Measurement campaign and mathematical model construction for the ship *Zodiak* magnetic signature reproduction

J. Tarnawski [a],[*], K. Buszman [b], M. Woloszyn [a], T.A. Rutkowski [a], A. Cichocki [b], R. Józwiak [b]

[a] *Faculty of Electrical and Control Engineering, Gdańsk University of Technology, Gdańsk, Poland*
[b] *Faculty of Navigation and Naval Weapons, Polish Naval Academy, Gdynia, Poland*

ABSTRACT

The paper presents the partial work done within the framework of the EDA Siramis II project focused on magnetic signature reproduction of ships. Reproduction is understood here as the ability to determine the magnetic anomaly of the local Earth magnetic field in any direction and at any measurement depth due to the presence of the analysed object. The B-91 type hydrographic ship *Zodiak* was selected as the real case study. The work was divided into two main stages: the development of a measurement campaign taking into account physical measurements, and the development of a mathematical model on the basis of the measured values. The measurement campaign included: preparation of the measuring range, selection of equipment for the measurement of magnetic quantities and geographical location, and data recording while the ship passes the measuring point according to the designated course. As a result of the measurement campaign, magnetic flux density components were collected in different positions in relation to the measuring instruments and the ship's heading. A multi-dipole model was used to build the mathematical model in accordance with the idea of inverse modelling. The effectiveness of this model was previously checked on synthetic data of virtual ships generated using the finite element method. Experiments performed with simulation models were helpful in determining the structure of the model, the nature of the data, and the number of samples needed to properly determine the multi-dipole model parameters. The parameters were determined using the nonlinear least squares method according to the idea of data fitting. The classical Ridge and Lasso regularization methods were applied to prevent the developed multi-dipole model from overfitting. Other regularization methods based on GPS accuracy marks and modification of fitness functions were also considered. The verification was done using real data: the data generated by the model was compared with patterns recorded during the *Zodiak* measurement campaign. High degree of conformity of the shape of characteristics was obtained. Moreover, the correctness of model execution was confirmed by low values of quantitative indices such as RMSE and MAE representing modelling errors. The methodology presented in the paper is quite universal and can be used to determine the signatures of other ferromagnetic objects.

1. Introduction

Nowadays in the coastal area, there is an increasing emphasis on maritime transportation safety due to the permanent increase in vessel traffic [1]. The information about moving ships in a certain area is essential in the navigation process [2,3]. Active and passive traffic monitoring systems are used to acquire such information [4,5]. A passive system can use the magnetic field disturbance described by the magnetic flux density vector as the carrier of information about an object [6–8]. When moving in the earth's magnetic field, the object made of ferromagnetic material is the source of local magnetic field disturbance. The magnitude and character of this disturbance depends on the size and shape of the object, its course relative to the Earth's magnetic field, its permanent magnetization resulting from the manufacturing process, and its mechanical impact on the seawater surface [9]. The magnetic field of the object includes induced and permanent magnetism [10]. The object analysed in the paper is a ship moving in a given area, the hull of which is made of ferromagnetic steel.

* Corresponding author at: Gdańsk University of Technology, Faculty of Electrical and Control Engineering, 11/12 Gabriela Narutowicza Street, 80-233 Gdańsk, Poland.
E-mail addresses: jaroslaw.tarnawski@pg.edu.pl (J. Tarnawski), k.buszman@amw.gdynia.pl (K. Buszman), miroslaw.woloszyn@pg.edu.pl (M. Woloszyn), tomasz.adam.rutkowski@pg.edu.pl (T.A. Rutkowski), a.cichocki@amw.gdynia.pl (A. Cichocki), r.jozwiak@amw.gdynia.pl (R. Józwiak).

Fig. 1. UMM view (left) and its transportation on RIB (right).

Research on magnetic signatures of ships can be found e.g. in [11,12] but the most comprehensive approach is presented in [13]. Along with the civil safety of maritime traffic, magnetic signatures are also important for the navy [14]. In military applications, various methods of modifying the magnetic signature are used [9,15].

The key issue in this article is the magnetic signature measurement campaign of the type B-91 hydrographic ship *Zodiak* [16,17] which operated in the specially prepared measuring range in the Gulf of Gdansk. The literature offers descriptions of both the organization of stationary testing sites [18], and signature measurements with the use of mobile devices [19]. To determine the magnetic signature in the reported measurement campaign, the magnetic flux density vector of the ship moving in four cardinal geographical directions was measured. The measurement was performed using a specially designed and developed underwater system equipped with one triaxial magnetometer. During the measurements, the ship's position was recorded using an on-board satellite navigation system, commonly used on civilian vessels. The position determination was subject to an error resulting from the accuracy of the on-board GPS system.

The obtained real magnetic field measurement results were used, along with the vessel trajectories, for the synthesis in which the multi-dipole numerical model describing the analysed vessel was determined. The multi-dipole model has prediction properties and allows to reconstruct the ship signature for arbitrary ship course and at specific depth. This knowledge makes it possible to identify a vessel from the magnetic signature measured during its single pass over a passive measurement system. In the literature, the topic of modeling magnetic signatures of ships is widely covered. A general overview of approaches to modeling is presented in [20]. The description of the multi-dipole model developed by the authors can be found in [21,22] while the methodology for determining its parameters is presented in [23] and [24], among other sources. An integrated multi-dipole approach with material sensitivity analysis can be found in [25].

The main goal of this paper was to create a multi-dipole model that would be able to reproduce magnetic signatures of the ship *Zodiak*, i.e., to develop a model with predictive capabilities based on limited but real measurement data. This process consists of three stages: model structure defining, raw data pre-processing, and model learning and validation. The authors have demonstrated that synthetic data can be used to reproduce magnetic signatures with very high accuracy in arbitrary directions and depths, even those for which the model was not tuned [22]. Using synthetic data from simulations based on the finite element method [26] is an important element of magnetic field research conducted by various authors. The magnetic disturbance data generated this way is an invaluable source of virtual measurement data with any desired resolution and accuracy for other analyzes.

This paper presents the problem of synthesizing a multi-dipole model using the measurements in real sea conditions, in which determining geographic position with high accuracy is extremely difficult. Certainly, the quality of the measurements will have a key impact on the quality of the model and the reproduction of magnetic signatures - the multi-dipole model prediction capability.

This paper is organized as follows. The description of the measurement campaign conducted with the real ship *Zodiak* and the measurement system specially designed for this purpose is presented in Section 2, which also discusses aspects of satellite navigation accuracy that are critical in determining the ship's position and the quality of the multi-dipole model in the magnetic signature reconstruction task. Section 3 presents the structure of the mathematical multi-dipole model, describes its previous applications based on the synthetic data, and presents the pre-processing of raw, real-world magnetic measurement data of *Zodiak* for the multi-dipole model purpose. Next, the formulation of the optimisation task (inverse model approach) is presented that leads to the identification of parameters of the multi-dipole model for *Zodiak* based on the concept of leave-one-out cross-validation (LOOCV). In view of natural contradictions in the data resulting from ambiguous determination of object's position during the measurements, an analysis considering rejection of entire sets representing single ship passage was also applied. This section also describes aspects related to fitting and validating the model on available measurement data, as well as presents the assessment and discussion of the obtained results. Section 4 describes additional methods of regularization verified by the authors as part of the experimental research which were applied for identifying parameters of the multidimensional model to prevent it from overfitting and to improve its predictive efficiency. Short descriptions of all verified methods are presented and the obtained results are analysed. Additionally, an approach incorporating additional measurement data quality information, i.e. GPS FIX and GPS HDoP, is presented and discussed. Section 5 concludes the paper, while the included Appendix contains additional figures illustrating cross-validation results (described in Section 4) for all twelve data sets acquired during the measuring campaign with *Zodiak*.

2. The *Zodiak* measurement campaign

The measurements of *Zodiak* magnetic field disturbances were carried out as a part of the project, which also measured hydroacoustic, hydrodynamic, electric, and seismic fields. The measurements were made in the Gulf of Gdansk, in a water depth of 20 m and 2 km from the fairway. The location of the measurement range allowed for safe ship's maneuvering and reduced the impact of disturbances from other objects moving on the fairway to the Port of Gdynia. The measurements of

Table 1
General measurements information.

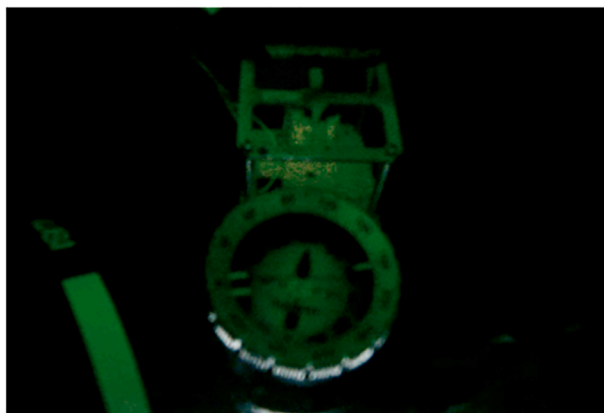| UMM position and depth	Depth			Latitude	Longitude	
	19,7 m			54° 33,205	18° 36,924	
Weather conditions	Pressure	Wind direction		Wind speed	Temperature	Sea state
	1029,6 hPa	NE		2 m/s	15 °C	1 – 2
Features of *Zodiak*	Type general	Type detailed		IMO	MMSI	GT
	Anti-Pollution	Buoy-Laying Vessel		8,030,908	261,194,000	751
	DWT	Length	Breadth	Draught	Heading	Speed
	243 t	61,33 m	10,8 m	3,2 m	N – E – S – W	5 kn – 7 kn – 10 kn

Fig. 2. Correction of UMM heading using underwater compass.

physical fields were carried out using the Underwater Measurement Module (UMM, presented in Fig. 1) [27]. The UMM has been designed to measure all underwater physical fields generated by the ship moving on the water surface. The module was equipped with four hydrophones, a hydrodynamic pressure sensor, gradient electric field sensor, seismic accelerometer, and magnetic flux density sensor.

The UMM was additionally equipped with a probe with a built-in hydrostatic pressure sensor Valeport miniSVS [28] to measure sound velocity in water, and with an inclinometer with a compass Honeywell HMR3300 [29]. The inclinometer with compass recorded the module's tilt (pitch and roll correction) and its position in relation to the north direction (yaw correction).

To measure magnetic field disturbances, a Bartington Grad-03 gradiometer [30] with 2 built-in three-axis fluxgate magnetometers. Notice that the data recorded from the first three-axis fluxgate magnetometer of Grad-03 were used for the analysis presented in the paper. The suitable analogue outputs of Grad-03 were connected to the National Instruments NI 9220 measurement card of the system with a built-in A/D converter [31]. The applied measurement card had 16 analog inputs. It converts signals into digital form with 16-bit resolution and sampling frequency of up to 100 kSa/s/ch. The described measurement card and other cards used for recording the remaining signals were managed by an industrial computer with the real-time system, National Instruments cRIO 9068 [32]. Communication between the industrial computer and the operator's console located on the base ship was carried out via the Gigabit Ethernet standard. This ship was anchored approximately 900 m away from the position of the UMM. To minimize magnetic influence, the UMM was built of non-magnetic materials. The base ship engines and all other noise sources were turned off. All measurement devices were supplied from batteries to eliminate 50-Hz disturbances from mains supply. This setting limited the influence of the base ship on the measurement results of *Zodiak* magnetic signature. To compensate Earth magnetic field variations, current values of each Earth's magnetic field component were recorded just before the measurement. Then, the obtained results were subtracted from the magnetic field measured during ship movement.

The measurements described in the article were carried out at the end of September 2019. Very good weather conditions occurred at the selected date: the sea state was 1–2, the temperature 15 °C, and the wind speed 2 m/s. All detailed measurement information is collated in Table 1. To carry out the measurements, in the first phase, the module was transported on the rigid inflatable boat (RIB) to the target geographical position. The boat with the crane installed at the bow is shown in Fig. 1. After reaching the geographical position, the UMM was placed on the seabed and the exact position of the module on the sea bottom was determined. After placing the module, the diver corrected the X-axis position of the magnetometer and the compass to the north, as shown in Fig. 2. Once the UMM position was corrected, a signaling buoy was launched on the surface to read the geographical position using the Garmin GPSMAP GPS receiver. This method of positioning, with strong tension of the rope connecting the buoy with the measuring module in good weather conditions and applying the DGPS differential correction, allowed to determine the position of the underwater measuring module with an accuracy of 5 m. It is noteworthy here that there are also known

Fig. 3. Draught of tested vessel during measurements (left) and UMM sea bottom placement (right).
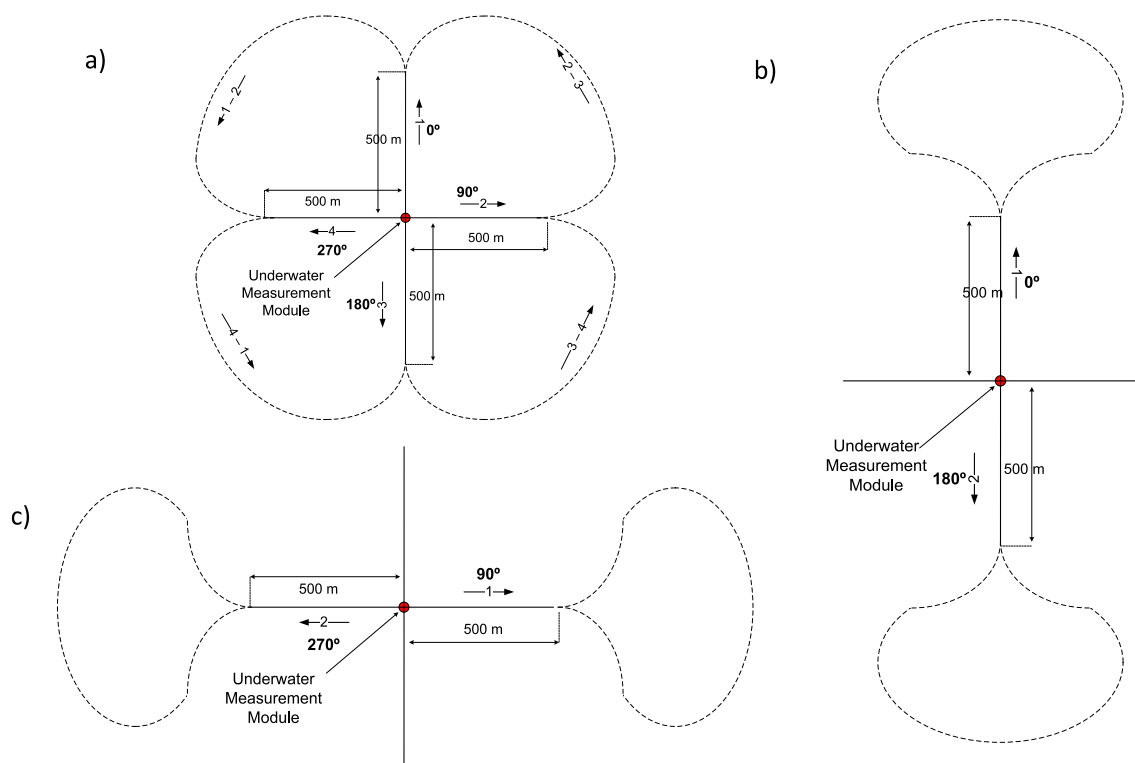
Fig. 4. Three configurations of ships' passages described in Table 2.

Table 2
Description of ship passages consistent with Fig. 4.

Four cardinal directions N, S, E, W (Fig. 4a)	Two cardinal directions N, S (Fig. 4b)	Two cardinal directions E, W (Fig. 4c)
From S to N (arrow 1)	From N to S (arrow 2)	From E to W (arrow 2)
From W to E (arrow 2)		
From N to S (arrow 3)	From S to N (arrow 1)	From W to E (arrow 1)
From E to W (arrow 4)		

Table 3
Parameters of ship runs.

Run name	Run No.	Speed at CPA [kn]		CPA [m]	Heading [°]
		Target	Real		
C01N05E	1	5	4,6	1,4	0
C01E05A	2	5	5,3	12	90
C01S05F	3	5	5,1	9,4	180
C01W05A	4	5	4,9	1,4	270
C01N07C	5	7	7,8	7,3	0
C01E07A	6	7	7,8	6,7	90
C01S07C	7	7	7,2	4,6	180
C01W07A	8	7	7,7	9,4	270
C01N10C	9	10	9,9	7	0
C01E10A	10	10	9,8	10	90
C01S10C	11	10	9,7	6	180
C01W10A	12	10	10	8,4	270

active underwater positioning methods [33,34] however, they require an additional positioning system and their accuracy for the shallow water of the Baltic Sea is not greater than that obtained using a differential satellite navigation receiver. After placing the module in its final position with depth 19,7 m (Table 1, Fig. 3), the cable was transported to the base ship, on which the physical field measurements were supervised.

During the measurement campaign, 12 measurements were made for 4 geographical directions and 3 ship's speeds (3 measurements for one geographical direction, and 4 measurements for one ship's speed). Different ship's speeds were performed for hydroacoustic and hydrodynamic pressure field measurements. It should be noted that the velocity, in this case, has no direct effect on the recorded magnetic field. These scenarios are presented in Fig. 4 and Table 2. Additionally, 5 physical field background measurements were performed using UMM (without *Zodiak* presence). The detailed measurement campaign schedule is shown in Table 3. The closest point of approach (CPA) in Table 3 denotes the minimum distance between UMM and *Zodiak* during particular passes (Fig. 4, Table 2).

Magnetic signatures obtained during the measurements for all 12 runs are presented in Fig. 5.

As mentioned above, during the measurement campaign of the ship *Zodiak*, the Earth's magnetic field background compensation was conducted independently for each ship passage by measuring range. Thus, the Earth's magnetic background was measured directly before each ship passage. During the background measurement, the ship was at a distance of more than 1 km from the center of the measuring range (measurement sensor location). Each time all three components of magnetic induction were recorded for 10 s, and its average values were used to determine/estimate the background, Earth's magnetic field. Obtained results are presented in Table 4 and shown in Fig. 6 according to the time of each pass, where all fluctuations of Earth magnetic fields variations during the measuring campaign can be seen.

It can be seen (Table 4, Fig. 6) that the fluctuation ranges of magnetic induction values during the measurements was 72 nT for the Bx component, 34 nT for the By component and 114 nT for the Bz component. In addition, the noise level of individual components of background magnetic fluxes was analysed, and its example results for 30 s of raw measurements are presented in the Fig. 7, where the deviation from average values (Table 4) may be observed. It should be noted that the observed background noise level (Fig. 7) is much lower than the magnetic field recorded in the presence of the ship *Zodiak* - a few thousand of nT (Fig. 5).
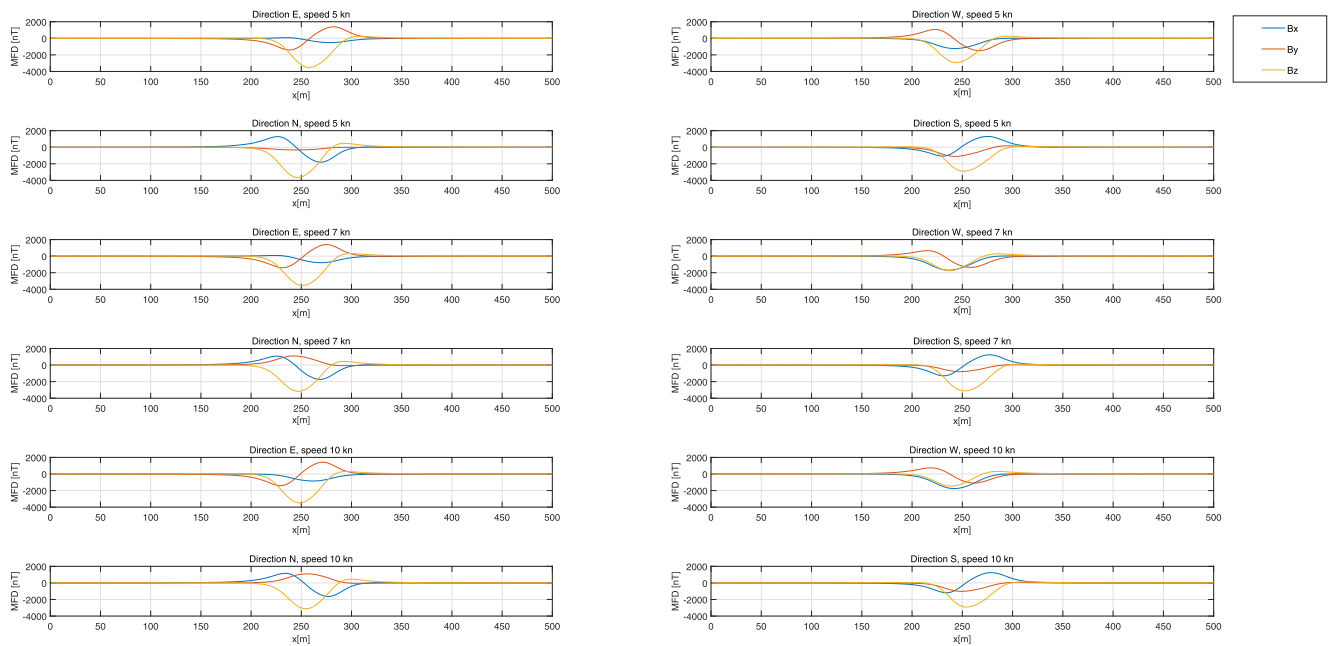
Fig. 5. Magnetic flux density (MFD) results of all ship's runs.

Table 4

The background, Earth magnetic field variations, which were determined directly before each passage of Zodiak through.

Path/Time	Bx [nT]	By [nT]	Bz [nT]	Path/Time	Bx [nT]	By [nT]	Bz [nT]
N5/8:17	16,071	−840	−45914	N10/13:02	16,084	−809	−45940
E5/8:38	16,066	−840	−45914	S10/13:27	16,132	−806	−46027
S5/9:06	16,060	−834	−45914	E7/14:30	16,080	−825	−45952
W5/9:28	16,060	−826	−45913	E10/14:55	16,100	−823	−45975
S7/11:58	16,073	−810	−45930	W7/15:05	16,085	−827	−45956
N7/12:11	16,077	−812	−45933	W10/15:59	16,082	−827	−45955
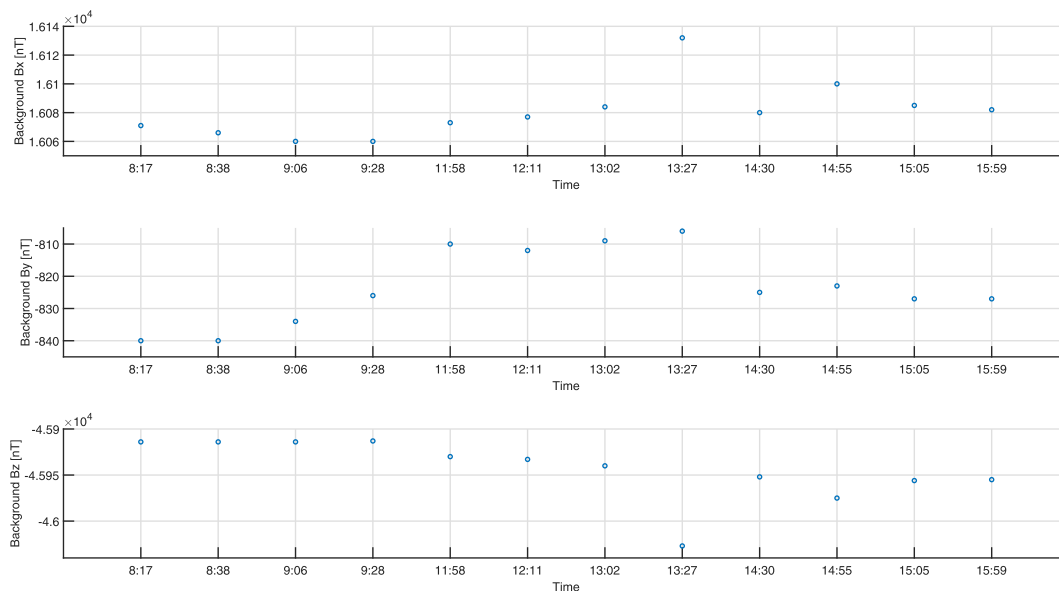
Fig. 6. The background, Earth magnetic field variations, which were determined directly before each passage of Zodiak through measuring range.

2.1. Accuracy aspect of the satellite navigation system

Determining the geographical position using such systems as GPS, GLONASS, GALILEO or BEIDOU requires appropriate calculations based on the signals emitted by navigation satellite transmitters and received by the antenna receiver. The method to calculate the geographical position based on the received signals is presented in [35]. The author describes the principle of operation of the satellite navigation receiver, focusing on signal parameters, i.e. frequency, amplitude and phase. The book indicates possible causes of errors in determining the geographical

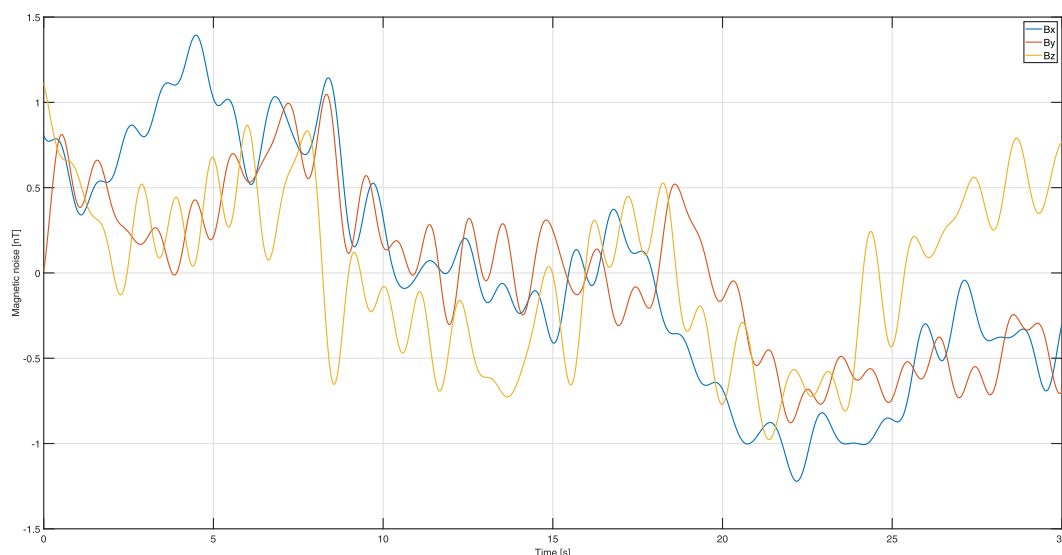**Fig. 7.** Example of the magnetic noise level of individual components of background magnetic fluxes measured at measuring range without the presence of a ferromagnetic object (Zodiak).
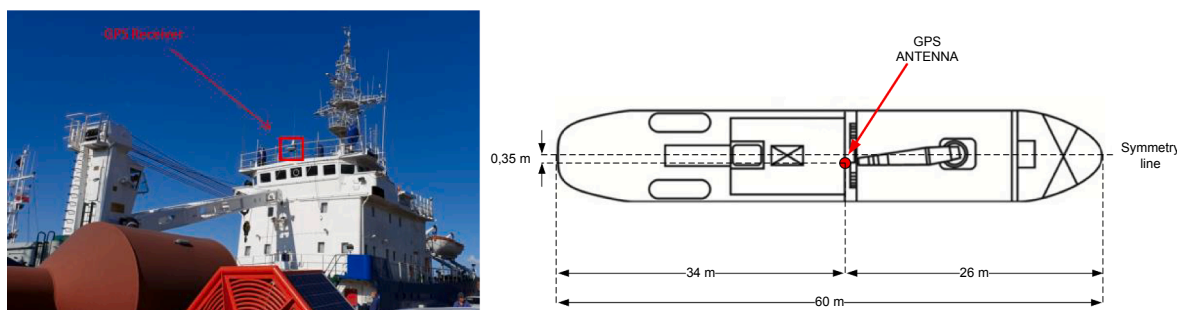

Fig. 8. Position of GPS antenna.

position in relation to the actual position.

The knowledge of exact geographical position is required in many aspects and areas of technology. This topic has been described in [36], in which the author indicated the problem of implementation limitations in new technologies using satellite navigation. The identified reason is the lack of long-term reliability of GNSS systems. The article describes the results of estimating moving object's course and speed in real time. NMEA 0183 protocols such as GPGGA and GPVTG were used for this purpose.

The use of NMEA protocols for satellite navigation systems is presented in [37]. The authors indicate the increasing demand for information, including: geographical position, speed, and direction of a moving object. The paper presents an algorithm increasing the efficiency of processing the data obtained from the GPS system, and thus reducing the time required for dynamic positioning.

The accuracy of geographical position determination depends on many factors. Increasing the real-time determination accuracy is possible with multiple GPS receivers. This technique for up to 8 receivers is described in [38]. The authors indicate the possibility of obtaining a better accuracy of the determined position through the method of averaging the results obtained from an individual receiver. The described method can be an alternative to differential systems and other advanced systems improving the accuracy of the determined position.

The positioning accuracy is closely related to the positioning method and systems used for that purpose. A solution to increase the accuracy of the geographical position may be the use of a system with radio corrections in the coastal zone. In Poland, such systems include the ASG-

EUPOS system used for precise measurements of the position of an object in space [39,40]. There is some limitation in the case of objects such as ships, where traditional GPS receivers are used to determine the instantaneous position with much lower accuracy than geodetic systems, including ASG-AUPOS.

Precise determination of the geographical position is used during the movement of ships in the area of increased activity, especially in port waters. Particular emphasis is placed on the mooring of large commercial ships carrying dangerous goods, such as tankers and LNG carriers. It is also important to observe possible movements of the ship while anchoring. This topic is discussed in [41], where the author describes the use of differential satellite navigation systems (DGPS) to improve positioning accuracy. This accuracy depends on factors such as satellite ephemeris, position of the receiver relative to the reference station, and changes in the troposphere and ionosphere. This issue is described and analyzed in [42]. The same aspect is cited by the authors of [43]. Various causes of errors in determining the geographical position, especially pseudo-range measurements and the instantaneous geometric configuration of satellites, are indicated.

The users of satellite navigation systems, in particular GPS, must be aware of certain limitation of dynamically determining the object's position. This limitation depends mainly on the time of observation of changes in the geographical position. This problem was analyzed by the authors of [44]. A specific item was verified in the period from 10 to 30 consecutive days, without taking into account the seasons. Based on the conducted experiments, it was determined that the accuracy deteriorates when observations are carried out in the time interval from 8

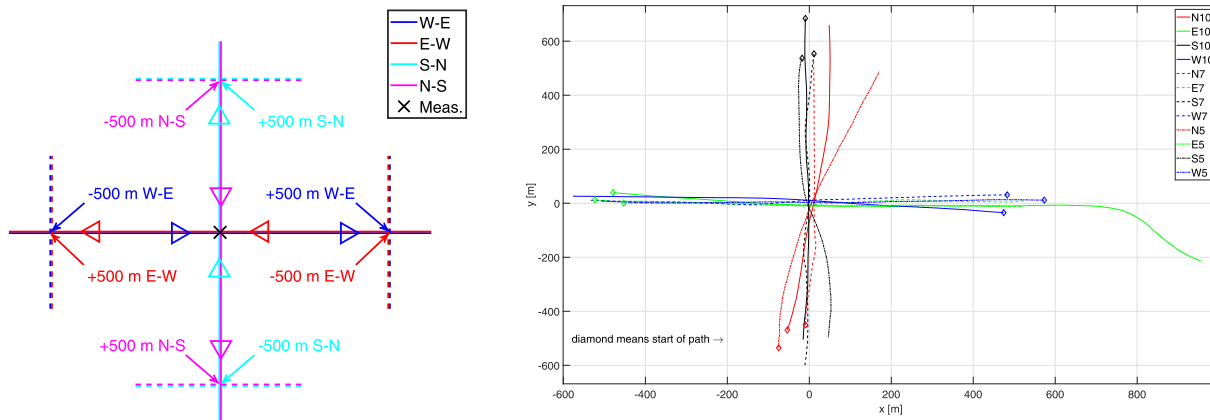**Fig. 9.** Real-time satellite navigation data recording system.

Fig. 10. Ship scenarios of transitions over the measurement module (left), all ZODIAK runs in Gdynia measurement range (right).

to 24 h.

Carrying out real-time measurements using multiple measurement subsystems requires reliable information about the exact start time for each subsystem. This information can be obtained from the received GPS time. The problem of time synchronization is discussed in [45,46]. In [45], the authors used logic circuits as field-programmable gate arrays (FPGAs) for time synchronization with an accuracy of single μs. Similar synchronization accuracy was obtained for seismic measurements [46].

The above brief presentation of accuracy aspects of using the existing satellite navigation systems for geographic positioning, especially in the marine environment, shows the need to analyze the influence of navigation data quality on the magnetic signature model synthesis in real sea conditions.

2.2. Navigating Zodiak

The measurements of magnetic field disturbances of *Zodiak* at the measuring range in Gdynia were accompanied with simultaneous recording of vessel's motion parameters, the most important of which were the current geographical position, speed, and course. The ship being the object of measurement was equipped with a satellite navigation system which provided differential position correction with a satellite compass. Used satellite navigation system Furuno SC-50 included an antenna with three differential receivers placed at equal distances around its circumference. The main technical parameters of the used satellite navigation system [47] are: heading accuracy 0.5°; heading

resolution 0,1°, and position accuracy 10 m (GPS), 5 m (DGPS), 3 m (EGNOS). The antenna of the satellite navigation system was placed amidships, in the area where there was no interference from other radio systems. The position of the antenna is shown in Fig. 8.

The main part of the satellite navigation system was the Furuno SC-501 processing unit, to which the Furuno SC-303 antenna and the Furuno SC-502 display were connected. The data output was connected to the computer via the Moxa NPort 5400 RS/Ethernet converter [48]. Two data outputs of the processing unit were used for synchronous data recording on a computer and ship's trajectory preview on the plotter screen. The entire system worked on a common Ethernet subnet. The navigation data recording system is shown in Fig. 9.

The application developed for data recording enabled real-time preview of the received NMEA 0183 protocols while saving data to files. The data extracted from the available protocols included [36,37]: global positioning system fix data (GPGGA), track made good and ground speed (GPVTG), date and time (GPZDA), true heading (GPHDT).

The vessel's satellite navigation data recording system required the operator's presence. By analyzing the current geographical position of the vessel, the operator decided when to start saving the file. At the same time, he passed, using VHF radio, the information about reaching the starting position to the operator of the magnetic field measurement system.

The measured vessel was moving with the set speed and course from at least 500 m before to 500 m behind the position of the measuring module. This distance concerned all directions: N, S, E, and W. The
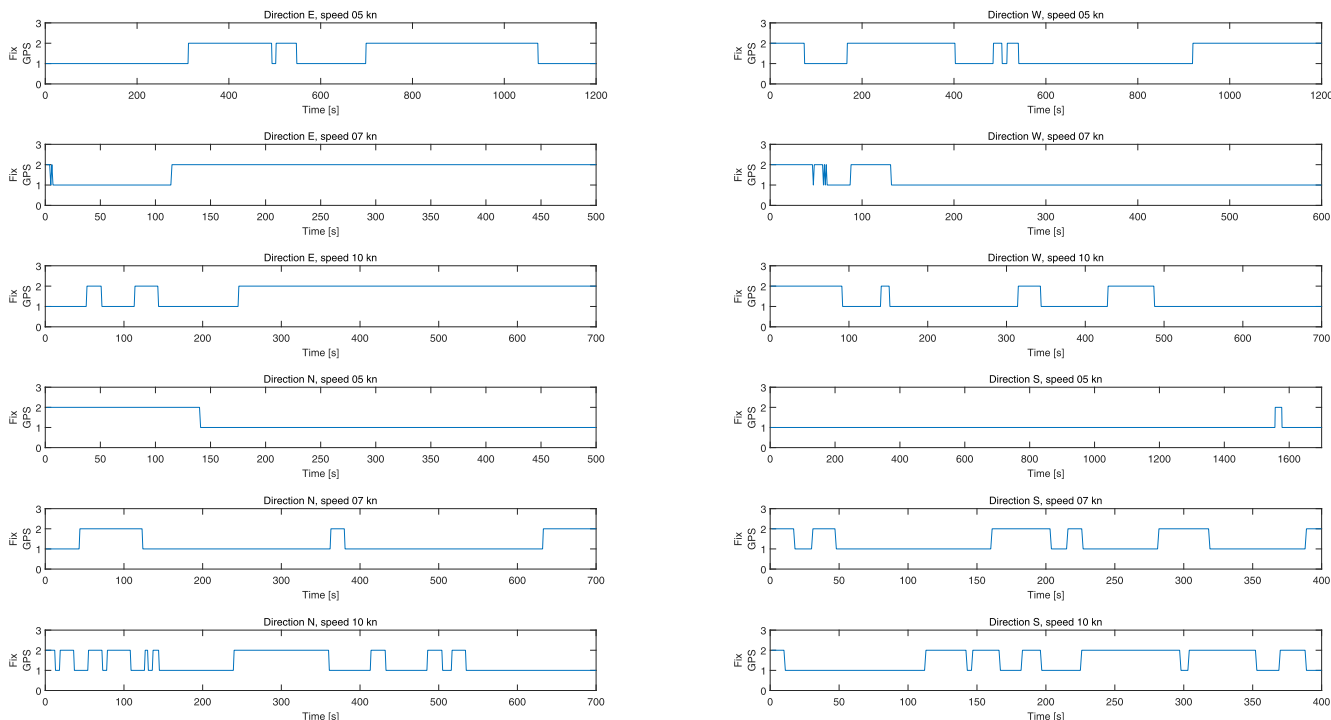
Fig. 11. GPS Fix parameter, 0 – incorrect, 1 – GPS correction, 2 – DGPS correction.
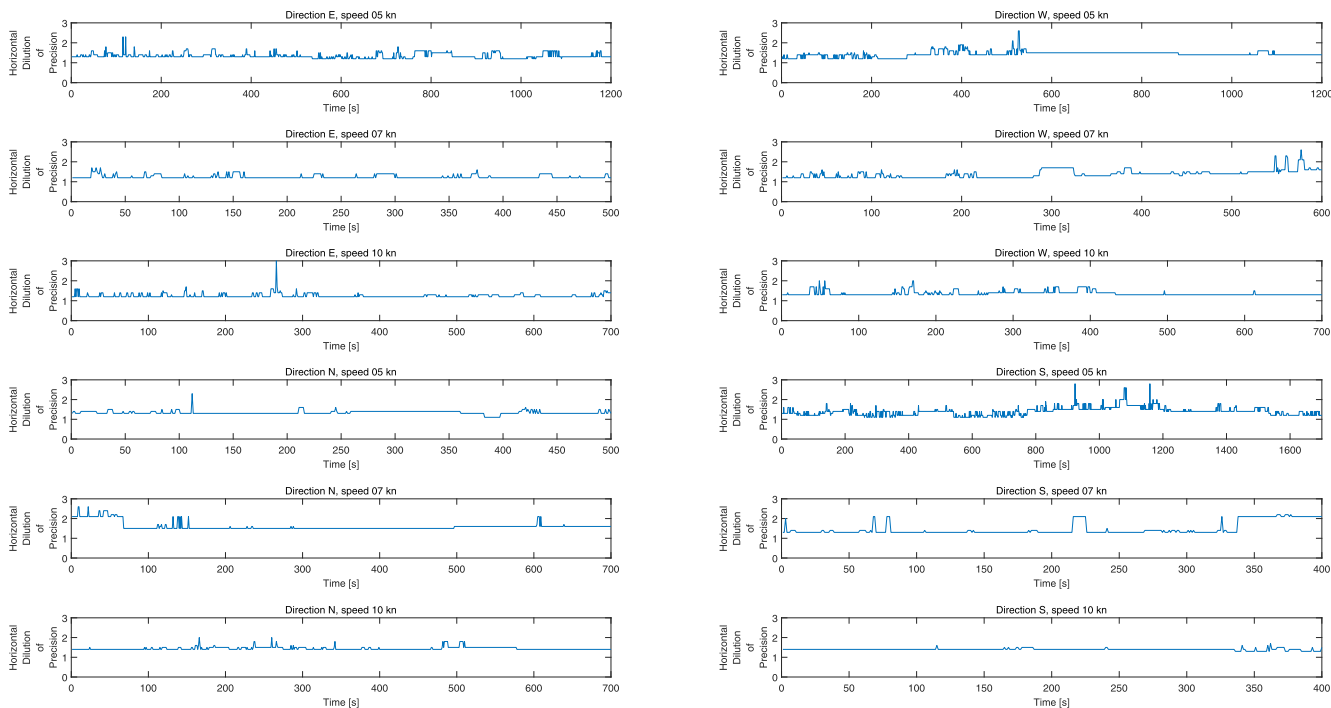
Fig. 12. GPS Horizontal Dilution of Precision (HDoP) parameter.

scenarios of the navigation data recording starts and stops are shown in Fig. 10. The run of the vessel over the measurement module with the set motion parameters was associated with the trajectory record from the satellite navigation system. A separate navigational data file was saved for each transition. The recorded data was analyzed and processed to obtain a series of navigation information, which included ship trajectories converted from geographic positions expressed in angular units to the metric scale, as shown in Fig. 10. For this purpose, the ready-made

functions of the Matlab environment were used.

The qualitative analysis of the recorded data was also performed. The value specified in that case was the type of the currently applied fix, i.e. GPS Fix. The results of this analysis are shown in Fig. 11.

Fig. 11 shows dynamic changes of the GPS Fix parameter between 1 (GPS correction) and 2, relating to the DGPS differential correction. For GPS Fix = 1, the accuracy is up to +/- 12 m, while for GPS Fix = 2, it is up to +/- 5 m [49], which means the increase by more than twice. The

Fig. 13. Number of satellites parameter.

transition for which the GPS Fix value was 0 was interpreted as the position determination error and was not recorded. Out of 12 transitions, those can be selected for which the correction value changed dynamically during the measurement. These transitions include: Direction N, speed 10 kn, Direction S, speed 7 kn, and Direction S, speed 10 kn. The transitions for which these changes showed lower dynamics were: Direction E, speed 5 kn, Direction W, speed 5 kn, and Direction W, speed 10 kn. The corrections for the remaining transitions showed slight changes. For example: GPS Fix = 1 was recorded 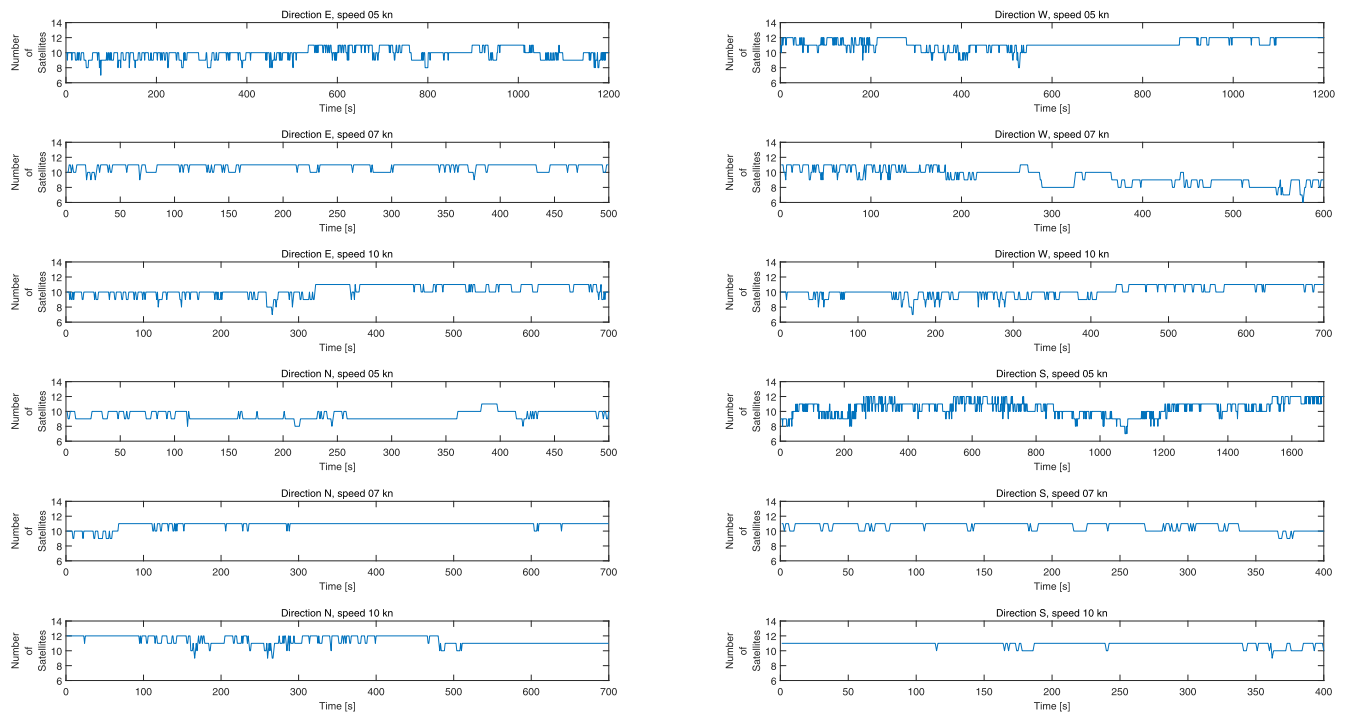in a significant part of the transition for Direction W, speed 7 kn, Direction N, speed 5 kn, Direction S, speed 5 kn, and Direction N, speed 7 kn. The GPS Fix = 2 (DGPS correction), was recorded in the most part only for the transitions: Direction E, speed 7 kn, and Direction E, speed 10 kn. Analyzing the data presented in Fig. 6, it can be concluded that in most cases the accuracy of a given position was determined by GPS correction, i.e. it was about 12 m. A similar analysis was performed for the parameter Horizontal Dilution of Precision (HDoP). The results for all 12 passes are presented in Fig. 12.

When analyzing changes in the HDoP parameter value during the measurements, it can be noticed that these changes ranged from 1 to 3. The higher the value, the greater the probability of an error in determining the geographical position [50]. It also means a reduction in the accuracy of the designated position. Values <1 indicate the ideal positioning of the satellites in the sky. HDoP is determined based on the mutual position of the satellites in relation to the satellite navigation receiver [51–53]. For the 12 transitions presented, the assumed HDoP value meant relatively good geometric conditions for the arrangement of the satellites in relation to the receiver. The last analyzed parameter concerning the quality of determining the geographical position was the number of satellites. The summary of changes in the number of satellites is shown in Fig. 13.

During the measurements, the number of satellites from which the geographical position was determined varied from 6 to 12. Some transitions, such as Direction N, speed 7 kn, and Direction S, speed 10 kn, were characterized by low variability of this value. In contrast, high variability was observed for the Direction S, speed 5 kn transition, for which this value varied from 7 to 12 throughout the range. A similar

situation was with the Direction W, speed 7 kn transition, where the number of satellites varied from 6 to 11. The analysis of the quality of determining the geographical position showed high dynamics of changes in individual parameters, including Fix GPS, HdoP, and the number of satellites. One of the main reasons was the duration of the measurements. All of these measurements lasted over 8 h, during which the accuracy of determining the geographical position changed [44].

2.3. Measurements applicability aspects

The distance of the ship's trajectory from the extreme sensor undoubtedly affects the information value of the disturbance and affects the quality of the model. A typical stationary measurement range is built with the use of several sensors. There are different topologies for sensor placement, but the most common are line or cross shapes. In the case of approaches to ports or straits, the sensors are usually located in the center of the approach fairway, which results in the fact that most of the units pass within the sensor line.

In the case considered in the article, i.e. a single sensor, the situation with registering the appropriate signature quality directly depends on the distance between the ship and the only sensor. When planning a measurement experiment, it is assumed that the ship will pass exactly over the sensor, but it is not always possible due to the accuracy of the GPS locating device and the hydrometeorological conditions.

The experience of measurement crew, included in informal heuristics, indicates that the passage of the vessel from the side sensor should not exceed two widths of the tested vessel for a measurement depth comparable to the vessel's width. If the measurement depth is much greater than the ship's breadth, then the distance between the extreme sensors should be equal to twice the measurement depth.

This estimate is quite crude and in practice an attempt is made to get the closest possible transitions to the sensor. Therefore, in an earlier article by the authors, based on synthetic data, data from a line of 3 sensors were taken into account, the extreme magnetometers of which were located at a distance equal to half the width of the analyzed object. This allowed for very good results. The experience gained in the SIRAMIS II project shows that the good quality of the mapping of magnetic
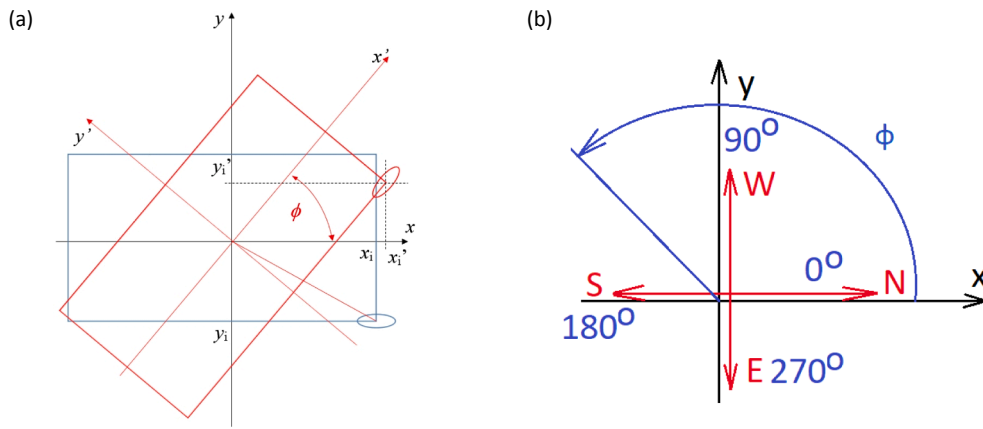
9

Fig. 14. The coordinate system and direction of angle progress (a), and the orientation of geographic directions used in the multi-dipole model (b).

signatures in terms of mean square error on the paths of ship crossings, defined as the difference between the measured values and those determined from the model of individual magnetic induction components, can be maintained when the ship transitions in relation to the extreme sensor are not greater than the breadth of the ship.

In the specific case analyzed in this article, we had only one sensor, and the maximum CPA value (Table 3) of the ship's trajectory from the sensor was 12 m, the average of all analyzed passages was ∼ 7 m, with the ship's width being 10.8 m.

3. The multi-dipole model

3.1. Mathematical formulation of the multi-dipole model

The multi-dipole model [21,22] provides flexibility in reconstructing the entire magnetic field disturbance map around a ferromagnetic source object located at the origin of the Cartesian coordinate system. The model is based on the aggregation of partial results of individual magnetic flux densities generated by m permanent and n induced dipoles defined in the multi-dipole model. Each i-th dipole, permanent or induced, within the multi-dipole model is characterized by six variables which describe its location in Cartesian coordinates (x_i, y_i, z_i) and magnetic moments ($m_{x,i}$, $m_{y,i}$, $m_{z,i}$) in each orthogonal direction (x, y, z). The general structure of the multi-dipole model is presented as follows [10,54]

$$\mathbf{B} = \sum_{i=1}^{m+n} \mathbf{B}_i(\mathbf{M}_i, \mathbf{R}_i) = \sum_{i=1}^{m+n} \frac{\mu_0}{4\pi} \left(\mathbf{R}_i^T \mathbf{M}_i \mathbf{R}_i \cdot \frac{3}{\mathbf{R}_i^5} - \frac{\mathbf{M}_i}{\mathbf{R}_i^2} \right),$$ (1)

$$\mathbf{B} = \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix}, \quad \mathbf{B}_i = \begin{bmatrix} B_{x,i} \\ B_{y,i} \\ B_{z,i} \end{bmatrix}, \quad \mathbf{M}_i = \begin{bmatrix} m_{x,i} \\ m_{y,i} \\ m_{z,i} \end{bmatrix}, \quad \mathbf{R}_i = \begin{bmatrix} (x - x_i) \\ (y - y_i) \\ (z - z_i) \end{bmatrix},$$ (2)

$$R_i = |\mathbf{R}_i| = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2},$$ (3)

$$B_T = \sqrt{\sum_{i=1}^{n+m} (B_{x,i})^2 + \sum_{i=1}^{n+m} (B_{y,i})^2 + \sum_{i=1}^{n+m} (B_{z,i})^2},$$ (4)

where: **B** is the magnetic flux density vector, $\mathbf{B}_i$ is the magnetic flux density vector of i-th magnetic dipole, where $i = 1,...; m, m + 1,...,m + n$, m is the number of permanent dipoles, n denotes the number of induced dipoles, $\mathbf{M}_i$ is the vector moment of i-th magnetic dipole in each orthogonal direction (x, y, z), $\mathbf{R}_i$ denotes the distance vector of the analysed point (x, y, z) from the i-th dipole with coordinates (x_i, y_i, z_i) along (x, y, z) directions, $R_i = |\mathbf{R}_i|$ denotes the distance, as the vector length, between two points (x, y, z) and (x_i, y_i, z_i), B_T denotes the total magnetic flux density, and $\mu_0 = 4\pi \cdot 10^{-7}$H/m is the absolute magnetic permeability of space.

Ships are dynamic objects and change their position and course intentionally or due to external disturbances related to e.g. sea currents, wind, or sea waves. Therefore, it is necessary to consider in the multi-dipole model, Eqs. (1)–(4), appropriate corrections of positions and moments of individual dipoles depending on ship course changes.

The proposed correction mechanism related to appropriate transformation of i-th dipole position according to the change of ship heading φ is graphically presented in Fig. 14(a) and can be described by the following equations [22]

$$x_i^{'} = x_i \cdot cos(\phi) - y_i \cdot sin(\phi),$$ (5)

$$y_i^{'} = x_i \cdot sin(\phi) + y_i \cdot cos(\phi),$$ (6)

where: φ is the object's heading, and $x_i^{'}$ and $y_i^{'}$ denote the transformed coordinates of the i-th dipole location (rotation does not change the z coordinate). Fig. 14(a) shows how the dipole position (x_i, y_i) for direction N (course φ = 0°) is transformed to the new location ($x_i^{'}, y_i^{'}$) in direction φ = 45°. Particular geographical directions in the multi-dipole model are defined as follows: N (North, 0°), W (West, 90°), S (South, 180°) and E (East, 270°), – Fig. 14(b) [22]. Finally, the new coordinates ($x_i^{'}, y_i^{'}$) update the old coordinates (x_i, y_i) of the i-th dipole in the multi-dipole model – Eqs. (1)–(4).

The values of permanent magnetic moments of the identified permanent dipoles characterizing the ship strongly depend on its 'magnetic history' during exploitation and should be estimated based on real measurements. In the paper, it is assumed that the permanent magnetic moments do not change significantly in a short period of time, and that they only depend on ship's heading φ and not on its geographical position [22].

Hence, as a result of changes of ship's heading φ, Eqs. (5) and (6), the estimated components of each permanent dipole should be transformed as well. The following correction mechanism is proposed to describe this transformation in the Cartesian coordinate system for all permanent dipoles ($i = 1,...,m$) [22]

$$\mathbf{M}_i^{'} = \begin{bmatrix} m_{x,i}^{'} \\ m_{y,i}^{'} \\ m_{z,i}^{'} \end{bmatrix} = \begin{bmatrix} m_{xP,i} \cdot cos(\phi) - m_{yP,i} \cdot sin(\phi) \\ m_{xP,i} \cdot sin(\phi) + m_{yP,i} \cdot cos(\phi) \\ m_{zP,i} \end{bmatrix},$$ (7)

where: $m_{x,i}^{'}$, $m_{y,i}^{'}$, $m_{z,i}^{'}$ denote the transformed permanent magnetic moments of i-th permanent dipole, and $m_{xP,i}$, $m_{yP,i}$, $m_{zP,i}$ denote the components of permanent magnetic dipole moments in orthogonal x, y, z directions before transformation ($m_{x,i}$, $m_{y,I}$ and $m_{z,i}$ in multi-dipole model – Eqs. (1)–(4)). Finally, the new vector $\mathbf{M}_i^{'}$ of permanent magnetic moments of i-th permanent dipole updates the vector $\mathbf{M}_i$ in the multi-dipole model – Eqs. (1)–(4). Notice that the base heading of the ship in the multi-dipole model, Eqs. (1)–(4), is defined in the direction N (North, 0°) – Fig. 14(b).

The values of induced magnetic moments of the considered induced dipoles are directly related to ship's geometry, geographical position, and heading [22]. In the multi-dipole model, Eqs. (1)–(4), they are represented by a combination of induced moments $m_{I1,i}$, $m_{I2,i}$ and $m_{I3,I}$ according to the following formula [22]

$$\mathbf{M}'_i = \begin{bmatrix} m'_{x,i} \\ m'_{y,i} \\ m'_{z,i} \end{bmatrix} = \begin{bmatrix} m_{I1,i} + m_{I2,i} \cdot cos^2(\phi) \\ m_{I2,i} \cdot sin(\phi) \cdot cos(\phi) \\ m_{I3,i} \end{bmatrix}, \tag{8}$$

where: $m'_{x,i}$, $m'_{y,i}$, $m'_{z,i}$ denote the transformed induced magnetic moments of i-th induced dipole. Finally, like in the previous case, the new vector $\mathbf{M}'_i$ of induced magnetic moments of i-th induced dipole updates the vector $\mathbf{M}_i$ in the multi-dipole model – Eqs. (1)–(4).

It should be noted that the multi-dipole model was established to describe the magnetic field in a mathematical manner, it is based on the physical description of a single magnetic dipole, but it is not a strict physical model in the sense of multiple dipoles. The heterogeneity of magnetic field synthesis means that there can be many different realizations of the arrangement and values of the magnetic moments of the dipoles describing the same magnetic field. Therefore, in general, parameter values cannot be expected to have a physical representation related to the object structure. They will be selected by the optimisation procedure in a mathematical way to provide the best fit of the modeled field to the reference data.

Due to the non-physical of the model, the number of dipoles needed to reproduce the magnetic signature in terms of a given field/path is unknown. It is an important and necessary part of the synthesis of the multi-dipole model. The choice of the number of dipoles depends on e.g. on the following factors: vessel size, vessel distance from sensors, vessel shape and structure. When using the multi-dipole model, it should be kept in mind that a single dipole can reproduce with satisfactory accuracy the magnetic signature of a vessel located far enough from the sensor. In the case of closer distances, a single dipole starts to be insufficient, and it is necessary to use more of them.

3.2. Previous applications of the multi-dipole model

The multi-dipole model has been verified both with the use of synthetic data and, to a limited extent, with the real data. The articles, widely available according to the idea of [22], considered the construction of the multi-dipole model for a virtual object in the form of a symmetrical ellipsoid. The object had both permanent and induced magnetization. Research was conducted to study both the quantity and diversity of the data needed for model learning, and the structure of the multi-dipole model. The research of the input data package on which the model will be taught included the number of directions from which the measurements come, as well as the resolution and, consequently, the number of data necessary for correct determination of signatures. The research of the multi-dipole model structure concerned the number and permissible positions of the dipoles. The research started with the possibility of locating the dipoles along the ship's keel and sides (1D), and was completed in 3D space, in the contour of the ellipsoid approximated by means of a cuboid. The number of dipoles in the model should be large enough to enable high-quality mapping of the signature, but it should not be too large, so as not to increase unnecessarily the computation time, and to avoid the phenomenon of model overparameterization. The issue of determining the number of dipoles needed to describe a specific object is difficult to analyse using a mathematical formula. Rather, it is possible to formulate heuristics and obtain this information through simulation experiments.

As a result of the computational work, it was found that in the model learning phase, it was sufficient to provide data from four cardinal directions and 3 object speeds in each direction, from the range −100 to 100, every meter. In total, therefore, the model was supplied with 4 * 3 * 201 samples from triaxial magnetometers containing the information

about MFD components Bx, By, Bz. Such a set of data turned out to be sufficient in terms of quantity and diversity to build a model capable of reproducing magnetic signatures in any direction and at any depth with very high accuracy. Additionally, tests of resistance to disturbances caused by potential measurement noise related to MFD were carried out. The multi-dipole model was very resistant to noise at the level guaranteed by the manufacturers of real measuring devices, and also at the level many times greater [22]. To describe an object of this size, with no sophisticated shapes and symmetry, it was enough to use 20 dipoles: 10 of induced and 10 of permanent type. This number was determined experimentally by increasing the number of dipoles as part of the simulation until the structure was so flexible that it was possible to achieve very low errors at the stages of both model training and validation.

The Authors made heavy use of the multi-dipole model during the SIRAMIS II project. The activities included calculations using synthetic data from simulators such as Opera 3D [55] and real data recorded on three main stationary and well-equipped test sites in Germany, Norway and Sweden. Additionally, it was supplemented by the use of data obtained at port container ship. In the case of synthetic data, we had an ellipsoid-shaped object, a corvette and a container ship at our disposal. Each of these objects differed in size, structure and shape. In the case of data recorded on the training grounds, it was a dozen different merchant ships. In the case of data from port entries, there were several dozen units of a very different nature: RoRo Vessel, Passenger Ferry, Bulk Carrier, Cargo Ship, Container Ship, Tanker, Tug. As soon as the ship moved over the area covered by the sensors, it was possible to reproduce its magnetic signature with at least satisfactory quality. Due to partial confidentiality of the survey ranges data, no details can be presented, but general features may be discussed. Unlike the approach with synthetic data and virtual testing grounds, the real ranges are usually located in straits, canals, fjords in such a way that it is possible to pass the vessel only in two directions, and it is impossible to make measurements in perpendicular directions. This results in much less data variation. To compensate for this, the measuring ranges are equipped with a number of sensors, for example 5 or 7, arranged in a line. An additional interesting feature of real ranges is the location of magnetometers at different measurement depths. The multi-dipole model, trained on multiple sensor data but taken only from one direction, revealed good reproduction properties of magnetic signatures. Significant errors appeared mainly when the model was trained with the data obtained from measuring devices located closest to ship path while the verification took place on data from magnetometers located extremely outside. However, when the dual-directional data model learning was applied, it significantly improved the reproduction quality. It can be treated as a feature or property of the multi-dipole model that for correct mapping of signatures it requires an appropriate amount, but also a variety of training data, e.g. the data from various directions.

Research with the use of synthetic data provides greater flexibility because you can create an arbitrary virtual object with the use of FEM, and the test site with an arbitrary set and configuration of measuring devices and disturbance profile. On the other hand, research with the use of real data allows to determine the usefulness of the model for practical applications burdened with natural measurement errors. Having obtained very good results of mapping the magnetic signatures with the multi-dipole model for both simulation scenarios and real measurement campaigns, it was decided to use the experience gained earlier to apply this model to research with *Zodiak*.

3.3. Preprocessing data measured on Zodiak for modelling purposes

3.3.1. Transferring measured data into model space

The data obtained from the measurements contained the synchronized Lat, Lon, Bx, By, Bz, Head, GpsFix, and GpsHDoP vectors on individual paths. The first step in adjusting the measured data to the needs of the model was the conversion from the description of longitude and
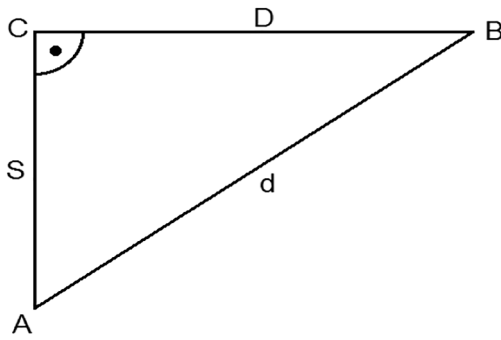
Fig. 15. Rhomboid triangle approximated to flat triangle.

Table 5
Information about X and Y ranges and data resolutions used in different scenarios.

Number of probes on path (dataset name)	X_{min} or Y_{min} [m]	X_{max} or Y_{max} [m]	Resolution [m]
8001	−400	400	0.1
5001	−250	250	0.1
3001	−150	150	0.1
501	−250	250	1
301	−150	150	1

latitude to the Cartesian X-Y coordinate system.

The first method used to calculate was based on the following formula:

$$d = \sqrt{D^2 + S^2}, \tag{9}$$

$$D = 1852 \cdot (\lambda_1 - \lambda_2) \cdot \cos(\varphi), \tag{10}$$

$$S = 1852 \cdot (\varphi_1 - \varphi_2), \tag{11}$$

where: d denote distance between two points A and B expressed in meters, described by geographical coordinates $A(\varphi_1, \lambda_1)$ and $B(\varphi_2, \lambda_2)$, D is a length of line between meridians λ_1 and λ_2, S is a line along the meridian λ_1, between latitudes φ_1 i φ_2, λ_1 and λ_2 rae longitudes of points A and B in arc seconds, φ_1 and φ_2 are latitudes of points A and B in arc seconds, φ is a latitude of points C and B.

The method is based on a rectangular flat triangle being the approximation of the rhomboid triangle. The flat triangle is presented in Fig. 15.

The second method makes use of the Matlab Mapping Toolbox [56]. The procedure *grn2eqa* converts the data from Greenwich coordinates to equal area coordinates. The compliance of both methods was very high. Determining the ship's position in relation to the measurement column is potentially the largest source of model error, therefore, the accuracy of this conversion was taken care of to marginally affect the quality of the model. After the conversion of geographic coordinates to the X-Y system, there was still the question of adapting to the coordinate system used in the multi-dipole model (Fig. 14(b)). By analysing the axes and directions in both systems, it was found that the necessary conversion is described by:

$$\begin{aligned} x_{MD} &= -y_{XY}, \\ y_{MD} &= x_{XY}, \end{aligned} \tag{12}$$

where: x_{MD} and y_{MD} denotes x and y coordinate in the coordinate system of the multi-dipole model, x_{XY} and y_{XY} denotes x and y coordinate in the X-Y coordinate system.

3.3.2. Data area selection and downsampling

In the measurement campaign, the data is recorded with maximum accuracy and resolution of the measurement devices. Usually, the recorded data is redundant, i.e. it exceeds the amount needed by the model. For example, the *Zodiak* data were recorded within the range of −500 m to 500 m, every 1 ms, which at a speed of 5 kn gave a new measurement sample every 0.26 cm and a total of more than 1E5 samples on one path. In total, 12 paths were recorded (4 directions with 3 speeds), which gives quite a large portion of data. The measurement data used to develop the model will be subject to many iterations of optimisation calculations. Mainly for this reason, the number of data is limited to one that is significantly smaller than the set obtained from the measurements, and at the same time carries the necessary information value. In the synthetic data approach, good results were obtained using data ranging from −250 m to 250 m with 1 m step, i.e. a total of 501 samples on a single path. The same conditions were used as the basic approach in the calculations reported in this article. The relevant sections also analysed the impact of using 8001 samples in the −400 m on the path, and the sparse datasets. All combinations of the used numbers of probes, resolutions, and data ranges are presented in Table 5.

3.3.3. YawPitchRoll compensation of MFD components

Another difference between real data and synthetic data is the need to take into account the so-called 'Local coordinate system'. Despite attempts to set it perfectly in the X-Y-Z order in which the measurements are to take place, the head (stand) with the measuring devices may have a certain error in setting, which should be compensated before downloading the MFD data to the model. Compensation takes place using Euler's formulas: the values of MFD components recorded by the magnetometers are corrected in such a way as to obtain the values that the magnetometers would show if they were set in the ideal X-Y-Z order (Fig. 16).

3.3.4. Using heading of ship for magnetic signature synthesis

The multi-dipole model requires ship's heading to calculate its magnetic signature. The ship was assumed to move on a given heading but in practice its heading differs from that assumed. This information is entered into the model in the form of the φ parameter (6, 7), as presented in Fig. 14. The ship's captain receives the trajectory to be followed and tries to keep it as accurately as possible, but due to drift, wave and inaccuracies of on-board measurement devices, he achieves a passage consisting of many sections with different geographic directions. There are three main approaches to this situation: (1) ignoring all deviations and treating the assumed course as the complete course, (2) using the information from each step of ship's journey to calculate the average heading and substituting it in the model, or (3) including heading in the model at each measurement step, which requires adjusting the model. The article uses the last approach as the most accurate and using the maximum amount of information available.

3.4. The multi-dipole model parameters identification fitting/validation combinations

Determining model parameters by way of optimisation is called model training and is marked as FIT (from parameter matching), while model verification, i.e. checking the quality of the model on data that was not used for training, is called CV (cross validation). To summarize the information collected in the section 3.3 12 data vectors, with different number of samples, representing the ship's passage in different ranges in four directions with three different speeds of 5 kn, 7 kn, and 10 kn are available for analysis (Table 4).

These data must be sufficient both to teach the FIT model and to verify the CV. Therefore, a necessary step is to split the data set into separate parts for teaching and validation. There are several methods and canons of data division. The key factor that determines how the data set is divided is its size. If we have a redundant amount of data, i.e. sufficient for excellent learning from a subset, we can separate a rich test set and comprehensive tests. The optimisation process results based on
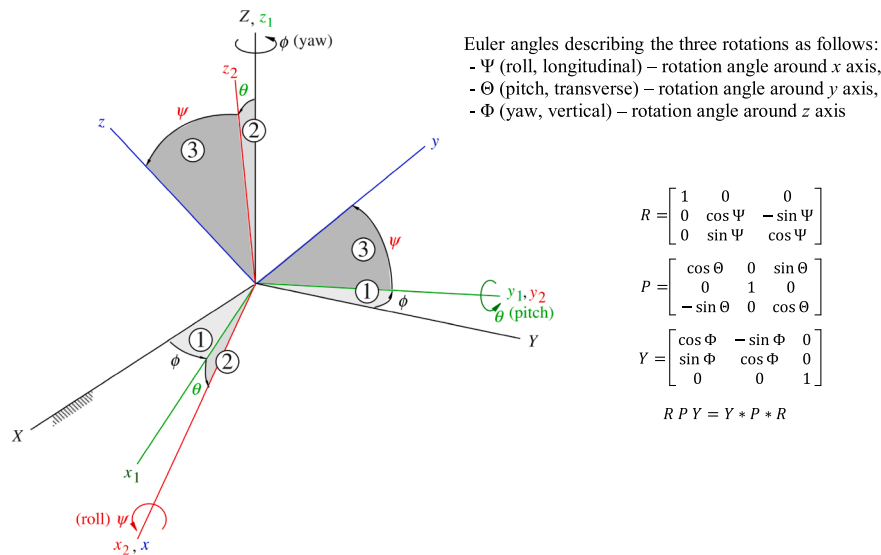
Fig. 16. Visualization of Euler angle rotations [57,58].

synthetic data showed that using 12 paths provided very good conditions for determining the parameters of the model and obtaining its robustness [22]. Data diversity, i.e. providing data from different directions, is also important. Reducing the available set will result in loss in teaching, robustness, and predictive ability of the model. Taking this into account, this article is based on the leave-one-out (LOOCV) solution [59], which consists in using all data from the available set, except for one path on which the verification will be performed.

The cross-validation according k-fold approach needs a k-hyperparameter to be specified [60]. The whole dataset is then divided into k subsets. The LOOCV is a variant of k-fold cross-validation where k means all subsets in the dataset. An example illustrating this approach and the notation used in the article will be the situation in which from the data set of all 12 paths will be separate 11 paths for fitting (FIT) and one for cross-validation (CV), i.e. E10-W10-S10-N7-E7-S7-W7-N5-E5-S5-W5 paths for FIT and one path N10 for CV. There are 12 such combinations with the data from 11 paths used for training and 1 for validation, and they are the 'richest' and most representative forms of experiments that can be performed. It is worth emphasizing that when there are 11 paths in the training set, the data from these sets are treated jointly and the parameters are matched with the same weight for all these data. However, it was decided to make an additional division of the data set for the purpose of checking the integrity of the data obtained as a result of the measurements. Therefore, all possible combinations, in the scheme 'number of FIT paths and one CV path', have been identified for data sets consisting of 4 (3–1), 8 (7–1) and 12 paths (11–1). These combinations will constitute the main material for inferring about the used data quality, the learning process, and the ultimate predictive ability of the model. In addition, sets of 4, 8 and 12 paths with data were separated as a set of auxiliary data, the latter cannot be verified because it includes all available data, so it is entirely intended only for the assessment of the teaching process and comparison with other analysed scenarios. The first dataset was based on the scheme of 3 paths to fitting (FIT) the multi-dipole model and 1 path for its cross-validation (CV) within the same speeds. Finally 12 datasets were investigated. The second dataset was composed of three combinations of data from four directions at the same speeds. The third dataset consisted of 24 combinations of data from the total combination of two speeds in scheme of 7 paths for FIT and 1 path for CV. The fourth dataset consisted of three combinations of data from four directions at a total of 2 out of 3 available speeds. The fifth dataset consisted of 12 combinations of 11 paths for FIT and 1 path for CV. The last element of the simulation dataset is FIT on all 12 data vectors. Therefore, in total, 12 + 3 + 24 + 3 + 12 + 1

= 55 simulation scenarios were defined. All described path combinations are presented thereafter in Table 9. The selected simulation scenarios, although requiring a huge amount of calculations, should nevertheless provide comprehensive information about the quality of the input data, the learning process, and the learned multi-dipole model.

3.5. Optimisation procedure – Fitting phase

The nonlinear least squares procedure *lsnonlin* from the Matlab package was used to determine the parameters of the multi-dipole model. In the constraint version, it uses the trust-region-reflective [61,62] algorithm.

To clearly describe the conditions of the optimisation process, the following should be provided: input data, optimisation (quality) criterion, number of decision variables, constraints on decision variables, starting point, stopping conditions, and specific features of the optimisation process.

The data sets specified in the previous chapter were used as input. The number of input data in each scenario depends of paths number multiplied by number of samples for each of the three MFD components Bx, By, Bz.

One of the more important parameters of the multi-dipole model is the number of dipoles used to describe the MFD. This number depends on many factors: the depths at which the sensors are located, the ship's draft (i.e. the distance between the sensors and the ship), the size of the ship, and its so-called magnetic history. The model includes permanent and induced magnetism consisting of two sets of dipoles - induced and permanent. For the sake of simplicity, it has been assumed that the number of induced and permanent dipoles is the same. There are no mathematical formulas or recognized heuristics to define a priori the number of dipoles that represent the optimal structure of the model. The determination of this number is experimental, and in this article simulation was used for this purpose. In the presented model, each dipole is described by six parameters (mx, my, mz, x, y, z). The number of model parameters determined by the optimisation procedure results therefore from the sum of induced and permanent dipoles numbers, which is multiplied by six (number of parameters characterising each dipole). Certainly, for the optimisation procedure to be properly conditioned, it is desirable that the number of input data (the number of constraints in the form of equations) is significantly greater than the number of the determined model parameters.

An exhaustive search, calculations in the joined space of the number of parameters (1–100) and the number of iterations (1–1000) of the
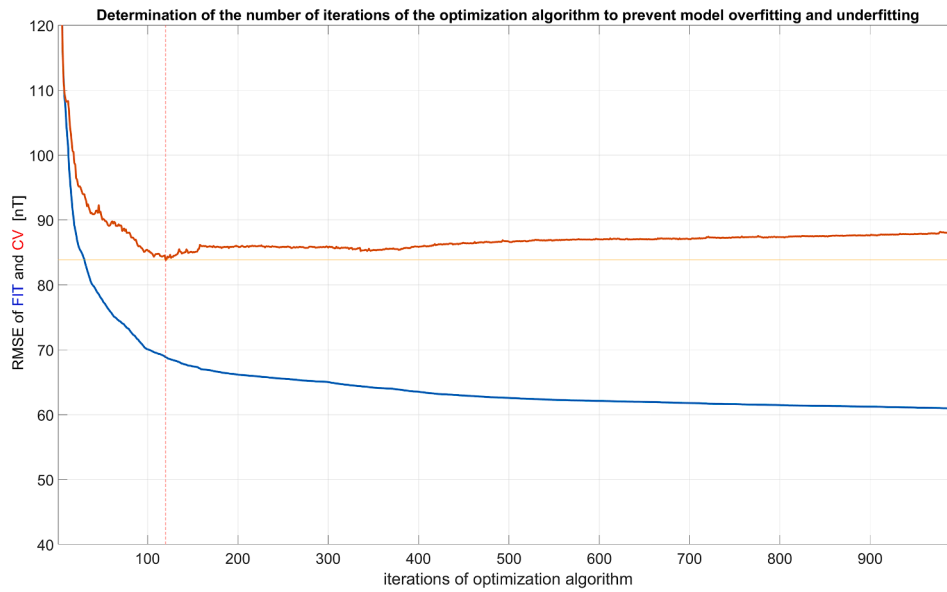
Fig. 17. Fit and Cross Validation results of the optimisation algorithm iteration function.

Table 6

Optimisation process termination parameters.

Stop condition 'TolX', 1E-30, 'MaxFunEvals', 1E30, 'TolFun', 1E-30, 'MaxIter', 124

Table 7

Zodiak dimensions and limitations in X, Y and Z axis directions as input parameters of multi-dipole synthesis process.

Length = 61.3; Breadth = 10.8; Draught = 3.2; opt.LB = [−5E6, −5E6, −5E6, −1.1*Length/2, −1.1*Breadth/2, −1.1*Draught]; opt.UB = [5E6, 5E6, 5E6, 1.1*Length/2, 1.1*Breadth/2, 15];

optimisation algorithm have been done. The observed quality index was the mean value of the RMSE_CV coefficient calculated on all 12 datasets (11-paths FIT). These calculations were very time consuming, but the results of this brute-force technique allowed to determine a meaningful multi-dipole model structure. Finally, the parameterization of the multi-dipole model was obtained in the form of 52 dipoles (26 induced and 26 permanent). Furthermore, the reference number of the optimisation algorithm iterations was set as equal to 124 steps.

The stopping conditions of the optimisation procedure consist of four parameters that can be divided into two groups. The first group has to do with the accuracy or step of the procedure. *StepTolerance* i, meaning the norm of $x_i - x_{i+1}$, is the lower bound on the step size. If the solver attempts to take a step that is smaller than *StepTolerance*, the iterations end. *StepTolerance* is generally used as a relative bound, meaning iterations end when $|(x_i - x_{i+1})| < StepTolerance*1 + |(x_i)|$, or a similar relative measure. The second group of stopping conditions is the number of iterations of the algorithm's calculations. We can limit the number of evaluations of the objective function or the number of steps in the optimisation procedure. Ultimately, through experiments with the optimisation procedure, it was decided to set very small tolerance values - so as to virtually exclude this kind of stoppage of the procedure. On the other hand, the parameter number of evaluations of the objective function was set very large. With these settings, in each case presented in the article the procedure stopped due to the same number of iterations, which resulted in the same process conditions and the possibility of comparisons. An important issue in the learning process is selecting the appropriate number of iterations. When the model whose parameters

are determined in the learning process has a large number of parameters that cannot be 'satisfied' with the information from the input data, the selection of parameter values has a random nature. Increasing the number of iterations of the optimisation process may lead to a deterioration of the performance criterion on the testing dataset, despite continuous correcting on the trained dataset. In order to prevent the phenomena of overfitting or overtraining, a series of exhaustive simulations were carried out. The average fit rates on the FIT dataset and the CV validation dataset were taken into account. All 12 scenarios of 11-path combinations were considered. The results of such analysis are shown in Fig. 17. It is very clearly visible that up to 124 iterations there is total improvement in fit (reduction of the fit error), while starting from iteration 125 there is overfitting in the form of simultaneous reduction of FIT and increase of CV.

Thus, the final stop condition of the optimisation process was set at 124 iterations. The set of all parameters associated with the termination of the optimisation process is presented in Table 6.

The constraints on the decision variables were determined on the basis of the physical dimensions of *Zodiak*. The dipoles are to describe the magnetic field around the ship, so their allowed position is determined by a cuboid with sides of 110% of the appropriate dimension among length, breadth and draft. Apart from draft, the vessel also has the height of 15 m. A set of appropriate constraints for optimisation problem is given in Table 7.

3.6. Objective function form

The task of the optimisation procedure is to minimize the distance between the reference and model data on all data paths available from the measurement. The basic optimisation task for a chosen model structure with m permanent dipoles and n induced dipoles is defined as follows:

$$\min_{\Omega \in \{\Omega_1, \cdots, \Omega_{n+m}\}} J_G = \sum_l \sum_d \sum_{j=1}^{501} \left(B_{l,d}^{ref}(j) - B_{l,d}^{model}(j, \Omega, \phi(d)) \right)^2, \tag{13}$$

subject to:

$$\underset{i \in (1, m+n)}{\forall} \Omega_i^{min} \le \Omega_i \le \Omega_i^{max}, \tag{14}$$

where:

Table 8

Starting point for optimisation process for permanent and induced dipoles as $[m_x, m_y, m_z, x, y, z]$

opt.M0P = [30000, −5000, −60000, 0, 0, 0]; % permanent dipoles
opt.M0I = [90000, 25000, −4000, 0, 0, 0]; % induced dipoles

$$\mathop{\forall}_{i \in (1, m+n)} \Omega_i \in \{m_{x,i}, m_{y,i}, m_{z,i}, x_i, y_i, z_i\}, \tag{15}$$

$$l \in \{x, y, z\}, \tag{16}$$

$$d \in \{N10, E10, S10, W10, N7, E7, S7, W7, N5, E5, S5, W5\}, \tag{17}$$

and J_G is the objective function, l denotes a set of x,y, and z magnetic fields components, d denotes an analysed dataset - in that case, all available data for all 12 paths, j denotes the number of analysed samples of data of each kind, $\varphi(d)$ is the ship heading related to the j-th sample of a analysed data set d, Ω is the vector of all decision variables defined for

all considered dipoles, Ω_i^{min}, Ω_i^{max} are the vectors of minimal and maximal constraint values for the decision variables subset Ω_i defining i-th dipole. For the clarity of the presented notation of the optimisation task (13)-(17), it should be mentioned that to determine the value of the objective function J_G (13) correctly, the ship's course φ related to the analysed data set d should be properly considered in the multi-dipole model B^{model} (see Fig. 12, Eqs. (5)–(8), and Eqs. (1)–(3)). The ship's course plays the role of the depending parameter, not a decision variable, which value depends on analysed data set d.

The objective function J_G (13) defines the difference in matching of the reference B^{ref} and model B^{model} data in all considered directions d, for all paths (17), over the length of 200 m, with the resolution of 1 m for magnetic field components Bx, By, Bz (16) − 501 samples of data of each kind. Depending on the simulation scenario, the objective function takes a different form – the dataset d is modified according to the analysed scenarios described in the previous section. Inside the criterion function, there is the sum of squares of model B^{model} and measured B^{ref} data

Table 9

Multi-dipole model parameters.

Parameter	m_x [Am^2]	m_y [Am^2]	m_z [Am^2]	x [m]	y [m]	z [m]
Constraints	$<-5 \cdot 10^6 \div 5 \cdot 10^6>$	$<-5 \cdot 10^6 \div 5 \cdot 10^6>$	$<-5 \cdot 10^6 \div 5 \cdot 10^6>$	$<-33.7 \div 33.7>$	$<-5.9 \div 5.9>$	$<-3.5 \div 15>$
Permanent dipoles	227211,1	267899,3	1094000,0	5,4	−5,7	14,0
	−170009,4	32186,3	139761,0	0,6	−2,1	14,8
	353167,5	255492,3	28583,3	8,7	−5,6	11,6
	−198853,3	834747,5	332873,7	6,0	−5,9	8,8
	298289,0	422219,8	−170639,5	21,4	3,4	11,4
	−180479,9	366620,2	200479,1	3,4	−4,8	11,8
	−193991,4	22576,2	−520171,8	0,5	−3,8	6,4
	−210550,6	−14936,6	−30445,5	−1,6	3,1	13,7
	−215413,1	−727829,3	−47927,7	0,8	5,9	15,0
	−86821,3	131940,9	−233069,4	4,0	0,2	9,1
	63663,6	241346,9	377417,2	6,0	−4,0	12,3
	56237,8	164876,9	−83562,9	7,5	0,4	10,9
	358713,0	159592,7	629381,6	9,5	−4,8	12,5
	−48941,3	19309,2	26989,6	−0,7	0,0	13,9
	81251,4	−625996,9	−123242,0	7,5	−3,0	0,8
	−95318,9	−119569,4	30610,6	1,9	1,1	13,9
	−102355,6	−86034,7	39052,3	1,0	0,9	13,6
	584337,1	−385295,5	96840,7	−15,5	−5,2	10,3
	−99235,8	197640,2	−218583,8	4,1	1,5	9,2
	−101905,9	−52496,5	−12436,1	2,0	0,8	13,6
	−204882,9	−785413,7	−208219,1	22,4	−4,6	7,0
	−19396,3	117966,3	141845,8	3,9	−2,7	13,4
	6303,7	116687,0	56849,4	3,5	−2,6	12,5
	−175271,7	113869,3	−194212,4	3,1	−2,3	9,9
	25309,2	126049,7	210052,9	5,9	−2,9	13,0
	150764,9	−707235,6	−408487,0	3,7	4,5	11,3
Induced dipoles	−30913,6	−128328,8	245815,7	−10,3	0,5	9,0
	60961,3	−304561,3	−195596,4	8,8	3,9	5,7
	211908,1	−54126,6	185178,5	−8,8	2,3	9,7
	372885,0	547157,1	−346222,3	9,2	5,8	11,8
	274503,9	−470654,2	−294328,0	−19,8	−5,4	13,0
	351871,0	−169001,9	−521734,4	−2,2	−4,7	9,2
	360078,3	−386855,1	−21350,5	−19,7	5,9	8,3
	26788,5	77399,4	6002,0	−18,6	−0,6	12,2
	2142,9	241214,9	13846,5	−20,7	−0,7	9,6
	339609,7	−225805,3	−230643,4	−3,2	5,8	6,7
	131092,5	−440189,2	−424342,6	−20,5	−5,1	13,3
	−155468,2	63649,1	382619,9	−2,6	1,7	3,1
	−606675,6	−1028398,5	−276939,1	12,9	1,8	15,0
	−273251,1	−381606,0	−340922,1	−18,7	5,6	12,0
	7926,5	322609,2	34136,4	−16,7	1,4	14,5
	20199,3	590834,8	205284,2	−24,4	2,8	15,0
	−162135,1	187561,1	−66732,4	−20,7	1,1	9,7
	−396864,6	−382634,4	223554,4	−21,9	−3,3	14,9
	−69894,9	−52123,4	95817,6	−11,8	−1,1	11,8
	152985,9	406227,1	−26827,8	−17,6	2,2	14,8
	26611,6	276963,9	−10050,7	−17,3	1,7	11,6
	289736,3	942025,8	−6642,2	12,5	−3,8	15,0
	51866,9	−47945,8	115849,3	−12,4	−3,1	10,4
	−93075,2	143709,2	34823,1	−18,1	1,6	13,3
	−512112,5	36353,4	−47736,1	−2,6	4,7	14,9
	−324597,6	328960,9	158845,9	−8,2	0,7	15,0

Permanent and induced dipole locations with scaled magnetic moments
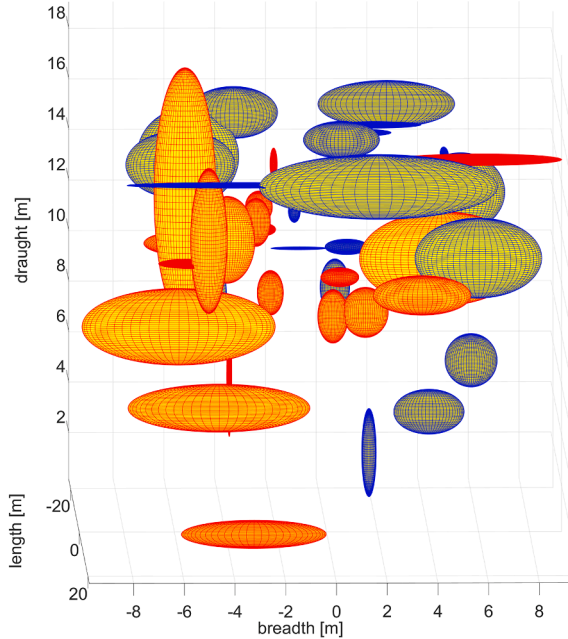

Fig. 18. Dipole locations and values (bow view).

Permanent and induced dipole locations with scaled magnetic moments
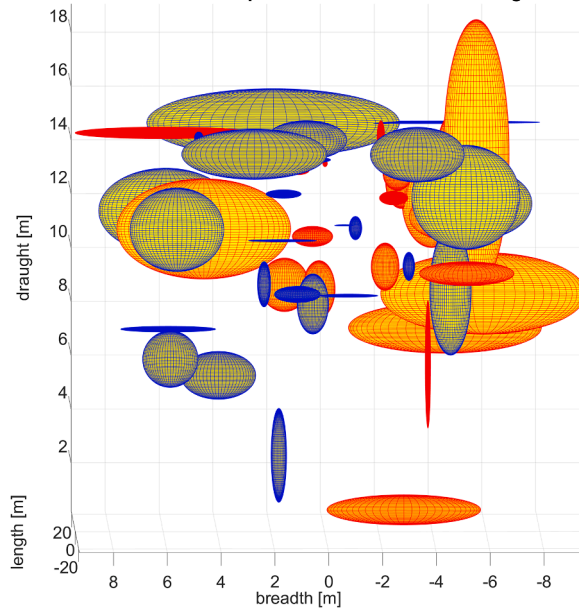

Fig. 19. Dipole locations and values (stern view).

differences for individual magnetic field components.

The optimisation with conservative box constraints (Table 7) for all dipole locations and magnetic moments was proposed to solve the problem defined in the paper. Limiting the space of dipole locations to the parallelepiped volume seems natural, and the limitation of the magnetic moments has been treated liberally, i.e. in the range of +/- $0.5 \cdot 10^6$ Am^2, based on the observation of many typical naval ferromagnetic objects modelled with the use of the multi-dipole model. The unconstrained version of optimisation for this problem leads to longer computation time and much larger errors with the same stop conditions for the optimisation procedure.

The starting point for the optimisation process, determined arbitrarily and replicated to all dipoles, is given in Table 8.

3.7. Results and assessment

The direct result of the optimisation process is 312 parameters of the multi-dipole model. The optimisation procedure also gives the FIT result in the form of residues. In the approach used, this result is related to the amount of data used in the optimisation process. Thanks to this, the individual variants of calculations may be compared.

The list of model parameters is presented in Table 9. Figs. 18–20 show symbolic dipoles in the form of ellipsoids visualizing the position and scaled moment value of the dipole. While in Figs. 21 and 22 presents the vector representation of each dipole. Observing those representation for individual dipoles does not reveal a direct physical interpretation. However, their superposition reproduces the magnetic signature with high accuracy.

Using these dipoles, the MFD values of *Zodiak* signature at any course and measurement depth (deploying depth of magnetometer) may be reproduced. To visualize the power of the predictive capabilities of the multi-dipole model, Figs. 23–25 show reproduced three main components Bx, By, Bz of MFD fields. Notice the marked lines that correspond to the input data used for the teaching phase of the model.

The results of individual simulation scenarios can be compared qualitatively in the form of graph observations, and quantitatively in the form of RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) fit indicators on paths, as well as by comparing the maximum modelling errors (Table 10). The root mean square error is given as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(ref_i - model_i)^2}, \tag{18}$$

and the mean absolute error is given as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|ref_i - model_i|, \tag{19}$$

where: $model_i$ is the vector of N signature values at i-th position coordinate counted by the model, and ref_i is the vector of N reference signature values at the same position. The RMSE definition is derived from [63], and comprehensive discussion can be found in [64].

The first column in Table 10 is the number of the calculation scenario. The second column contains the description of the category or the description of the calculation scenario in the FIT - CV scheme. The third column, labelled FIT_RMSE, contains the RMSE of the alignment results on the paths given in the second column by FIT. The fourth column, labelled FIT_MAE, contains the MAE of the alignment results on the paths given in the second column by FIT. The fifth column, labelled CV_RMSE, contains the RMSE of the path alignment results given in the second column by CV. The sixth column, labelled CV_MAE, contains the MAE of the path match results given in the second column by CV. Columns 7–18 of Table 10 contain the match result on each path (with data with much higher resolution – 5001 samples). Depending on whether the data for learning included a given path or not, these data should be treated as FIT or CV. For each dataset with 501 samples and with 5001 samples, mean values were also determined and are bolded in the corresponding rows of Table 10. The values in the table have the matching error dimension: the smaller the values, the better the matching process was.

For further analysis, it is worth plotting the main results from Table 10, and they are shown in Figs. 26 and 27.

Fig. 26 summarizes the MAE and RMSE error values for the learning (FIT) and cross-validation (CV) phases for the different scenarios analysed (number of learning paths - Section 4.2, set d in the formulated optimisation task, Eqs. (13)–(17)), with an equal number of data per each path (501 samples). While in Fig. 27 are shown the FIT and CV errors along each of the analysed paths (with the number of data per each path equal 5001) for each of the analysed scenarios.

Permanent and **induced** dipole locations with scaled magnetic moments
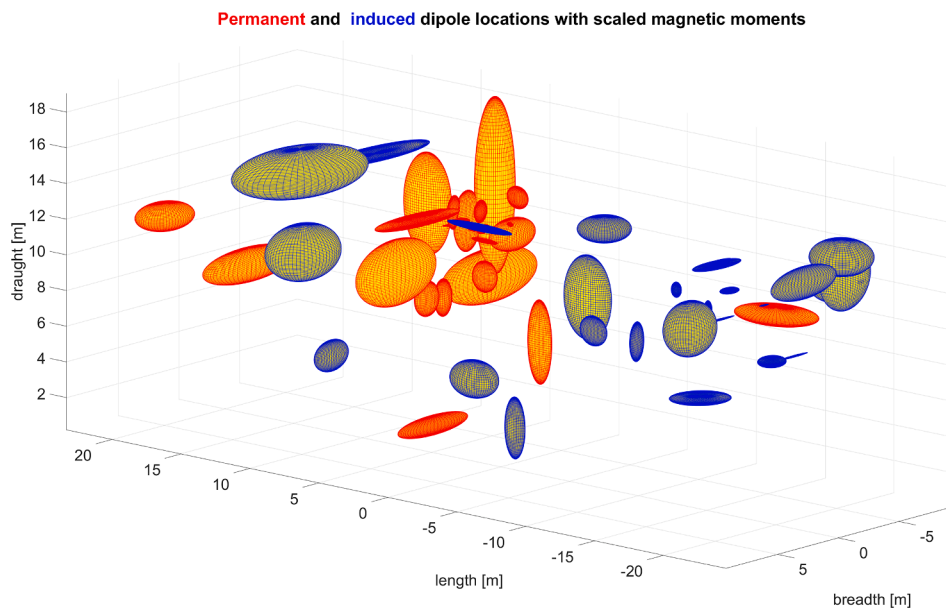
Fig. 20. Dipole locations and values (port/stern view).

Permanent and **induced** dipole locations with scaled magnetic moments
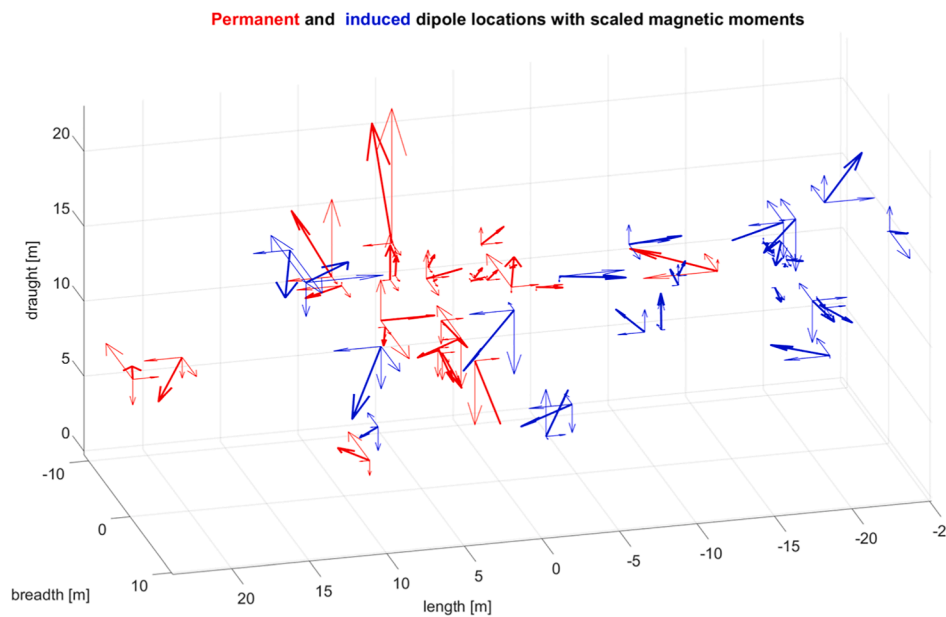
Fig. 21. Permanent and induced dipole location with scaled magnetic moments (port/stern view).

3.8. Discussion of results

When analyzing the results presented in Table 10, Figs. 26, and 27, the following can be noticed:

- the greatest differences in the results occur in the least numerous datasets, in terms of the amount of data, 3-path simulation scenarios (Fig. 22);
- the CV-RMSE indicator, due to the effect of exponentiation, strengthens large errors and thus more explicitly presents the differences between simulation scenarios (Fig. 26);
- the best (smallest) FIT error was for the FIT-MAE indicator, in the scenarios with the least data and increased with the increase in the amount of data (Fig. 26). It is logical because the optimisation procedure performed the calculations so as to select the parameters of the multi-dipole model describing all the data used for the

calculations with the smallest mean square error. So, the more data, the more difficult it is to do and the error value grows;
- the best CV verification scenario was for the largest number of data and deteriorated with the decreasing number of data, because in this situation the data from all FIT paths were used to determine the dipoles, which were then verified on only one CV path (Fig. 27);
- the data on individual paths highly varied, which may indicate their different quality and suitability for the the multi-dipole model structure and parameters creation process. In particular, high values for the most interesting 11-path scenarios (orange bars in Fig. 27) are characteristic of paths S5, N5, E10, and S10. Statistically, the group of data recorded at speed of 5 kn is the worst.

The basic quantitative indicator describing the quality of the developed model is CV. It is the average error on all magnetic field components Bx, By and Bz. For qualitative assessment, CV charts of model fit to

17

Permanent and **induced** dipole locations with scaled magnetic moments
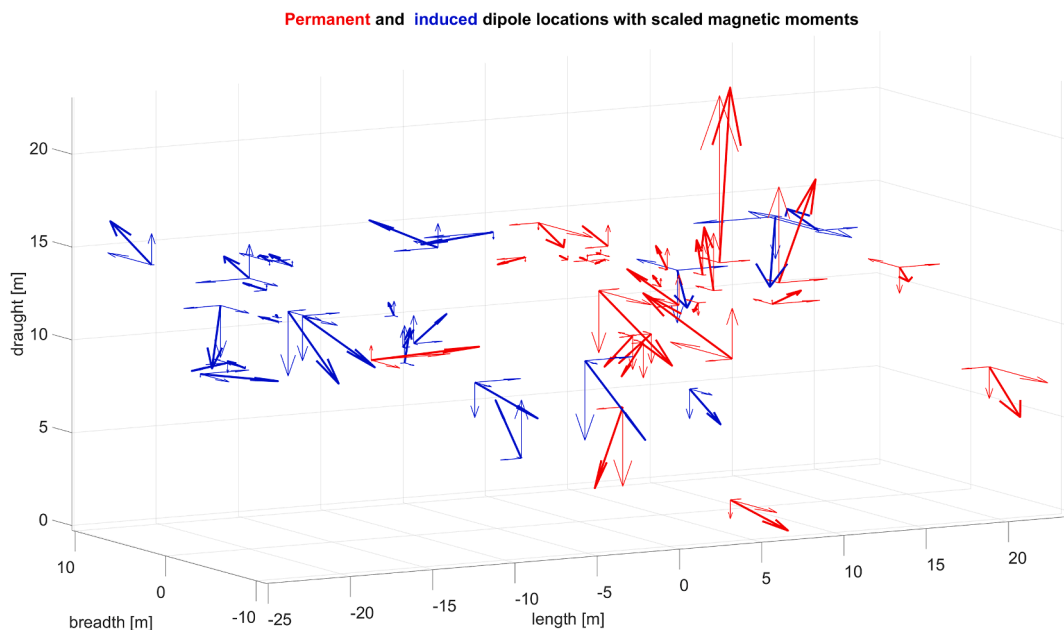
Fig. 22. Permanent and induced dipole location with scaled magnetic moments (stern/port view).

Bx component of MFD generated from multi-dipole model [nT]
FIT: E10, S10, E10, N7, E7, S7, W7, N5, E5, S5, W5 CV: N10
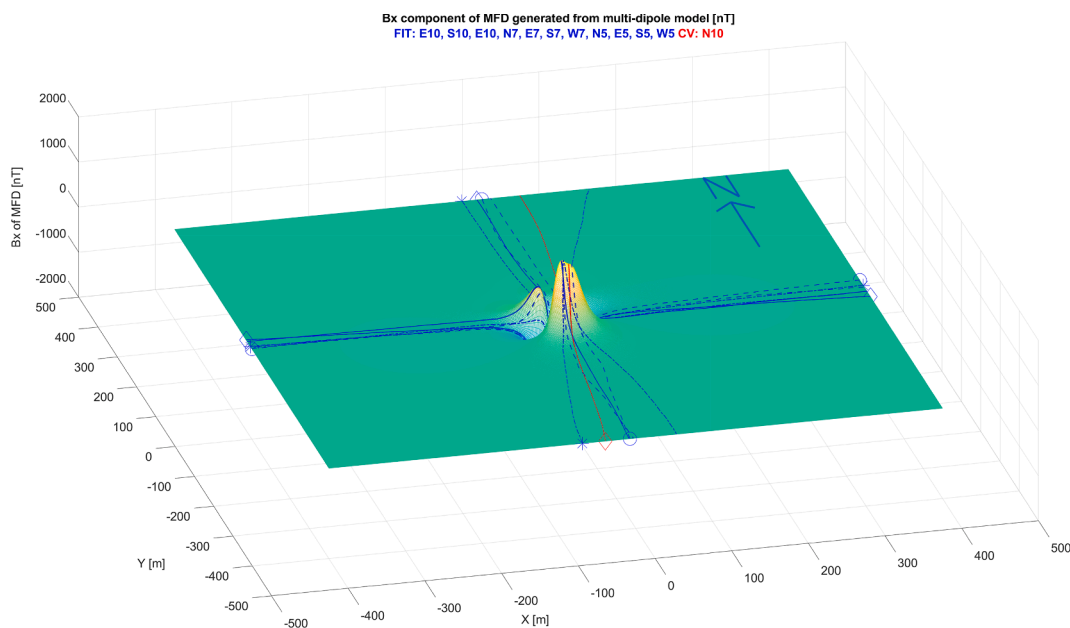
Fig. 23. Bx component of MFD generated from multi-dipole model and measurement results of all paths.

the values measured on all paths can be used. Individual plots for each path and component as well as an integrated plot containing data from all paths against the components Bx, By, Bz and BT are presented in Figs. 28–31, while the cumulative results of Bx, By, Bz and BT from all paths are presented in Fig. 32. The assessment in terms of BT is important because, having scalar magnetometers, the information about the values of individual components are not available. The information about the maximum error we are dealing with may also be interesting. Therefore, a drawing was prepared that contains graphs of relative values of the BT error from all paths (Fig. 33) and the average of the maximum values is calculated as another quantitative parameter describing the quality of the model. Only the combined analysis of CV with the matching graphs of BT component and the BT relative error graph allows us to form an opinion on the quality of the multi-dipole model. Graphs of fields generated on the basis of the model with

marked FIT and CV paths can also be assessed. This allows to assessing whether the multi-dipole model fitted in places of "constraints" do not produce abstract and unrealistic values outside the regions from which it received the information.

The evaluation of the obtained results must take into account several aspects. First, the dipoles were obtained to enable predicting MFD components in the form of fields, i.e. in any direction in the area under consideration. The fields of Bx, By, Bz components for direction 0, presented in Figs. 23–25, were calculated for the scenario in which N10 was the direction of validation. However, the authors also reviewed all remaining fields coming from the calculated dipoles in all cardinal directions. In all scenarios, the disturbance of the field occurred only in the center, which means that in the multi-dipole model there are no anomalies in the form of disturbances outside the place where it should occur.
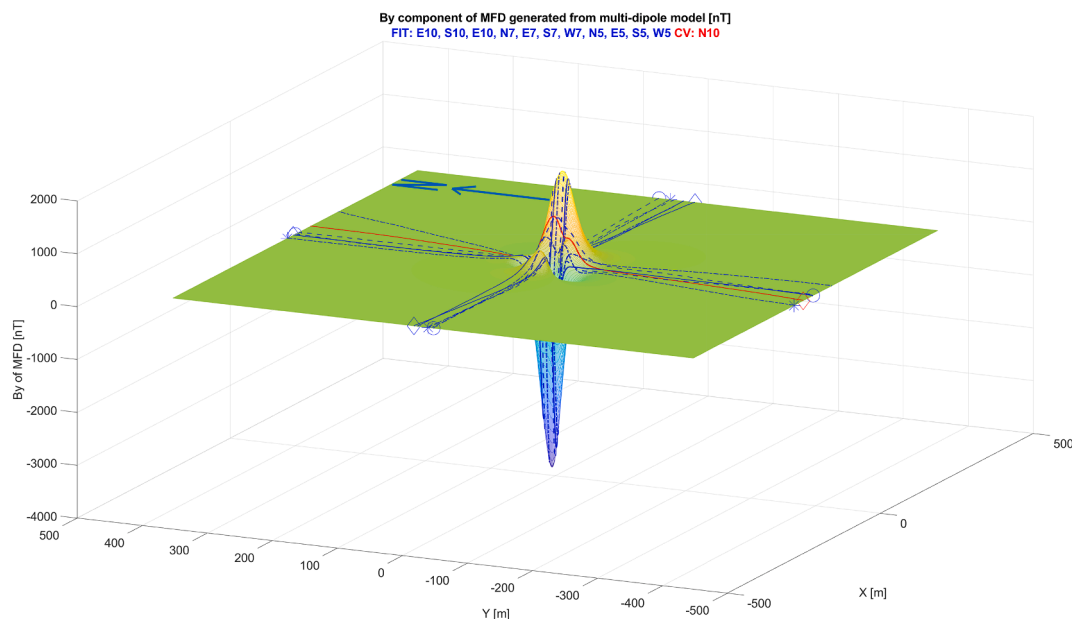
18

By component of MFD generated from multi-dipole model [nT]
FIT: E10, S10, E10, N7, E7, S7, W7, N5, E5, S5, W5 CV: N10

Fig. 24. By component of MFD generated from multi-dipole model and measurement results of all paths.

Bz component of MFD generated from multi-dipole model [nT]
FIT: E10, S10, E10, N7, E7, S7, W7, N5, E5, S5, W5 CV: N10
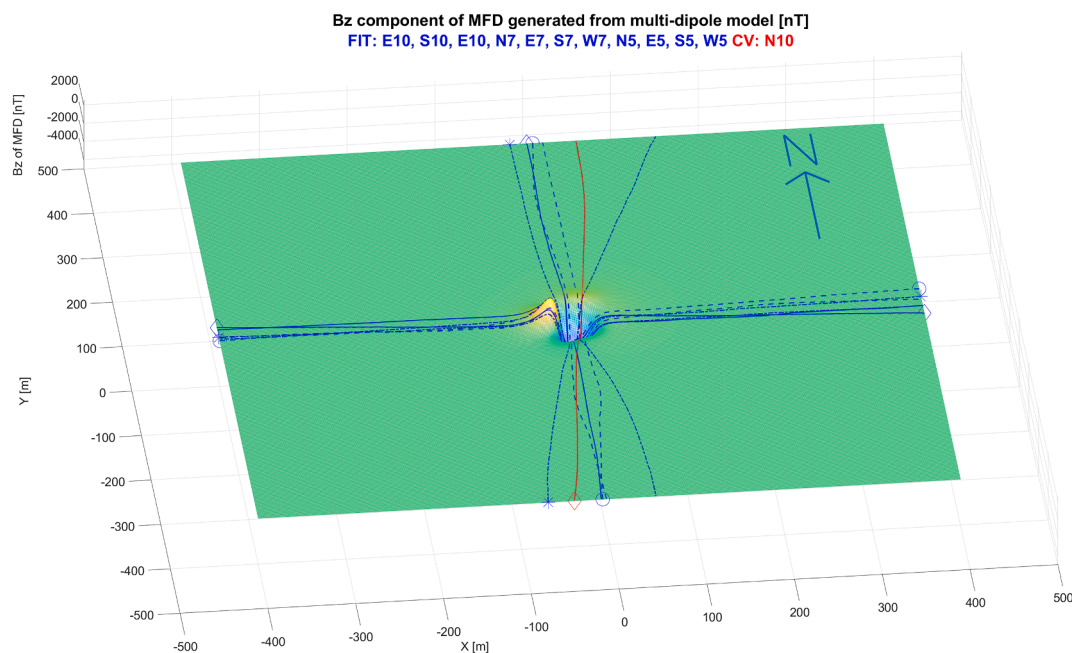
Fig. 25. Bz component of MFD generated from multi-dipole model and measurement results of all paths.

When the nature of the field prediction was positively assessed, the analysis of the fit on the paths was started. For Bx component, there are clear errors in the form of deflection in the opposite direction for the path E10 and especially for the path E5 (Fig. 28). For By component, the most visible errors occurred for paths N5 and S5 (Fig. 29). These errors existed in Bx component for E – W direction and By for N – S direction. They might be caused by inaccuracy of geographical position determined by the GPS receiver. It should be also noted that these errors occur for the paths that make the smallest quantitative contribution to the total BT area. The Bz components fit very well (Fig. 30), which in combination with the minor contribution from Bx and By results in excellent BT mapping (Fig. 31).

The Authors undertook additional simulation studies using regularization, i.e. modifying the manipulation of input data and the optimisation criterion in such a way as to obtain an improvement in the quality of the model. This will be described in detail in the next section.

4. Regularization

An important problem in the area of development of models based especially on the measured data is to protect them against overfitting/underfitting phenomena [65]. These phenomena are related to the number of available data, their diversity and representativeness, the model structure, and the applied method of learning and verification [66]. The main goal of learning is to learn as much as possible from the fit dataset, but at the same time the generalization property of the model needs to be preserved to enable its use for prediction purposes and to work with unseen data [67]. This can be reached by compromising

Table 10
MAE and RMSE values of fit and CV for different numbers of dipoles and various runs.

No.	ZODIAK DIPOLES	Optimisation report based on residuals Size of data paths 501 samples				Fitting validation on referenced data 5001 samples per path												mean
		FIT-RMSE [nT]	FIT-MAE [nT]	CV-RMSE [nT]	CV-MAE [nT]	N10 [nT]	E10 [nT]	S10 [nT]	W10 [nT]	N7 [nT]	E7 [nT]	S7 [nT]	W7 [nT]	N5 [nT]	E5 [nT]	S5 [nT]	W5 [nT]	[nT]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(!3)	(14)	(15)	(16)	(17)	(18)	(20)
	for 12 paths																	
1	**FIT** N10-E10-S10-W10-N7-E7-S7-W7-N5-E5-S5-W5	62,3	26,9			58,36	70,67	60,98	50,03	42,39	40,79	52,99	42,88	83,93	59,33	100,38	57,35	**60,01**
	for 11 paths																	
2	**FIT** E10-S10-W10-N7-E7-S7-W7-N5-E5-S5-W5 **CV** N10	63,2	26,6	76,7	34,1	76,73	70,32	62,65	45,97	57,86	43,83	54,07	38,69	81,75	56,48	102,14	54,76	–
3	**FIT** N10-S10-W10-N7-E7-S7-W7-N5-E5-S5-W5 **CV** E10	62,6	26,7	97	37,8	52,6	97,07	63,96	46,48	45,97	38,76	55,19	38,93	87,96	82,32	92,5	53,46	–
4	**FIT** N10-E10-W10-N7-E7-S7-W7-N5-E5-S5-W5 **CV** S10	62,9	26,3	77,3	35,1	57,41	72,75	77,37	48,63	46,64	40,57	64,75	40,58	73,89	60,5	108,7	45,26	–
5	**FIT** N10-E10-S10-N7-E7-S7-W7-N5-E5-S5-W5 **CV** W10	64,4	27,2	78	31,7	55,73	73,96	67,41	78,1	34,2	44,72	54,94	48,53	85,48	59,28	101,67	53,86	–
6	**FIT** N10-E10-S10-W10-E7-S7-W7-N5-E5-S5-W5 **CV** N7	66,2	27	57,2	26,3	61,82	76,42	66,91	40,35	57,29	42,93	58,46	43,74	88,11	61,96	101,41	59,05	–
7	**FIT** N10-E10-S10-W10-N7-S7-W7-N5-E5-S5-W5 **CV** E7	64,8	27,6	56,3	24,3	56,62	68,5	65,24	48,24	44,45	56,35	53,6	39,42	85,91	58,05	106,8	57,31	–
8	**FIT** N10-E10-S10-W10-N7-E7-W7-N5-E5-S5-W5 **CV** S7	75,1	31,3	106,5	43,9	57,13	86,18	107,86	48,93	65,83	53,52	106,64	48,34	75,16	77,42	118,86	46,59	–
9	**FIT** N10-E10-S10-W10-N7-E7-S7-N5-E5-S5-W5 **CV** W7	61,7	27,4	69	31,1	57,66	71,98	58,51	48,84	46,24	41,61	46,96	69,04	77,08	58,48	95,77	55,41	–
10	**FIT** N10-E10-S10-W10-N7-E7-S7-W7-E5-S5-W5 **CV** N5	61,8	24,8	112,4	46,1	53,36	77,23	56,51	44,38	52,01	49,03	43,25	39,65	112,52	66,3	102,29	68,48	–
11	**FIT** N10-E10-S10-W10-N7-E7-S7-W7-N5-S5-W5 **CV** E5	59,8	26,2	93,6	34,8	50,65	90,69	56,1	43,7	38,64	37,58	43,74	40,32	79,06	93,64	90,21	52,63	–
12	**FIT** N10-E10-S10-W10-N7-E7-S7-W7-N5-E5-W5 **CV** S5	75	33,7	130,5	48	59,3	79,57	107,91	53,19	67,39	64,94	99,94	49,77	83,93	72,46	130,61	64,07	–
13	**FIT** N10-E10-S10-W10-N7-E7-S7-W7-N5-E5-S5 **CV** W5	80,4	36,9	105,6	48,7	66,94	86,72	82,98	60,61	76,77	62,51	77,07	68,39	100,01	78,78	110,08	105,66	–
	mean	**66,49**	**28,48**	**88,34**	**36,83**	**58,83**	**79,28**	**72,78**	**50,62**	**52,77**	**48,03**	**63,22**	**47,12**	**85,91**	**68,81**	**105,09**	**59,71**	**66,01**
	standard deviation	**6,30**	**3,44**	**22,03**	**7,87**	**6,78**	**8,58**	**17,41**	**9,61**	**12,07**	**8,86**	**20,01**	**10,40**	**10,50**	**11,45**	**10,76**	**15,17**	**11,80**
	for 8 paths																	
14	**FIT** N10-E10-S10-W10-N7-E7-S7-W7	43	19,6	–	–	34,18	62,32	44,52	34,78	39,38	51,43	30,94	37,95	140,87	73,34	123,81	165,92	–
15	**FIT** N10-E10-S10-W10-N5-E5-S5-W5	64,9	27,3	–	–	66,51	62,26	62,08	42,08	63,31	47,87	72,59	70,35	75,93	46,87	95,63	52,41	–
16	**FIT** N7-E7-S7-W7-N5-E5-S5-W5	66,4	28,7	–	–	69,78	92,49	82,39	72,74	51,46	46,26	74,91	35,5	80,79	72,88	100,53	41,92	–
	mean	**58,10**	**25,20**	**–**	**–**	**56,82**	**72,36**	**63,00**	**49,87**	**51,38**	**48,52**	**59,48**	**47,93**	**99,20**	**64,36**	**106,66**	**86,75**	**67,19**
	standard deviation	**10,69**	**4,00**	**–**	**–**	**16,07**	**14,24**	**15,47**	**16,45**	**9,77**	**2,16**	**20,20**	**15,88**	**29,53**	**12,37**	**12,29**	**56,15**	**18,38**
	for 7 paths																	
17	**FIT** N5-E5-S5-W5-N7-E7-S7 **CV** W7	62,5	27,5	131,8	52,7	64,49	81,55	80,56	147,59	42,24	44,05	71,71	131,92	73,01	58,67	90,89	38,28	–
18	**FIT** N5-E5-S5-W5-N7-E7-W7 **CV** S7	65,5	26,4	115,7	46,1	64,23	92,5	113,18	65,94	55,85	42,87	115,76	31,23	66,08	75,06	114,89	31,42	–
19	**FIT** N5-E5-S5-W5-N7-S7-W7 **CV** E7	60,6	27,6	52,6	25	76,94	87,23	64,52	78,83	46,51	52,62	52,85	41,09	72,56	63,62	92,44	34,86	–
20	**FIT** N5-E5-S5-N7-E7-S7-W7 **CV** N7	55,6	25,8	96,8	45	122,06	83,5	61,9	69,75	96,89	38,03	56,15	29,64	69,41	55,29	82,74	38,56	–
21	**FIT** N5-E5-S5-N7-E7-S7-W7 **CV** W5	66,1	28,8	91,4	37,2	67,9	65,57	64,94	68,2	51,85	41,35	55,63	41,42	85,87	68,33	96,73	91,52	–
22	**FIT** N5-E5-W5-N7-E7-S7-W7 **CV** S5	47,9	22,8	137,2	53,6	68,95	67,62	91,72	80,57	40,21	47,13	58,34	33,69	63,38	51,2	137,36	32,32	–
23	**FIT** N5-S5-W5-N7-E7-S7-W7 **CV** E5	56,4	25,2	152,4	47,2	55,04	143,46	73,99	68,37	47,29	37,12	63,45	26,49	77,33	152,59	82,24	35,16	–
24	**FIT** E5-S5-W5-N7-E7-S7-W7 **CV** N5	57,4	25,4	103	44	62,59	91,24	60,5	73,07	44,62	46,68	49,03	36,35	103,04	68,76	92,6	43,63	–
25	**FIT** N10-E10-S10-W10-N7-E7-S7 **CV** W7	39,2	19,2	58,6	25	34,46	60,47	35,74	23,57	33,59	44,68	31,06	58,61	144,64	79,4	116,25	187,56	–
26	**FIT** N10-E10-S10-W10-N7-E7-W7 **CV** S7	37,4	18,6	112	37,4	34,55	52,27	22,97	35,89	36,78	39,78	112,07	33,43	159,19	65,35	126,68	210,35	–
27	**FIT** N10-E10-S10-W10-N7-S7-W7 **CV** E7	33,8	16,7	111,1	39,8	33,11	31,66	43,56	34,54	28,99	111,24	29,74	33,29	157,74	79,34	138,14	190,39	–
28	**FIT** N10-E10-S10-W10-E7-S7-W7 **CV** N7	39,8	18,4	49,8	20,9	31,85	59,45	39,32	30,3	49,87	41,15	31,49	37,47	145,36	69,72	121,54	175,19	–
29	**FIT** N10-E10-S10-N7-E7-S7-W7 **CV** W10	40,3	18,6	56,4	25,6	35,41	59,81	42,71	56,45	28,43	47,58	27,39	29,85	139,52	69,11	121,3	165,8	–
30	**FIT** N10-E10-W10-N7-E7-S7-W7 **CV** S10	39	18,6	61,3	26	34,64	55,15	61,37	38,05	36,13	47,15	18,34	33,16	145,81	66,79	141,54	169,91	–
31	**FIT** N10-S10-W10-N7-E7-S7-W7 **CV** E10	32,5	16,5	115,7	45,3	24,65	115,84	36,36	33,66	39	27,9	27,76	35,77	125,5	131,2	94,85	150,75	–
32	**FIT** E10-S10-W10-N7-E7-S7-W7 **CV** N10	40,7	18,9	51,9	22,9	51,91	55,29	41,71	34,68	35,64	45,75	30,81	35,78	130,39	66,93	125,09	172,9	–
33	**FIT** N10-E10-S10-W10-N5-E5-S5 **CV** W5	59,8	26,7	201	69,7	61,64	42,61	46,5	41,66	60,48	78,83	89,73	76,72	77,89	37,62	89,96	201,19	–
34	**FIT** N10-E10-S10-W10-N5-E5-W5 **CV** S5	55,3	23,5	139,2	53,8	62,16	42,08	83,72	32,39	59,85	69,4	73,91	59,25	66,07	35,34	139,37	46,39	–
35	**FIT** N10-E10-S10-W10-N5-S5-W5 **CV** E5	66,6	29,1	80,9	30,3	55,64	81,77	63,45	34,52	56,78	50,57	62,72	72,37	78,41	80,93	89,17	44,74	–
36	**FIT** N10-E10-S10-W10-E5-S5-W5 **CV** N5	60	25,5	138	55,9	61,66	67,04	42,36	41	66,63	80,26	75,02	68,42	138,16	50,78	90,23	52,17	–
37	**FIT** N10-E10-S10-N5-E5-S5-W5 **CV** W10	61,8	27,6	239,2	81,3	38,47	55,86	69,28	239,4	75,02	67,92	63,71	245,51	78,4	39,61	83,92	51,35	–
38	**FIT** N10-E10-W10-N5-E5-S5-W5 **CV** S10	72,8	28,7	132,1	44,9	59,46	66,92	132,18	42,3	72,97	79,83	137,91	68,44	61,05	60,05	136,22	38,28	–
39	**FIT** N10-S10-W10-N5-E5-S5-W5 **CV** E10	66,4	28,2	98,3	38,7	61,24	98,4	73,9	34,15	65,97	45,29	76,97	68,29	79,73	73,52	85,74	37,47	–
40	**FIT** E10-S10-W10-N5-E5-S5-W5 **CV** N10	60,3	26,6	134,9	55,9	135,07	57,09	64,24	26,04	113,19	52,65	80,32	70,04	76,6	39,29	89,16	45,38	–

Table 10 (*continued*)

| No. | ZODIAK DIPOLES | Optimisation report based on residuals Size of data paths 501 samples | | | | Fitting validation on referenced data 5001 samples per path | | | | | | | | | | | | mean |
| | | FIT-RMSE [nT] | FIT-MAE [nT] | CV-RMSE [nT] | CV-MAE [nT] | N10 [nT] | E10 [nT] | S10 [nT] | W10 [nT] | N7 [nT] | E7 [nT] | S7 [nT] | W7 [nT] | N5 [nT] | E5 [nT] | S5 [nT] | W5 [nT] | [nT] |
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(!3)	(14)	(15)	(16)	(17)	(18)	(20)
	mean	**53,24**	**23,79**	**110,89**	**42,68**	**58,26**	**72,27**	**62,95**	**59,62**	**53,53**	**53,33**	**62,16**	**58,26**	**100,63**	**68,02**	**107,46**	**95,23**	**70,98**
	standard deviation	**11,96**	**4,25**	**45,88**	**14,81**	**25,65**	**24,65**	**24,77**	**45,91**	**20,35**	**18,34**	**29,87**	**45,56**	**34,31**	**26,02**	**21,00**	**67,80**	**32,02**
	for 4 paths																	
41	**FIT** N10-E10-S10-W10	22,7	12,6	–	–	25,58	25,17	16,87	22,25	52,57	113,98	92,17	58,53	170,38	80,81	135,92	244,85	–
42	**FIT** N7-E7-S7-W7	49,2	24	–	–	49,48	114,25	76,38	67,63	55,84	46,95	53,85	38,23	123,93	107,74	112	97,04	–
43	**FIT** N5-E5-S5-W5	57,5	25,7	–	–	93,46	81,88	105,16	124,03	71,78	50,07	120,17	133,99	56,5	46,16	84,55	27,66	–
	mean	43,13	20,77	–	–	56,17	73,77	66,14	71,30	60,06	70,33	88,73	76,92	116,94	78,24	110,82	123,18	82,72
	standard deviation	14,84	5,82	–	–	28,11	36,82	36,76	41,63	8,39	30,89	27,18	41,20	46,75	25,21	20,99	90,57	36,21
	for 3 paths																	
44	**FIT** E5-S5-W5 **CV** N5	40,1	20,2	177,5	70,2	118,31	99,91	72,4	192,76	114,67	115,99	101,71	204,12	177,67	46,26	43,4	28,34	–
45	**FIT** N5-S5-W5 **CV** E5	54,8	23,2	196,3	58,2	72,5	184,26	104,28	42,67	96,09	75,8	122,58	75,91	66,55	196,44	67,14	9,84	–
46	**FIT** N5-E5-W5 **CV** S5	19,3	10,2	228,8	72,8	57,31	51,72	237,41	58,84	68,53	150,34	256,73	82,43	20,4	19,49	229,02	17,84	–
47	**FIT** N5-E5-S5 **CV** W5	61,4	25,2	397,1	144,8	92,72	49,31	99,34	399,18	161,14	175,38	113,42	416,64	76,39	27,16	69,03	397,42	–
48	**FIT** E7-S7-W7 **CV** N7	20,3	11,2	85,6	34,3	58,73	138,36	67,37	61,09	85,72	20,38	20,11	20,54	140,09	158,6	124,21	151,48	–
49	**FIT** N7-S7-W7 **CV** E7	55,6	29,9	96,4	51,2	60,82	171,31	75,6	72,31	67,13	96,44	58,5	36,83	135,67	147,98	104,36	76,32	–
50	**FIT** N7-E7-W7 **CV** S7	35,1	17,9	177	63,6	40,19	119,06	162,44	79,7	46,91	28,83	177,11	25,73	73,23	123,8	144,41	60,22	–
51	**FIT** N7-E7-S7 **CV** W7	20,9	12,1	153,3	46,7	38,17	146,31	75,32	159,04	18,73	22,26	21,53	153,42	131,45	188,62	143,18	121,48	–
52	**FIT** E10-S10-W10 **CV** N10	13,4	8,6	72,9	29,5	73	15,52	10,38	13,83	78,84	105,81	104,78	60,31	147,75	45,56	132,22	214,3	–
53	**FIT** N10-S10-W10 **CV** E10	24,4	13,8	138,4	50,7	26,33	138,48	29,01	15,95	63,66	54,7	52,8	64,06	145	137,91	84,53	154,65	–
54	**FIT** N10-E10-W10 **CV** S10	16	9,4	237,1	84	17,95	15,74	237,29	14,28	61,82	169,83	251,99	61,39	166,6	79,25	240,79	124,31	–
55	**FIT** N10-E10-S10 **CV** W10	25	13,4	144,2	51,1	29,18	19,18	25,69	144,34	80,9	137,78	52,94	161,89	141,92	56,92	118,29	287,55	–
	mean	32,19	16,26	175,38	63,09	57,10	95,76	99,71	104,50	78,68	96,13	111,18	113,61	118,56	102,33	125,05	136,98	103,30
	standard deviation	16,19	6,66	83,61	28,79	28,03	59,93	72,61	105,21	33,91	53,80	77,21	106,71	45,64	61,07	57,90	111,48	67,79
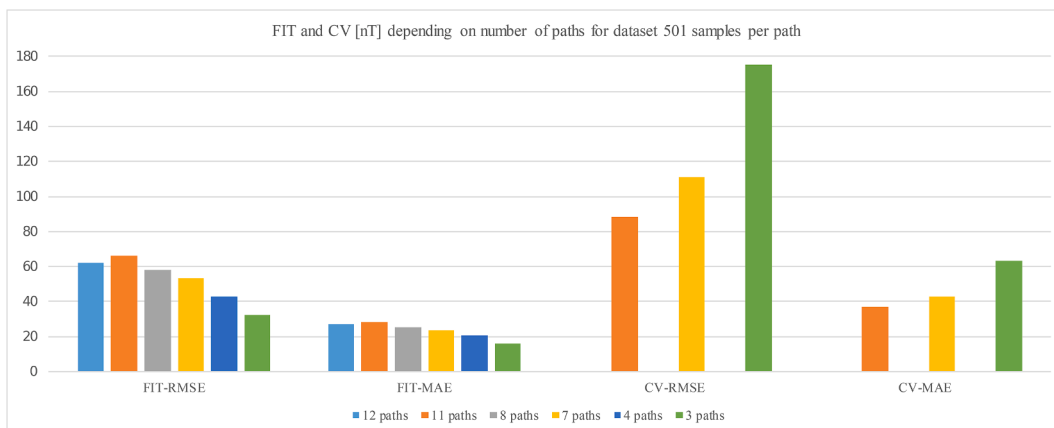
Fig. 26. MAE and RMSE values of fit and CV for different numbers of runs.
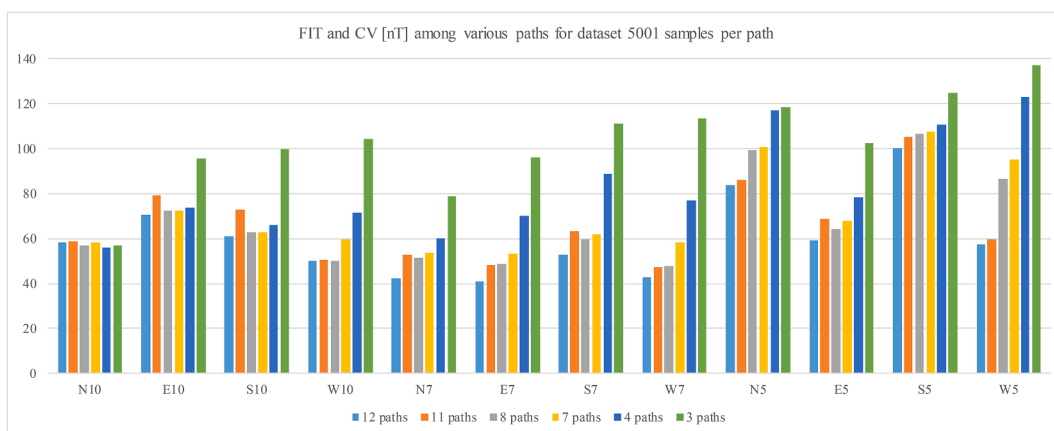
Fig. 27. FIT and CV values for all analyzed datasets.
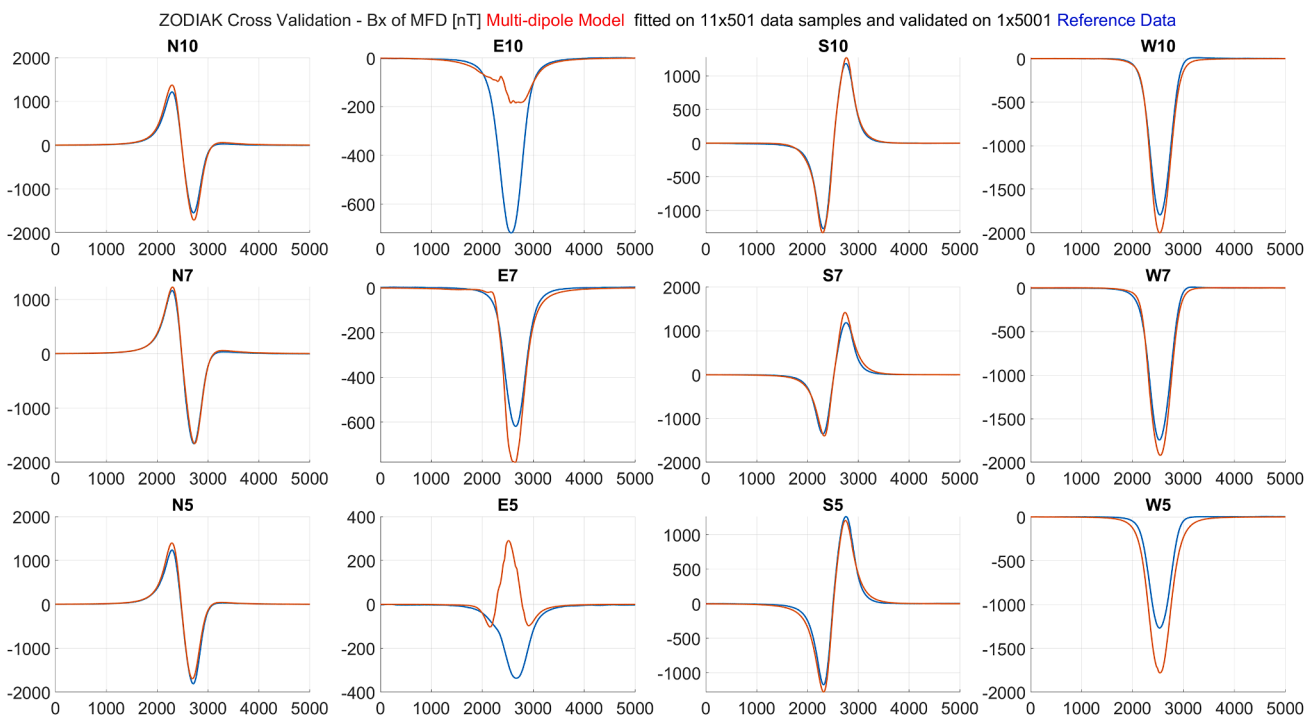
Fig. 28. CV results of MFD Bx component in comparison with reference data from measurements. Fitting on 501 probes from 11 paths, validated on 5001 probes from remaining path.

ZODIAK Cross Validation - By of MFD [nT] Multi-dipole Model fitted on 11x501 data samples and validated on 1x5001 Reference Data
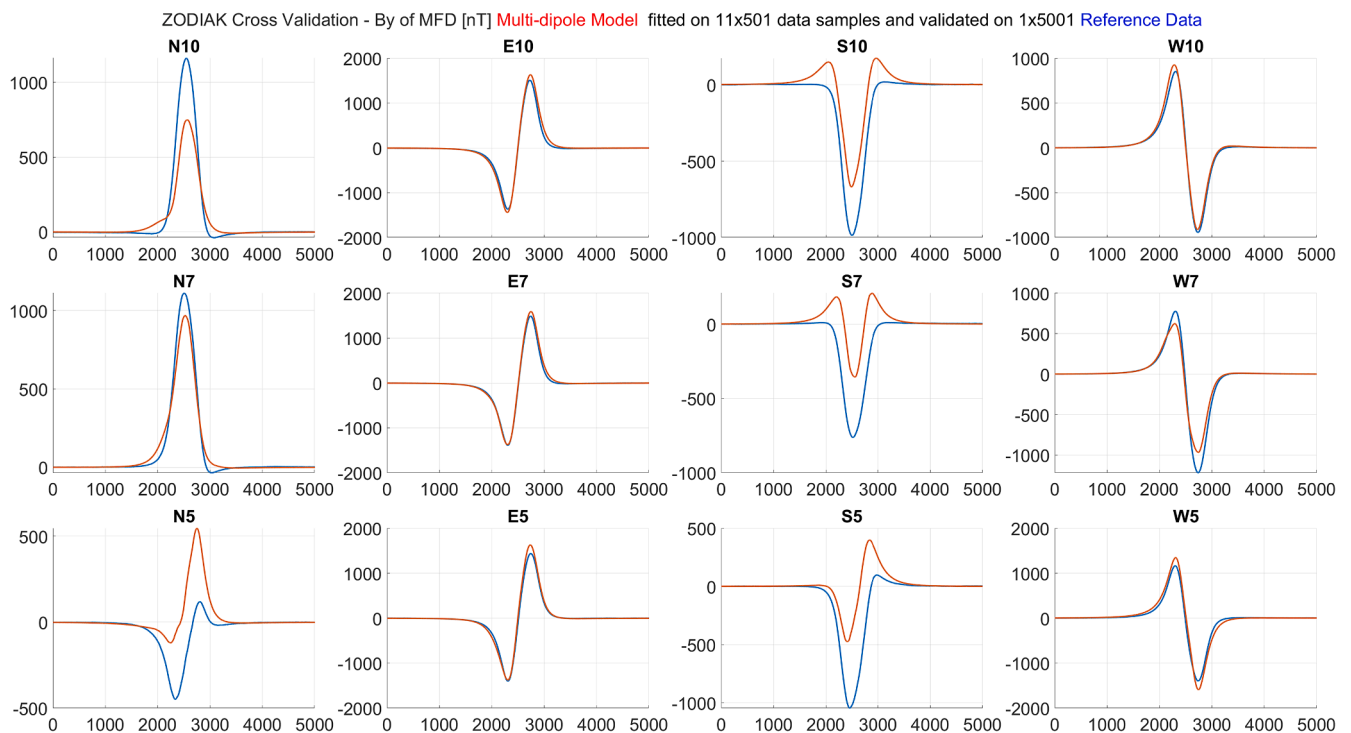
Fig. 29. CV results of MFD By component in comparison with reference data from measurements.

between the bias and the variance of the developed model [68]. Some of these issues were presented when the FIT waveform was considered in relation to the CV waveform in Fig. 17. This section presents other techniques to prevent overfitting and to improve model quality.

Regularization is usually applied to ill-posed optimisation problems as objective function modification. A classic approach in statistics and machine learning applications is adding a regularization term to the optimisation criterion to penalize the optimised variables and prevent

them from overgrowing. The two most common approaches to regularization are L1-Lasso and L2-Ridge [69].

In this paper, regularization is considered as another method that, operating on cost function modification, can potentially improve the quality of the model. The full list of regularization techniques considered in this paper includes:

ZODIAK Cross Validation - Bz of MFD [nT] Multi-dipole Model fitted on 11x501 data samples and validated on 1x5001 Reference Data
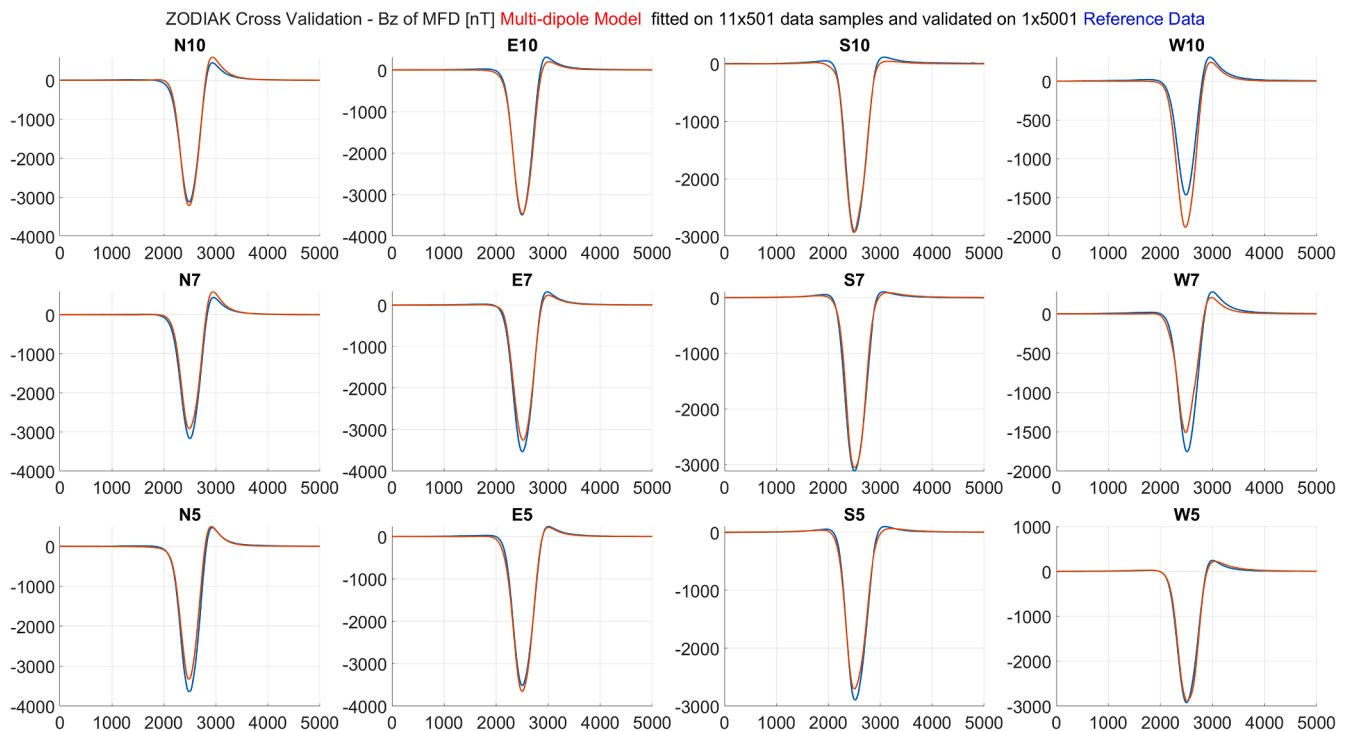
Fig. 30. CV results of MFD Bz component in comparison with reference data from measurements.
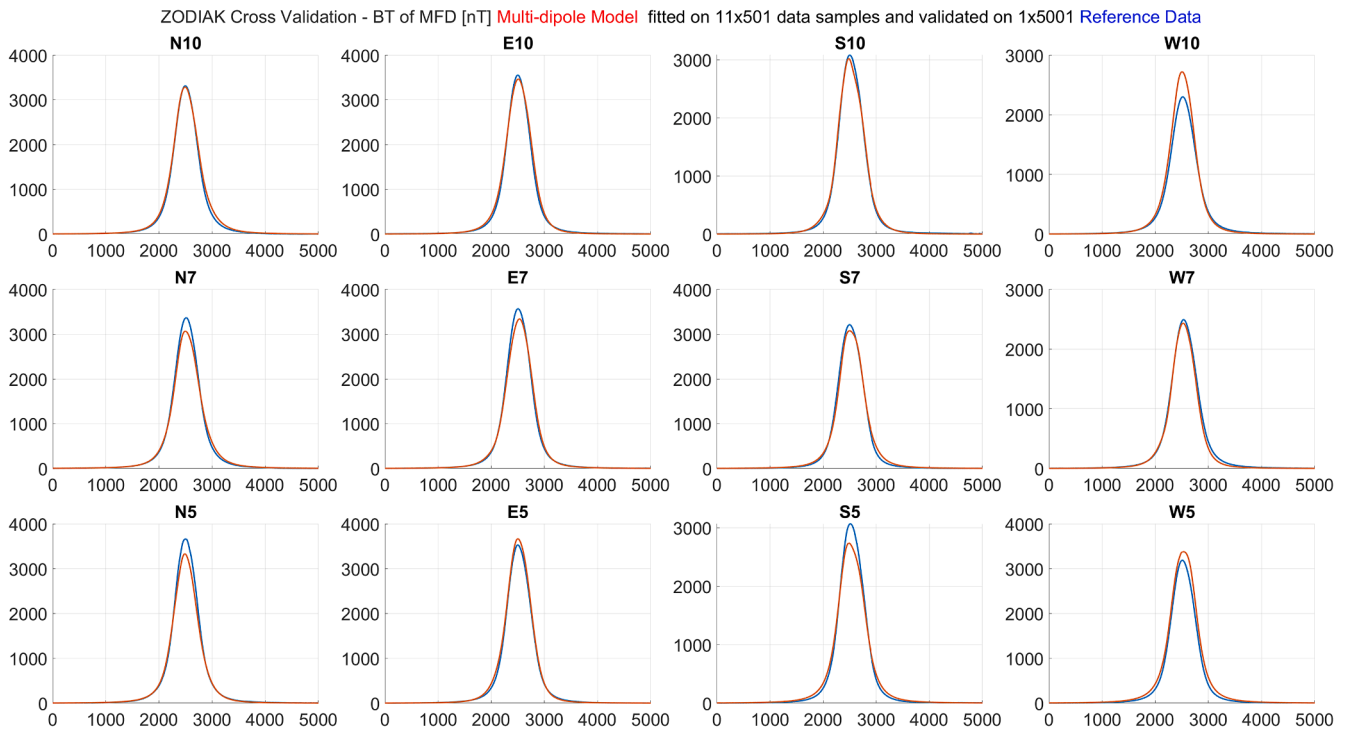
Fig. 31. CV results of MFD BT in comparison with reference data from measurements. Fitting on 501 probes from 11 paths, validated on 5001 probes from remaining path.

- simultaneous consideration of FIT and CV function to find the number of dipoles and the number of iterations for optimisation routine,
- L1 and L2 regularizations,
- increasing and decreasing the number of samples on the paths,
- strengthening and weakening the impact from individual paths in the optimisation criterion - integral and min–max approaches,

- strengthening and weakening the impact of data within the paths according to the GPS quality tags: FIX and HDoP,
- eliminating worst-fit paths from the dataset.

The quantitative results of the applied regularization methods are shown in Table 15 just before summarizing, while their qualitative evaluation is provided in Appendix in the form of an aggregated fit waveform.
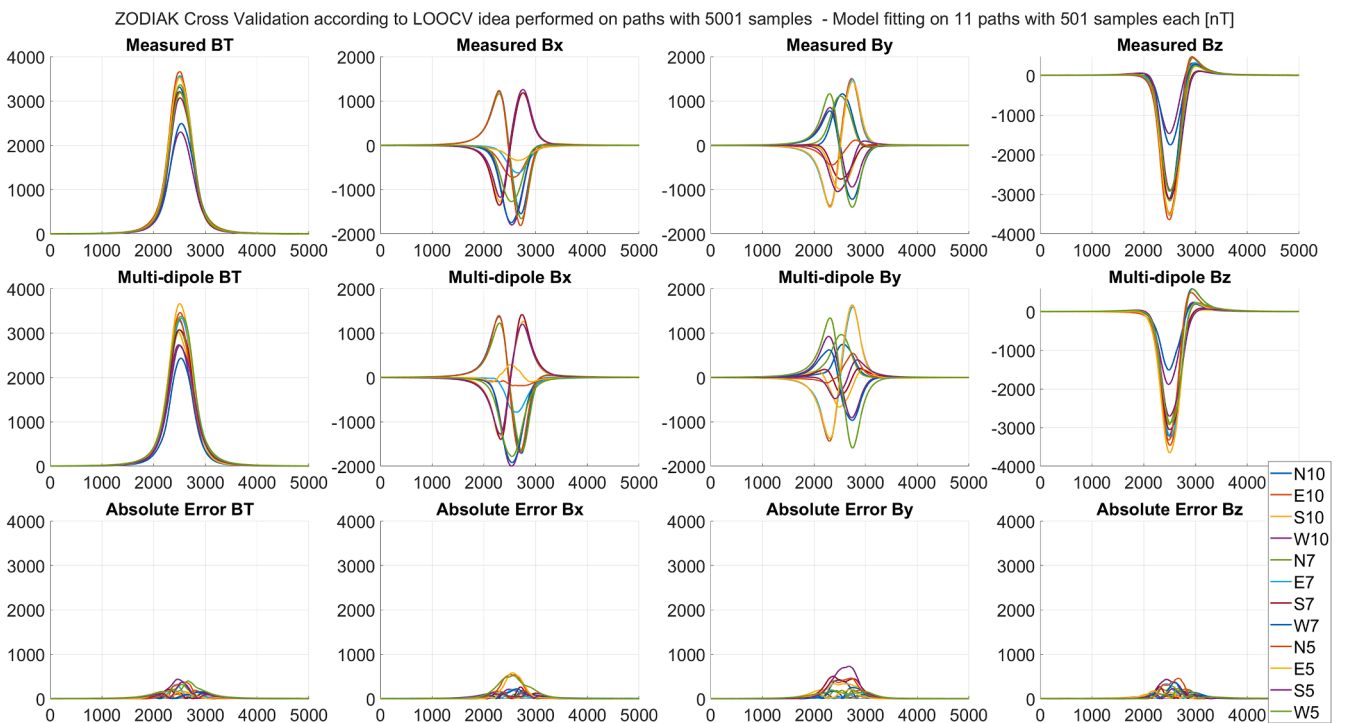
Fig. 32. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run. Fitting on 501 probes from 11 paths, validated on 5001 probes from remaining path.
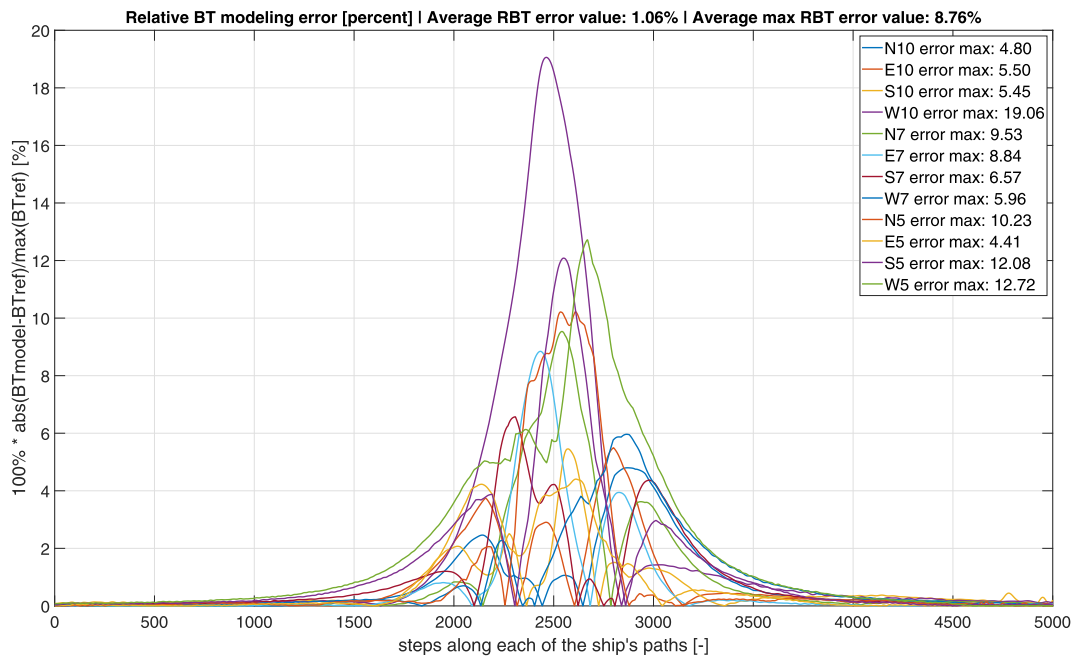
Fig. 33. Graphical summary of relative BT error distribution for each run.
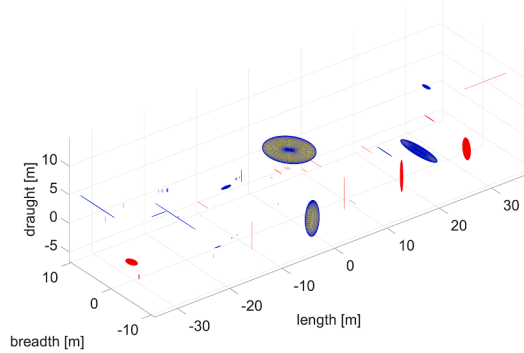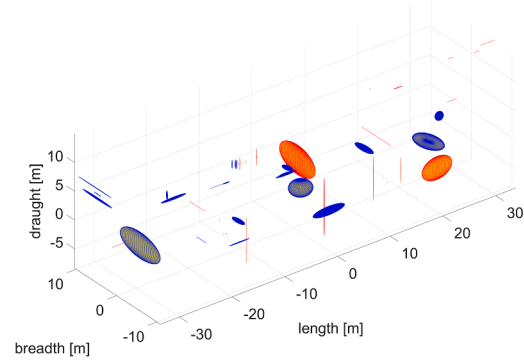
Fig. 34. Dipole order obtained with L1 Lasso regularization.
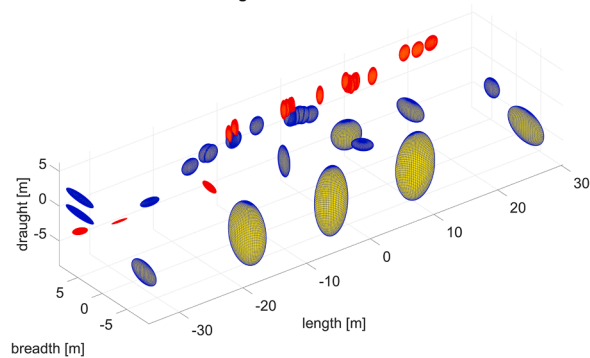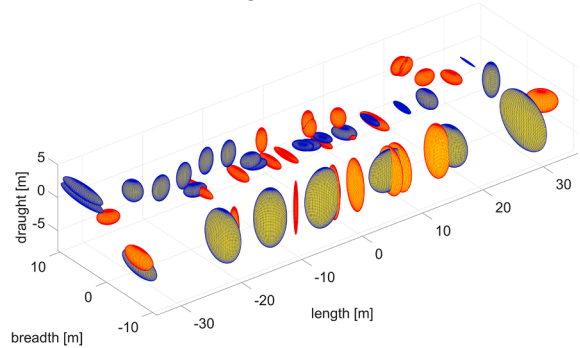
Fig. 35. Dipole order obtained with L2 Ridge regularization.

25
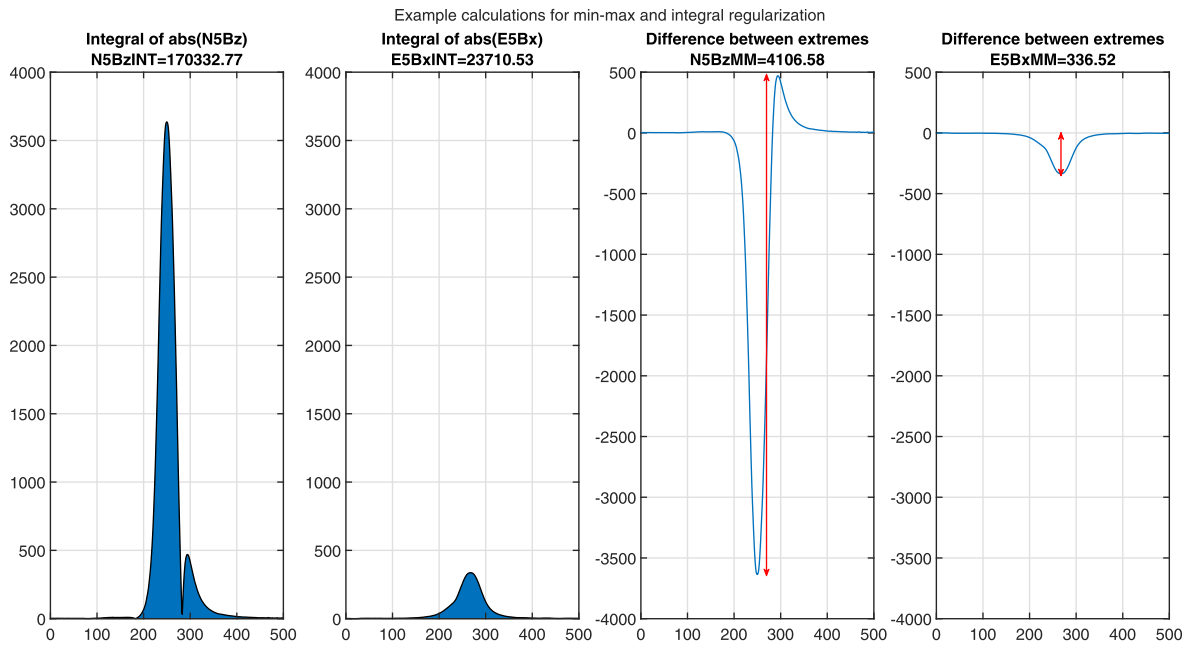
Fig. 36. Examples of calculating signal parameters for the need of integral (left pic) and min–max (right) regularization.

4.1. L1 and L2 regularizations

Both regularization methods L1 and L2 assume that the optimised cost function is supported with the term with optimised parameters. The L1 regularization, also called the Lasso regularization, works with absolute values of the optimised parameters, so the cost function (13) with L1 mechanism can be written as follows:

$$\min_{\Omega \in \{\Omega_1, \cdots, \Omega_{n+m}\}} J_{L1} = \sum_l \sum_d \sum_{j=1}^{501} \left(B_{l,d}^{ref}(j) - B_{l,d}^{model}(j, \Omega, \phi(d)) \right)^2 + \ldots$$
$$\ldots + \lambda \sum_i \sum_d \left| w \bullet m_{l,i} \right| \tag{20}$$

where:

$$l \in \{x, y, z\}, \qquad \text{and} \quad \mathop{\forall}_{i \in (1, m+n)} m_i \in \{m_{x,i}, m_{y,i}, m_{z,i}\}, \tag{21}$$

While, the L2 regularization, also called the Ridge regularization, uses squared values of the optimised parameters as penalty, so the cost function (13) with L2 mechanism can be written as:

$$\min_{\Omega \in \{\Omega_1, \cdots, \Omega_{n+m}\}} J_{L2} = \sum_l \sum_d \sum_{j=1}^{501} \left(B_{l,d}^{ref}(j) - B_{l,d}^{model}(j, \Omega, \phi(d)) \right)^2 + \ldots$$
$$\ldots + \lambda \sum_i \sum_d (w \bullet m_{id})^2 \tag{22}$$

Both methods need standardization of the optimised parameters in the regularization term to keep the balance between two terms in the optimisation criterion, (20) or (22). There are two main standardization techniques, one based on standard deviation and the other based on equalizing ranges of operations. The parameters being optimised in criteria (13), (20), and (22) are locations and magnetic moments of dipoles. The ranges of these variables are completely different, what can

Table 11

Min-max extreme differences and derived weighting factors among all data paths.

Path	MMBx [nT]	MMBy [nT]	MMBz [nT]	MMwBx [-]	MMwBy [-]	MMwBz [-]
N10	2760.82	1200.11	3568.00	1.49	3.42	1.15
E10	719.61	2874.39	3792.53	5.71	1.43	1.08
S10	2449.04	1001.60	3032.18	1.68	4.10	1.35
W10	1806.90	1794.25	1775.90	2.27	2.29	2.31
N7	2820.82	1145.87	3609.79	1.46	3.58	1.14
E7	622.09	2878.92	3854.11	6.60	1.43	1.07
S7	2538.54	772.61	3223.45	1.62	5.32	1.27
W7	1753.57	1995.43	2034.30	2.34	2.06	2.02
N5	3049.67	567.08	4106.58	1.35	7.24	1.00
E5	336.52	2838.17	3751.48	12.20	1.45	1.09
S5	2432.17	2070.40	2994.90	1.69	3.59	1.37
W5	965.80	2070.40	3170.51	4.25	1.98	1.30

Table 12

Integrals of absolute values and derived weighting factors among all data paths.

Path	INTBx [nT]	INTBy [nT]	INTBz [nT]	INTwBx [-]	INTwBy [-]	INTwBz [-]
N10	99377.68	53615.04	145537.20	1.71	3.18	1.17
E10	43657.78	101013.17	156285.24	3.90	1.69	1.09
S10	89948.85	47048.43	135352.42	1.89	3.62	1.26
W10	88241.76	69243.20	71784.64	1.93	2.46	2.37
N7	102698.66	53294.23	147373.63	1.66	3.20	1.16
E7	33706.82	102728.62	163492.68	5.05	1.66	1.04
S7	92167.58	36981.11	145774.20	1.85	4.61	1.17
W7	88268.61	77159.67	83621.87	1.93	2.21	2.04
N5	109537.38	24712.61	170332.77	1.56	6.89	1.00
E5	23710.53	101438.32	160077.60	7.18	1.68	1.06
S5	90865.67	94111.30	135598.81	1.87	3.23	1.26
W5	63271.89	94111.30	134966.86	2.69	1.81	1.26

Fig. 37. CV results after min–max and integral regularization for Bx MFD component.

be seen from Table 9. Since the dipole locations do not directly affect the value of the cost function, only magnetic moments will be applied in the penalty term. Based on observations of residual values, these magnetic moments should be transferred in range −1:1. Table 9 was established to search for expected values of magnetic moment weight w equal to 1E-6. The λ hyperparameter was chosen using a minimum value of CV with new optimisation calculations. The results of model quality improvement using parameters obtained from regularizations L1 and L2 are

presented in Table 15, rows E5 and E6.

What is worth additional presentation is the fact that the regularizations L1 and L2 have different impact on the dipole order in the dipole space. L1 forces some dipoles to vanish, saving those with bigger values as most important ones. That's why this method can be used for model shrinkage [70]. Sample dipole locations and magnetic moments regularized by the L1 method are presented in Fig. 34.

A different dipole behaviour can be observed with the L2

Fig. 38. CV results after min–max and integral regularization for By MFD component.

Bz MFD Component [nT] Cross Validation Reference, MIN-MAX and INTEGRAL Regularization
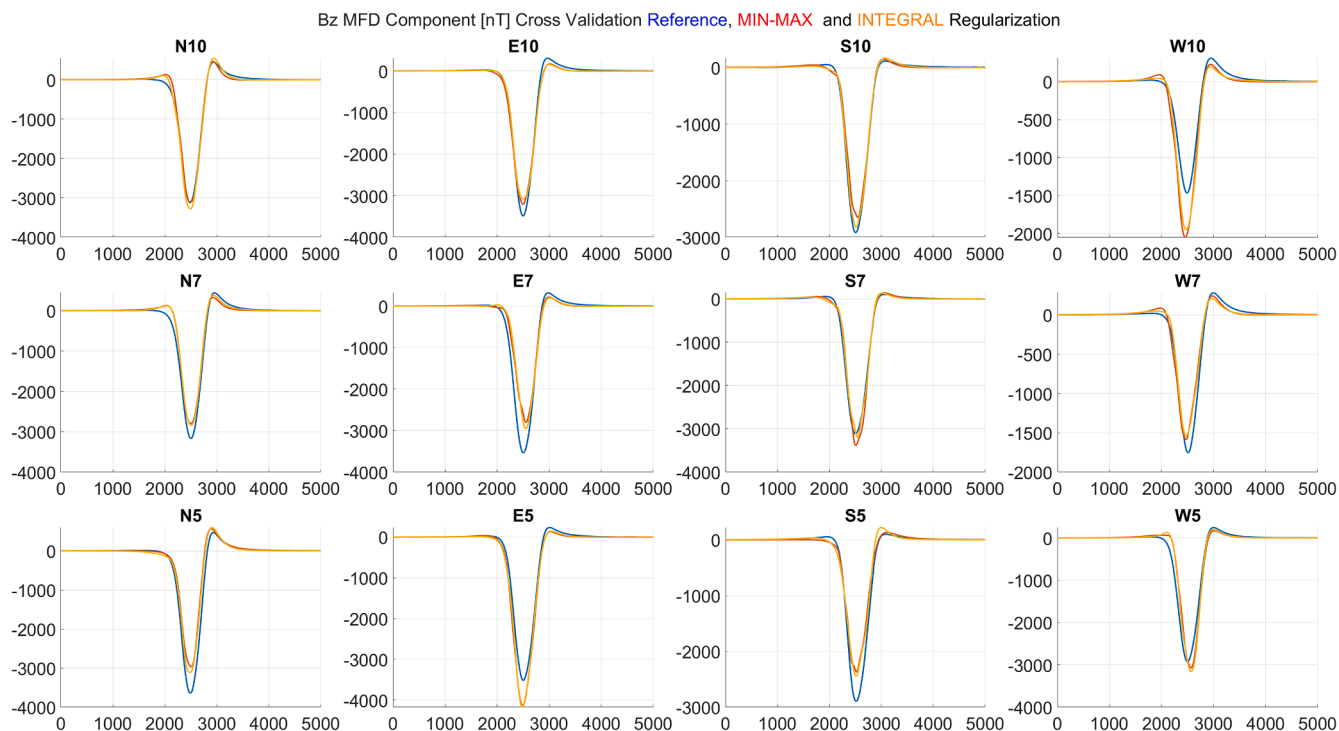
Fig. 39. CV results after min–max and integral regularization for Bz MFD component.

regularization. All dipoles were forced to reduce their magnetic moments almost equally. All dipoles are still present with similar magnetic moment values, and the regularization introduced an interesting regular order presented in Fig. 35.

The dipoles presented in Figs. 34 and 35 were obtained for large value of λ hyperparameter, what presents the nature of each regularization method. However, it has to be pointed that the minimum values of CV and best regularization scenarios are rather closer to these presented in Figs. 18–20.

Table 13
Number of simulation scenarios for different numbers of rejected paths.

Number of paths rejected from set	1	2	3	4
Number of paths remaining in set	11	10	9	8
Number or path combinations	12	66	220	495
Number of single calculations to perform	12*11 = 132	66*10 = 660	220*9 = 1980	495*8 = 3960

BT of MFD [nT] Cross Validation Reference, MIN-MAX and INTEGRAL Regularization
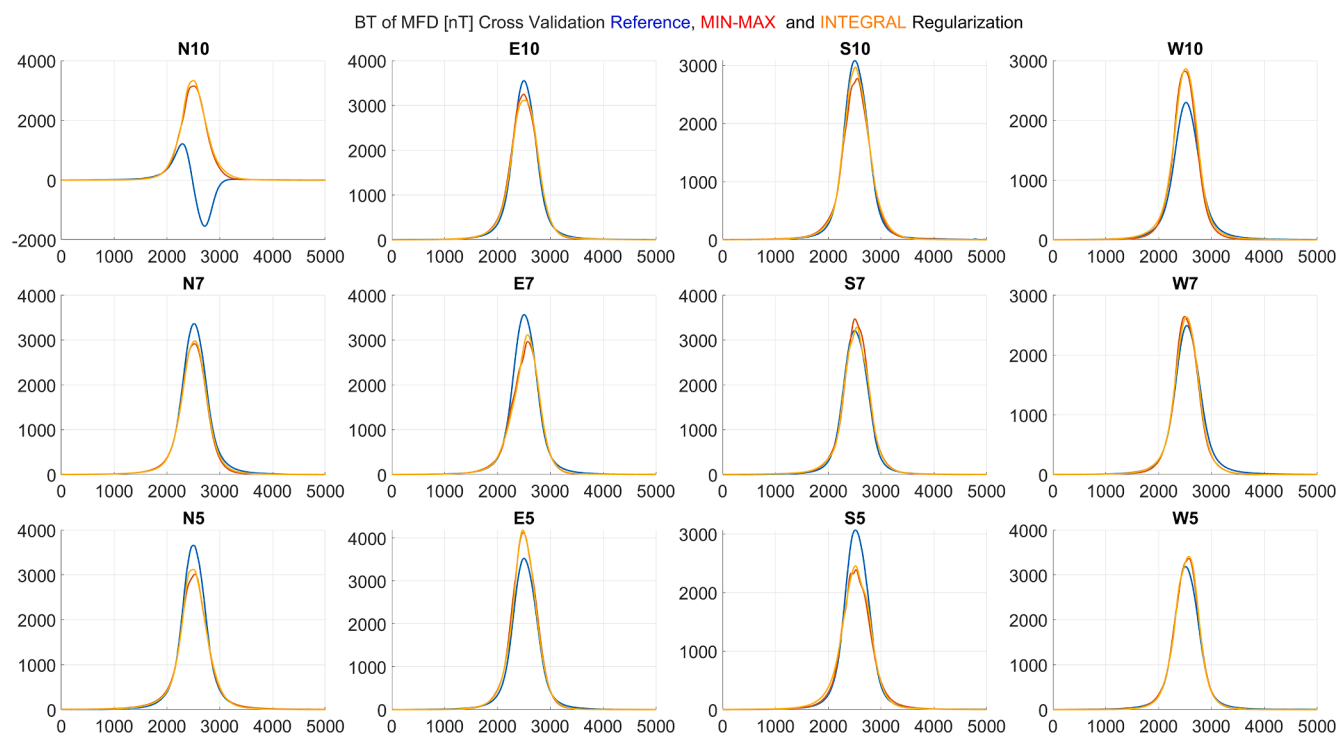
Fig. 40. CV results after min–max and integral regularization for MFD BT.
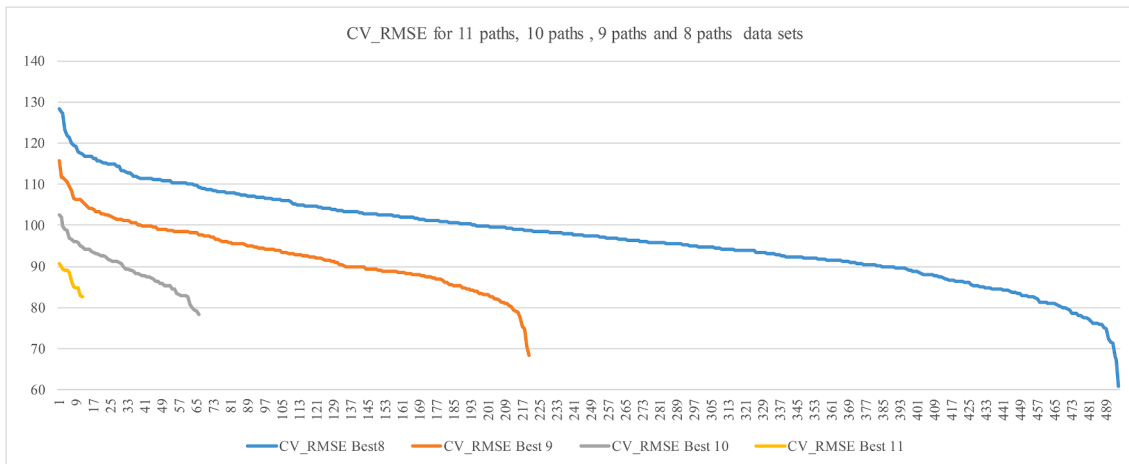
Fig. 41. CV-RMSE results for all combinations of sets with 8, 9, 10 and 11 paths.

Table 14
The best results obtained in different experiments without paths.

Name	Best results obtained without paths
Best of 8	S10, E7, S7 and S5
Best of 9	S10, S7 and S5
Best of 10	S5 and W5
Best of 11	W5

4.2. Strengthening and weakening the impact from individual paths in the optimisation criterion: Integral and min–max approaches

Modelling errors on paths presented in Figs. 28–30 exists mainly in situations where the values of one MFD component are significantly different in value from those of the other component. In an attempt to eliminate this error, it was worth analysing whether the normalization of component values would improve the results. Two different methods were applied according to the idea presented in Fig. 36.

The first scaling method was derived as the difference between extremes of signals. The $MinMax_{Global}$ indicator was found among all signals according to formula (23), and then it was used as a scaling term. Then, the $MinMax_{Local}$ terms were established with (24) for all signals and components. Hence, the optimisation criterion is described by the objective function expressed as (24).

$$MinMax_{Global} = \forall_{l,d,j} \max\left(\max\left(B_{l,d}^{ref}(j)\right) - \min\left(B_{l,d}^{ref}(j)\right)\right), \quad (23)$$

$$MinMax_{Local}(l,d) = \max\left(\max\left(B_{l,d}^{ref}(j)\right) - \min\left(B_{l,d}^{ref}(j)\right)\right), \quad (24)$$

$$\min_{\Omega \in \{\Omega_1, \cdots, \Omega_{n+m}\}} J_{MinMax} = \sum_l \sum_d \sum_{j=1}^{501} \left(\frac{MinMax_{Global}}{MinMax_{Local}(l,d)}\right) \cdots$$
$$\cdots \left(B_{l,d}^{ref}(j) - B_{l,d}^{model}(j, \Omega, \phi(d))\right)^2. \quad (25)$$

The *Min–Max* coefficients (MMBx, MMBy, MMBz) and final weights (MMwBx, MMwBy, MMwBz) of each signal are gathered in Table 11.

Due to the fact that signals have not only different heights but also

Table 15
Summarized results of regularization methods in terms of model quality.

Experiment	Name	Paths no.	Paths for FIT	No.of data per path	FIT/CV with original data size				FIT/CV with 5001 data size				AR BT [%]	AMR BT [%]
					FIT-RMSE [nT]	FIT-MAE [nT]	CV-RMSE [nT]	CV-MAE [nT]	FIT-RMSE [nT]	FIT-MAE [nT]	CV-RMSE [nT]	CV-MAE [nT]		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
E1	data 501	12	11	501	66,49	28,48	88,34	36,83	66,56	28,54	88,41	36,89	1,06	8,76
E2	data 301	12	11	301	82,1	43,6	111,8	58,5	63,73	27,14	86,73	36,19	1,06	8,76
E3	data 5001	12	11	5001	64,97	27,27	87,18	35,83	64,97	27,27	87,18	35,83	1,00	8,78
E4	data 8001	12	11	8001	52,28	18,10	71,33	23,67	66,11	28,07	90,19	36,89	1,01	8,57
E5	L1 Regularization	12	11	501	73	30,15	87,88	36,07	67,53	28,43	87,95	36,13	1,05	8,83
E6	L2 Regularization	12	11	501	72,73	29,94	86,97	34,81	67,52	27,79	87,03	34,87	1,01	8,17
E7	integral regularization	12	11	501	79,76	33,89	106,67	44,83	79,83	33,97	106,76	44,93	1,41	14,10
E8	min–max regularization	12	11	501	85,18	35,42	110,77	46,63	85,27	35,47	110,87	46,73	1,55	15,81
E9	GPS FIX	12	11	3043 *	–	–	–	–	168,28	46,32	184,27	59,56	1,78	27,20
E10	GPS HDoP<=1.4	12	11	6321 *	–	–	–	–	71,90	29,08	102,07	40,62	1,17	11,29
E11	GPS weighted 1_2	12	11	8001	54,95	17,69	73,33	23,44	69,5	27,4	92,7	36,6	1,00	9,04
E12	best of 11 paths	11	10	501	62,60	25,00	82,60	32,20	140,55	90,42	90,69	35,86	1,06	10,88
E13	best of 10 paths	10	9	501	54,80	22,40	78,40	31,30	127,37	84,53	87,69	34,87	1,05	10,81
E14	best of 9 paths	9	8	501	45,30	17,60	68,50	25,40	102,4	71,3	80,5	30,6	0.99	9,62
E15	best of 8 paths	8	7	501	32,70	14,10	60,80	23,00	87,9	65,0	73,3	28,5	1,19	11,31

* avg number, different number of data across the paths.

widths, the min–max approach might be considered too simple and inappropriate for the present regularization task. That was why another scaling method based on integral operation with the right rectangular approximation method was introduced, as expressed in (26) and (27). In this case, the objective function with integral scaling is defined by (28):

$$Integral_{Global} = \forall_{l,d} \max \sum_{j=1}^{501-1} \left| B_{l,d}^{ref}(j) \cdot h(j+1) \right|, \tag{26}$$

$$Integral_{Local}(l,d) = \sum_{j=1}^{501-1} \left| B_{l,d}^{ref}(j) \cdot h(j+1) \right|, \tag{27}$$

$$\min_{\Omega \in \{\Omega_1, \cdots, \Omega_{n+m}\}} J_{Integral} = \sum_{l}\sum_{d}\sum_{j=1}^{501} \left(\frac{Integral_{Global}}{Integral_{Local}(l,d)} \right) \cdots$$

$$\cdots \left(B_{l,d}^{ref}(j) - B_{l,d}^{model}(j, \Omega, \phi(d)) \right)^2, \tag{28}$$

where $h(j+1)$ denotes the Euclidean distance between locations x_j, y_j and x_{j+1}, y_{j+1} of samples j and $j+1$, respectively. The *Integral* coefficients (INTBx, INTBy, INTBz) and final weights (INTwBx, INTwBy, INTwBz) of each signal for all analysed datasets are gathered in Table 12.

The results of min–max and integral regularization are presented in Figs. 37–40. Despite almost equal representation of all considered signals in the cost function, the errors still exist on paths (Bx E5, By N5, and By S5). The shape of By N5 and By S5 signal waveforms reinforce the thesis that the source of the problem is the error related to determining ship's position. Moreover, new errors appeared in BT which had not been seen before min–max and integral regularization. The remaining results in terms of model quality can be found in Table 15, rows E7 and E8.

4.3. Increasing and decreasing the number of samples on the paths

The base simulation scenario assumes 501 samples of data for each of the 12 paths and for 3-MFD components Bx, By, Bz. This value was confirmed to be sufficient in many simulations with synthetic data, including artificially noised data which are close to the real ones. The amount of data used for fitting is always an issue that is worth examining during model development. Calculations with 301, 501, 5001 and 8001 data per each path were conducted. Details of fitting and validation are presented in Table 15, rows from E1 to E4. Considering the amount of fitting data, it should be reminded that the optimisation procedure needs to find 312 model parameters.

4.4. Strengthening and weakening the impact of data within the paths according to the GPS FIX and HDoP quality tags

During the recording of ship's position, additional qualitative GPS indicators were also saved. Since the accuracy of determining ship's position seems to be one of the crucial factors of model quality, GPS FIX and HDoP should be taken into consideration. The waveforms for GPS FIX, HDoP, and the number of satellites are presented in Figs. 11–13, respectively.

The GPS FIX scenario was conducted only for data with mark GPS FIX = 2. As can be seen in Fig. 11, the distribution of data with higher GPS precision is heterogeneous. For N5 and N7 data sets, there are only a few data with mark 2 of GPS FIX. Therefore, when using this criterion, some paths will be much less represented in the objective function. In order to prevent this, but at the same time to use the GPS FIX tag, another scenario, called 'GPS weighted 1_2', was implemented. This scenario consisted in taking an equal number of data on each path and considering the data from FIX 2 as more reliable (with weight 2 in the criterion). HDoP is also a qualitative marker of the accuracy of GPS data. After observing the distribution of these data, those with HDoP <= 1.4 were selected for the analysis. Details of fitting and validation for scenarios related to GPS quality indicators are presented in Table 15.

4.5. Evaluation with paths exclusions

The measurements on individual paths were recorded independently of each other at different times of the day, with the ship travelling at different speeds. In view of the errors of matching as a result of optimisation, a hypothesis was formulated that some of the measurements are contradictory, i.e. mutually exclusive. This may be, for example, due to errors in determining ship's position and other GPS interference profiles during the measurement campaign. If these errors are evenly distributed in all paths, they will not be detected. However, if the error was particularly large in a particular transition (or generally spread across the paths), removing this transition from the set on which the calculations are performed should indicate particularly error-prone paths and correct the rest of the path after discarding them. Therefore, it was decided to perform calculations for a situation where all scenarios were tested with the rejection of 1, 2, 3, or 4 paths (rejecting a larger number of paths would lead to significant reduction of information in the model teaching phase, along with excessive multiplication of combinations). Formula (29) for combination without repetitions gives the number of unique combinations for *k* of *n* elements. Using this formula we can calculate the number of combinations resulting from the rejection of a certain number of paths from the full set of 12 paths, (30).

$$C_n^k = \frac{n!}{k!(n-k)!} \tag{29}$$

$$C_{12}^{11} = \frac{12!}{11!(12-11)!} = 12, \qquad C_{12}^{10} = \frac{12!}{10!(12-10)!} = 66,$$

$$C_{12}^{9} = \frac{12!}{9!(12-9)!} = 220, \qquad C_{12}^{8} = \frac{12!}{8!(12-8)!} = 495. \tag{30}$$

For example, eliminating one path from 12 paths with data N10, E10, S10, W10, S7, E7, S7, W7, N5, E5, S5, W5 we get 12 possible combinations of 11 paths, while by eliminating 2 paths from the 12-path set we obtain 66 unique 10-path combinations. Using the *leave-one-out CV* approach, another path can be selected from the remaining paths for verification and re-examining the number of combinations. The final number of simulation scenarios is presented in Table 13. By applying the LOOCV approach to each of these unique sets and considering the appropriate number of remaining paths, we obtain the number of individual computational tasks.

The CV_RMSE results for all combinations of sets with 8, 9, 10 and 11 paths are shown in Fig. 41. More detailed results with best co``mbinations are shown in Appendix, Experiments E12 - E15 (Figs. A9–A12). When assessing the results, it can be seen that the principle is maintained where the less data, the worse the average fit. However, it can also be seen that individually the best scenarios occur for the fewest sets. The best results obtained in different experiments without given paths are presented in Table 14.

The low evaluation of data on paths S5 and W5 which was obtained as a result of other analyzes is confirmed here. The best results obtained for scenarios with 8 and 9 paths do not necessarily translate into the best quality of the model. Excluding successive paths reduces the generality of the model.

4.6. Summary of regularization experiments

Table 15 presents the results of all experiments conducted to enable comparison of the results among different configurations. Thus, the table includes the FIT/CV sections for the experiments conducted for the original number of data, and also for the reference data set containing 5001 data per single path. As already mentioned, it isn't easy to evaluate a given experiment based on a single result. The quality of the model is mainly confirmed by the parameters CV_RMSE and CV_MAE, but also by the error related to BT.

Comparing the amount of data per path used in the FIT phase

calculations, the results are very close. The quality of model fitting increased with the growth of data used for calculations, but this does not carry over to the validation phase, where the results are similar. The best results of CV are obtained for the smallest number of data, but the smallest BT error occurrs for the largest data set. An improvement in the results can be seen in the regularization experiment L2, where the best CV_MAE and AMR_BT values were obtained. The regularization with penalization of magnetic moment values turned out to be more effective than the prevention of overfitting by the method of diverging FIT and CV curves. Additionally, L1 regularization can be used to reduce the number of dipoles and make the model more sparse when needed. Regularization in the form of amplification of errors from the weakest signals in the optimisation criterion improved the waveforms on these weakest paths, but ruined the final result. The GPS FIX and HDoP regularizations were based on pure GPS quality signal, and therefore the accuracy of determining the position should be impruved but the results were deteriorated probably by hetrogenous data distribution. However, the results obtained with weighted values depending on the GPS FIX mark may be classified as reasonable. The path removal analyzes lead to significant improvement in the CV indicator, accompanied by deterioration of BT. It should be noted that despite the attempt to reduce the comparison of sets with less than 12 paths to the same conditions, they could not be directly compared with the rest of the rate.

All unpresented previously scenarios are provided in Appendix (Figs. A1–A12) to enable qualitative comparison of experiments E1 – E15. It should be emphasized that most of them show a very good match. Compared to mathematical models based mainly on synthetic data generated in simulators, visible model unperfections resulting from real measurement data usage have to be mentioned. The applied real data is burdened with errors, mainly in determining the position. Considering that each registered position may contain location errors of +/-5 m and +/- 12 m, some of the data may even be contradictory. Mean-square adjustments of model parameters make the model partially resistant to such errors, but an attempt to improve the representation of magnetic signatures quality should be related to higher quality of determining ship's position in relation to the magnetometer.

Due to the uncertainty in determining the ship's position in relation to the UMM, basic analysis related to the influence of those errors on the quality of the magnetic signatures reproduced by the multi-dipole model were carried out. They were based on disturbing the ship position by the selected minimum and maximum values equal to +/- 5 m and +/- 12 m, respectively. The obtained results of such experiments are presented in Appendix - Figs. A13–A20. Their analysis leads to the conclusion that the observed magnitude of the relative errors associated with the reconstruction of the magnetic signatures by the multi-dipole model varies within the range of 22% and 50% for position disturbances at the level of +/- 5 m and +/- 12 m, respectively. Furthermore, the values of the CV_RMSE and CV_MAE indicators in these cases were several times larger than the values observed in Table 15, for the basic undisturbed scenario E1. The issue of the effect of position disturbance on the resulting magnetic signatures will be comprehensively analyzed by the Authors in further research work.

5. Final conclusions

As a result of the measurement campaign followed by the synthesis and tuning of the multi-dipole approach, a predictive model was obtained that can reproduce magnetic signatures of the ship *Zodiak*. The comprehensive analysis and verification was performed to show the rational behaviour of the model, its high accuracy in reproducing the Bz MFD component and BT, and mostly at least good accuracy for the Bx and By components. The key parameter influencing the quality of reproduction of magnetic signatures seems to be the accuracy in

determining the vessel position in relation to the UMM. The standard GPS accuracy is a limitation, especially when the measurement concerns the position very close to the UMM. A special feature of the data available for *Zodiak* is that they were measured with a single sensor at different times, at different vessel speeds and trajectories. Usually, for magnetic signature reproduction purposes, the measurements consist of simultaneous recording from several sensors. The approach used at the test site in Gdynia, while maintaining the maximum meticulousness and accuracy of the measurement crew, was exposed to errors resulting from inaccuracies in determining the location and synchronizing all measurements. In a few cases, applying the data from real measurement to the multi-dipole model revealed errors that had not been found in calculations with synthetic data. Several post-processing regularization methods were applied to check if the modified optimisation criterion or different sets of data will allow adjusting the multi-dipole model to the real data burdened with error. The evaluation of the obtained model quality is multi-criterial, and the most common effect was that when improving one factor, the deterioration of another factor was observed. Suspecting that some of the data had a greater error than the rest, they were removed from the analysed dataset. Brute-force search or exhaustive search calculations were performed to find the best combination of data and exclude the least matching ones. An improvement was obtained in the model validation index. However, it is difficult to determine whether getting rid of some data to improve the efficiency of the model in a limited set does not lead to the loss of quality of the model in general.

Summing up, despite problems with data integrity, the multi-dipole model with the data from the test site in Gdynia allowed to build a model that would enable the reconstruction of magnetic signatures of the ship *Zodiak* with high accuracy. In order to improve the quality of reproducing magnetic signatures in the future, measurements with a larger number of sensors should be made to provide more diversified data but recorded under the same conditions. All possibilities of improving the determination of ship's geographical position should also be taken into account.
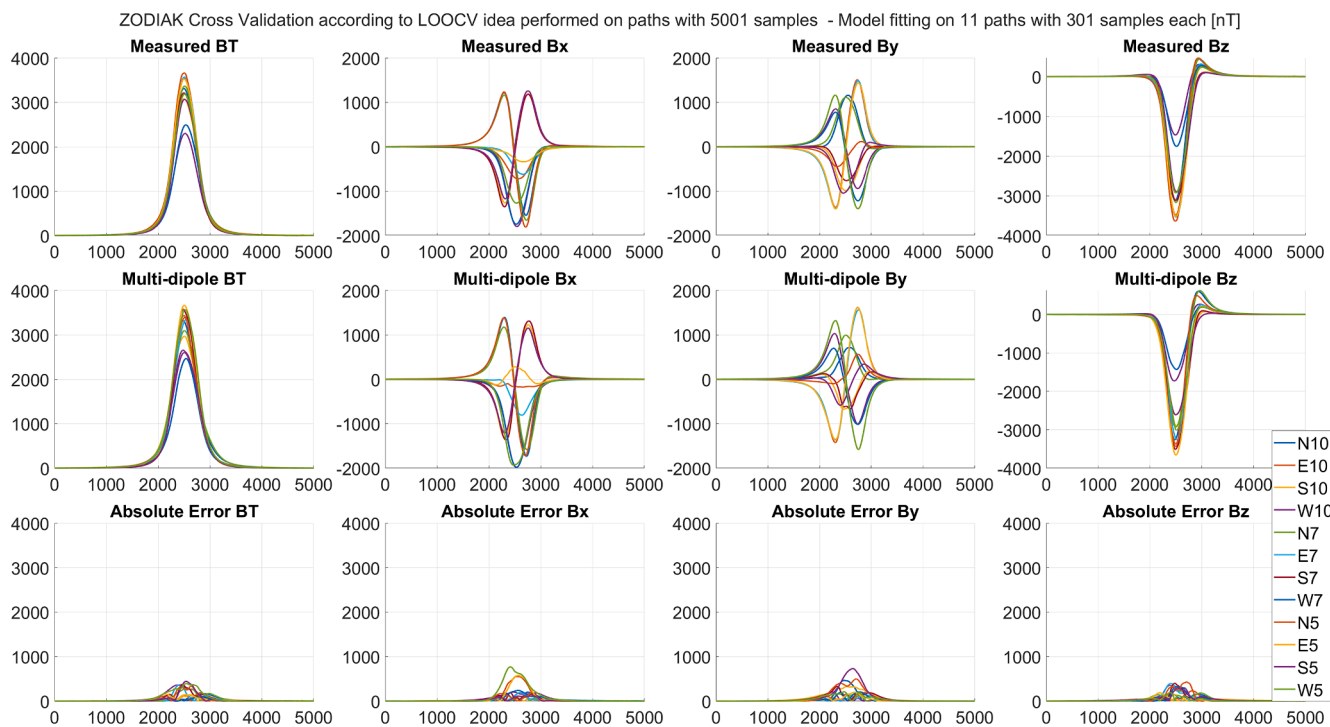
Appendix A

See Figs. A1–A20.

Fig. A1. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E2.
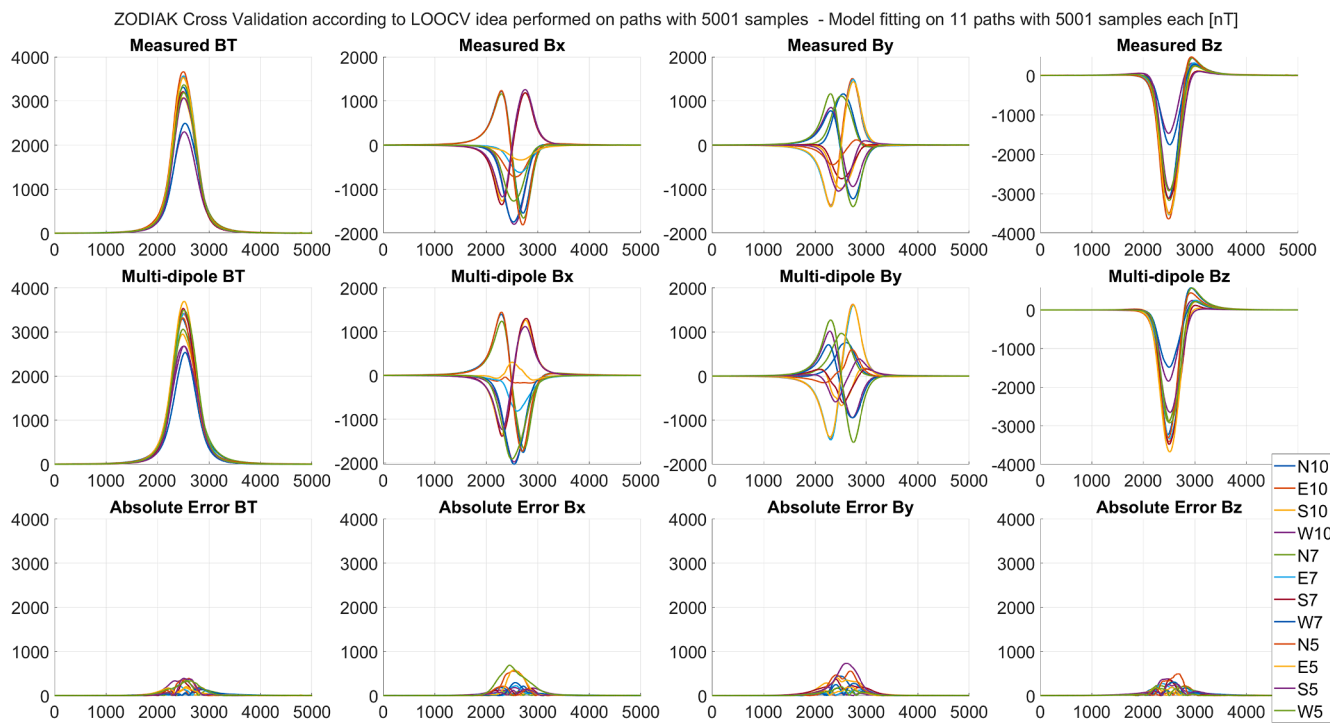

Fig. A2. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E3.
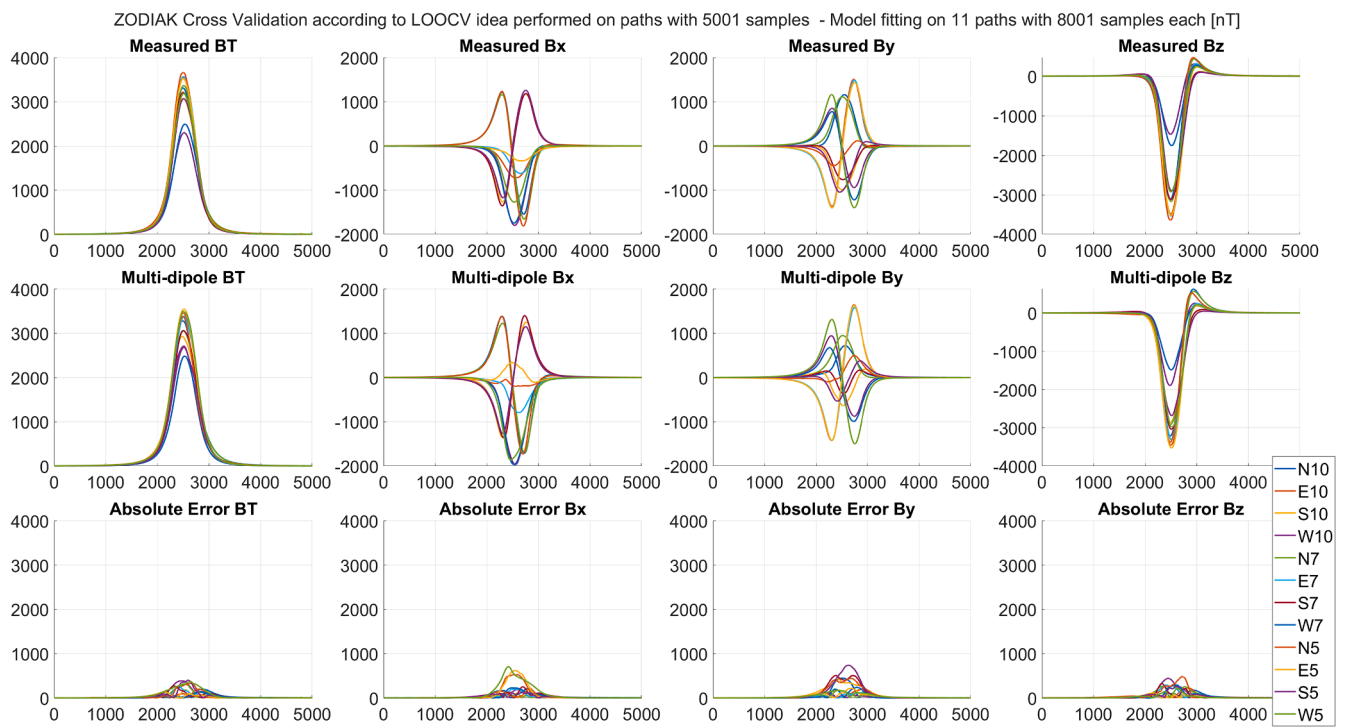
Fig. A3. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E4.
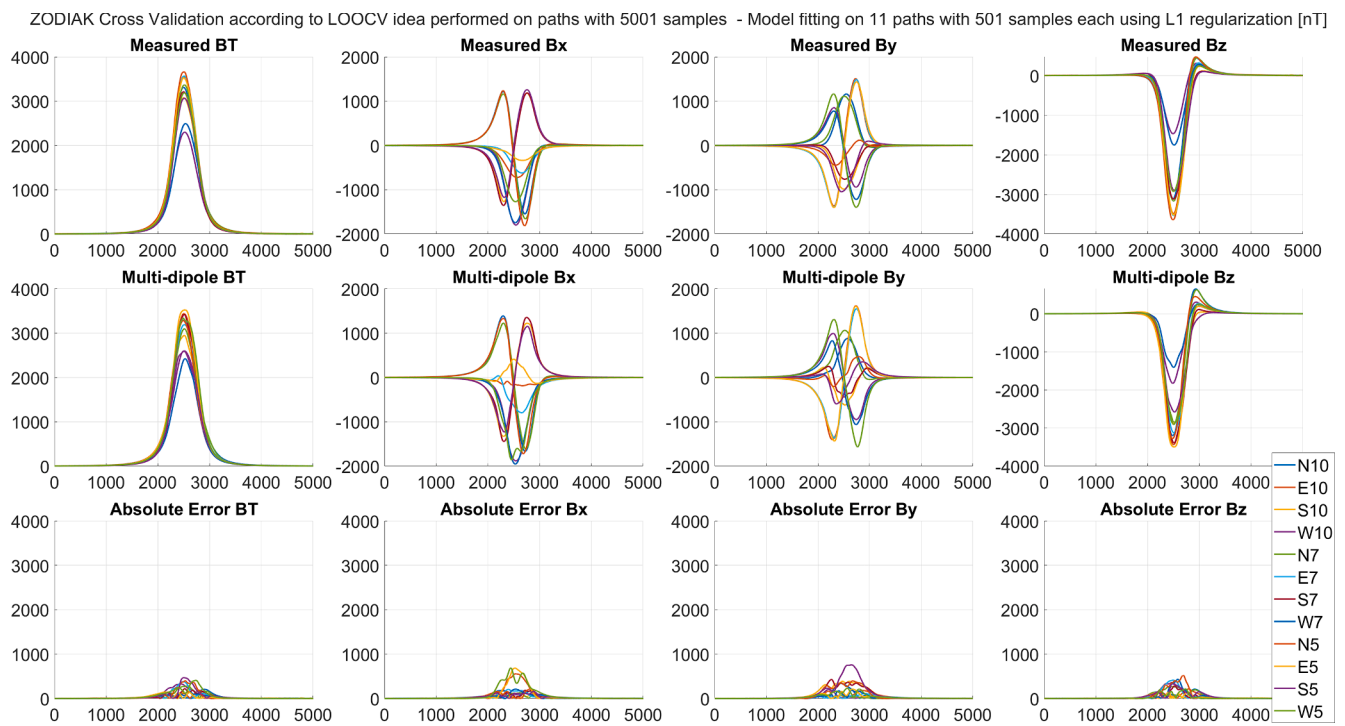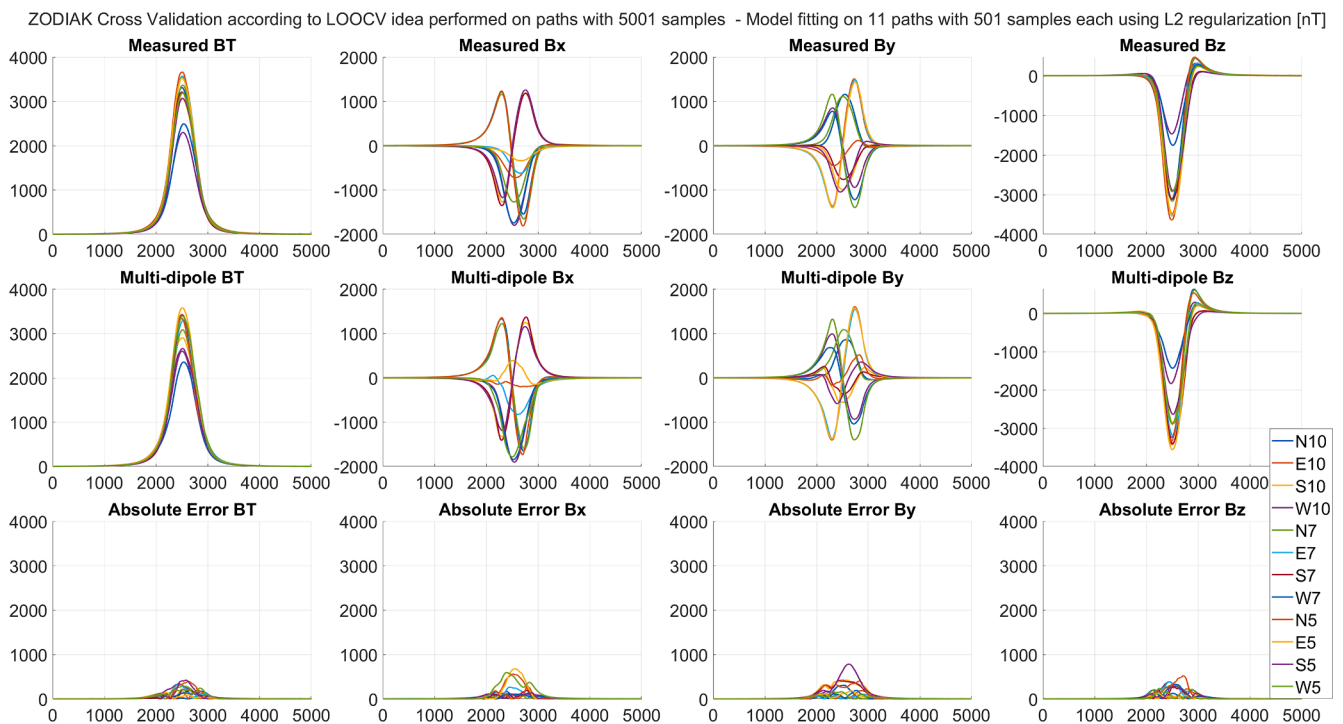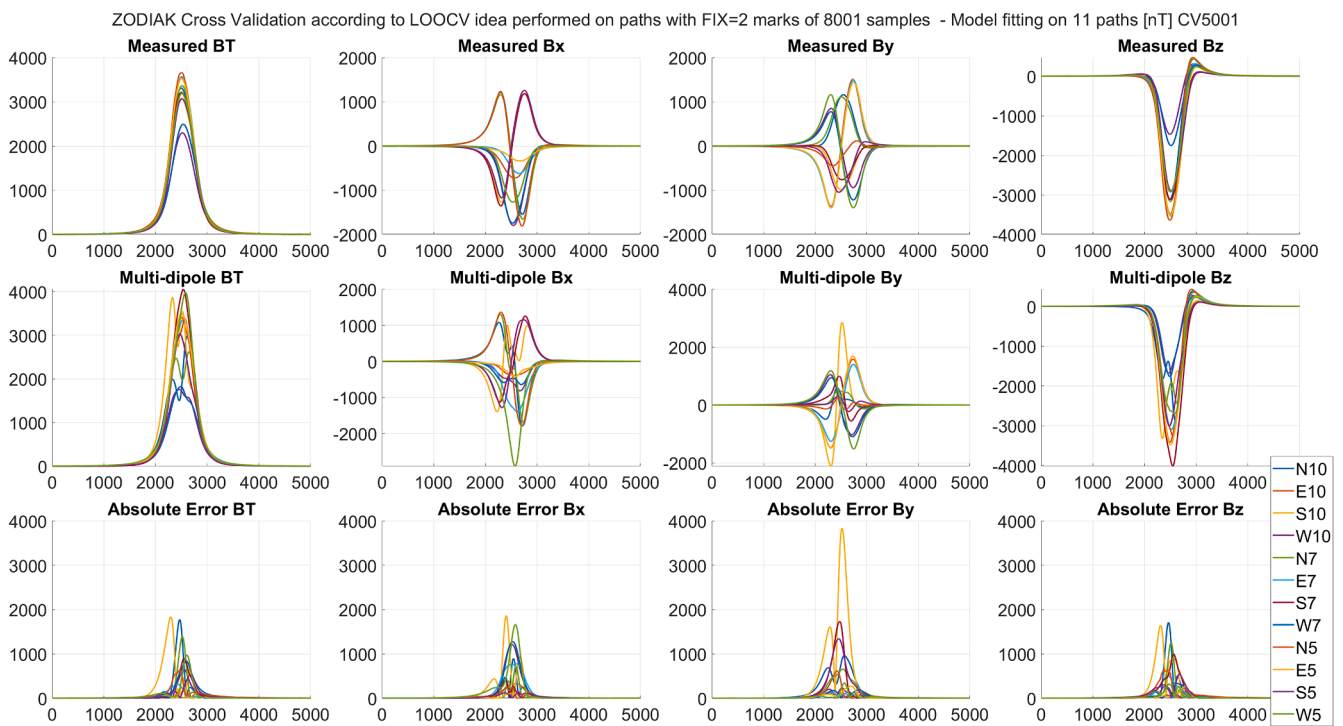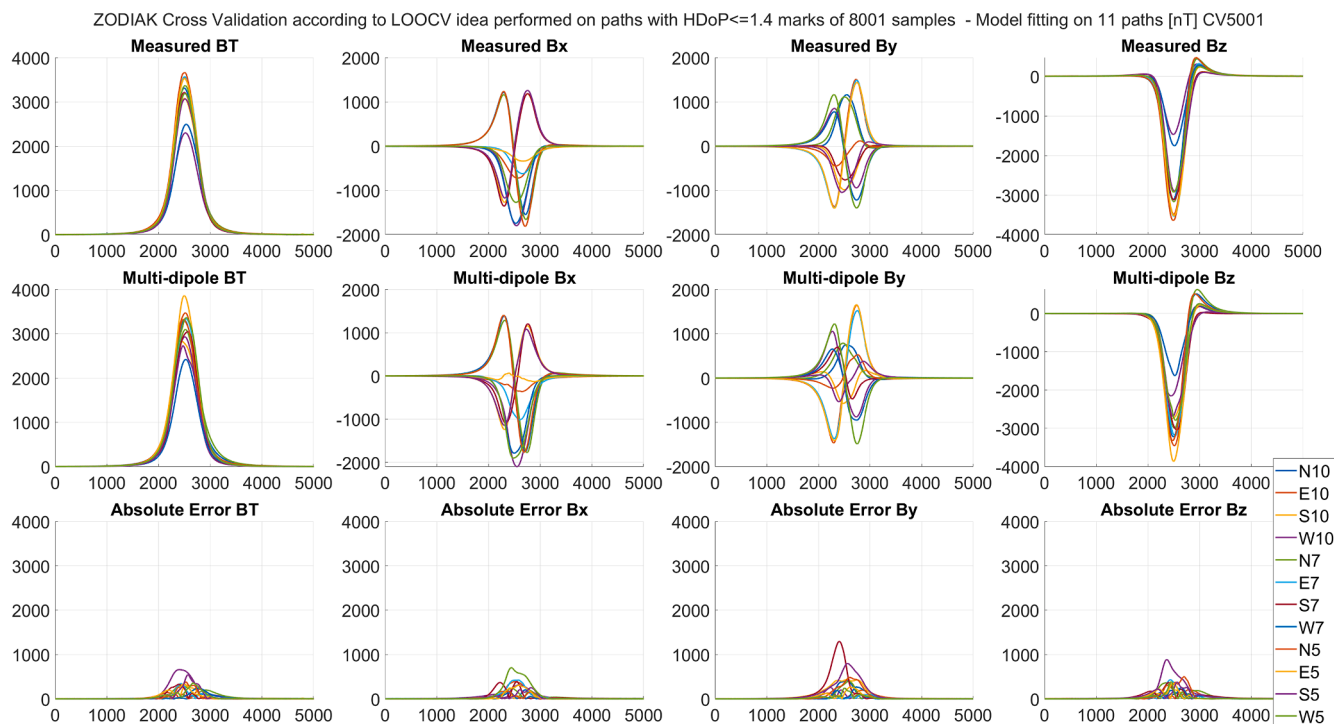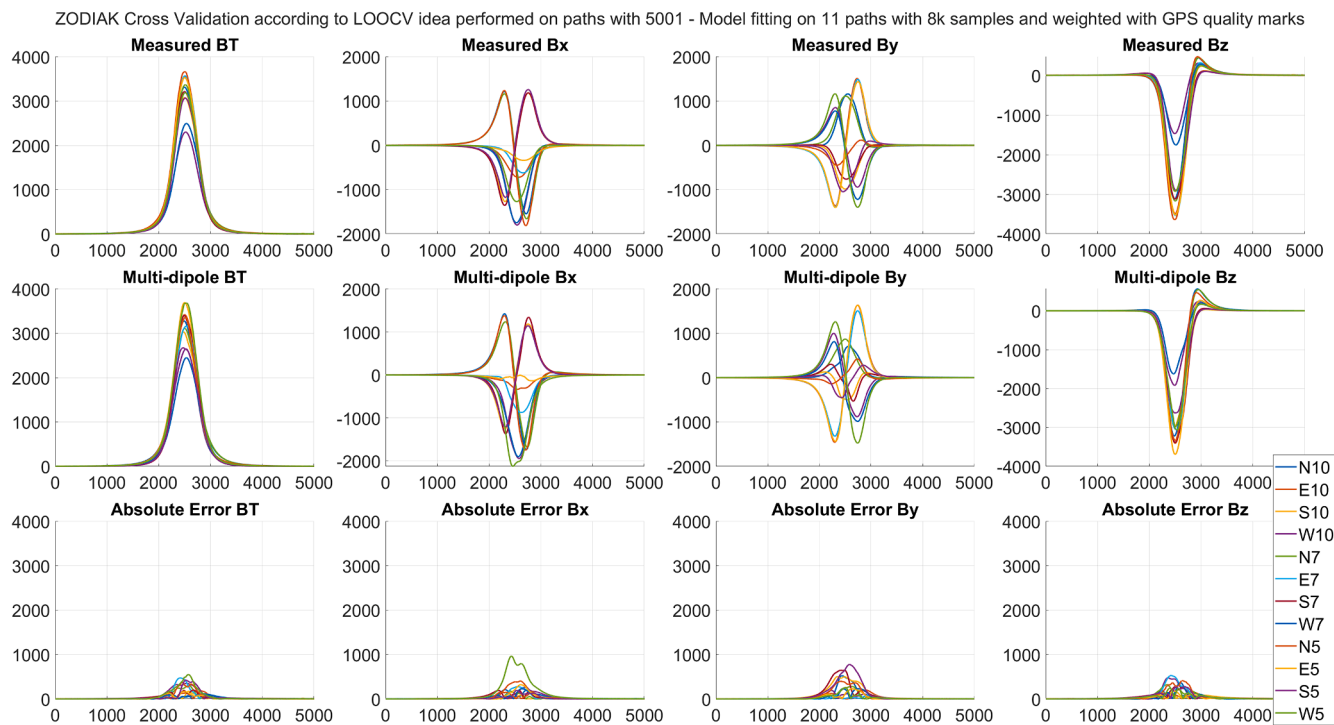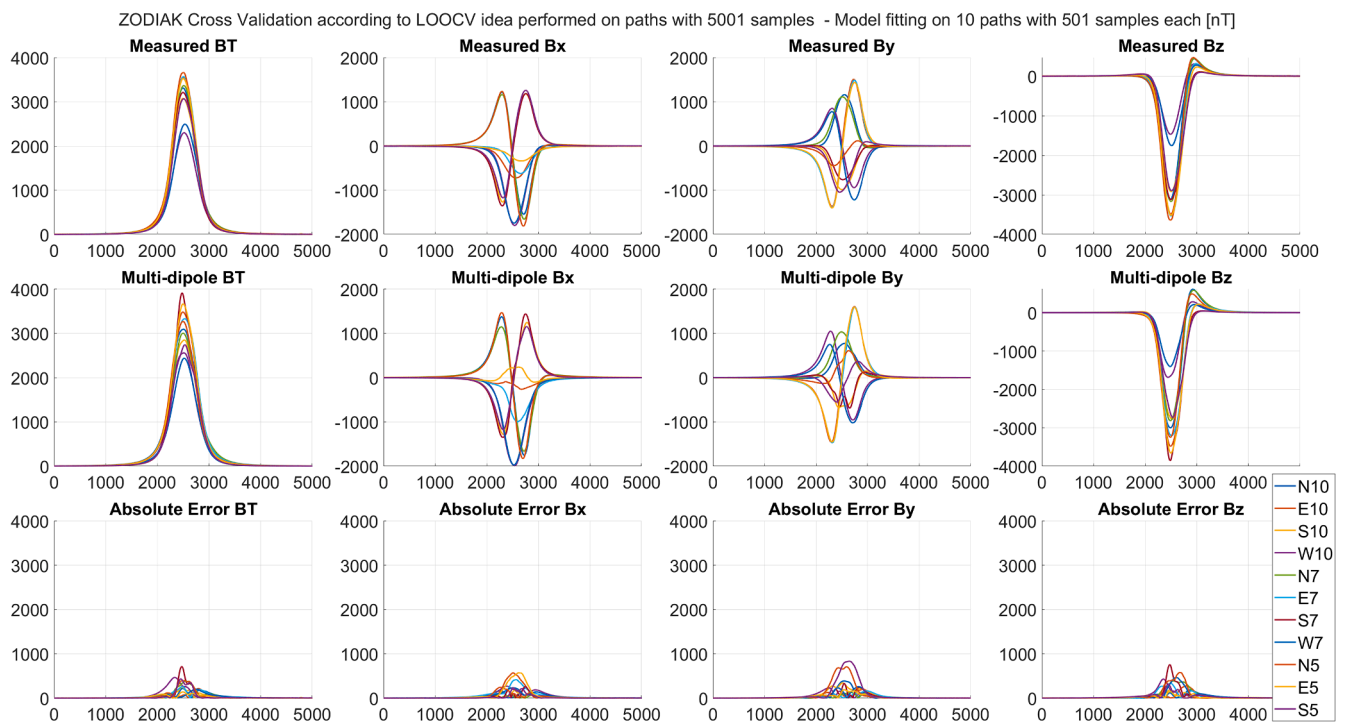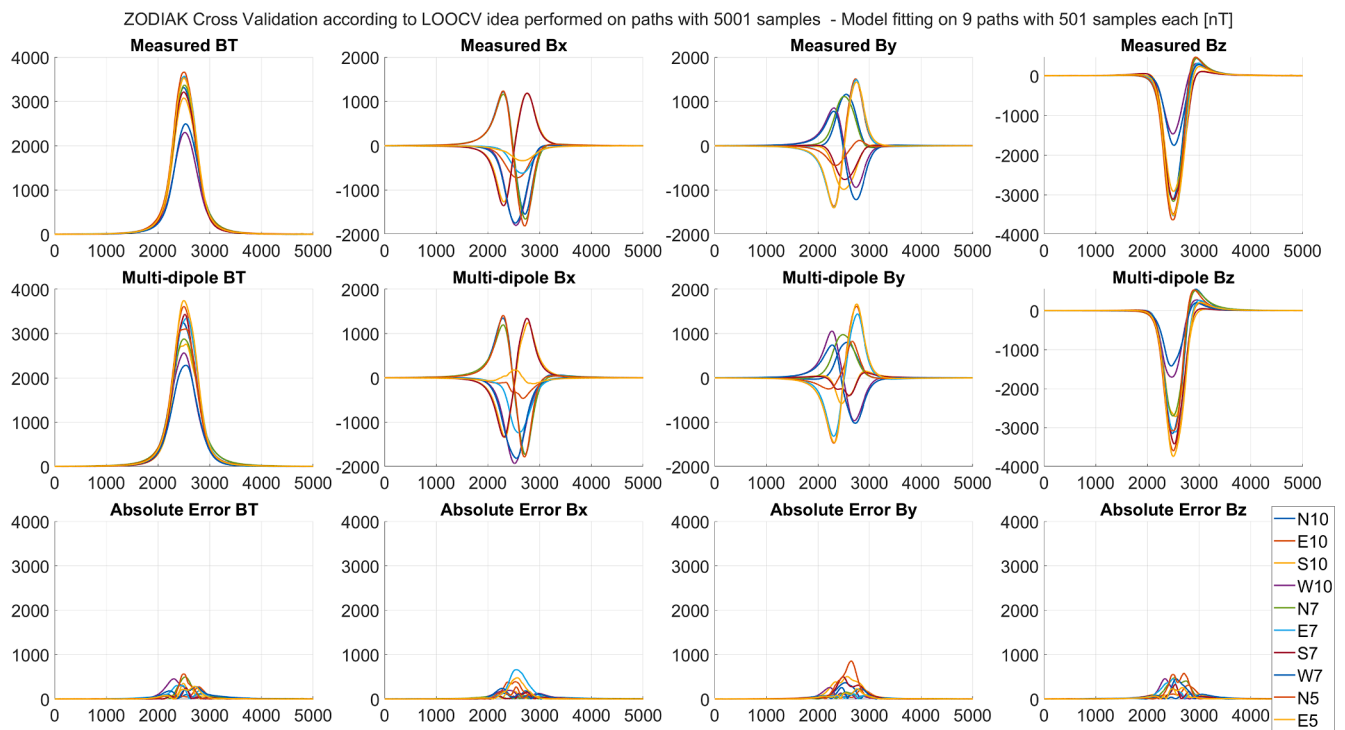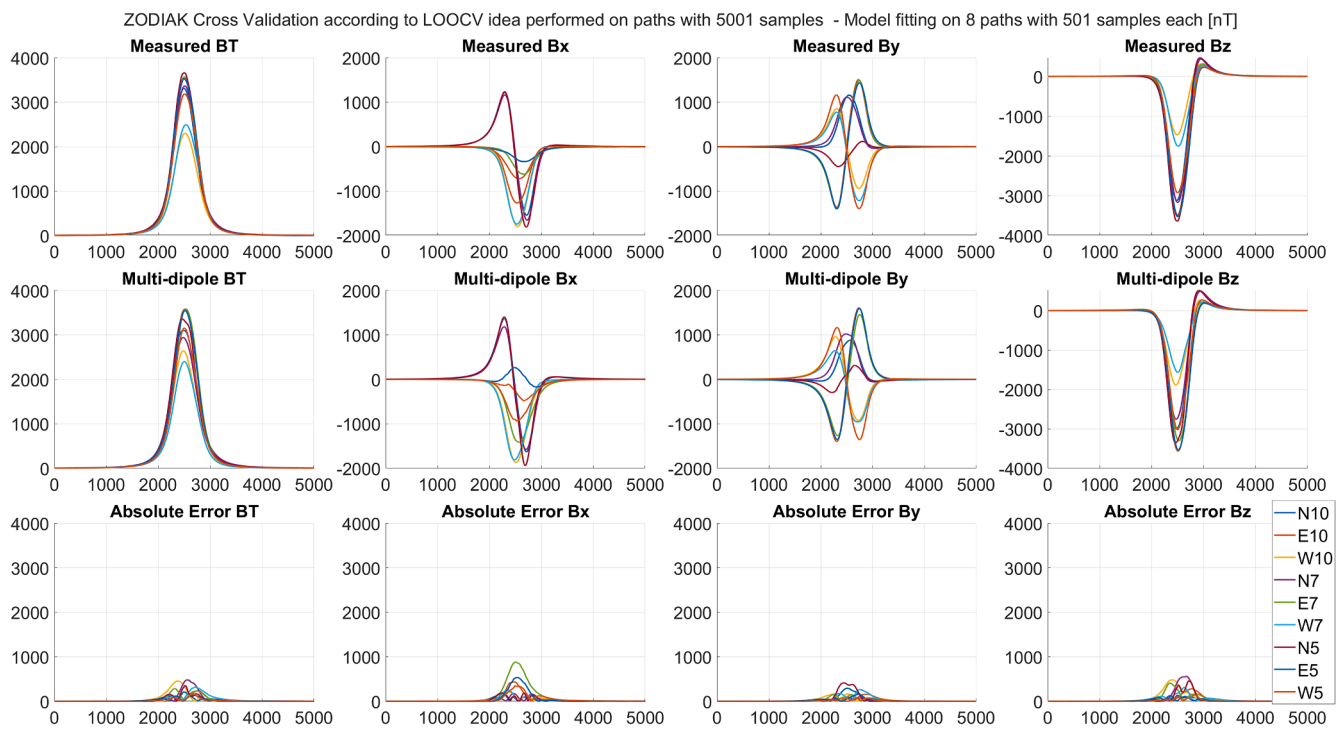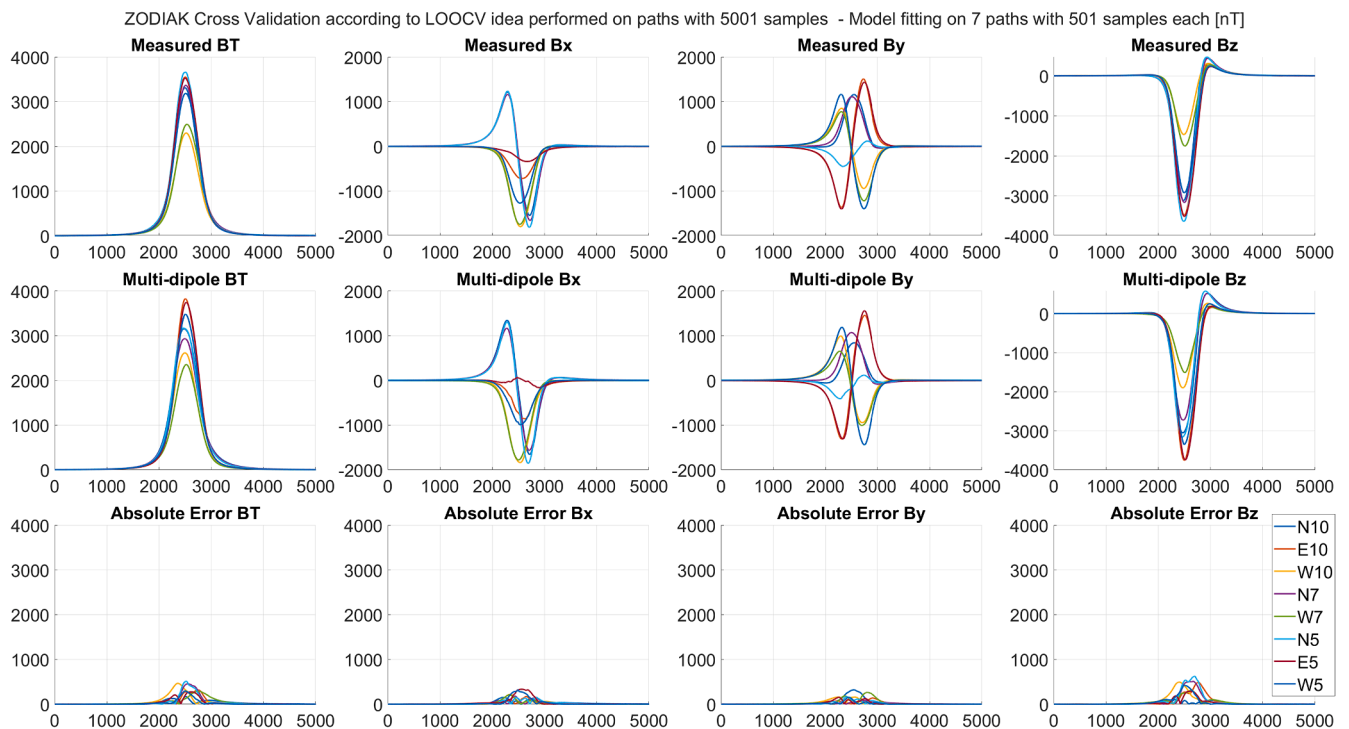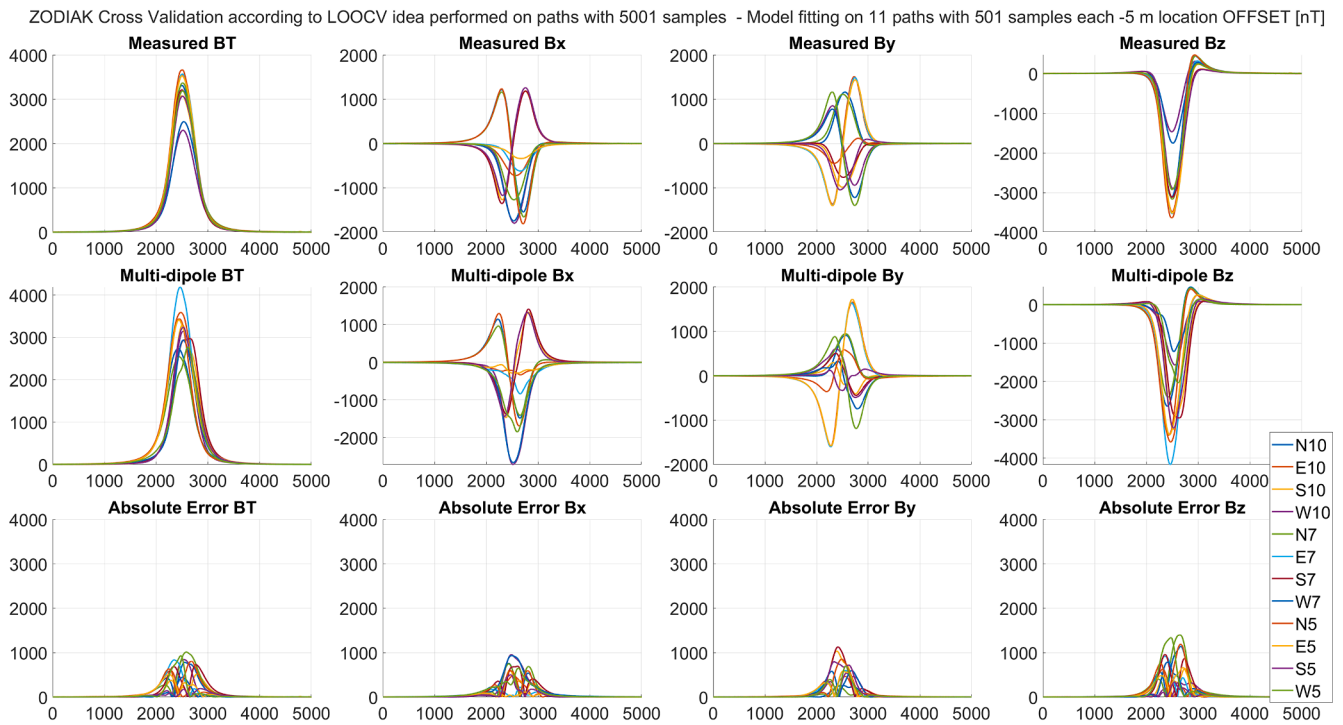

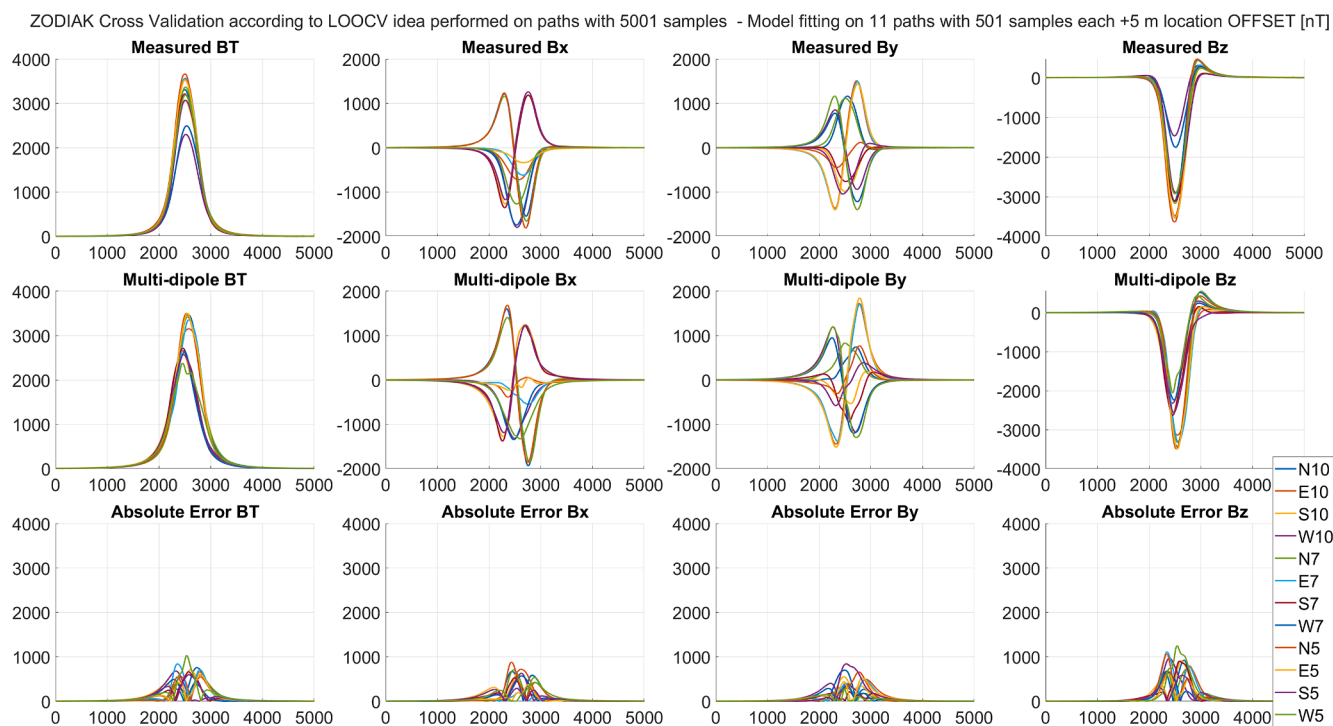**Fig. A4.** Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E5.

Fig. A5. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E6.

Fig. A6. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E9.

Fig. A7. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E10.

Fig. A8. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E11.

Fig. A9. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E12.

Fig. A10. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E13.

Fig. A11. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E14.

Fig. A12. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT for each run of Experiment E15.
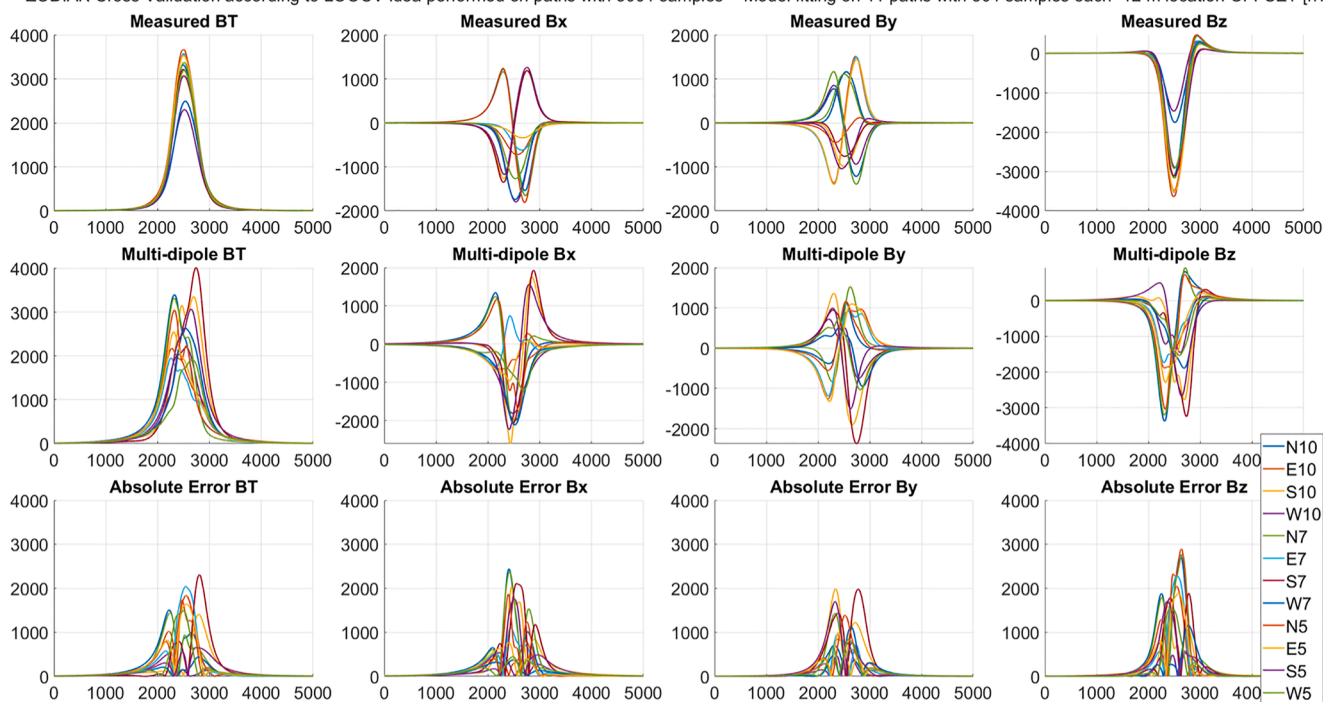
Fig. A13. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT (experiment related to the propagation of the error in position: −5 m).
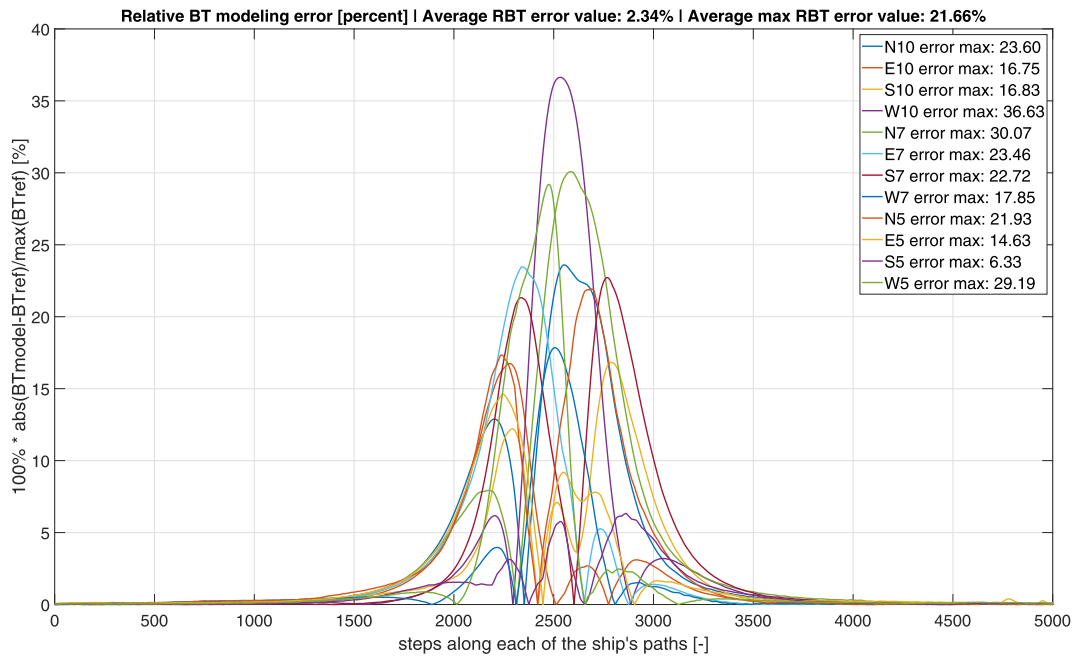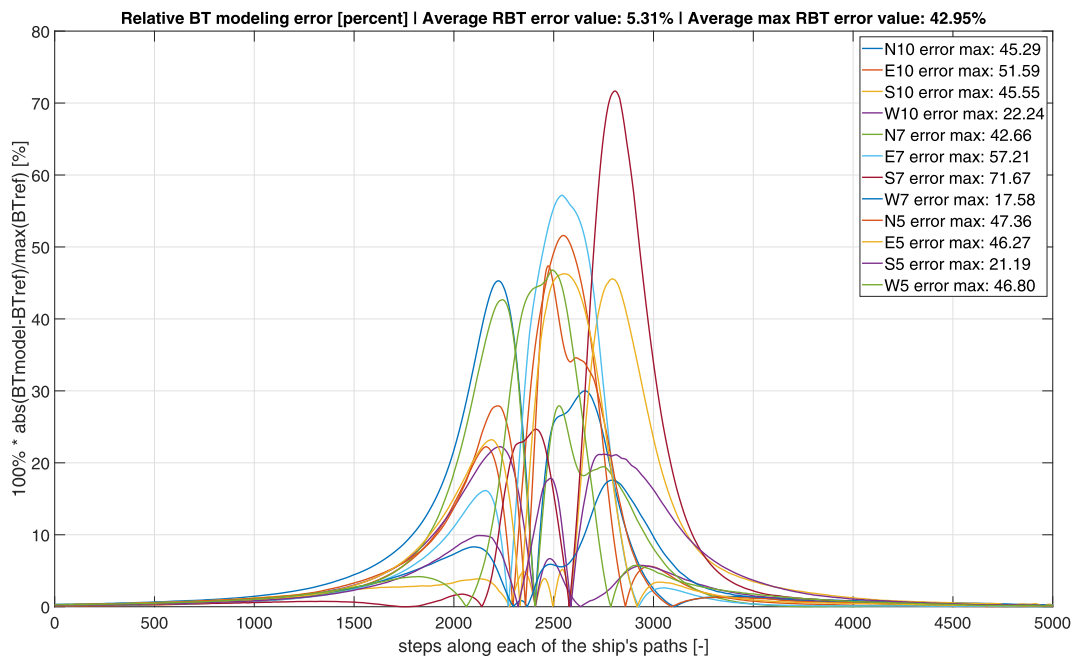

Fig. A14. Graphical summary of relative BT error distribution (experiment related to the propagation of the error in position: −5 m).
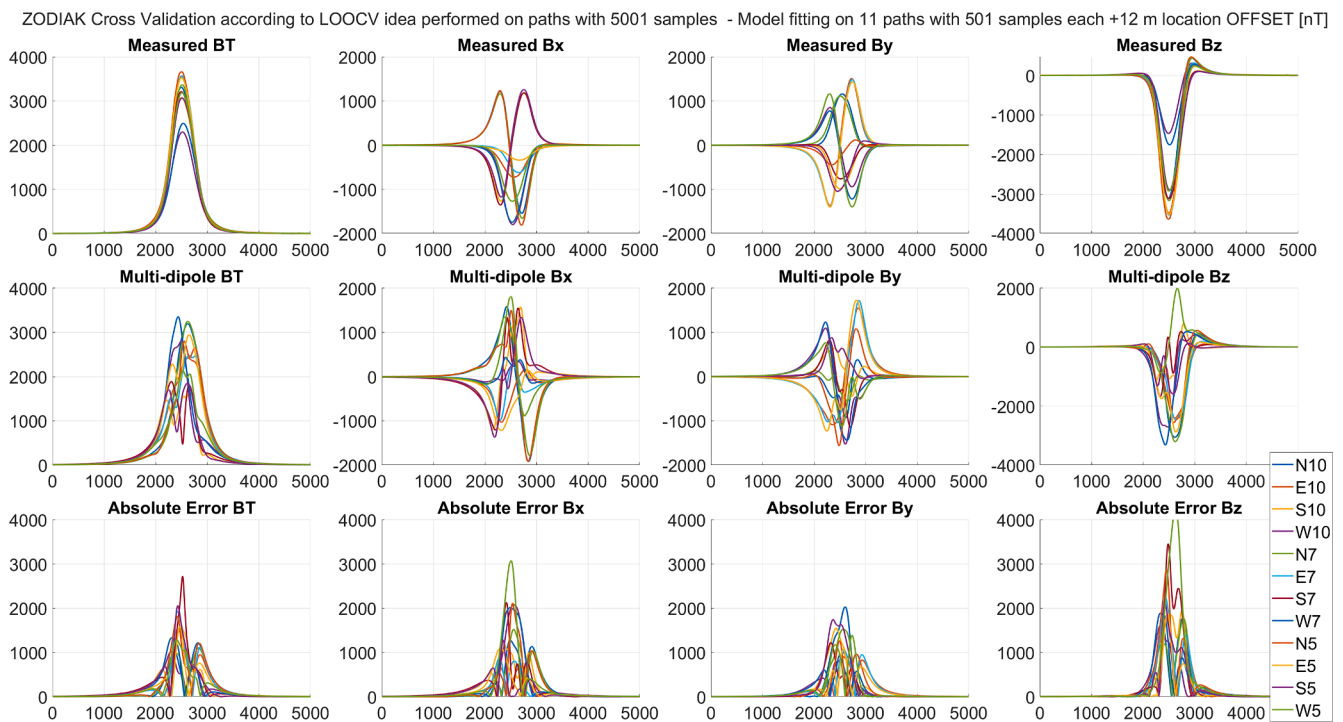
Fig. A15. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT (experiment related to the propagation of the error in position: +5 m).
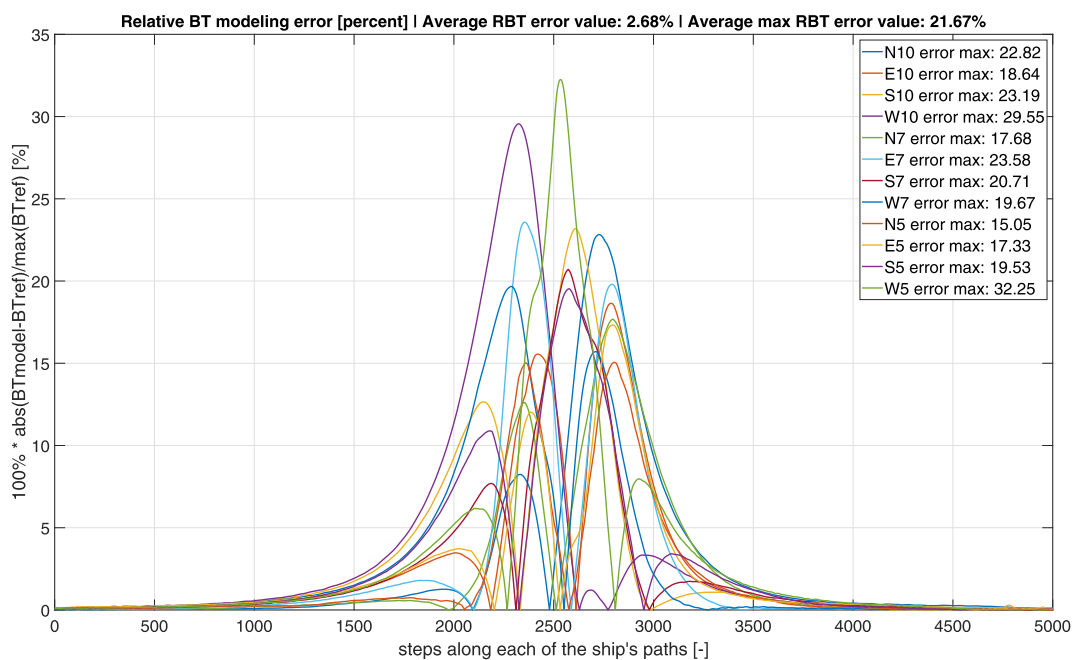
Fig. A16. Graphical summary of relative BT error distribution (experiment related to the propagation of the error in position:+-5 m).

Fig. A17. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT (experiment related to the propagation of the error in position: −12 m).
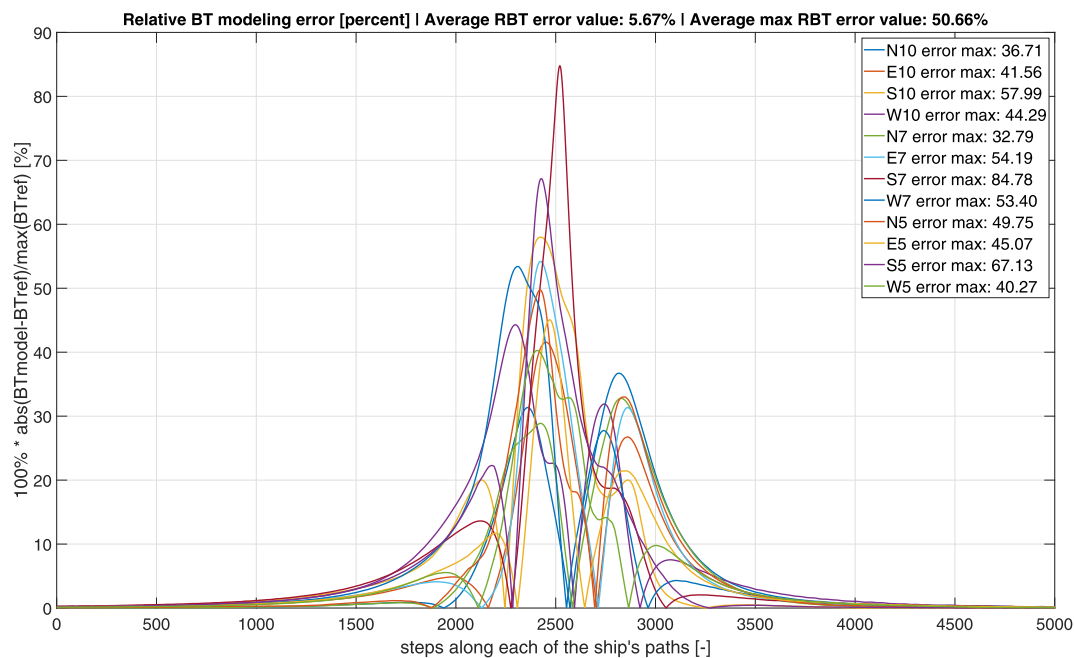

Fig. A18. Graphical summary of relative BT error distribution (experiment related to the propagation of the error in position: −12 m).

ZODIAK Cross Validation according to LOOCV idea performed on paths with 5001 samples - Model fitting on 11 paths with 501 samples each +12 m location OFFSET [nT]

Fig. A19. Cumulative plots of measured, multi-dipole and absolute error results of MFD Bx, By, Bz and BT (experiment related to the propagation of the error in position: +12 m).

Fig. A20. Graphical summary of relative BT error distribution (experiment related to the propagation of the error in position: +12 m).

References

[1] J.M. Mou, P.F. Chen, Y.X. He, T.L. Yip, W.H. Li, J. Tang, H.Z. Zhang, Vessel traffic safety in busy waterways: A case study of accidents in western shenzhen port, Accid. Anal. Prev. 123 (2019) 461–468, https://doi.org/10.1016/j.aap.2016.07.037.

[2] L. Zhang, H. Wang, Q. Meng, Big Data-Based Estimation for Ship Safety Distance Distribution in Port Waters, Transp. Res. Rec. J. Transp. Res. Board 2479 (1) (2015) 16–24, https://doi.org/10.3141/2479-03.

[3] O.V. Smirnova, Situation Awareness for Navigation Safety Control, TransNav Int. J. Mar. Navig. Saf. Sea Transp. 12 (2) (2018), https://doi.org/10.12716/1001.12.02.20.

[4] J.A. Kuma, Vessel traffic service as a maritime security tool: vessel traffic management information systems (VTMIS) in Ghana, undefined, 2015, Accessed: Jun. 06, 2021. [Online]. Available: /paper/Vessel-traffic-service-as-a-maritime-security-tool%3A-Kuma/a6f4009477b63dd28484c463f5705e32f1a557f2.

[5] S. Guze, M. Ledóchowski, Ship Traffic and Port Operation Information Critical Infrastructure Network, J. Pol. Saf. Reliab. Assoc. 7 (2) (2016). Accessed: Jun. 06, 2021. [Online]. Available: http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.baztech-4fd3bf05-ca89-4872-a332-3978a60ca456.

[6] G. Ioannidis, Identification of a Ship or Submarine from its Magnetic Signature, IEEE Trans. Aerosp. Electron. Syst. AES-13 (3) (1977) 327–329, https://doi.org/10.1109/TAES.1977.308404.

[7] T.R. Clem, D.J. Overway, J.W. Purpura, J.T. Bono, Magnetic detection of underwater targets in very shallow water for searches at high speeds, in: MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No.01CH37295), vol. 1, Nov. 2001, pp. 50–58. http://dx.doi.10.1109/OCEANS.2001.968670.

[8] D.V. Lukyanenko, A.G. Yagola, N.A. Evdokimova, Application of inversion methods in solving ill-posed problems for magnetic parameter identification of steel hull vessel 18(9) (2011) 1013–1029. http://dx.doi.org/10.1515/jiip.2011.018.

[9] J.J. Holmes, Reduction of a Ship's Magnetic Field Signatures, Morgan & Claypool Publishers, 2008.

[10] J.-O. Hall, H. Claésson, J. Kjäll, G. Ljungdahl, Decomposition of Ferromagnetic Signature Into Induced and Permanent Components, IEEE Trans. Magn. 56 (2) (2020) 1–6, https://doi.org/10.1109/TMAG.2019.2953860.

[11] G. Rosu, G. Samoilescu, O. Baltag, A. Bordianu, A. Ciuculin, Assessment of the ellipsoidal shell model for ship magnetic signature, in: in *2015 9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, May, 2015, pp. 413–416, https://doi.org/10.1109/ATEE.2015.7133839.

[12] P. Jankowski, M. Wołoszyn, 'Applying of thin plate boundary condition in analysis of ship's magnetic field', *COMPEL- Int*, J. Comput. Math. Electr. Electron. Eng. 37 (5) (2018) 1609–1617, https://doi.org/10.1108/compel-01-2018-0032.

[13] J.J. Holmes, Modeling a Ship's Ferromagnetic Signatures, Synth. Lect. Comput. Electromagn. 2 (1) (Jan. 2007) 1–75, https://doi.org/10.2200/S00092ED1V01Y200706CEM016.

[14] H. Nain, M. Isa, M.M. Muhammad, N. Hassanuddin, M. Yati, I. Nor, Management of Naval Vessel's Electromagnetic Signatures: A Review of Sources and Countermeasures, Def. T Tech. Bull. 6 (Nov. 2013) 93–110.

[15] F. Le Dorze, J.P. Bongiraud, J.L. Coulomb, P. Labie, X. Brunotte, Modeling of degaussing coils effects in ships by the method of reduced scalar potential jump, IEEE Trans. Magn. 34 (5) (Sep. 1998) 2477–2480, https://doi.org/10.1109/20.717570.

[16] Statki Urzedu Morskiego w Gdyni, 'Statki Urzedu Morskiego w Gdyni - Zodiak (in polish)', Statki Urzedu Morskiego w Gdyni. Accessed: Jun. 05, 2021. [Online]. Available: https://bazagdansk.pl/index.php?news&nid=35.

[17] Polski Rejestr Statków S. A. (PRS), 'Register Of Ships 2021', Register Of Sea-Going Ships 2021, 2021, pp. 1–190. [Online]. Available: https://www.prs.pl/wydawnictwa/rejestry-statkow-jachtow-lodzi.

[18] I. Gloza, S.J. Malinowski, B. Marchalewski, Ranges and equipment for the measurement of the ship's underwater signatures, Hydroacoustics 15 (2012) 39–48.

[19] B. Armstrong, et al., Field measurement of surface ship magnetic signature using multiple AUVs, in: OCEANS 2009, Oct. 2009, pp. 1–9. http://dx.doi.10.23919/OCEANS.2009.5422197.

[20] K. Jakubiuk, P. Zimny, M. Wołoszyn, 'Model of ship's magnetic signature', Poznan Univ. Technol. Acad, J. Electr. Eng. 65 (2011) 7–14.

[21] K. Jakubiuk, P. Zimny, M. Wołszyn, Multidipoles model of ship's magnetic field, Int. J. Appl. Electromagn. Mech. 39 (1-4) (2012) 183–188.

[22] J. Tarnawski, A. Cichocki, T.A. Rutkowski, K. Buszman, M. Woloszyn, Improving the Quality of Magnetic Signature Reproduction by Increasing Flexibility of Multi-Dipole Model Structure and Enriching Measurement Information, IEEE Access 8 (2020) 190448–190462, https://doi.org/10.1109/ACCESS.2020.3031740.

[23] J.E. McFee, Y. Das, Fast Nonrecursive Method for Estimating Location and Dipole Moment Components of a Static Magnetic Dipole, IEEE Trans. Geosci. Remote Sens. GE-24 (5) (1986) 663–673, https://doi.org/10.1109/TGRS.1986.289612.

[24] J. Wang, Y. Shen, R. Zhao, C. Zhou, J. Gao, Estimation of dipole magnetic moment orientation based on magnetic signature waveform analysis by a magnetic sensor, J. Magn. Magn. Mater. 505 (2020) 166761, https://doi.org/10.1016/j.jmmm.2020.166761.

[25] G. Jeung, C.-S. Yang, H.-J. Chung, S.-H. Lee, D.-H. Kim, Magnetic Dipole Modeling Combined With Material Sensitivity Analysis for Solving an Inverse Problem of Thin Ferromagnetic Sheet, IEEE Trans. Magn. 45 (10) (2009) 4169–4172, https://doi.org/10.1109/TMAG.2009.2021853.

[26] X. Brunotte, G. Meunier, J. Bongiraud, Ship Magnetizations Modelling by Finite Element Method, in: Digest of the Fifth Biennial IEEE Conference on Electromagnetic Field Computation, Aug. 1992, pp. TOD2–TOD2. http://dx.doi.10.1109/CEFC.1992.720698.

[27] I. Gloza, K. Buszman, The Multi-Influence Passive Module For Underwater Environment Monitoring, Hydroacoustics 14 (2011) 47–54.

[28] Valeport miniSVS datasheet, 'miniSVS Sound Velocity Sensor - Valeport'. [Online]. Available: https://www.valeport.co.uk/products/minisvs-sound-velocity-sensor/.

[29] Honeywell HMR3300 datasheet, 'Integrated Compassing Solutions'. [Online]. Available: https://aerospace.honeywell.com/en/learn/products/sensors/integrated-compassing-solutions.

[30] Bartington Grad 03 datasheet, 'Grad-03 Three-Axis Fluxgate Gradiometers'. [Online]. Available: http://www.bartington.kr/presentation/grad-03-three-axis-gradiometer/index.html.

[31] National instruments NI9220 datasheet, 'NI-9220 - NI'. [Online]. Available: https://www.ni.com/en/pl-pl/support/model.ni-9220.html.

[32] 'NI cRIO-9068'. Accessed: Aug. 04, 2021. [Online]. Available: https://www.ni.com/pdf/manuals/376007a_02.pdf.

[33] A. Tesei, et al., 'Real-time underwater positioning and navigation of an AUV in deep waters', 2018. http://dx.doi.10.1109/OCEANSKOBE.2018.8558876.

[34] A. Tomczak, Modern Methods Of Underwater Positioning Applied In Subsea Mining, Gór. Geoinżynieria R. 35 (2011) 381–394.

[35] A. Oxley, Uncertainties in GPS positioning: A mathematical discourse, first ed. Academic Press, 2017, pp. 180. [Online]. Available: https://www.elsevier.com/books/uncertainties-in-gps-positioning/oxley/978-0-12-809594-2.

[36] F. Rovira-Más, Vulnerability of GPS to provide vehicle states in real time, in: IFAC Proceedings Volumes (IFAC-PapersOnline), vol. 46(18 PART 1), 2013, pp. 207–212. http://dx.doi.10.3182/20130828-2-SF-3019.00002.

[37] K. Shi, M. Xu, H. Jin, T. Qiao, X. Yang, N. Zheng, J. Xu, K.-K. Raymond Choo, A novel file carving algorithm for National Marine Electronics Association (NMEA) logs in GPS forensics, Digit. Investig. 23 (2017) 11–21, https://doi.org/10.1016/j.diin.2017.08.004.

[38] D.K. Schrader, B.C. Min, E.T. Matson, J. Eric Dietz, Real-time averaging of position data from multiple GPS receivers, Meas. J. Int. Meas. Confed. 90 (2016) 329–337, https://doi.org/10.1016/j.measurement.2016.04.028.

[39] P. Hordyniec, J. Bosy, W. Rohm, Assessment of errors in Precipitable Water data derived from global navigation satellite system observations, J. Atmospheric Sol.-Terr. Phys. 129 (2015) 69–77, https://doi.org/10.1016/j.jastp.2015.04.012.

[40] B. Calka, E. Bielecka, M. Figurski, Spatial pattern of ASG-EUPOS sites, Open Geosci. 9 (1) (2017) 613–621, https://doi.org/10.1515/geo-2017-0046.

[41] H. Oda, E. Okuyama, E. Shimizu, Ship to ship operations monitoring system using high accuracy DGPS, in: IFAC Proceedings Volumes (IFAC-PapersOnline), vol. 43 (20), 2010, pp. 62–66. http://dx.doi.10.3182/20100915-3-DE-3008.00021.

[42] L.S. Monteiro, T. Moore, C. Hill, What is the accuracy of DGPS? J. Navig. 58 (2) (2005) 207–225, https://doi.org/10.1017/S037346330500322X.

[43] Y. Jiang, Y. Cui, B. Li, Study on differential GPS (DGPS): Method for reducing the measurement error of CNNS, Adv. Mater. Res. 482–484 (2012) 75–80, https://doi.org/10.4028/www.scientific.net/AMR.482-484.75.

[44] A. Saracoglu, D.U. Sanli, Effect of meteorological seasons on the accuracy of GPS positioning, Meas. J. Int. Meas. Confed. 152 (2020) 107301, https://doi.org/10.1016/j.measurement.2019.107301.

[45] Z. Slanina, V. Kasik, K. Musil, GPS synchronisation for FPGA devices, in: IFAC Proceedings Volumes (IFAC-PapersOnline), vol. 11(PART 1), 2012, pp. 337–340. http://dx.doi.10.3182/20120523-3-cz-3015.00064.

[46] Ruyun Tian, Junjie Zhang, Shuai Zhang, Longxu Wang, Hongyuan Yang, Yuda Chen, Yuanjie Jiang, Jun Lin, Linhang Zhang, A high-precision energy-efficient GPS time-sync method for high-density seismic surveys, Appl. Sci. Switz. 10 (11) (2020) 3768, https://doi.org/10.3390/app10113768.

[47] Satellite Compas Furuno SC-50 technical documentation, 'SC50'. [Online]. Available: https://www.furunousa.com/en/support/SC50.

[48] Serial Port Server Moxa NPort 5400 technical documentation, 'NPort 5400 Series - General Device Servers | MOXA'. [Online]. Available: https://www.moxa.com/en/products/industrial-edge-connectivity/serial-device-servers/general-device-servers/nport-5400-series.

[49] C. Adrados, I. Girard, J.P. Gendner, G. Janeau, Global Positioning System (GPS) location accuracy improvement due to selective availability removal, Comptes Rendus - Biol. 325 (2) (2002) 165–170, https://doi.org/10.1016/S1631-0691(02)01414-2.

[50] J. Haifu, L. Ma, G. Ai, M. Wang, The Distributions of HDOP and VDOP in GNSS and a Corresponding New Algorithm of Fast Selecting Satellites, 2012. http://dx.doi.10.1007/978-3-642-29175-3_37.

[51] R.B. Langley, Dilution of Precision, GPS World 10 (May) (1999) 52–59.

[52] J.J. Spilker, Global Positioning System: Theory and Applications, Volume I Chapter 5: Satellite Constellation and Geometric Dilution of Precision | Engineering360. American Institute of Aeronautics and Astronautics, Inc. Accessed: Aug. 05, 2021. [Online]. Available: https://www.globalspec.com/reference/35484/203279/chapter-5-satellite-constellation-and-geometric-dilution-of-precision.

[53] E.R. Swanson, Geometric Dilution of Precision, Navigation 25 (4) (1978) 425–429, https://doi.org/10.1002/j.2161-4296.1978.tb01345.x.

[54] A. Kaufman, D. Alekseev, M. Oristaglio, Principles of Electromagnetic Methods in Surface Geophysics. Newnes, 2014.

[55] 'Simulia Opera Simulation Software'. Accessed: Aug. 04, 2021. [Online]. Available: https://www.3ds.com/products-services/simulia/products/opera/.

[56] Mathworks, 'Matlab Mapping Toolbox'. Mathworks. Accessed: Jun. 06, 2021. [Online]. Available: https://www.mathworks.com/products/mapping.html.

[57] K. Dokkum, H.T. Katen, K. Koomen, J. Pinkster, Ship Stability, sixth ed., DOKMAR Maritime Publishers VS, Vlissingen, 2018 [Online]. Available: https://www.dokmar.com/previews/ship-stability/#fb0=5.

[58] H.D. Curtis, *Orbital Mechanics for Engineering Students*, vol, Chapter 9. Elsevier Ltd (2013) 751, https://doi.org/10.1016/C2011-0-69685-1.

[59] B. Ghojogh, M. Crowley, The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial, ArXiv190512787 Cs Stat, May 2019, Accessed: Jun. 06, 2021. [Online]. Available: http://arxiv.org/abs/1905.12787.

[60] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning: with Applications in R, first ed. 2013, Corr. 7th printing 2017 ed., Springer, New York, 2013.

[61] Thomas F. Coleman, Yuying Li, An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds, SIAM J Optim 6 (2) (1996) 418–445, https://doi.org/10.1137/0806023.

[62] D.W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, J. Soc. Ind. Appl. Math. 11 (2) (Jun. 1963) 431–441, https://doi.org/10.1137/0111030.

[63] J.F. Kenney, Mathematics of statistics, third ed., Van Nostrand, 1964.

[64] Z. Wang, A.C. Bovik, Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures, IEEE Signal Process. Mag. 26 (1) (Jan. 2009) 98–117, https://doi.org/10.1109/MSP.2008.930649.

[65] D.T. Larose, Data Mining Methods and Models, first ed., Wiley-IEEE Press, Hoboken, NJ, 2006.

[66] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, New York, NY, 2016.

[67] L. Breiman, Statistical Modeling: The Two Cultures, Stat. Sci. 16 (3) (2001) 199–215.

[68] D. Barber, Bayesian Reasoning and Machine Learning, first ed., Cambridge University Press, Cambridge; New York, 2012.

[69] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, fourth ed., Morgan Kaufmann, Amsterdam, 2016.

[70] Robert Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. Ser. B Methodol. 58 (1) (1996) 267–288.