Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

STATISTICS OF SIMULATED IRREGULAR BLOCK SHAPES FROM REMOTE SENSED IMAGERY FOR DIGITAL GEOGRAPHIC INFORMATION SYSTEMS (GIS)

M.C. MUEKSCH

University of Bonn/Department of Geography
Bonn, Federal Republic of Germany

I. ABSTRACT

Building up Geographic Information Systems (GIS) on basis of remotely sensed images(air photos of larger scale up to high altitude scenes) for an urban block layout of irregular structure,involves statistics to quantify the data,i.e. coordinate storage in respect to machine plotted thematic maps with a cartographically generalised topography.

Different ways of block coordinate recording and generation are investigated regarding platform altitude,scale and generalisation using shape descriptors and indices.

A minimum amount of coordinate storage is produced by street polygons,surrounding urban blocks for high altitude scenes(scales 1:50 000 to 1:500 000). A medium amount will be got from computed polygons introducing linear distances (street widths) for medium altitude scenes(scales 1:5 000 to 1:50 000). The largest amount follows from the exact polygons,describing the real nature of the blocks and ploygonal sequences of building center points for low altitude scenes(scales 1:1 000 to 1:5 000).

A statistic analysis of different simulated block shapes with base geometries,like circular and elliptical,leads to the geometrical type of block,the distribution,the dependency from the number of coordinates,scale and generalisation.

The relationship between scales and storage amount is found and a function and computation proposal for quantifications are given.
Finally it is shown that such investigations are of an advantage for a better planning of GIS under financial restrictions,regarding the cost-benefit factor,especially for smaller urban areas under 100 000 inhabitants with an irregular block layout.

II. INTRODUCTION

Running GIS on computer basis demands knowledges about storage capacities,especially for coordinates of planimetric base topography.
In old middle-aged european towns the urban block layout is irregular,non squared and can only be described by larger amounts of block points without lost of base structure and sufficient cartographic accuracy.
The relationship between image or map scale and amount of points leads to the method of image sensing and observation from different platform altitudes.
In combination with feature detection,methods of remote sensing are applicable according to the detectivities and sensitivities of the sensor.
Another important criterium for the kind of computerized block layout is the way of recording and generating the coordinates from the images, their computational and graphical use in planning process and automatic thematic mapping.

III. POINT RECORDING AND GENERATION

A. RECORDING BLOCK TOPOGRAPHIES

Different ways of recording and generating block topographies were investigated (Table I):

1. Blockwise digitizing of closed coordinate sequences by exact polygons(EP)
2. Calculation of polygon points by distances from a street or topography modelling network by distances of corner point in direction of their centroids(CPI)
3. Calculation of polygon points by distances of blocksides(street widths) from the network polygon(CPII)
4. Digitizing a topography modelling network polygon(NP)
5. Digitizing adress point(building centers) sequences as block modelling points(AP)

B. POLYGON COMPUTATIONS

The second and third method make it neccessary to compute the block points by central network points,gravity centers and spatial distances. For method 2 a minor point calculation and for method 3 a section point calculation as they are applied in surveying were used.

IV. SHAPE MODELS FOR GENERALISED TOPOGRAPHIES

Shape descriptors in form of ratios,usually taken for pattern recognition and feature selection in image processing were tested for the different recorded and generated block layouts (Table I).

The mostly spread shape descriptors are the form ratio(FR),circularity ratio(CR),elongation ratio(ELO) and ellipticity ratio(ELI),defined by different authors and used in morphological analyses of quantitative geography,geomorphology and geology.

Areas(A),distances(L) and perimeter(P) are parameters of the ratios. The distances are defined as the largest diameter of the layouts which was introduced manually from a linear measurement of the images. All other parameters were computed from coordinates.

V. STATISTICAL ANALYSES

Results of block layout and shape descriptor simulations were subjected to a statistic analysis. The sample size was nearly 30 blocks,describing all sorts of layouts within an urban structure under influence of scale(platform altitude),sensing system and generalisation.

A. FREQUENCY DISTRIBUTION OF RATIOS

The differences between the ratios are seen from histograms,comprising values from 0 to 3 for the first 3 and 0 to 8 for the last descriptor, the ellipticity ratio(Table I).
The first two ratios present a relatively flat distribution over all block layouts,while the last two are concentrating their maxima around the indices 1 to 2 with frequencies up to 50%.

This can be attributed to the fact that elongation and ellipticity ratios are describing best all shapes from different stage of abstraction and image scale. Although a large part of the blocks are irregular squared,they can be detected as elongated and elliptical. Ambiguity is not excluded.

The second remarkable result is found in the differences of the ratios between the polygon methods(NP) and the others. While the 3 methods give a relatively equal distribution of amplitude, the last two have higher amplitude differences and more significant periods. From this follows that the higher and more complicated the edge locations are the more uniform is the class distribution of the ratios. Network polygon and adress point polygon are less accurate methods and show a more amplitude modulated distribution.

For circularity ratio a significant tendency to increasing amplitudes for higher classes of indices can be concluded. It supports the fact that the more abstract the polygon is the more it approximates to a rounded,circular to elliptical shape.

The elongation ratio has its highest amplitude between 1 to 2 for all methods which expresses the "dynamics of layout variations" from circular to elliptical shape passing the transitional stage of elongation.
This seems to be independent from the recording or generation method and is therefore valid for all.
The ellipticity ratio has its maximum between 1 and 2 for all methods and non-linear decreasing amplitudes with higher indices. It confirms the tendency to elliptical shapes against higher stages of abstraction.

The exact polygon presents the "true shape" and forms a layout as it is to be imaged in a large scale air photo.
A comparison of the four other more abstract shapes leads to a frequency distribution of standard errors from the true shape(Table II).
Circularity ratio shows the lowest differences in amplitudes against the original block shape, the highest contains the ellipticity ratio.
The periods of standard errors are higher in all first three ratios,but for the elliptical one they are low or have no second periodic element.
From the relatively equal distribution of errors over all stages of abstraction follows that the accuracies of methods are equal.

There is no need to accept the exact polygon method as the only way of describing a block layout. Therefore high altitude scenes are sufficient for block oriented census studies or thematic mapping.
By this fact an automatic reduction of point numbers will be originated and coordinate storage problems will decrease.

B. SPATIAL AUTOCORRELATION

In all the recording and generation methods lies a spatial sequence,concerning diameter and area. This is trivial and can easy be seen from the area amounts of different block layouts such as:

$$A_{AP} < A_{EP} < A_{CPI} \text{ or } A_{CPII} < A_{NP}$$

From this sequence also a spatial autocorrelation for the ratios can be concluded,as they are area and diameter dependent.
A simple measure for autocorrelation is the form

$$k = m^2 / s^2$$

with m^2 as mean squared successive differences and s^2,the variance.
All autokorrelation coefficients have been tested with a χ^2- test for their significance and distribution.

The more significant and positive the autocorrelation is between successive ratios,the smaller it will be. The spatial trend from accurate block polygon(EP) to network(NP) is significant, although there are partly regressive amounts of indices in between,caused by irregualities in the point selection during generalisation.
Therefore a frequency distribution was designed and the class under 1.0 was observed with higher priority.

With a 95% significance level the amplitudes of the frequency distribution for the classes between 1.0 and 2.0 for all ratios,included diameter and area is highest with nearly 40-50%.
All distributions lie very close under the χ^2-limit for the significance of 95% and a degree of freedom of 4.

These results show that area,diameter,form ratio and ellipticity ratio has the highest probability of autocorrelation.

To see,wether there is also a possible autocorrelation of the other ratios,alternative tests with higher exspected frequencies were made.
The results of the first were confirmed and showed that form ratio and ellipticity ratio under the upper limit of 1.0 can have a range of 20% between the amplitudes,if the exspected frequency is taken at a maximum value for the same degree of freedom as above.

If the ratios will be ordered by its amplitudes of distribution of autocorrelation for class 0 to 1,there is the sequence:

F or DIA(60%) $>$ FR or ELI(50%) $>$ ELO(40%) $>$ CR(30%)

The mentioned transition from circular over elongation to ellipticity can be concluded from this sequence as well.

C. EMPIRIC CORRELATION COEFFICIENTS AND FACTOR ANALYSIS

The autocorrelation test gave a certain sequence for the ratios.
The idea that between the ratios at all exists also a correlation,lies near.
From the formulae follows that the area and largest diameter are dependant variables and originally correlated.

For the analysis a correlation matrice was constructed and the empiric correlation of the ratios were calculated(Table III).
The data for this analysis were taken as mean ratios of all recording methods.
Before starting this calculation,the mean errors of the means were observed.
They showed a variation of 20 to 40% of the means for all ratios with a concentration around 20%.
The low mean errors made it possible to calculate those correlations as it was not to fear that they would influence the results.

The shaded squares present the originally correlated ratios.
Remarkable is the high correlation of the form ratio and circularity ratio and those of form and circularity together with elongation ratio.
The others were less correlated.

To confirm the correlations,the influences between them were investigated.
Setting the correlation coefficients as coefficients for unknowns in a set of normal equations, their eigenvalues were determined with Gauß elimination method. Multiplication of eigenvalues with the correlation coefficients gave the matrice of factor load(Table IV).

The correlation between form and circularity ratio is confirmed,while the others are less significant against those of the first tableau.
Regarding the correlation of diameter with ellipticity ratio,the relative high value in the first tableau has decreased to an irrelevant correlation of negative value now.
Further studies of these highly factor loaded coefficients in a partial correlation analysis show the same results.

From the original correlation matrice and the factor analysis there could be related some interdependencies for multiple regression analysis.
For example the above approximation of the form ratio can be got by this analysis with the same coefficients.
Other possible highly correlated ratios are ellipticity with form ratio and diameter and form ratio in dependence of diameter and area.

D. MULTIPLE REGRESSION AND RELATIONSHIPS BETWEEN THE RATIOS

Analysing the mean error of arithmetic mean for ratios of all methods,probable simple relationships are immediately seen.
The mean error for circulatity ratio is the smallest of all others.
A sequence of ratios and descriptors ordered after their increasing mean errors,expressed in percent is:

CR(11%) $<$ ELO(13%) $<$ F(17%) $<$ DIA(18%) $<$ FR(25%) $<$ ELI(47%)

If one observes the errors in per cent and the absolute means,one can deduce an empiric relationship such as:

$$FR \approx CR + ELO - 0.3$$

with 0.3 as biased error for all studied samples.

From further studies this relationship could be confirmed with nearly identic values for FR and variations of 14%.
A small variation lies in the value of 0.3.

For mixed structured block layouts other relationships,combining area,diameter and shape ratios were found:

1. FR = 1.086 + 0.000069 F + 0.253 DIA (10%)
2. FR = -1.365 - 0.202 CR + 2.498 ELO (5.6%)
3. ELI = -0.254 + 1.027 DIA + 0.771 FR (10%)

The percentage values are standard errors.

VI. GENERATION OF SYNTHETIC BLOCK LAYOUTS FOR POINT STORAGE QUANTIFICATIONS

From the analysis foll owing conclusion can be drawn:

1. The largest amount of block shapes are elongated and elliptical
2. A medium amount is circular
3. A relatively small amount is totally irregular and not to subsummize under a regular shape
4. Squared block layouts are negligible in the type of irregular urban structure

Several experiments for elongated and elliptical shapes give point numbers lying between 6 and 10,around 6 for elongated shapes and 8 for elliptical shapes. In case of irregular,elongation or ellipticity generalised shapes the number jumps less over 10.
Very long curved block shapes have to be treated separately.

Circularity gives lesser difficulties:
A minimum of 6 points ordered in a hexagon shape is necessary to regard a shape as circular or approximately circular.
Totally irregular block shapes are handled separately or can be divided in sub-shapes of the shapes above,substracting all surplus points.

Squared layouts give no problems in point quantifications. Form ratio give a general description and says nothing detailed about the approximated shape geometry.

To compute points for storage quantifications a random generation is best suited. The random numbers are multiplied with the base number for the respective shape structure.
Following steps for a computation can be assumed:

1. Generalize the block layout from the scene
2. Measure the urban area of total block layouts
3. Estimate the number of blocks within this area
4. Generate point numbers as many times as numbers of blocks have been estimated
5. Take the arithmetic mean
6. Multiply it with the number of blocks and compute the storage amount,regarding the digital places of coordinates to be taken manually or with digitizer

The following program,runned on an APPLE II (48k) computer performs the computational steps:

```
1  REM  SIMULATION DES SPEICHERPL
     ATZBEDARFS FUER BLOCKBEGRENZ
     UNGSPUNKTE EINER IRREGULAERE
     N STADTSTRUKTUR
2  REM  ECKP=GRUNDANZAHL DER ECKP
     UNKTE
3  REM  STW=SIMULATIONSSTARTWERT
     0 STW 1
4  REM  KODIG=ANZAHL DER STELLEN
     DES KOORDINATENPAARES
5  REM  FLAECHE IN QUADRATKILOMET
     ER
6  DIM Z(1000)
7  INPUT ECKP,STW,KODIG,AREA
8 N = AREA / 0.0025
9 N =  INT (N)
10 SUMN = 0.
11  FOR I = 1 TO N
12 Z(I) =  RND (STW)
13 SUMN = SUMN + Z(I) * ECKP
14  NEXT I
15 NP = SUMN / N
16 NP =  INT (NP)
17 BYTES = (KODIG * 16 * NP * N) /
     1000
18 BYTES =  INT (BYTES)
19  PRINT "MITTLERE ANZAHL DER BL
     OCKPUNKTE",NP
20  PRINT "SPEICHERPLATZBEDARF",B
     YTES,"KB"
21  END
```

VII. CONCLUSION

The largest amount of block shapes in an urban structure of old kind (Fig.1,2) are the elongated and elliptical ones(with the definition that the largest part of the block area is falling into the shape boundaries),dependant from stage of abstraction,generalisation and scale.

It is up to the interpreter to recognize an elongation or ellipticity structure. If so then a fixed point number can be used to describe a general block layout of this kind.

Circular or near circular shapes are highly dependent from scale and density of points,less from generalisation. There is a tendency to count more points in larger scales. From a certain point on there is no higher gain of roundness and accuracy,neither in shape,nor in the visual image by adapting a higher number of points.
For example at ascale of 1:5 000 a circular shape of 5 m diameter,generalized with a pen of 0.5 mm is vanishing to one point(Fig.3).

The statistics of simulated block shapes leads to base numbers of block points which are in case of ellipticity relatively independent from scale,sensing system and platform altitude. Using these numbers,new numbers can be generated within a random process which simulates an approximate number,regarding the connectivity in case of a closed network and the stage of generalisation in case of isolated polygons.
With the number of base points,starting simulation value,digits of coordinates,mean block area and the investigated urban area,the capacity for coordinate storage can be simulated(Fig.4).
The reliability of such a generation lies under 30%.

Storage problems seems today less important, because of high available capacities. Thinking that even a medium populated town under 100 000 inhabitants comprises an average area of several squarekilometers with adense very irregular block structure,the amounts of coordinates are increasing rapidly over the capacity of desk computers.

But towns of mentioned kind cannot install large storage computers for their coordinate registers and centralisation of such registers are becoming more expensive and difficult in access, besides the fact,that there is no balance in cost-benefit situation in using them.
Only desk computers are finally possible to finance and run,therefore storage problems are still remaining,dependent of origin of datas,sensing systems and methods,platform altitudes,scale and generalisation.

VIII. REFERENCES

1. Hagget,P.,R.J.Chorley,Network Analysis in Geography,Edward Arnold,London 1972

2. Castleman,K.R.,Digital Image Processing, Prentice-Hall,Inc.,Englewood Cliffs,New Jersey 07632

3. ASP,Manual of Remote Sensing,Falls Church, Virginia,Chapter 12,p.620,Geocoding,First edition 1975

4. Blair,D.J. and T.H. Biss,The measurement of shape in Geography,Nottingham University, Department of Geography,Qantitative Bulletin 11,1967

5. Bunge ,W.,Theoretical geography,Lund studies in geography,Series C,General and Mathematical Geography,1.,1962

6. Stoddart,D.R.,The shape of atolls,Marine Geology,3,369-83,1965

7. Pedersen,P.P.,On the geometry of administrative areas,Copenhagen,MS.Report,1967

8. Pavlidis,T.,A review of algorithms for shape analysis,Computer graphics and image processing,7,243-258,1978

AUTHOR BIOGRAPHICAL DATA

Michaela C. Mueksch studied geosciences and astronomy at University of Bonn,F.R.G. from 1964-69, MSc. in geodesy and Phd. in geosciences, several years working in surveying and photogrammetry of civil service and industry,research of application of GIS in urban and regional planning and thematic mapping on computer basis from 1973-75,lectures and researches in photogrammetry,cart ography and remote sensing in West Africa(Cameroon) and East Africa(Tanzania) from 1976-82,presently working on remote sensing applications in ecological monitoring of tropical zones,
member of American Society of Photogrammetry(ASP) and German Society of Photogrammetry and Remote Sensing(DGPF).

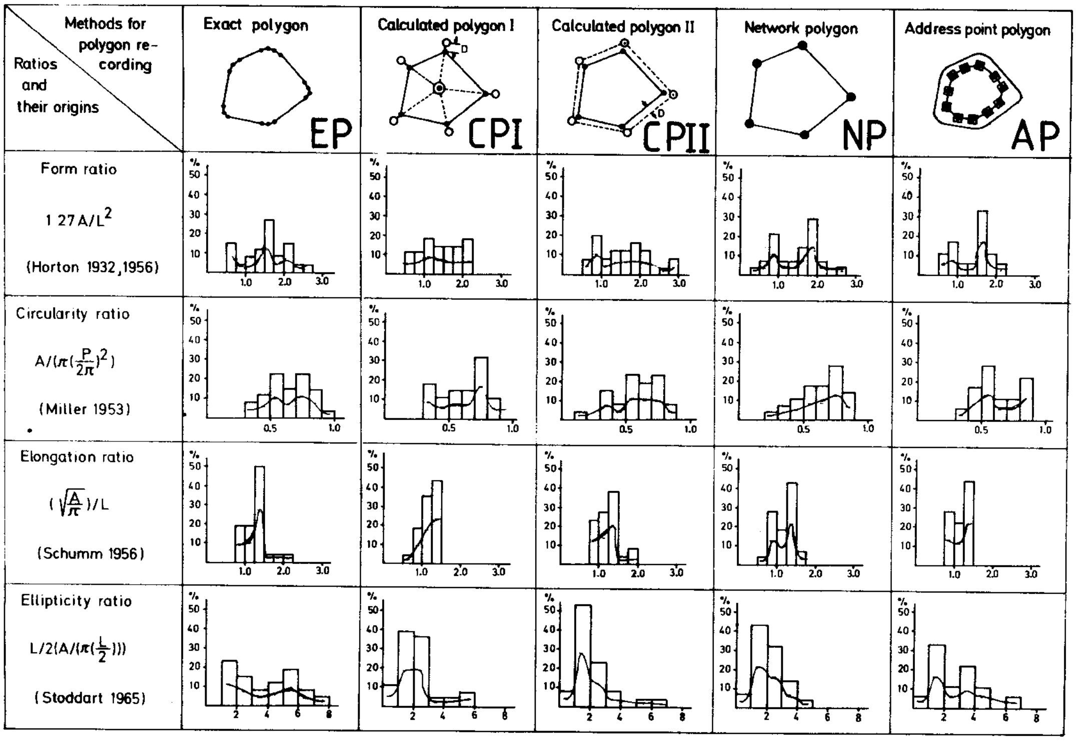

Table I : Methods of polygon recording, shape ratios and their distributions

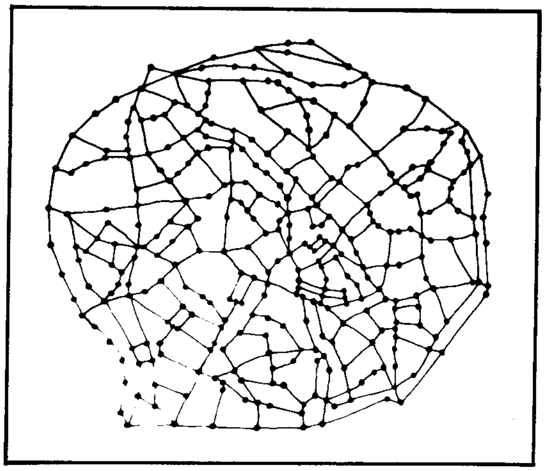

Fig. 1 : Network structure of a middle-aged german town(Aachen,the residence of Charles the Great)

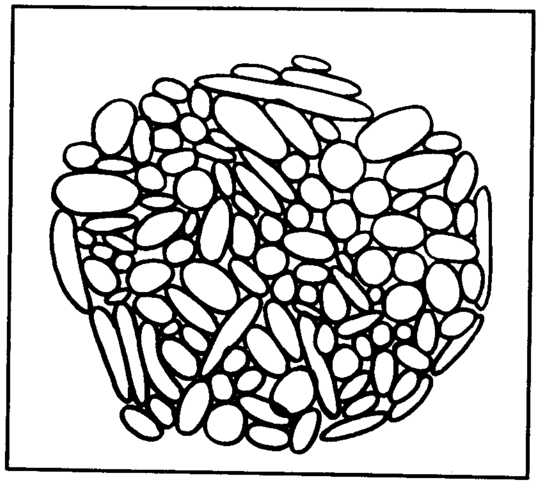

Fig. 2 : Ellipticity and circularity simulated blocks of the network of Fig. 1

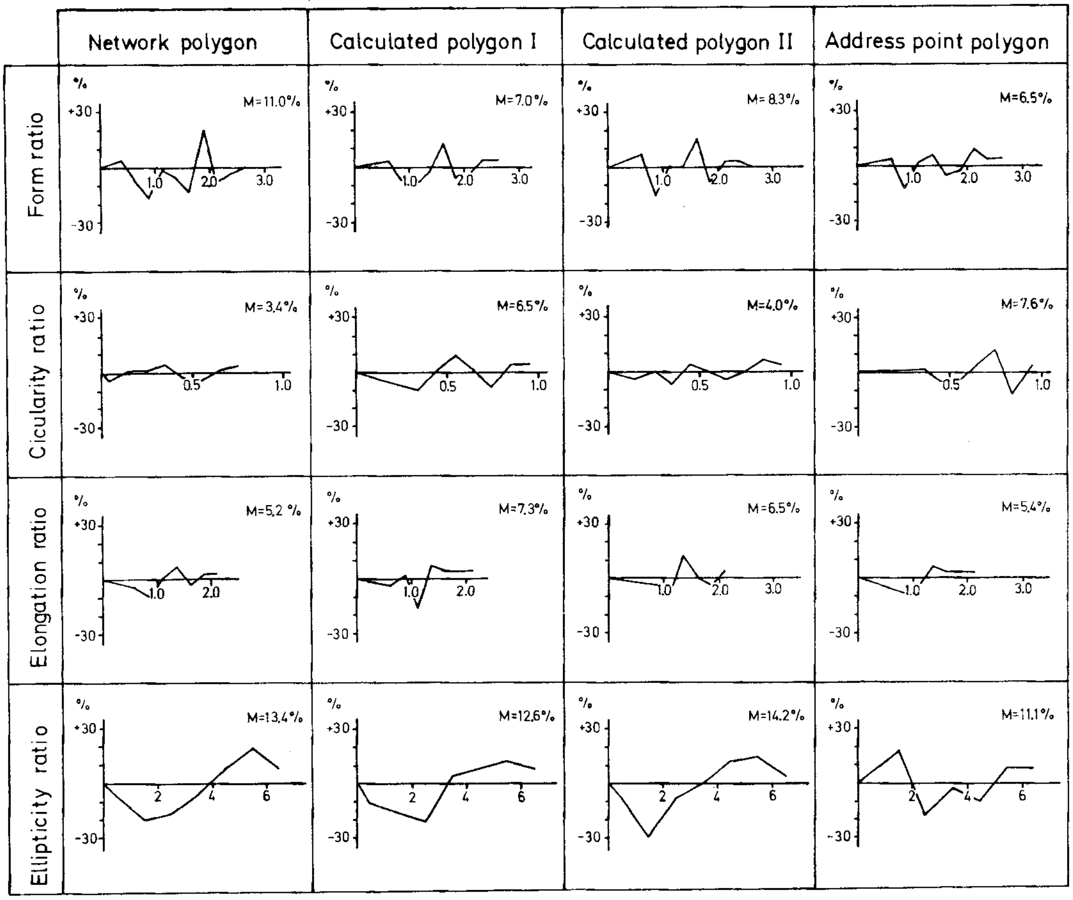

Table II : Frequency distribution of standard errors from the true shape

F	DIA	FR	CR	ELO	ELI
1	0.917	0.437	0.068	0.070	0.038
	1	0.319	0.026	0.197	0.826
		1	0.971	0.635	0.406
			1	0.646	0.125
				1	0.337
					1

originally correlated (0.917)

assumed correlation (0.826, 0.971, 0.635, 0.646)

Table III : Correlation matrice

F	DIA	FR	CR	ELO	ELI
-22.9	142.3	-127.8	17.1	4.6	-2.3
	155.2	-93.3	6.5	13.2	-51.7
		-292.5	243.4	42.6	-25.4
			250.7	43.3	7.8
				67.0	-21.1
					-62.6

Table IV : Matrice of factor load

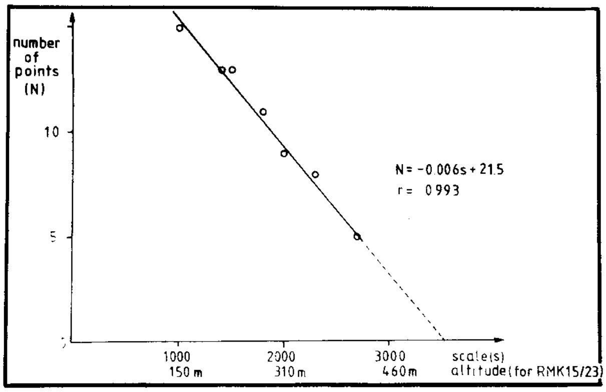

Fig. 3 : Relationship between number of points and scale(altitude) for a circular shape

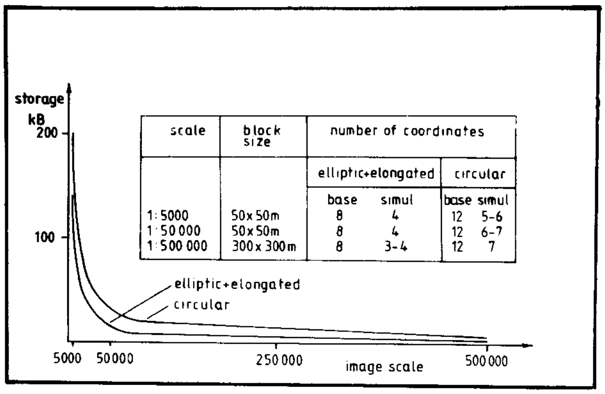

Fig. 4 : Relationship between necessary storage capacity for coordinates and scales for elliptical+elongated and circular shapes