Estimating error statistics for Chambon-la-Forêt observatory definitive data

Vincent Lesur, Benoît Heumez, Abdelkader Telali, Xavier Lalanne, Anatoly Soloviev

Annales
Geophysicae

Estimating error statistics for Chambon-la-Forêt observatory definitive data

Vincent Lesur[1], Benoît Heumez[1], Abdelkader Telali[1], Xavier Lalanne[1], and Anatoly Soloviev[2,3]

[1]Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Univ. Paris Diderot, UMR 7154 CNRS, 75005 Paris, France
[2]Geophysical Center of the Russian Academy of Sciences (GC RAS), Molodezhnaya Str. 3, 119296 Moscow, Russia
[3]Schmidt Institute of Physics of the Earth of the Russian Academy of Sciences (IPE RAS), Bolshaya Gruzinskaya Str. 10-1, 123242 Moscow, Russia

Correspondence to: Vincent Lesur (lesur@ipgp.fr)

Abstract. We propose a new algorithm for calibrating definitive observatory data with the goal of providing users with estimates of the data error standard deviations (SDs). The algorithm has been implemented and tested using Chambon-la-Forêt observatory (CLF) data. The calibration process uses all available data. It is set as a large, weakly non-linear, inverse problem that ultimately provides estimates of baseline values in three orthogonal directions, together with their expected standard deviations. For this inverse problem, absolute data error statistics are estimated from two series of absolute measurements made within a day. Similarly, variometer data error statistics are derived by comparing variometer data time series between different pairs of instruments over few years. The comparisons of these time series led us to use an autoregressive process of order 1 (AR1 process) as a prior for the baselines. Therefore the obtained baselines do not vary smoothly in time. They have relatively small SDs, well below 300 pT when absolute data are recorded twice a week – i.e. within the daily to weekly measures recommended by INTERMAGNET. The algorithm was tested against the process traditionally used to derive baselines at CLF observatory, suggesting that statistics are less favourable when this latter process is used. Finally, two sets of definitive data were calibrated using the new algorithm. Their comparison shows that the definitive data SDs are less than 400 pT and may be slightly overestimated by our process: an indication that more work is required to have proper estimates of absolute data error statistics. For magnetic field modelling, the results show that even on isolated sites like CLF observatory, there are very localised signals over a large span of temporal frequencies that can be as large as 1 nT. The SDs reported here encompass signals of a few hundred metres and less than a day wavelengths.

Keywords. Geomagnetism and paleomagnetism (instruments and techniques)

1 Introduction

Geomagnetic field models have drastically improved over the last decades because of the availability of satellite data (e.g. Olsen et al., 2006, 2014; Lesur et al., 2008, 2015; Sabaka et al., 2015; Finlay et al., 2016). Observatory data are nonetheless necessary because satellites are never sampling the magnetic field at the same point twice, and that of course limits our ability to separate spatial and temporal variations of the magnetic field. To the contrary, in observatories the field is sampled at a single place and its temporal variations are continuously recorded on timescales ranging from seconds to decades. Therefore most of the modellers utilise observatory data to derive their field models – e.g. they are heavily used in the level 2 products of the European Space Agency magnetic satellite mission: Swarm (Macmillan and Olsen, 2013).

When using exclusively observatory data (e.g. Wardinski and Holme, 2006; Wardinski and Lesur, 2012; Lesur et al., 2017), or when combining them with satellite data to model the different contributions to the geomagnetic field, it is always difficult to estimate the level of noise in the data. This is because we are unable to model some contributions to the field that ultimately get into the error budget (e.g. Finlay et al., 2016). However, as the scientific community pro-

gresses in its modelling effort, it becomes increasingly difficult to separate genuine magnetic large- or medium-scale signals in observatory data from drifts and discontinuities generated by local signals or, possibly, observatory operations. In order to robustly model the rapid core magnetic field variations, or describe the external field variations during magnetic quiet times, the scientists, engineers or technicians in charge of observatory operations should provide their data with estimates of their accuracy. This is of course very challenging because observatory data are combinations of magnetic measurements, where calibration operations – the so-called baseline estimations, play a crucial role.

The goal of this paper is to describe an algorithm for baseline estimations that provides information on the observatory definitive data error statistics. Of course, as input this process needs an estimation of the accuracies of the different data acquired, and these depend on the observatory setup. The algorithm was applied to provide error estimates for the French national observatory in Chambon-la-Forêt (CLF) for the year 2016. This time period was chosen because three sets of instruments were available on site, recording simultaneously data for most of the year. It should be noted that the observatory setup has been significantly modified during that same year, introducing several jumps and difficulties. In the next section is described the instrument setting in CLF and then, in Sect. 3, is explained how the noise in each individual data type collected on site has been estimated. The fourth section is dedicated to the description of the baseline calculation. Finally, the results are discussed and our conclusions given in the last two sections.

2 Observatory setup

Chambon-la-Forêt magnetic observatory (CLF) has been running since 1936 as a replacement of the Val-Joyeux (1901–1935) and previously St. Maur (1883–1900) observatories. They had to be closed because of the increasing level of noise in the data due to the construction of urban electric train lines nearby the observatories.

In 1936 the instruments used to measure continuously the magnetic field were photographic recorders of the field direction, but at the beginning of the 1970s, digital instruments were installed. They became the reference instruments only in 1986 (Bitterly et al., 1986). Typically they consist in a pair of magnetometers: a "variometer" made of fluxgates along three orthogonal axes and an absolute scalar instrument measuring the strength of the field. These instruments are combined with a data recording system that also controls the electric power provided. The fluxgates and parts of the recording system are sensitive to temperature variations and possibly other environmental variables such as humidity. The output values of the fluxgates are therefore prone to drifts that have to be accounted for when deriving definitive (i.e. calibrated) observatory data. The observatory buildings are such that

variations of the temperature are damped, or actively controlled, for the drift of the instruments to remain relatively smooth and small.

Another characteristic of the variometer is that it is most accurate on a limited range of values. Therefore, the fluxgates are traditionally oriented so that they are aligned along the magnetic north, east and vertical-down directions. The recorded values have therefore to be rotated in the local geodetic north, east, down reference frame. This is also part of the calibration process. In the local geodetic north, east, down reference frame, the magnetic field components are respectively called X, Y and Z components.

In 1936, the CLF observatory instruments were set in a vault. This vault was still housing in 2015 two pairs of instruments that are referenced here as CL1 and CL2. Their positions are indicated on the site plan – Fig. 1. The types and characteristics of the instruments are given in Table 1. The notation CLx refers to a combination of location, pillar and instrument, since a modification of any of these elements changes the characteristics of the measurements. Due to the ageing of the building, water infiltration became a difficult issue in recent years. Therefore, two new shelters were established to house the magnetometers. CL1 remained in the same place, in the vault, as a reference. CL2 instruments were moved to CL4. A new pair of instruments was set in CL3.

The CL1 has been continuously recording the magnetic field throughout 2014 to 2016, but pumping (on 17 March 2016 and 26 July 2016) was necessary to keep the water at low level inside the vault. The pumping generates only a small amount of noise visible in CL1 data, but it is likely that the water level itself has an influence on the recorded values. On 23 September 2016, perturbations in the data series are associated with building works done to restore the ventilation system of the vault.

The CL3 instruments were running for more than 6 months in 2015 – during that time we tested the performances of the new shelters. The CL3 instruments were set in their final positions on 8 January 2016 and not further disturbed before the end of November 2016.

The CL2 set of instruments that were originally in the vault a few metres away from CL1 were moved to their new position in CL4 on 17 February 2016. However, it took some time for the instruments to settle in their new locations. Variometers were re-oriented on 25 March 2016. We consider that the data are reliable for CL4 from 13 April 2016. The CL2 instruments up to 17 February 2016 and the CL4 instruments from 13 April 2016 are the CLF observatory main instruments – in between, data have been patched with the CL3 instrument outputs.

Absolute measurements used to calibrate variometer data are collected on a pillar roughly mid-way between CL1 and CL3 – see green labels in Fig. 1. During normal operations, absolute measurements are made twice a week, during the quietest days, early in the morning or late in the afternoon to avoid rapid variations of the magnetic field strength and

Table 1. List of instruments in place at CLF observatory in 2016. CL1–4 correspond to the different sets of instruments described in the main text; Abs. is the set of instruments used to acquire absolute data.

	Vector (Fluxgate)			Scalar (Overhauser)		
	Type	Start	End	Type	Start	End
CL1	triaxial fluxgate MAGNOLIA	03/1990	–	SM100	01/2002	–
CL2	VM-391 HOMOCENTRIC	10/2004	02/2016	SM100	01/2008	02/2016
CL3	VM-391 HOMOCENTRIC	04/2015	12/2016	SM100	04/2015	12/2016
CL4	VM-391 HOMOCENTRIC	02/2016	–	SM100	02/2016	–
	Theodolite		Fluxgate		Scalar	
Abs.	ZEISS-010A		BARTINGTON		SM90	

Figure 1. Plan of the Chambon-la-Forêt observatory site. Positions of the instruments used in this study are indicated. Distances are shown on the picture side in metres. Roughly 200 m separate CL3 from CL4.

direction associated with ionospheric signals. Again, the instruments used are listed in Table 1.

3 Observatory data

In order to estimate baseline values (i.e. calibration parameters) that are used to compute *definitive* observatory data values, the following data are available:

– Absolute measurements of the declination (Da), inclination (Ia) and strength (Fa) of the magnetic field on the absolute pillar. These values are measured nearly independently, and therefore it is assumed that the errors associated with these data are not correlated.

– Vector variometer data (X_v, Y_v, Z_v). These data are not fully independent since the three fluxgates are not exactly orthogonal, but the correlation of the associated errors is likely to be small and will be neglected here.

– Absolute scalar measurements in the variometer building (F_v). The errors for these data are clearly independent from other data types. It can be noted that these data are generally not used for the baseline estimations.

Our goal is to use all these data in a large inverse problem, and derive from them the baseline values together with an estimate of their variances. To be able to achieve this, we need first to estimate the raw data error statistics. We describe how that has been done in the two following subsections.

3.1 Absolute measurement

Absolute measurements are made twice a week in CLF, typically once in the morning and once in the late afternoon. Different types of errors are associated with these measurements

– e.g errors in pointing the reference mark, levelling errors, collimation errors, timing errors and errors associated with the changes in magnetic field strength and direction during measurements; the list is not exhaustive. Studying independently all these sources of errors is a complex task, so here we tried to estimate an overall error budget for absolute measurements.

Of course, the acquisition of absolute measurements has been going on for years and there is a large collection of data that can be used to estimate their noise level. Measurements are made of the declination, inclination and total intensity but are reported in yearbooks (http://www.bcmt.fr/bulletins.html) as declination, horizontal component and vertical component time series. When compared with variometer data and estimated baselines, these absolute measurements appear to be contaminated by errors with zero means and standard deviation (SD) values around 2 arcsec for the declination, 0.2 nT for the horizontal component and 0.1 nT for the vertical component. Estimates of the SD for the total intensity errors are also available, with values around 0.2 nT. Of course, these values vary along the years and also depend on the observers and equipment used. It should be first noted that these errors estimates are correlated and redundant since only three measurements are really done. Finally, these estimates rely on strong assumptions on the baseline behaviour (it has always been assumed that the baseline variations should be smooth in time). Although these values give us an idea of the absolute measurement noise level, alternative ways of deriving error SDs should be investigated.

In this work, the approach used to derive estimates of the SDs associated with absolute measurement errors rely on the record of several absolute measurements in a single day and the hypothesis that baselines have constant values over that day. To estimate these error SDs it is first necessary to compute for each absolute measurement a difference relative to variometer data. The daily mean of differences give an idea of the baseline value for that day. Deviations from that mean value lead to an estimate of the SD. There are two difficulties associated with these calculations. The first is due to the orientation of the variometers that is only partially known. To solve this problem it is assumed that the vertical component is truly aligned with the local vertical, and we impose that the daily mean declination given by the variometer matches the mean observed absolute declination – i.e. we apply a rotation on the variometer data around its Z axis so that the mean declination differences between absolute and variometer data are zero. The second difficulty is associated with the time necessary to take absolute measurements: absolute inclination data are obtained a few minutes (typically 3 min) after declination data. Furthermore, it takes a few minutes to make the four measurements entering in the estimation of the inclination and declination. Finally, total intensity measurements are made either during the inclination measurements, or after. These time shifts are corrected by tracking temporal variations of the horizontal and vertical components of

the variometer data, but they contribute significantly to the budget error of absolute measurements.

In 2016, we made twice a series of 24 absolute measurements within a 24 h period. The first series, hereinafter Series 1, started on 29 August 2016 at 15:07, and finished at 14:00 on 30 August 2016. The magnetic activity remained weak during this period with a Dst index staying between 2 and −7 nT. Over the 24 measurements of this series, two appeared to be anomalous and have not been used to estimate statistics. They correspond to time intervals when the magnetic field varied rapidly at CLF observatory. The second time series, hereinafter Series 2, started on 27 October 2016 at 07:30 and finished the same day at 11:30. The Dst index stayed between −34 and −45 nT. All 24 measurements are valid. For both time series we used variometer data from CL4. Results are summarised in Table 2. The absolute measurements over these 2 days have been made in the same way as absolute measurements are made throughout the year.

Although the estimated SDs actually represent the cumulated errors of the absolute measurements and those of the variometer data, in the following these estimates are used as the SDs of the absolute measurements alone. Furthermore, having less than 50 samples to estimate the SDs is clearly not enough, and other campaigns of measurements will have to be organised. In this work, the following values are used: $\sigma_{D_a} = 1.4\,10^{-3}$ deg (i.e. $\simeq 5$ arcsec), $\sigma_{I_a} = 0.6\,10^{-3}$ deg (i.e. $\simeq 2.2$ arcsec) and $\sigma_{F_a} = 0.3$ nT. These values correspond to the estimates derived from Series 2 even if the field was more active for that day. We made this choice because for this series the observations were made exclusively during daylight, as is usually the case during normal observatory operations. Finally, it should be pointed out that these values are valid only for CLF observatory and can be used here because the data set of absolute observations is homogeneous. However, in the algorithm described below there is no difficulty in introducing different SD values, if that is necessary, for some of the observations.

3.2　Variometer data

The fluxgates provide data in millivolts (mV). These data need to be scaled to obtain the usual nanotesla (nT). Further, the three orthogonal fluxgates forming the variometer are not exactly orthogonal, and this has to be corrected using so-called "non-orthogonality" angles. Scaling values and angles have to be estimated in an independent calibration process that we do not consider here. We note that any errors in these estimates affect the quality of the data and enter in their error budget, as would do a poor levelling of the variometer in the process described below.

Variometer and scalar instruments have a very good resolution but the vector components and total intensity data are only reported with 10 pT resolution. The true level of noise in the data is, however, much higher. To find SDs for these data errors, variometer data sets obtained with the different pairs
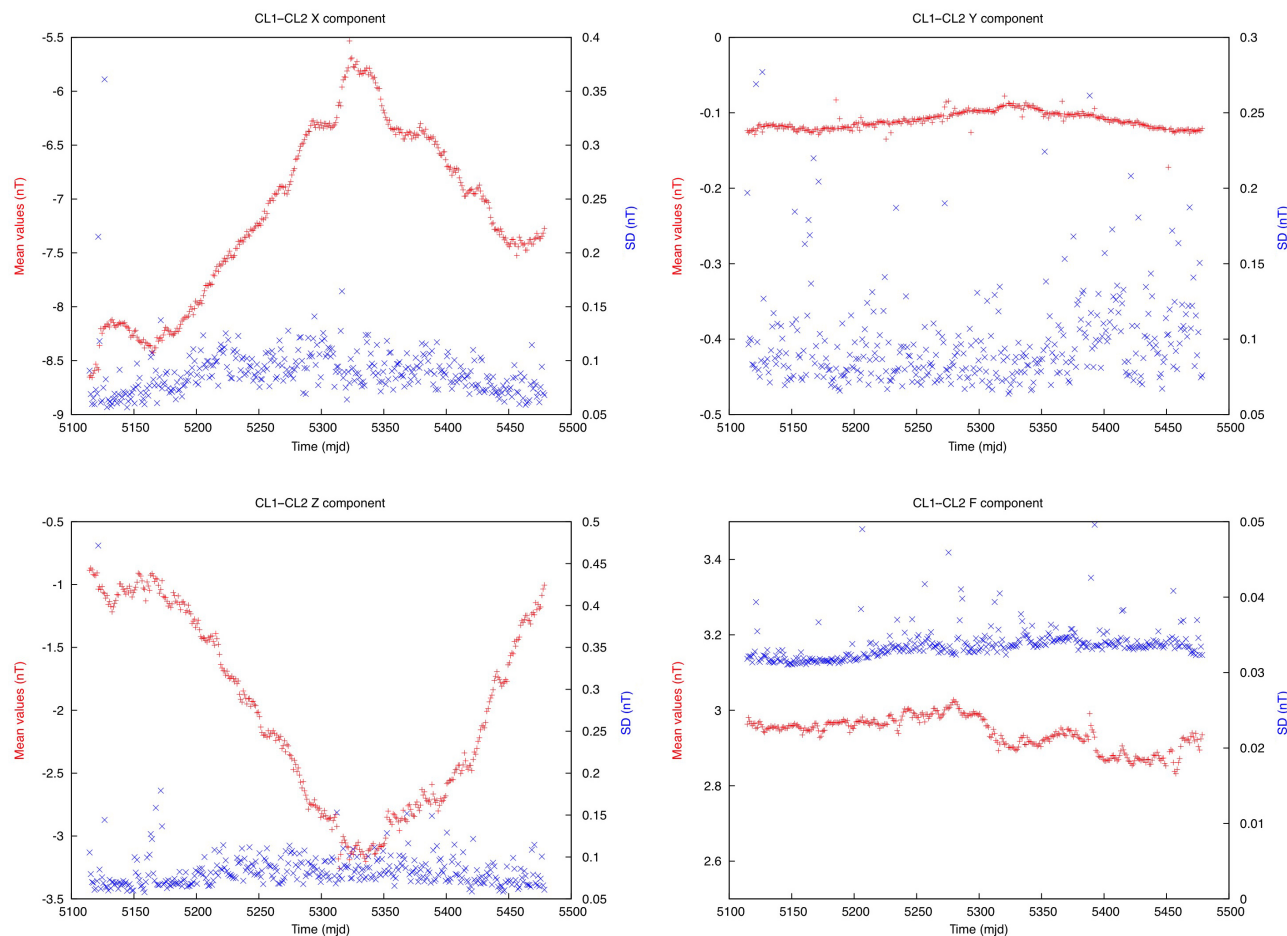
Figure 2. Mean differences (in red) and SDs (in blue) of the three components of CL1 and CL2 variometers and associated scalar instruments for year 2014. The red crosses can be seen as the cumulated baseline values of the CL1 and CL2 instruments.

Table 2. Statistics obtained from series of absolute magnetic data during 2016, assuming constant baselines for all data of a series. The number of valid data for each time series is N.

	Start	End	N	Declination (arcsec)		Inclination (arcsec)		Intensity (nT)	
				Mean	SD	Mean	SD	Mean	SD
Series 1	29/08 15:07	30/08 14:05	22	2.0	3.1	31.1	1.8	−15.2	0.3
Series 2	27/10 07:30	27/10 11:30	24	2.8	5.0	34.9	2.2	−15.5	0.3

of instruments available on CLF site are compared. These operations are easily done but again it is necessary to make the hypothesis that baseline values are constant over a day for all instruments. As for absolute data, variometer data have to be realigned so that they can be compared. Again it is assumed that vertical components are always truly vertical, and the X_v and Y_v components are rotated so that the measured declinations are the same for the two compared variometers. Ultimately, for each magnetic field component of the variometers, and for the total field intensity, we obtain per day

a mean value of the differences between the studied instruments pair, and a SD.

Figure 2 presents the results obtained for the CL1 and CL2 pair for the year 2014. The obtained mean values can be seen as the cumulated baseline drifts of the two sets of instruments. These mean values vary relatively smoothly in time, and are in the range of 3 nT maximum drift per year for the X and Z components. For the Y component, the variations remain very small. This is an effect of the way the two variometers are re-aligned. Differences between scalar instruments present a drift in time that does not exceed 100 pT,
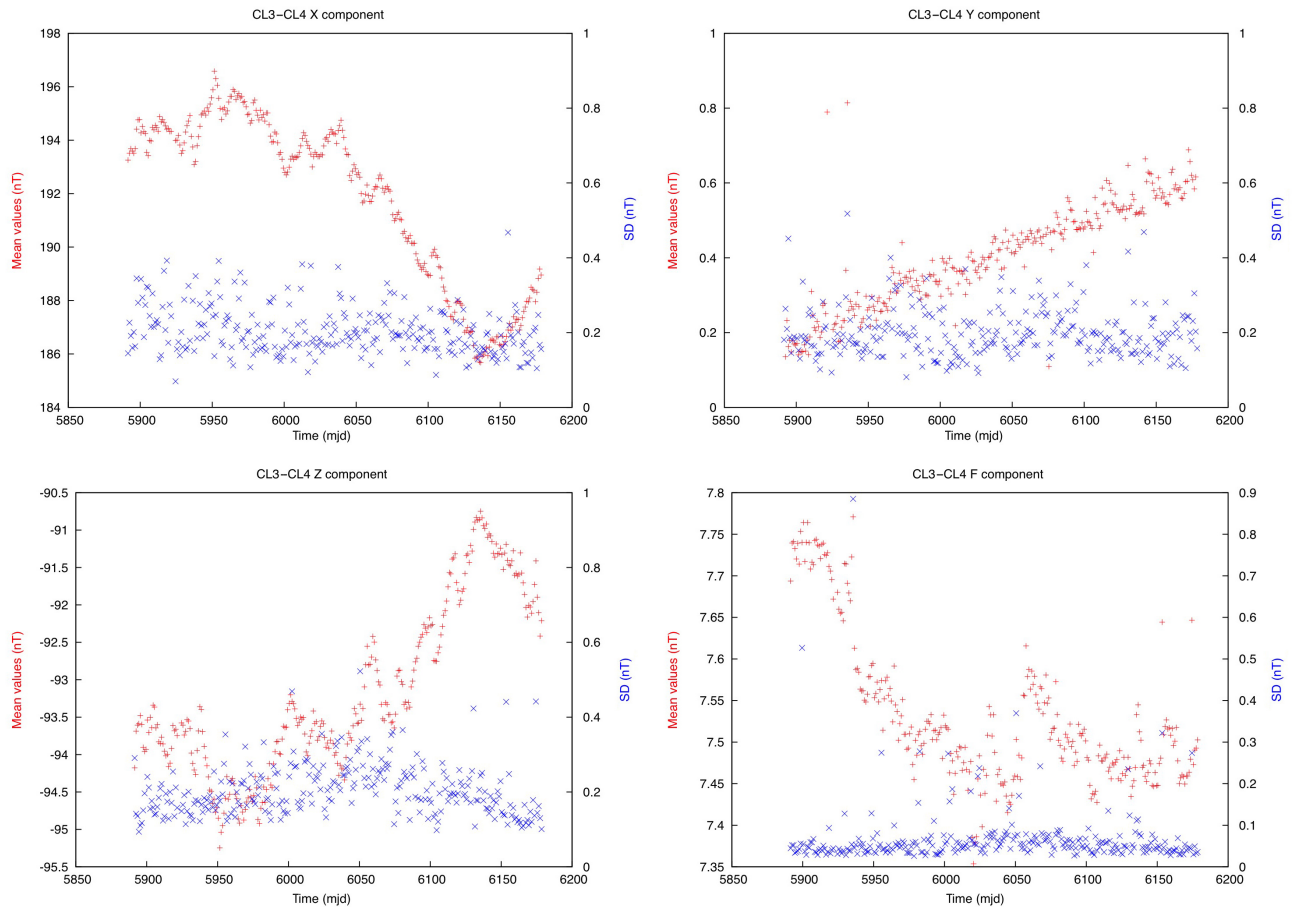
Figure 3. Mean differences (in red) and SDs (in blue) of the three components of CL3 and CL4 variometers and associated scalar instruments for year 2016. The red crosses can be seen as the cumulated baseline values of the CL3 and CL4 instruments.

with very small SDs, typically less than 50 pT. However, for these instruments there is a noticeable tendency to present large outliers, leading for some days to large SDs (outside the range of values presented here). Over that year the mean SDs are 89, 98, 84, 34 pT, for the X, Y, Z components and F respectively.

The results presented in Fig. 2 differ from those obtained in 2016 between CL3 and CL4, and presented in Fig. 3. Before day 104 of that year, the CL4 instruments were not yet in their optimal settings. It is nonetheless clear that the drifts between CL3 and CL4 instruments are much larger than those observed in 2014 between CL1 and CL2. Also, the cumulated baselines that represent the mean differences are no longer smooth. Finally, the SDs are at least twice as large than those obtained in 2014. The obtained averaged SDs are 204, 199, 204, 63 pT, for the X, Y, Z components and F respectively.

Intercomparisons between CL1, CL4 and CL3 show that the large drifts observed are mainly due to CL3 instruments. However, the increases of the SDs and the roughness of the cumulated baselines are associated with the increase distance

between sets of instruments. Clearly, the distance between the absolute pillar and the CL4 set of instruments is one of the main parameters controlling the noise level in CLF observatory data. In the remainder of this work, the following values are used for the SDs: $\sigma_{X_v} = \sigma_{Y_v} = \sigma_{Z_v} = 0.2$ nT, $\sigma_{F_v} = 0.06$ nT. As for the absolute measurements, the SD estimates provided here are valid only for CLF observatory, and are assumed constant through the year because the same instruments, in the same configuration, are used. Of course, as indicated before, the algorithm presented in the next section does not preclude modifying the error statistics of some of the provided data.

4 Baseline estimation

4.1 Theory

Baseline values are quantities that are added to variometer data to correct for potential drifts of the instruments and account for the site differences between the variometer building and the absolute pillar. The definitive data provided to the sci-

entific community are nothing else than variometer data that have been calibrated by adding baseline values, and rotated to be set in the local geodetic reference frame.

On the other hand, absolute data are snapshot values, contaminated by noise, of the direction and strength of the magnetic field as measured on the absolute pillar. Absolute data are declination, inclination and total intensity data that can be transformed in absolute X, Y, Z data through the relations

$$
\begin{aligned}
X_a &= F_a \cos(I_a) \cos(D_a) \\
Y_a &= F_a \cos(I_a) \sin(D_a) \\
Z_a &= F_a \sin(I_a).
\end{aligned}
\tag{1}
$$

Therefore the vectorial relation linking variometer data, baseline values and absolute data is

$$
\begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix} = \mathcal{R}_\theta \begin{bmatrix} X_v + X_\mathrm{b} \\ Y_v + Y_\mathrm{b} \\ Z_v + Z_\mathrm{b} \end{bmatrix} + \epsilon_{xyz},
\tag{2}
$$

where X_b, Y_b, Z_b are the baseline values estimated at the time the absolute data were recorded. $\mathcal{R}_\theta$ is the rotation matrix that rotates variometer data in the geodetic reference frame. Finally, ϵ is the cumulated noise contributions from the absolute and variometer data. This relation implies that the absolute total intensity data are such that

$$
F_a = \sqrt{(X_v + X_\mathrm{b})^2 + (Y_v + Y_\mathrm{b})^2 + (Z_v + Z_\mathrm{b})^2} + \epsilon,
\tag{3}
$$

and it follows that the absolute scalar data measured in the variometer building F_v are such that

$$
F_v = \sqrt{(X_v + X_\mathrm{b})^2 + (Y_v + Y_\mathrm{b})^2 + (Z_v + Z_\mathrm{b})^2} - F_\mathrm{b} + \epsilon_F,
\tag{4}
$$

where F_b is a site difference, but we refer to it as a baseline value for the variometer scalar data because it varies with time – e.g see Figs. 2 and 3. The misclosure error ϵ_F is the cumulated contributions of the F_v, X_v, Y_v, Z_v data errors. It is often called ΔF by the observers.

In the following, the relations (2) and (4) are used to derive estimates of the baseline values X_b, Y_b, Z_b and F_b. To solve this inverse problem, the same hypothesis as for the SD estimation is considered and it is assumed that a good description of baseline evolution is possible by taking one constant value per day. During a year there are therefore 365 (or 366 for leap years) unknowns to estimate for each baseline. In vector notation, all these values are organised in a vector $\boldsymbol{m}$ of maximum length $M = 4 \times 366$.

There are typically two sets of absolute measurements per week, and one value of F_v every second. There is no need to have so many F_v values, and to have an inverse problem of acceptable size, the F_v data are decimated to one value per day. All these data are collected in a data vector $\boldsymbol{d}$. It is clear that the inverse problem is heavily underdetermined.

The relation between $\boldsymbol{m}$ and $\boldsymbol{d}$ is non-linear because of Eq. (4). An iterative approach is therefore necessary to solve the problem, and at each step small perturbations of the baseline values have to be estimated. Then Eq. (2) becomes

$$
\begin{aligned}
\begin{bmatrix} \delta X_a \\ \delta Y_a \\ \delta Z_a \end{bmatrix} &= \begin{bmatrix} X_a \\ Y_a \\ Z_a \end{bmatrix} - \mathcal{R}_\theta \begin{bmatrix} X_v + X_\mathrm{b}^e \\ Y_v + Y_\mathrm{b}^e \\ Z_v + Z_\mathrm{b}^e \end{bmatrix} \\
&= \mathcal{R}_\theta \begin{bmatrix} \delta X_\mathrm{b} \\ \delta Y_\mathrm{b} \\ \delta Z_\mathrm{b} \end{bmatrix} + \epsilon_{xyz},
\end{aligned}
\tag{5}
$$

where $X_\mathrm{b}^e, Y_\mathrm{b}^e, Z_\mathrm{b}^e$ are estimates of the baseline values. The equation corresponding to relation (4) is

$$
\begin{aligned}
\delta F_v &= F_v - F_v^e + F_\mathrm{b}^e \\
&= \frac{X_v + X_\mathrm{b}^e}{F_v^e} \delta X_\mathrm{b} + \frac{Y_v + Y_\mathrm{b}^e}{F_v^e} \delta Y_\mathrm{b} + \frac{Z_v + Z_\mathrm{b}^e}{F_v^e} \delta Z_\mathrm{b} - \delta F_\mathrm{b} + \epsilon_F,
\end{aligned}
\tag{6}
$$

where $F_v^e = \sqrt{(X_v + X_\mathrm{b}^e)^2 + (Y_v + Y_\mathrm{b}^e)^2 + (Z_v + Z_\mathrm{b}^e)^2}$ and F_b^e is an estimate of the baseline value for F_v. In vector notation these equations correspond to a linear problem:

$$
\delta \boldsymbol{d} = \mathbf{H} \cdot \delta \boldsymbol{m} + \epsilon.
\tag{7}
$$

This underdetermined problem is solved in the usual way (see Eq. 41 in Tarantola and Valette, 1982):

$$
\delta \boldsymbol{m} = \mathbf{C}_m^e \mathbf{H}^t \left(\mathbf{H} \mathbf{C}_m^e \mathbf{H}^t + \Sigma_d \right)^{-1} \delta \boldsymbol{d},
\tag{8}
$$

where the superscript t is indicating a transpose matrix. $\mathbf{C}_m^e$ is the prior covariance matrix of the model perturbation $\delta \boldsymbol{m}$, whereas Σ_d is the covariance matrix of the data errors. This solution fits the data to its expected level while minimising the norm: $\Phi = \delta \boldsymbol{m}^t \cdot \mathbf{C}_m^{e\ -1} \cdot \delta \boldsymbol{m}$. Since the problem is underdetermined, the definition of the prior covariance matrix of the model perturbation has a very strong impact on the solution. The matrices $\mathbf{C}_m^e$ and Σ_d are described in the next subsections. The information carried by the data allows updating the prior covariance matrix of the model perturbation using the relation

$$
\mathbf{C}_m = \mathbf{C}_m^e - \mathbf{C}_m^e \mathbf{H}^t \left(\mathbf{H} \mathbf{C}_m^e \mathbf{H}^t + \Sigma_d \right)^{-1} \mathbf{H} \mathbf{C}_m^e,
\tag{9}
$$

where $\mathbf{C}_m$ is the posterior covariance matrix of the model perturbation $\delta \boldsymbol{m}$ (see Eq. 42 in Tarantola and Valette, 1982).

Since the inverse problem is non-linear, the matrix $\mathbf{H}$ depends on the a priori baseline values $X_\mathrm{b}^e, Y_\mathrm{b}^e, Z_\mathrm{b}^e, F_\mathrm{b}^e$. Regrouping these values in the vector $\boldsymbol{m}^e$, the iterative process leading to our best estimate of the baseline values $\boldsymbol{m}^*$ starts by choosing an initial vector $\boldsymbol{m}^e$ and defining the matrix $\mathbf{C}_m^e$. Both quantities are updated for the next iteration by estimating $\boldsymbol{m} = \boldsymbol{m}^e + \delta \boldsymbol{m}$ and $\mathbf{C}_m$ through Eqs. (8) and (9). This iterative process stops when the baseline values do not change significantly from one iteration to the next.

4.2 Prior covariance matrix of the baselines

In this section we explain how we set the prior covariance matrix for the baselines: $\mathbf{C}_m^e$. When starting this work, it was planned to derive the prior covariance matrix for the new baselines from the baselines calculated over the last few years. However, this appears to be unwise because comparison of the CL3 and CL4 data in Fig. 3 suggests that the baseline curves, which can be seen as the red dots, should not be smooth. Therefore, our hypothesis is that the baselines are autoregressive processes of order 1 (AR1 processes). Such a process follows the recursive equation

$$V_{i+1} = \alpha + \beta V_i + \gamma, \tag{10}$$

where the V_i are elements of a time series, α is a constant and γ is a random variable. The scaling β can be related to the time decay of an exponential: $\beta = 1 - \frac{\delta t}{\tau}$, where δt is the time interval between two samples of the time series, and τ is the time decay – see for example Evensen (2003). In this application τ is set to 50 days. It mainly controls how fast the time series of the baseline values goes back to its initial values, set in $\boldsymbol{m}^e$, in the absence of absolute data.

The correlation between two points of a time series following an AR1 process is

$$C_{i,j} = \beta^{|i-j|}. \tag{11}$$

For the specific case of baseline estimation, there are four series of baseline values (X_b, Y_b, Z_b and F_b). Under the hypothesis that these series are not correlated, the prior correlation matrix $\mathbf{C}_m^e$ is block diagonal if the unknowns are properly organised in the vector $\delta \boldsymbol{m}$. In each of these blocks the correlation coefficients take the form of Eq. (11), but are scaled such that the prior on the baseline variation amplitudes corresponds to chosen values. Here, these values are the observed variations of the baseline estimates (the red dots) in Fig. 2 around their average – i.e. variations are expected to be of the order of 1 nT for X_b, Y_b and Z_b. Of course, in Fig. 2, the variations in the Y component are much smaller but, as explained in Sect. 3.2, that is likely an effect of the data processing. For F_b values, the variations should be much smaller, around 0.5 nT (again, see Fig. 2).

The hypothesis that the different baseline time series are not correlated may not be valid, particularly if the baseline variations are due to temperature changes. If the baseline are correlated, the matrix is no longer block-diagonal. In principle the covariance matrix should carry the prior information we have regarding the baselines. The one we used here is particularly simple as we know little about what should be the baseline temporal behaviours at CLF observatory.

4.3 Covariance matrix of the data errors

In this section we show how the covariance matrix for the data errors Σ_d is defined. In Sect. 3, SD values were set for the errors in D_a, I_a, F_a and X_v, Y_v, Z_v, F_v data. These quantities are listed again below (angle values are given in degrees but are transformed into radians to be used in the equations):

$$\sigma_{D_a} = 1.4\ 10^{-3}\ \text{deg} \quad \sigma_{I_a} = 0.6\ 10^{-3}\ \text{deg} \quad \sigma_{F_a} = 0.3\ \text{nT}$$
$$\sigma_{X_v} = 0.2\ \text{nT} \quad\quad\quad \sigma_{Y_v} = 0.2\ \text{nT} \quad\quad\quad\quad \sigma_{Z_v} = 0.2\ \text{nT}$$
$$\sigma_{F_v} = 0.06\ \text{nT}$$

It is assumed that these errors are uncorrelated between themselves and in time. Therefore, the three-component variometer data X_v, Y_v, Z_v, measured at a given time, have a diagonal 3×3 error correlation matrix $\mathbf{C}_v$ in the instrument reference frame, with $\sigma_{X_v}^2$, $\sigma_{Y_v}^2$, $\sigma_{Z_v}^2$ on its diagonal. In the geodetic north, east, vertical-down reference frame, this covariance matrix is no longer diagonal and is calculated by

$$\mathbf{C}_{xyz} = \mathcal{R}_\theta\ \mathbf{C}_v\ \mathcal{R}_\theta^t, \tag{12}$$

where $\mathcal{R}_\theta$ is defined in Eq. (2).

Similarly, the absolute data D_a, I_a, F_a, measured at a given time, have a diagonal covariance matrix $\mathbf{C}_{\text{dif}}^a$ with $\sigma_{D_a}^2$, $\sigma_{I_a}^2$, $\sigma_{F_a}^2$ on its diagonal, but the covariance matrix of the absolute vector element X_a, Y_a, Z_a is such that

$$\mathbf{C}_{xyz}^a = \mathcal{T}\ \mathbf{C}_{\text{dif}}^a\ \mathcal{T}^t, \tag{13}$$

where $\mathcal{T}$ is the matrix that transforms small perturbations in declination, inclination and total intensity, in perturbations of vector elements:

$$\mathcal{T} = \begin{bmatrix} -Y_a & -\dfrac{Z_a X_a}{H_a} & \dfrac{X_a}{F_a} \\[2mm] X_a & -\dfrac{Z_a Y_a}{H_a} & \dfrac{Y_a}{F_a} \\[2mm] 0 & H_a & \dfrac{Z_a}{F_a} \end{bmatrix}. \tag{14}$$

It follows that the 3×3 covariance matrix of errors associated with δX_a, δY_a, δZ_a in Eq. (5) is

$$\mathbf{C}_{\delta x \delta y \delta z}^a = \mathbf{C}_{xyz}^a + \mathbf{C}_{xyz}. \tag{15}$$

To estimate the variance associated with Eq. (6), we have first to estimate the variance of F_v^e errors. The latter is

$$\sigma_{F_v^e}^2 = \left(\frac{(X_v + X_b^e)\ \sigma_{X_v}}{F_v^e} \right)^2 + \left(\frac{(Y_v + Y_b^e)\ \sigma_{Y_v}}{F_v^e} \right)^2 + \left(\frac{(Z_v + Z_b^e)\ \sigma_{Z_v}}{F_v^e} \right)^2.$$

Therefore the variance of δF_v errors in Eq. (6) is

$$\sigma_{\delta F_v}^2 = \sigma_{F_v}^2 + \sigma_{F_v^e}^2. \tag{16}$$

Since the temporal correlation of the errors is ignored, the covariance matrix of the data errors Σ_d is block diagonal, where, depending upon a proper organisation of the data in the vector $\boldsymbol{b}$, the blocks are defined either by Eq. (15) or Eq. (16).
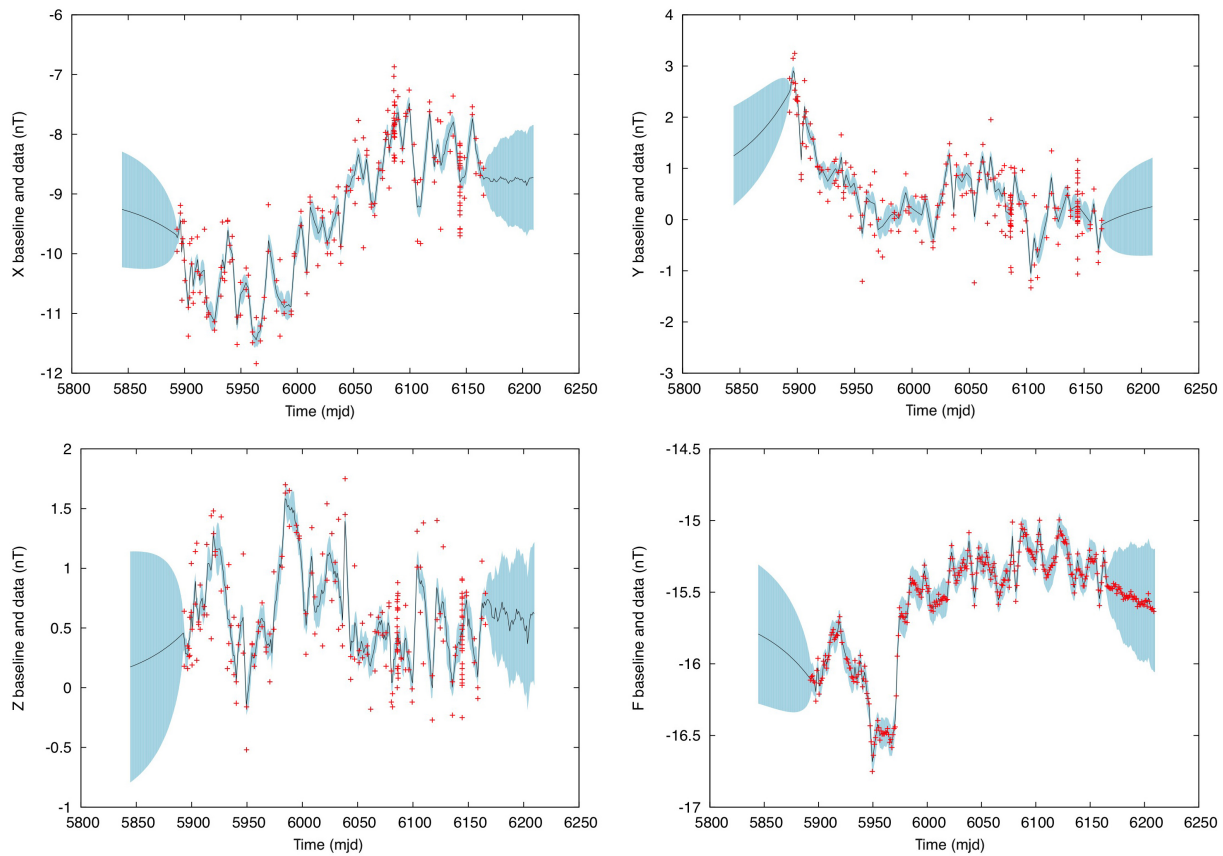
Figure 4. Estimated baseline values in black for the X, Y, Z components and F values. The 1σ interval is shown in light blue. Differences between absolute and variation data are shown in red. Series 1 and Series 2 were made during mjd $= 6086$ and mjd $= 6144$ respectively.

4.4 Application

The algorithm described above has been applied to estimate baseline values associated with CL4 for the year 2016. There were no CL4 variometer data before 17 February 2016, and absolute data after 16 November 2016 were not used. These features give the possibility to describe the effect of the lack of different types of data on the baseline values and their SDs.

For the iterative process, the initial values for the X_b, Y_b, Z_b and F_b baselines were constants set to -9.0, 0.5, 0.0, -15.6 nT respectively. To align variometer data to the geodetic local reference frame, the rotation angle was imposed to $\theta = 0.272465$ deg. There is some freedom on the way this angle can be defined. In principle, it should be co-estimated with the baselines such as to minimise the scatter of data. However, with the linearised approach we use here (see Eqs. 5 and 6), small variations of the baselines or of the angle cannot be distinguished. The angle value given here was set such that the mean declination value of the variometer data matches the mean absolute declination values measured.

Figure 4 shows the calculated baselines X_b, Y_b, Z_b and F_b, with their 1σ interval. The latter are always less than 1 nT,

and are reduced to less than 0.3 nT for X_b, Y_b, 0.23 nT for Z_b and 0.15 nT for F_b when absolute measurements are available. These baselines are rough compared to those normally set in CLF observatory. This roughness is controlled by the data and their associated variances; this is in agreement with our choice of prior covariance matrix for the model. Choosing a baseline by drawing a smooth curve through the set of data points may lead locally to differences as large as 1 nT compared with a rough baseline.

Up to 19 February (mjd $= 5893$) there are no data available. The estimated baselines slowly drift from a value close to their initial values on 1 January (mjd $= 5844$) to reach the first values imposed by the data on 19 February. Meanwhile the associated SDs decrease. The rate of change of the baselines during this period is fully controlled by the decay time τ imposed – here, $\tau = 50$ days. In a more general way, during a period without absolute data, the baselines tend to fall back to their initial values, and the SDs increase up to those set to scale the prior covariance matrix of the model (here 1 nT for X, Y, Z, and 0.5 nT for F).

After mjd $= 6164$ (16 November) no absolute data are used, but total intensities are still measured in the variometer shelter. These data partially constrain the baselines which
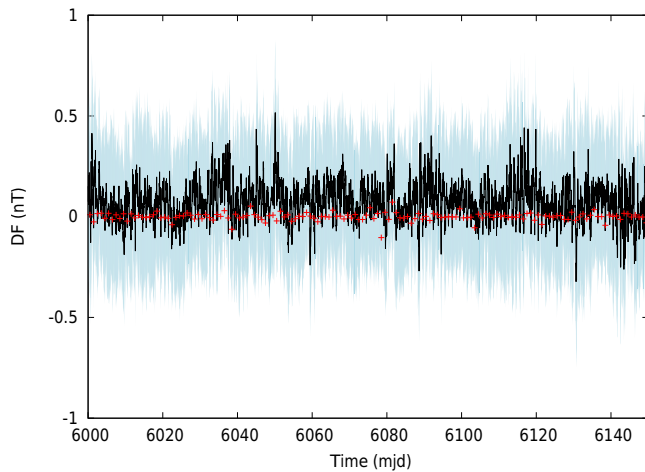
Figure 5. Estimates of the ΔF values (see Eq. 4) for 150 days in 2016. In black is one ΔF value per hour, in light blue is their 1σ interval, and in red are the ΔF values that have been minimised during the optimisation process leading to the baseline values.

slowly change in time in agreement with the prior set through the covariance matrix of the model. The Y baseline seems not to be affected by these data. This is not a surprise since the Y component is very small compared to X and Z components in CLF observatory and therefore it has only a very small contribution to the total intensity of the field.

Finally, it should be recalled that the CL4 data used here are not the observatory main instruments before 13 April 2016 (mjd = 5947) and therefore significant discontinuities are possible. The large jump in the F baseline on the mjd = 5971 (6 May 2016) is due to a change of the scalar instrument.

5 Discussion

The algorithm described and applied in the previous sections is giving as output a set of baseline values. The time series of the baseline values is not smooth in time. However, these values have been derived accounting for absolute and variometer data errors that have been estimated independently, so there is no reason to reject this rough baseline. In the following we verify that the baseline values are appropriate. First are given the ΔF values (see Eq. 4) obtained using CL4 data. Second the baseline values are compared with those obtained with the algorithm used in CLF observatory for several years. This algorithm is based on smoothing splines Silverman (1985). Then, finally, definitive values obtained with CL3 and CL4 sets of instruments are compared.

5.1 ΔF estimates

The ΔF estimates – i.e. the misclosure errors defined above in Eq. (4) – are often used in observatory operations to ver-

ify that the baseline calculations are correct. They also give an idea of the level of noise in the observatory data. The baseline values have been calculated for the CL4 data, and the estimated ΔF values are shown in Fig. 5. Values remain small, less than 0.5 nT, with large variances, but also with a clear bias. This is not a problem linked with the algorithm but comes from the fact that only one value per day of scalar variometer data (F_v) has been used to estimate the baseline. The values used correspond to ΔF estimates that are shown in red on Fig. 5. For these values, it is clear that there is no bias, and that ΔF estimates are very small. This suggests that it would be more appropriate to use several scalar variometer data per day to estimate the baselines. However, it is clear that the estimated ΔF are well inside their expected errors.

5.2 Comparison with other baseline estimates

In the algorithm that has been used this last decade in CLF observatory, the H_b and Z_b baseline values are calculated independently for the magnetic north direction (this is the X_v direction) and Z_v direction, respectively. Then, rather than estimating the baseline in the Y_v direction and applying a rotation of angle θ, a baseline D_b is estimated (in degrees) as a rotation angle around the Z_v axis. The relation linking the variometer reading and baseline values to the absolute values are

$$H_a = \sqrt{(X_v + H_b)^2 + Y_v^2}$$
$$D_a = D_b + \tan^{-1}(Y_v/(X_v + H_b)) \qquad (17)$$
$$Z_a = Z_v + Z_b.$$

For each set of absolute data, it is straightforward to derive baseline values H_b, D_b and Z_b. Over a year, a smooth spline is drawn through the cloud of H_b, D_b, Z_b values to serve as daily calibration parameter (Silverman, 1985). In this approach, the continuous recording of the total field intensity in the variometer house is used to check that the baseline estimates are correct.

In Fig. 6, the baselines calculated using both methods, for the CL4 set of instruments, are compared. General trends are similar, but differences are significant and within a range of $[-1 : 0.86]$ nT for H_b, Z_b and $[-6 : 11]$ arcsec for D_b. At several epochs, the baselines based on splines stay for a few days outside the 3σ interval of the baseline defined by the new algorithm. These large differences are only due to the assumption of smoothness that goes with the traditional way of computing baselines.

The results shown in Figs. 2 and 3 strongly support a rough baseline. This is not due to the variometer building or to the instruments themselves, but rather to the distance of the variometer building to the absolute pillar. The larger the distance, the rougher the baseline is likely to be. Of course this relationship is also dependent on the observatory environment.

Following our choice of prior on the baseline, a question arises on the information carried by observatory data. Most of the signal recorded can be associated with relatively large-
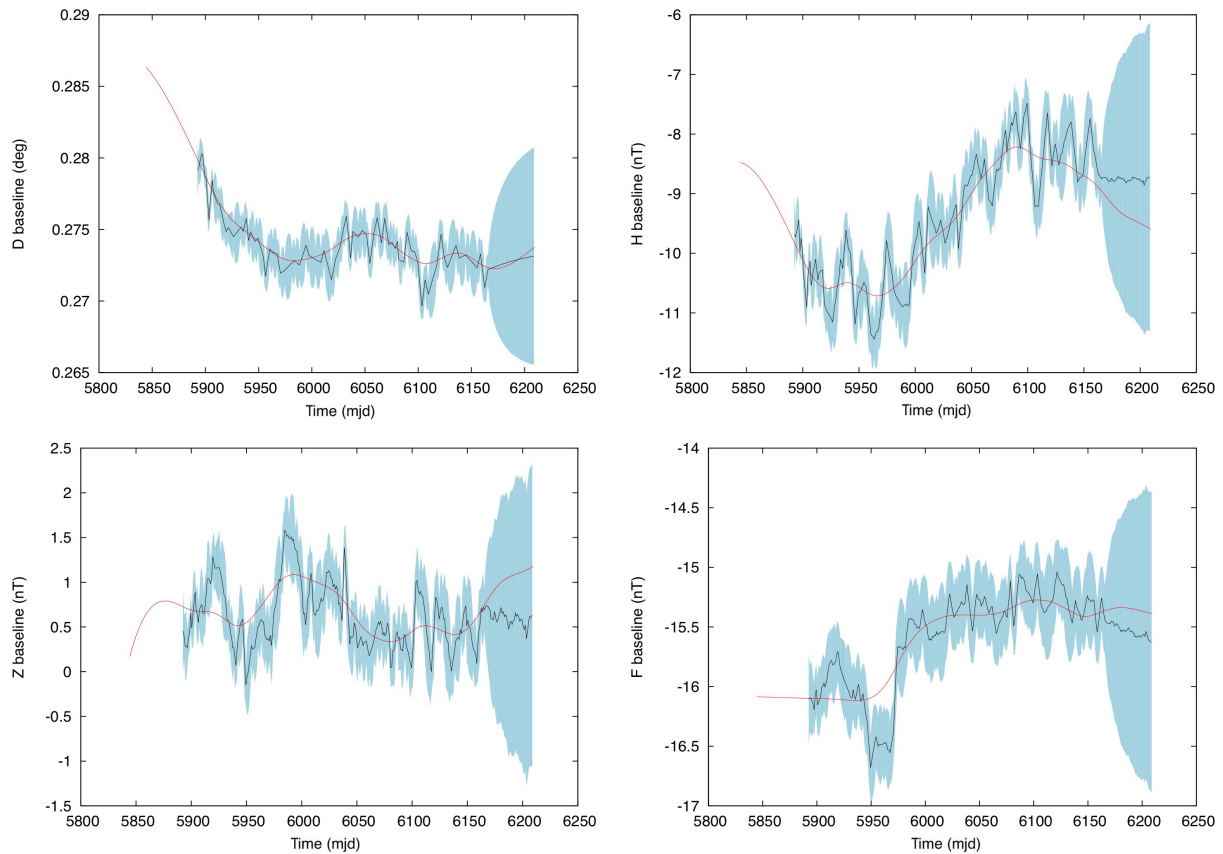
Figure 6. Estimated baseline values in black for the H, D, Z components and F values. Their 3σ interval shown in light blue. In red are shown the provisional baselines as estimated with smooth splines in CLF observatory.

scale sources of the magnetic field variations – e.g. the core, ionosphere and magnetosphere. Nonetheless, recorded signals at two places a few hundred metres away differ with amplitudes of the order of a few nanotesla – see e.g variations of the red dots in Fig. 3. It is very difficult for modellers to handle such signals unless they have some idea of their amplitudes and their frequency content. Such signals with large amplitudes on yearly timescales would be a difficulty for modelling rapid variations of the core field. Signals with significant amplitudes and frequency content of a few tenths of a hertz would not be a major problem for core field studies, but may be a limiting factor for external field or space weather studies. It would be beneficial for the scientific community to better characterise in space and time the local signals contributing to observatory data. The SDs values provided here give an idea of the strength of the signals for spatial and temporal wavelengths of 100–200 m and less than a day, respectively.

5.3 Comparison of CL3 and CL4 definitive data

The baseline estimation algorithm described in Sect. 4 has been applied to compute definitive values for the CLF observatory from both the CL3 and the CL4 set of instruments. In both cases a SD for the definitive data can be calculated from the combination of the baseline SDs obtained through our algorithm and the variometer data SDs estimated in Sect. 3.2. In principle, the two definitive data series obtained from CL3 and CL4 should nearly agree, and the quality of the baseline calculation can be estimated through the distribution of their differences.

For these days where both CL3 and CL4 data are available, the time series were subsampled to one value per hour leading to two sets of 6777 definitive vector data with the same time sampling. (The sets are large enough for the statistics ultimately obtained to be robust. There is no need to use the full 1 Hz data set.) For each sample, differences between definitive values obtained from CL3 and CL4 data were estimated together with a SD. Figure 7 presents the histograms of these differences, divided by their SDs (these ratios have no dimension). These histograms agree reasonably well with a Gaussian distribution of SD = 0.9. This value is lower than 1, in-
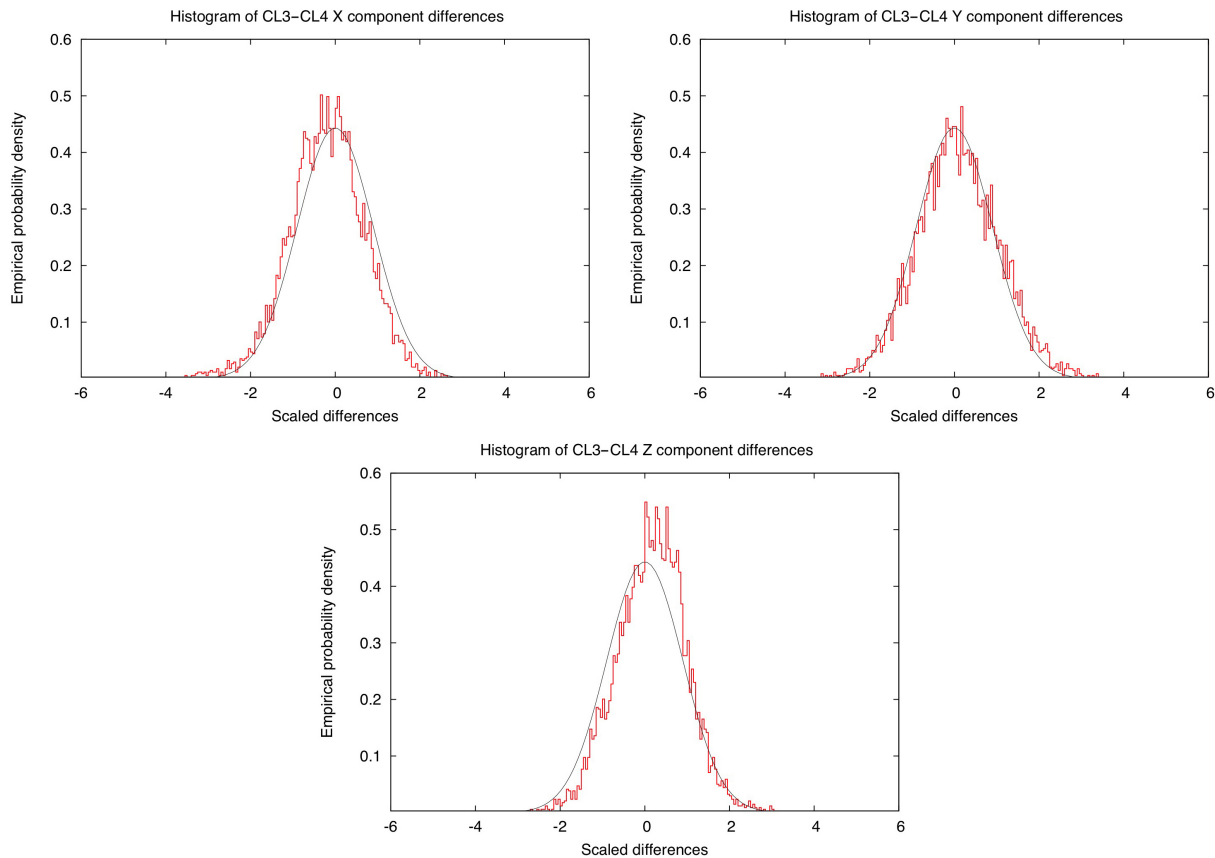
Figure 7. Rescaled histograms of weighted differences between CL3 and CL4 definitive vector data series. The weights used are the inverse of the estimated difference SDs. In black are shown the corresponding Gaussians for a SD of 0.9. Note that there are no units here since differences have been divided by their SDs and both are in nanotesla.

dicating that the SDs of differences, and therefore the SDs of the definitive values, have been slightly overestimated. This is possibly a consequence of the way SDs have been estimated for the absolute and variometer data, or a consequence of the relatively small number of absolute data available to estimate their variances.

In these histograms, the Y component seems to have nearly the expected distribution. For the X and Z, the main anomalies are associated with an excess of values for slightly negative scaled differences in the X direction, and slightly positive scaled differences in the Z direction. This is clearly due to the magnetic activity. The fact that the baseline is assumed to be a constant during a day has a strong influence here. On the same histograms, there is a small bias (< 0.2) that is due to the last 20 days of the time series where no absolute data were used (the baselines are then only controlled by the F_v values). Also, after mjd $= 6178$ there are still CL4 variation data – and therefore a small control on the baseline, whereas there are no further CL3 variation data. The absence of control on the baseline affects the baseline SDs and makes baseline drifts possible. It therefore generates the bias in the distributions of differences shown here. The deviations from the expected Gaussian distribution show the necessity to have a good control on the baseline evolution through regular absolute data measurements.

6 Conclusions

A new algorithm for producing calibrated definitive observatory data has been described. The algorithm sets a large inverse problem for the estimation of the baselines where all available data – i.e. absolute data, variometer data and scalar data acquired in the variometer building – are used as input. To proceed, data error SDs have been estimated. The SDs are realistic for the variometer data ($\sigma_{X_v} = \sigma_{Y_v} = \sigma_{Z_v} = 0.2$ nT, $\sigma_{F_v} = 0.06$ nT), but more work is required for the absolute data. For the latter, the data set from which the error statistics are estimated is, for now, too small. Furthermore, our estimates of these absolute data SDs also account partly for the variometer data errors. Overall, the absolute data error SDs ($\sigma_{D_a} = 5$ arcsec, $\sigma_{I_a} = 2.2$ arcsec, and $\sigma_{F_a} = 0.3$ nT) are probably overestimated as the statistics for the baseline, output of the inversion process, are also slightly overestimated.

Another important input to our algorithm is the prior covariance matrix of the baselines. To build it, we assumed that the baselines are stochastic auto-regressive processes of order 1 (AR1). These processes were tuned by choosing a decay time $\tau = 50$ days. This roughly corresponds to posterior least-squares estimates of this quantity although the four baselines do not give exactly the same value. The choice of an AR1 process for a baseline is supported by the observed daily mean differences between three sets of variometer instruments. Ultimately, our obtained baselines are not smooth in time.

As an output to our algorithm, each day a baseline value is set for the X, Y, Z directions, and F scalar data. These values are associated with a large baseline covariance matrix that is far from being diagonal – i.e. the baseline estimates are correlated in time. In a more mathematical view, our solution is actually a set of acceptable baseline values defined by a Gaussian distribution characterised by a mean and a covariance. The variance associated with one baseline value for a given day depends mainly on the density and quality of absolute data recorded in the few previous and following days. With systematically two absolute measurements per week, the baseline SDs for CLF observatory are less than 0.3 nT in the horizontal components, and slightly less for the vertical component. This variance drops to 0.15 nT for the F baseline. It is possible to combine these estimates with those of the variometer data SDs to derive definitive data SDs. These stay under 0.4 nT in the horizontal components, less than 0.3 nT in the vertical component. Clearly, at CLF observatory, the data errors are far from reaching the maximum of 5 nT error set in the INTERMAGNET standard (see page 29 of the INTERMAGNET technical manual v4.6 at http://www.intermagnet.org/publications/intermag_4-6.pdf).

The variance estimates are representative of the level of noise for the CLF observatory site, assuming constant baselines in a day. As explained above, baselines are not varying smoothly in time due to the distance between different data recording systems. This is an indication that there are relatively strong signals at short spatial wavelengths on the observatory site. These signals have a temporal wavelength longer than a day. Therefore they are not due to short-lived disturbance as, for example, a passing large vehicle. Such signals, like other *very small* scale signals in the data (where *very small* is 100–200 m in space, and less than a day in time) are accounted for in the error statistics of the baseline, and therefore of the definitive data. It is of course difficult in a global description of the magnetic field to handle such signals, or even larger-scale signals. There is, however, no consensus in the scientific community on what are the spatial and temporal minimum wavelengths under which observatory signals should be described as noise. These minimum wavelengths are probably not the same depending on the geophysical phenomenon studied. Therefore, It should be an objective of the scientists in charge of observatories to investigate the small-scale spatial and temporal content of the geomagnetic signal in the vicinity of the observatory site.

References

Bitterly, J., Cantin, J., Burdin, J., Schlich, R., Folques, J., and Gilbert, D.: Digital recording of variations in the Earth's magnetic field in French observatories: Description of equipment and results for the period 1972–1986, in: Proceedings of the International Workshop on Magnetic Observatory Instruments: Ottawa, Canada, 30 July–9 August 1986, 88, 75–81, 1986.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam., 53, 343–367, https://doi.org/10.1007/s10236-003-0036-9, 2003.

Finlay, C. C., Lesur, V., Thébault, E., Vervelidou, F., Morschhauser, A., and Shore, R.: Challenges handling magnetospheric and ionospheric signals in internal geomagnetic field modelling, Space Sci. Rev., 206, 157–189, https://doi.org/10.1007/s11214-016-0285-9, 2016.

Finlay, C. C., Olsen, N., Kotsiaros, S., Gillet, N., and Tøffner-Clausen, L.: Recent geomagnetic secular variation from Swarm and ground observatories as estimated in the CHAOS-6 geomagnetic field model, Earth Planet. Space, 68, 112 pp., https://doi.org/10.1186/s40623-016-0486-1, 2016.

Lesur, V., Wardinski, I., Rother, M., and Mandea, M.: GRIMM – The GFZ Reference Internal Magnetic Model based on vector satellite and observatory data, Geophys. J. Int., 173, 382–394, https://doi.org/10.1111/j.1365-246X.2008.03724.x, 2008.

Lesur, V., Whaler, K., and Wardinski, I.: Are geomagnetic data consistent with stably stratified flow at the core-mantle boundary?, Geophys. J. Int., 201, 929–946, https://doi.org/10.1093/gji/ggv031, 2015.

Lesur, V., Wardinski, I., Baerenzung, J., and Holschneider, M.: On the frequency spectra of the core magnetic field Gauss coefficients, Phys. Earth Plan. Int., in press, https://doi.org/10.1016/j.pepi.2017.05.017, 2017.

Macmillan, S. and Olsen, N.: Observatory data and the Swarm mission, Earth, Planet. Space, 65, 1355–1362, https://doi.org/10.5047/eps.2013.07.011, 2013.

Olsen, N., Lühr, H., Sabaka, T., Mandea, M., Rother, M., Tøffner-Clausen, L., and Choi, S.: CHAOS – A model of the Earth's magnetic field derived from CHAMP, Oersted, and SAC-C magnetic satellite data, Geophys. J. Int., 166, 67–75, https://doi.org/10.1111/j.1365-246X.2006.02959.x, 2006.

Olsen, N., Lühr, H., Finlay, C., Sabaka, T. J., Michaelis, I., Rauberg, J., and Tøffner-Clausen, L.: The CHAOS-4 Geomagnetic Field Model, Geophys. J. Int., 197, 815–827, https://doi.org/10.1093/gji/ggu033, 2014.

Sabaka, T. J., Olsen, N., Tyler, R. H., and Kuvshinov, A.: CM5, a pre-Swarm comprehensive geomagnetic field model derived from over 12 yr of CHAMP, Ørsted, SAC-C and observatory data, Geophys. J. Int., 200, 1596–1626, https://doi.org/10.1093/gji/ggu493, 2015.

Silverman, B.: Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion), J. R. Stat. Soc., 47, 1–52, 1985.

Tarantola, A. and Valette, B.: Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion, Rev. Geophys. Space Phys., 20, 219–232, 1982.

Wardinski, I. and Holme, R.: A time-dependent model of the Earth's magnetic field and its secular variation for the period 1980 to 2000, J. Geophys. Res., 111, B12101, https://doi.org/10.1029/2006GB004401, 2006.

Wardinski, I. and Lesur, V.: An extended version of the C^3FM geomagnetic field model – Application of a continuous frozen-flux constraint, Geophys. J. Int., 189, 1409–1429, https://doi.org/10.1111/j.1365-246X.2012.05384.x, 2012.