Mandic, I., Korte, M. (2017): On the possibility of producing definitive magnetic observatory data within less than one year. - *Acta Geophysica*, *65*, 2, pp. 275—286.

DOI: http://doi.org/10.1007/s11600-017-0026-9

Acta Geophys. (2017) 65:275–286
DOI 10.1007/s11600-017-0026-9

RESEARCH ARTICLE

On the possibility of producing definitive magnetic observatory data within less than one year

Igor Mandić[1,2] · **Monika Korte**[2]

Received: 1 March 2017 / Accepted: 3 March 2017 / Published online: 20 March 2017

Abstract Geomagnetic observatory data are fundamental in geomagnetic field studies and are widely used in other applications. Often they are combined with satellite and ground survey data. Unfortunately, the observatory definitive data are only available with a time lag ranging from several months up to more than a year. The reason for this lag is the annual production of the final calibration values, i.e. baselines that are used to correct preliminary data from continuously recording magnetometers. In this paper, we will show that the preparation of definitive geomagnetic data is possible within a calendar year and presents an original method for prompt and automatic estimation of the observatory baselines. The new baselines, obtained in a mostly automatic manner, are compared with the baselines reported on INTERMAGNET DVDs for the 2009–2011 period. The high quality of the baselines obtained by the proposed method indicates its suitability for data processing in fully automatic observatories when automated absolute instruments will be deployed at remote sites.

Keywords Geomagnetic observations · Definitive data · Quasi-definitive data · Baseline fitting · Data processing

✉ Igor Mandić
mandici@gfz.hr

Monika Korte
monika@gfz-potsdam.de

[1] Department of Geophysics, Faculty of Science, University of Zagreb, Zagreb, Croatia

[2] Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Introduction

Magnetic observatories are designed to perform continuous and accurate measurements of the strength and direction of the Earth's magnetic field over many years, or even centuries. Nowadays, the time resolution of these measurements is one minute or less. The observatory data are essential to understand how the field is changing on a wide range of scales from seconds to centuries, and to get an insight into dynamic processes inside and outside the Earth that generate the observed geomagnetic field. Ground observatories serve as base stations and are an important complement for various types of geomagnetic surveys (e.g. Mandea and Korte 2011). They provide data from a different observation altitude and pure time series in contrast to the data obtained for example from airborne or marine surveys and satellite missions, containing both temporal and spatial variations. Observatory data help to better constrain survey data and fill the gap between the present and future satellite missions.

Continuously recording vector magnetometers do not provide absolute field component values, but might be influenced by temperature variations, instable pillars or instrumental drifts. Manual absolute observations are still mandatory at modern magnetic observatories to calibrate the continuous recordings and are usually done once a week. These observations are used in conjunction with the data from the vector magnetometer (variometer) and an independent scalar magnetometer to derive observed base values. These values later serve for calculation of adopted baselines which are used to obtain absolute variometer data free from drift or environmental variations. Adopted values are derived by fitting, interpolation or hand scaling from the observed base values. Today, final baseline fitting is performed on an annual basis and data calibrated with these final baselines are labelled as definitive (D) data. The final D data should be also free from spikes, jumps and other degradations in continuous recordings. For details of preparation of the final observatory data, see Jankowski and Sucksdorff (1996).

Since the establishment of INTERMAGNET (International Real-time Magnetic Observatory Network, http://www.intermagnet.org) in 1991, INTERMAGNET Magnetic Observatories (IMOs) reported three types of magnetic data until recently: D data (described above), reported (R) data, and adjusted (A) data. R data are raw data of the geomagnetic field variation measured with the variometer and should be available within 3 days from acquisition. These data are only suitable for studies of the rapid field variations (and may contain spikes, gaps and discontinuities). Within 7 days from transmission, observatories are allowed to modify R data to produce A data by removing spikes, filling gaps or to modify baselines. However, A data do not have sufficient absolute accuracy and are not free from drift. The D data required in particular for studies of secular variation and the geodynamo process in the Earth’s core are reported only after the end of a calendar year with a time lag from several months up to more than a year. Unlike observatory data, fully calibrated data from satellites are available within a few days from acquisition.

Based on previous satellite missions (CHAMP, Ørsted, SAC-C), several studies demonstrated (Lesur et al. 2006; Macmillan and Olsen 2013; Olsen et al. 2006) the importance of the absolute observatory data for the ongoing Swarm mission, a constellation of three satellites (Friis-Christensen et al. 2006). The data from the Swarm satellites enable separation of the internal and external sources better than ever before. Processing, selection and validation of satellite data are usually based on data from ground observatories and their products, like geomagnetic activity indices. Absolute hourly mean values from world-wide observatories also play an important role in hour-by-hour spherical harmonic analysis used for deriving sophisticated magnetospheric models.

To increase the usefulness of observatory data and to fulfil the needs of the community for global modelling and other user groups, INTERMAGNET has defined the new type of “quasi-definitive” data (Clarke et al. 2013; Peltier and Chulliat 2010; Matzka 2013). Quasi-definitive (QD) data are corrected with temporary baselines, should be close to expected definitive values and are delivered much sooner (1–3 months) than the final D data. The variation part of QD data should have the same (or similar) quality as D data, i.e. without spikes or noise. QD data bridge the gap between preliminary and D data and pave the way to more efficient combination of ground observatory data and satellite, marine and airborne survey data.

Temporary quasi-definitive baselines (QDBLs) are estimated from all absolute measurement results performed in the current year. Some observatories report their QD data in near-real time using baselines obtained by extrapolation. In these cases, the annual baseline amplitude should not exceed 5 nT and this also requires prompt processing of the variation data by observatory staff. Nowadays, de-spiked and magnetically clean variation data and other observatory products like geomagnetic activity indices can be obtained relatively promptly, within several days from recording. Of course, this depends on the available observatory staff, amount and complexity of the encountered problems in the data and efficiency of processing protocols. The same is true for the absolute control of data, only on longer time scales. Human intervention is essential for the absolute control of observatory data. This includes manual absolute observations, retyping and processing of the observational results, quality checks and validation of results, decisions during baseline calculation, etc., both for temporary QDBLs and DBLs. For the production of QD and D data, those timescales correspond to several months and more than a year, respectively.

Today, most observatories use parametric (e.g. polynomial) or semi-parametric (e.g. smoothing splines) fits in the global domain, i.e. on annual basis to derive DBLs. Obtaining baselines in such manner has some disadvantages: (1) A full year of observed values has to be collected before calculating DBLs, (2) In the case of parametric fits, the convergence of temporary QDBLs toward the DBL is slow, i.e. the overlap between cumulative QDBLs and the DBL might not be good. The same is true for semi-parametric fits in the case when the distribution of observed values is non-uniform or the smoothing parameter of the spline functions is changed as the observational dataset is updated. This means that we need to have good baseline stability (i.e. high quality observatory infrastructure) to produce high-quality QD data. In general, global fits are good for estimating general trends in the observational dataset. On the other hand, if we have rapid trending of samples during a short period of time (for example, due to temporary problems with the mechanical stability of the variometer sensor), global fits may not be the appropriate tool to fit these short-term variations without affecting the baseline quality in other parts of the domain. Increasing the polynomial degree or flexibility of smoothing splines, to track samples when the baseline varies more rapidly, can cause “overfitting” or artificial variations in other parts of the domain, particularly in parts where observations are sparse (an example is given in Sect. “Special cases: data gaps and disagreeing intensity in scalar baselines”).

In this paper, we first show that D data can be produced within the timespan less than a year. We show that it is enough to collect a reasonable number of observed values during some period that enable us to do an accurate local fit and to determine DBL for this period. With absolute sampling (usually) once per week, we can collect enough samples within a few months to calculate one piece of DBL within a year. By constructing piecewise continuous fits, with smooth transitions on the edges of these pieces (windows), we are effectively constructing DBL by pieces. The advantage of the piecewise fitting is that observations in the future will not affect the shape of a baseline in the past, which is clearly justified from a physical point of view. Observed base values reflect the instrument response in the current environmental conditions and they are in fact independent from those in the future. We only need a reasonable number of good quality observations for some period of a time to obtain correct estimation of the baseline trend for days without observations. Correct approximation of samples in the periods when the baseline varies more rapidly and avoiding overfitting when we have good baseline stability are further advantages of the piecewise fitting.

We propose a new method for producing DBLs within a year. The developed routine can determine the "best fit" as candidate for QDBL and DBL without human intervention, although we emphasize that QDBL or DBL should be always reviewed by experienced observatory staff. We use the published INTERMAGNET observatory data from 2009 to 2011 to show that the proposed method gives high-quality results in a nearly automated manner. It thus can find its applicability in continuous baseline fitting when automated absolute instruments (Gonsette et al. 2013; Korte et al. 2013) become supplement or standard observatory equipment.

Methods

Proposed method for automatic production of temporary baselines

Motivated by the work and results obtained by Peltier and Chulliat (2010), we decided to design a routine based on cubic smoothing splines (De Boor 1978). They tested their method for producing QDBLs on nine IPGP (Institut de Physique du Globe de Paris) observatories (AAE, BOX, CLF, KOU, LZH, MBO, PHU, PPT, TAM) for the year 2008 and showed that the production of accurate QD data is possible even in the cases where baselines vary more rapidly through a year. The differences between definitive and temporary baselines, calculated after the end of each month, were within a fraction of nT. However, the obtained results might be too promising and unrealistic. The authors determined an optimal smoothing parameter for the dataset on an annual basis. Then, they simulated the production of QDBLs in retroactive manner with a priori known optimal smoothing parameters (see Peltier and Chulliat 2010, for details). In reality, in most cases the optimal smoothing parameter cannot be known in advance and in many cases simple smoothing splines will not give satisfactory results. This will be demonstrated in the next subsection.

To obtain high-quality baselines in an automatic manner, we use cubic smoothing splines in conjunction with additional constrains. First, we calculate a baseline by pieces (windows). We empirically found a length of eight observational days ($L = 8$) to be a good choice in most cases (one observational day can have several observation sets). Depending on the observation frequency, the time span of a window may vary from a week up to several months depending on the observatory. Additionally, samples from two ($d = 2$) observation days before and after the window are used for calculation of a baseline within the window. This alleviates edge effects between neighbouring baseline segments. The baseline within a window is a cubic smoothing spline function $f(t)$ that minimizes the expression

$$p\sum_{i=1}^{n} w_i(y_i - f_i)^2 \quad + \quad (1-p)\int_a^b \{f''(t)\}^2 dt, \tag{1}$$

where $f_i = f(t_i)$ are values of the baseline $f(t)$ at the observation times $t_i, f''(t)$ stands for the second derivative. In the integral expression, a and b are the limits of the extended window containing $L + 2d$ observations. Observations are denoted with y_i and w_i are corresponding weights. The smoothing parameter $p \in [0, 1]$ controls the smoothness of $f(t)$. For higher values of p, the baseline will be closer to the observations. In the case $p = 0$, the baseline will be a straight line and for $p = 1$ the interpolating baseline will pass exactly through the data points. To obtain weights first, we estimate a preliminary baseline $\tilde{f}(t)$ where all observations have weights equal to 1. Then, using the residuals $y_i - \tilde{f}_i$, weights are estimated according to the most frequent value (Steiner 1988). This way relatively flexible baseline segments (high values of p) are robust to bigger outliers or low-quality observations. Determination of the optimal smoothing parameter is performed automatically according to relation (De Boor 2003)

$$p = 1/\left(1 + h^3/c\right), \tag{2}$$

where h is the average sampling, i.e. average period between observation days. The current version of the program assumes the same coefficient c for three vector components while the scalar c may be the same or different. Coefficient c is varied in the range from 10^{-3} up to 10^2 and for each c (vector and scalar) baselines are calculated.

To determine the suitable value for c and check the quality of the determined baselines, the so-called "Delta F" values are calculated. Delta F presents the difference between the total field intensity obtained from vector and independent scalar magnetometers. Here, the scalar total field is corrected for the field difference between the sensor site and the absolute pier (for details see St-Louis 2012). The best c corresponds to the minimum value of Delta F. For this purpose, the routine requires the variometer and scalar daily mean values in addition to the observations itself. In case no scalar magnetometer data are available, it is not possible to perform the Delta F check and the default value of $c = 0.5$ is used.

In some cases of IMOs, the annual long-term baseline drift is comparable or order of magnitude larger than random fluctuations of samples. On the other hand, some observatories have very frequent absolute observations (several sets, almost every day) in some time intervals. In this case, observations can show random fluctuations, which are not expected in the nature of the underlying baseline. Note also that with decrease of h (frequent observations), p increases. According to Eq. (1), this means that the baseline will try to follow the same random fluctuations by remaining close to the observations. Thus, in general we prefer "smoothness" over "staying close to the data". By choosing the default $c < 1$, we restrain the spline function to be smooth enough to capture the trend and not to follow random scattering or erroneous observations.

Once we have obtained piecewise baselines according to Eq. (1), with several constrains as described above, the resulting baseline pieces will have similar (but not equal) values on the edges to neighbouring windows. Thus, a final smoothing by a spline function with the default $c = 0.005$ for vector and $c = 0.0005$ for scalar is applied to piecewise baselines to obtain smooth transition on the edges of the windows for which the baselines are calculated. In this way, we obtain smooth, continuous baselines over the whole domain. Optionally, the final vector and scalar c can be set manually by the user. Except in the last window, the shape of a baseline obtained in this manner will not be affected as we add new observations.

Realignment or relocation of a magnetometer sensor (or some other reasons) may result in discontinuities, i.e. jumps in the variometer or scalar baseline. Generally, the number of discontinuities may be different for the vector and scalar instrument. These discontinuities are usually known at the observatories and are marked with "d" in the INTERMAGNET baseline files. The marker "d" and originally reported baseline values are used to identify jumps and their magnitudes. The routine requires a reference list with times and values of discontinuities. In our test cases, this list is created from the information provided in the INTERMAGNET baseline files. Observations are then corrected to a single reference level to obtain accurate estimations of the new adopted values.

Special cases: data gaps and disagreeing intensity in scalar baselines

The BOX example from Peltier and Chulliat (2010) was chosen to demonstrate how non-uniformity of sampling and assumption about an a priori "known" smoothing parameter can introduce artificial baseline variation that is not supported by samples. In Fig. 1, only the horizontal (H) and vertical (Z) baseline components are displayed, together with Delta F. Original baselines and Delta F (reported on the INTERMAGNET DVD) are obtained taking into account all observations. Now, let us imagine that the observations were not possible during spring due to some operational problems (e.g. the observatory location was unreachable for some reason or problems with the declination–inclination magnetometer occurred). Results obtained under the assumption that some observations in spring are missing with the IPGP and the newly proposed method are presented by blue and black lines, respectively. Vertical grey lines designate the automatically defined windows in the proposed method (Sect. "Proposed method for automatic production of temporary baselines "). The IPGP and original baselines overlap perfectly everywhere except in the third window. Clearly, we can see that the IPGP method gives a more erroneous estimation of the baseline than the new method, if observations in spring are omitted. The initially chosen smoothing parameter in the

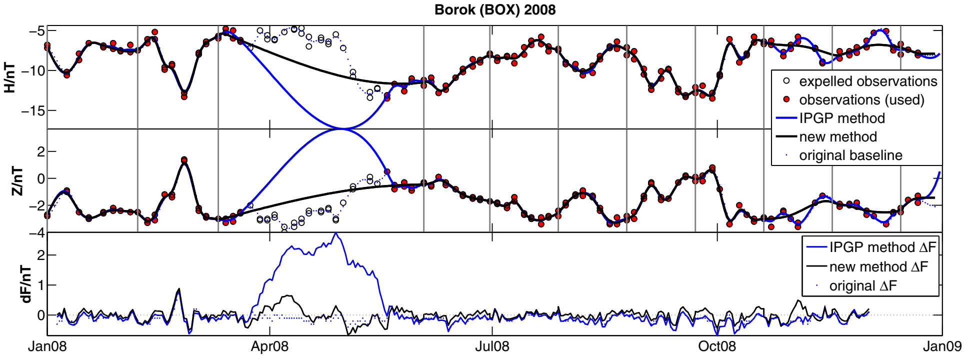


Fig. 1 Baselines calculated with the IPGP and new method (default settings) under the assumption that observations from spring (*white circles*) are missing. Original baselines and Delta F (*blue dotted line*) were obtained by taking all observations (*red and white circles*)

IPGP method will not be a good choice anymore, since the smoothing splines in the BOX case are quite flexible and we do not have any samples in spring to restrain the spline wiggling. The new baselines give more reasonable values due to several constraints that are described in Sect. “Proposed method for automatic production of temporary baselines ”.

On the other hand, Fig. 1 shows somewhat smoother baselines obtained with the new method during November and December. Too smooth values of the new baselines are caused by missing information about the total field difference between the scalar sensor and absolute (i.e. reference) site. Prior to 2009, observatories did not report this information, named as “scalar baseline”. Also, there is no information about Delta *F* in December, i.e. data from the scalar magnetometer are missing. Thus, the proposed baselines in the last two windows were obtained without any information about the scalar baseline and Delta *F* in December.

In many cases, the scalar baseline has the biggest impact on the Delta *F* residual and sometimes the baseline obtained from scalar samples cannot properly reduce the total field recordings from scalar magnetometer to the reference site. Although scalar magnetometers in general are absolute instruments, a potentially time-varying baseline results come from the fact that vector, scalar and absolute observations are carried out at different sites. All measurements in the end are referred to the same reference site. The difference between values at the site of the scalar instrument and the reference site determines the scalar baseline. As demonstrated in Fig. 2 (“*S*” diagram) in the case of the reported data, we have an obvious degradation in Delta *F* in this example. (IMOs on Figs. 2, 3 and 4a, b are left as anonymous). The major reason is the presence of a constant offset (approximately −0.5 nT) in the reported scalar baseline.

We presume that the presence of the offset in scalar samples on Fig. 2 is a consequence of erroneous or outdated information about the field difference between the scalar and reference site in the protocol for processing absolute observations. Generally speaking, the representativeness (offset, variability and scattering) of the scalar samples also depends on the magnetic homogeneity of the observatory site. We need to keep in mind that recordings and observations in the observatory are performed at three different sites. The gradiometer difference and its variation over a time, between the reference and scalar site, should be as small as possible. The same stands for the difference between the variometer and scalar site. Also, scalar and vector recordings during absolute observations must be free from spikes and the differences between natural short-term variations at all observatory sites should be negligible. These are basic prerequisites for obtaining quality base values and a proper reduction of the variometer and scalar data to the reference site.

In situations when we have disagreeing intensity in the scalar baseline (i.e. offset in Delta *F*) or increased scattering of *S* samples, the new method estimates the *S* baseline as a de-trended value of Delta *F* obtained in the

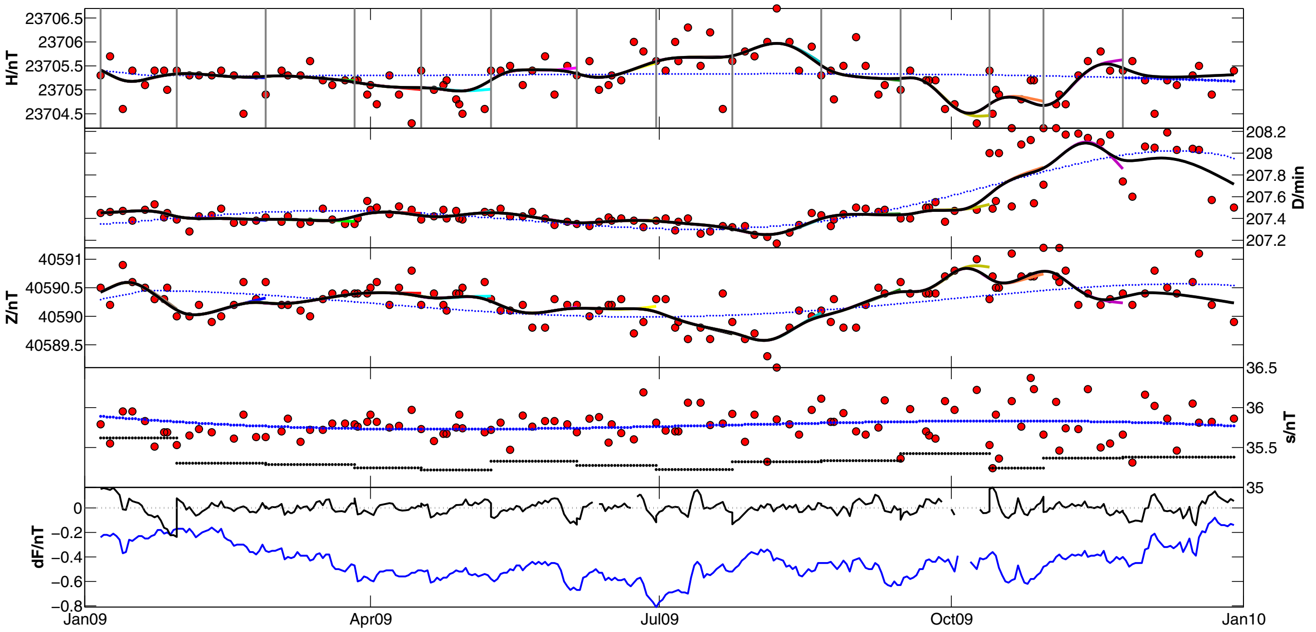


Fig. 2 Vector baselines obtained with the new method (*black line*) and original baselines (*blue dotted line*). Scalar baseline obtained with the new method is displayed by *black dotted lines* and original by *blue dotted line*. Observations are denoted with *red circles*. Original Delta *F* (*blue line*) corresponds to original baselines (*blue dotted lines*)

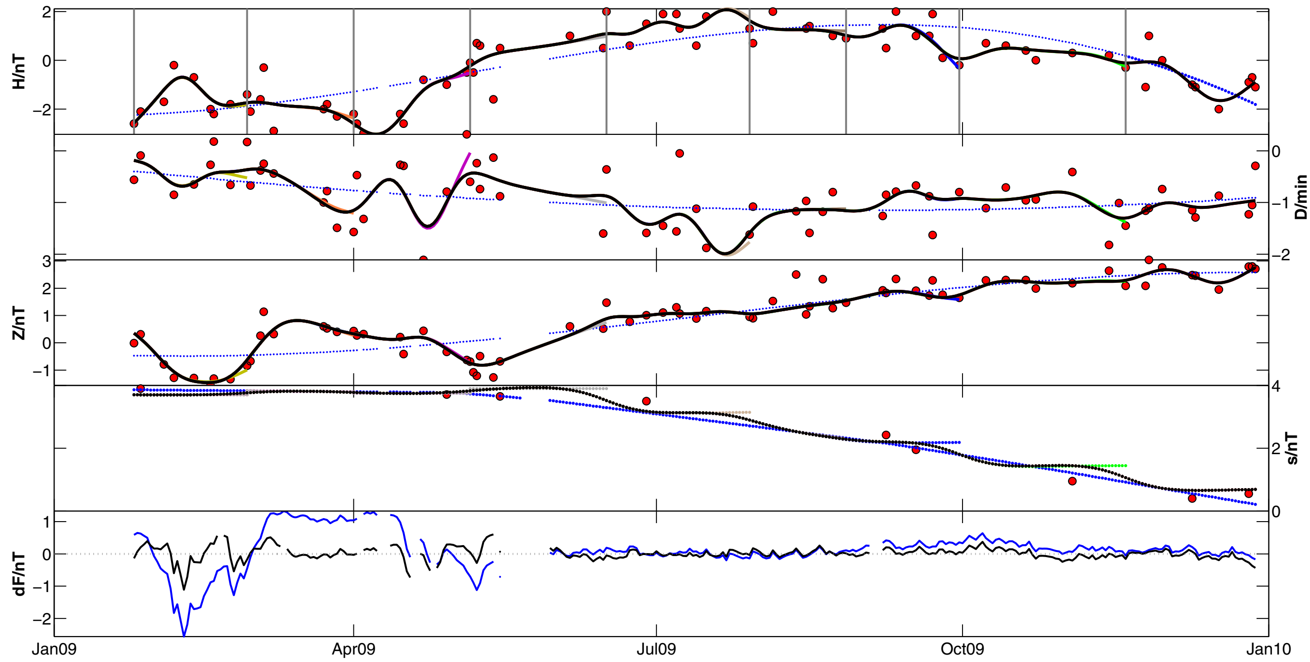


Fig. 3 Vector baselines obtained with the new method (*black line*) and original baselines (*blue dotted line*). Scalar baseline obtained with the new method (*black dotted line*) and original baseline (*blue dotted line*). The new temporary scalar baselines (within windows) are displayed by *coloured constant lines*. Observations are denoted with *red dots*. Original Delta *F* (*blue line*) corresponds to original baselines

iterative manner described in Sect. "Proposed method for automatic production of temporary baselines" under the initial assumption that *S* baseline is zero. The baseline obtained in this way is shown by black dotted lines in Fig. 2 ("*S*" diagram). Each piecewise constant segment of the annual baseline is estimated for each of the windows shown in the "*H*" diagram (Fig. 2, grey lines).

In case we do not have any information on the scalar baseline or less than 3 samples, necessary for spline calculation within the window, we assume that the scalar baseline is a constant. This assumption is also valid if Delta *F* calculated using the spline scalar baseline, obtained from scattered *S* samples, is bigger than Delta *F* obtained with a constant scalar baseline.

Furthermore, if we obtain discontinuities bigger than 0.5 nT between piecewise constant scalar baselines, then these constant baselines are smoothed by a final spline with default $c = 0.0005$ (see Sect. "Proposed method for automatic production of temporary baselines") to avoid sudden jumps in *F* from the independent scalar instrument. An example of a small jump (<0.5 nT) can be seen in Fig. 2 in Delta *F* and *S* (on the edge of the first and second window). In Fig. 3 ("*S*" diagram), we can see preliminary piecewise constant baselines (coloured dotted lines). The final adopted baseline obtained by smoothing the constant baseline segments is represented by a black dotted line.

Looking at the fit of the vector baselines in Figs. 2 and 3, we see that new baselines capture the trend of the samples much better than the original baselines. This is especially pronounced in the *Z* component in Fig. 3. In the first half of the year, the originally adopted values could not completely compensate a drift in the *Z* baseline. As a consequence, the Delta *F* residuals are increased in this period. Note that in this period, the original and new values of the scalar baseline are practically identical. In Fig. 2, we see a long-term drift through the whole year in addition to a constant offset in the original Delta *F*. The reported baselines in this case also underestimate the baseline trends indicated by samples in *H* and *Z*.

Manual adjustment and quasi-definitive baselines

A single computational method may not be appropriate to handle all different dataset types and in some cases might not give satisfactory results. This can occur especially in cases where we have a relatively small number or different numbers of samples in each component, non-uniform sampling, sudden discontinuities and when every dataset has its own characteristics (like trends, amplitudes and sampling quality). All these are potential features of observational datasets obtained from standard measurements in geomagnetic observatories. Testing our new routine for all IMOs showed that in some cases improved results are obtained with manual adjustments of the automated procedure following visual inspection of the data and calculated baselines.

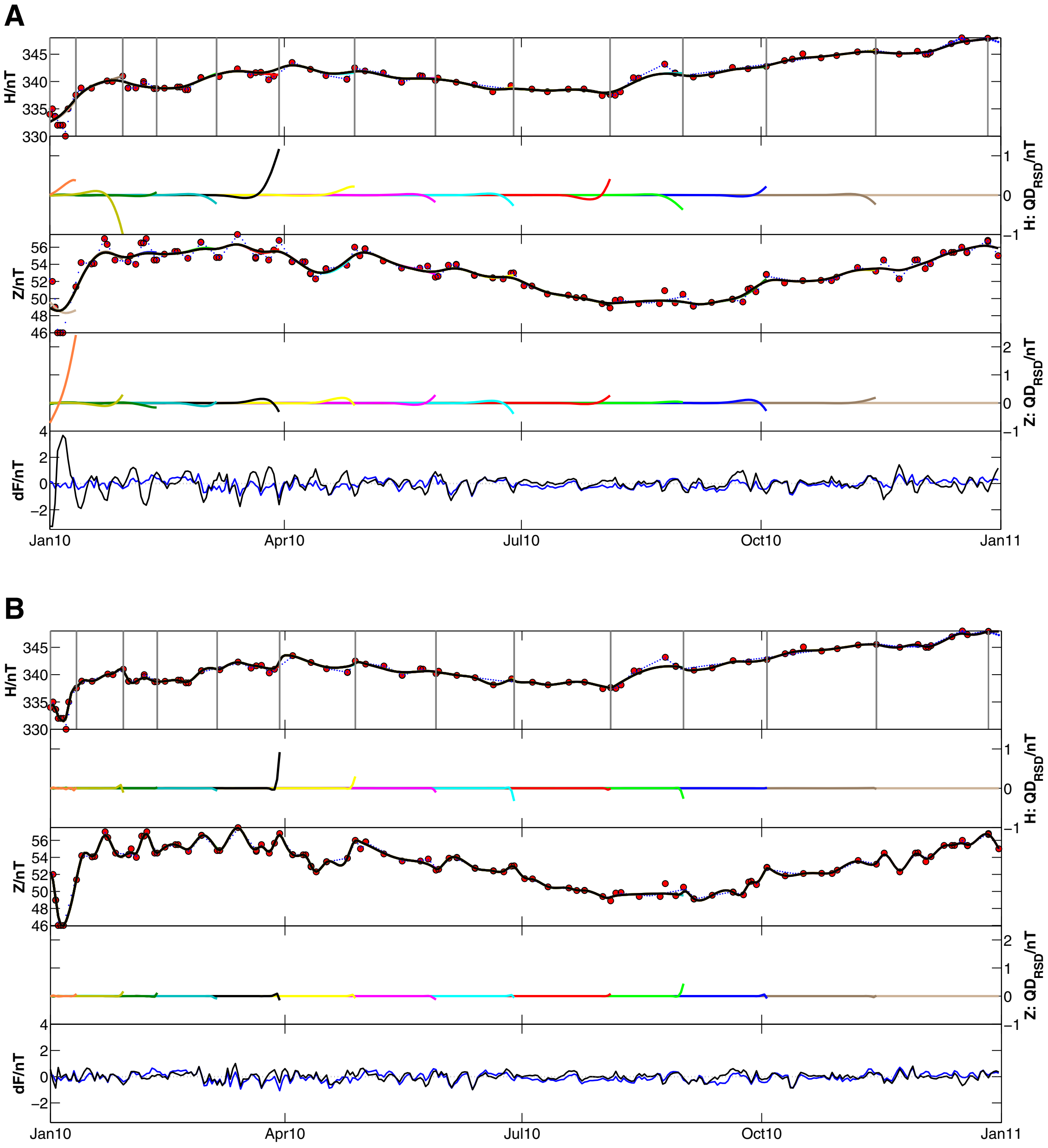


Fig. 4 **a** H and Z baselines obtained with the new method (*black lines*) and original baselines (*blue dotted lines*). Residuals between QDBLs and DBLs are shown with *coloured lines*. Observations are denoted with *red dots*. Original Delta F (*blue line*) corresponds to original baselines. **b** Same as **a** only the new H and Z baselines are obtained with manual adjustment. Results are presented with the same ordinate scaling as used in **a**

We manually varied L and d parameters (see Sect. "Proposed method for automatic production of temporary baselines") or the smoothing parameter of the final spline which is used to join sub-splines within windows. This procedure is demonstrated in Fig. 4a, b. Results obtained with default settings are presented in Fig. 4a. In this case, the new baselines are not flexible enough to fit rapid baseline variability at the start of the year. This clearly leads to degradation in Delta F.

If we manually increase coefficient c in relation (2) of the final spline that smooths the baselines within individual windows, a more flexible baseline will be obtained. This is illustrated in Fig. 4b. With this minor intervention (L and d were not changed), significantly better results were obtained.

In Figs. 2, 3 and 4a, b, the new temporary QDBLs (coloured lines) are simultaneously plotted under DBL. However, due to practically perfect overlapping between cumulative QDBLs and the final baseline, only small segments of QDBLs are visible in Figs. 2 and 3. A fraction of QDBLs, barely noticeable, can be seen on the right edges of some windows. For example in Fig. 2 near the edge of eleventh window part of the vector, QDBLs (orange line) are noticeable, which are slightly different from final DBLs calculated from all observations in the year. Similarly, in Fig. 3 this can be seen in declination (*D*) near the edge of the third window. The residuals between QDBL and DBL ($\mathrm{QD_{RSD}}$) are plotted separately in Fig. 4a, b for better visibility. Here, we see that small differences between QDBL and DBL exist only near the right edge of a window. If we denote the number of windows with *N*, then we have perfect overlapping QDBLs with DBL within all windows from 1 to $N-1$, i.e. $\mathrm{QD_{RSD}}$ are exactly zero. This means that DBLs are calculated within all windows, from first up to $N-1$. Only near the right edge of *N*-th window, small differences exist and till we define the $N+1$ window the baseline is labelled as QD. Generally, for all observatories used in this study these small differences are within fractions of an nT.

Verification of the proposed method

Data selection and preparation

To compare our baseline calibration results with the ones reported by IMOs, the data in the period 2009–2011 were used. This period was chosen because calibration data from the independent scalar magnetometer were not available before 2009. Also during the preparation of this paper, the last available INTERMAGNET DVD was published for 2011. In this period, we have used only data from observatories that report Delta *F* results. This way it is possible to estimate the adoption quality according to the Delta *F* check. The overall number of observatories that reported Delta *F* in this 3 year period is 255.

Standardly, the baseline adoption for a year (YY) includes observations from December of $\mathrm{YY}-1$ and January $\mathrm{YY}+1$. Since baseline results are reported in standardized yearly IBF V2.0 files (http://www.intermagnet.org/data-donnee/formats/ibfv200-eng.php), new baselines are calculated using only observations within YY. Then, the original and new baseline results are compared for a time period between the first and last observational day of year.

Statistical analysis of results

In Sect. "Methods", we have presented only a few examples of the 255 analysed cases. To show all results in summarized form, some statistics are shown in the following.

To obtain an idea about the goodness of fit with the new and reported DBLs, we have simply used squared correlation coefficients (R^2) between observed and adopted values. Table 1 shows results of examples presented in Figs. 2, 3, and 4a, b. From the numerical values in Table 1 and from the plotted results, we can see that R^2s are quite good indicators of the goodness of fit.

Table 1 Squared correlation coefficients (R^2) between observed and adopted values presented in Figs. 2, 3, and 4a, b

Case/comp.	DBL type	R_1^2	R_2^2	R_3^2	R_4^2
Figure 2	Reported DBL	0.036	0.643	0.261	0.001
	New DBL	0.450	0.724	0.653	0.003
Figure 3	Reported DBL	0.674	0.198	0.787	0.978
	New DBL	0.869	0.526	0.943	0.972
Figure 4a, b	Reported DBL	0.990	–	0.990	–
	New DBL	0.949	–	0.899	–
	New manual DBL	0.986	–	0.990	–

Four coefficients (R_1^2,… R_4^2) correspond to three vector and one scalar component. R_2^2 and R_4^2 are not given for Fig. 4a, b because examples of these plots are not shown there

Then, the maximal absolute residuals ($\max|\mathrm{RSD}|$) and the average absolute residuals ($\mathrm{avg}|\mathrm{RSD}|$) between the new and reported baselines were determined on yearly basis. This was done for each of the 255 cases, for vector and scalar baselines and Delta *F* residuals. We have rejected 5 cases where $\max|\mathrm{RSD}|$ and $\mathrm{avg}|\mathrm{RSD}|$ were outside two standard deviations from the average results. In all rejected cases, the new baselines gave better results in Delta *F* and according to squared correlation coefficients between observed and adopted values as a measure of the fit. The remaining residuals are sorted into 0.5 nT bins and presented in the form of histograms.

A similar analysis was done for the new, cumulative temporary baselines (i.e. QDBLs) within a year. Here, the residuals ($\mathrm{QD_{RSD}}$) were determined only with respect to the new DBL. In the QD case, only the maximal absolute residuals within windows are determined ($\max|\mathrm{RSD}^i_{\mathrm{QD}}|$) where $i=1,\ldots,N$ and *N* is the number of windows. Then, the maximal ($\max(\max|\mathrm{RSD}^i_{\mathrm{QD}}|)$) and average ($\mathrm{avg}(\max|\mathrm{RSD}^i_{\mathrm{QD}}|)$) values from all windows within a year were determined. By considering only $\max|\mathrm{RSD}^i_{\mathrm{QD}}|$, we have excluded periods of perfect overlapping among the new QDBL and DBL (Fig. 4a, b, residual diagrams). Otherwise, we would obtain unreasonable small residuals (less than 0.01 nT), i.e. much smaller than the achievable absolute accuracy of the baseline calibration procedure. Residuals between the new QDBLs and reported DBLs were not calculated separately. These residuals correspond approximately to residuals between the new DBLs and reported DBLs. In other words, $\mathrm{avg}|\mathrm{RSD}^i_{\mathrm{QD}}|$ are much smaller than $\max|\mathrm{RSD}|$ and $\mathrm{avg}|\mathrm{RSD}|$.

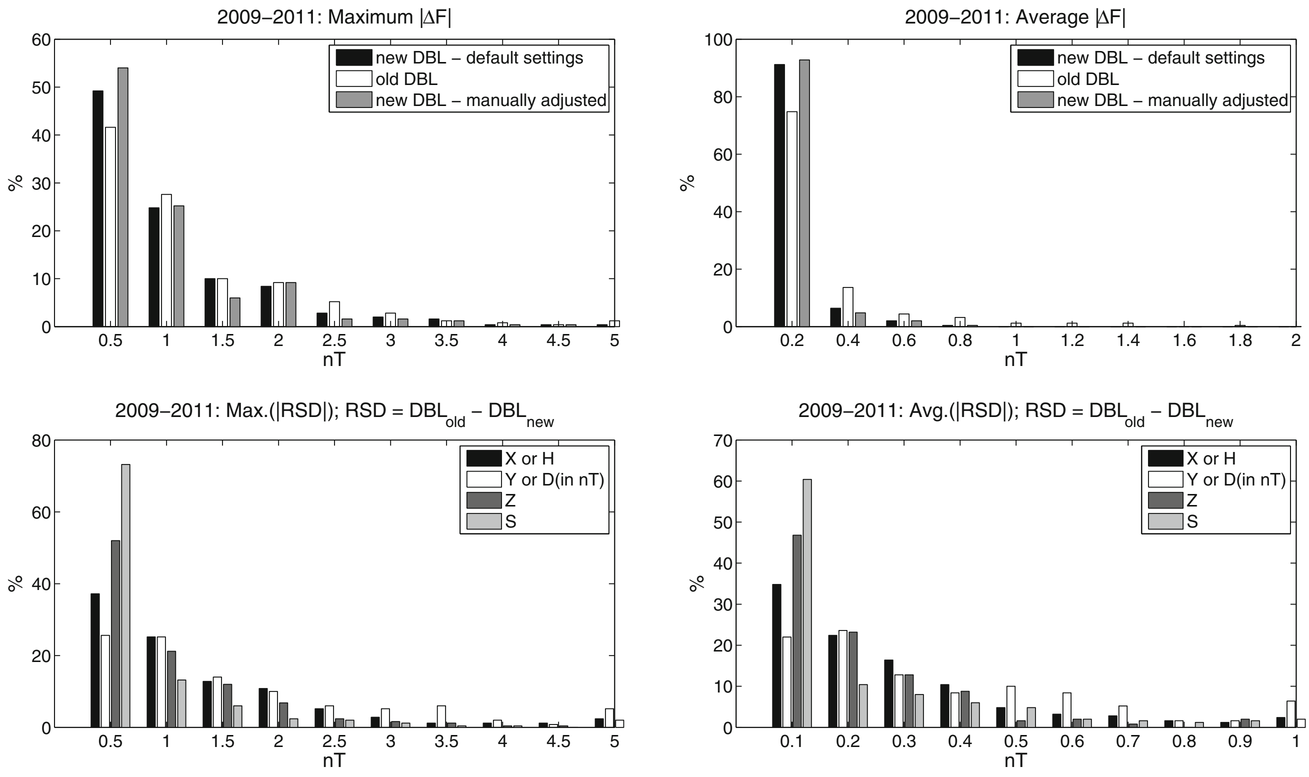
Results and discussion

In Fig. 5 (upper panels), the maximum and average absolute Delta *F* values are presented. New baselines, automatically calculated, with default settings are displayed with black histograms. Results of originally reported Delta *F* are presented with white histograms. We clearly see that new baselines in general give slightly better results according to the Delta *F* check. However, here we must emphasize that in some individual cases human intervention is crucial to improve the baseline fit. Minor manual adjustments were done in 46 cases for this purpose, based on visual inspection. After re-calculation of baselines, small improvements can be seen in Delta *F* (Fig. 5, grey histograms).

Histograms of differences between the new DBLs, obtained in automatic manner, and reported DBLs show that max|RSD| are within 1 nT and avg|RSD| are within 0.2 nT in approximately 80% of the cases. The biggest residuals are obtained in the *Y* (or *D*) component and the smallest are in the *S* component. These results are expected because usually the baseline variability is biggest in *Y* (or *D*) while the *S* baseline is the most stable. Stability of the *S* baseline is also evident from Fig. 6 where around 30% of *S* baselines are approximated by constant or piecewise constant fits, i.e. R^2s are ≤ 0.1. For the vector components, we clearly find a higher number of cases with R^2 between 0.5 and 0.9 obtained with the new baselines than with the originally reported baselines.

A somewhat higher number of observatories with R^2 around 1 in the case of originally reported baselines is mostly due to observatories that determine their DBLs by linear interpolation or piecewise linear fits. In this case, human intervention is essential. Here, all outliers and low-quality observations are excluded from the observational datasets. Also, the data processor makes decisions during piecewise fitting, for example which observations should be used to calculate one piece of a baseline, where to join the baseline pieces, which may be defined differently for different components, etc. These cases also show that to obtain high-quality data in some cases the intervention from a skilful and experience data processor is essential.

The same test was performed with the new baselines that are calculated using default *c* parameters for the splines within windows. This means fitting was not performed iteratively to minimize Delta *F,* i.e. to find an optimal *c*. The overall average differences between baselines with default *c* and automatically determined *c* are quite small. For all components, more than 90% of average/maximal residuals are <0.1 nT/1 nT while 0.3 nT/1 nT are rarely excited. However, in some cases significant improvements can be seen in the fit quality for baselines that are determined by minimization of Delta *F*.


2009–2011: Maximum |ΔF|

2009–2011: Average |ΔF|

2009–2011: Max.(|RSD|); RSD = DBL$_{old}$ − DBL$_{new}$

2009–2011: Avg.(|RSD|); RSD = DBL$_{old}$ − DBL$_{new}$

Fig. 5 *Upper panels* histograms of the absolute maximum and average absolute Delta *F*. *Lower panels* residuals between the new DBL (without manual adjustment) and reported DBL for three vector and the scalar component

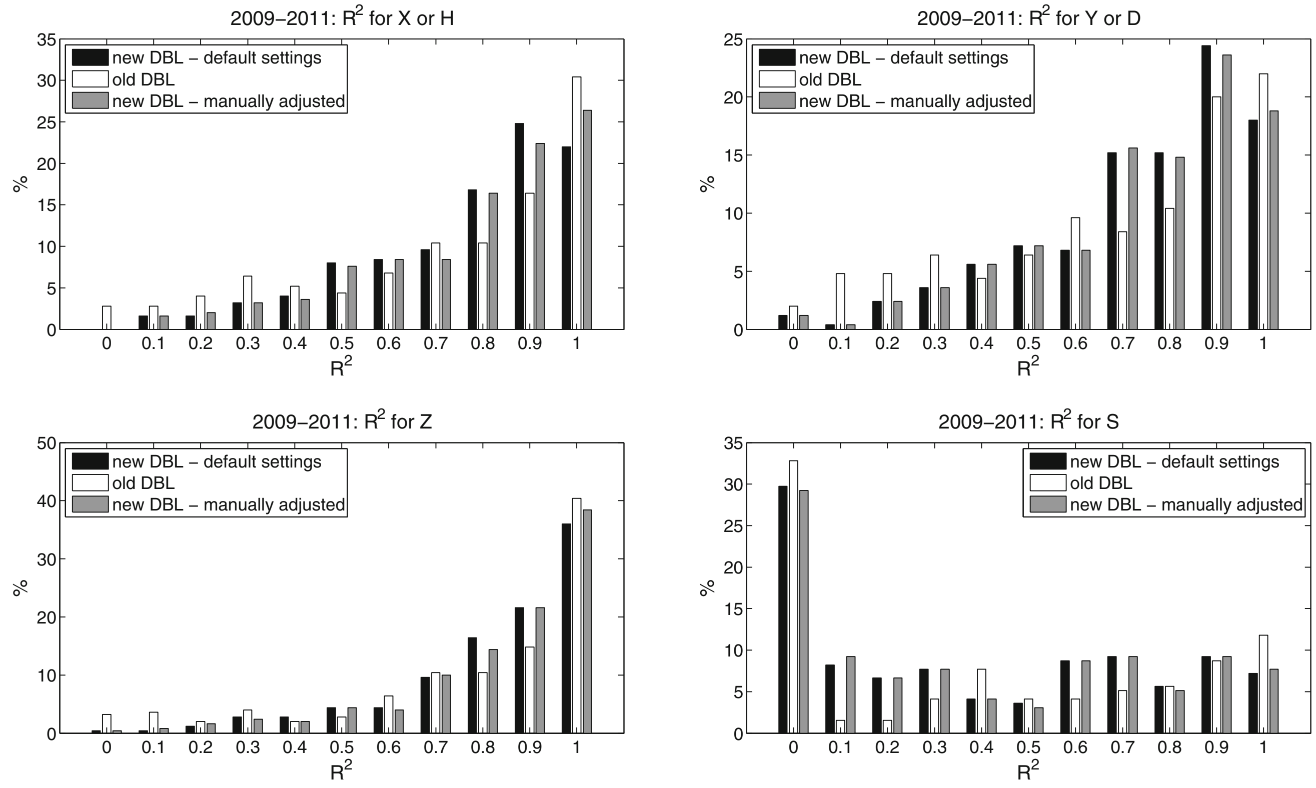


Fig. 6 Squared correlation coefficients (R^2) between adopted and observed values for three vector and the scalar component. Three cases are presented. R^2 obtained for newly automatically calculated baselines (*black*), reported baselines (*white*) and newly calculated after manual adjustment for some observatories (*grey*)

The new method thus has several advantages compared to traditional ones. First, the new baselines are indeed more accurate, since in most cases they track the observations better than the original baselines and consequently give a smaller Delta F (Figs. 2, 3). Secondly, residuals between the new DBL and QDBLs are practically negligible (Fig. 7). Small differences exist only near the right edge of the last window for which baseline is calculated. In average, more than 90% of these residuals are <0.3 nT (Fig. 7right) while yearly maximum residuals in all 250 cases rarely exceed 1 nT (Fig. 7left). Note that this discrepancy occurs only for a small time period, as demonstrated in Fig. 4a, b. Immediately after calculation of a new baseline using samples from the consecutive window (i.e. baseline update), this discrepancy is corrected. Last, but most important, the new DBLs are determined within a year. The time delay of DBL production may vary from several weeks up to several months and depends only on the frequency of absolute observations.

Our new temporary baselines also clearly fulfil the requirements to define them as QDBLs with respect to the reported DBLs. According to INTERMAGNET standards, the (average) differences between QDBLs and DBLs should be less than 5 nT nd for many observatories; these differences are within 1 nT. From Fig. 5, we can see that average avg|RSD| are within 1 nT in all components and that about 80% are within 0.2 nT. Here, we made the assumption that new temporary baselines have perfect overlapping with the new DBLs. This assumption is valid because differences between the new DBLs and its temporary segments are much smaller than differences between the new and original DBLs. Note that Fig. 5 shows maximum and average residuals while Fig. 7 shows annual maxima and averages of the maximal residual within all windows.

Conclusions

Traditionally definitive geomagnetic observatory data are published annually with a delay from several months up to more than a year. The reasons are either that it may be difficult to produce high-quality recordings (visually checked, de-spiked, corrected for jumps, etc.) soon after acquisition due to the lack of observatory staff and processing protocols or simply because the calculation of definitive baselines and calibration of recorded data on yearly basis is currently standard procedure.

In this study, we have investigated the possibility to produce definitive observatory data more continuously with

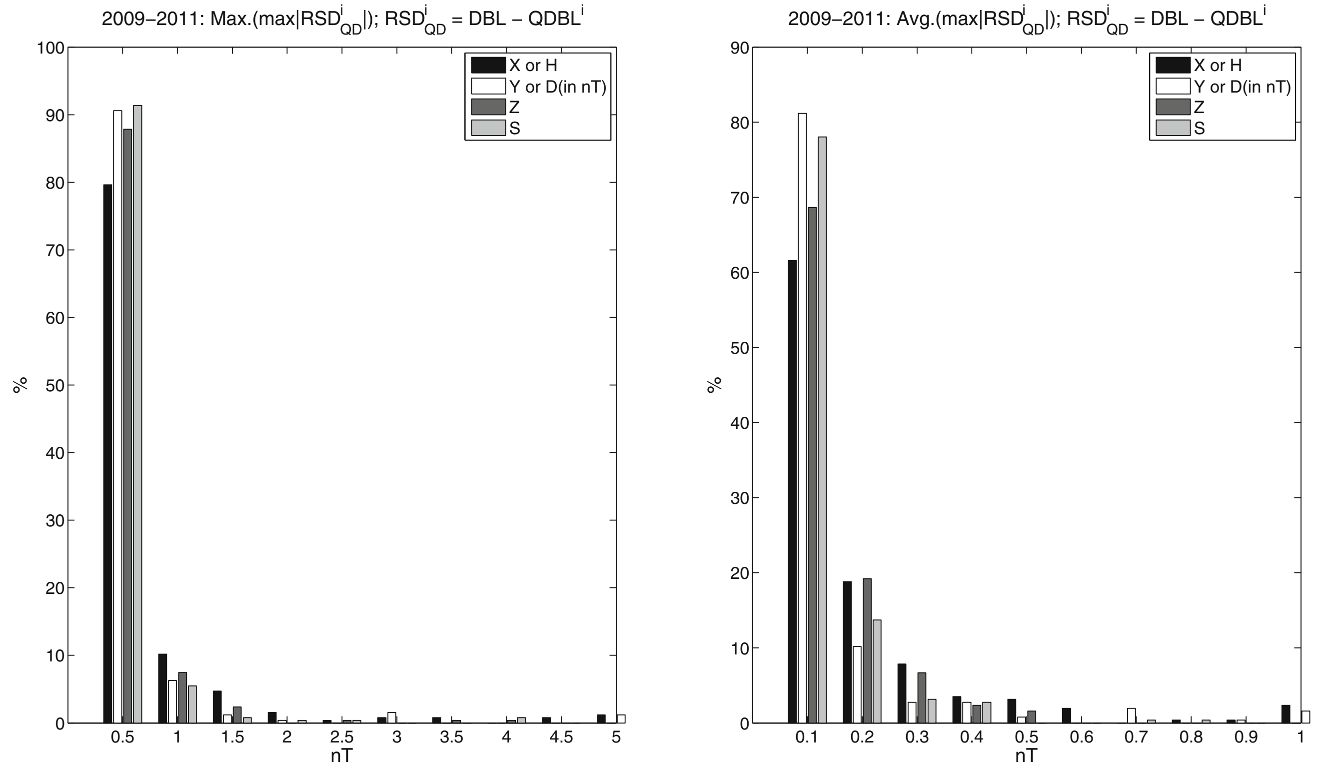


Fig. 7 *Left* annual maxima of the maximal residuals between the new DBL and its temporary baseline segments (QDBL), i.e. each maximal residual correspond to one baseline window. *Right* annual averages of the same residuals presented on the left subplot

smaller time lags. We have proposed a new method for reliable baseline adoption in a nearly fully automatic manner. Definitive baselines can be estimated by pieces from a sufficient number of observations in a way that the future results will not affect the shape of baselines from the past. This way it is also much easier to fit flexible baselines in case of rapid variation of the absolute observations within some period without overfitting surrounding observations that are stable. By re-processing the published baseline observations of the INTERMAGNET observatories from 2009 to 2011, we have shown that DBLs calculated with the new method in most cases give the same or better results (as indicated by Delta F and R^2) than DBLs determined by experienced observatory staff. However, in some cases manual intervention is necessary to obtain good results with the new method. We plan to develop a graphical user interface (GUI) that will allow others to use the method with options for manual intervention to improve results. For example, before fitting the user will be able to remove big outliers, define knots (i.e. window edges) differently in each component and modify the automatically determined smoothing parameter within the window and for each component separately. (For availability of GUI, please contact the corresponding author).

Most results in this study are obtained relying on the Delta F check. Continuous and accurate scalar observations are essential to obtain correct scalar F at the reference site to obtain reliable Delta F. Due to increased scattering or absolute inaccuracy in many cases, the proposed method estimates scalar baselines as a piecewise constant functions that do not fit observations. Examples similar to one shown in Fig. 2 (results obtained with the proposed method) can be also found on INTERMAGNET DVDs 2009–2011. Thus, the best observatory practice would be to set scalar baseline to zero, i.e. to perform continuous scalar measurement at the reference site (except during manual DI-flux observations). Alternatively, the gradiometer difference between the scalar and reference site, which is required for processing of the DI-flux observations and obtaining accurate scalar observations, should be regularly updated after each absolute DI-flux observation.

We have demonstrated that most of INTERMAGNET observatories should be able to distribute definitive data relatively soon after acquiring recent absolute measurements. Availability of the best available data, i.e. D data, within less than a year will be highly appreciated from the side of the observatory data users. With weekly absolute observations, most observatories should be able to produce their definitive data more continuously with a delay between 2 and 3 months. Increasing the number of observations, or shortening the window size, could significantly shorten this delay. In addition, with the proposed technique maximal discrepancies between D and QD data (which can be available within a few days/weeks) rarely exceed 0.3 nT.

The fact that in many cases the new method gives good results without human intervention paves the way towards automating the whole data processing chain, and towards fully automated observatories when automated absolute instruments become supplement equipment or the only absolute instrument at remote (unmanned) observatories (Korte et al. 2009; Mandić et al. 2016). These instruments will provide information about the absolute values of the field and variometer base values much more frequently than today. The increase in the number of automatically collected observations (let say every 30 min, Gonsette et al. 2013) will require a routine for baseline fitting and calibration of variometer data continuously in near-real time. However, the automatic instruments show a larger scatter and we have to expect increased number of outliers. Thus, an automated processing protocol should be robust to outliers. In future work, we will show that our new method produces robust and reliable results with data from an automated absolute instrument producing many absolute observations per day, so that final definitive data could be available within a few days from acquisition.

Acknowledgements We would like to thank two anonymous reviewers for constructive suggestions that helped us to improve the quality of the paper. The results presented in this paper rely on the data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (http://www.intermagnet.org). This work was supported by the DAAD (No. 91529536-50015537) and ERASMUS (No. ERA-2013-166) exchange programme.

References

Clarke E, Baillie O, Reay SJ, Turbitt CW (2013) A method for the near real-time production of quasi-definitive magnetic observatory data. Earth Planets Space 65(11):1363–1374. doi:10.5047/eps.2013.10.001

De Boor C (1978) A practical guide to splines. Springer, New York

De Boor C (2003) Spline toolbox for use with MATLAB, The MathWorks, Inc

Friis-Christensen E, Lühr H, Hulot G (2006) Swarm: a constellation to study the Earth's magnetic field. Earth Planets Space 58(4):351–358. doi:10.1186/BF03351933

Gonsette A, Rasson J, Marin JL (2013) AUTODIF: automatic absolute DI measurements. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 16–19

Jankowski J, Sucksdorff C (1996) Guide for magnetic measurements and observatory practice. International Association of Geomagnetism and Aeronomy, Boulder

Korte M, Mandea M, Linthe HJ, Hemshorn A, Kotzé P, Ricaldi E (2009) New geomagnetic field observations in the South Atlantic Anomaly region. Ann Geophys Italy 52(1):65–81

Korte M, Brunke HP, Bronkalla HP, Pulz E (2013), Status of the geomagnetic automated system GAUSS. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 20–23

Lesur V, Macmillan S, Thomson AWP (2006) Deriving main field and secular variation models from synthetic Swarm satellite and observatory data. Earth Planets Space 58(4):409–416. doi:10.1186/BF03351937

Macmillan S, Olsen N (2013) Observatory data and the Swarm mission. Earth Planets Space 65(15):1355–1362. doi:10.5047/eps.2013.07.011

Mandea M, Korte M (2011) Geomagnetic Observations and Models, IAGA Special Sopron Book Series, Springer, Dordrecht, DOI: 10.1007/978-90-481-9858-0

Mandić I, Vujić E, Heilig B, Pelajić I, Herak D (2016) Recent efforts toward the establishment of the Lonjsko Polje geomagnetic observatory. Acta Geophys 64(5):1311–1339. doi:10.1515/acgeo-2016-0051

Matzka J (2013) Preparation of quasi definitive (QD) data for the observatories Narsarsuaq, Qeqertarsuaq and Tristan Da Cunha. In: Proceedings 15th IAGA Workshop on Geomagnetic Observatory Instruments, Data Acquisition, and Processing, Boletin ROA, No. 03/13, 50–53

Olsen N, Haagmans R, Sabaka TJ, Kuvshinov A, Maus S, Purucker ME, Rother M, Lesur V, Mandea M (2006) The swarm end-to-end mission simulator study: a demonstration of separating the various contributions to Earth's magnetic field using synthetic data. Earth Planets Space 58(4):359–370. doi:10.1186/BF03351934

Peltier A, Chulliat A (2010) On the feasibility of promptly producing quasi-definitive magnetic observatory data. Earth Planets Space 62(2):e5–e8. doi:10.5047/eps.2010.02.002

Steiner F (1988) Most frequent value procedures. Geophys Trans 34(2–3):139–260

St-Louis B (2012) Intermagnet technical reference manual (Version 4.6)