

# Application of the polynomial adjustment to the aeromagnetic survey of the Spanish Mainland; Requirements and shortcomings

J. Ardizone<sup>1</sup> and M. Herraiz<sup>2</sup>

<sup>1</sup>National Geographic Institute of Spain, c/General Ibáñez Ibero, 3, 28003, Madrid

<sup>2</sup>Department of Geophysics and Meteorology, Facultad de Ciencias Físicas, Universidad Complutense, 28040, Madrid

(Received August 15, 1998; Revised November 24, 1999; Accepted December 20, 1999)

Polynomial surfaces are a very useful tool in order to represent the geomagnetic field over small area of the Earth's surface. Nevertheless, the method has not always been applied with enough precision, nor its limitations are considered in detail. This work analyses the conditions of applicability of this method to data from the aeromagnetic survey of the Spanish mainland accomplished in 1986, and the fulfilment of the necessary conditions to enhance a merely descriptive model to produce another of inferential character. The best statistical adjustment with only descriptive character, valid for the whole Spanish mainland, is a third degree model. The study proves that the conditions of validity for inferential models are met using polynomials of second and third degree in restrictive areas, but not over the whole Iberian Peninsula. In this case, the model of second degree is proposed as the best option because of its simplicity and the extension of its region of validity. To a great extent, the restrictions to enhance the descriptive models are due to the own nature, specially to their geographical dependence, and can also affect the analysis of other kinds of geophysical data.

## 1. Introduction

Sometimes, an airborne magnetic survey not only aims at obtaining maps of the Earth's main magnetic field, but also to look for a mathematical model that allows its description and representation in the surveyed area as well as the determination of the anomalous field. In this way, it is possible to perform a geophysical interpretation of the region of interest. Haines (1990) classifies the analytical procedures to reach such objective, into five main groups: Polynomial Surfaces, Methods of Equivalent Dipolar Sources, Spherical Harmonic Analysis, Rectangular Harmonic Analysis, and Spherical Cap Harmonic Analysis. The last four methods are based upon the Magnetic Field being derived from a potential whereas the first method, which has been chosen for this study, does not require such a condition. This final methodology has been previously applied to the magnetic survey of Norway, Sweden, Finland and Denmark (Haines, 1968; Haines *et al.*, 1970); Canada (Haines and Hannaford, 1972, 1974) and Japan (Tanaka *et al.*, 1986).

A polynomial surface used in this kind of study is a surface that only depends only on the latitude,  $\varphi$ , and the longitude,  $\lambda$ , and it is obtained by fitting 3rd or lower degree polynomials to observations of the magnetic field. Hence, for a generic magnetic component,  $P$ , the model of the polynomial surface is

$$P = p_0 + p_1\lambda + p_2\varphi + p_3\lambda\varphi + p_4\lambda^2 + p_5\varphi^2. \quad (1)$$

Mathematically, it can be considered as a power series truncated in the highest order term. Despite the simplicity of this expression and due to the relations between its components, these models can present geometrical incongruities when more than three individual components of the magnetic field are measured. In these cases the geometrical restrictions can be satisfied by imposing conditions on the coefficients, although the procedure becomes more complicated as coefficients become non linear within the fitting functions. Another problem affecting the use of least squares is the uncertainty associated with the limits of the region of interest, due to the lack of observation points outside it. This fact brings as a consequence that the areas where the adjustments can be considered reliable are smaller than those where the data sampling takes place.

In addition, models of  $\mathbf{B}$  components must satisfy Maxwell equations, i.e., to be consistent from the physical point of view. For an area free from electromagnetic sources, the well known equations:  $\nabla \cdot \mathbf{B} = 0$  and  $\nabla \times \mathbf{B} = 0$  determine the conditions to be met by the coefficients of the models for each component to ensure consistency (Haines, 1967). Even if the curl of  $\mathbf{B}$  is not null, it is possible to study statistically the existence of an inconsistency in the area (Haines, 1968).

The bidimensional character of the polynomial surfaces model does not exclude the possibility of extending the anomalies derived to an altitude different from that chosen in survey. In this way, those details that introduce errors in the measurements are attenuated (upward prolongation), while some anomalies of small amplitude for which the resolution is difficult because of the flight altitude, are stressed (downward prolongation). In both cases it is

possible to match anomaly maps of adjacent regions which have been laid up to different altitudes, so that a unique map describing a wider area and of the same altitude is obtained (Galdeano and Rossignol, 1977)

In this work, we look upon the applicability of the polynomial surfaces model to the data array of the aeromagnetic survey carried out on the Spanish mainland. As only the field intensity is considered, no geometrical incongruities are present. The conditions to be met by data in order that the obtained models are not merely descriptive but have inferential character as well, are highlighted. As it will be proved, this objective raises strong restrictions in the extension of areas in which models can be considered valid. The resultant model is compared to the IGRF85 and to the representation obtained by Torta (1992) applying Spherical Cap Harmonic Analysis (SCHA), after moving them to 1987.0 and 3,000 m of altitude above sea level.

## 2. Data

The dataset used in this analysis consists of a sample of 5,316 points out of the 2,019,690 that make up the measurements obtained during the aeromagnetic flight over the Iberian Peninsula. This flight was performed from September 13th, 1986, to June, 4th, 1987. Measurements of the magnetic field intensity were obtained at a barometric altitude of 3,000 m, except for the mountain

areas for which the altitude was increased up to 500 m over the peak. The horizontal distribution of the measurement points spread over a net consisting of 104 lines of flight in the N-S direction, 10 km apart; 22 lines of control in the E-W direction, 40 km apart, and 40 perimetric lines that closed most of the net. The measurements were performed by a double-resonance Overhauser effect magnetometer with a sensibility of 0.01 nT. The samples were taken every half second, which implies a measurement every 34 m. A detailed description of the aeromagnetic flight and the data processing can be found in Ardizone *et al.* (1989), Socías *et al.* (1991) and Ardizone (1996).

The data used in this study were selected from the total Data File in order to assure that they were representative of the whole survey. They were taken every 10 km along the flight lines and were located between parallels 36°0'3.6" N and 43°48'0.0" N and meridians 3°15'3.6" E and 9°17'13.6" W. The comparison of the cartographic representation of these intensity values (Fig. 1) with the corresponding original map published by the National Geographic Institute of Spain (Ardizone *et al.*, 1989), proves that, in spite of the smoothing introduced by the sampling, the most important details of the survey are preserved. The number of data points may appear to be sparse in absolute terms but it is mathematically sufficient

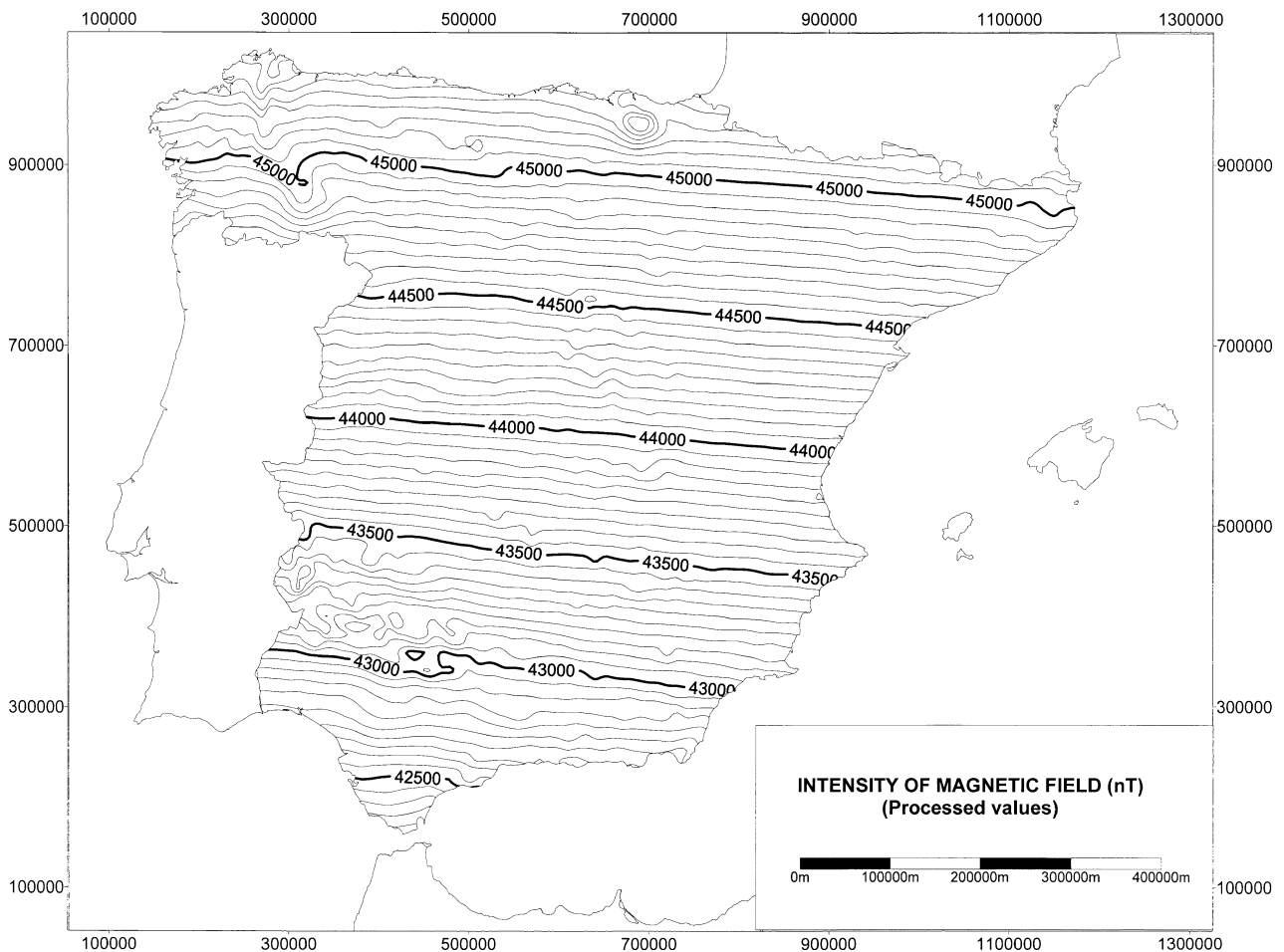


Fig. 1. Intensity of Magnetic Field obtained from processed data. Values are given in nT and contour intervals correspond to 50 nT (Spanish Official Lambert Projection).

and even high if compared to that used to determine the 3rd order polynomial models for Scandinavia. In that case 1,400 points of observation were applied to an extension of 1,500,000 km<sup>2</sup>, more than three times the Iberian Peninsula territory (493,486 km<sup>2</sup>). The distance between two points was 30 km (Haines, 1968).

**3. Model Determination**

According to what has previously been stated, in order to represent the intensity of the earth's magnetic field over the Spanish Peninsular territory, we chose a model whose generic form is:

$$B = d_{00} + d_{10}(\varphi - \bar{\varphi}) + d_{01}(\lambda - \bar{\lambda}) + d_{11}(\varphi - \bar{\varphi})(\lambda - \bar{\lambda}) + d_{20}(\varphi - \bar{\varphi})^2 + d_{02}(\lambda - \bar{\lambda})^2 + \dots + \varepsilon_j \quad (2)$$

In this expression ( $\bar{\varphi}$ ,  $\bar{\lambda}$ ) are the geographical coordinates of the gravity centre of the sample

$$\begin{aligned} \bar{\varphi} &= 40.427^\circ \\ \bar{\lambda} &= -3.537^\circ \end{aligned} \quad (3)$$

and  $\varepsilon_j$  are the perturbations or errors introduced when

using any polynomials approach instead of the real explicit form  $B = B(\varphi, \lambda)$ , which is actually unknown.

Through simple changes like:

$$w_{ij} = (\varphi - 40.427)^i (\lambda + 3.537)^j, \quad (4)$$

expression (2) is reduced to a first degree polynomial:

$$B = d_{00} + d_{10}w_{10} + d_{01}w_{01} + d_{11}w_{11} + d_{20}w_{20} + d_{02}w_{02} + \dots + \varepsilon_j \quad (5)$$

To determine the coefficients, we apply the least squares method to the set of data described in the previous topic. Nevertheless, because this set is a sample of the measurements, this method only provides estimates of such coefficients that we name as  $\hat{d}_{ij}$ . Therefore, the models (5) will be now:

$$B = \hat{d}_{00} + \hat{d}_{10}w_{10} + \hat{d}_{01}w_{01} + \hat{d}_{11}w_{11} + \hat{d}_{20}w_{20} + \hat{d}_{02}w_{02} + \dots + \varepsilon_j = \hat{B} + e_j \quad (6)$$

where the  $e_j$  term represents the residual, whose values are an estimation of the errors.

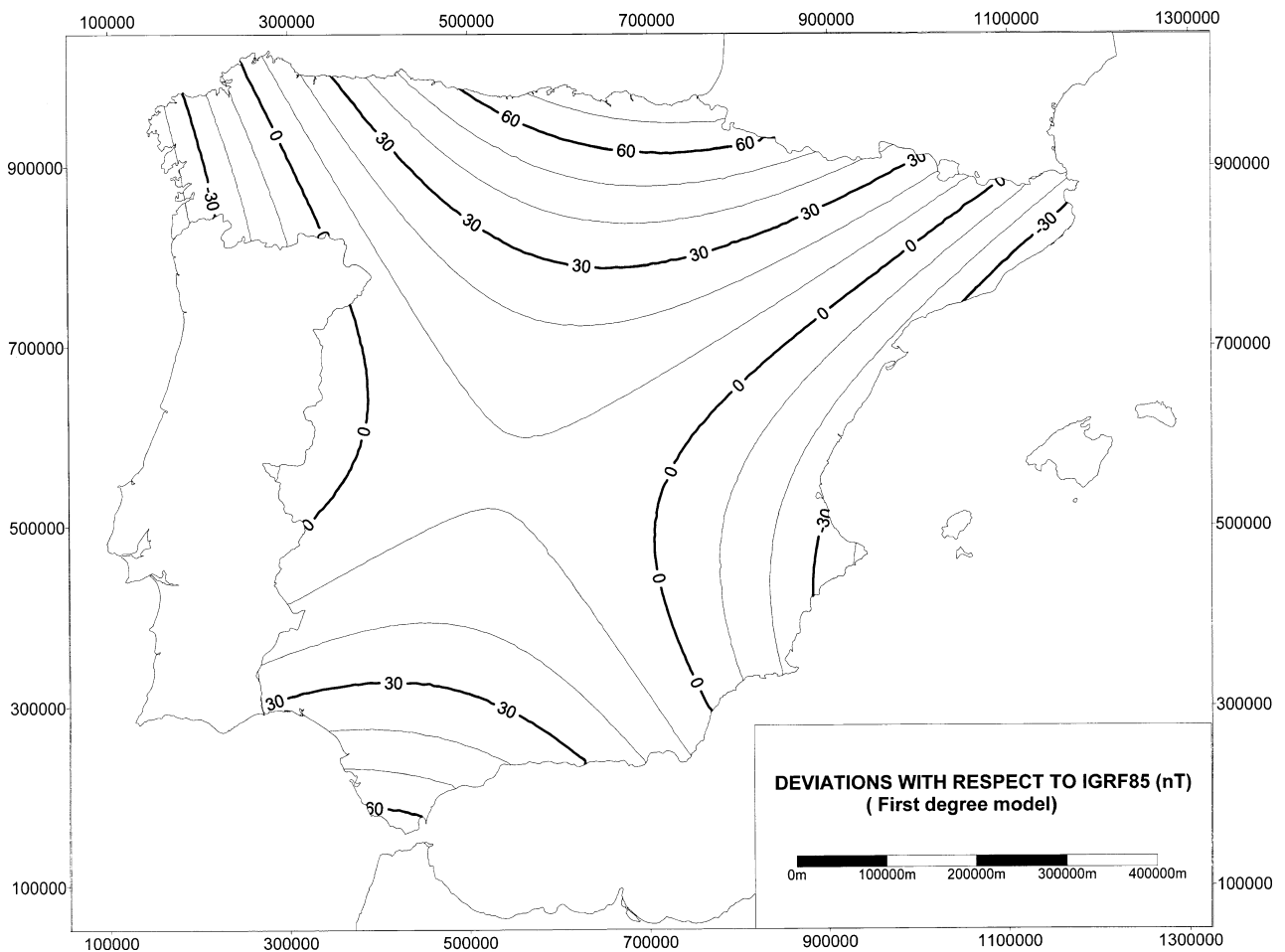


Fig. 2. Map of differences between the first degree polynomial model and the IGRF85 moved to the reference date (1987.0). Values are given in nT and contour intervals correspond to 10 nT (Spanish Official Lambert Projection).

Table 1. Estimates of the first, second and third degree coefficients.

Polynomial degree	Estimation of coefficients		Standard error of coefficients
	Estimate	Standard error	
1	$\hat{a}_{00}$	44149.600	0.446
	$\hat{a}_{10}$	401.536	0.228
	$\hat{a}_{01}$	21.791	0.172
2	$\hat{a}_{00}$	44151.000	0.464
	$\hat{a}_{10}$	395.362	0.153
	$\hat{a}_{01}$	21.602	0.116
	$\hat{a}_{11}$	-2.914	0.061
	$\hat{a}_{20}$	-4.216	0.078
	$\hat{a}_{02}$	2.346	0.036
3	$\hat{a}_{00}$	44151.812	0.522
	$\hat{a}_{10}$	396.123	0.366
	$\hat{a}_{01}$	21.918	0.233
	$\hat{a}_{11}$	-2.938	0.076
	$\hat{a}_{20}$	-4.425	0.093
	$\hat{a}_{02}$	2.279	0.053
	$\hat{a}_{21}$	-0.097	0.040
	$\hat{a}_{12}$	0.023	0.028
	$\hat{a}_{30}$	-0.124	0.043
	$\hat{a}_{03}$	-0.0002	0.012

The results of the least squares procedure applied to first, second and third degrees models are summarised in Table 1.

To estimate the quality of the adjustment, we use the standard error of regression ( $\hat{S}_R$ ) and the coefficient of determination ( $R^2$ ).  $\hat{S}_R$  equals 32.42, 19.66 and 19.63 for the first, second and third degree polynomials, respectively.  $R^2$  is 0.9983 for the first model and 0.9994 for the two other cases. These results indicate that if we consider only the group of observations fitted in the sampling, the accuracy of the adjustment increases more when switching from the first to the second degree models than when changing from the second to the third.

To see the differences and similarities between the three models, Figs. 2, 3, and 4 respectively show the deviations between them and the IGRF85 moved to the reference date of data, 1987.0. As it can be seen, the two last models have a very similar behaviour, clearly different from the first degree.

#### 4. Inferential Analysis of the Model

The polynomial models determined previously fit in the data sample and have a descriptive character. This is already a useful result but if we want them to possess an inferential character, they have to fulfil the seven classic

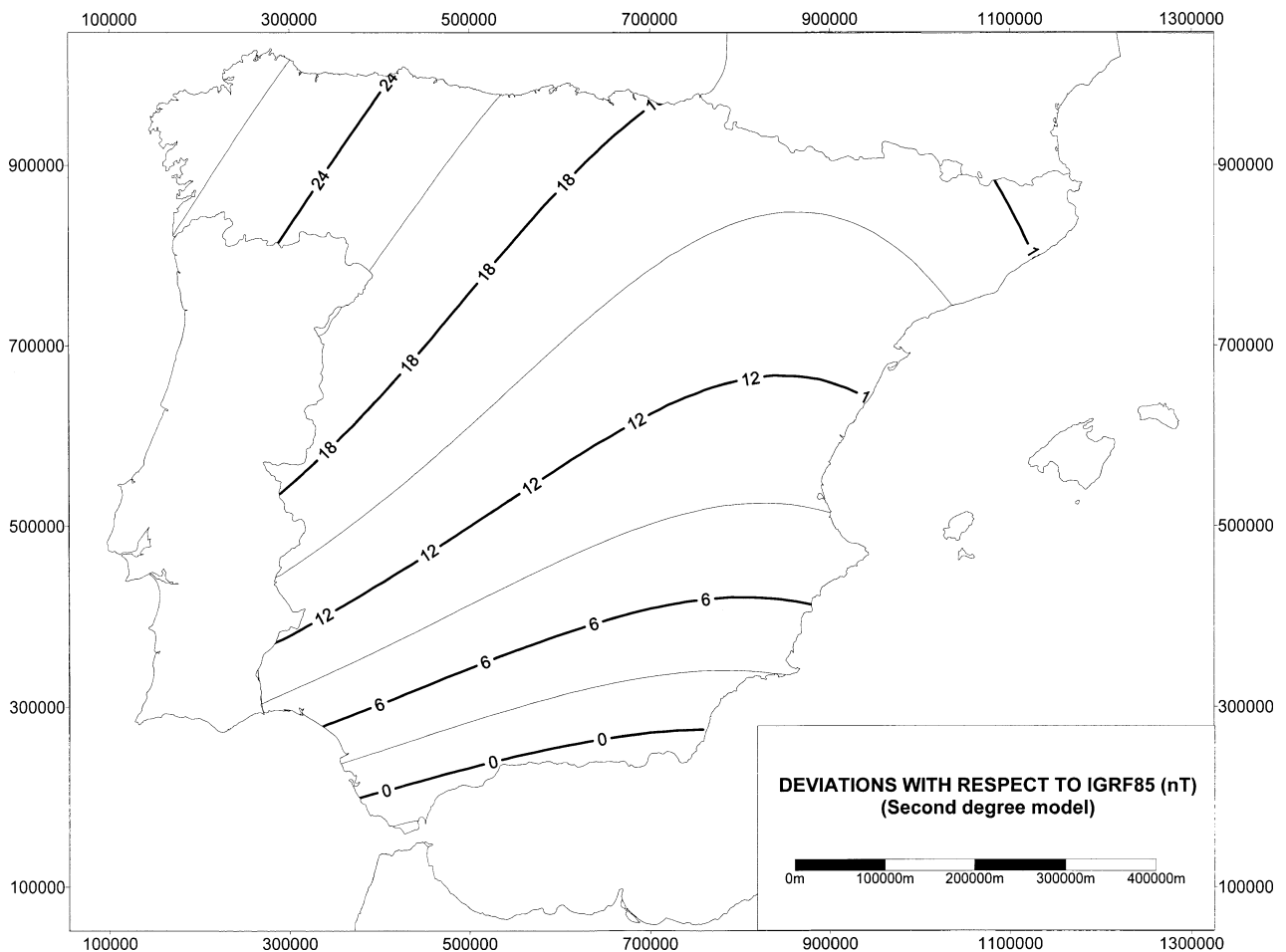


Fig. 3. Map of differences between the second degree polynomial model and the IGRF85 moved to the reference date (1987.0). Values are given in nT and contour intervals correspond to 3 nT (Spanish Official Lambert Projection).

conditions of validity demanded by the Regression Theory: Homogeneity, Robustness, Noncollinearity, Linearity, Homoscedasticity, Normality and Independence (Morrison, 1983; Weisberg, 1985; Peña, 1989b). The problems arising as a consequence of ignoring these requirements are numerous and have been studied in detail in the Regression literature (Weisberg, 1985). The main one is the loss of accuracy when making the estimation of the model provided by the least squares method. In the next paragraphs, we summarize the analysis carried out to verify the degree of fulfilment of these conditions, starting with the most general ones. A more detailed study can be found in Ardizone (1996).

Homogeneity can be defined as the lack of observations with abnormal residuals that may cause important distortions in the model. Experience shows that the residuals which distant three times the standard deviation from the mean, fall into this category. In our case, the analyses have shown that for the first, second and third degree models there are 1.03%, 1.90% and 1.84% values which respectively exceed the limits indicated. So, the residual values should be considered heterogeneous.

The robustness of the model, i.e., its independence of any particular set of data, may be affected by the presence of influential observations that for the  $i$ -th observation can be detected through a statistical test based on the

Cook's distance,  $D_i$ . This statistic can be defined as (Weisberg, 1985),

$$D_i = \frac{(\hat{d}_{-i} - \hat{d})^T (X^T X)^{-1} (\hat{d}_{-i} - \hat{d})}{(p+1)\hat{\sigma}_R^2} \quad (7)$$

where  $\hat{d}$  is the vector of coefficients estimated by least squares;  $\hat{d}_{-i}$  is the vector of coefficients estimated using all of cases except the  $i$ -th;  $p + 1$  is the number of coefficients of the model and  $X$  is the matrix of observations of independent variables which first column is completed with 1.

Therefore, it is usually accepted that observation "i" is influential when its Cook's distance exceeds the critical value  $F_c$ , for a significant level, which is equal to 5% in this particular analysis. Applying this test to the three models the results for  $F_c$  are of 2.60, 2.10 and 1.83. The maximum values of Cook's distance found in our case are 0.01 for the first degree model and 0.03 for the two other cases. Thereby prove that there are not influential observations.

For its part, the noncollinearity means the absence of any intense correlation between independent variables. The simplest way of looking into this property is to

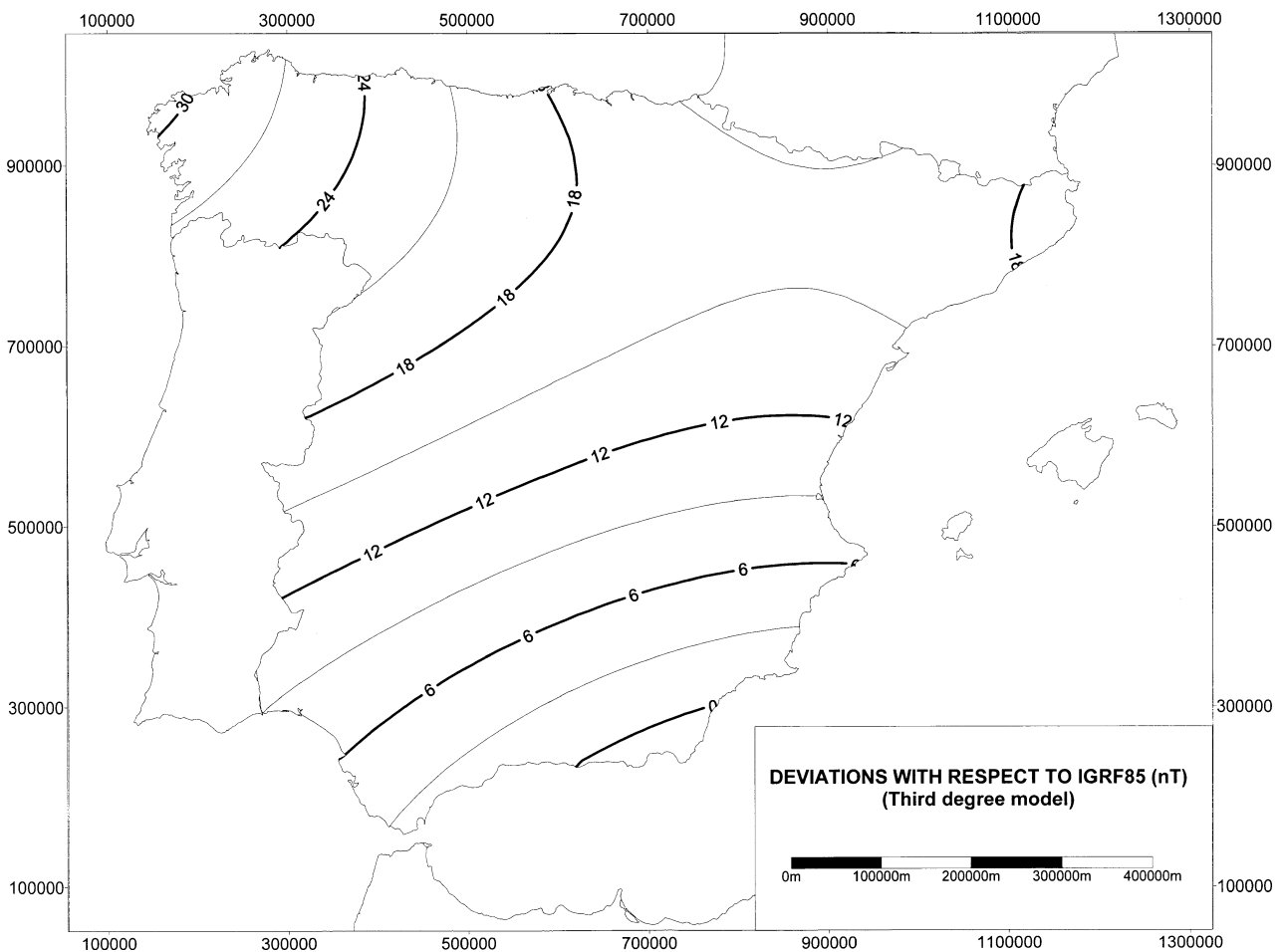


Fig. 4. Map of differences between the third degree polynomial model and the IGRF85 moved to the reference date (1987.0). Values are given in nT and contour intervals correspond to 3 nT (Spanish Official Lambert Projection).

examine the correlation matrix of independent variables. In our case, the correlation matrices corresponding to each model are those shown in Table 2. In the first two cases the coefficients of linear correlation between couples of independent values are very low. Nevertheless, some elements of the third, mainly  $r_{81}, r_{92}, r_{62}, r_{71}, r_{75}$ , are high, which indicates that the correlation is quite important in some cases. This fact may be due to the change of variables (4) performed to make the models linear. To quantify the effect we use the conditioning index defined as:

$$IC = \sqrt{\frac{L_M}{L_m}}, \tag{8}$$

where  $L_M$  and  $L_m$  are respectively the maximum and minimum eigenvalues of the correlation matrix of independent values. When the conditioning index is less

than 10, multicollinearity does not exist or it is considered to be very low. A value between 10 and 30 shows that the multicollinearity is moderate and when it is bigger than 30 it is deemed high (Belsley *et al.*, 1980). In our case,  $IC$  values are 1.073 for the first degree model, 1.685 for the second degree and 6.508 for the third one. Therefore, the first two models do not have multicollinearity and that of the last one is very low. It can be stated that the least square matrix is well conditioned in the three cases.

Only if the relationship of the magnetic field values with the independent variables has been correctly established, it is possible to say that our model fulfils the linearity condition. Probably, the quickest method of testing it is to analyze the plot of residuals versus the estimated values. When the cluster of points depicted by this graphic has a tendency to make a turn, a lack of linearity, which possibly is due to the omission of any relevant variable in the model, can affect the data. In our

Table 2. Matrixes of correlation between the independent values of each model.

Degree 1									
	$w_{10}$	$w_{01}$							
$w_{10}$	1.0000								
$w_{01}$	0.0706	1.0000							

Degree 2					
	$w_{10}$	$w_{01}$	$w_{11}$	$w_{20}$	$w_{02}$
$w_{10}$	1.0000				
$w_{01}$	0.0706	1.0000			
$w_{11}$	-0.2145	0.3478	1.0000		
$w_{20}$	-0.2457	-0.2157	0.0474	1.0000	
$w_{02}$	0.2848	0.1598	0.0564	-0.0654	1.0000

Degree 3									
	$w_{10}$	$w_{01}$	$w_{11}$	$w_{20}$	$w_{02}$	$w_{21}$	$w_{12}$	$w_{30}$	$w_{03}$
$w_{10}$	1.0000								
$w_{01}$	0.0706	1.0000							
$w_{11}$	-0.2145	0.3478	1.0000						
$w_{20}$	-0.2457	-0.2157	0.0474	1.0000					
$w_{02}$	0.2848	0.1598	0.0564	-0.0654	1.0000				
$w_{21}$	0.0940	0.7635	0.2571	-0.3895	-0.0567	1.0000			
$w_{12}$	0.6563	0.0950	-0.2288	-0.0361	0.6659	-0.0079	1.0000		
$w_{30}$	0.8862	0.5229	-0.3094	-0.4176	0.2055	0.1588	0.5233	1.0000	
$w_{03}$	0.047	0.8312	0.5190	-0.1738	0.3348	0.6186	0.1718	-0.0062	1.0000

Table 3. Asymmetry and kurtosis tests for each model residual distribution.

Polynomial Degree	Skewness	Statistic	Critical Values $\alpha=5\%$
1	0.30	9.07	$\pm 1.96$
2	2.25	66.91	$\pm 1.96$
3	2.32	69.20	$\pm 1.96$
Polynomial Degree	Kurtosis	Statistic	Critical Values $\alpha=5\%$
1	1.81	26.92	$\pm 1.96$
2	18.45	274.55	$\pm 1.96$
3	18.99	282.64	$\pm 1.96$

case, the graphics are presented in Figs. 5(a), (b), and (c). The first of them, corresponds to the first order polynomial model. There, it can be noticed that the cluster of points tends to turn to the negative values of the y-axis. On the contrary, Figs. 5(b) and (c), that belong to the polynomial models of second and third degree respectively, do not show any significant curvature. In these cases the zone of

the highest density of observations spread around the x-axis residual mean. We can conclude that the first degree model shows a lack of definition which may stem from the absence of variables (in our case, those terms of a degree higher than the first one). On the other hand, the two remaining models are well defined and have a very similar behaviour.

The normality of the residuals of the polynomial models has been studied with the Kolmogorov-Smirnov contrast that compares the cumulative distribution of residuals with a normal distribution of the same mean and standard deviation (Siegel, 1986). If the statistical values of the contrast  $|D|_M$  exceed a critical value  $|D|_{MC}$ , the distribution lacks of normality. According to this test,  $|D|_{MC}$  equals 0.019 when a 5% level of significance is considered. As the results obtained are 0.040 for the first degree model and 0.160 for the two others, the hypothesis of normality is rejected in the three cases. To look into the reasons for this lack of normality we resort to testing for skewness and kurtosis (Afifi and Azen, 1977), whose results are shown in Table 3. They prove that the null hypothesis of asymmetry and kurtosis is rejected for the chosen level of 5%. Consequently, in a more rigorous way, we can conclude that the reasons why the hypothesis of normality is abandoned are the asymmetry and the kurtosis of the distributions. The first is due to the instability caused by positive values of high residuals, whereas the second can be attributed to the strong concentration of these kinds of values in the tails and in the middle of the distribution. All this is confirmed when we compare the histograms of the residuals from each distribution with the curve of density of normal probability and with the maps of residuals for each model (Ardizone, 1996).

Since the lack of normality is a result not only of the asymmetry in the distribution of residuals but also of the strong kurtosis originated by the residuals of high values, the change of variable which is the normal procedure adopted to overcome the skewness, is not useful in this case. As will be seen later in this paper, the solution is linked to the search of areas where the variability of residuals is lesser.

The homoscedasticity, that is, the homogeneity of the model perturbation variances, can be studied through contrasts of hypothesis and graphics of residuals against estimated values of the dependent variable. When, as it happens in this case, only one measurement corresponds to each point, the most efficient method is the last one.

Graphics of Figs. 5(a), (b), and (c) clearly show the heterogeneity of variances with no need to use any test of hypothesis. Indeed, although no systematically increment or decrement with the estimated value of the earth's magnetic field intensity is observed, there exist groups whose variability is by far greater than the rest because they present residuals with very high values (greater than two and a half times the residual standard deviation). In order to reduce the effect of these data with atypical residuals, the Huber's weighting function (Huber, 1981) has been used. Nevertheless, the improvement has been insignificant which indicates the need to use more robust techniques.

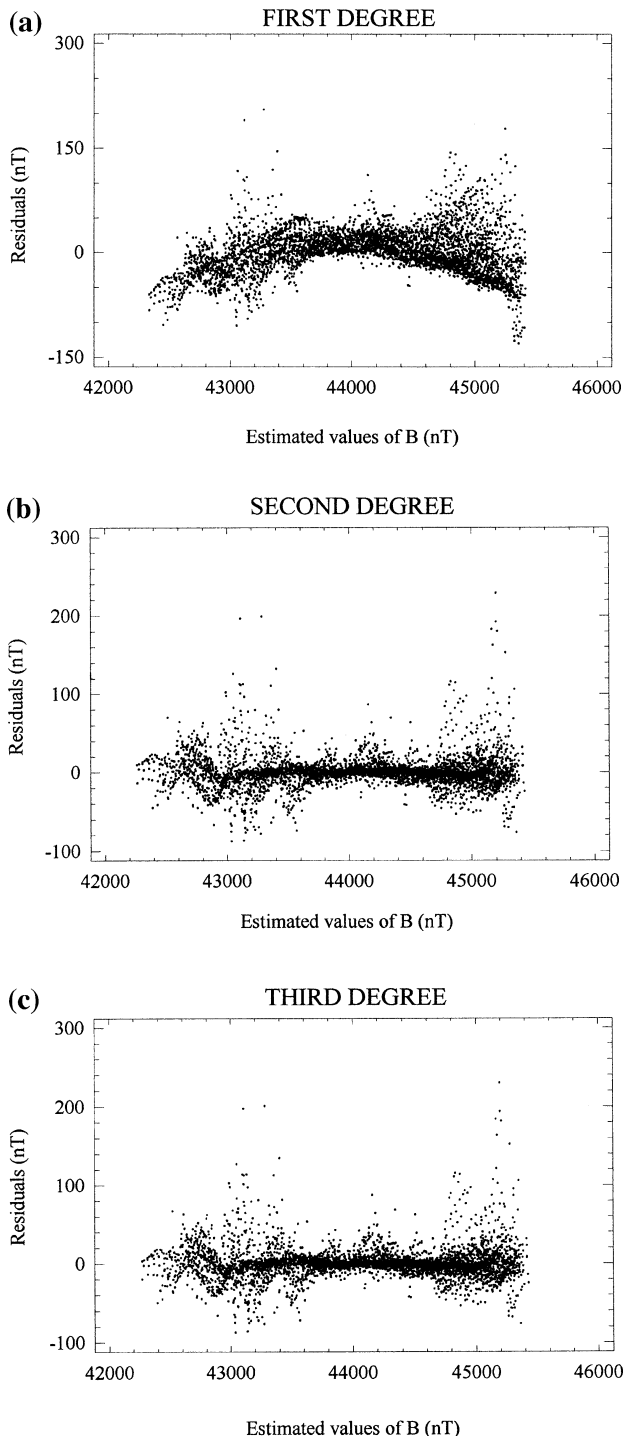


Fig. 5. Plot of residuals versus estimated values of magnetic field for the first (a), second (b) and third (c) degree models. A curvature indicating lack of linearity in the first degree model can be clearly observed in (a). On the contrary, in (b) and (c) the cloud of points is close to the zero line which reveals that every trend has been removed by the model.

Finally, the independence of the residuals (that is to say, the lack of autocorrelations among them) can be studied through the contrast of runs. With this last expression we mean a series of values whose residuals are either over or below the average. The hypothesis of independence is rejected when the number of runs is significantly high or low. In our case, because of the large amount of data, this

technique has been performed through the normal approximation (Siegel, 1986). The results of the analysis applied to the residuals of the three polynomial models indicate a critical value of  $\pm 1.96$  that corresponds to the usual 5% level of significance, statistical values of  $-3.77$ ,  $-37.86$  and  $-37.64$ . The respective numbers of runs are 699, 1,297 and 1,278. These results show that the

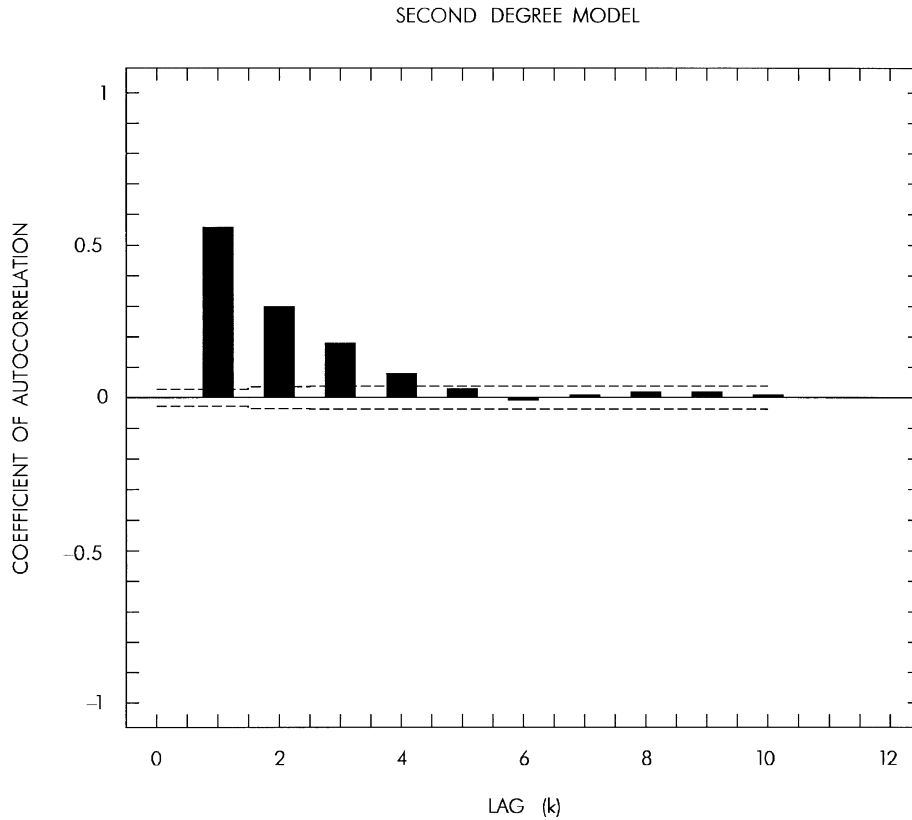


Fig. 6. Autocorrelation plot of the residuals corresponding to the second degree model. Horizontal dotted lines indicate the level from which values are significantly different from zero for the usual confidence level of 5%. As it can be observed, autocorrelation exists only in the first four lags there and even then, values are low.

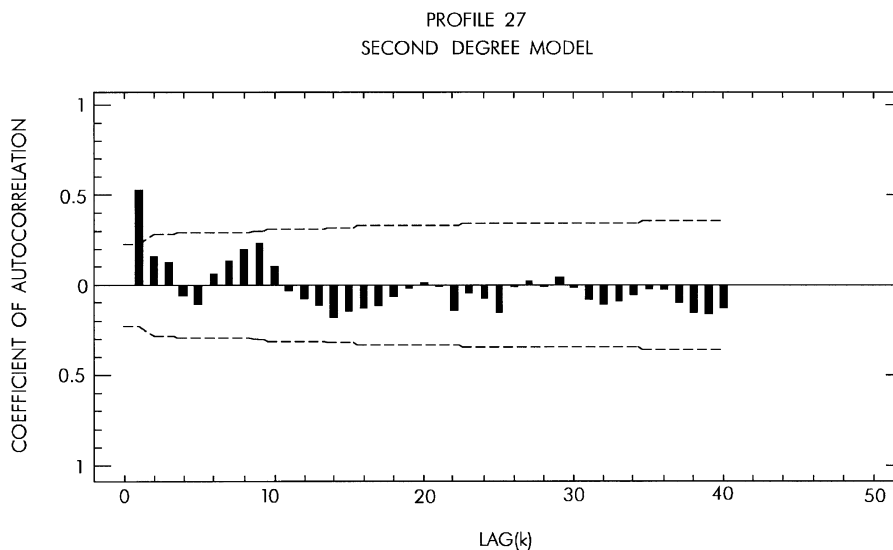


Fig. 7. Autocorrelation plot of the residuals corresponding to the longest (752,7 km) North-South profile, named Profile 27. In this case, only the coefficient of autocorrelation for  $k = 1$  is significantly different from zero.



hypothesis of independence must be rejected in the three cases, which indicates that there exists an autocorrelation in the residuals. However, as yet we do not know whether this one is high or low. This quantitative aspect has a very special importance because the accuracy of the estimates of the model coefficients will be low if the coefficient of autocorrelation is high. To study quantitatively this coefficient, we make use of the residual autocorrelation plot. As an example, the case of the second degree polynomial model it is represented in Fig. 6 for the lag ( $k$ ) varying from 1 to 10. Values of the correlation coefficient for each lag  $k$ ,  $r_k$ , which exceed the pair of horizontal dotted lines are significantly different from zero for a level of 5%. As can be noticed, the overcoming takes place only in the first four lags. In addition, values attained are very low (the highest one is  $r_1 = 0.5542$ ) and a considerable decrease occurs from  $k = 2$  upwards. The behaviour for the residuals of third degree model is very similar but in the case of the first degree, the result is considerably worse and  $r_1$  reaches a value of 0.8009 (Ardizone, 1996).

In any of the three cases, the explanation for the existence of the positive autocorrelation is to be found in the intrinsic nature of data and in their local distribution. In fact, because they are geographical data, the value of the intensity of the earth's magnetic field and the residual in each point, are related to those surrounding them. This influence comes from the own nature of these data and can not be avoided. In addition, the profiles correspond to lines flown in direction North-South but are arranged from West to East. As a consequence, there will exist another source of autocorrelation in the data originating from the sequential order of the profiles. According to this, the values of the coefficients of autocorrelation should decrease when this influence is eliminated because only one profile is considered. This can be seen in Fig. 7, which corresponds to the autocorrelation plot of the longest (757.2 km) North-South profile. This time, only the  $r_1$  value can be considered notably different from zero, unlike the case of the autocorrelation plot of the whole survey (Fig. 6) where this was the case with three more values. From this last result, we can conclude that the effect of the autocorrelation on the efficiency of the coefficients can be reduced limiting the profile arrangement in a predominant geographical direction.

**5. Search for a Region where the Hypotheses of the Model are Verified**

Considering the results obtained in the previous section, we can deduce that the heterogeneity of the residuals and the intrinsic nature of data are the main reasons for the deviation from the hypotheses of normality, homoscedasticity and independence. A possible solution to these problems is suppressing the data that cause the deviations already mentioned. One way to accomplish it is to detect the residuals whose values are far from the media and whose Mahalanobis distance is high (Afifi and Azen, 1977).

Mahalanobis distance from the  $i$ -th observation to the mean, is defined as (Weisberg, 1985),

Table 4. Estimates of the coefficients of the second degree model with inferential character.

Variables	Coefficients	Confidential Intervals (95%)
Intercept.	44150.211	(44149.925, 44150.496)
$w_{10}$	394.797	(394.6397, 394.954)
$w_{01}$	23.291	(23.132, 23.451)
$w_{11}$	-3.268	(-3.379, -3.157)
$w_{20}$	-4.407	(-4.502, -4.312)
$w_{02}$	1.985	(1.903, 2.066)

$$MD_i = (x_i - m)^T S^{-1} (x_i - m) \tag{9}$$

where  $x_i$  is the vector of observations of a  $n$ -dimensional variable;  $m$  the vector of their means and  $S$  is the covariance matrix of their components.

This last statistic is introduced in order to strengthen the accuracy of the model because the greater the Mahalanobis distance of an observation, the lesser the accuracy of the estimation of the predicted value. In order to estimate the critical values of the Mahalanobis distances we have employed the contrast of signification proposed by Afifi and Azen (1977) and Jobson (1991). The analysis has been applied only to the second and third degree models because, as it has been already explained, the model corresponding to the first degree does not even fulfil the noncollinearity condition. The Mahalanobis distances obtained for the second and third degree polynomial models, for a level of signification of 5%, are 11.10 and 16.92, respectively. For the second degree, after ten iterations, we have reached the new model summarised in Table 4. This one yields the region shown in Fig. 8 which contains 1,872 points (marked with dots) and where the conditions of normality and homoscedasticity are already verified. In this area, the statistical independence is not yet satisfied but the results of the test of runs are better than in the case of the whole Peninsula (Ardizone, 1996). For this region, the differences between the model of Table 4 and that of the same degree defined for the whole group are lower than 6.68 nT and their rms equals 2.46 nT.

The third order model does not meet the hypothesis of normality in this region yet and the procedure has to continue although now with a Mahalanobis critical distance of 16.92. The final result is a new region of validity included within the area plotted in Fig. 8 and with only 962 observations. Since the second degree polynomial model listed in Table 4 is more simple and its region of validity is larger, it can be considered as the best inferential model.

**6. Discussion**

The analyses performed so far confirm the necessity of distinguishing between polynomial surfaces used to represent the geomagnetic field that only have a descriptive character and those that seek to include an inferential value. Models listed in Table 1 belong to the first group and provide a worthy description of the geomagnetic field according to the processed information. Although they represent a very interesting result, it is necessary be aware

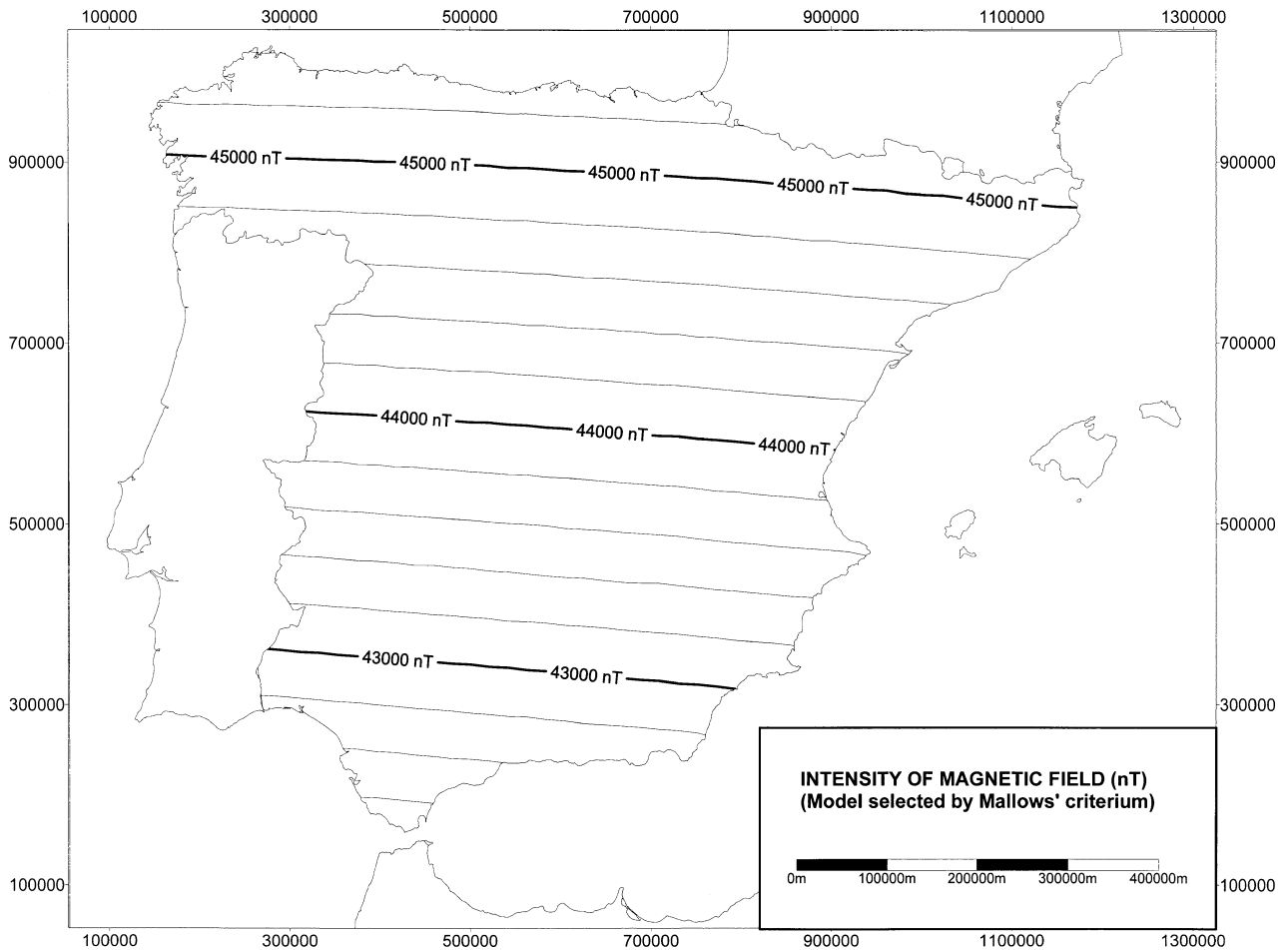


Fig. 8. Map of the magnetic field intensity corresponding to the model selected by Mallows' criterium. This model constitutes the best non inferential model for the whole Spanish Mainland.

that they are not able to make predictions of values different to those used as starting point. The enhancement of these models to give them this ability requires the fulfilment of the seven previously described conditions, which can introduce important restrictions in the area of applicability.

With reference to these descriptive models (Table 1), the values obtained for the standard error of regression indicate that those of second and third degree represent a remarkable improvement with regard to that of first degree. This result has been confirmed by performing a spectral analysis of the longest North-South profile. The spectra corresponding to the models of second and third degree share the same shape that, in agreement with the global geomagnetic field characteristics, includes increasing values for higher latitudes. This trend is not present at the first degree model spectrum whose amplitudes also reach different values (Ardizone, 1996). Moreover, the root mean square of differences between the spectrum corresponding to the first degree model and those of second and third degree are 769.98 and 768.73, respectively, whereas the value decreases to 8.12 when the comparison is made between the second and the third degree models. The similarity between these last two models is also clearly revealed when they are compared with the IGRF85 translated to the reference date (Figs. 3

Table 5. Optimal values of the model without inferential character deduced from the Mallows' criterium.

Variables	Coefficients	Standard Error of Coefficients
Intercept.	44151.6	0.49
$w_{10}$	396.261	0.324
$w_{01}$	21.9155	0.1769
$w_{11}$	-2.9567	0.0638
$w_{20}$	-4.4102	0.091
$w_{02}$	2.308	0.0381
$w_{21}$	-0.0934	0.0394
$w_{30}$	-0.1286	0.0425

and 4). In any case, this last argument must be taken with care because the data source of IGRF is clearly different from that of the aeromagnetic survey.

Several criteria can be used to select which of these two models can be considered more accurate. The most practical is to compare the standard errors of regression and the coefficients of determination. Nevertheless, in this case this criterion is not useful because, as has been already mentioned, these statistics produce very similar

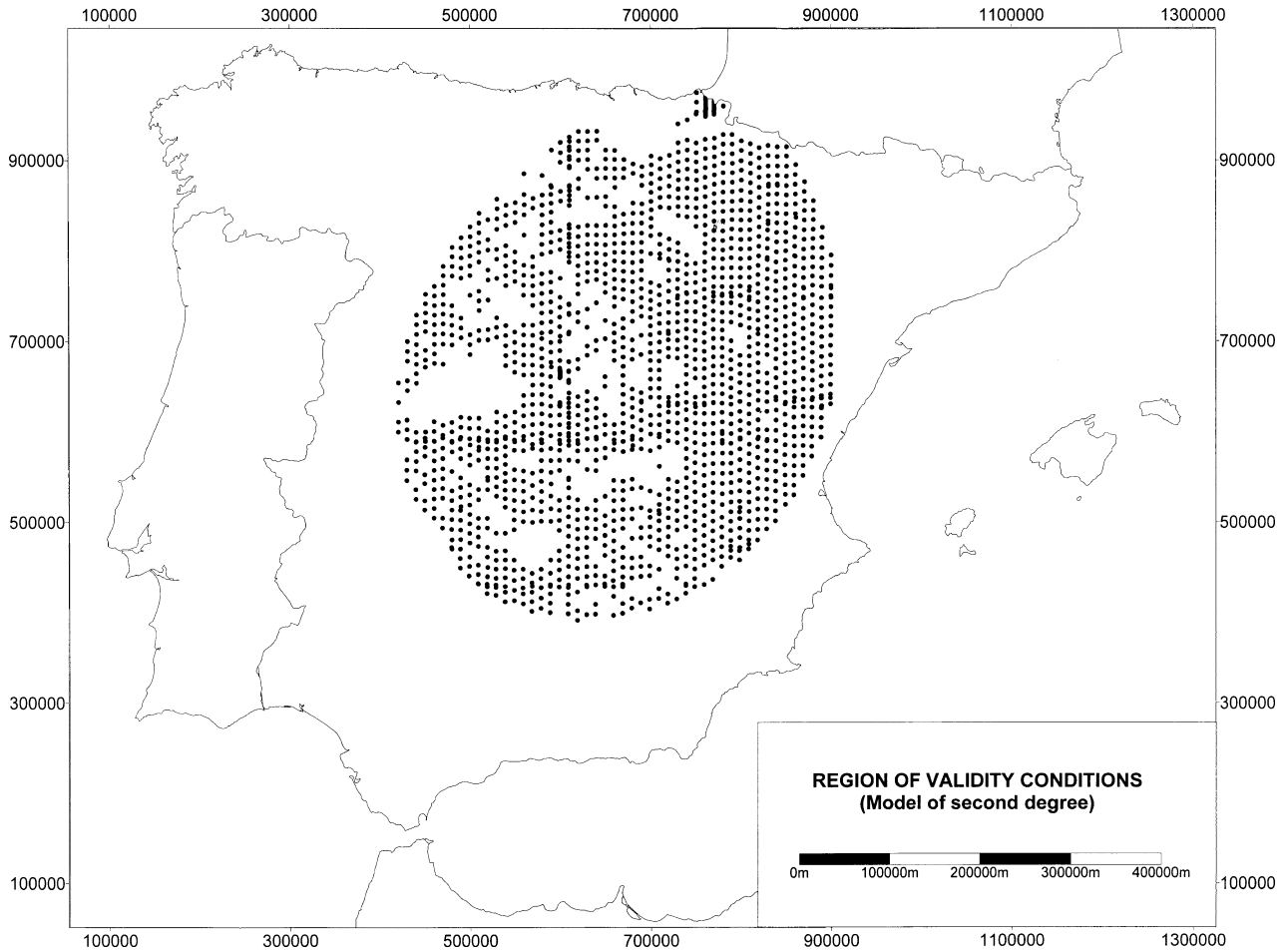


Fig. 9. Area where validity conditions are satisfied for the second degree model and an inferential character can be established. Dots indicate the observation points (Spanish Official Lambert Projection).

values for both models. To overcome this situation we have also applied the Mallows' criterium (Morrison, 1983; Weisberg, 1985) that is based on the minimisation of the root mean square error of the prediction in the points of observation. In our case, the model with the lowest value of the Mallows' coefficient is listed in Table 5 and plotted in Fig. 8. It gives the best statistical adjustment to the analysed data that can be applied to the whole Spanish Mainland without inferential character. Its explicit expression is:

$$\begin{aligned} \hat{B} = & 44151.60 + 396.261(\varphi - 40.427) + 21.9155(\lambda + 3.537) \\ & - 2.9567(\varphi - 40.427)(\lambda + 3.537) - 4.4102(\varphi - 40.427)^2 \\ & + 2.308(\lambda + 3.537)^2 - 0.0934(\varphi - 40.427)^2(\lambda + 3.537) \\ & - 0.1286(\varphi - 40.427)^3. \end{aligned} \quad (10)$$

The analysis summarised in Chapter 3 proves that the descriptive polynomial models are far from satisfying the necessary statistical requirements to provide them with a predictive capacity. The hypotheses of robustness and noncollinearity are fulfilled by the three models but only those of second and third degree satisfy the linearity and

the homoscedasticity. Conditions of homogeneity, normality and independence of residuals are not met by any of them. The reasons for this disagreement arise mainly from the own nature of data and their geographical distribution, aspects that can not be avoided because they have their origin in the characteristics of the Earth's magnetic field. As the influence of these factors can be increased by the way they are compiled and processed, the measures adopted to introduce randomness in these steps can attenuate the lack of independence. Some other statistical tools can be used to diminish the skewness and the lack of normality but lessening the area of application is the only robust procedure to obtain an inferential model. It implies an impoverishment of the models and represents a shortcoming of the polynomial adjustments. As a consequence, the second degree model listed in Table 4 constitutes the best solution with inferential character, but its area of applicability is restricted to the region shown in Fig. 9. Now, the explicit expression is

$$\begin{aligned} \hat{B} = & 44150.211 + 394.797(\varphi - 40.427) + 23.291(\lambda + 3.537) \\ & - 3.268(\varphi - 40.427)(\lambda + 3.537) - 4.407(\varphi - 40.427)^2 \\ & + 1.985(\lambda + 3.537)^2. \end{aligned} \quad (11)$$

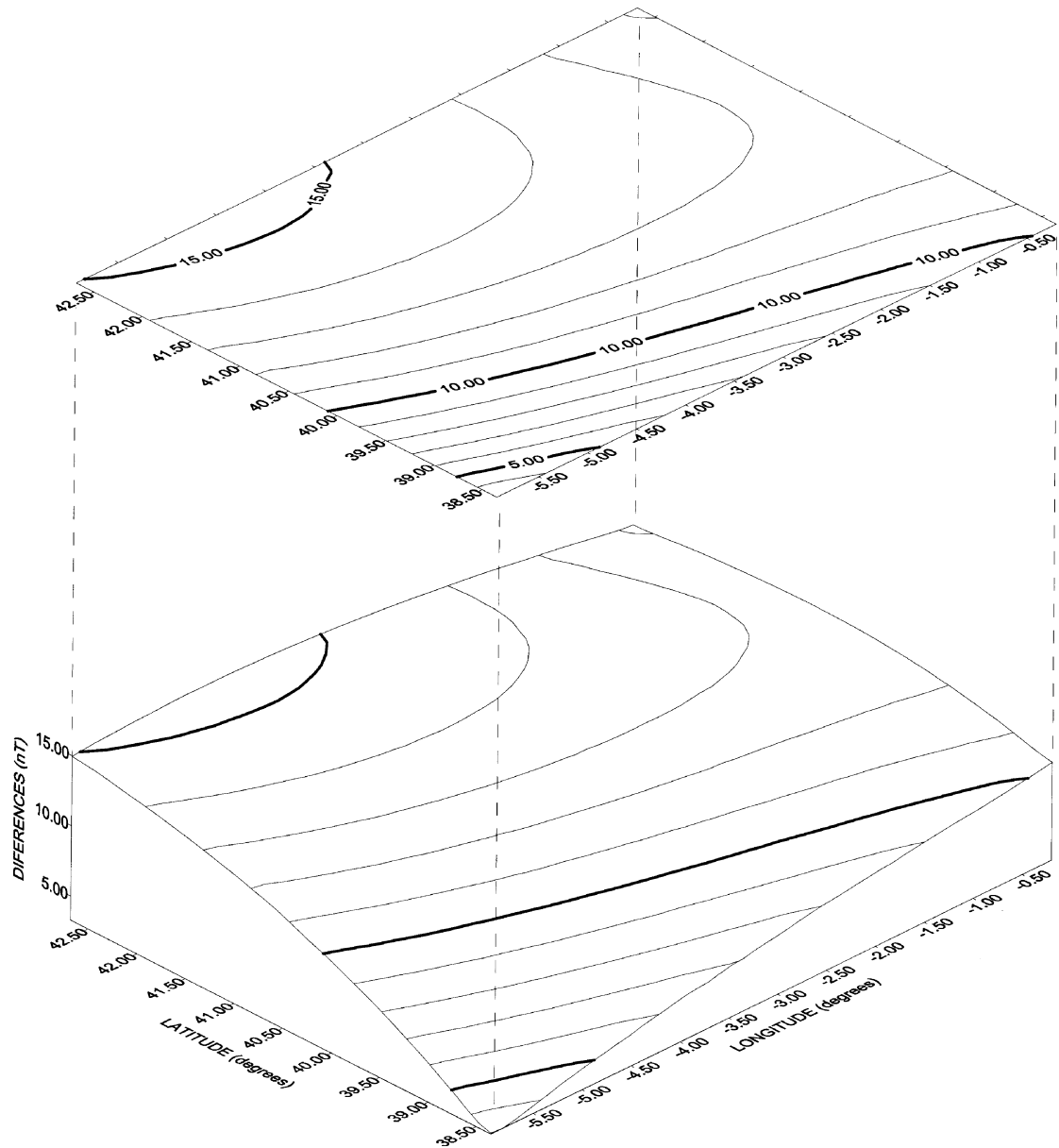


Fig. 10. 3D (bottom) and 2D (top) representations of the differences between the second degree model and the IGRF85 moved to the reference date (1987.0) and to an altitude of 3,000 m above sea level. Contour intervals correspond to 1 nT.

Because the method of polynomial surfaces is an alternative procedure to those based on potentials, it could be interesting to compare the results obtained by both ways. The two most important potential models proposed for Spain are the IGRF85 and the representation obtained by the Spherical Cap Harmonic Analysis (SCHA) (Torta, 1992). The comparison is carried out here by means of an statistical study of the differences between the second order inferential model and these potential models moved to the year of 1987.0 and to an altitude of 3,000 m above sea level. When doing this comparison it is necessary be aware of the differences between the source of data used in the polynomial model and those of potential models. In the first case data come from an aeromagnetic flight and their minimum wavelength corresponds to about 510 km. The SCHA model is mainly based on MAGSAT data (although it includes a significant amount of data taken on

the earth surface) and the minimum wavelength can be estimated in 416 km. In the case of the IGRF, this wavelength is 4,000 km, approximately. The IGRF represents, above all, variations originated by very deep sources (close to the earth's core), while those of the two other models are found at a depth that is considerably lower, presenting the possibility of having its origin in the inferior part of the earth's crust. These facts must be kept in mind when analysing the results.

Figures 10 and 11 respectively plot the differences between the second degree model and the IGRF and the SCHA models. The resulting root mean squares of these are 12.22 nT when the comparison is made with the IGRF and 20.47 nT when the SCHA model is taken. These results indicate that, although there exists a small improvement when the comparison is made with the IGRF, the behaviour of the polynomial surface is very

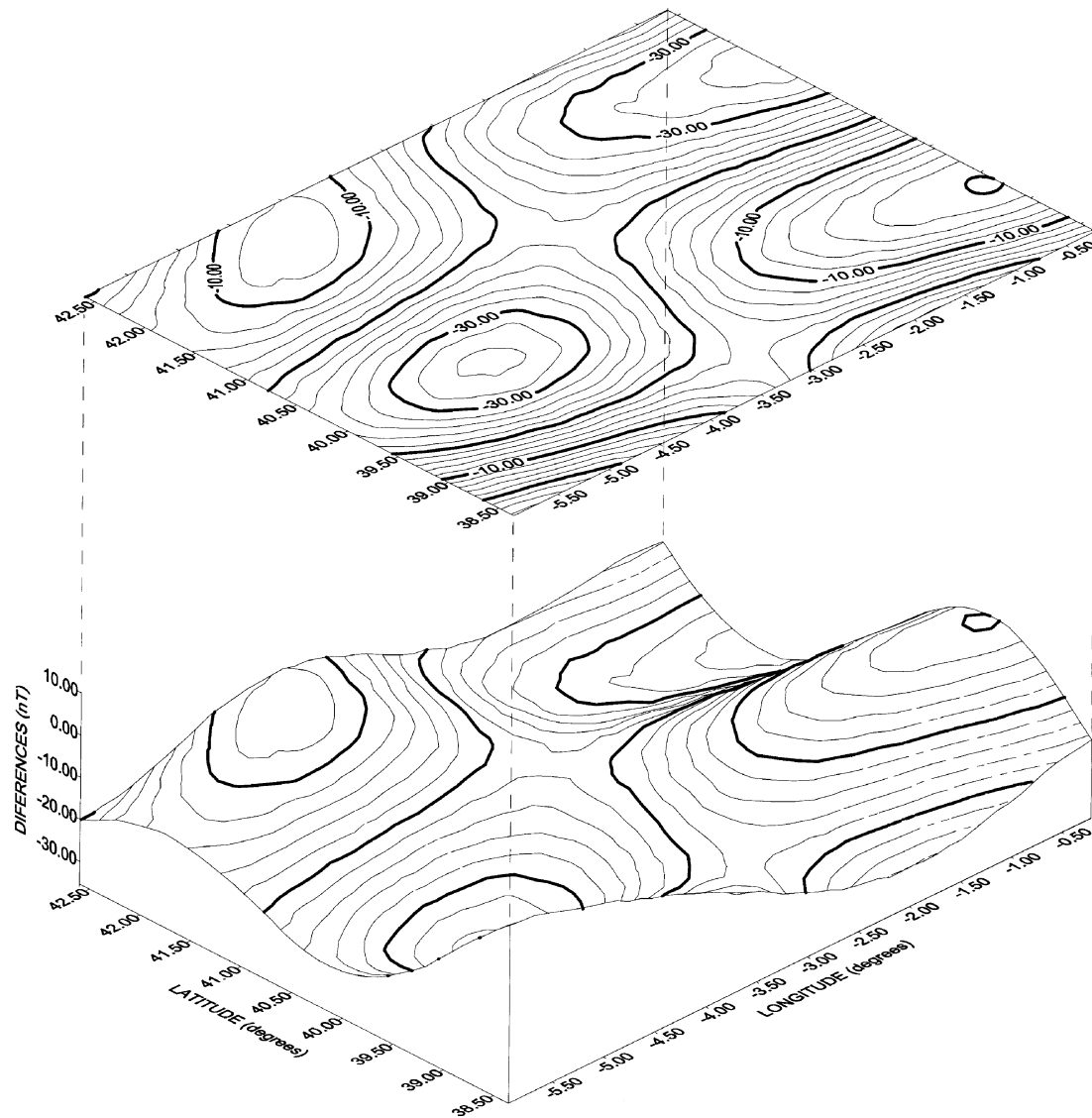


Fig. 11. 3D (bottom) and 2D (top) representations of the differences between the second degree model and the SCHA (Torta, 1992) model moved to the reference date (1987.0) and to an altitude of 3,000 m above sea level. Contour intervals correspond to 2 nT.

similar no matter which model is chosen. On the other hand, when the mean values are calculated, the differences with respect to the SCHA tend to keep a sign opposite to that referred to the IGRF. This result suggests that the polynomial model to represent a middle state between both solutions.

The discrepancies that exist between the polynomial and the potential solutions must be considered in the light of the previously commented different source of data. In addition, the time shift of the models to 1987.0 and the prolongation up to 3,000 m of altitude, introduce uncertainties in the result because both models have been built with data mostly coming from a satellite (between 500 and 300 km. of nominal altitude). So, this comparison should be considered mainly as a mathematical exercise.

## 7. Conclusions

The method of polynomial surfaces stands as a useful alternative to the potential methods because it does not necessarily have to derive from a potential function

(Haines, 1968) and provides simpler models. In addition, this method simplifies the procedure and, if some requirements are satisfied, allows for the prediction (evaluation of the model at those points where there are no measurements) and the estimation of the corresponding errors. When applied correctly, it makes possible the inferential analysis of the model coefficients, and stands as a way to obtain the optimal method by following rigorous criteria.

In this study, the best descriptive model for the whole Spanish mainland correspond to the third degree adjustment given by expression (10).

The polynomial adjustment also has some disadvantages because the uncertainty in the boundaries can be high and it requires a rigorous statistical analysis of the conditions of validity if the highest precision for efficient estimates of the model is to be achieved. Due to its 2-dimensional character, the analytical prolongation of the field is complicated and when applied to different components, geometrical and physical incongruities can arise. Some

limitations of the polynomials models can be originated from the data themselves. For instance, the lack of independence is due to the latitudinal variation of the geomagnetic field and can be attenuated by reducing the region of analysis. In our case, the precise application of polynomial analysis to geomagnetic data proves that the validity conditions are met by the second and third degree models in two restricted areas but not in the whole Spanish mainland. The second degree model, given by expression (11), has been chosen as the best solution because is simpler than the third degree polynomial adjustment and has a larger region of validity.

It is important to bear in mind that the characteristics of geomagnetic data can also be found in other geophysical information. Therefore, the advantages and disadvantages of polynomial surfaces mentioned so far can be present when these models are applied to other fields of the Geophysics.

**Acknowledgments.** We thank the Geomagnetism and Aeronomy Service of the National Geographic Institute of Spain, for providing the data of the Aeromagnetic Survey of the Spanish mainland. Dr. M. Torta kindly provided data to perform the comparison of our results with his SCHA model. We also are very grateful to Dr. L. Alberca and two anonymous reviewers for their comments and suggestions on an early version of the manuscript.

## References

- Afifi, A. A. and S. P. Azen, *Statistical Analysis. A Computer Oriented Approach*, 442 pp., Academic Press, London, 1977.
- Ardizone, J. A., Análisis de datos aeromagnéticos. Metodologías y aplicación al levantamiento aeromagnético de España Peninsular, Ph.D. Thesis, 225 pp., Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 1996 (in Spanish).
- Ardizone, J. A., J. Mezcuca, and I. Socías, *Mapa Aeromagnético de España Peninsular*, 29 pp. and 3 maps, Instituto Geográfico Nacional, Madrid, 1989 (in Spanish).
- Belsley, D. A., A. Khu, and R. E. Welsch, *Regression Diagnostics*, 292 pp., John Wiley & Sons, New York, 1980.
- Galdeano, A. and J. C. Rossignol, Assemblage à altitude constante des cartes d'anomalies magnétiques couvrant l'ensemble du bassin occidental de la Méditerranée, *Bull. Soc. Géol. France.*, **XIX**(3), 461–468, 1977 (in French with English abstract).
- Haines, G. V., A Taylor expansion of the geomagnetic field in the Canadian Arctic, *Publications of the Dominion Observatory*, **XXXV**, 115–140, 1967.
- Haines, G. V., Polynomial estimation of certain geomagnetic quantities, applied to a survey of Scandinavia, *Publications of the Dominion Observatory*, **XXXVII**, 79–112, 1968.
- Haines, G. V., Regional magnetic field modelling: A review, *J. Geomag. Geoelectr.*, **42**(9), 1001–1018, 1990.
- Haines, G. V. and W. Hannaford, Magnetic anomaly maps of British Columbia and the adjacent Pacific Ocean, *Publications of the Earth Physics Branch, Department of Energy, Mines and Resources, Ottawa*, **42**(8), 215–228, 1972.
- Haines, G. V. and W. Hannaford, A three component aeromagnetic survey of the Canadian Arctic, *Publications of the Earth Physics Branch, Department of Energy, Mines and Resources, Ottawa*, **44**(8), 213–234, 1974.
- Haines, G. V., W. Hannaford, and P. H. Serson, Magnetic anomaly maps of the Nordic Countries and the Greenland and Norwegian seas, *Publications of the Dominion Observatory*, **39**(5), 1970.
- Huber, P. J., *Robust Statistics*, 308 pp., John Wiley & Sons, New York, 1981.
- Jobson, J. D., *Applied Multivariate Data Analysis, Vol. I, Regression and Experimental Design*, 621 pp., Springer-Verlag, 1991.
- Morrison, D. F., *Applied Linear Statistical Methods*, 562 pp., Prentice-Hall Inc., New Jersey, 1983.
- Peña, D., *Estadística Modelos y Métodos, V. 1 Fundamentos*, 402 pp., Edit. Alianza Universidad Textos, Madrid, 1989a (in Spanish).
- Peña, D., *Estadística Modelos y Métodos, V. 2 Modelos Lineales y Series Temporales*, 745 pp., Edit. Alianza Universidad Textos, Madrid, 1989b (in Spanish).
- Siegel, S., *Estadística no Paramétrica*, 562 pp., Trillas, Méjico, 1986 (in Spanish).
- Socías, I., J. Lynam, and R. del Potro, Interpretation of an Aeromagnetic Survey of the Spanish Mainland, *Earth Planet. Sci. Lett.*, **105**, 55–64, 1991.
- Tanaka, M., K. Hiroishi, S. Ando, and S. Matsumura, Aeromagnetic surveys in the Geographical Survey Institute and the magnetic anomalies in and around Japan, *Bull. of the Geological Survey of Japan*, **XXX**, 1–14, 1986.
- Torta, J. M., *Modelización regional del campo geomagnético sobre España: Campo anómalo, variación secular y campo de referencia*, 139 pp., Publicaciones del Observatorio del Ebro, 15, Tortosa, 1992 (in Spanish with English abstract).
- Weisberg, S., *Applied Linear Regression*, 324 pp., John Wiley & Sons, New York, 1985.