Statistical Tests of the Distribution of Errors in Manually Digitized Cartographic Lines

Xiaolin Meng*, Wenzhong Shi* and Dajie Liu†

*Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong
†Department of Surveying and Geomatics Engineering, Tongji University
Shanghai, P. R. China

Abstract
This paper presents a study of error distribution of GIS data from manual digitization based on experiments. First, an experiment is conducted by digitizing selected features on a cadastral map by several operators. The experimental data sets with random error are generated by eliminating the effect of the systematic error and blunder error. Second, several statistical tests are conducted to analyze the statistical distribution of the map digitization error. It is found that, in the said conditions, the nature of manual digitization error is different from the normal distribution but closer to a new random error distribution — NL distribution. Finally, the functions of the NL distribution are given.

I. INTRODUCTION

Data capture is one of the most essential steps in developing a GIS. This is not just because the data capture is the most expensive. Furthermore, the GIS-based decision making is very much dependent on the quality of the data in GIS. Error modeling for GIS data, therefore, has been identified as one of focus research issue in GIS research society and been investigated by many researches (Goodchild et al. 1992; Hunter et al. 1996; Shi 1997). According to the errors introduced to a GIS, the methods for capturing data for GIS can be classified as direct and indirect methods (Shi 1994). The direct method of spatial data capture refers to deriving data directly from the field (such as from GPS or a total station). On the other hand, an indirect method refers to the processes in which data are derived from existing document, such as maps, charts, and graphs etc. The advantages of indirect methods are faster and less expensive compared with the direct data collection methods. Within the indirect methods, manual digitization is the common one for GIS data capture.

People gradually realize the importance of data quality to a GIS, especially for the decision making from the difficulties and failure examples of GIS applications. Data capture, such as manual digitizing, has been recognized as a significant error source in GIS data generation (Keefer et. al. 1988). On the other hand, there is still room for further understanding the nature of error in manual digitization. This paper presents a research nature of error from statical test point of view.

Because the complex objects in GIS, such as lines and polygons, are composed of points. Error distribution of points is thus the most essential in dealing with data quality in GIS. There are three kinds of error existed in map digitization, i.e. systematic error, random error and gross error. Systematic error and blunder can be identified or even removed by using appropriated mathematical models and data snooping techniques. This study focuses on analyzing characteristics of random error in GIS data capture.

Bolstad and et al. (1990) studied the characteristics of error distribution for point in manual digitization by an empirical method. They found that statistical distribution of signed distance deviation was Leptokurtic relative to a random normal variant. Walsby (1993) studied the reasons and effects of error in manual digitizing to a GIS.

Caspary and Scheuring (1993) showed the methods of describing the accuracy of geometric data digitized from existing maps via statistic derivation. Under the assumptions of independent and normally distributed coordinate errors with the expectation of zero, the distribution of positional errors was derived.

The *p*-norm distribution (Sun 1995) is a more general distribution. When *p* equals to 1 or 2, the corresponding distribution is Laplace distribution or Normal distribution.

The selection of error model for data processing is very much dependent on the nature of error distribution. For example, when error obeys normal distribution, we can use the least squares adjustment to deal with random error of the observations. On the other hand, we might not be able to get the optimal solution by the least squares adjustment when the observation error does not obey normal distribution. Therefore, the study of the error distribution is a fundamental issue for error modeling for GIS data.

1082-4006/98/0401~2-52$3.00

It is generally acceptable that the error of the GIS data follows normal distribution (Hunter et al 1996 and Shi 1998). This is true, especially when there are many error sources that contribute to the error of GIS data and the value of these errors are similar to each. This study will discuss the nature of the one of these error sources — digitization. Under certain conditions, for example the experimental conditions of this study, the error distribution may not follow normal distribution.

In the following, we will first describe the generation of a test data set of manual digitization and then test the nature of the error distribution by various test methods.

II. NORMALITY FITNESS TESTS OF MANUAL MAP DIGITIZATION ERRORS

The Experiment of Digitization

An experiment was conducted to investigate the nature of manual digitization error. Three 50cm by 50cm mylar cadastral maps with the known analytic coordinates were used, which strictly followed cartography specifications. The map scale was 1:500. The experiments were carried out within a laboratory with constant temperature and humidity. The deformation rate of the mylar map media was less than 0.2 ⁰/oo after a special anti-deformation processing.

A Summagrid III digitizer with nominal precision of 0.01mm was used for digitization. This nominal precision has been tested before this experiment and proved that the precision of the digitizer can reach the nominal level. Each of the three operators digitized one cadastral map independently. A total of nine examples were obtained. All the coordinates of the features were obtained within three days. During the experiment, the maps were not removed or remounted from the digitizer.

On these maps, the analytic coordinates of control points were known. All the analytic coordinates of feature points, such as well, post, building concerns etc., were measured by a total station in field. These were considered as the 'true' values. The field measured data were then input into a computer system by using the field data collection software. The captured digital cadastral data can be further output via computer cartography software.

The observation points were independent points, such as well, post and building concerns. Point mode digitizing was used. Error between points was considered as independent to each other.

To eliminate the effect of systematic errors due to map deformation, the affine coordinate transformation was used. The transformation with 6 unknown parameters was used to transfer the digitized coordinates to Gauss plane coordinates. Meng, Cao and Liu (1996) explained the reason why the affine transformation was selected among various coordinate transformation methods. Four corner points on each map were used to calculate transformation parameters. The RMS of the coordinate transformation was 0.056mm. By using these transformation parameters with the precision calculated, the observations of each digitized point were rectified. By comparing the 'true' value of the coordinates with the observations digitized and transformed coordinates, the error of the coordinates in X and Y directions were obtained.

By conducting correlation hypothesis tests, i.e., the student test, the independence in X and Y directions was investigated. The average correlation parameter for nine sets of sample equal 0.15. This indicates that error in X and Y direction is weakly correlated. For the shake data processing, we assumed the error was independent in X and Y directions.

To limit influence of the blunders within the observations to the final results, the tolerance of 0.5m of ground value was settled. Any coordinate with error larger than 0.5m, would be rejected in the observations as a blunder. This was because that 0.5m equivalent to 1mm on the 1:500 cadastral map. This was the distance between features on map distinguishable by the operators' eyes.

After pre-processing (coordinate transformation, systematic error correction and blunder processing) to the observations, the nature of the errors of the observations was considered as random in nature.

By using modern statistical analyses, the error distribution of manual map digitization was visualized in Figure 1 (The unit of X axis is mm). From Figure 1, it can be seen that the error distribution of map digitization falls between the Normal distribution and Laplace distribution. Here, we define this combined distribution as NL distribution. In the following, the

Normality tests, such as combined Kurtosis and Skewness test and Kolmogorov-Smirnov test, were used for investigating the nature of the statistical distribution of the random error.

Combined Kurtosis and Skewness Tests of Normality

Kurtosis and Skewness test is often used to test normality of a set of observed samples. Suppose random variable X obey normal distribution, the coefficients of Kurtosis and Skewness are r_1 and r_2 (Fang et al. 1987).

$$r_1 = B_4 / B_2^2 - 3 \quad (1)$$

$$r_2 = B_3 / B_2^{3/2} \quad (2)$$

where B_i (i=2, 3, 4) is the *ith* order central moment. When the sample comes from a normal general population, the estimated values of r_1 and r_2 should equal to zero. As one of two parameters is larger than zero, we can reach the conclusion that the sample is different from the normal distribution. But this approach has the default of lacking of clear statistic significance. Combined Kurtosis and Skewness tests were founded on following premise, that is when the general population obeys normal distribution and the sample volume n is larger than 100, the estimated values of kurtosis and skewness coefficients obey normal distribution approximately (Pan et al. 1993).

$$r_1 \sim N(0, 6/n) \quad (3)$$

$$r_2 \sim N(0, 24/n) \quad (4)$$

Under the significant level α and in case of

$$\mu_{r_1} = \sqrt{\frac{n}{6}} * | r_1 | > \mu_{1-\alpha} \quad (5)$$

or

$$\mu_{r_2} = \sqrt{\frac{n}{24}} * | r_2 | > \mu_{1-\alpha} \quad (6)$$

We can say the sample does not come from a normal

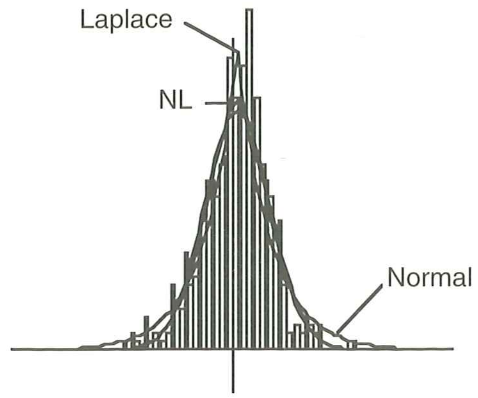

Figure 1. A comparison of Laplace and NL and Normal Distribution with Histogram of manual map digitization.

distributed population. Table 1 is the results of the Kurtosis and Skewness tests and the combined tests (where α=0.05 and $\mu_{1-\alpha} = 1.645$). Comparing the statistic value calculated with the critical value 1.645, we can easily find that the error distribution of map digitization does not obey a normal distribution. It might obey some other random distribution.

Kolmogorov-Smirnov Test of Normality

The basis of Kolmogorov-Smirnov test lies on the empirical distribution (Fang et al. 1987). It can be used to test whether a set of samples comes from a known distribution $F_0(x)$.

Assume $Z_1, Z_2, ..., Z_n$ are the sample points from an unknown distribution $F(z)$ and we want to test following hypotheses:
null hypothesis $H_0 : F(z) = F_0(z)$;
alternative hypothesis $H_1 : F(z) \neq F_0(z)$.

The empirical distribution is denoted by $F_n(z)$ and the statistical value is by D_n^*.

Table 1. The results of Kurtosis and Skewness tests

S. N.	S. S.	r_{1x}	r_{2x}	r_{1y}	r_{2y}	μ_{1x}	μ_{2x}	μ_{1y}	μ_{2y}
1	448	0.34	1.79	0.95	5.43	2.96	7.74	8.26	23.5
2	333	-0.25	0.97	-0.34	3.06	1.92	3.63	2.6	11.4
3	379	0.90	3.87	-0.46	3.74	7.21	15.4	3.73	14.9
4	559	-0.41	4.98	-0.78	3.94	4.02	24.1	7.62	19.2
5	427	0.85	4.38	0.79	3.41	7.19	18.5	6.67	14.4
6	660	0.83	5.41	0.03	0.46	8.78	28.4	0.35	2.00

* r_{ix} is the calculated coefficient of X direction (i=1,2).
r_{iy} is the calculated coefficient of Y direction (i=1,2).
μ_{ix} is the calculated coefficient of X direction (i=1,2).
μ_{iy} is the calculated coefficient of X direction (i=1,2).
S. N. is the sample group number
S. V. is the sample size

$$D_n^* = \sup_{-\infty<z<\infty} |F_n(z) = F_0(z)| \quad (7)$$

In above equation, *Sup* is the sign of suprenum of an expression. The critical value corresponding to the significant level α is $D_n^*(\alpha)$. When $D_n^* < D_n^*(\alpha)$, the null hypothesis is accepted, i.e., the sample $Z_1, Z_2, ..., Z_n$ comes from a known distribution $F_0(x)$. Table 2 is the results of normality Kolmogorov-Smirnov test of 6 sets of samples. From the analysis of Table 2, we can find that each statistical value D_n^* is larger than the critical value $D_n^*(0.05)$. This indicates that random errors of manual map digitization do not obey normal distribution under significant level α.

III. LAPLACE DISTRIBUTION TESTS OF MAP DIGITIZATION ERRORS

From Figure 1, we can see that Laplace distribution closer to the error distribution of map digitization. Thus, it is necessary to further verify whether the error distribution of map digitization obeys Laplace distribution. According to numerical characteristic of Laplace distribution, the authors conducted Kurtosis and Skewness tests and coefficient formula was derived.

$$r_1 = B_4 / B_2^2 - 6 \quad (8)$$

$$r_2 = B_3 / B_2^{3/2} \quad (9)$$

The calculated results are listed in Table 3.

The meaning of the symbols in Table 3 is the same as those in Table 2. From Table 3, we can also find that the random error distribution of manual map digitization is different from that of Laplace.

IV. NL DISTRIBUTION AND THE PROBABILITY DISTRIBUTION TABLE

Based on the above statistical analyses, we can find that the error distribution of the map digitization does not obey normal or Laplace distribution. By analyzing Figure 1, we can see that the error distribution of map digitization is a distribution between the two distributions: normal and Laplace distributions. By a mathematical derivation, we established NL distribution function that is the combination of the Normal and Laplace distributions.

Table 2. The results of Kolmogorov normality test (a=0.05)

S. N.	S. S.	D_{nx}^*	D_{ny}^*	$D_n^* = 1.22/\sqrt{n}$
1	448	0.105	0.079	0.058
2	333	0.098	0.084	0.067
3	379	0.068	0.105	0.062
4	559	0.097	0.090	0.052
5	427	0.103	0.076	0.059
6	660	0.084	0.056	0.052

* D_{nx}^* is the sample statistic value in X direction
D_{ny}^* is the sample statistic value in Y direction

Table 3. The results of Laplace Kurtosis and Skewness tests

S. N.	S. S.	r_{1x}	r_{2x}	r_{1y}	r_{2y}
1	448	0.34	4.21	0.96	-0.56
2	333	-0.25	-5.02	-0.34	-2.93
3	379	0.90	-2.12	-0.47	3.75
4	559	-0.42	-1.01	-0.79	-2.06
5	427	0.85	-1.61	0.79	-2.58
6	660	0.84	-0.58	0.03	-5.53

Assume: $X \sim N(\mu, \sigma^2)$, $Y \sim La(\alpha, \beta)$
where, X and Y are the normal and Laplace random variables respectively, $\mu, \sigma, \alpha, \beta$ are the position and scale parameters of the two density functions ((11) and (12)). We further assume,

$$Z = X + Y \quad (10)$$

Z is the random variable which obeys NL distribution.

The Normal distribution function of random variable X is

$$f_1(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (11)$$

and the Laplace distribution function of random variable Y is

$$f_2(y) = \frac{1}{2\beta} \exp\left\{-\frac{| y - \alpha |}{\beta}\right\} \quad (12)$$

By using convolution integral, we have the NL distribution function $\varphi(z)$.

$$\varphi(z) = \int_{-\infty}^{\infty} f_2(y) * f_1(z-y) dy \quad (13)$$

We can approve that the NL distribution function is described by (14). The detailed description of the formula is in the appendix.

$$\varphi(z) = A * \Phi((z - \mu - \alpha)/\sigma - \sigma/\beta) + B * \Phi(-(z - \mu - \alpha)/\sigma - \sigma/\beta) \quad (14)$$

where

$$A = \exp(-(z - \mu - \alpha)/\beta + \sigma^2/2\beta^2)/2\beta \quad (15)$$

and

$$B = \exp((z - \mu - \alpha)/\beta + \sigma^2/2\beta^2)/2\beta \quad (16)$$

In (14), $\Phi(\cdot)$ is the normal probability distribution. Through the mathematical derivation, we can also approve that the mathematical expectation and variance of NL random variable are:

$$E(z) = \mu + \alpha \quad (17)$$

$$Var(z)=2\beta^2+\alpha^2 \quad (18)$$

When $\alpha=0, \mu=0, 2\beta^2=1/2, \sigma^2=1/2$, we have E(z)=0, Var(z) = 1 and

$$Z \sim \text{NL}(0, 1) \quad (19)$$

The distribution of an NL random variable with mean 0 and variance 1 is called standard NL distribution. The density function of the standard NL distribution is described by the formula:

$$\varphi(z)=e\times(e^{-2z}\Phi(\sqrt{2}z-\sqrt{2})+e^{2z}\Phi(-\sqrt{2}z-\sqrt{2})) \quad (20)$$

The probability expression of standard NL distribution is

$$F(z)=P(Z\le z)=\int_{-\infty}^{z}\varphi(t)dt \quad (21)$$

Combining equation (20) and equation (21), we have

$$F(z)=P(Z\le z)$$

$$=\int_{-\infty}^{z} e\times(e^{-2t}\Phi(\sqrt{2}t-\sqrt{2})+e^{2t}\Phi(-\sqrt{2}t-\sqrt{2}))dt \quad (22)$$

By using numerical integral calculus, we can calculate the probability of standard NL distribution (see Table 4). From Table 4, we can find that when the probability of NL distribution equals 0.955 and 0.997 respectively. The corresponding critical value z equals 1.7 and 3.2 when the critical values of normal distribution are 2 and 3, that of Laplace are 2.12 and 4.10 respectively. From the probability calculation of the three distributions, we can also find that NL distribution is situated between the latter two distributions and sharper than the normal distribution. These results coincide with the error histogram of map digitization in Figure 1. For further investigating the characteristics of NL distribution, p-norm probability distribution are compared with the NL distribution (Table 5). The density function of p-norm distribution is equation 23 (Sun 1995).

$$f(x)=\frac{p^{(1-1/p)}}{2\Gamma(\frac{1}{p})}e^{-\frac{1}{p}|x|^p} \quad (23)$$

From Table 5, we can easily find that NL distribution is higher identical with p-norm distribution when p is located in the interval [1.6, 1.75]. It has different probability at the same critical value with the normal distribution (p=2) and Laplace distribution (p=1). This is the same as the results of the following Chi-square and Kolmogorov-Simirnov tests.

The reason why p-norm distribution is compared with NL distribution is because p-norm adjustment is based on solid theoretic foundation and simple for programming. If we can substitute NL adjustment for p-norm adjustment, many regular features in cadastral map can be corrected by adjustment and relevant quality index can also been got from the adjustment process.

V. THE CHI-SQUARE TEST OF NL DISTRIBUTION

Assume the random variable Z has the unknown distribution function F(z), and $z_1, z_2, ..., z_n$ are sample points with sample volume n. The purpose of Chi-square test lies on testing whether the unknown distribution is identical to the known distribution $F_0(z)$.

The hypotheses of Chi-square test are:
null hypothesis $H_0: F(z)=F_0(z)$;
alternative hypothesis $H_1: F(z)\neq F_0(z)$.

For simplifying distribution test, we divide the X-axis into s intervals and use v_i to represent the sample number which falls into the interval i. When v_i< 5, we combine the adjacent intervals as one interval. The unknown parameter l of $F_0(z)$ equals to 2 (position and scale parameters). $\bar{Z}$ and S^2 are the estimation values of parameter μ and σ^2.

$$\bar{Z}=\frac{1}{n}\sum_{i=1}^{s}Z_i \quad (24)$$

$$S^2=\frac{1}{n}\sum_{i=1}^{s}(Z_i'-\bar{Z})^2v_i \quad (25)$$

Where Z_i' usually equals $(Z_i+Z_{i+1})/2$. Under the condition of null hypotheses, the theoretical probability is p_i (equation 26).

$$p_i=P(Z_i<z\le Z_{i+1})=F_0(Z_{i+1})-F(Z_i) \quad (26)$$

Table 4. NL Probability Distribution

z	F(z)	z	F(z)
0. 0	0. 5000	2. 0	0. 9756
0. 1	0. 5427	2. 1	0. 9799
0. 2	0. 5848	2. 2	0. 9835
0. 3	0. 6259	2. 3	0. 9864
0. 4	0. 6654	2. 4	0. 9889
0. 5	0. 7030	2. 5	0. 9909
0. 6	0. 7383	2. 6	0. 9925
0. 7	0. 7711	2. 7	0. 9939
0. 8	0. 8011	2. 8	0. 9950
0. 9	0. 8284	2. 9	0. 9959
1. 0	0. 8529	3. 0	0. 9967
1. 1	0. 8746	3. 1	0. 9972
1. 2	0. 8937	3. 2	0. 9978
1. 3	0. 9104	3. 3	0. 9982
1. 4	0. 9248	3. 4	0. 9985
1. 5	0. 9372	3. 5	0. 9988
1. 6	0. 9477	3. 6	0. 9990
1. 7	0. 9566	3. 7	0. 9992
1. 8	0. 9641	3. 8	0. 9993
1. 9	0. 9703	3. 9	0. 9995

Table 5. The comparison of NL distribution and p-norm distribution

z	p=2	p=1. 75	P=1. 7	p=1. 65	p=1. 6	p=1. 5	p=1. 4	p=1. 35	p=1. 3	p=1	NL
0. 0	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000	0. 5000
0. 5	0. 6915	0. 6919	0. 6921	0. 6921	0. 6924	0. 6928	0. 6934	0. 6936	0. 6940	0. 6967	0. 7030
0. 7	0. 7581	0. 7564	0. 7561	0. 7557	0. 7554	0. 7548	0. 7542	0. 7539	0. 7538	0. 7517	0. 7711
1. 0	0. 8414	0. 8360	0. 8349	0. 8336	0. 8325	0. 8301	0. 8276	0. 8263	0. 8249	0. 8161	0. 8529
1. 5	0. 9332	0. 9242	0. 9223	0. 9202	0. 9183	0. 9140	0. 9096	0. 9072	0. 9047	0. 8884	0. 9372
2. 0	0. 9773	0. 9696	0. 9678	0. 9658	0. 9640	0. 9598	0. 9553	0. 9529	0. 9503	0. 9323	0. 9756
2. 5	0. 9938	0. 9893	0. 9882	0. 9868	0. 9855	0. 9825	0. 9791	0. 9772	0. 9751	0. 9590	0. 9909
3. 0	0. 9987	0. 9967	0. 9962	0. 9954	0. 9947	0. 9929	0. 9907	0. 9894	0. 9879	0. 9751	0. 9967
3. 5	0. 9998	0. 9991	0. 9989	0. 9985	0. 9982	0. 9973	0. 9961	0. 9952	0. 9943	0. 9849	0. 9988
3. 9	1. 0000	0. 9997	0. 9996	0. 9994	0. 9993	0. 9989	0. 9981	0. 9976	0. 9970	0. 9899	0. 9995

Note: The probability equals 0. 9995. The corresponding critical values of the above distributions are 3.3, 3.7, 3. 8, 4. 0, 4. 1, 4. 3, 4. 6, 4. 8, 5. 0, 6. 9 and 3. 9.

where i = 1, 2, . . . s.

Assume $Z_0=-\infty$ and $Z_s=\infty$; We have $P(Z_0)=0, P(Z_s)=1$. Each interval¢s probability equals np_i, i=1, 2, . . . s. The statistical value of Chi-square test is χ^2(Fang et al. 1987).

$$\chi^2=\sum_{i=1}^{s}\frac{(v_i-np_i)^2}{np_i} \quad (27)$$

Under the significant level α (α=0.01), we can look up $\chi^2_{s-l-1}(\alpha)$ from a Chi-square table (Fang 1987). When statistical value χ^2 is larger than $\chi^2_{s-l-1}(\alpha)$, then we reject null hypothesis H_0 and vice versa. The test results are listed in Table 6. Table 7 is the results of normality test that can be used as the reference of Chi-square test.

From Table 6 and Table 7, we can reach the conclusion that the error distribution of manual map digitization obeys the NL distribution and not the Normal distribution. For the further investigating the random error characteristics of map digitization, we also used Kolmogorov-Smirnov test (α=0.01).

VI. THE KOLMOGOROV-SMIRNOV TEST OF NL DISTRIBUTION

Table 8 is the results of Kolmogorov-Smirnov test of 9 sets of samples. In Table 8, the interval number equals 20 and significant level equals 0.01. With those conditions, the critical value $D_n^*(0.01)$ of Kolmogorov-Smirnov equals 0.36. From the analysis of Table 8, each statistical value D_n^* is less than critical value $D_n^*(0.01)$, i.e., the random errors of manual map digitization obey NL distribution.

VII. CONCLUSIONS

This paper presented a study to the error distribution of digitization for GIS data capture using statis-

Table 6. The Chi-square test of NL distribution

Interval	Interval average	Frequency	Theoretical probability	Theoretical frequency np_i
0.50~0.40	0.45	2	0.0000	
0.40~0.30	0.35	11	0.0018	5.314
0.30~0.20	0.25	47	0.0147	43.394
0.20~0.10	0.15	295	0.1089	321.473
0.10~0.00	0.05	958	0.3319	979.769
0.00~-0.10	-0.05	1239	0.3956	1167.811
-0.10~-0.20	-0.15	338	0.1270	374.904
-0.20~-0.30	-0.25	59	0.0173	51.069
-0.30~-0.40	-0.35	5	0.0028	8.266
-0.40~-0.50	-0.45	3	0.0000	
sum		2952	1.0000	2952.000

Note: α=0. 01, s=8, l=2, S=0.0937, $\bar{Z}$=-0.0075,S=0.0937, χ^2=14.586, $\chi^2_5(0.01)=15.09$, $\chi^2>\chi^2_5(0.01)$.

Table 7. The Chi-square Test of Normal Distribution

Interval	Interval average	Frequency	Theoretical probability	Theoretical frequency np_i
0. 50~0. 40	0. 45	2	0. 0000	
0. 40~0. 30	0. 35	11	0. 0000	
0. 30~0. 20	0. 25	47	0. 0132	38. 966
0. 20~0. 10	0. 15	295	0. 1098	324. 129
0. 10~0. 00	0. 05	958	0. 3410	1006. 927
0. 00~-0. 10	-0. 05	1239	0. 3699	1091. 940
-0. 10~-0. 20	-0. 15	338	0. 1453	428. 925
-0. 20~-0. 30	-0. 25	59	0. 0207	61. 106
-0. 30~-0. 40	-0. 35	5	0. 0000	
-0. 40~-0. 50	-0. 45	3	0. 0000	
sum		2952	1. 0000	2952.000

Note: α=0. 01, s=6, l=2, S=0.0937, $\bar{Z}$=-0.0075,S=0.0937, χ^2=56.508, $\chi^2_3(0.01) = 6.25$, $\chi^2 > \chi^2_3(0.01)$.

Table 8. Kolmogorov-Smirnov Test of NL Distribution (α=0. 01)

Map Serial	1		2		3	
Digitizer	x	y	x	y	x	y
1	0.262	0.264	0.225	0.309	0.291	0.243
2	0.207	0.283	0.275	0.335	0.264	0.254
3	0.266	0.258	0.252	0.287	0.281	0.206

tical test approach. Normal distribution is a normally acceptable distribution for describing error in GIS data. In this study, however, it was found that error of digitization is closer to another newly developed distribution — NL distribution under the experimental conditions. The functions of the NL distribution were also derived. The corresponding error models need to be redeveloped based on the NL distribution for handling error in measured GIS data.

ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong SAR (Project No. HKP65/95E) and research grant from The Hong Kong Polytechnic University (Project No. G-S 407) and and National Natural Sciences Foundation of China (Project No. 49671065). The authors would like to express thanks to Professor Z.H. Zhu and Dr. D.G. Cao and Mr. H.C. Fan for their assistance.

REFERENCES

[1] Bolstad, P. V., Gessler, P. and Lillesand, T., 1990, Positional Uncertainty in Manually Digitized Map Data, *International Journal of Geographical Information Systems*, Vol. 4, No. 4, pp. 399~412.

[2] Caspary, W. and Scheuring, R., 1993, Positional Accuracy in Spatial Databases, *Comput., Environ. and Urban systems*, Vol. 17, pp.103~110.

[3] Fang, K. and Xu, J., 1987, *Statistical Distribution*, Science Press, China.

[4] Goodchild, M. F., Sun, Guoqing and Yang, Shiren, 1992, Development and Test of an Error Model for Categorical Data, *International Journal of Geographical Information Systems*, Vol. 6, No. 2, pp.87~104.

[5] Hunter, G. J. and Goodchild, M. F., 1996, A New Model for Handling Vector Data Uncertainty in Geographic Information Systems, *URISA Journal*, Vol. 8, no. 1, pp. 51~57.

[6] Meng, Xiaolin, Cao, Dongguo and Liu, Dajie, 1996, The Statistic test of the Error Distribution on Map Digitization, In: Ma,W.J.(Ed) , *Proceedings of PhD Candidate of Tongji University,* Tongji University Press, pp.145~150.

[7] Meng, Xiaolin and Liu, Dajie, 1996b, NL distribution of Surveying Errors, In: Ou J.K.(Ed.), *New Development of Geodesy*, Surveying and Mapping Press, China, pp. 81~85.

[8] Pan, C., and He, Y., 1993, *Principle and Methodology of Statistics*, Tongji University Press, P. R. China.

[9] Shi, W. Z., 1994, *Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and GIS*, ITC publication, No. 22, Enschede, pp.148.

[10] Shi, W. Z., 1998, A Generic Statistical Approach for Modeling Error of Geometric Feature in GIS, *International Journal of Geographical Information Science*.Vol.12, No.2, 1998, pp.131-143.

[11] Sun, H., 1995, *The Theory of the P-th Norm Distribution and Application in the Surveying Data Processing*, Ph.D. Thesis, Wuhan Technical University of Surveying and Mapping, China.

[12] Walsby, J., 1995, The Causes and Effects of Manual Digitizing on Error Creation in Data Input to GIS, In: P. Fisher (Ed), *Innovations in GIS 2*, Taylor & Francis.

Spatial Structure of Accommodation Costs in the Madison Area

Suh-Cem Pang* and A-Xing Zhu†

* Water Chemistry Program, University of Wisconsin-Madison. 660 North Park Street, Madison, WI 53706

†Department of Geography, University of Wisconsin-Madison. 384 Science Hall, 550 North Park Street, Madison, WI 53706-1491

Abstract

The spatial structure of accommodation costs in the Madison areas with respect to the direct distance from the University of Wisconsin-Madison campus was investigated using GIS and spreadsheet software. Major factors influencing the spatial variation of accommodation costs were discussed.

The spatial structure of accommodation costs indicated that the preferred and convenient locations for students were within two kilometres from the campus boundary. Factors such as proximity to social amenities, commercial centres, recreation parks and accessibility generally appeared to have significant impacts on the spatial structure of accommodation costs beyond two kilometres distance from the campus boundary.

I. INTRODUCTION

The University of Wisconsin-Madison has a student population of about 37,000 which comprises about 10% of international students from countries all over the world [1]. The high demand for student housing and accommodation each academic year has always exceeded supply. Existing hostels and dormitories and other university residential facilities located on campus can only accommodate about 6,800 students each academic year [2]. All of these residential facilities are normally fully occupied, and indeed, far from being adequate in accommodating the large student population. Fortunately, the proximity of UW-campus to the Madison City downtown areas has greatly alleviated the pressing accommodation needs of UW students. Rental services provided by many real estate agencies have made available a full range of off-campus accommodation and housing units within the vicinity of the campus to cater for the various needs of students. Various accommodation types ranging from dormitory room, efficiency to one- to more than three-bedroom apartment units are available for lease to students.

Affordability, location and safety are reckoned to be the foremost important factors that each student seeking accommodation invariably need to take into consideration before deciding on any particular unit [3]. One of the major concerns of a student prior to arriving at Madison was to secure a suitable and affordable accommodation. Most students, particularly those who have travelled half way round the globe to Madison have shared similar daunting and exasperating experience in their search for suitable accommodation. None of them feel assured, lest very much disheartened, on being informed by the University Housing Office that no accommodation could be arranged or guaranteed to new students due to limited space in existing housing facilities on campus. Being a newcomer to Madison, the search for a suitable and affordable accommodation proved to be overwhelming and frustrating, despite the compiled list of vacant accommodation units that was made available by the Campus Assistant Centre. In their eagerness to settle-in promptly, most new students would just grab the first available accommodation units that were seemingly within their budget limits.

Given the above scenario, it is therefore imperative for the relevant university authorities to provide adequate information and appropriate advisory services pertaining to the accommodation needs of newcomers. The main objective of the present study was to determine the spatial structure of off-campus accommodation costs in the Madison areas using GIS. Results from this preliminary study yield baseline information on the basic relationship between accommodation types and costs, and distance from the UW campus, which could be a useful basis for further in-depth studies.

1082-4006/98/0401~2-59$3.00