Accuracy Estimation of DTM Using High Sampling Density Profiles

Zhang Yaonan

Zhengzhou Institute of Surveying and Mapping
Zhengzhou,P.R.C.
Commission III

1.Introduction

A digital terrain model(DTM) represents a topographic surface in terms of a set of spatial coordinates.In sampling the surface to establish a DTM,we often face on such problems:the determining of adequate sampling density in order to meet a given specifications and the evaluating of the accuracy of the obtained DTM. How accurately a topographic surface is represented by a DTM depends essentially on the such several factors:sampling density, measuring errors,interpolation method and terrian classification, etc.It is important to develop a method of the accuracy estimation and find the relationship between the accuracy and the these factors.

In the attempt to solve the problem,many experimental and theoretical investigations have been completed.F.Ackermann[1],from his experimental research work,worked out that the main factor determining the accuracy of a DTM is the acquisition of the data, the height accuracy could be described approximately as a linear function of the average sampling interval.In 1972,B.Makarovic[2] presented the concepts of fidelity and transfer function,the transfer function can be computed according to the fidelity of different sampling density.It allows for comparative evaluation of different interpolation procedures,and also be used for planning purposes.K.Tempfli[3] applied the theory of spectral analysis to the problem of accuracy estimation.A theoretical formula was developed,inwhich the topographic surface was described by its spectra and measuring errors were considered.Jacobi and Frederiksen[4] also used the spectra to present the topographic relief.The spectra was showed in a Log/Log coordinates system,and was approximated by a straight line.The accuracy estimation was made based on the line.

Theoretical accuracy studies,which wre based on simplifying assumptions such as homogeneity of the terrain relief,have a limited significance since these assumptions are hardly ever fulfilled.A more general approach is desirable.This article will present a method in which the high sampling density profiles are used to complete the evaluation of the accuracy of a DTM.The experiment has been carried out to judge the feasibility of the approach.The effect of measuring error are also discussed.

2.method

First of all,we present a theoratical formula for computing the interpolated height accuracy.

Based on the theory of spectral analysis, we can develop a fomula in case of one dimension as follows:

$$\sigma^2=\frac{1}{L}\int_0^{\infty}\left(\left(A(f)-R(H^*(f))\right)^2+\left(I(H^*(f))\right)^2\right)|F_1(f)|^2\,df \qquad (1)$$

$$\overset{*}{H}(f)=\frac{F(f)}{F_1(f)} \qquad (2)$$

where: σ^2: variance of interpolated heights.

2L: the length of sampling profile.

$F_1(f)$: Fourier transformation of the discrete sampling heights.

A(f): the transfer function of the given interpolation method.

$R(\overset{*}{H}(f))$: the real part of $\overset{*}{H}(f)$.

$I(\overset{*}{H}(f))$: the imagenary part of $\overset{*}{H}(f)$.

F(f): Fourier transformation of the "true" topographic profile.

The formula (1) is different from the one of F.Tempfli.The reason is that F.Tempfli's formula was developed under the condition that sampling interval Δx must meet the sampling theorem,which formula (1) has no such limit.

From the spatial coordinates obtained by sampling the surface and the given interpolation method,A(f) and $F_1(f)$ can be determined.In getting H(f),F(f) is desirable.But it is impossible to get F(f) because the "true" terrain is unknown in most cases. F(f) can be estimated from $F_1(f)$.From the knowledge of spectral analysis,we know that $F_1(f)$ is deformed in the range of $f\geq\frac{1}{2\Delta x}$. Therefore,only the low frequency($f<\frac{1}{2\Delta x}$) information of F(f) can be obtained from the discrete sampling height.If we want to produce σ^2 accurately,the spectra of topographic surface with higher freqency range is needed.This lead to our idea of using high sampling density profiles.

Let h_{ij}(i=0,1,...M-1,j=0,1,...N-1) to be a grid DTM with the sampling interval Δx,which has been designed to meet an application.A direct method to get high frequency information is to increase the sampling density,that is to say,to decrease sampling interval Δx to $\frac{1}{2}\Delta x$.But in this way the total sampling points will increase four times than normal sampling procedure.It is not allowable for most applications.Instead, we can increase the sampling density only on several profiles distributed on the whole area of the terrain(called high sampling density profiles).The accuracy estimation is first made on these profiles.Then the total accuracy can be obtained according to the estimation of the profiles.In this approach,it is obvious that only little sampling work is added.

Surpose that terrain can be regarded as a homogenous surface. Let m_{00}^2 to be the variance of terrain relief,m_0^2 to be the variance of profile relief.Then we can get:

$$m_{00}^2=m_0^2 \qquad (3)$$

Though above formula is under the condition of homogeneity,it tell us that the variance of surface relief can be predicted using the variance of profile relief.We can anticipate that there is a relationship between the interpolation accuracy of profiles and the one of surface.It has been proved by the experiments that the relation between u_1(interpolation accuracy of profiles) and u_2(the interpolation accuracy of surface) can be expressed as follows:

$$u_2=\sqrt{\lambda}\,u_1 \qquad (4)$$

$$\text{or} \quad u_2^2=\lambda\,u_1^2 \qquad (5)$$

Usually the topographic surface is not homogenous surface, so in formula (5) the u_1^2 is always taken by the average u_1^2 of several profiles.

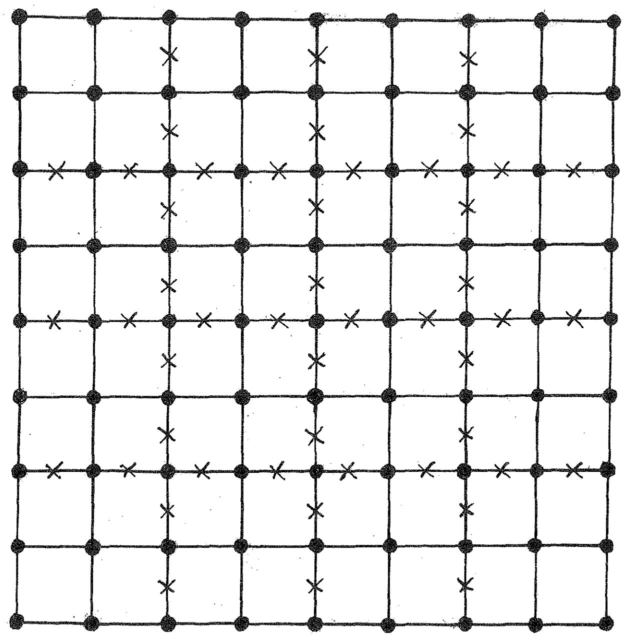

Fig.1

As showed in Fig.1, point • is the normal sampling point with the sampling interval Δx, corresponding height can be expressed as h_{ij} $(i=0,1,\cdots M-1, j=0,1,\cdots N-1)$. The profiles containing point × (point × is additional sampling point) are high sampling density profiles. The total number of high sampling density profiles is q. Each high sampling density profile contains N_l points $(l=1,2,\cdots q)$. The corresponding sampling interval in these profiles is Δ $(\Delta=\Delta x/2)$. It means that in these profiles, the half of points are the normal sampling points, and other half are additional sampling points. The additional sampling points are used to as checking pointsfor the computation of profile accuracy. The one dimension interpolation is applied to each high sampling density profile in the sampling interval of Δx. The mean error of l^{th} high sampling density profiles is calculated by the formula (6).

$$u_{1,l}^2=\frac{1}{N_l}\sum_{j=1}^{N_l}\Delta_j^2 \qquad (6)$$
$$(l=1,2,\cdots q)$$

where Δ_j is the difference between the interpolated height and "true" height.

u_1^2 results in the average of $u_{1,l}^2$

$$u_1^2=\frac{1}{q}\sum_{l=1}^{q}u_{1,l}^2 \qquad (7)$$

One kind of arrangement of high sampling density profiles has been shown in Fig.1, Fig.2 gives another example of arranging high sampling profiles. This mode has strong control on the central part of the terrain, but weak on the edge. In our experiments, we has chosed the mode showed in Fig.1 to arrang the high sampling profiles. In this mode, high sampling density profiles coincide with some of normal sampling profiles. Therefore, the additional sampling needs only little work.

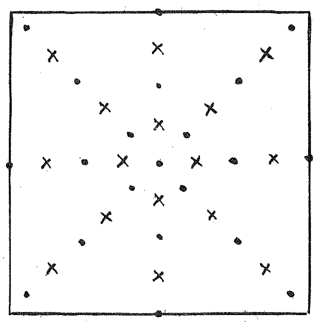

Fig. 2

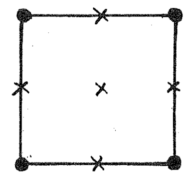

Fig. 3

We can see that, the smaller the Δ (interval of high density profiles), the higher frequency information can we get. But with the reduction of Δ, the total number of sampling points will increase. If Δ is chosed to be $\frac{1}{2}\Delta x$, the information in the range $f<\frac{1}{\Delta x}$ can be obtained, the useful frequency range is twice as large as in normal sampling procedure. It has been proved that $\Delta=\frac{1}{2}\Delta x$ is a suitable choice for accuracy estimation.

u_2 is defined as follows

$$u_2^2=\frac{1}{M_x N_y}\sum_{i=0}^{M_x-1}\sum_{j=0}^{N_y-1}(f_{ij}-\hat{f}_{ij})^2 \qquad (8)$$

where

f_{ij} is "true" height.
$\hat{f}_{ij}$ is interpolated height.

$M_x=2M$
$N_y=2N$

that is to say, in each grid, there are 4 checking points, as showed in Fig. 3.

3. Experiment and Result

In our experiments, there are two kind of original data: modelling data and real data. Modelling data is expressed as D1, D2, D3, D4. The real data is expressed as T1, T2, T3. T1, T2, T3 were measured on Topocart B. The Model scale was 1:5000, while photo-scale was 1:18000. Both T1 and T2 are 41 in line and 41 in row, with sampling interval of 5 meter. T3 is a 41 x 41 DTM, with sampling interval of 20 meter. In the original data, some of the data were used as the known heights, other were regarded as checking points.

It has been proved by the test that when $\lambda=1.5$, the optimal result could be got.

So formula (5) can be written as

$$u_2^2=1.5u_1^2 \qquad (9)$$

3.1 the influence of q (the number of high sampling density profiles)

We have chosed several different q to investigate the influence of q, the result is showed in table 1. In table 1, we use

$$e=\left|\frac{\hat{u}_2-u_2}{u_2}\right|$$

to evaluate how $\hat{u}_2$ is close to u_2.

surface	u_2	q=10		q=6		q=2	
		$\hat{u}_2$	e(%)	$\hat{u}_2$	e(%)	$\hat{u}_2$	e(%)
D1	0.1398	0.1415	1.2	0.1434	2.6	0.1465	4.8
D3	0.1401	0.1415	1.0	0.1435	2.4	0.1465	4.6
D2	0.1219	0.1286	5.5	0.1258	3.2	0.1331	9.2
D4	0.1218	0.1288	5.7	0.1257	3.2	0.1331	9.3
T1	0.5873	0.5677	3.3	0.6020	2.5	0.4770	18.8
T2	0.5514	0.5577	1.0	0.5630	2.1	0.4740	14.0
T3	1.2485	1.2169	2.5	1.0298	17.5	0.8340	33.2

Table 1
Influence of q

As to D1 and D3,there are no evident difference between the e, The same situation occurs on D2 and D4,when q is 6 and 10 respectively.But it doesn't so when q=2.When q is 6 and 10 respectively, e is almost same as to T1 and T2.But q=2 and q=6 produce quitely different e.In the case of T3,q=6 and q=10 correspond to the 17.5% and 33.2% of e respectively.Modelling surface is homogenous surface,but T1 and T2 are not such surfaces.In the case of T3,left part is reservior,the right part is mountainous region.We can conclude that we can use less number of high sampling density profiles for homogenous terrain ,but more number for other kind of terrain.

3.2 Inluence of different Δx

surface		D1	D2	D3	D4
$\Delta x=0.2$	u_2	0.0744	0.0744	0.0745	0.0744
	$\hat{u}_2$	0.0753	0.0751	0.0753	0.0751
	e	1.2%	0.9%	1.1%	1.0%
$\Delta x=0.4$	u_2	0.1398	0.1219	0.1401	0.1218
	$\hat{u}_2$	0.1415	0.1286	0.1415	0.1288
	e	1.2%	5.5%	1.0%	5.7%
$\Delta x=0.8$	u_2	0.2350	0.2779	0.2350	0.2778
	$\hat{u}_2$	0.2239	0.2467	0.2240	0.2472
	e	4.7%	11.2%	4.7%	11.0%

Table 2
Influence of different Δx

From the table 2,we can see that with the increasing of Δx,e slight increases, so it can be used for different sampling interval.

3.3 Comparison between the Tempfli's method and presented method

In order to comparise the result of Tempfli's method and the method we have just presented,an experiment has been carried out. The result is showed in table 3 and table 4.In table 3 and table 4,NM stands for the new method,and TM stands for F.Tempfli's method.

It is obvious that the new method this article has presented can achieve much better estimation than F.Tempfli's.

Besides,our method costs less computation time than Tempfli's. Because we only need the computation of one dimension interpolation in order to calculate the accuracy.But for Tempfli's method, two-dimension interpolation and Fourier transformation are needed. According to our experiments,Tempfli's method costs 10 times as much as our method.

$e(\%)$		D1	D2	D3	D4
$\Delta x=0.2$	NM	1.2	0.9	1.1	1.0
	TM	9.4	10.8	9.2	11.1
$\Delta x=0.4$	NM	1.2	5.5	1.0	5.7
	TM	27.1	12.1	27.2	11.9

Table 3 q=10

$e(\%)$	I1		I2	
	T1	T2	T1	T2
NM	3.3	1.1	1.7	3.4
TM	29.1	41.3	36.3	46.8

Table 4 x=10m,q=10

I1 stands for interpolation method 1.

I2 stands for interpolation method 2.

In above discussion,We haven't considered the effect of measuring error on accuracy.We'll discuss the problem in following.

4.Interpolation accuracy estimation considering measuring error

Let $h_{i,j}$ be a height in a grid DTM, $f_{i,j}$ is the corresponding "true" height.

Let

$$\varepsilon_{i,j}=h_{i,j}-f_{i,j}$$
$$(i=0,1,\cdots M-1,j=0,1,\cdots N-1)$$

$\varepsilon_{i,j}$ is the measuring error caused by many factors.For simplicity, in our research, ε is surposed to be white noise.
The effect of measuring error on the accuracy has been investigated by many experts.In B.Makrovic's opinion,the total variance σ_T^2 can be expressed as

$$\sigma_T^2=\sigma_0^2+u_2^2 \qquad (10)$$

where

σ_0^2 is the variance of measuring error.
u_2^2 is the variance of sampling error.

The similar formula was developed by F.Ackermann.But in formula (10),the interpolation method hasn't been taken into consideration.In following,We are going to present several formulae to estimate the accuracy considering the measuring error.
u_2,the interpolation accuracy with no influence of measuring error,is defined as

$$u_2^2=\frac{1}{M_x N_y}\sum_{i=0}^{M_x-1}\sum_{j=0}^{N_y-1}(f_{i,j}-f'_{i,j})^2 \qquad (11)$$

σ_T,the interpolation accuracy considreing the measuring error, is defined as

$$\sigma_T^2=\frac{1}{M_x N_y}\sum_{i=0}^{M_x-1}\sum_{j=0}^{N_y-1}(f_{i,j}-z_{i,j})^2 \qquad (12)$$

where

$f'_{i,j}$ is the interpolated height with no influence of measuring error, $z_{i,j}$ is the interpolated height with the effect of measuring error.

Let $h_i(i=0,1,\cdots,N_l-1)$ be the sampling height with measuring error in high sampling density profiles, f_i is the corresponding "true" height. Also let z_i be the interpolated height of profiles considering measuring error, f_i' be the interpolated height with no influence of measuring error.

Let

$$\sigma_{1,l}^2 = \frac{1}{N_l}\sum_{i=0}^{N_l-1}(h_i - z_i)^2 \tag{13a}$$

$$\sigma_1^2 = \frac{1}{q}\sum_{l=1}^{q}\sigma_{1,l}^2 \tag{13b}$$

$$\sigma_{f,l}^2 = \frac{1}{N_l}\sum_{i=0}^{N_l-1}(f_i - z_i)^2 \tag{14a}$$

$$\sigma_f^2 = \frac{1}{q}\sum_{l=1}^{q}\sigma_{f,l}^2 \tag{14b}$$

$$u_{1,l}^2 = \frac{1}{N_l}\sum_{i=0}^{N_l-1}(f_i - f_i')^2 \tag{15a}$$

$$u_1^2 = \frac{1}{q}\sum_{l=1}^{q}u_{1,l}^2 \tag{15b}$$

where σ_f: profile interpolation accuracy with the consideration of ε.

u_1: profile interpolation accuracy with no consideration of ε.

Since h_i is the known value, z_i can be produced from the interpotion computation, σ_1 can be got according to formula (13). $f_{i,j}$, f_i is unknown. σ_T is the finally required result. How can we obtain σ_T from σ_1?

We solve the question in following procedure:

a. find the relation between σ_1^2 and σ_f^2.

b. find the relation between σ_f^2 and u_1^2.

c. find the relation between u_2^2 and σ_T^2.

Considering

$$u_2^2 = \lambda u_1^2$$

if above three relations can be maken out, the relation between σ_1^2 and σ_T^2 can be developed. Above procedure can be showed as follows

$$\sigma_1^2 \dashrightarrow \sigma_f^2 \dashrightarrow u_1^2 \dashrightarrow u_2^2 \dashrightarrow \sigma_T^2$$

The detail in the development of formulae is ignored, only the main formulae are presented here.

$$\sigma_f^2 = \sigma_1^2 + (a(0)-1)\sigma_0^2 \tag{16}$$

$$a(0) = \int_{-\infty}^{\infty} H(f)\,df \tag{17}$$

$$u_1^2 = \sigma_f^2 - R_1\sigma_0^2 \tag{18}$$

$$R_1 = 2\int_0^1 |H(f)+H(2-f)|^2\,df \tag{19}$$

$$\sigma_T^2 = u_2^2 + R_2\sigma_0^2 \tag{20}$$

$$R_2 = 4\int_0^1\int_0^1 |H(u,v)+H(2-u,v)+H(u,2-v)+H(2-u,2-v)|^2\,du\,dv \tag{21}$$

In formula (17) and (19), H(f) is the transfer function of one-dimension interpolation method, H(u,v) is the transfer function of two-dimension interpolation method.
From above formulae and formula (5), we can develop

$$\sigma_T^2 = \lambda\sigma_1^2 - R\sigma_0^2 \tag{22}$$

$$R = \lambda(R_1 - a(0) + 1) - R_2 \tag{23}$$

From formula (23), we know that R depends only on interpolation method.
We has arranged an experiment to judge the feasibility of formula (22) and (23). In the case of D1, σ_1^2 was calculated with the influence of measuring error. The $\hat{\sigma}_T^2$, the estimation of σ_T^2, is calculated according to formulae (22) and (23). We use formula (12) to produce "true" σ_T^2. As to the given interpolation method, we have

$R_2=0.62$
$R_1=0.8$
$a(0)=1.0$
$\lambda=1.5$

Then we get

$R=0.58$

The result is showed in table 5.

	x=0.4				x=0.8			
σ_0	σ_1	$\hat{\sigma}_T$	σ_T	e	σ_1	$\hat{\sigma}_T$	σ_T	e
0.422	0.3747	0.3276	0.3619	9.5%	0.4043	0.3767	0.4070	7.5%
0.211	0.2133	0.2060	0.2184	5.7%	0.2589	0.2733	0.2880	5.1%
0.105	0.1469	0.1612	0.1636	1.5%	0.2059	0.2392	0.2494	4.1%

Table 5 $\quad e = \left|\dfrac{\hat{\sigma}_T - \sigma_T}{\sigma_T}\right|$

From the table 5, we can make out that formula (22) and (23) are right.

5. Conclusion

This article presents a new method of estimating accuracy of DTM using high sampling density profiles. The method has been proved to be feasible for application. The effect of measuring error is also taken into consideration.
Finally, we must point out that in this new method, if only accuracy estimation is concerned, only several high sampling profiles are needed, i.e. it is not necessary to get the whole DTM. So it can be used as a method to determine the sampling interval before the acquisition of DTM. It is very useful for the establishment of LIS/GIS.

References

1, F. Ackermann, The accuracy of digital height models, Vorträge der 37 photogrammischen woche an der Universität Stuttgart, Sep. 1979.
2, B. Makarovič, Information transfer in reconstruction of data from sampled points, Phia, 1972, 4.
3, K. Tempfli, Spectral analysis of terrain relief for the accuracy estimation of digital models, ITC Journal, 1980, 3.
4, P. Frederiksen, Accuracy prediction for digital terrain models, XVth ISPRS.
5, Zhang Yaonan, The thesis of M.Sc, 1985