SPATIAL DATA POSITIONAL ACCURACY ANALYZING AND TESTING

ZENG Yanwei

National Quality Supervision and Testing Center of Surveying and Mapping Product.
198,Renmin Beilu Er Duan,Chengdu,610081 - zengyw@sbsm.gov.cn

KEY WORDS: DLG, GPS, Quality Control, Accuracy, Testing, Error Distribution

ABSTRACT:

In this article 1:50000 DLG data is used as an example to analyze and test spatial data positional accuracy.At first 1:50000 DLG data acquisition technical process and data content is simply introduced. Then the error possibly introduced by various working procedures is analyzed, and the positional accuracy relative to map is estimated.Positional accuracy of 1:50000 DLG data is tested based on GPS data, and data error distribution is examined.Some conclusion and proposals are given at last.

Along with Geographic Information System (GIS) is widely applied day by day, spatial data quality control becomes a hot topic in domestic and foreign GIS fundamental research. In recent years, the domestic and foreign scholars have researched broadly in many fields such as the source of the spatial data error ,the quality evaluation model, the quality measurement index, quality testing method, error distribution and so on. some fruit was obtained. Such as the problem of spatial data quality is introduced and disseminated through the data acquisition, data processing and application. Positional accuracy, attribute accuracy, logical uniformity, integrity and correctness, time accuracy and semantic accuracy etc is used to describe the quality of spatial data. Some uncertainty measurement model has been established on the basis of the positional uncertainty of point,line and area (SHI Wenzhong, LIU Dajie and so on).Statistic Model of Rate of Disfigurement for attribute data accuracy with the general sampling principle (SHI Wenzhong ,LIU Chun and so on) is put. Data quality test method is suggested based on the rule and the model(Zeng Yanwei,etc).Map scanning digitization data error doesn't always obey the Gaussain Distribution (LIU Dajie, ZENG Yanwei and so on).

The national 1:50000 geographic information database has been basically completed at the end of 2005 in China. Among them, 1:50000 Digital Line Graphic (DLG) data is one of the important contents. 1:50000 DLG uses many kinds of source including map and so on, uses many kinds of software to scan and digitize map, to renew the data, to conformity GPS country and provincial road data, to conformity place name data, etc. Therefore, the problems of whether the 1:50000 DLG data positional accuracy conforms to our country's correlated standard, whether the data error obeys the Gaussain Distribution and so on, are worth going deep into analyzing and discussing.

The paper is organized as the following.1:50000DLG data acquisition technical process is firstly introduced. Then the error possibly introduced by various working procedures is analyzed, and the positional accuracy of the result data relative to map is discussed. Positional accuracy of 1:50000 DLG data is tested based on GPS data, and data error distribution is examined.According to the analyzing and testing results, some proposals is given.

1. ERROR SOURCE ANALYSIS AND ERROR CONTROL

1.1 Error Sources Analysis

The core essential feature data of 1:50000 DLG mainly includes river system, inhabitant, road, bridge, place name and so on. The data collection used many data source including 1:50000 map, 1:50000 Digital Raster Map (DRG), 1:50000 Digital Elevation Model (DEM), Arial Photogrammetry Digital Orthograph Map(APDOM), SPOT satellite image, LandSat image, 1:250000 GPS road differential data, 1:50000 place name database, 1:50000 river system name and code, 1:250000 DLG and other referencing material. Main technical process including map scanning and processing, image correction and processing, feature gathering and renewal, GPS road data conformity, place name data input and so on.

Looked from the data process, the error source of the positional accuracy includes the following factor.

1. The map inherent error, the primary factor is refer to the paper map distort, the map feature itself influence, the original map clarity influence.

2. Scanning error. Because of the scanning machine error adds on paper map distort, map scanning may have the bigger scanning error, the scanning error is the main error source of map scanning digitization.

3. Image processing error, which is mainly from geometry adjusting of map scanning image and the renewal image.

4. Map orientation error, which is influenced by the orientation mathematical model error and the control point error.

5. Data acquisition and renewal error, which mainly include operator's sampling error and error introduced by software.

6. Editing error. Data editing may eliminate or reduce the local error or the obvious error in the data acquisition and renewal process. Although it is important to insure the achievement quality, simultaneously also possibly introduces new error.

1.2 Error Control

In view of the many data source, complex data process, much error source and so on,strict error control measurement in the 1:50000 DLG data process is used to insure the result data accuracy.

1.2.1 Control of the Data Source Error:

1. Choosing maps which satisfy the request, guaranteing map scanning accuracy.

2. Controlling image process error. After map image and renewal image correction are completed, geometry correction accuracy must be tested and evaluated.

1.2.2 Controlling Data Gathering and Renewing Error:

1.Controlling map orientation error. Orientation accuracy is the basis of guarantee for accuracy of digitized data.When orientation is completed,orientation accuracy must be tested and evaluated.

2. Controlling screen digitizing error.Data acquisition and the renewal are completed on the screen.Screen digitizing error control is the key to guarante the accuracy of result data. They mainly include to choose software conformed the request map scanning digitization, make the unification work scheme (including layer, feature class,line width, line color, an attribute items definition and so on).Line tracking parameter, the density of collected points, the joint distance, hanging distance and so on are given before.Data is processed under the zooming conditions.

3.Controlling the error in edgematch to insure the continuity and consistency of geometry position of the same feature.

1.2.3 Testing and Evaluation of Gathering Accuracy：

During data collection, gathering accuracy is tested based on DOM,DRG etc.If it doesn’t accord with requirement,data must be modified. Some statistical result can be found in Table1.

From table 1,the gathering and renewing accuracy satisfied the demands.

Accuracy Type	Classification		Positional Accuracy	Standard
Gathering Accuracy Based on DRG	Flat,upland		±2.51m	±5 m
	Mountanious Region		±2.75 m	±5 m
Renewing Accuracy	APDOM	Flat,upland	±2.48 m	±5 m
		Mountanious Region	±1.64 m	±5 m
	RSDOM	Flat,upland	±7.57 m	±1 0 m
		Mountanious Region	±5.31 m	±1 0 m

Table 1. Data Gathering Accuracy Testing Result Statistical

1.2.4 Tolerance of the Positional Accuracy

We can know from the analysis above that the main factor which influence positional accuracy of 1:50000 DLG data includes scanning error, image processing error, map orientation error, screen digitization error.When DLG and DRG are overlayed in screen to test accuracy of DLG, map orientation error and screen digitization error are the main error source.When map scanning resolution takes 300DPI,screen digitization error is 0.08mm.When the map orientation error is 0.10mm, the standard difference of positional accuracy is

$$M_0 = \pm\sqrt{(0.1^2 + 0.08^2)} = \pm 0.13\text{mm}.$$

Gaining 1:50000 DLG based on map scanning digitization,the minimum value of DRG resolution is 317.5DPI.We control the map orientation accuracy within 0.10mm.We should strictly control the process data error and check data positional accuracy.The inspection statistic result of 291 maps is given in table 2. It can be infered from table 2 that all the data position accuracy is within standard difference.

Max value	Min value	0～0.66 M_0	0.66 $M_0 \sim M_0$
0.13mm	0.03mm	249	42

Table 2 Inspection Result Statistical table of Capture Accuracy

2 TESTING OF POSITIONAL ACCURACY

To get testing point coordinates,GPS control point of E class is firstly surveyed.Then RTK technique is applied to gain plane and elevation data based on those GPS control points. Positional accuracy (x,y,h) is computed from coordinate difference (Δx , Δy , Δh)of the same point by formula (1).Computing result is listed in table 3.

It can be known from table 3 that all data positional accuracy is within standard difference.

$$\sigma_x = \pm\sqrt{\frac{\sum_1^n \Delta x_i^2}{n}} \quad , \quad \sigma_y = \pm\sqrt{\frac{\sum_1^n \Delta y_i^2}{n}} \qquad (1)$$

$$\sigma_h = \pm\sqrt{\frac{\sum_1^n \Delta h_i^2}{n}}$$

sequence number	terrain type	Horizontal Accuracy	testing point number	Vertical accuracy	testing point number
1	Flat	±22.7m	192	±1.8m	136
2	Flat	±21.4m	149	±2.3m	107
3	Hilly terrain	±19.4m	25	--	--
4	Hilly terrain	±13.7m	41	--	--
5	Hilly terrain	±22.6m	20	--	--

Table3 Accuracy Testing Result Statistics

3 ERROR DISTRIBUTION TESTING

Literature [10] adopt four geographic limits point to rectify the scanning map，on the basis of the experimental data , the accuracy of map scanning digitizing and error distribution are discussed, and it considered that scanning digitizing error not always submit Gaussain Distribution.Literature [15]] draw the academic reseau of a 1:50000 map on the film ,then scan the film map with 400DPI.Consequently, reseau cross coordinate data is obtained by the way of linear mode on screen digitizing manually, data accuracy is discussed respectively with the single-point registration digitalization ,four point orientation correction and nine point orientation correction with Affine Transformation mode at both case of map not rectifying and map precisely rectifying geometrically one by one.The results indicate that data error disobeys Gaussain Distribution. This paper adopts GPS data to test accuracy of 1:50000 DLG data，the following part will examine the distribution of data error.

3.1 Normal Distribution Testing

Use χ^2 testing method to check if data error of 1:50000 DLG data submits to normal distribution.

Original hypothesis H_0 : $E(x) = E_0(x)$, obey normal distribution; standby hypothesis H_1 : $E(x) \neq E_0(x)$, don't obey normal distribution, and $E_0(x)$ obey normal distribution.

Strictly speaking, positional accuracy is 25m relative to control point, vertical accuracy is about 2.5m.According to error distributing, carve up coordinate error of X and Y direction to K sections, do some statistics to frequency n_i ($i = 1,2,3,...,k$)of error which drop into the inter-zone, and calculate average and variance of coordinate error. If H_0 is right, according to every section median, use normal distribution function to calculate p_i and np_i , make statistic parameter χ^2 as follows.

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \qquad (2)$$

When calculating horizontal position accuracy, points within 3 times of root mean square error must be contained. Therefore,the plane coordinates data error according to the 15m gap, the altitude error according to the 1m gap, the division is 10 sectors. Assigning the remarkable level α =0.01， we calculate the test statistics χ_x^2 and χ_y^2 according to the X coordinate error and the Y coordinate error.The results and $\chi^2_{k-2-1}(\alpha)$ which looks up in the distribution table named χ^2 are listed in table 4 and table 5.

Sample ID	$\chi_x^{\ 2}$	$\chi^2_{k-2-1}(\alpha)(x)$	$\chi_y^{\ 2}$	$\chi^2_{k-2-1}(\alpha)(y)$
1	91.212	23.2	69.498	23.2
2	58.0786	23.2	41.430	23.2

Table 4 The testing result table of the plane data error distribution

Sample ID	$\chi_h^{\ 2}$	$\chi^2_{k-2-1}(\alpha)(h)$
1	17.859	18.5
2	23.266	18.5

Table 5 The testing result table of the elevation data error distribution

From table 4, the sample 1 and sample 2 reject the original supposition, it indicates that neither X coordinate error or the Y coordinate error obey Gaussain Distribution .From table 5,the sample 1 accepts the original supposition, it indicates that the elevation data error obey the Gaussain Distribution. The sample 2 rejects the original supposition, it indicates that the elevation data error disobey the Gaussain Distribution.

3.2 Mean Testing

According to the formula (1) and (2), the mean of the sample data coordinates error ($\overline{x}$, $\overline{y}$, $\overline{h}$),the root mean square error (σ_x 、 σ_y 、 σ_h)and the covariance σ_{xy} are calculated and listed in table 6 and 7

$$\overline{x} = \frac{\sum_1^n \Delta x_i}{n}, \quad \overline{y} = \frac{\sum_1^n \Delta y_i}{n}, \quad \overline{h} = \frac{\sum_1^n \Delta h_i}{n},$$

$$\sigma_{xy} = \frac{\sum_1^n \Delta x_i \Delta y_i}{n} \tag{2}$$

Sample ID	Point Number	$\overline{x}$ (m)	$\overline{y}$ (m)	σ_x (m)	σ_y (m)	σ_{xy}
1	191	1.167	1.976	15.482	14.620	3.084
2	149	-3.253	4.106	14.292	15.797	4.426
3	25	-6.292	-0.591	13.699	13.771	3.783
4	41	1.361	-1.605	11.516	13.904	7.423
5	20	-5.345	4.162	17.808	13.902	10.123

Table 6 The Horizontal Coordinate Error Statistical Result

Sample ID	Point Count(unit)	$\overline{h}$ (m)	σ_h (m)
1	136	0.324	1.829
2	107	-1.57	2.339

Table 7 The Vertical Coordinate Error Statistical Result

From table 6 and table 7, all means of sample map coordinate error are not zero, in other words, there are system error in the data. According to the analysis, the error possibly comes from the residual error after the geometry rectification of map scanning image,map paper orientation error , vectorization software error,etc.

In order to test the influence of system error in the process of scanning digitizing, the supposition testing for mean sample data error by the method of nonnormal ensemble average testing is used. Supposing the distribution function of collectivity is $F(Z)$, ensemble mean is Z, ensemble standard deviation is S, the inspection statistics is constructed according to the following formula.

$$\mu_{\bar{z}} = \frac{\bar{Z} - E(Z)}{S / \sqrt{n}} \tag{3}$$

While n is so big，sample average value of arbitrary Z will approximately obey the Gaussian distribution N (0,1).The original supposition H_0 ： $E(Z) = 0$ .The prepared supposition H_1： $E(Z) \neq 0$ .Under the remarkable level α , when H_0 is true, and that n is big enough, there are:

.

$$P\left(\frac{\left|\bar{Z} - E(Z)\right|}{S / \sqrt{n}} \geq \mu_{1-\frac{\alpha}{2}} \right) = \alpha \tag{4}$$

If

$$\mu_{\bar{z}} = \frac{\left|\bar{Z}\right|}{S / \sqrt{n}} \geq \mu_{1-\frac{\alpha}{2}} \tag{5}$$

H_0 is rejected in remarkable level α ,otherwise it is received.

The sub- sample average value $\bar{x}$ 、 $\bar{y}$ 、 $\bar{h}$ is used to replace the collectivity average value $\bar{Z}$, the sub- sample mean square error σ_x 、 σ_y 、 σ_h is used to replace the collectivity standard error S , and statistic parameter of X Coordinate error, Y Coordinates error, elevation ($\mu_{\bar{x}}$ 、 $\mu_{\bar{y}}$ 、 $\mu_{\bar{h}}$) is calculated respectively and listed in table 8

Map Serial number	Total points	$\mu_{\bar{x}}$	$\mu_{\bar{y}}$	$\mu_{\bar{h}}$
1	191	1.04	1.87	7.441
2	149	9.671	3.173	5.117
3	25	2.296	0.214	--
4	41	0.756	0.734	--
5	20	1.342	1.339	--

Table 8 The Testing Result of Mean Coordinate Error

When the testing remarkable level α is 0.01, the rank value $U_{0.995} = 2.576$ by looking up the table. From table 8, among the five sample map, all map horizontal coordinate error obeys the Gaussain distribution except the second one. The two testing map's elevation error doesn't obey the Gaussian distribution.

4. CONCLUDING REMARK

Referencing data acquisition technical process of 1:50000 DLG, the possibly introduced errors in every process is analyzed, the main factor which may influence data accuracy is discussed and the positional accuracy of DLG data relative to map is estimated. Then GPS data is used to check the horizontal position accuracy of five maps, the vertical accuracy of two maps. From testing results, both the horizontal position accuracy and the vertical accuracy of the 1:50000 DLG data satisfy the standard requirements. The testing distribution of data error indicates that the data error disobey the Gaussain Distribution, in other words, the data exists system error. Therefore, it is necessary to go deep into researching of the digitizing data errors' distribution rule and the method of quality testing and accuracy evaluation,which will provide foundation for correlative standard's constitution.

REFERENCES

Caspary W,Scheuring.Error-bands as measures of geometrical accuracy. Proceedings of EGIS,1992:226～233.

Chrisman N R.Methods of spatial analysis based on error in categorical maps:[phD Thesis].Bristol:University of Bristol, 1982.

DAI Hong-lei.The Theory and Method of Positional Uncertainty Measurement and Propogatiing in GIS [PhD Thesis].Wuhan:Wuhan University of Surveying and Mapping Techology,2000. (in Chinese)

DU Dao-sheng,WANG Zhan-hong,MA Cong-li.Study on Spatial Data Quality Model.Journal of Image and Graphics. 2000，5(7)：559-562.(in Chinese)

LIU Wen-bao.GIS Spatial Data Uncertainty Theory.Wuhan: Wuhan Technical University of Surveying and Mapping, 1995.(in Chinese)

LI Da-jun,Gong Jian-ya,Xie Gang-sheng.Determination of Point Positional Uncertainty Entropy Index.Geotechnical Investgation and Surveying. 2002(2):33～36.(in Chinese)

LI Da-jun,Gong Jian-ya,Xie Gang-sheng. Research on Entropy Uncertainty Band of Linear feature in GIS. Geomatics and Information Science of Wuhan University, 2002，27(5):462～466.(in Chinese)

LI Da-jun,GONG Jian-ya,XIE Gang-sheng,DU Dao-sheng. The Model of Error Entropy of Area Unit inGIS.ACTA GEODAETICA et CARTOGRAPHICA SINICA, 2003，32(1):31～35.(in Chinese)

LIU Da-jie,SHI Wen-zhong,TONG Xiao-hua.Accuracy Analysis and Quality Control of Spatial Data in GIS[M]. Shanghai Science Literature Press. 1999，pp：68～71.(in Chinese)

LIU Chun,SHI Wen-zhong, LIU Da-jie.Statistical Model of Rate of Disfigurement for Attribute Data Accuracy Assessment in GIS. ACTA GEODAETICA et CARTOGRAPHICA SINICA, 2003，32(1):36～41.(in Chinese)

LI Qin-hai,TAO Ben-zao. Probability Statistics Principle and Its Application inSurvey. Beijing:Surveying and Mapping Press,1980..(in Chinese)

ZENG Yan-wei.Positonal Accuracy Evaluation of Vector GIS Data.Journal of Geomatics. 2002，27(3)：20～22.(in Chinese)

SHI Wen-zhong.Theory and Methods for Handling Errors inSpatial Data[M].Beijing:Science Press,1998.(in Chinese)

XU Jin-zhao. symbolic Statistics and Product Quality.Beijing:China Specification Press,1986.(in Chinese)

ZENG Yan-wei.Accuracy Analyzing and Error Distribution Testing of Map Scanning Digitization. Surveying and Mapping of Sichuan. 2006，29(4)：159～162.(in Chinese)

ZENG Yan-wei. Discussion of Orientation Accuracy of Map Scanning Vectorization.Bulletin of Surveying and Mapping. 2004(5)：49～52.(in Chinese)

ZENG Yan-wei. Research on Spatial Data Quality Control and Evaluation Technique System[phD Thesis]. Wuhan University,2004.(in Chinese)