國土測繪與空間資訊　民國一〇四年七月
第三卷　第二期　第 87 頁至第 98 頁
Taiwan Journal of Geoinformatics
Vol. 3, No. 2, pp.87 ~ 98

The Best Surface Fitting of Regional Geoidal Undulation- A Case Study of Taichung Area

Fang-Shii Ning[1*], Wen-Chieh Lee[2]

Received 25 December 2014
Revised 09 February 2015
Accepted 23 March 2015

Abstract

The geoidal undulation of points can be derived from orthometric height and GNSS geodetic height. The geometric method can produce geoidal undulation more accurate and faster than gravimetric method in the area with GNSS data and levelling data. In this research, we use different surface models to find out the best fitting surface based on geometric method to interpolate geoidal undulation of unknown point. The result shows quadratic surface method is the best one to solve regional geoidal undulation in terms of both costs and benefits.

Keyword: Fitting Surface, Geoidal Undulation, Orthometric Height

[1] Assistant professor , Department of Land Economics, National Chengchi University.
[2] Assistant professor , Department of Economics, National Chengchi University.
[*] Corresponding Author, TEL: +886-2- 29393091 #50741 , E-mail: fsn@nccu.edu.tw.

區域性大地起伏最佳曲面擬合之研究－以台中地區為例

甯方璽[1*] 李文傑[2]

摘 要

幾何法大地起伏可以由正高及全球導航定位系統的橢球高快速獲得，且其精度較重力法所得到之大地起伏高，因此本研究於不同的曲面模型中尋求最佳的大地起伏擬合曲面，利用此曲面能夠精確的內插出各未知點之大地起伏值，由研究結果顯示在經費及精度上考量以二次曲面為最佳。

關鍵詞：大地起伏、正高、擬合曲面

[1] 助理教授，國立政治大學地政學系。
[2] 助理教授，國立政治大學經濟學系。
[*] 通訊作者, TEL: +886-2- 29393091 #50741，E-mail: fsn@nccu.edu.tw。

1. Introduction

Taiwan is famous for its island-wide complex terrain. In the middle of the Taiwanese island, the highest point (Mt. Jade) is 3952 m at altitude. In Eastern Taiwan, the mountains a few kilometers away from the coast typically have 2000m in height. On the other side, the western part is a flat plain. Thus, the geoidal undulation in Taiwanese terrain plays an important role (You., 2006). The geoidal undulation can be computed using several techniques. For example, the computation of geoidal undulation can be done by using the numerical integration of Stokes' formula directly, fast Fourier transform, least squares collocation, spherical harmonic functions developed in a series, or by direct calculation of the difference between the ellipsoidal heights (from Global Navigation Satellite System – GNSS) and orthometric height (from spirit leveling). Relationship of the three height types is given by the equation, as shown in Eq. 1 and Figure 1:

$$H = h - N \tag{1}$$

Where H is orthometric height above the geoid; h is geodetic height above reference ellipsoid; and N is geoidal undulation.

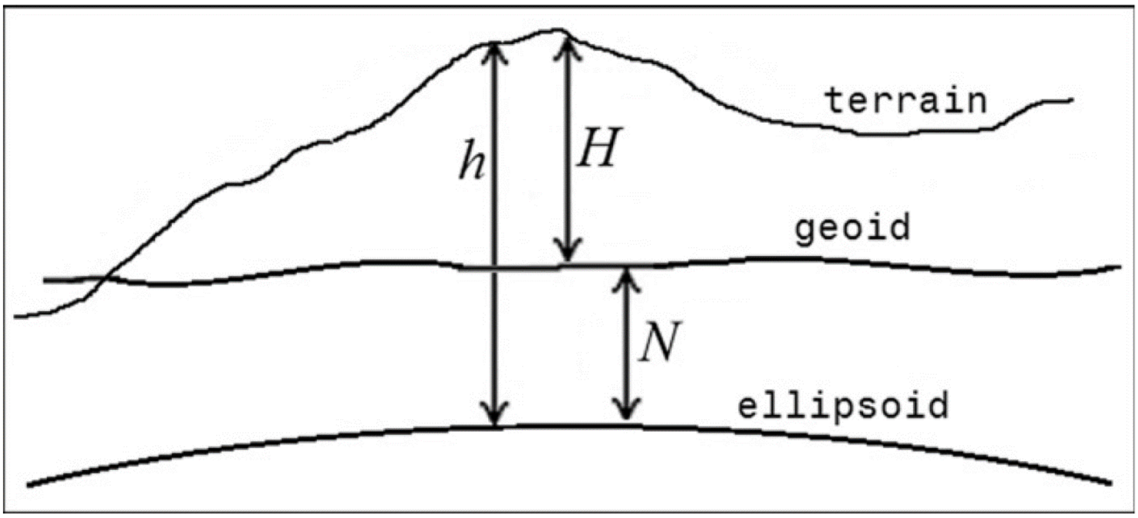


Figure 1 The geometrical relationship of three height types

Throughout this research, the geoidal undulation of points is derived from orthometric height, GNSS geodetic height, and surface model. We utilize different surface models to find out the best for Taiwan's geoidal undulation.

2. Methodology

2.1 Surface fitting

This research is first to adopt different surface models to conduct our analysis. The surface equation can be classified into many types including plane, quadratic, cubic, quartic and quintic surface (Lancaster and Salkauskas, 1986; Pottmann and Leopoldseder , 2003). The plane surface equation is shown as equation 2.

$$\mathrm{N} = a_0 + a_1 x + a_2 y + a_3 xy \tag{2}$$

Where $a_0 \sim a_3$ are unknown parameters; N is geoidal undulation; x and y are coordinates.

The plane surface contains four unknown parameters. The meaningful solution can only be found only if there are 4 points on the fitting geoidal undulation surface. The quadric surface equation (as equation 3) has 6 parameters and needs 6 points to be solved. The cubic surface equation (as equation 4) has 10 parameters and needs 10 points to be solved. The quartic surface equation (as equation 5) has 15 parameters and needs 15 points to be solved. The quintic surface equation (as equation 6) has 21 parameters and needs 21 points to be solved. Thus, the number of points within the fitting range has to be considered in selecting to solve surface equations (Awange et al., 2010) .

$$\mathrm{N} = a_0 + a_1 \mathrm{x} + a_2 y + a_3 x^2 + a_4 y^2 + a_5 \mathrm{xy} \tag{3}$$

$$\mathrm{N} = a_0 + a_1 \mathrm{x} + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 + a_6 x^3 + a_7 y^3 + a_8 x^2 y + a_9 x y^2 \tag{4}$$

$$\mathrm{N} = a_0 + a_1 \mathrm{x} + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 + a_6 x^3 + a_7 y^3 + a_8 x^2 y + a_9 x y^2 + a_{10} x^4 + a_{11} y^4 + a_{12} x^3 y + a_{13} x^2 y^2 + a_{14} x y^3 \tag{5}$$

$$\mathrm{N} = a_0 + a_1 \mathrm{x} + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 + a_6 x^3 + a_7 y^3 + a_8 x^2 y + a_9 x y^2 + a_{10} x^4 + a_{11} y^4 + a_{12} x^3 y + a_{13} x^2 y^2 + a_{14} x y^3 + a_{15} x^5 + a_{16} y^5 + a_{17} x^4 y + a_{18} x^3 y^2 + a_{19} x^2 y^3 + a_{20} x y^4 \tag{6}$$

Also, $\mathrm{a_n}$ indicates unknown parameters, N indicates geoidal undulation, and x and y indicates the components on abscissa and ordinate, respectively.

To derive perfect fitting data, the precision after fitting should require to be approximated to 0. We discuss the fitting result with simulated data, which are presented in plane equation, cubic surface and quintic surface equations, respectively. In the research, the simulation data are 9 by 9 grid points, and the z component is a random number between 0 to 1 units. The content is detailed in Table 1. As shown in Figures 2, 3 and 4, the data are fit to a plane surface with precision of 0.2723 unit. The data are fit to cubic surface with precision of 0.2703 unit. The data are fit to quintic surface with precision of 0.2637 unit. From above data and graphs, higher order surface equations may result in more fit data. However, higher order surface equations imply the risk of overfitting, that is, the prediction error is relatively high.

Table 1 Data Points of Surface Fitting Simulation (No Unit)

		Y								
		1	2	3	4	5	6	7	8	9
x	1	0.706	0.0344	0.7094	0.3404	0.5472	0.35	0.9172	0.7792	0.3112
	2	0.0318	0.4387	0.7547	0.5853	0.1386	0.1966	0.2858	0.934	0.5285
	3	0.2769	0.3816	0.276	0.2238	0.1493	0.2511	0.7572	0.1299	0.1656
	4	0.0462	0.7655	0.6797	0.7513	0.2575	0.616	0.7537	0.5688	0.602
	5	0.0971	0.7952	0.6551	0.2551	0.8407	0.4733	0.3804	0.4694	0.263
	6	0.8235	0.1869	0.1626	0.506	0.2543	0.3517	0.5678	0.0119	0.6541
	7	0.6948	0.4898	0.119	0.6991	0.8143	0.8308	0.0759	0.3371	0.6892
	8	0.3171	0.4456	0.4984	0.8909	0.2435	0.5853	0.054	0.1622	0.7482
	9	0.9502	0.6463	0.9597	0.9593	0.9293	0.5497	0.5308	0.7943	0.4505

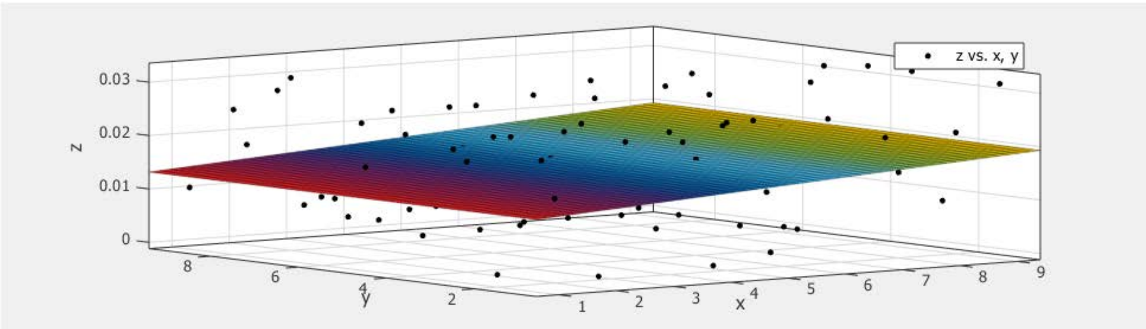


Figure 2 Result of Simulated Points from Plance Surfae Fitting with Precision of 0.2723 Unit

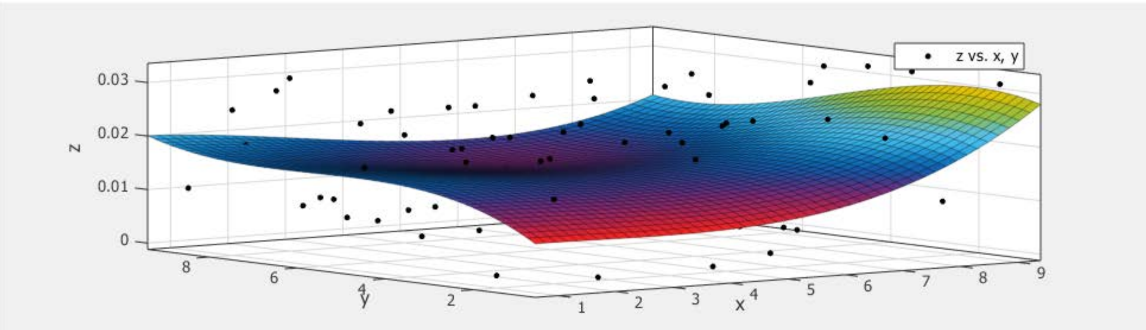


Figure 3 Result of Simulated Points from Cubic Surface Fitting with Precision of 0.2703 Unit

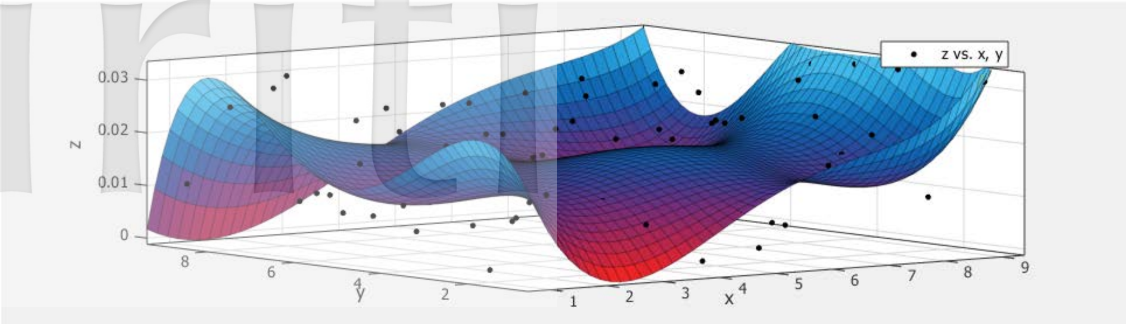


Figure 4 Result of Simulated Points of Quintic Surface with Precision of 0.2637 Unit

The comparison chart (Figure 5) shows training sample errors and test sample errors for different model complexities utilizing 100 groups of training data (there are 50 respective samples in each group of the training sets) (Hastie et al., 2009). The abscissa indicates the complexity of the model, the ordinates the prediction error, the pale blue curves indicate training errors, the reddish curves indicate test errors, and the solid lines indicate the expectation values of training errors and test errors. From the graph, higher complexity model result in lower training errors and test errors. However, as the model complexity is higher, the difference between test errors and training errors increase instead. As the complexity is increasing until the training error reaches zero, it indicates the case of overfitting for training samples.

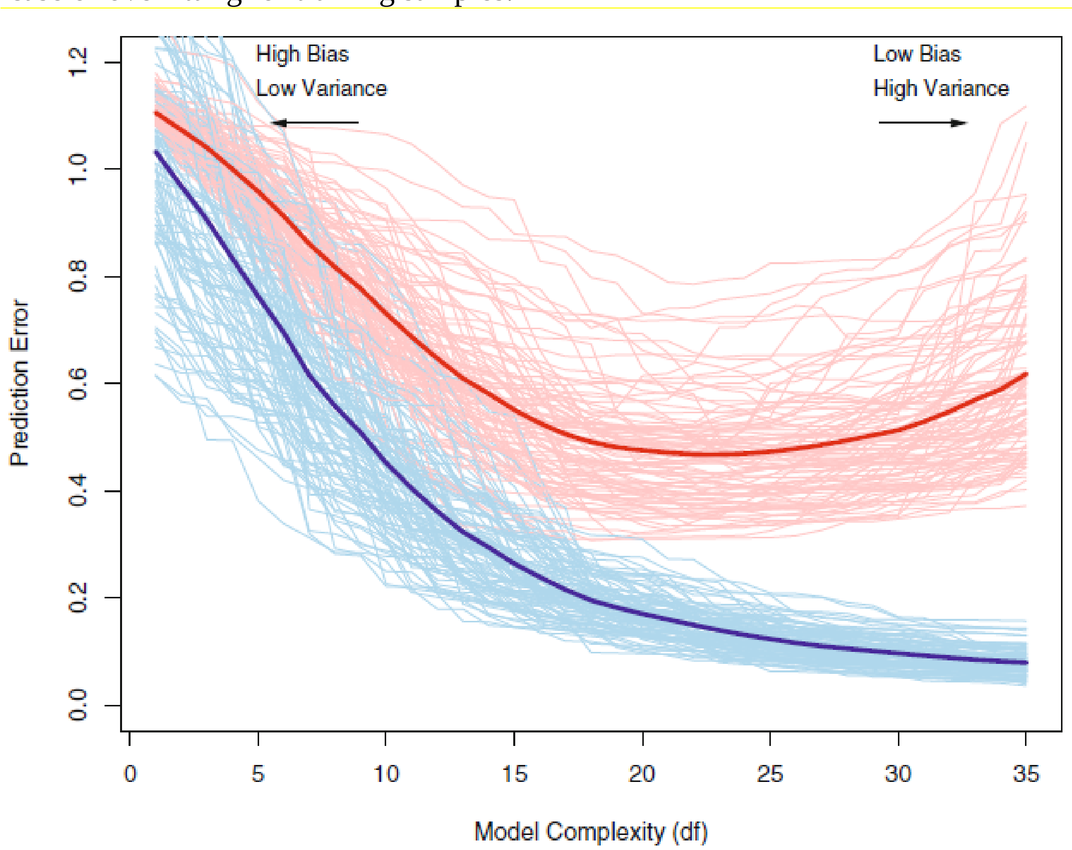


Figure 5 Training Sample Errors and Test Sample Errors under Different Model Complexities (Hastie et al., 2009)

2.2 Cross Validation

Cross validation is probably the most widely used and the easiest tool for evaluation prediction errors of model (Hastie et al., 2009). The cross validation is used to determine prediction errors of model. It classified the original data into test data and training data, followed by validating data quality with cyclic analysis and calculation. As shown in Figure 6, total 20 data are assumed for model creation. At first, all data are classified into 5 subsets, each of which has 4 data. In cross validation, one subset is used as the validation data after model creation and does not join training model for every calculation. After 5 iterations, all subsets are used as validation data to evaluate prediction errors of model. The data are classified into 5 subsets in Figure 6. Such method of classification into multiple subsets is referred to as K-fold cross validation in cross validation methods, wherein K indicates the number of subsets, which is 5 in the example.

	Part 1		Part 2		Part 3		Part 4		Part 5	
	Train		**Train**		**Train**		**Train**		**Validation**	
Step 1	*data 1*	*data 2*	*data 5*	*data 6*	*data 9*	*data 10*	*data 13*	*data 14*	*data 17*	*data 18*
	data 3	*data 4*	*data 7*	*data 8*	*data 11*	*data 12*	*data 15*	*data 16*	*data 19*	*data 20*

	Part 1		Part 2		Part 3		Part 4		Part 5	
	Train		**Train**		**Train**		**Validation**		**Train**	
Step 2	*data 1*	*data 2*	*data 5*	*data 6*	*data 9*	*data 10*	*data 13*	*data 14*	*data 17*	*data 18*
	data 3	*data 4*	*data 7*	*data 8*	*data 11*	*data 12*	*data 15*	*data 16*	*data 19*	*data 20*

	Part 1		Part 2		Part 3		Part 4		Part 5	
	Train		**Train**		**Validation**		**Train**		**Train**	
Step 3	*data 1*	*data 2*	*data 5*	*data 6*	*data 9*	*data 10*	*data 13*	*data 14*	*data 17*	*data 18*
	data 3	*data 4*	*data 7*	*data 8*	*data 11*	*data 12*	*data 15*	*data 16*	*data 19*	*data 20*

	Part 1		Part 2		Part 3		Part 4		Part 5	
	Train		**Validation**		**Train**		**Train**		**Train**	
Step 4	*data 1*	*data 2*	*data 5*	*data 6*	*data 9*	*data 10*	*data 13*	*data 14*	*data 17*	*data 18*
	data 3	*data 4*	*data 7*	*data 8*	*data 11*	*data 12*	*data 15*	*data 16*	*data 19*	*data 20*

	Part 1		Part 2		Part 3		Part 4		Part 5	
	Validation		**Train**		**Train**		**Train**		**Train**	
Step 5	*data 1*	*data 2*	*data 5*	*data 6*	*data 9*	*data 10*	*data 13*	*data 14*	*data 17*	*data 18*
	data 3	*data 4*	*data 7*	*data 8*	*data 11*	*data 12*	*data 15*	*data 16*	*data 19*	*data 20*

Figure 6 Example of Cross Validattion Flow

The cross validation is a relative conservative estimation method for evaluating prediction errors of model, and would take considerable computation time. Since advanced computing capability nowadays, the cross validation would not consume too much cost with the reasonable amount of data number and the less complex model. Thus, the cross validation may be used to determine prediction errors of model in a relatively simple manner.

In the research, LOOCV (Leave One Out Cross-Validation), one of cross validation methods, is used to determine prediction errors of surface fitting model. LOOCV is the extreme form in K-fold cross validation methods, wherein K is the total number of data. One datum is extracted to be validation datum every time, while other data are trained and iterated until all data have been used as validation data for one time (Kearns, M. and Ron, D., 1999).

The evaluation equation for prediction errors of cross validation is as shown in equation 7 (Hastie et al., 2009):

$$\mathrm{CV}(\hat{f}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}^{-i}(x_i)\right]^2} \tag{7}$$

The data are classified into N subsets; $\hat{f}$ indicates fitting surface equation, while $\hat{f}^{-i}$ is the model obtained by using the ith group of the subsets as validation data while other data are trained; y is a dependent variable, which is the value of geoidal undulation as surface fits geoidal undulation; while x is an independent variable, which is a plane coordinate as surface fits geoidal undulation. Equation 7 is the prediction error formula defined on the basis of K-fold cross validation. LOOCV is used in the research. The prediction error evaluation of LOOCV may be calculated simply by setting N as the total number of data.

3. Data and Results

In the research 78 level points in total comprising ellipsoid height surveyed in Taichung City are utilized with the point distribution profile shown in Figure 7. The data are used because the points are distributed rather uniformly in Taichung City, and are ideal data for fitting geoidal undulation. The mean geoidal undulation is 19.798 m (Maximum 20.818 m, Minimum 19.214 m).

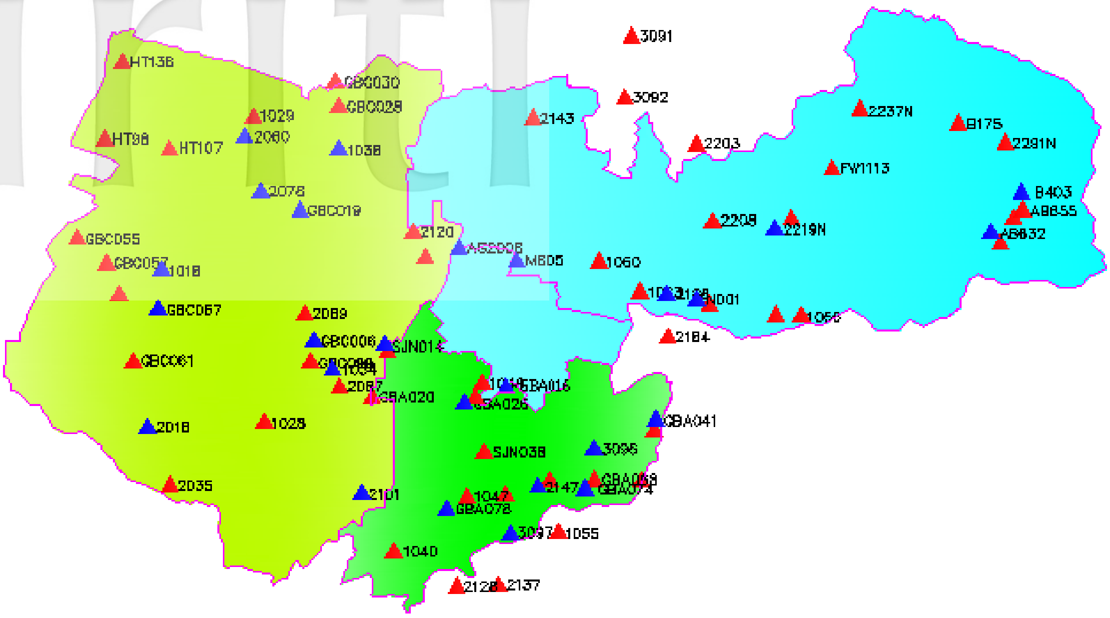


Figure 7 The distribution of leveling data

Plane, quadric, cubic, quartic and quintic surface equations are used to fit geoidal undulations in Taichung experimental area. In addition, the precision of models are evaluated by using LOOCV. The evaluation purpose of the experiment is to come out the best surface equation for fitting. Furthermore, such surface equation has to control the prediction errors in a reasonable range in order to prevent overfitting condition. The flow chart of the research is as shown in Figure 8.

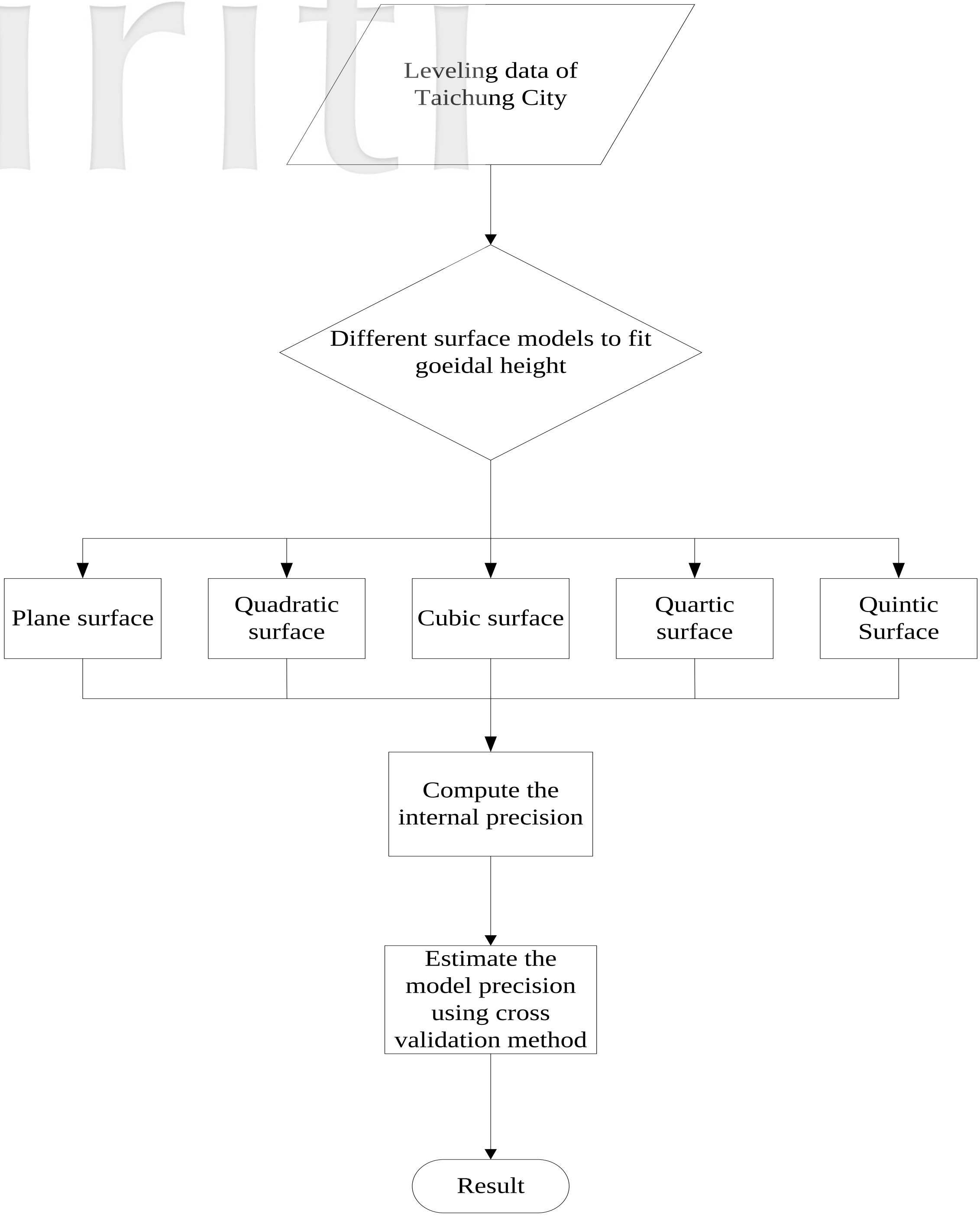
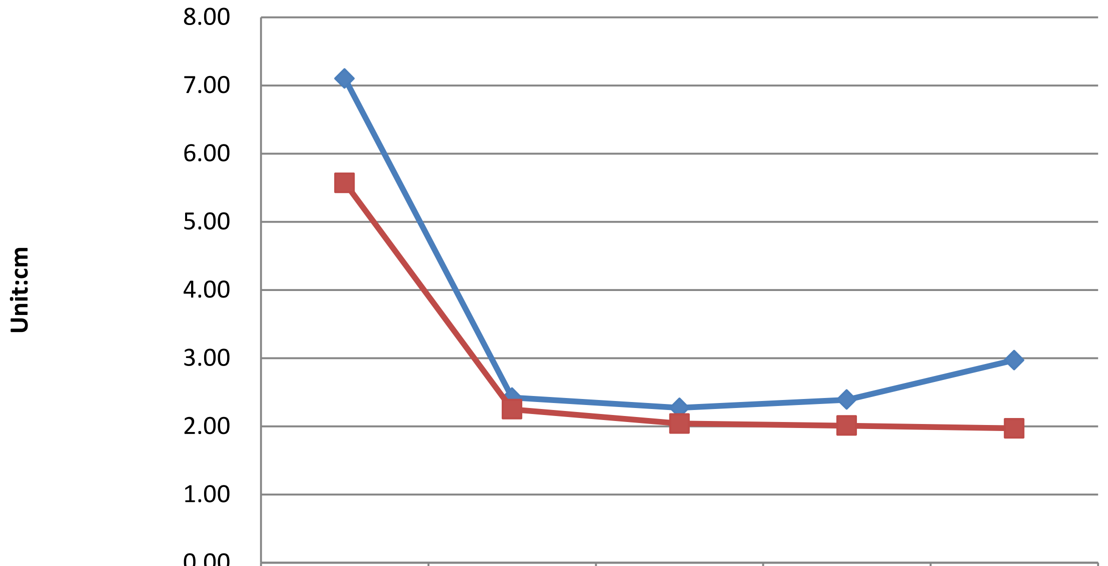


Figure 8 The flowing chart of surface's model analysis

The result of the experiments is shown in Figure 9: the model precision for the plane surface is 5.57 cm, that for the quadratic surface is 2.25cm, that for cubic surface is 2.04cm, that for quartic surface is 2.01 cm, and that for the quintic surface is 1.91cm; the prediction error (evaluated by LOOCV) for the plane surface is 7.10 cm, that for the quadratic surface is 2.42cm, that for cubic surface is 2.27cm, that for quartic surface is 2.39 cm, and that for the quintic surface is 2.97cm.

The experimental result is nearly compliant with Figure 2. As the model complex increases, the model precision decreases gradually, which might approximates to 0;

however, the prediction errors would bounce as valley is reached as the model complex decreases. Although the most complex model in the experiment is only a quintic surface equation, it is reasonable to estimate that the model precision of higher order surface equation would be higher than 1.91cm, and its prediction errors would be higher than 2.97cm. From the figure, the model precisions and prediction precisions of model complexity are most proximate between quadratic surface and quartic surface.



	Plane Surface	Quadratic Surface	Cubic Surface	Quartic Surface	Quintic Surface
LOOCV	7.10	2.42	2.27	2.39	2.97
Model Precision	5.57	2.25	2.04	2.01	1.97

Figure 9 Comparison of Prediction Errors and Internal Precisions (Taichung Experimental Area)

4. Conclusion

From the results, both training and testing error of the quartic surface equation are smaller compared with the quadratic surface equation. However, the quartic equation needs 15 necessary observations, which are far more than 6 and 10 necessary observations for the quadratic surface equation and the cubic surface equation. In view of the potential measurement errors generating from less density of levelling points distributed in full Taiwan compared with Taichung experimental area, we need 15 or more necessary observations in the fitting range to produce results with better accuracy. This is more difficult. From the obtained fitting results, the increase is not large. Thereby, there is no reason to accept the quartic surface equation. Therefore, in terms of costs and benefits, the quadratic surface method is the best one to solve regional geoidal undulation.

References

Awange, J. L., E. W. Grafarend, B. Paláncz, and P. Zaletnyik, 2010, *Algebraic Geodesy and Geoinformatics*, 2nd edition, Berlin Heidelberg : Springer-Verlag.

Hastie, T., R. Tibshirani, and J. Friedman, 2009, *The elements of statistical learning*, 2nd edition, New York: springer.

Kearns, M., and D. Ron, 1999, Algorithmic Stability and Sanity-Check Bounds for Leave-One-Out Cross-Validation, *Neural Computation*, 11(6), 1427-1453.

Lancaster, P., and K., Salkauskas, 1986, *Curve and Surface Fitting: An Introduction*, 1st edition, London: Academic Press.

Pottmann, H., and S., Leopoldseder, 2003, A concept for parametric surface fitting which avoids the parametrization problem, *Computer Aided Geometric Design*, 20(6), pp.343-362.

You, R. J., 2006, Local Geoid Improvement Using GPS and Leveling Data: Case Study, *Journal of Surveying Engineering*, 132(3), pp.101-107.