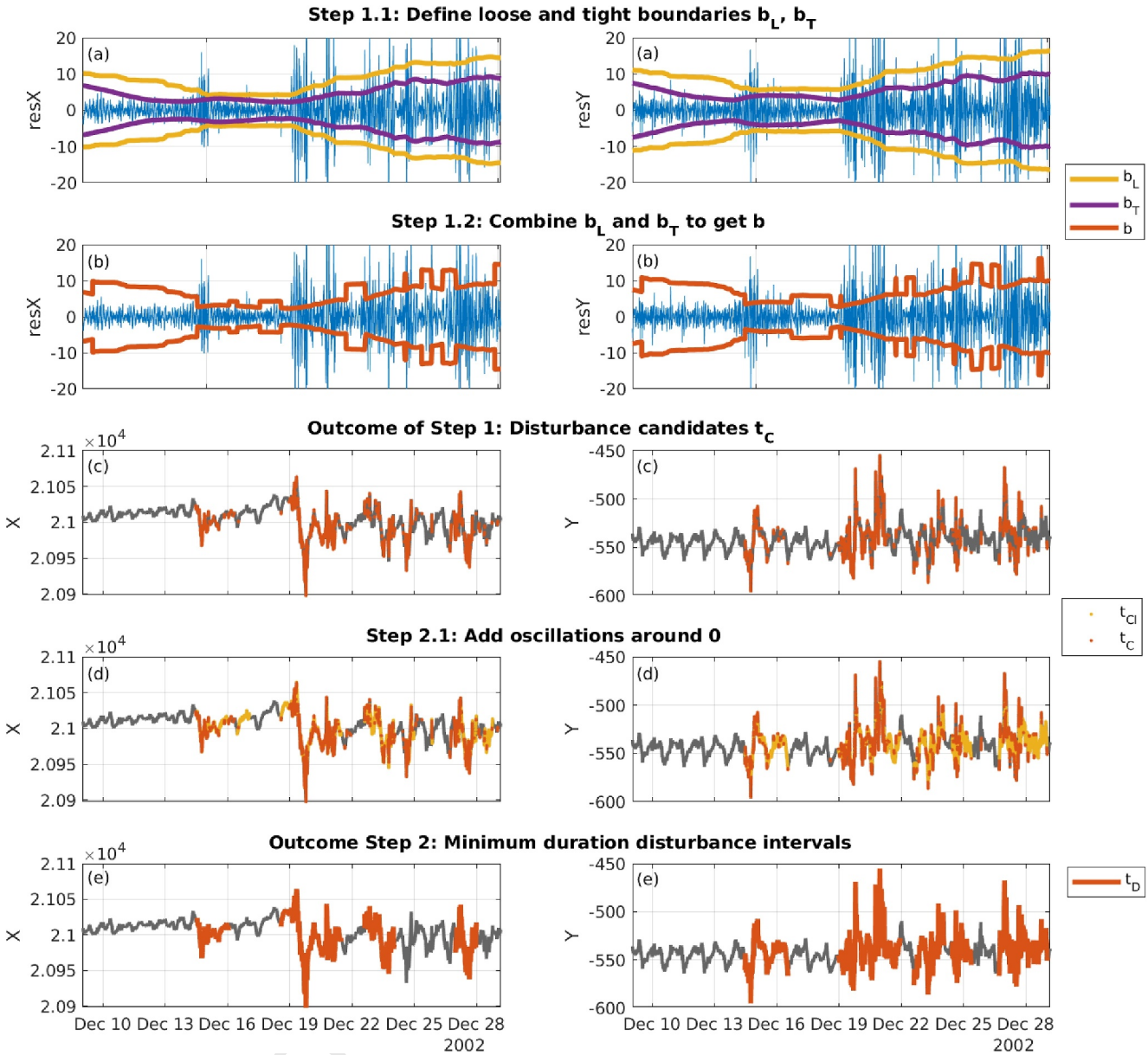


**Figure 2.** Disturbance Detection using the residuals: Visualization of the two-step algorithm that is used to identify disturbance intervals for the residuals of $X$ on the left and $Y$ on the right. Residuals are in blue and components are in black. Panels (a) show the loose and tight boundaries $b_L$ and $b_T$ in yellow and violet, respectively, as defined in step 1.1. Panels (b) show the final upper and lower boundaries $b$ (in orange) as defined by the combination of $b_L$ and $b_T$ in step 1.2. Panels (c) show the identified disturbance candidate minutes $t_C$ (in orange dots) on the component (black) as outcome of step 1. Panels (d) indicate the added minutes $t_{CI}$ (in yellow dots) during the residuals' oscillations around zero, as defined in step 2.1. Panels (e) show the outcome of Step 2, and thus the final disturbance intervals $t_D$ (in orange), after removing intervals with lengths below the minimum duration as defined in step 2.2. All vertical axes are in nanotesla.

depicted in panels (b) of Figure 2 in orange. Each residual that is outside of the upper and lower $b$ boundaries is considered a disturbance candidate minute $t_C$. The resulting candidates on each of the components are indicated as orange dots labeled $t_C$ in panels (c) of Figure 2.

- **Step 2: Defining disturbance intervals** The outcome of step 1 leaves us with an irregular collection of storm candidate minutes $t_C$. Unsurprisingly and from panels (c) in Figure 2, it is evident that occurrences of $t_C$ are clustered. We will first create full disturbance candidate intervals by adding values that belong to the same disturbance, that is, when the values oscillate in between the upper and lower quiet level boundaries $b$. We then check if these intervals belong to an actual disturbance by defining the minimum disturbance duration $\Delta t_{\min}$.

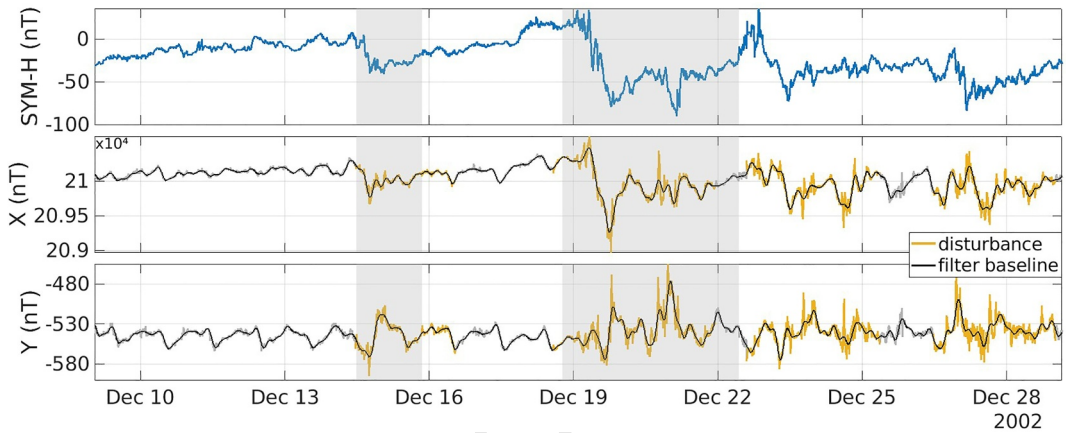


**Figure 3.** Full disturbance detection: The detected disturbance intervals are indicated by the X and Y components in yellow, otherwise in gray. The filter baseline is in black. The first panel indicates the SYM-H index and gray shaded areas correspond to the SYM-H definition of the moderate and strong events respectively.

– **Step 2.1 Incorporation of oscillations**

The typical disturbance event induces residuals that fluctuate from above to below the quiet boundaries $b$, implying that values within the quiet boundaries are missed although they belong to the same event. Such that, during the event's duration $\Delta t$, the values inside $|b|$ should be incorporated within disturbance candidates. Therefore, we define maximum duration $\Delta t_{CI}$ that relates two disturbance candidate minutes to the same event $t_{Ci}$ to $t_{Cj}$. If the duration $\Delta t$ between $t_{Ci}$ and $t_{Cj}$ is less than $\Delta t_{CI}$ then $t_{Ci}$ and $t_{Cj}$ are considered to belong to the same event. All minutes $t_{CI}$ (shown in panels (d) in yellow) within $\Delta t$ are then added to the storm candidates, resulting in full disturbance candidate intervals (or events). By empirical refinement, we determined $\Delta t_{CI} = 8hours$.

– **Step 2.2 Final Disturbance intervals**

It is possible that the method of using a maximum $\Delta t_{CI}$ between two disturbance candidates marks quiet periods. This is the case for for example, the interval around sixteenth of December on the $X$ component in panel (d) right after the moderate event. For the identification of such intervals, we define a minimum disturbance time $\Delta t_{\min} = 15h$. Any disturbance candidate interval that is longer in duration than $\Delta t_{\min}$ is considered a final disturbance interval, otherwise it is removed. The time of 15hr corresponds to a typical duration of shorter storms and agrees well with empirical refinement.

The final disturbance intervals $t_D$ produced by this algorithm are illustrated in panels (e) for the $X$ component (left) and the $Y$ component (right) in Figure 2. Please note that the marked intervals are not necessarily identical for both components and do not need to be geoeffective disturbances as is the case for the detection with the long-term filters. Rather this method identifies intervals for which the filter baselines show consistent discrepancies, not following quiet variations for longer periods.

### 2.2.3. Full Storm and Disturbance Detection

Creating the union of the disturbance detection using the long-term filter response from 2.2.1 and using the residuals from both components from 2.2.2 holds the final disturbance interval detection. This approach also guarantees that disturbance intervals are the same for each of the two components $X$ and $Y$. The full detection is illustrated in Figure 3 in yellow for December 2002 for the magnetic field measurements $X$ and Y at CLF (in gray if not detected) together with the filter baseline in black. The SYM-H index during this period is indicated in the first panel and the moderate and strong events as defined by SYM-H are indicated with gray shaded areas on all panels.

In this figure, there is a quiet period during the first days, also clear in SYM-H values in the first panel, for which the filter baseline follows the smooth $X$ and $Y$ variations as expected. Between thirteenth and sixteenth of December, the moderate event occurs which effects would be underestimated by removing our initial filter baseline. This event is detected by both residuals (see panels (e) in Figure 2), but is too weak for detection within

the long-term filter (see Figure 1). The start of the detection is aligned with the SYM-H storm definition of the event but the presented detection is longer in duration. Afterward, another rather quiet period of a few days is observed until the second, stronger event occurs on December 19th and disturbs the components significantly for several days. For the onset and main phase of this storm, it is clear that using our filter baseline would result in underestimation of its strength. The start of the event is captured, while the presented detection ends earlier than the SYM-H storm definition suggests. From this figure it is also clear that disturbances in local geomagnetic field measurements do not necessarily overlap with storm onsets and duration defined by global magnetic indices like the SYM-H. This is due to the fact that these indices are a combination of various observatories indicating a global, planetary state of the geomagnetic field. The strong event is followed by a rather disturbed phase for which no distinct quiet patterns can be recognised within the X,Y components. The disturbance detection however identifies two intervals which align with observable storm signatures which are on top of a general elevated activity level in SYM-H. This indicates that local effects which do not follow typical storm signatures are caught with the introduced method. For a 5-months period between August 2002 to January 2003, a total of 26 disturbance events were detected at CLF. Of these, 1 event has a duration lower than 1 day; 9 events have a duration between 1 and 2 days; 7 events between 2 and 3 days; 4 events between 3 and 4 days; 3 events between 4 and 5 days; 1 event has a duration of 5–6 days and the single longest event has a duration of over 9 days.

### 2.3. Quiet Variations During Storm and Disturbance Time

The disturbance detection algorithms for the $X$ and $Y$ components provides intervals during which the filter baseline follows the disturbance induced variations too closely and a replacement in the form of a substitution baseline during these periods is sought after. There is no quantitative ground truth on how quiet variations are supposed to look like during disturbances and storm-time. The very least we can, however, impose on the substitution, is that it incorporates typical quiet variations. As has been done in the definition of the filter baseline (Equation 1) and (Equation 2), we split the substitution baseline into a long-term and a daily part.

For the substitution of the long-term filter response, we use the moving average over the past 3 days in accordance with the filter window-size and denote it $\tilde{x}_{>24}$ and $\tilde{y}_{>24}$, respectively. The moving average window size of 3 days smooths disturbance variations in the long-term filters while preserving local trends that can be induced by disturbances that otherwise would be concealed in longer moving windows.

For the daily variations $x_D$ and $y_D$ we would like to have typical solar quiet current system footprints within the substitution. One possibility would be to use existing Sq models. However, any (statistical) model may not be fully adapted for current local geomagnetic conditions. Another possibility is to use surrounding quiet days to extrapolate quiet variations for the time-interval in question as has been done by Kamp (2013). In order to keep the real-time aspect, however, we cannot use quiet data that lie in the future. At any given timestep, the latest information on the Sq current system can be found within the variations in $x_D$ and $y_D$ from the quiet day/period before. In the following we look at two approaches on how to derive possible daily variations during the disturbance interval.

#### 2.3.1. Using (Sub-) Diurnal Filter Responses From Previous Quiet Periods

Leveraging each of the (sub-) diurnal filter responses $x_P, y_P$ with $P = 24, 12, 8, 6$, we can use their values from previous quiet periods and extrapolate them into the future. This means that the signal is time-shifted from the past to represent the theoretical baseline during the identified disturbance period. The used time-shift $\delta t$ hereby depends upon the period $P$ of the (sub-) diurnal filter. Concretely, this means that for the diurnal filter $\delta t$ is 24 hr and thus the signal is time-shifted into the future for the next 24 hr. Analogously, the semi-diurnal filter reconstruction is time-shifted 12 hr into the future. The same for the 8 and 6 h filters. Such that, at time $t$ and filter with period $P$ the reconstructions per (sub-) diurnal filter per component are defined as.

$$\hat{x}_{P,shift}(t) = x_P(t - P), \text{for X} \tag{7}$$

$$\hat{y}_{P,shift}(t) = y_P(t - P), \text{for Y} \tag{8}$$

Their superposition defines the baseline substitution with forward propagation and writes as.

$$\hat{x}_{D,shift} = \hat{x}_{24,shift} + \hat{x}_{12,shift} + \hat{x}_{8,shift} + \hat{x}_{6,shift} \quad (9)$$

$$\hat{y}_{D,shift} = \hat{y}_{24,shift} + \hat{y}_{12,shift} + \hat{y}_{8,shift} + \hat{y}_{6,shift}. \quad (10)$$

This approach has the important limitation that at time-step $t_0$, the substitution baseline can only be determined for the next 6 hr.

Assuming that a disturbance starts at $t_0$, after $t_0$ + 6h this substitution would be generated on disturbed information, namely the disturbance-affected 6h-filter response $x_6$. To avoid this issue, we do not use the interval directly before the start of the disturbance interval, but use the baseline data from 2 days before the start of the disturbance $t_0$, that is, $t_0$ − 48h to $t_0$ − 24h. Additionally, this allows avoiding any possible storm onsets that were missed by the detection method described in the previous chapter. Note that the baseline data from 48 to 24hr before the start of a disturbance $D_0$ at time-step $t_0$ may contain a disturbance $D_1$ itself. In that case, the filter baseline has been already been replaced for $D_1$ by data 48-24 hr before $D_1$ started. Thus, the substitution for $D_0$ is build upon an already substituted baseline and contains only quiet variations. In this way, as long as the replacement of the very first detected disturbance has been done on quiet data, and such a situation is always possible to find by going back sufficiently long, the baseline data used for replacement contains quiet variations independently whether a disturbance is present or not during the time-window of 48–24hr before.

Eventually, this gives us a 24-hr template with potential quiet day variations. As disturbances can last longer than a day, we repeat this template and extend it to fit the full disturbance length. The obtained replacement for the combined (sub-) diurnal variations with the shift model is denoted as $\tilde{x}_{D,shift}$ and $\tilde{y}_{D,shift}$.

### 2.3.2. Using Daily Variations From Previous Quiet Periods

We know that the daily filter responses $x_D$ and $y_D$ follow the quiet variations well during quiet periods. A second possibility to replace them during disturbances is thus to use their variations from a previous, quiet interval. We can derive such intervals as we have done for the shift model by taking the variations from the previous 48–24hr as a 24-hr template and repeat it to fit the length of the disturbance similarly as in the previous subsection. We denote the direct quiet variation substitution baseline as.

$$\tilde{x}_D(t) = x_D(t - 24h) \quad (11)$$

$$\tilde{y}_D(t) = y_D(t - 24h). \quad (12)$$

### 2.3.3. Comparison of Quiet Variation Replacements

Using the superposition of the long-term filter replacement and the two daily filter replacements, we end up with two possible substitution baselines which write as, for the $X$ component.

$$\tilde{x}_{B,shift} = \tilde{x}_{>24} + \tilde{x}_{D,shift} \quad (13)$$

$$\tilde{x}_B = \tilde{x}_{>24} + \tilde{x}_D, \quad (14)$$

and for the $Y$ component.

$$\tilde{y}_{B,shift} = \tilde{y}_{>24} + \tilde{y}_{D,shift} \quad (15)$$

$$\tilde{y}_B = \tilde{y}_{>24} + \tilde{y}_D. \quad (16)$$

To avoid discontinuities at the end points of the disturbance intervals, we apply a cubic spline at 3 hr before the start and 3 hr after the end of the disturbance interval to plug in the substitutions. Figure 4 illustrates the resulting geomagnetic baselines. Note here the smooth transition between the filter baseline and substituted interval due to the spline. Overall both replacements are quite similar and differ mainly on smaller scales. The differences manifest in small-scale ripples within the shift model $\tilde{x}_{B,shift}$ and $\tilde{y}_{B,shift}$ (in orange) compared to the smoother $\tilde{x}_B$ and $\tilde{y}_B$ (in green). For the moderate event between thirteenth to 16th December both replacements show smooth

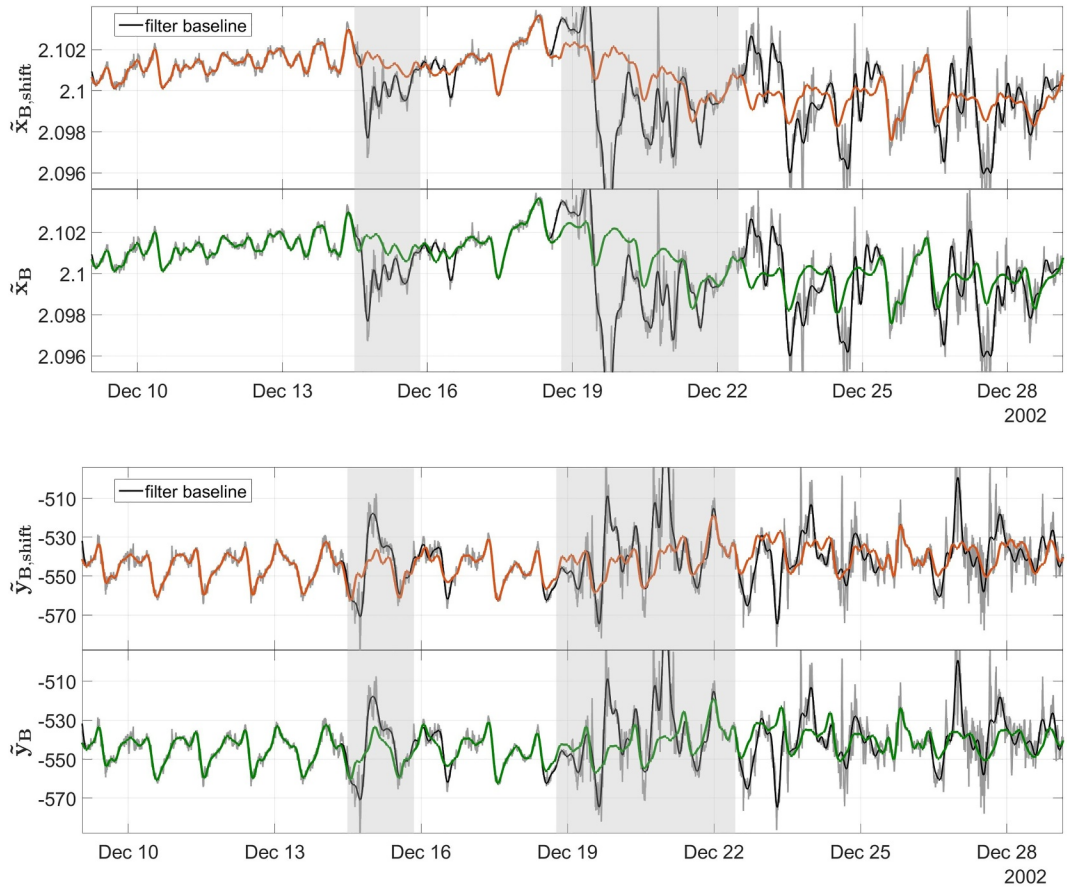


**Figure 4.** Quiet variation replacement comparison: Comparison between the two baseline replacements for the X (upper panels) and Y (lower panels) components. The components are in gray, the filter baseline in black. The baseline substitution with forward propagation $\tilde{x}_{B,shift}, \tilde{y}_{B,shift}$ are in orange and the direct quiet variation substitution baseline $\tilde{x}_B, \tilde{y}_B$ are in green. The gray shaded areas correspond to the SYM-H definition of the moderate and strong events respectively. The vertical axes are in nanoteslas.

variations. The increase before the event is not fully detected, such that the peak in $X$ is preserved and will not be fully present in the residuals (after removing each baseline from the measurements). For the intense event between 19th to 22nd December, the variations are also smooth for both replacements and we see that the general trend of both components, especially the one of $X$, are systematically followed. As for the rather disturbed period following the strong event, the baseline is replaced according to the identified intervals and smooth, recurrent patterns are found within the baseline replacements. Based on these discussions, we can make the qualitative choice of using $\tilde{x}_D$ and $\tilde{y}_D$ as they incorporate less small-scale features which are not necessarily present in the typical quiet variations. Further examples of this final baseline for further magnetic observatories are discussed in Section 4.

## 3. Comparison With Existing Baselines

For the validation of geomagnetic baselines no ground truth is available, especially not during storm-time. Kamp (2013) and Gjerloev (2012) use a spectral analysis to validate that only sources acting on relevant frequencies are contained within the baseline. Haberle et al. (2022) showed that quiet sources like the solar quiet current system and the secular variation are imprinted within the filter baseline. Indeed, as the filter baseline is based on frequency regimes, only the fast variations below 6 hr are part of the residuals. All the other, above 6 hr variations, are retained within the baseline. Another option to validate geomagnetic baselines is to compare them to existing ones, as has been done in Kamp (2013). We compare the final baseline with the original filter baseline, the SuperMAG one from Gjerloev (2012) and the FMI one from Sucksdorff et al. (1991). The method from

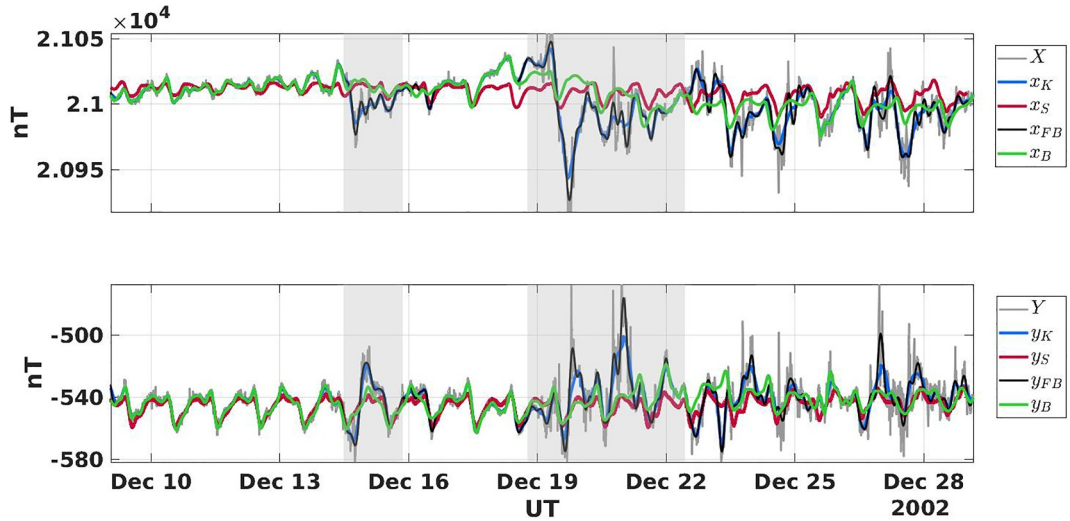


**Figure 5.** Comparison of baseline methods: The methods of the introduced baseline (green), the filter baseline (black), FMI (blue) and SuperMAG (red) for $X$, $Y$ components (gray) at CLF during winter 2002 are presented. The gray shaded areas indicate the moderate and longer, strong event. All vertical axes are in nanotesla.

Kamp (2013) and the one for the PC indices from Troshichev and Janzhura (2012) are designed for auroral and polar observatories and thus are considered out of range for mid-latitudes studied in this work.

Hereafter we compare our method to the filter baseline $x_{FB}, y_{FB}$, FMI method, referred to as $x_K$, $y_K$; and to the SuperMAG method, referred to as $x_S$, $y_S$. As such, Figure 5 demonstrates our baseline (green) in comparison with the $X$, $Y$ components (gray), the original filter baseline (black), the FMI (blue) and the SuperMAG (red) baselines, whereby the gray shaded time intervals indicate the moderate and strong events as seen by SYM-H, respectively. During magnetically quiet periods, the FMI filter and our introduced method produce very similar baselines, while the SuperMAG one differs from them. A detailed discussion during quiet periods is treated in Haberle et al. (2022), including the details on the reconstruction of FMI and SuperMAG baselines for this work, and will not be repeated here.

The greatest differences between the methods occur during disturbed times. Two types of baselines can be readily distinguished: the first one that follows storm activity rather closely (FMI and original filter baseline, respectively blue and black) and the second one that follows quiet variations from previous activity levels during storms (SuperMAG and our new introduced baseline, respectively red and green). For the first type, the main risk is that storm effects are underestimated as they react to the strong deflection. In Figure 5 this is clearly visible during both of the indicated storm events.

For the second type, during the moderate event, the new introduced $x_B, y_B$ and SuperMAG $x_S, y_S$ baselines are very similar showing minor differences. The main differences for this event are found in the $X$ component that shows an increase just before the SYM-H defined start. This increase is closely followed by $x_B$ while $x_S$ depicts a smaller amplitude. It cannot be ruled out that this increase is connected to the actual storm activity (storm onset) and in this case, the introduced methodology would miss this part.

During the strong event, even these two baseline methods differ significantly. During the initial phase $x_B$ starts from a higher background level than $x_S$ and stays higher until the middle of the recovery phase where they are the most similar. Afterward, $x_S$ shows higher levels than $x_B$. The offset before the event can be related to the observed increasing activity before the event, which is not recognised by $x_S$, while $x_B$ still includes this activity. For $Y$, both methods are similar until the middle of the recovery phase where they start to diverge, as $y_B$ follows the observed activity closely again. This can be related to the disturbance detection algorithm, see Figure 3. As noted there, it is not surprising that using the SYM-H index to define start and ending times of geomagnetic storms does not necessarily align with locally observed variations.

The period after the strong event is distinguished by strong variability that persists several days after the event ended. While these variations are barely imprinted in the SuperMAG baseline $x_S, y_S$, the final baseline $x_B, y_B$

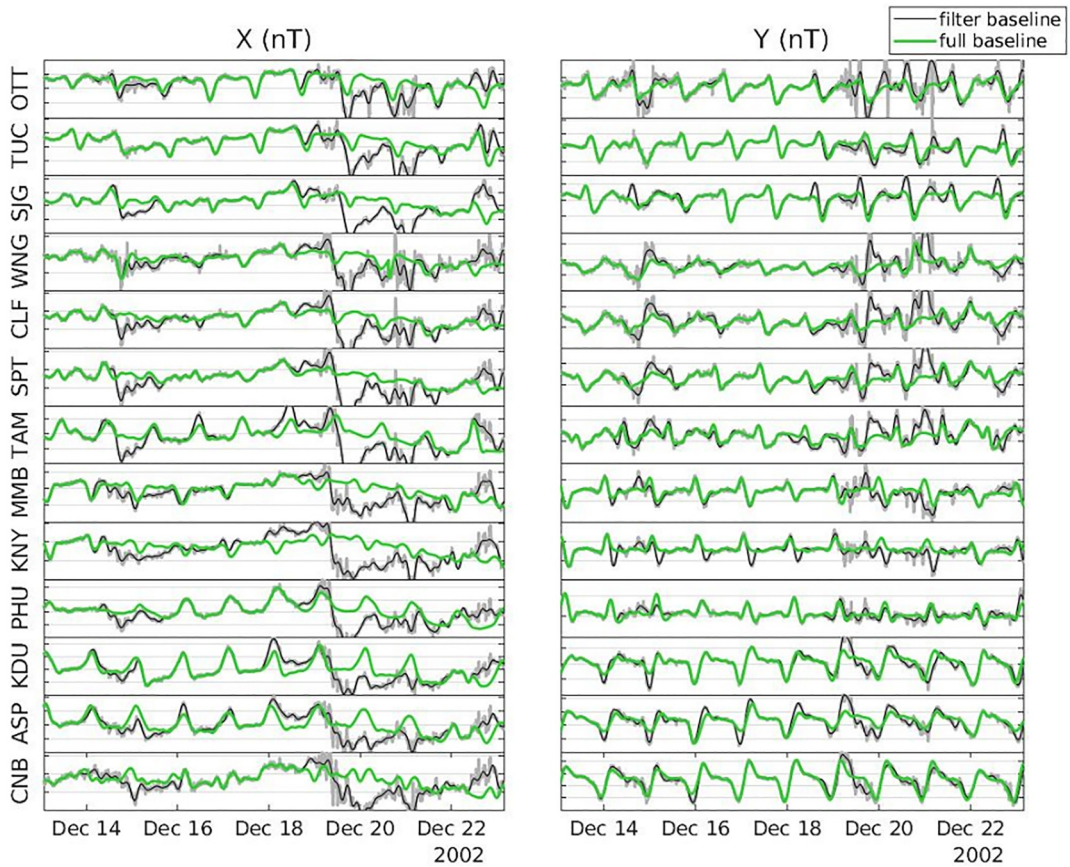


**Figure 6.** Final Baseline Event Study: Final geomagnetic baselines in green for further magnetic observatories located in geographic sectors North America (OTT, TUC, SJG), Europe (WNG, CLF, SPT), Africa (TAM), Asia (MMB, KNY, PHU) and Australia (KDU, ASP, CNB) during moderate and strong geomagnetic storms. The measurements $X, Y$ are shown in gray and the filter baseline in black. The abscissas represent time in UT days, ordinates are in nanotesla with 30 nT between consecutive tickmarks.

follows variations except when the storm detection algorithm flags a time-interval and quiet variations are used instead.

## 4. Final Baseline Event Study

We now present and discuss the final baseline for further observatories than CLF. Figure 6 shows the final baseline (green), together with measurements $X, Y$ (gray) and the original filter baseline (black) for 13 magnetic observatories. Their geographic locations are indicated in Figure 7 and given in table S1 in Supporting Information S1. For visualization purposes the time-interval is shortened with respect to previous figures as we want to highlight method performances during the moderate and strong events.

The introduced geomagnetic baseline follows the magnetic field variations closely during non-disturbed periods and produces plausible daily variations in general for all stations. A closer analysis of Figure 6 verifies that the effects of disturbances are a local property. Though similar, for each station the filter baseline is replaced at different points in time. This is to be expected, due to spatio-temporal variations of geomagnetic storm impacts in longitude and local factors such as ground conductivity (Gough, 1973). Especially at Tucson (TUC), the measurements do not show significant variations during the moderate event such that the filter baseline has not been replaced. This is mainly the case when the event was not intense enough to create a significant variance within residuals. For the $X$ component of TUC, the diurnal pattern is slightly distorted by the event, which is not accompanied by a significant amplitude increase. For its Y component, the pattern is marginally distorted, and the filter baseline still produces credible variations. The strong event was detected and thus replaced with quiet

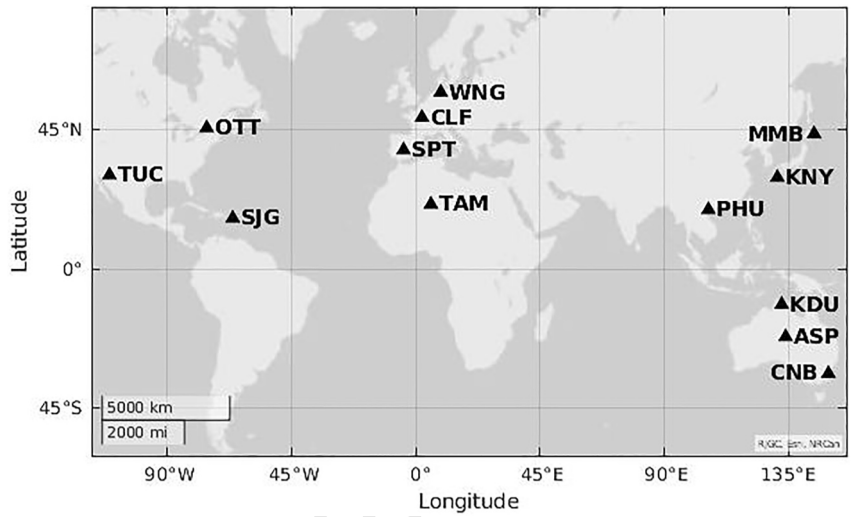


**Figure 7.** Geographic locations of magnetic observatories used in Figure 6.

variations for all stations. The $X$ components show stronger variations than $Y$ in general. Such that it can happen that the filter baseline in $Y$ is replaced even though it seemingly follows quiet patterns (e.g., at TUC, SJG, PHU, CNB). The replaced quiet variations, however, do not differ considerably which shows that the quiet variation replacement choice is robust.

Globally, daily variations are well captured during quiet periods and well reproduced during disturbed periods. Thus, we can conclude that the introduced algorithm is doing a functional job in producing appropriate geomagnetic baselines.

## 5. Discussion

In this section, we will first discuss advantages and considerations of the methodology proposed to detect geomagnetic disturbances and to replace the baseline. Then we will present the global implications for magnetic indices derivation and real-time application.

### 5.1. Detection of Geomagnetic Disturbances in the Filter Baseline

Considering the followed strategy for replacing the filter baseline, the first challenge lays within the detection of disturbances. For the identification of strong events, a limit $h_L$ needs to be derived for each station independently. Investigations show that $\bar{h}_{>24}$ is well linearly correlated with the SYM-H index at stations in mid-latitudes. As this first step of the detection is supposed to mark strong events, we choose for all stations to systematically use a SYM-H level that is commonly associated with strong events, meaning SYM-H= −50 nT. As is exemplified for CLF in Section 2.2.1 and in the supplementary material, this choice seems rather appropriate. With the determination of this threshold, the linear relationship can be used to derive such a limit per station. As much as SYM-H has been thoroughly investigated as a reliable indicator of magnetic activity in the past, it still implies that we accept to rely on external information (another index) to determine the corresponding threshold at each magnetic station.

Finally, in Section 2.2.1, we proposed an 11-year running mean with an update at each time step to actualize the $h_L$ threshold. However, we don't expect this threshold to change radically over a period of a few months or even years, so it's possible to reduce the frequency of this update. Thus, once the $h_L$ threshold is defined, it may be applied for periods over a few years but then would need re-calibration as the secular variation moves the station to different magnetic latitude, possibly changing its Sq pattern. Such a re-calibration, however, does not impact the near real-time operationality and can even be updated every year or even less for sake of precision.

For weak to moderate events, we introduced the sigma-clipping of the residuals. This method is based on the premise that quiet variations make up the majority of the signal and that storm variations are stronger and occur

less frequently. While this is certainly true during solar minimum, we cannot guarantee this holds during solar maximum when external driving is significantly increased and high level of perturbations could last for several days, naturally enlarging the clipping thresholds. This is also the reason why it is important to still use the long-term filter to identify strong events. By deriving upper and lower boundaries for the signal residuals, we also had to overcome the issue of the zero oscillations. We did so by defining two durations: a minimum duration for the effects of a disturbance $\Delta t_{CI}$ and a minimum duration of a disturbance to be considered $\Delta t_{\min}$. This implies that any disturbance that manifests itself beneath 15hr is missed, as well as events that cause zero-oscillations with recurrences above 8 h. Whereby the latter case seems to be less likely to happen, the minimum disturbance time puts a real restriction on the identified intervals. This should be kept in mind. The adapted sigma-clipping thus works as long as residuals react clearly to disturbances which is largely guaranteed by the signal filtering itself. Moreover, the free parameters ($\Delta t_{CI}$, $\Delta t_{\min}$) were empirically determined by an educated guess informed by manual investigation and discussion of the events detected. In order to confirm this, we chose to derive key performance indicators (KPIs) which were used in a systematic parameter search to identify the best values. Among others, these KPIs included the detection of pre-selected events. Although not shown in this paper, a thorough discussion on these is available within the doctoral thesis of Haberle (2023). Finally, the residuals are generated by removing all variations above 6 hr from the measurements. This means, that all faster variations, below 6 h, are contained within the residuals and will be considered as perturbations whatever their origin. This two-step detection method has been thoroughly checked on several multiple stations and is working relatively well, although we saw that onsets of weak to moderate storms are not always fully reflected within the residuals and may then be missed. Nevertheless, the disturbance determination as introduced here is at least consistent as it is derived on objective rules that treat each data point the same way. Additionally, its derivation is clearly presented and can directly be applied to any magnetic field measurements independent of location.

### 5.2. Quiet Magnetic Variation Replacement During Disturbances

Here, we adapt the filter baseline only during non-quiet times, the capturing of quiet sources within the geomagnetic baseline still holds true outside perturbations. The used replacement during disturbances is based on variations from 2 days before the start (to avoid introducing perturbations coming from the non-detected onset) and then duplicated for the duration of the disturbance. Thus during magnetically quiet periods, the introduced baseline follows a smooth curve as continuity is an intrinsic property of the applied temporal filter technique (see Haberle et al. (2022) for further information) and during storm-time, when the filter baseline is replaced, the algorithm uses interpolation to make sure the transition is smooth. Additionally, Figures 4–6 demonstrate that the substitution during storm-time does not affect the smooth curve criteria. The possible replacement duplication for several days also implies that the baseline assumes the same daily variations during the entire disturbance duration. We have seen that there is an intrinsic day-to-day variability within quiet variations. Such a variability is not taken into account by the introduced replacement. Fair enough, as there is no knowledge of a baseline during storm-time, we may choose to leave out day-to-day variability during storm-time. Such an approach has been also followed by Gjerloev (2012) and Kamp (2013). Additionally, Kamp (2013) improved his templates by using variations from before and after the event. For an operational, real-time determination, we have no information about conditions after a disturbance event and cannot apply similar strategies. However, for post-processing such methods should be considered.

### 5.3. Future Validation of the Filter Baseline

The final validation of an appropriate baseline presents a significant challenge due to the absence of a definitive ground truth. This may have profound implication for subsequent magnetic indices derivation, as well as (near) real-time computation. This problem is already introduced in Section 2.3 and was originally motivated by the desire to improve the new generation of magnetic indices such as in Chambodut et al. (2015). In Figure 5, our new adapted filter baseline $x_B$, $y_B$ and the SuperMAG baseline $x_S$, $y_S$ during the disturbed days are very similar to their quiet curve of the preceding day, which can be interpreted physically as a fully developed Sq current cell. Such a full system does not necessarily form during a disturbance event (Blanc & Richmond, 1980; Huy & Amory-Mazaudier, 2008), however, as the corresponding baseline is then unknown, our choice to keep a quiet pattern similar to surrounding days may be considered as a valid possibility, especially when looking at figure S1 in Supporting Information S1 for CLF. In order to characterize the impact of our baseline derivation and of our choice of baseline replacement during disturbance-time, the calculation of the residuals, followed by new

magnetic indices with different baselines, including the ones presented in this work, may be helpful. The thorough comparison of these different baselines during quiet periods and various strong and moderate events, ideally covering all solar cycle phases and seasons, will help to evaluate quantitatively the discrepancies and to document the strengths and weaknesses of our new baseline for such future magnetic indices derivation. This is left for future study.

When we first considered a new method for baseline derivation based on finite impulse response (FIR) filters (Haberle et al., 2022), a few of our main objectives were to enable a computation solely based on the magnetic measurements themselves and easily applicable in real-time. Unfortunately, it is impossible to fully derive a baseline without a priori information related to the secular variation of the geomagnetic field. For example, the SuperMAG and FMI methods need the magnetic latitude as an input, to produce either the local magnetic field orientation or the K9 lower limit, which differs for each magnetic observatory and is evolving over time according to secular variation. Moreover, the SuperMAG rigid method suppresses also all day-to-day variability of the baseline. Although we also need additional information in the calculation of our filter baseline, we have tried to keep it to a minimum. Indeed, we only need one identical parameter for all stations, namely SYM-H, which is used to derive the $h_L$ thresholds at each station via a very simple analytical regression as a pre-treatment that can be applied once a year or even less frequent, meeting standards of quasi-definitive magnetic observatory data (Clarke et al., 2013). After this pre-treatment, the filtering method together with the replacement during disturbances is able to produce baselines without any further information than the magnetic measurements themselves at all mid-latitude stations. This property gives the introduced method the main advantage of being directly applicable in operational settings where (near) real-time applications are a requirement, since the baseline can be computed as soon as the geomagnetic field data is available.

## 6. Conclusion and Further Steps

This paper continues the work from Haberle et al. (2022) of deriving geomagnetic baselines to characterize the geomagnetic field with enhanced temporal resolution by improving the filter baseline with the goal of properly determining the amplitude of space weather events and other disturbances and replacing them with a quiet counterpart.

In order to remove storm contributions from the filter baseline a two-step approach is applied. The first step is to detect storm and disturbance intervals for which the filter baseline performs poorly. The output of this step are identified time intervals that contain storm and disturbance information within magnetic field measurements. These intervals are characterized by increased geomagnetic activity and the likelihood for the filter baseline following its effects too closely is increased. The second step is dedicated to quiet variations that can be used instead of the filter baseline during identified disturbances. Ideally such variations are consistent with patterns preceding and following the corresponding quiet periods.

To validate the final baseline it is compared to other methods, showing overall good performance, following quiet variations well, while representing them adequately during detected disturbances. Additionally the final baseline is presented and evaluated during moderate and strong space weather events for 13 magnetic observatories distributed globally.

This leads to the conclusion that the introduced geomagnetic baseline can effectively be used to determine the storm influences within high-quality geomagnetic field measurements of observatories located in mid-latitudes and is suitable for future derivation of new magnetic indices with improved temporal resolution. An essential point of this baseline is that the introduced procedures can be executed in near real-time as it works solely on the filter/magnetic data itself. Further input (the SYM-H index) is only required to derive thresholds which can be done for several months in advance. As such, the introduced geomagnetic baseline can be leveraged for purposes of characterizing the geomagnetic field, including the derivation of magnetic indices in operational settings.

One of the main challenges encountered during this work is the absence of a ground truth for validation of geomagnetic baselines. One important aspect of validating geomagnetic baselines is to quantitatively show the contained sources, as has been done in Haberle et al. (2022). The other one is the comparison with existing methodologies and presentation for further data sets/observatories. Future validation strategies in the form of validation scores or key performance indicators (KPIs) including event lists may prove supportive in the future.

## Data Availability Statement

The magnetic observatory data are available from BCMT and INTERMAGNET data repositories (BCMT, 1921; INTERMAGNET, 2024). The magnetic activity indices are available from WDC Kyoto and ISGI data repositories (ISGI, 1906; WDCKyoto, 2022).

**Acknowledgments**
The authors would like to thank the reviewers for their meaningful and constructive feedback which helped the improvement of this manuscript. The authors would like to thank Roman Leonhardt for his valuable discussions and input for this manuscript. The results presented in this paper rely on: (a) data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET (International Real-time Magnetic Observatory Network) for promoting high standards of magnetic observatory practice. (b) Magnetic activity indices, and related data products, calculated and made available by ISGI (International Service of Geomagnetic Indices) Collaborating Institutes from data collected at magnetic observatories. We thank the involved national institutes and ISGI. This work was supported by the PNST (Programme National Soleil Terre), from CNRS-INSU (Institut des Sciences de l'Univers of Centre National de la Recherche Scientifique) co-funded by CNES (Centre National d'Etudes Spatiales) and CEA (Commissariat à l'Energie Atomique et aux Energies Alternatives), and by TOSCA committee from CNES focused on Space Weather and Geomagnetism. This work was supported by CNES. Veronika Haberle was sponsored by Thales Alenia Space, Toulouse. Veronika Haberle thanks GeoSphere Austria for the support to publish this paper. The authors acknowledge the support of the AID (Agence de l'Innovation et de la Defense) and the French ANR (Agence Nationale de la Recherche), under grant ANR-20-ASTC-0030 (project ANR ASTRID PRISMS).

## References

Bailey, R. L., & Leonhardt, R. (2016). Automated detection of geomagnetic storms with heightened risk of gic. *Earth Planets and Space*, *68*, 1–13. https://doi.org/10.1186/s40623-016-0477-2

Bartels, J. (1949). Appendix b of iatme bulletin no. 12b: The standardized index, ks, and the planetary index, kp. *IUGG*. https://doi.org/10.25577/xkjt-w404

Bartels, J., Heck, N. H., & Johnston, H. F. (1939). The three-hour-range index measuring geomagnetic activity. *Journal of Geophysical Research*, *44*(4), 411–454. https://doi.org/10.1029/te044i004p00411

BCMT. (1921). Bureau central de magnétisme terrestre, french global network of geomagnetic observatories [dataset]. *BCMT, France*. Retrieved from https://doi.org/10.18715/BCMT.MAG.DEF

Blanc, M., & Richmond, A. (1980). The ionospheric disturbance dynamo. *Journal of Geophysical Research*, *85*(A4), 1669–1686. https://doi.org/10.1029/JA085iA04p01669

Chambodut, A., Marchaudon, A., Lathuillère, C., Menvielle, M., & Foucault, E. (2015). New hemispheric geomagnetic indices \alpha with 15 min time resolution. *Journal of Geophysical Research: Space Physics*, *120*(11), 9943–9958. https://doi.org/10.1002/2015JA021479

Clarke, E., Baillie, O., Reay, S. J., & Turbitt, C. W. (2013). A method for the near real-time production of quasi-definitive magnetic observatory data. *Earth Planets and Space*, *65*(11), 1363–1374. https://doi.org/10.5047/eps.2013.10.001

Davis, T. N., & Sugiura, M. (1966). Auroral electrojet activity index ae and its universal time variations. *Journal of Geophysical Research*, *71*(3), 785–801. https://doi.org/10.1029/jz071i003p00785

Echer, E., Gonzalez, W. D., & Tsurutani, B. T. (2011). Statistical studies of geomagnetic storms with peak dst<=-50nt from 1957 to 2008. *Journal of Atmospheric and Solar-Terrestrial Physics*, *73*(11–12), 1454–1459. https://doi.org/10.1016/j.jastp.2011.04.021

Gjerloev, J. W. (2012). The supermag data processing technique. *Journal of Geophysical Research*, *117*(A9), 9213. https://doi.org/10.1029/2012JA017683

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, *99*(A4), 5771–5792. https://doi.org/10.1029/93ja02867

Gough, D. I. (1973). The interpretation of magnetometer array studies. *Geophysical Journal of the Royal Astronomical Society*, *35*(1–3), 83–98. https://doi.org/10.1111/j.1365-246X.1973.tb02416.x

Haberle, V. (2023). *Determination of space weather effects on the geomagnetic field: An automatic derivation of geomagnetic baselines containing quiet variations (theses)*. Université Paul Sabatier - Toulouse III. Retrieved from https://theses.hal.science/tel-04573929

Haberle, V., Marchaudon, A., Chambodut, A., & Blelly, P. L. (2022). Direct determination of geomagnetic baselines during quiet periods for low- and mid-latitude observatories. *Journal of Geophysical Research: Space Physics*, *127*(8), e2022JA030407. https://doi.org/10.1029/2022JA030407

Haines, C., Owens, M. J., Barnard, L., Lockwood, M., & Ruffenach, A. (2019). The variation of geomagnetic storm duration with intensity. *Solar Physics*, *294*(11), 154. https://doi.org/10.1007/s11207-019-1546-z

Hutchinson, J. A., Wright, D. M., & Milan, S. E. (2011). Geomagnetic storms over the last solar cycle: A superposed epoch analysis. *Journal of Geophysical Research*, *116*(A9), 9211. https://doi.org/10.1029/2011JA016463

Huy, M. L., & Amory-Mazaudier, C. (2008). Planetary magnetic signature of the storm wind disturbance dynamo currents: Ddyn. *Journal of Geophysical Research*, *113*(A2). https://doi.org/10.1029/2007JA012686

INTERMAGNET. (2024). International real-time magnetic observatory network reference data set (irds) 2020 [dataset]. INTERMAGNET. GFZ, Germany. Retrieved from https://doi.org/10.5880/INTERMAGNET.1991.2020

ISGI. (1906). International service of geomagnetic indices [collection]. EOST, Strasbourg, France. Retrieved from https://doi.org/10.17616/R3WS49

Iyemori, T. (1990). Storm-time magnetospheric currents inferred from mid-latitude geomagnetic field variations. *Journal of Geomagnetism and Geoelectricity*, *42*(11), 1249–1265. https://doi.org/10.5636/jgg.42.1249

Kamp, M. V. D. (2013). Harmonic quiet-day curves as magnetometer baselines for ionospheric current analyses. *Geoscientific Instrumentation, Methods and Data Systems*, *2*, 289–304. https://doi.org/10.5194/gi-2-289-2013

Lehmann, G. (2006). Kappa sigma clipping. *The Insight Journal*. https://doi.org/10.54294/39QEKN

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. Retrieved from https://doi.org/10.1029/2020SW002641

Mayaud, P.-N. (1967). *Iaga bulletin no. 21: Atlas of k indices, text and figures*. IUGG Publication Office. https://doi.org/10.25577/gb4j-zn77

Mayaud, P.-N. (1972). The aa indices: A 100-year series characterizing the magnetic activity. *Journal of Geophysical Research*, *77*(34), 6870–6874. https://doi.org/10.1029/ja077i034p06870

Mayaud, P. N. (1980). Derivation, meaning, and use of geomagnetic indices. *American Geophysical Union*, *22*. https://doi.org/10.1029/GM022

Sucksdorff, C., Pirjola, R., & Häkkinen, L. (1991). Computer production of k-indices based on linear elimination. *Geophysical Transactions*, *36*, 333–345.

Sugiura, M. (1964). Hourly values of equatorial dst for the igy. *Annals of the International Geophysical Year*.

Sugiura, M., & Kamei, T. (1991). *Iaga bulletin no. 40: Equatorial dst index 1957-1986*. ISGI Publications Office. https://doi.org/10.25577/z1kh-9196

Troshichev, O., & Janzhura, A. (2012). *Space Weather Monitoring By Groundbased Means: Pc index*, *287*.

Vennerstrom, S., Lefevre, L., Dumbović, M., Crosby, N., Malandraki, O., Patsou, I., et al. (2016). Extreme geomagnetic storms – 1868 – 2010. *Solar Physics*, *291*(5), 1447–1481. https://doi.org/10.1007/s11207-016-0897-y

Walach, M. T., & Grocott, A. (2019). Superdarn observations during geomagnetic storms, geomagnetically active times, and enhanced solar wind driving. *Journal of Geophysical Research: Space Physics*, *124*(7), 5828–5847. https://doi.org/10.1029/2019JA026816

WDCKyoto. (2022). Mid-latitude geomagnetic indices asy and sym (asy/sym indices) [dataset]. World Data Center for Geomagnetism Kyoto et al. In *Data analysis center for geomagnetism and space magnetism, graduate school of science*. Kyoto University. Retrieved from https://doi.org/10.14989/267216

Yamazaki, Y., Matzka, J., Stolle, C., Kervalishvili, G., Rauberg, J., Bronkalla, O., et al. (2022). Geomagnetic activity index hpo. *Geophysical Research Letters*, *49*(10), e2022GL098860. https://doi.org/10.1029/2022GL098860