6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Compliance analysis in polynomial surface determination with ANOVA

Ulku Kirici Yildirim[*1], **Yasemin Sisman** [1]

[1]Ondokuz Mayis University, Engineering Faculty, Department of Geomatics, Samsun, Türkiye

Keywords
Geoid
Polynomial geoid determination
Significant test
Regression analysis

Abstract

Geoid determination has always been an important subject of study. Geoid determination is the modelling that enables us to determine the height of a point whose position is known. While determining the geoid, it is very important to state the degree of the surface. At first sight which degree polynomial surface will be used or appropriate point distribution is unknown for the work area. A significance test should be performed to determine the most appropriate degree of polynomial with regression analysis. This study tried to determine the best fit polynomial geoid for the region where Ondokuz Mayis University is located in Samsun.

1. Introduction

The geoid determination is the most important problem for scientist interested in the earth. There are a lot of areas interested in geoid like geodesy, geophysics, geography etc. (Akçın, 2001). The geoid called the surface closed the average sea surface and formed by the combination of the points have got sea level (Sjöberg, 2023). The geoid is a complex surface and it is not easy defined as mathematically. In the geodesy the measurements on the physical earth, but the calculation of measurements is done on the reference surface (Bolat, 2011).

2. Method

2.1. The determination of the best suitable polynomial

This method is the most widely used surface fitting procedure. The function of surface is determined with basic definition of orthogonal polynomials: (Cakır, Yilmaz, 2014).

$$N_{(x,y)} = \sum_{i=0}^{m} \sum_{\substack{j=k-i \\ i=0}}^{k} a_{ij} x^i y^j \quad (1)$$

where (x, y) is the position coordinates of points, aij the constants of the polynomial and m the order of the chosen polynomial. 2nd order polynomial equation can be written for the polynomial:

$$N_{(x,y)} = a_{00} + a_{01}y + a_{10}x + a_{02}y^2 + a_{11}xy + a_{20}x^2 \quad (2)$$

Equation 2, the measurement and unknown numbers are equal to the point and constants number. If the measurement number (n) is bigger than the unknown number (u), the solution must be realized by using adjustment procedure. When the Equation (2) are designed according to indirect measurement adjustment mathematical model, the following equations are obtained:

$$V = AX - \ell \qquad P_{ll} = Q_{ll}^{-1} \qquad X = \begin{bmatrix} a_{00} \\ a_{01} \\ a_{10} \\ a_{02} \\ a_{11} \\ a_{20} \end{bmatrix} \quad A = \begin{bmatrix} 1 & y_1 & x_1 & y_1^2 & x_1y_1 & x_1^2 \\ 1 & y_2 & x_2 & y_2^2 & x_2y_2 & x_2^2 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & y_n & x_n & y_n^2 & x_ny_n & x_n^2 \end{bmatrix} \quad \ell = \begin{bmatrix} N_1 \\ N_2 \\ . \\ . \\ N_n \end{bmatrix} \quad (3)$$

This model can be solved by objective function of the least square adjustment method. The unknown parameters are obtained following equation (Sisman, 2014).

$$X = \left(A^T PA\right)^{-1} A^T P\ell \quad (4)$$

The root mean square error formula as follows;

* **Corresponding Author**

*(ulku.kirici@omu.edu.tr) ORCID ID 0000 – 0002 – 3569 – 4482
(ysisman@omu.edu.tr) ORCID ID 0000 – 0002 – 6600 – 0623

Cite this study

Kirici Yildirim U, & Sisman Y (2023). Compliance analysis in polynomial surface determination with ANOVA. Intercontinental Geoinformation Days (IGD), 6, 228-232, Baku, Azerbaijan

$$m_0 = \sqrt{\frac{V^T PV}{n-u}} \tag{5}$$

The measurement group has got the outliers inevitably. These outliers can adversely affect the adjustment. Therefore, the outlier detection test must be done to determine the outliers' measurements, (Aksoy, 1984; Ayan, 1992; Uzun 2003; Bayrak, 2003; Teke and Yalçınkaya, 2005; Bektaş, 2005, Sisman et al. 2012). The outlier detection test is realized according to hypothesis is used for outlier detection. The test size is calculated by using the residuals of measurements and their standard deviation.

2.2. Regression analysis and ANOVA

Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables. It is one of the most important statistical tools which is extensively used in almost all sciences. A model of the relationship is hypothesized, and estimates of the parameter values are used to develop an estimated regression equation. Various tests are then employed to determine if the model is satisfactory. Model validation is an important step in the modelling process and helps in assessing the reliability of models before they can be used in decision making (Ostertagova, 2012).

The degree of the polynomial was determined by regression analysis in the MiniTab program. MiniTab is data analysis, statistical and process improvement software tool used by organizations worldwide to improve quality and reduce costs. Minitab provides users with tools to perform statistical analysis, including hypothesis testing, regression analysis, and ANOVA. Additionally, Minitab provides various graphical tools to help users visualize data (Guide to the BASIC Programming Language, 03.06.2023).

Analysis of variance (ANOVA) is a technique originally developed by Fisher (1925). It has widespread applications. Its purpose is to predict a single dependent variabla on the basis of one or more predictor variables and to establish whether those predictors are good predictors (Cardinal and Aitken, 2013). ANOVA is a statistical test used to analyze the difference between the means of more than two groups. A one-way ANOVA uses one independent variable, while a two-way ANOVA uses two independent variables. One-way ANOVA is used when you have collected data about one categorical independent variable and one quantitative dependent variable. The independent variable should have at least three levels. ANOVA uses the F test for statistical significance. This allows for comparison of multiple means at once, because the error is calculated for the whole set of comparisons rather than for each individual two-way comparison (which would happen with a t test). The F test compares the variance in each group mean from the overall group variance. If the variance within groups is smaller than the variance between groups, the F test will find a higher F value, and therefore a higher likelihood that the difference observed is real and not due to chance (Büyüköztürk, 2012).

3. Results

In this study, a data set of 1765 points related to Ondokuz Mayis University was used. The points have X,Y,Z coordinates.

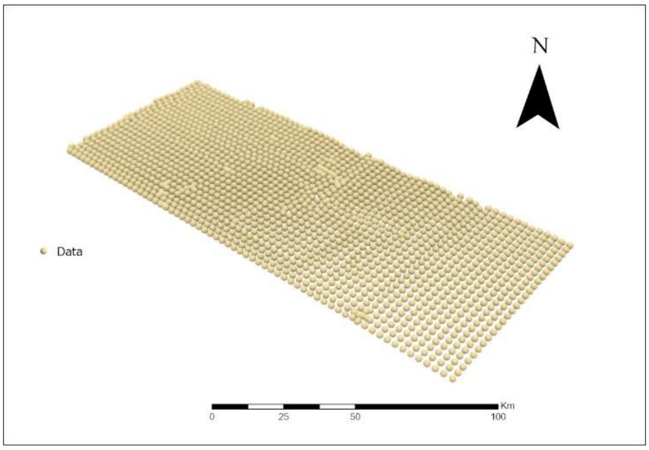

Figure 1. Data set

The data set was first tested for significance with regression analysis using the Minitab program. Polynomial degrees are tested. As a result of the regression analysis, the match and significance of the model were examined.

Then outlier measurement test was performed with the data set. The test was repeated beyond the outlier points were removed from the data set. This process was repeated until there were no outlier measurements.

3.1. The determination of the best polynomial

Firstly, the regression equation in the Minitab program was chosen as a 1st degree polynomial. The model has been found to be % 38.30 compatible. All the variables in the polynomial equation were significant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	97,00	48,5009	548,45	0,000
X	1	82,66	82,6578	934,71	0,000
Y	1	13,12	13,1179	148,34	0,000
Error	1762	155,82	0,0884		
Total	1764	252,82			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,297375	38,37%	38,30%	38,16%

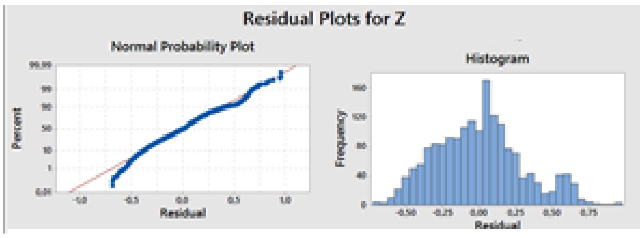

Figure 2. 1st degree polynomial regression equation

The regression equation was chosen as a 2 nd degree polynomial. The model has been found to be % 87,46 compatible. All the variables in the polynomial equation were significant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	221,202	44,2405	2461,34	0,000
X	1	84,136	84,1357	4680,93	0,000
Y	1	13,087	13,0867	728,08	0,000
X*X	1	15,929	15,9295	886,24	0,000
Y*Y	1	6,925	6,9245	385,25	0,000
X*Y	1	0,086	0,0865	4,81	0,028
Error	1759	31,617	0,0180		
Total	1764	252,819			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,134068	87,49%	87,46%	87,42%

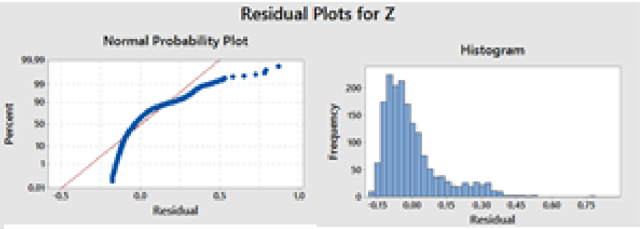

Figure 3. 2nd degree polynomial regression equation

The regression equation was chosen as a 3rd degree polynomial. The model has been found to be % 89,47 compatible. The xy^2 variable in the polynomial equation was found to be insignificant.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	9	226,338	25,1487	1666,70	0,000
X	1	14,702	14,7017	974,34	0,000
Y	1	6,590	6,5904	436,77	0,000
X*X	1	16,040	16,0401	1063,04	0,000
Y*Y	1	6,749	6,7493	447,30	0,000
X*Y	1	0,097	0,0973	6,45	0,011
X*X*X	1	0,269	0,2693	17,85	0,000
Y*Y*Y	1	0,224	0,2245	14,88	0,000
X*X*Y	1	0,108	0,1078	7,15	0,008
X*Y*Y	1	0,007	0,0074	0,49	0,484
Error	1755	26,481	0,0151		
Total	1764	252,819			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,122837	89,53%	89,47%	89,43%

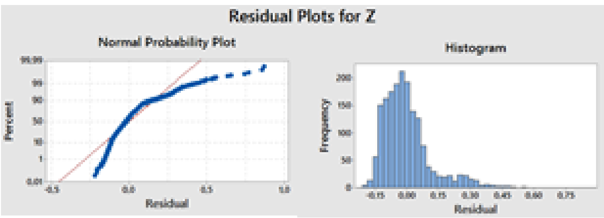

Figure 4. 3rd degree polynomial regression equation

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	226,331	28,2913	1875,53	0,000
X	1	30,058	30,0579	1992,64	0,000
Y	1	14,907	14,9072	988,25	0,000
X*X	1	16,047	16,0466	1063,78	0,000
Y*Y	1	6,749	6,7489	447,41	0,000
X*Y	1	0,099	0,0990	6,56	0,010
X*X*X	1	0,733	0,7333	48,61	0,000
Y*Y*Y	1	1,746	1,7464	115,77	0,000
X*X*Y	1	0,778	0,7778	51,56	0,000
Error	1756	26,488	0,0151		
Total	1764	252,819			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,122819	89,52%	89,48%	89,44%

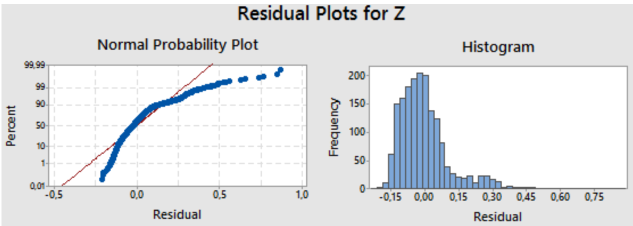

Figure 5. 3rd degree polynomial (without xy2) regression equation

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	13	231,950	17,8423	1497,05	0,000
X	1	13,180	13,1795	1105,82	0,000
Y	1	27,229	27,2292	2284,66	0,000
X*X	1	1,260	1,2604	105,76	0,000
Y*Y	1	3,774	3,7744	316,69	0,000
X*Y	1	0,023	0,0232	1,95	0,163
X*X*X	1	0,220	0,2199	18,45	0,000
Y*Y*Y	1	0,296	0,2956	24,81	0,000
X*X*Y	1	0,521	0,5206	43,68	0,000
X*X*X*X	1	0,000	0,0004	0,03	0,862
Y*Y*Y*Y	1	0,119	0,1188	9,97	0,002
X*X*X*Y	1	0,028	0,0280	2,35	0,125
X*X*Y*Y	1	0,013	0,0130	1,09	0,296
X*Y*Y*Y	1	0,021	0,0212	1,78	0,182
Error	1751	20,869	0,0119		
Total	1764	252,819			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,109171	91,75%	91,68%	91,64%

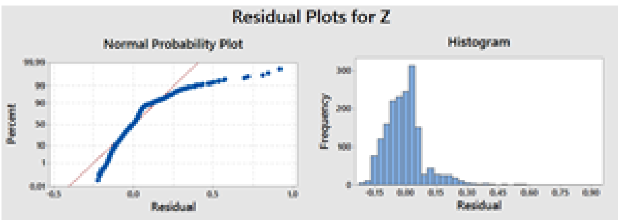

Figure 6. 4th degree polynomial regression equation

The variable xy^2 in the 3rd degree polynomial equation was removed from the equation and the regression analysis was repeated. The model has been found to be % 89,48 compatible. All the variables in the polynomial equation were significant.

The regression equation was chosen as a 4th degree polynomial. The model has been found to be % 91,68 compatible. The xy, x^2y^2, x^3y, xy^3, x^4 variable in the polynomial equation was found to be insignificant.

The variable xy, x^2y^2, x^3y, xy^3, x^4 in the 4th degree polynomial equation was removed from the equation and the regression analysis was repeated. The model has been found to be % 90,67 compatible. All the variables in the polynomial equation were significant.

3.2. Outlier measurement test

The polynomial function was obtained from 1 to 4 in order by using an adjustment solution according to the least square method. The outlier detection was realized in all solutions until there were no outlier measurements in the data. When the 1st order polynomial equation is used, 8 outlier measures are found. The root mean square error was calculated as 0.2921 m.

When the 2nd order polynomial equation is used, 353 outlier measures are found. The root mean square error was calculated as 0.0454 m.

When the 3rd order polynomial equation is used, 439 outlier measures are found. The root mean square error was calculated as 0.0266 m. When the xy^2 variable was removed from the 3rd degree polynomial equation and the outlier measurement test was repeated and 439 outlier measurement were found. The root mean square error was calculated as 0.0268 m.

When the 4th order polynomial equation is used, 403 outlier measures are found. The root mean square error was calculated as 0.0261 m. When the xy, x^3y, x^2y^2, xy^3, x^4 variables were removed from the 4th degree polynomial equation and the outlier measurement test was repeated and 284 outlier measurement were found. The root mean square error was calculated as 0.0414 m.

4. Discussion

Table 1. Significant test results

Polynomial Degree	Model fit	Insignificant Coef.
1st degree	%38,30	-
2nd degree	%87,46	-
3rd degree	%89,47	xy^2
3rd degree_withoutxy^2	%89,48	-
4th degree_withoutxy^2	%91,68	xy,x^3y,x^2y^2,xy^3,x^4

Table 2. Outlier Measurement test results and calculated mo values

Polynomial Degree	Outlier Point	m_0(m)
1st degree	8	0,2921
2nd degree	353	0,0454
3rd degree	439	0,0266
3rd degree_withoutxy^2	439	0,0268
4th degree_withoutxy^2	403	0,0261

5. Conclusion

In this study, it was tried to determine the appropriate polynomial surface model Ondokuz Mayis University in Samsun. 1765 points were used in the application. Firstly, the significance of the polynomial was tested using regression analysis. Then, an outlier measurement test was performed and m_0 was compared. As a result of these process, it was decided that it would be appropriate to use the 3rd degree polynomial with the xy^2 variable removed from the equation as the surface.

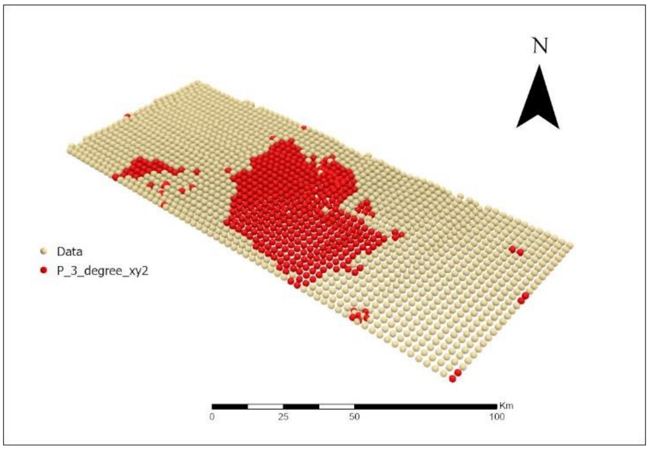

Figure 7. Outlier points in fit polynomial surface

It was observed that there was no big alteration in m_0 values after the 3rd degree. In this case, it is decided that the best suitable geoid determination function was the 6th order polynomial function for this application. The study can be improved by using different and increasing data set.

References

Akçin, H. (2001). Jeoit kavramı ve belirleme teknikleri üzerine bir inceleme. Niğde Üniversitesi Mühendislik Bilimleri Dergisi, 4(1), 37-50.

Aksoy, A. (1984). Uyuşumsuz ölçüler testi. Harita Dergisi, 93, 15-24.

Ayan, T. (1992). Uyuşumsuz ölçüler testi. Harita ve Kadastro Mühendisliği. Dergisi, (72), 38-46.

Bayrak, T. (2003). Heyelanlar için bir dinamik deformasyon ve bir dinamik hareket yüzeyi modellerinin oluşturulması. PHD Thesis, Karadeniz Teknik University, Trabzon.

Bektaş, S. (2005), Dengeleme Hesabı. Samsun: Ondokuz Mayıs Universitesi Yayınları, Samsun

Bolat, S. (2011), Lokal Jeoid Belirlemede Kullanılan Enterpolasyon Yöntemleri Samsun İli Örneği. Yüksek Lisans Tezi, Ondokuzmayıs University, Samsun.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, E. A., Karadeniz Ş. & Demirel, F. (2012). Bilimsel Araştırma Yöntemleri (11. Baskı). Ankara: Pegem Yayınları

Cardinal, R. N., & Aitken, M. R. (2013). ANOVA for the behavioral sciences researcher. Psychology Press.

Cakir, L., & Yilmaz, N. (2014). Polynomials, radial basis functions and multilayer perceptron neural network methods in local geoid determination with GPS/levelling. Measurement, 57, 148-153.

Guide to the BASIC Programming Language, (2023). This guide provides an overview of the built-in BASIC programming language available within SPM®,

Ostertagová, E. (2012). Modelling using polynomial regression. Procedia Engineering, 48, 500-506.

Sjöberg, L. E. (2023). On the topographic bias by analytical continuation in geoid determination. Studia Geophysica et Geodaetica, 67(1), 27-38. https://doi.org/10.1007/s11200-022-0337-4.

Sisman, Y., Dilaver, A., & Bektas, S. (2012). Outlier detection in 3D coordinate transformation with fuzzy logic. Acta Montanistica Slovaca, 17(1), 1-8

Sisman, Y. (2014). Coordinate transformation of cadastral maps using different adjustment methods. Journal of the Chinese Institute of Engineers, 37(7), 869-882.

Teke, K., & Yalçinkaya M. (2005). Yerel jeoid yüzeyinin belirlenmesinde kullanılan enterpolasyon yöntemleri. TUJK Düşey Datum Çalıştayı, Karadeniz Teknik Üniveristesi, Trabzon.

Uzun, Y. (2003). Üç boyutlu astrojeodezik dik koordinat sistemlerinde dönüşüm modelleri ve uyuşumsuz ölçülerin belirlenmesi yöntemlerinin karşılaştırılması. PHD Thesis, Karadeniz Teknik University, Trabzon.