

FULL PAPER

Open Access



# On the feasibility of routine baseline improvement in processing of geomagnetic observatory data

Anatoly Soloviev<sup>1,2\*</sup> , Vincent Lesur<sup>3</sup> and Dmitry Kudin<sup>1,4</sup>

## Abstract

We propose a new approach to the calculation of regular baselines at magnetic observatories. The proposed approach is based on the simultaneous analysis of the irregular absolute observations and the continuous time-series  $\Delta F$ , widely used for estimating the data quality. The systematic  $\Delta F$  analysis allows to take into account all available information about the operation of observatory instruments (i.e., continuous records of the field variations and its modulus) in the intervals between the times of absolute observations, as compared to the traditional baseline calculation where only spot values are considered. To establish a connection with the observed spot baseline values, we introduce a function for approximate evaluation of the intermediate baseline values. An important feature of the algorithm is its quantitative estimation of the resulting data precision and thus determination of the problematic fragments in raw data. We analyze the robustness of the algorithm operation using synthetic data sets. We also compare baselines and definitive data derived by the proposed algorithm with those derived by the traditional approach using Saint Petersburg observatory data, recorded in 2015 and accepted by INTERMAGNET. It is shown that the proposed method allows to essentially improve the resulting data quality when baseline data are not good enough. The obtained results prove that the baseline variability in time might be quite rapid.

**Keywords:** Geomagnetism, Observatories, Instruments and techniques

## Introduction

The Earth's magnetic field (EMF) is the one of the most important sources of information about the physical processes occurring inside the Earth and in the circumterrestrial space. The data on the magnetic field's state and its temporal variations are registered using modern magnetometers installed on special satellites and ground-based magnetic observatories. Despite the advantages of satellite measurements, there are a number of limitations that do not allow us to abandon the classical magnetic field measurements of stationary observatories. The measuring satellites are relatively recent and inevitably have a limited lifetime. The typical duration of the continuous satellite data time series does not exceed

10 years, whereas the oldest geomagnetic observatories provide continuous series of observations lasting more than a 100 years. Such records are of exceptional value for the fundamental researches in the field of geomagnetism assuming the study of the evolution of the Earth's magnetic field and the associated dynamic processes in the outer core over longtime intervals. The development of science in the study of the magnetic fields of the Earth and the Sun requires a permanent improvement of the quality of the data provided by the observatories as the adequacy of reproducing the characteristics of the magnetic field outside the points of actual measurements using appropriate mathematical models depends directly on it.

Over the past decades, the highest international standard for the quality of geomagnetic observations was elaborated within the framework of the INTERMAGNET program (Love and Chulliat 2013). To date, this standard

\*Correspondence: a.soloviev@gcras.ru

<sup>1</sup> Geophysical Center RAS, Moscow, Russia

Full list of author information is available at the end of the article

corresponds to about 130 observatories united into a single network and providing continuous, quasi real-time data online (<http://www.intermagnet.org>). The methods for the data preparation established in INTERMAGNET are the de facto standard for any geomagnetic observatory included in the network. The quality of the magnetic field component records relates to the two main components: the hardware used for the measurements and the processing of the collected data. The main instrument of the magnetic observatory, providing the continuity of the measurements, is a quartz-of fluxgate-type vector magnetometer, also called a variometer. At the INTERMAGNET observatories, variation measurements are carried out with a sampling rate varying from  $\sim 250$  to 0.05 Hz. Despite the development of methods for measuring the magnetic field, the present high-precision vector magnetometers do not allow automatic measurements of the total values of the magnetic field components in an autonomous mode. Temperature changes, pillar movements, aging of electronics and many other factors inevitably influence the magnetic field vector measurements. To ensure the precision needed for contemporary researches, a periodic calibration of the variation magnetometer should be carried out at every observatory. Such calibration is provided as a result of observation of the absolute magnetic field values (absolute observations), performed by a trained specialist using the non-magnetic theodolite, on which a single-axis fluxgate magnetometer is mounted. The corresponding instrument is called an absolute magnetometer or a declinometer/inclinometer. A single absolute observation series is a fairly long routine process, which takes 10–15 min, at best; in addition, the result directly depends on the accuracy of the procedure performed. However, even with the strict fulfillment of all the requirements, the resulting measurements can be still burdened with inaccuracies due to increased geomagnetic activity or weather conditions during the absolute observations. The resulting measurements make it possible to calculate the calibration corrections, the so-called baseline values, for each component of the vector magnetometer (Jankowski and Sucksdorff 1996).

Spot (or observed) baseline values are calculated using the results of absolute observations, as well as spot values of the component variations and the field modulus (total intensity) registered with two independent instruments at times of absolute observations. Further, they are used to obtain a regular series of baseline values applied for the correction of 1-min (or 1-s) data, which are derived from continuous variometer recordings. According to the INTERMAGNET rules for the data processing, such regular series should represent daily values resulting from the interpolation of the observed baseline values. The

interpolation algorithm is selected by each observatory independently, ensuring the smoothness of the resulting baseline. Most observatories solve this problem by using cubic smoothing splines or the approximation with polynomials.

Nevertheless, the described approach has some serious disadvantages. First of all, application of interpolation to obtain the baseline values for the periods of the absence of the absolute observations does not take into account the physical effects caused by the behavior of the magnetic field during the corresponding time intervals. That is, the common agreement that the baseline variations should be described by smooth functions (quadratic polynomials, splines, etc.) based on the absolute observation values generally has no rigorous physical justification. At the same time, the field behavior between the time moments of absolute observations can be fully taken into account, as observatories normally provide continuous and independent registration of both relative variations of the field components and its modulus values. In the current observatory practice, when calculating the regular baselines, continuous records of the components and the modulus of the magnetic field are not used. Taking into account the detected EMF signals with the amplitude of several nT with a spatial scale of some hundred meters and less than 1-day duration (Lesur et al. 2017), this circumstance becomes especially critical. Indeed, the distance between the vector and the scalar magnetometers at observatory is often about a hundred meters, so such signals can affect the record of only one of the two instruments. Also, the selection of baseline values at each time moment is carried out for each component independently, which does not allow to take into account the accuracy of orientation and orthogonality of the vector magnetometer sensors. Another disadvantage of the widely accepted method is that it does not admit baseline value variations within a day, which may well occur in actual practice due to an abrupt temperature change in the pavilion or small-scale EMF signals described above. The mentioned disadvantages necessarily lead to the deterioration of the quality of the definitive data, which in turn introduces errors in model predictions built using observatory data, especially when studying rapid (shorter than a year) core field variations.

In the present paper, we propose a new approach to the calculation of regular baselines in which both of the mentioned disadvantages are eliminated. The proposed approach is based on the simultaneous analysis of the results of absolute observations and the values of the regular time series  $\Delta F = \Delta F(t)$  which is widely used for estimating the data quality (Reda et al. 2011). The latter represents a series of differences between the field modulus recorded with a scalar magnetometer and calculated

from the full component values, which are derived from the vector magnetometer data. In the proposed method, the systematic  $\Delta F$  analysis allows to take into account all available information about the operation of observatory instruments in the intervals between the moments of absolute observations. Of course, it implies a continuous recording of the absolute field modulus at observatory. To establish a connection with the observed baseline values, we introduce a function for approximate evaluation of the intermediate baseline values. This paper continues the series of studies set out in (Lesur et al. 2017), where a method is proposed for the estimation of the error statistics of resulting data by solving the inverse problem in a linear form. In our case, the solution of the inverse problem has an initial, nonlinear form.

In the first part of the paper, the description of the algorithm for the calculation of the regular baselines, based on the minimization of the proposed functional, is given. In the second part, we analyze the results of the algorithm operation by the examples of synthetic data. The third part is dedicated to the algorithm application to real data obtained at the Saint Petersburg INTERMAGNET observatory. In the two final sections, we provide the results of comparison between the 2015 definitive data obtained by the classical and proposed methods, as well as their discussion and conclusions.

### Calculation of regular baselines

The traditional method for the definitive data preparation does not require a continuously recording scalar magnetometer at the observatory. Despite this, many INTERMAGNET observatories (more than 85 according to the 2015 definitive data statistics) (<http://www.intermagnet.org>) perform continuous recordings of the total intensity of the magnetic field by means of a proton-type scalar magnetometer.

The difference  $\Delta F$  between the calculated value of the total Earth magnetic field intensity according to the baseline-corrected vector magnetometer data and the scalar magnetometer record is the one of the basic quality criteria provided by the observatory (Reda et al. 2011). Thus, the presence of approximate full values of the three components of the field vector and the time-synchronized values of its modulus allows the operational calculation of the  $\Delta F$  record. The corresponding time series reflects the consistency of the operation of vector and scalar magnetometers, which in turn enables continuous monitoring of the quality of work of the devices. At the same time, this approach ensures the quality control of the variometer to a greater extent because current scalar magnetometers are characterized by a long-term stable operation and practically no sensitivity to external climatic influences (Hrvoic and Newitt 2011). It should be

noted that such approach to quality control is not applicable for error detection in the variometer data resulting from whole and permanent motions like drifts due to pier inclinations or rotations. Also, since the proposed method for estimating baselines involves the  $\Delta F$  time-series analysis, it can be applied only to the data from the observatories that carry out the continuous registration of total intensity  $F$ . Usually, the vector and the scalar magnetometers sample geomagnetic field with different frequencies; therefore, they must first be brought to a single timeline by averaging. In particular, time-synchronized min data are calculated by averaging the original, noise-cleared data using a Gaussian filter (St-Louis 2012).

Let us analyze the calculation procedure of the intermediate (regular) baseline values using the proposed method. For a calculation, the following initial data sets are used:

1. Three orthogonal component variations of the EMF  $\vec{V} = (X_\nu, Y_\nu, Z_\nu)$ , measured with a vector magnetometer with a frequency  $l_1$ ;  
The EMF total intensity values EMF  $F_\nu$ , measured with a scalar magnetometer with a frequency  $l_1$ ;
2. The observed baseline values  $\vec{B}_0 = (X_b^0, Y_b^0, Z_b^0)$ , calculated using the results and at time moments of the absolute observations carried out with a non-regular frequency of an average value  $l_2 \ll l_1$ .

In the real practice of the geomagnetic observatory, the  $l_1$  value is within 1/60–1 Hz, and the  $l_2$  is about 1–2 measurements per week. The observed baseline values are preliminarily processed for the elimination of erroneous measurements leading, in particular, to spikes. Using the processed series, the range of acceptable baseline values is determined for each component. The similar procedure is carried out for all fragments of the observed baseline values separated by abrupt changes (jumps) due to modifications in the observatory environment; we give the algorithm description for one such fragment.

At the first stage, for the calculation of a regular series of approximate baseline values, a grid  $S = \left\{ X_b^i \Big|_0^I, Y_b^j \Big|_0^J, Z_b^k \Big|_0^K \right\}$  is constructed for three components within the ranges of acceptable values. Definition of the grid step is based on the required accuracy of the baseline calculation. The informal essence of the algorithm is in the selection of such a combination of baseline values for every time moment that (1) the resulting  $\Delta F$  tends to zero, (2) their discrepancy with the adjacent observed baseline values is minimal. Let us formalize the first criterion as a function  $G$  tending to zero, and the second one—as the minimization of the function  $A$ . We introduce the target function  $\phi$ , which is a linear

combination of the  $G$  and  $A$  functions connected with a weighting factor  $\lambda$ . For every time moment  $t$  in every node of the grid  $\vec{S}^{i,j,k} = (X_b^i, Y_b^j, Z_b^k)$ , the function  $\phi$  will have the following form:

$$\phi(t, \vec{S}^{i,j,k}) = \lambda G + (1 - \lambda)A, \quad \lambda \in [0, 1]. \quad (1)$$

To optimize the computational process, while calculating the  $\phi$  function, we set the time step equal to 1 h. For this, it is necessary to assume that the baseline values remain unchanged during an hour, which is quite permissible due to the assumption of the baseline constancy during a day accepted by INTERMAGNET (St-Louis 2012). Then, the  $G$  function will be calculated using the formula:

$$G = \sum_{1h} \left( \sqrt{(X_v + X_b^i)^2 + (Y_v + Y_b^j)^2 + (Z_v + Z_b^k)^2} - F_a \right)^2, \quad (2)$$

where  $F_a = F_v + F_b^0$  and  $F_b^0 = \text{const}$ —the difference of the total field intensity between the site of installation of the scalar magnetometer and the site of installation of the absolute magnetometer. Therefore, when analyzing the 1-min  $X_v, Y_v, Z_v$  and  $F_v$  data, at every step the  $G$  function is calculated from 60 values of the three components and the total intensity of the magnetic field.

The  $A$  function, which is responsible for the correspondence of the considered combination of baseline values to two neighboring observed values, has the following form for each time moment:

$$A = \left( |\vec{B}_0^p - \vec{S}^{i,j,k}| w_p \right)^2 + \left( |\vec{B}_0^n - \vec{S}^{i,j,k}| w_n \right)^2, \quad (3)$$

where  $\vec{B}_0^p$  is the previous observed baseline vector,  $\vec{B}_0^n$  is the next observed baseline vector,  $\vec{S}^{i,j,k}$  is the current vector of the grid,  $w_p$  is a factor of closeness to  $\vec{B}_0^p$ ,  $w_n$  is a factor of closeness to  $\vec{B}_0^n$ . Herein, we use the factors of closeness within  $[0, 1]$  that vary linearly:

$$w_p = 1 - \frac{\Delta t_p}{T}, \quad w_n = 1 - \frac{\Delta t_n}{T}, \quad (4)$$

where  $\Delta t_p$  is a distance in time from the current moment to the moment of the previous observed baseline vector,  $\Delta t_n$  is a distance in time from the current moment to the moment of the next observed baseline vector, and  $T = \Delta t_n + \Delta t_p$  is a distance in time between the previous and the next observed baseline vectors in relation to the current time moment. In principle, nonlinear factors of closeness can also be used, such as  $\frac{1}{1+\Delta t}$  or  $e^{-\Delta t/T}$ .

Hence,

$$\vec{B}(t) \stackrel{\text{def}}{=} (X_b, Y_b, Z_b) = \arg \min_{\vec{S}^{i,j,k}} \phi(t), \quad \forall t \in T, \quad (5)$$

where  $\vec{B}(t)$  is the desired baseline vector for the time moment  $t$  on a regular time interval  $T$  with a step of 1 h and “argmin” returns the value of  $\vec{S}^{i,j,k}$ , which minimizes  $\phi(t)$  over the set of candidates.

### Validation of the method on synthetic data

Let us analyze the application of the method to partially synthesized data. As the initial variation values  $\vec{V} = (X_v, Y_v, Z_v)$ , we use the variations of the three orthogonal field components registered by the vector magnetometer at the Saint Petersburg INTERMAGNET magnetic observatory (IAGA code SPG, Leningrad Region, Russia), cleaned from anthropogenic noise. The initial digital registration frequency is 1 Hz, so for our purposes the data are preliminarily averaged to 1-min values according to the Gaussian filter using the weight factors recommended by INTERMAGNET (St-Louis 2012), with centering to the beginning of the UTC min. Also, it is necessary to define a priori a certain synthetic regular baseline  $\vec{B}(t) = (X_b, Y_b, Z_b)$ . In the future, the accuracy of the solution found by the algorithm is estimated from its proximity to this series. Generally, a regular baseline series can be set by means of an arbitrary function defined on a time interval  $T = \{t_k = kh\}, k = 1, \dots, N, h$  is 1 h. In our case, we use a sine function with a phase and frequency shift between the components and an offset with respect to zero:

$$\begin{aligned} X_b &= a_x \sin(f_x t) + c_x, \\ Y_b &= a_y \sin(f_y t) + c_y, \\ Z_b &= a_z \sin(f_z t) + c_z, \end{aligned} \quad (6)$$

where  $a_x, a_y, a_z$ —the amplitude factors of the baseline,  $f_x, f_y, f_z$ —the factors of the alternation frequency,  $c_x, c_y, c_z$ —the baseline offsets. The values of the applied factors are:

- $a = 2, f = 0.2, c = -22$  for  $X$  component,
- $a = -5, f = 0.1, c = 40$  for  $Y$  component,
- $a = 2, f = 0.1, c = 120$  for  $Z$  component.

The factors  $c_x, c_y, c_z$  for each component are chosen equal to mean values calculated from real observations at the Saint Petersburg observatory. The factors  $f_x, f_y, f_z$  are arbitrary.  $a_x, a_y, a_z$  are defined from real dispersions of the baseline measurements at the observatory. The series generated this way adequately reflect the baseline behavior for the Saint Petersburg observatory.

The set of the pseudo-observed baseline values  $\vec{B}_0 = (X_b^0, Y_b^0, Z_b^0)$  is defined as a subset of regular values for the predefined irregular time moments  $T_0 = \{t_m^0\} \subset T$ , where  $|T_0| \ll |T|$ .

Further, from the 1-min values  $\vec{V}$  and the 1-h values  $\vec{B}$ , we calculate 1-min values  $F_a$ , which represent synthetic total intensity, using the formula:

$$F_a = \sqrt{(X_v + X_b)^2 + (Y_v + Y_b)^2 + (Z_v + Z_b)^2} + W_{0.1}^f(t), \quad (7)$$

where  $W_{0.1}(t)$  is an additive white Gaussian noise with the RMS  $\sigma = 0.1$  and the mean  $\mu = 0$ .

Now let us analyze the results of the algorithm application to the initial data  $\vec{V}$ ,  $F_a$  and  $\vec{B}_0$  on the observation interval  $T = 21$  days with  $\lambda = 0.5$ . In Fig. 1, the baseline values of the EMF components are presented. The synthetically generated regular baseline is marked with green, and the selected pseudo-observed values are marked with an asterisk. The obtained hourly baseline values are marked with blue. The step of the grid  $S$  in the  $\phi$  function minimization is taken 0.25 nT for each component.

An important indicator of the quality of the calculated baseline values is the difference function  $G$  (Fig. 1). As it is seen, its values are within the  $[-0.5, 0.5]$  interval, which indicates the high definitive data quality according to the INTERMAGNET regulations. The values of the minima of the target function  $\phi$  for every hour of the processed data are also presented in Fig. 1. Let  $Q_\phi$  be the time series of the found minima of the function  $\phi$  values:

$$Q_\phi(t) = \min_{\vec{S}^{i,j,k}} \varphi(t, \vec{S}^{i,j,k}). \quad (8)$$

It is seen that during the periods of increasing variability of the baseline curve and at the moments, which are remote in time from the observed baseline values, the minimum of the  $\phi$  function increases. On the contrary, in between the two first observed baseline values  $Q_\phi$  does not show any significant peaks, as (1) these two values are very close for each component, and (2) vector/scalar measurements are in good agreement. We recall that the value of the weighting factor  $\lambda = 0.5$  means equal “confidence” of the algorithm both in the observed baseline values and in the coherence of vector and scalar measurements of the EMF.

The algorithm validation process consisted of 100 computational experiments based on the synthesized data. In every computational experiment, an additive Gaussian white noise with a RMS  $\sigma = 0.5$  and a mean  $\mu = 0$  was added to the pseudo-observed baseline values:

$$\begin{aligned} X_b^i(t) &= X_b^0(t) + W_{0.5}^x(t), \\ Y_b^i(t) &= Y_b^0(t) + W_{0.5}^y(t), \\ Z_b^i(t) &= Z_b^0(t) + W_{0.5}^z(t), \end{aligned} \quad (9)$$

where  $i$  is an experiment number.

During every computational experiment, a baseline calculation was performed sequentially for  $\lambda = 0.1$ , 0.5 and 0.9. Such values were selected in order to explore the computational side effects of the algorithm when

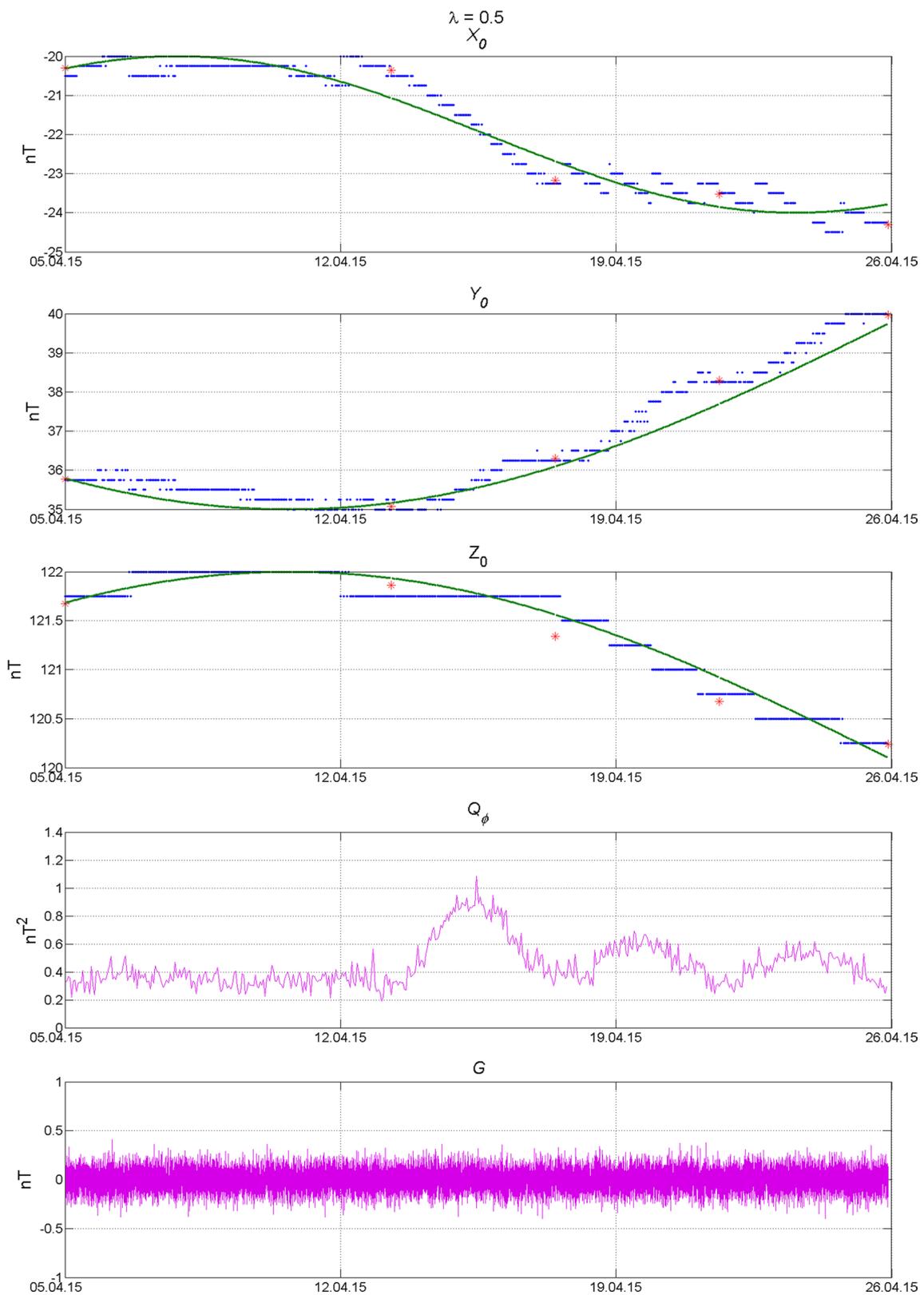
a confidence factor is selected incorrectly. The results of the algorithm work for the one of the computational experiments are given in Fig. 2.

As shown in Fig. 2a, when  $\lambda = 0.1$  the increased confidence of the algorithm in the observed baseline values reduces the dispersion of the computed baseline values; however, the dispersion of the  $G$  function values increases with that (STD  $\approx 0.19$  nT). On the other hand, when  $\lambda = 0.9$ , representing the excessive confidence of the algorithm in the quality of the work of the scalar and vector magnetometers of the observatory, the decrease in the dispersion of the  $G$  function (STD  $\approx 0.03$  nT, Fig. 2c) is seen. Accordingly, the lack of confidence in the observed baseline values decreases smoothness and adds a noise component to the calculated baselines, which can adversely affect the quality of the definitive data.

The hourly series  $Q_\phi$  is a main characteristic function for definitive data obtained using the proposed algorithm. For instance, the local  $Q_\phi$  maxima can indicate excessive variability of the observed baseline values over the corresponding time interval (Fig. 2) or the problems with synchronous data registration by the scalar and the vector magnetometers. The latter may include wrong scale factors, time desynchronization, sensor temperature effects and others. As shown in Fig. 2, the  $Q_\phi$  series depends on the predefined  $\lambda$  value.

During the algorithm validation, 100 sets of resulting data were received for each value of  $\lambda$ . In Fig. 3, the  $Q_\phi$  time series for all the computational experiments are plotted. As it is seen from the figure, the alternation of the confidence factor  $\lambda$  does not have a significant impact on the distribution of the local maxima of the  $Q_\phi$  functions. However, with increasing the confidence in the function  $G$  ( $\lambda \rightarrow 1$ ) the contribution of the noise component of the variometer values increases, which leads to the occurrence of noise effects in the  $Q_\phi$  values (Fig. 3c). With this, the order of magnitude of the total noise effect becomes comparable to the one of the noise rate of the observed baseline values, which almost completely masks possible problems with the baseline values on the  $Q_\phi$  data. Also note the overall increase in the absolute  $Q_\phi$  values due to the fact that by the definition of the function  $\phi$  [see formula (1)] and the functions  $G$  and  $A$  contained in it [see formulas (2) and (3)], the  $Q_\phi$  values are expressed in nT<sup>2</sup>. At the same time, the function  $G$  expresses the sum of the squares of the deviations for all values over an hour, while the function  $A$  is calculated for one baseline value per hour.

Let us consider the function  $Q_\phi$  in more detail for  $\lambda = 0.5$  (Figs. 2b, 3b). The local maxima of the function are clearly seen and strictly localized in time on all random sets of test data. This indicates that the change in noise characteristics in the initial data does



**Fig. 1** Calculated hourly baseline values (blue) with  $\lambda = 0.5$ , pseudo-observed values (red) and functionally defined hourly baselines (green) for the X (first plot), Y (second plot) and Z (third plot) components of the EMF. Fourth and fifth plots represent the corresponding  $Q_\phi$  and G functions

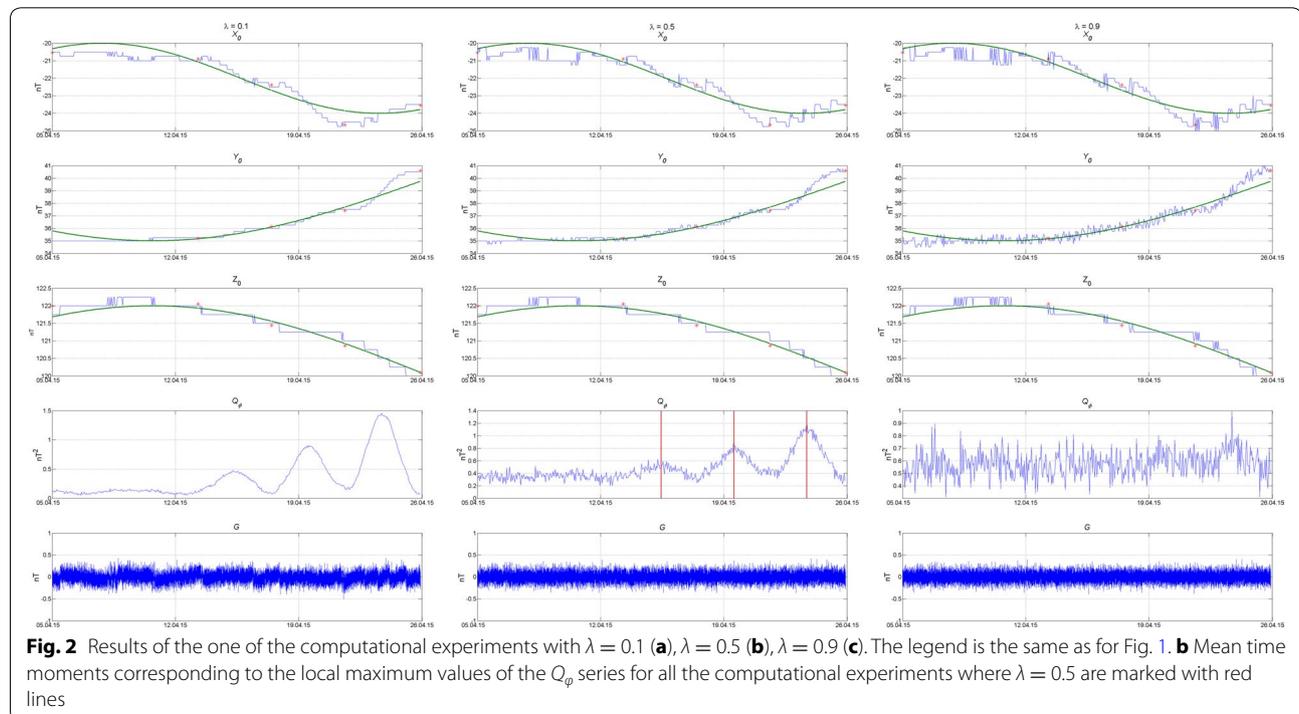
not significantly affect the result of the algorithm. The found local maxima of  $Q_\phi$  correspond to the intervals of low confidence of the algorithm in the obtained definitive data quality with respect to its general level. For instance, the local  $Q_\phi$  maxima on the time interval from 13.04 to 21.04 are caused by high variability of the baseline values. On the contrary, local  $Q_\phi$  minima indicate a higher quality of the resulting data. They occur at the time moments corresponding to the presence of pseudo-observed baseline values. However, the significant difference between the neighboring pseudo-observed values and their insufficient frequency cause the occurrence of local maxima in  $Q_\phi$  at the intermediate points; their time moments are close to the mid-points of the intervals between the pseudo-observed baseline values (Figs. 2b, 3b).

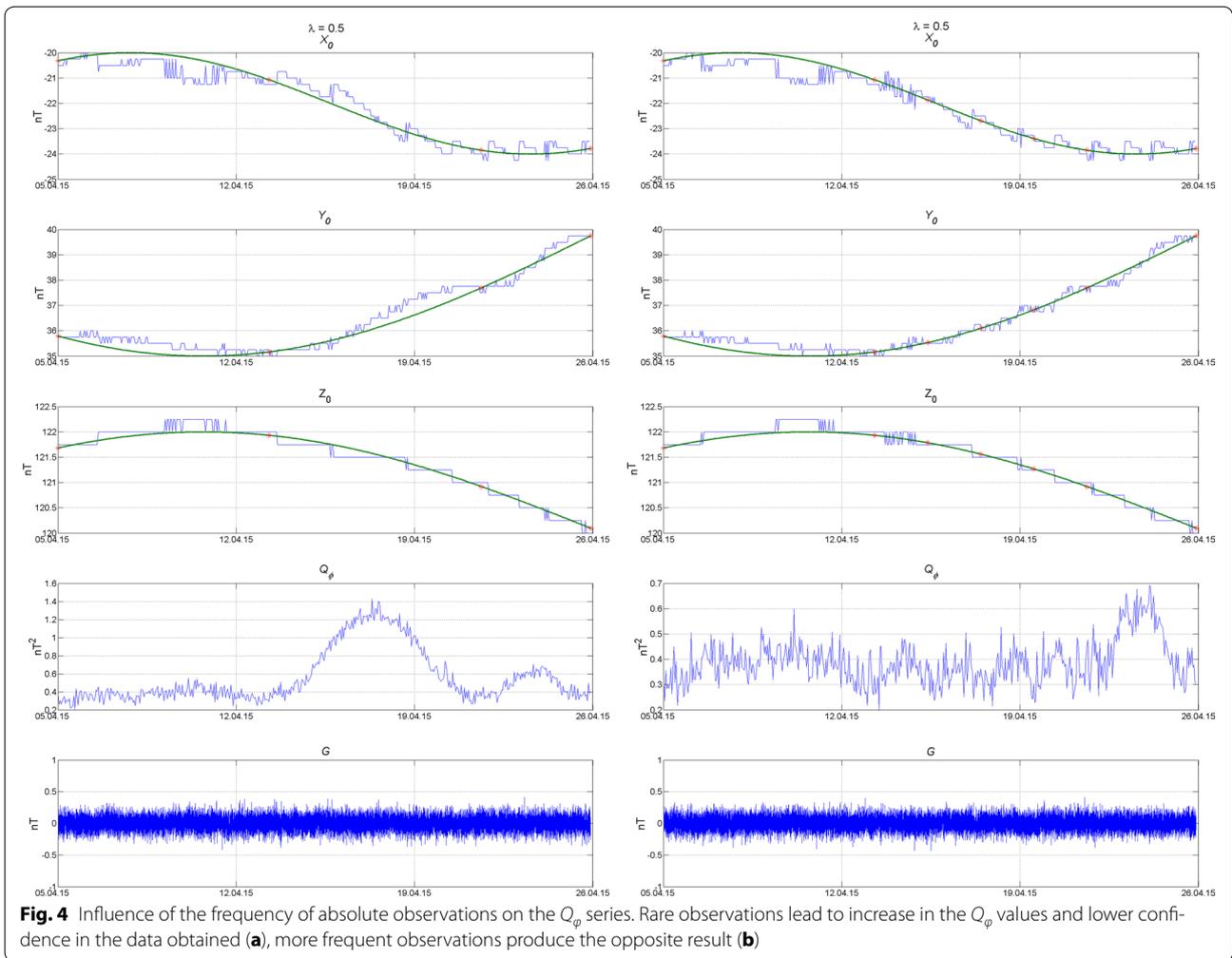
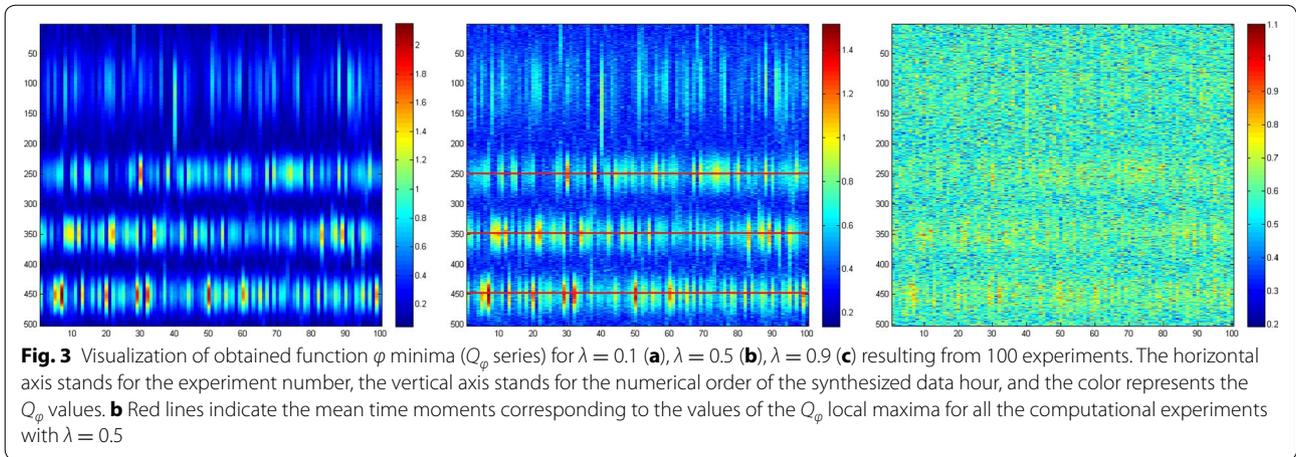
Let us illustrate the variability of the algorithm operation by introducing changes to the initial synthetic data. Figure 4a shows a result of the algorithm's operation after the pseudo-observed value on 17.04 was removed. Before the removal of this point, in its time neighborhood there were two local  $Q_\phi$  maxima with amplitudes of about  $0.9 \text{ nT}^2$  before 17.04 and  $0.6 \text{ nT}^2$  after 17.04 (Fig. 1). The point removal led to the appearance of one local  $Q_\phi$  maximum on the time interval from 13.04 to 21.04 with an amplitude approximately twice higher, about  $1.3 \text{ nT}^2$ . On the contrary, the increase in the frequency of the pseudo-observed

baseline values leads to the disappearance of local maxima on  $Q_\phi$  and, therefore, to increasing the level of confidence in the resulting data. This circumstance is illustrated in Fig. 4b, where two measurements on 15.04 and 19.04 are added, leading to the disappearance of local maxima on  $Q_\phi$ , as shown in Fig. 1. Thus, the function  $Q_\phi$  represents a quantitative measure of the quality of the initial data and, as a consequence, reflects the level of confidence in the resulting data. The latter is one of the most important features of the developed algorithm.

### Validation of the method on real data

To validate the method in real conditions, we used initial data from the Saint Petersburg geomagnetic observatory (IAGA code SPG), obtained during 2015. The observatory is located 100 km north from the city of Saint Petersburg (60.542N, 29.716E) and based at the site of the former Krasnoe Ozero (Red lake) observatory, where continuous observations were carried out from 1960 to 1990. Through the efforts of the GC RAS in 2014, it was equipped with the modern magnetometric instrumentation, and in 2016 the first definitive data set for 2015 was produced. These data were accepted by INTERMAGNET, and the observatory was officially included into the network (Soloviev et al. 2015; Sidorov et al. 2017). The continuous data recording at the observatory are provided by the following instruments:



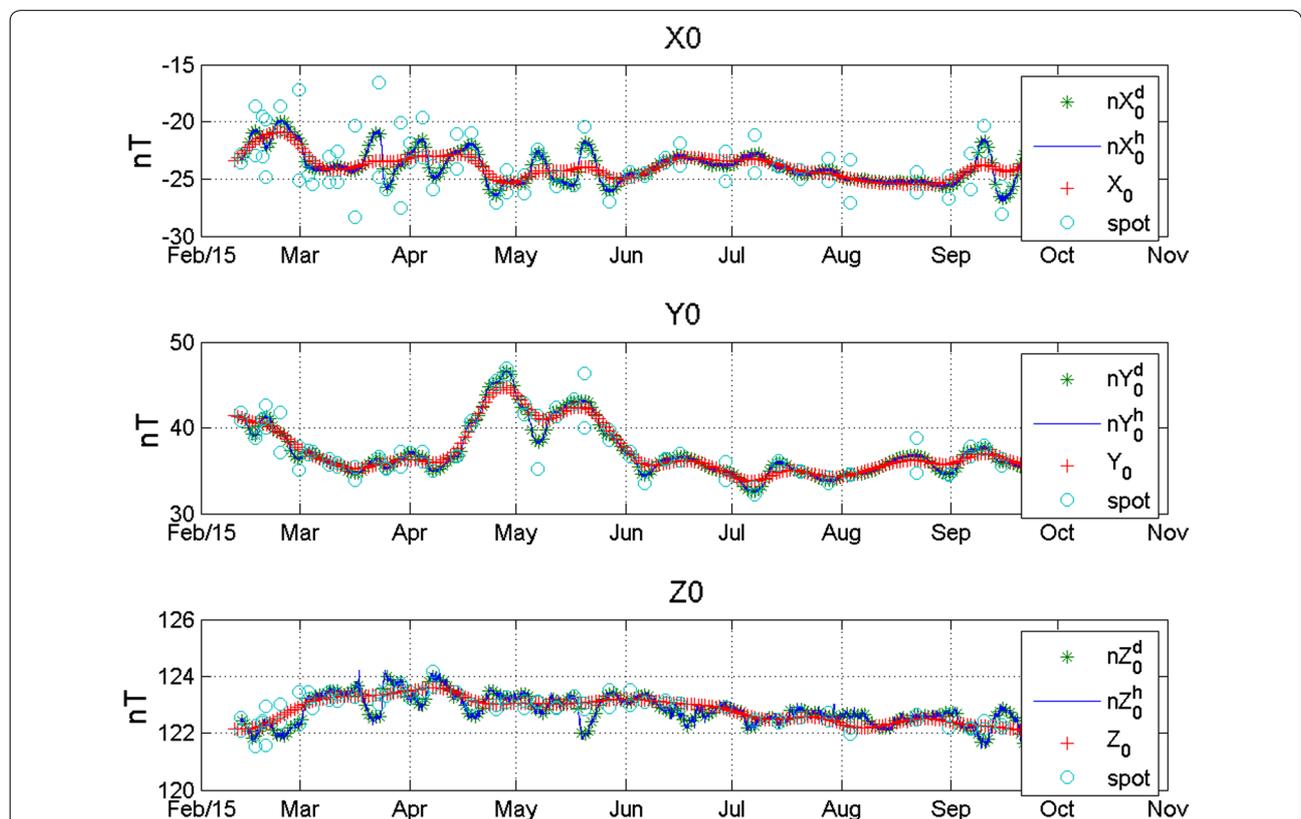


- FGE 3-axial fluxgate magnetometer (DTU, Denmark) for recording the variations of the EMF components;
- GSM-19 scalar magnetometer (Gem Systems, Canada) for recording the modulus of the EMF vector.

The fluxgate magnetometer is installed in the separate pavilion and provides measurements with the frequency of 1 Hz. It is oriented in such a way that the sensor axes are aligned along the directions to the geographic north, east and vertical line, so the measured components of the EMF are  $X$ ,  $Y$  and  $Z$ , respectively. The scalar magnetometer carries out measurements with the frequency of 0.33 Hz and is installed in another pavilion. In the same pavilion, absolute observations of magnetic declination and inclination are carried out by operator 1–2 times per week, using the MinGeo 020B declinometer/inclinometer (MinGeo, Hungary). The distance between the pavilions is about 25 m. The data from both magnetometers are continuously transmitted to the Magrec data logger (MinGeo, Hungary), where their temporal synchronization, min averaging by Gaussian filter, preliminary processing and further transmission to GC RAS

are performed (Soloviev et al. 2013; Gvishiani et al. 2014, 2016a, b).

To demonstrate the advantages of the described method, herein we provide its comparison with the traditional method used by the INTERMAGNET community by the example of the SPG data over 2015. Let us remind that the data obtained by the traditional method have the official status of definitive data, published on the INTERMAGNET Web site (<http://www.intermagnet.org>). According to the  $\Delta F$  record constructed for the 2015 definitive data, the processed data from the magnetometers have a high level of coherence, as the  $\Delta F$  amplitude does not exceed 1 nT during the whole year. The baselines over the whole period of absolute observations in 2015, which correspond to the definitive data, have the following maximal amplitudes: 5 nT for  $X$  component, 10 nT for  $Y$  component and 2 nT for  $Z$  component. Exceedance of 5-nT span in  $Y_0$  was due to short-period problems with the variometer electronics that were soon resolved. It was found out experimentally that during the year the baseline value  $F_b^0$  for the total field intensity record was constant and equal to -5.5 nT.



**Fig. 5** Comparison between the baseline values for the three EMF components, obtained by the classical and the proposed methods for the period from February 20 to October 1, 2015. The baseline according to the classical method with the use of the smoothing spline interpolation is marked with red ( $X_0$ ), the hourly baseline values obtained by the described algorithm are marked with blue ( $nX_0^h$ ), the daily values calculated by averaging the hourly values  $nX_0^d$  are marked with green ( $nX_0^d$ ), and the observed baseline values are marked with light-blue circles (spot)

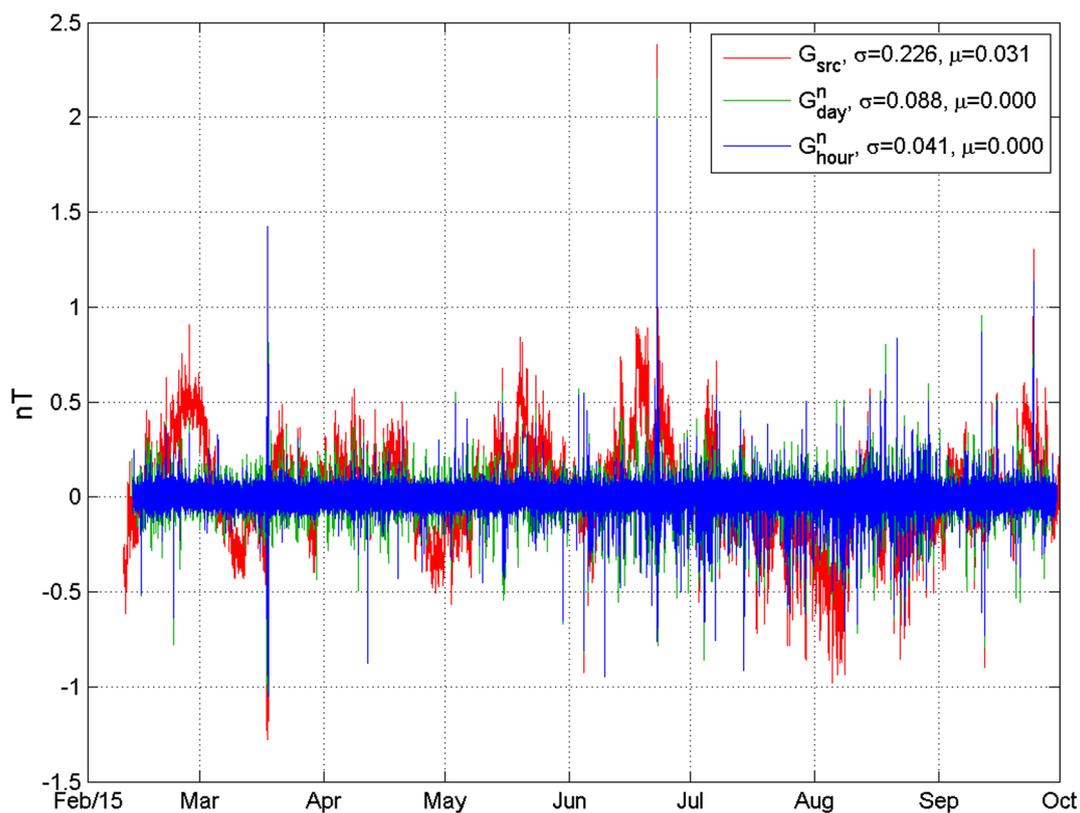
As the input data for the developed algorithm, we used the min variations of the components and the EMF modulus record, cleaned from anthropogenic disturbances using the algorithms (Bogoutdinov et al. 2010; Soloviev et al. 2012a, b; Sidorov et al. 2012), as well as the absolute observations after outlier removal. In the calculations, the confidence factor  $\lambda$  was set equal to 0.5; the observed baseline values, if done more than once a day, were averaged and bound to a nearest whole hour. If we keep several close in time measurements within a day, the resulting curve will tend to their average (given that  $\lambda = 0.5$ ). In the following section, we present the results of the comparison of the two methods by the example of the data for the period from February 20 to October 1, 2015, as it is the longest time interval of almost continuous measurements during a year. In early February and at the end of October, there are big losses of variometer data.

### Discussion

Figure 5 demonstrates the comparison between the daily baselines calculated according to the classical method using the smoothing spline interpolation ( $X_0$ ), the hourly

baselines obtained by the algorithm described in the paper ( $nX_0^h$ ), the daily baselines calculated by averaging the hourly values  $nX_0^h$  ( $nX_0^d$ ) and the observed baseline values (spot). Let us remind that we use the grid ( $S$ ) step 0.25 nT for calculating hourly baseline values (see “Validation of the method on synthetic data” section), whereas INTERMAGNET requires daily values. While averaging 24 values we get a higher accuracy exceeding 0.1 nT, which falls within INTERMAGNET standards. The comparative analysis of the results displays much more variability of the baseline values calculated by the new method both for the hourly and for the daily cases. At the same time, it is seen from the plot that the values of both series are very close to each other.

The main indicator for the proposed method efficiency is the resulting curve  $\Delta F$ , which is widely used in INTERMAGNET practice as the data quality measure (Reda et al. 2011). The general difference from the  $\Delta F$  series, derived from the approved definitive data of the SPG observatory, is the significant decrease in the root-mean-square deviation from  $\sigma = 0.226$  nT to  $\sigma = 0.041$  nT when using the hourly baseline values ( $G_{\text{hour}}^n$ ) and  $\sigma = 0.088$  nT when using the daily baseline values ( $G_{\text{day}}^n$ ), and the

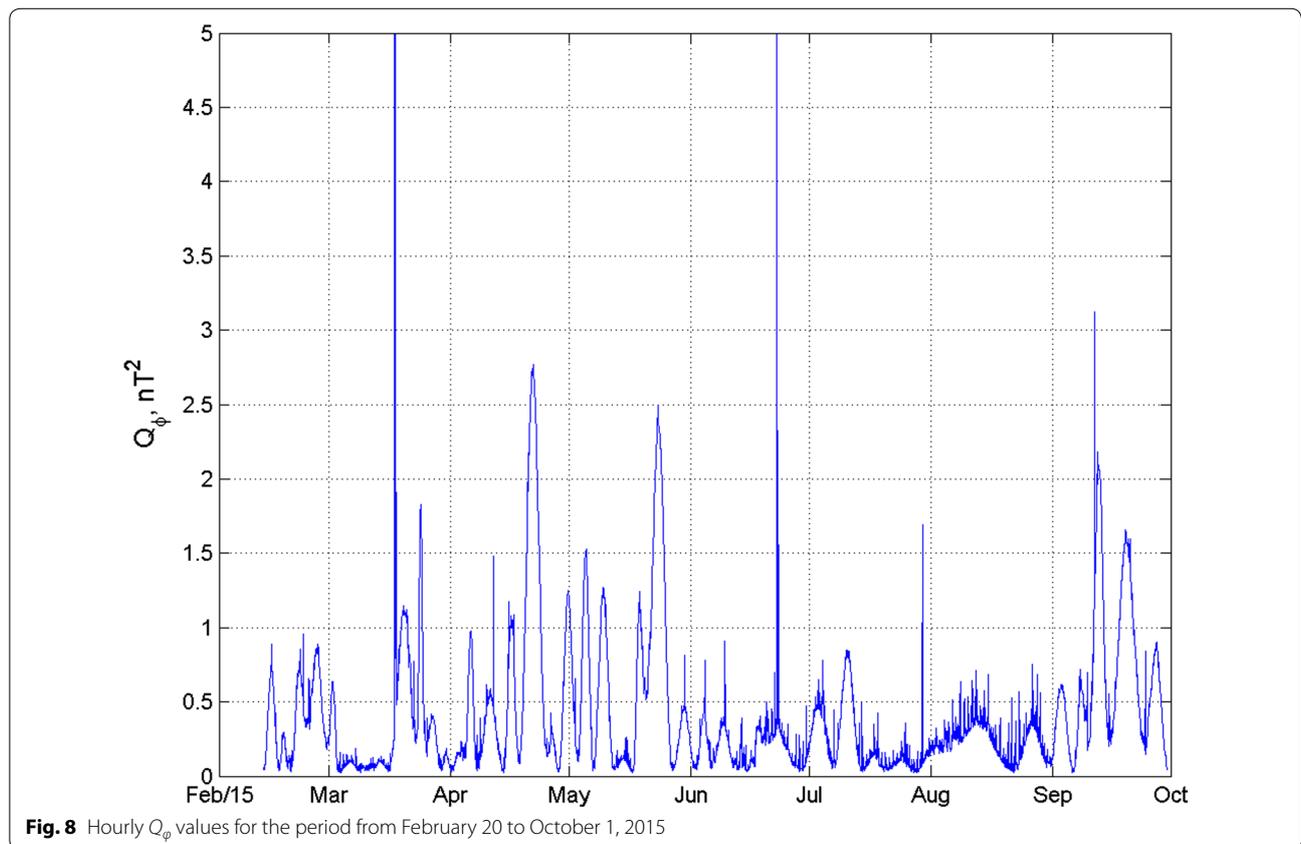
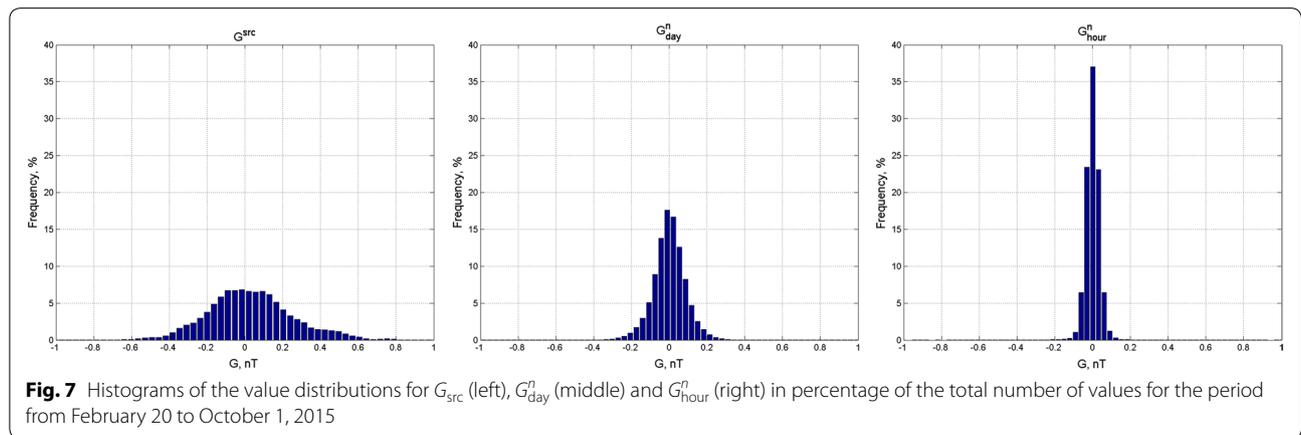


**Fig. 6** Comparison between the  $\Delta F$ , obtained from the definitive data accepted by INTERMAGNET ( $G_{\text{src}}$ , blue), and the  $\Delta F$  curves obtained using the new method from hourly ( $G_{\text{hour}}^n$ , green) and daily ( $G_{\text{day}}^n$ , red) baseline values for the period from February 20 to October 1, 2015

decrease in the mean to 0. In Fig. 6, the three corresponding  $\Delta F$  plots are presented ( $G_{src}$ ,  $G_{hour}^n$ ,  $G_{day}^n$ ) for the whole interval under consideration; Fig. 7 shows histograms of their value distributions in percentage.

Also, the  $Q_\phi$  series was plotted for the whole interval under consideration, characterizing the level of confidence in the definitive data for each value. The resulting  $Q_\phi$  plot is given in Fig. 8. As it is seen from the plot, the data with the highest confidence level (the lowest  $Q_\phi$

values) are obtained in July and August. High-frequency low-amplitude spikes around July–August are due to electric works that took place nearby and consequently affected  $G$  function. Low confidence level according to  $Q_\phi$  (the highest  $Q_\phi$  values) corresponds to the time intervals of the most intense baseline variations, for example, at the end of April, at the end of May and at the end of June. Single spikes in the  $Q_\phi$  values correspond to the intervals of technical gaps in the initial data and can be



excluded from consideration when evaluating the processing results.

Thus, the comparative analysis shows that the definitive data obtained using the developed algorithm meet the INTERMAGNET requirements. Moreover, due to the new method, the data quality indicators are essentially improved, comparing to the classical approach, and also a relevant estimation of the measurement quality during the analyzed time interval is provided. Taking this estimation into account, it becomes possible to set the minimal quality level required for the final data and, when processing the observations from several observatories, to rank the individual data fragments according to their quality.

## Conclusions

The developed approach allows to build a baseline in a semiautomatic mode using all the available data from the observatory. They include not only the observed baseline values derived from spot values of the magnetic field components and its total intensity at the times of absolute observations, but also the continuous data of vector and scalar magnetometers. The latter are not used in traditional observatory practice when calculating the regular baselines. The developed approach is based on the minimization of the target functional  $\phi$ , which is a linear combination of two functions  $G$  and  $A$ , connected with a weight factor  $\lambda$ . The  $G$  function is responsible for estimating the difference of the vector magnitudes obtained from the independent vector and scalar recordings, and the  $A$  function represents the conformity of the considered baseline vector to the observed ones at two nearest times. The resulting series of the found minima of the  $\phi$  function ( $Q_\phi$ ) represents a quantitative estimation of confidence in the obtained baseline values for each time moment. The method is adopted for Cartesian component measurements of the magnetic field and will not work in its present form for dIdD or LAMA variometers as they record spherical components.

It should be noted that the traditional approach to quasi-definitive (Peltier and Chulliat 2010) and definitive data preparation involves the interpolation of the observed baseline values. With that, the interpolation parameters are defined individually for each observatory and depend on the instrumentation used, the measurement quality, the typical baseline value dispersion, etc. In the proposed method, the free parameter  $\lambda$  has a clearer physical meaning, and its selection is less subjective. Moreover, the selection of the parameter can be automated, for example, by the analysis of statistical features of the initial data. An important feature of the algorithm is the possibility of a quantitative estimation of the resulting data precision and, thus, determination of the

problematic fragments in initial data without operator intervention.

The algorithm operation was evaluated both on synthetic examples and on real data registered in 2015 at the Saint Petersburg INTERMAGNET observatory. The data obtained using the proposed method were compared with the definitive data officially accepted by INTERMAGNET. The advantages of the new method are demonstrated, and it is shown that it allows to essentially improve the resulting data quality as compared to the classical approach. The obtained results prove that the baseline variability in time should not necessarily be smooth. In particular, this may be due to the distance between the variation and absolute pavilions: the more this distance, the less smooth is the baseline (Lesur et al. 2017). At the same time, the assumption of the baseline smoothness might lead to the loss of the information about the geomagnetic signals of small spatial scale (100–200 m) but quite lasting ( $\sim 1$  day long), with the amplitude of several nanoteslas. In turn, this will add significant distortion into the models of rapid core field variations built using observatory data.

## Abbreviation

EMF: Earth's magnetic field.

## Authors' contributions

AS performed the resulting analysis and drafted the manuscript. VL set the problem and advised for data interpretation. DK developed software and carried out data processing. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Geophysical Center RAS, Moscow, Russia. <sup>2</sup> Schmidt Institute of Physics of the Earth RAS, Moscow, Russia. <sup>3</sup> Institut de Physique du Globe de Paris, Paris, France. <sup>4</sup> Gorno-Altaysk State University, Gorno-Altaysk, Russia.

## Acknowledgements

The results presented in this paper rely on data collected at the INTERMAGNET magnetic observatories. We express our gratitude to the national institutes that support them, INTERMAGNET community for promoting the high standards of magnetic observatory practice, and Russian-Ukrainian Geomagnetic Data Center for making the data freely available online. The authors thank Dr. Roman Sidorov (GC RAS) for his help in preparing the materials. We are also grateful to the three anonymous reviewers for their valuable remarks that helped us to improve the paper.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

INTERMAGNET data and Russian observatory data are available from <http://www.intermag.net> and <http://geomag.gcras.ru>.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Funding

The research is funded by the Federal Agency for Scientific Organizations of Russia.

## Publisher's Note

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

Received: 14 November 2017 Accepted: 22 January 2018

Published online: 05 February 2018

## References

- Bogoutdinov SR, Gvishiani AD, Agayan SM, Soloviev AA, Kihn E (2010) Recognition of disturbances with specified morphology in time series. Part 1: spikes on magnetograms of the worldwide INTERMAGNET network. *Izvest Phys Solid Earth* 46(11):1004–1016
- Gvishiani A, Lukianova R, Soloviev A, Khokhlov A (2014) Survey of geomagnetic observations made in the Northern Sector of Russia and new methods for analysing them. *Surv Geophys* 35(5):1123–1154. <https://doi.org/10.1007/s10712-014-9297-8>
- Gvishiani A, Soloviev A, Krasnoperov R, Lukianova R (2016a) Automated hardware and software system for monitoring the Earth's magnetic environment. *Data Sci J* 15:18. <https://doi.org/10.5334/dsj-2016-018>
- Gvishiani AD, Sidorov RV, Lukianova RYu, Soloviev AA (2016b) Geomagnetic activity during St. Patrick's Day storm inferred from global and local indicators. *Russ J Earth Sci* 16:ES6007. <https://doi.org/10.2205/2016es000593>
- Hrvoic I, Newitt LR (2011) Instruments and methodologies for measurement of the Earth's magnetic field. In: *Geomagnetic observations and models* (IAGA special sopron book series), vol 5. Springer Science + Business Media B.V., pp 105–126
- Jankowski J, Sucksdorff C (1996) Guide for magnetic measurements and observatory practice. In: *International association of geo-magnetism and aeronomy*, Warsaw, pp 1–235
- Lesur V, Heumez B, Telali A, Lalanne X, Soloviev A (2017) Estimating error statistics for Chambon-la-Forêt observatory definitive data. *Ann Geophys* 35(4):939–952. <https://doi.org/10.5194/angeo-35-939-2017>
- Love JJ, Chulliat A (2013) An international network of magnetic observatories. *EOS Trans Am Geophys Union* 94(42):373–374. <https://doi.org/10.1002/2013EO42001>
- Peltier A, Chulliat A (2010) On the feasibility of promptly producing quasi-definitive magnetic observatory data. *Earth Planets Space* 62:e5. <https://doi.org/10.5047/eps.2010.02.002>
- Reda J, Fouassier D, Isac A, Linthe H-J, Matzka J, Turbitt CW (2011) Improvements in geomagnetic observatory data quality. In: *Geomagnetic observations and models* (IAGA special sopron book series), vol 5. Springer Science + Business Media B.V., pp 127–148
- Sidorov RV, Soloviev AA, Bogoutdinov SR (2012) Application of the SP algorithm to the INTERMAGNET magnetograms of the disturbed geomagnetic field. *Izvest Phys Solid Earth* 48(5):410–414
- Sidorov R, Soloviev A, Krasnoperov R, Kudin D, Grudnev A, Kopytenko Y, Kotikov A, Sergushin P (2017) Saint Petersburg magnetic observatory: from Voeikovo subdivision to INTERMAGNET certification. *Geosci Instrum Methods Data Syst* 6:473–485. <https://doi.org/10.5194/gi-6-473-2017>
- Soloviev AA, Agayan SM, Gvishiani AD, Bogoutdinov SR, Chulliat A (2012a) Recognition of disturbances with specified morphology in time series: part 2. Spikes on 1-s magnetograms. *Izvest Phys Solid Earth* 48(5):395–409
- Soloviev A, Chulliat A, Bogoutdinov S, Gvishiani A, Agayan S, Peltier A, Heumez B (2012b) Automated recognition of spikes in 1 Hz data recorded at the Easter Island magnetic observatory. *Earth Planets Space* 64(9):743–752. <https://doi.org/10.5047/eps.2012.03.004>
- Soloviev A, Bogoutdinov S, Gvishiani A, Kulchinskiy R, Zlotnicki J (2013) Mathematical tools for geomagnetic data monitoring and the INTERMAGNET Russian segment. *Data Sci J* 12:WDS114–WDS119. <https://doi.org/10.2481/dsj.wds-019>
- Soloviev A, Kopytenko Y, Kotikov A, Kudin D, Sidorov R (2015) Definitive data from geomagnetic observatory Saint Petersburg (IAGA code: SPG): minute values of  $X$ ,  $Y$ ,  $Z$  components and total intensity  $F$  of the Earth's

magnetic field. ESDB repository. Geophysical Center of the Russian Academy of Sciences (2016). <http://doi.org/10.2205/SPG2015min-def>  
St-Louis B (2012) INTERMAGNET technical reference manual, version 4.6, 92 pp

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---