An Elementary Program on Urban Description
The Cell-Counting Method

By Ole Hebin

Abstract

The program described in this paper makes simple, but very time-consuming sorting and summation operations on land-use material assigned to coordinates. The present version has been applied on urban material, but with only slight changes of the program many other fields may be treated.

The work has been completed as a smaller dissertation in connection with the author's M.Sc. degree.

Introduction and General Description

For the purpose of the 1967 course for students majoring in geography (HFK67) a computer-program was developed in the course of autumn 1967 and spring 1968. It makes a wide range of elementary, but very time-consuming additions and calculations on the basis of land-use information assigned to coordinates from town reconnaissance. These additions and calculations are made in a usual right-angled quadratic grid net with variable cell and net side.

The results appear as ordinary area data in the form of areas depicting characteristics, ratios of areas, and observation frequencies.

Input

The field survey should be made with all possible accuracy, and it was found to facilitate work if initially, in the field, the collected data were subdivided into 3 groups:

I. Data on the site as:

1. Year of registration.
2. Centre of site in the UTM-net.
3. Name of town.
4. Name of street.

5. House number and letter.
6. Title No. and litra in Land Register.
7. Site area.
8. User of site.
9. User's share of site.
10. Application.

II. Data on the building as:

1. Year of registration.
2. Centre coordinate of the building in the UTM-net.
3. Year of construction.
4. Built-up area.
5. Number of floors.
6. Type of house.
7. Condition of house.

III. Data on floors as:

1. Year of registration.
2. Centre coordinate of the floor in the UTM-net.
3. Number of the floor.
4. Number of floors under the ground floor.
5. Area of the floor.
6. Room type.
7. User.
8. Activity.
9. Share of activity.
10. ISIC (International Standard Industrial Classification).
11. Size classification of activity.
12. Assortment and range of activity.
13. Technical equipment of activity.
14. Number of households.
15. Number of persons.

In its present version the program only comprises allocation and size of site, building and floor areas and does not include the application of area within the three main groups. In a forthcoming version, however, the program will also include this subject.

All area data are stated in square metres and the coordinates in 1 (one) metre units. The registration in the field is entered on optical sheets, which are later converted electronically into punched cards to be run on the computer.

The Grid

The grid has the following characteristics:

1. Minimum and maximum coordinates must be specified (maximum stated as maximum + 1).

2. The grid must be specified quadratically (number of cells in each row and column).

3. Number of cells (in a row) should be indicated as an integer between 2 and 30 (incl.); any integer, however, between 2 and 30 may be defined as the starting number on the basis of which all calculations are made, and a third integer stipulated defining the number of cells to be added to the first number, before the whole calculation is repeated with the new number of cells, etc.
   Example 1.
   With a single integer as: 21, all calculations will be made in a grid with 21×21 cells.
   Example 2.
   With two integers as: 5 and 7, all calculations are made first in a grid of 5×5 cells, then all calculations are repeated in a grid of 12×12 cells, next time in a grid of 19×19 cells and finally in a grid of 26×26 cells.

4. The areal covering of the grid is without maximum or minimum limits, apart from the restriction that the maximum coordinates mentioned under item 1 should be given by a maximum of 8 digits (e.g. 99999999 m) and the minimum coordinates by at least 1 digit (e.g. 1 m). If necessary, all length and area records can rather easily be changed into more convenient units (μ or km).

5. When read in, all coordinates are transformed into a coordinate system with (0,0) in the minimum coordinates of the original UTM-grid, and the unit measure is transformed into K, where K = (MAXX-MINX)/MAXN.

   K – the unit measure in the new system (= a cell side length in the old system).
   MAXN – number of cells in a row.
   MINX – minimum for x-coordinate as specified in item 1.
   MINY – minimum for y-coordinate as specified in item 1.
   MAXX – analogous to MINX.
   MAXY – analogous to MINY.

(x', y') = ((x÷MINX)/K+1, (y÷MINY)/K+1). By this calculation, the coordinate system is transformed to (0,0) irrespective of the size of MINX and MINY, thereby simplifying the calculations. Specifically, the length of all cell sides will be equal to 1 after this transformation.

6. Values falling on coordinate lines are entered into the cell in the direction of the axis (i.e. towards east and/or towards north). If for example the coordinates are indicated by (s, t) we obtain $x_i \leq s < x_{i+1}$ and $y_j \leq t < y_{j+1}$ for i = 1, 2, 3, .. and j = 1, 2, 3, .. This method of calculating the coordinates requires that MAXX and MAXY be indicated one unit greater than the actual maximums.

All the specifications of the grid are then printed out by the program for control, and we obtain the following print outs: MINX, MAXX, MINY, MAXY, MAXN and K. Finally, number and area of cells as well as area of grid are calculated and printed out.

Sequence of Operations

The data are read in on punched cards and computed in the following order: data on sites, on buildings, and on floors.

The reading in of site data first results in a summation of all read in positions within the stated minimum and maximum limits, whereas those falling outside the limits are indicited as invalid positions. Invalid positions need not be misindicated coordinates, but may be due to the fact that more data than actually needed are available.

During this input phase the positions with their areas are distributed into the correct cells as described above, and the sum of areas and the number of positions (sites) of each cell as well as of the whole grid are computed. Next, the total site area of the grid, the average site size per cell, the standard deviation around the average, minimum and maximum for the cells are computed and printed out. Further, the average site size as well as average site area for those cells containing sites are computed. The values of the site areas found in the cells are then computed into totals, averages, standard deviation, maximum and minimum for all rows and columns. Finally, the contents of the grid are printed out as a squared matrix.

To facilitate interpretation, the distribution matrix is printed out as a corrogram showing the total site area of each cell. As seen from

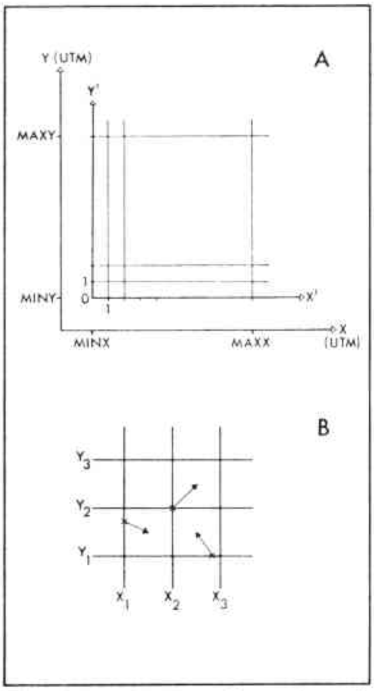

Fig. 1 A og B. Transformation of the coordinate system and assignment of border locations.

Fig. 1 A og B. Transformering af koordinatsystem og henføring af randpunkter.

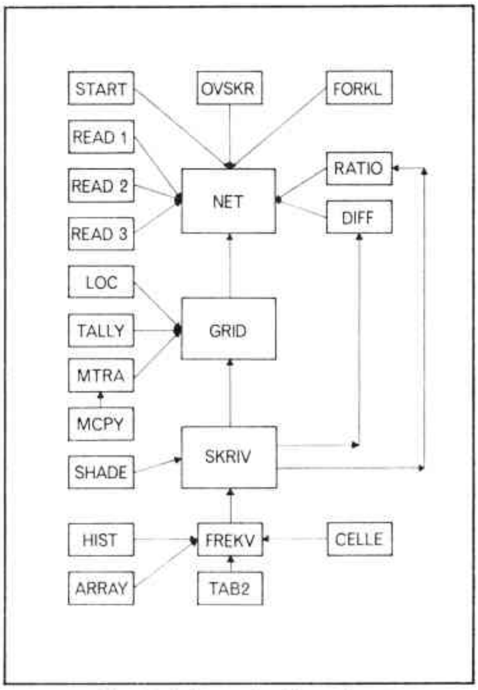

Fig. 2. Schematic flow-chart.

Fig. 2. Skematisk programoversigt.

the legend the hatching divides the area of cells into six classes of equal size.

Simultaneously with the computing of areas of the single cells the frequency of sites is counted and a frequency matrix printed out on the basis of the observations in the grid. As was the case for the areas, the frequencies are then calculated for totals, averages, and standard deviations for both rows and columns.

The freqency matrix is reproduced graphically as two histograms of row (horiz.) and column (vert.) frequencies. This matrix is then used to compute a matrix containing the relative frequencies and further, together with the area distribution matrix, to compute the average site size in each cell.

Next step in the process is the computation of the data on buildings, effected in exactly the same way as for the sites. This done, the relative built-up area is computed by means of the distribution matrices for the site and the building areas, respectively, by dividing the building areas in each cell by the site areas of same cell. Finally, the remaining area of each cell is computed by subtracting the built-up area from the site area.

Correspondingly, computations are made on the floor data. From the preceeding result the floor space ratio in each cell can be obtained by dividing the floor area by the site area. Further, the height index can be found by dividing the floor area by the built-up area of each cell.

This completes the computations, although they may be repeated on the basis of other cell specifications.

It should be added that all lengths are measured in metres, all areas in square metres and all ratios are expressed in thousandths.

The standard deviations are calculated by the following formula:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

The Composition of the Program

The program is constructed around a simple main program, which controls the computation process by calling subroutines in correct sequence. A detailed flow-chart will be omitted here, but in fig. 2 a diagram with the subroutines as elements shows the whole process.

The function of the individual routines briefly is as follows:

NET – main program, reads control parameters and calls subroutines.

START – reads all data cards and writes the information on binary tape.

OVSKR – administrates all headings.

FORKL – prints out a description of the program (a summary of the present paper).

READ1 – reads in site data, transforms the coordinates to a new system and distributes the data in the correct cells.

READ2 – same as READ1, but on building data.

READ3 – same as READ1, but on floor data.

GRID – controls the computation of totals, averages, standard deviations, minimum and maximum values. Computes elementary data on the whole grid.

LOC – changes matrices from 2 to 1 dimension prior to the call of TALLY.

TALLY – computes totals, averages, standard deviations, minimum and maximum values.
MTRA – transposes a two-dimensional matrix by means of MCPY on basis of which column and row computations can be made analogously.
MCPY – changes the name of the transposed matrix.
SKRIV – makes all larger print outs.
SHADE – computes and "draws" a corrogram.
RATIO – computes built-up ratio, floor space ratio and height index (matrix divisions).
DIFF – computes unbuilt areas (matrix subtractions).
FREKV – controls the computations of frequencies by means of ARRAY and TAB2 and computes column and row data of frequencies.
ARRAY – changes matrices from 1 to 2 dimensions.
TAB2 – computes frequencies and relative frequencies which are printed out in FREKV.
CELLE – computes and prints out the values per position per cell.

The whole program is controlled by means af a single parameter card containing the following information:

1. MAXX
2. MINX
3. MAXY
4. MINY
5. MAXA – number of cells per row at first computation (at least 2).
6. MAXB – number of cells per row to be added to MAXA in sequence of first computation.
7. MAXC – maximum number of cells per row (max. 30).
8. KOANT1 – number of punched cards with site data.
9. KOANT2 – number of punched cards with building data.
10. KOANT3 – number of punched cards with floor data.

The parameter card is punched according to the format (4I8, 3I2, 3I8).

The program is constructed in such a way that it lies permanently in core-storage, which is nearly filled up (I/O Buffers comprise only 1435 octal places). This heavy use of core-storage involves one single limitation in respect of the amount of data, namely that the

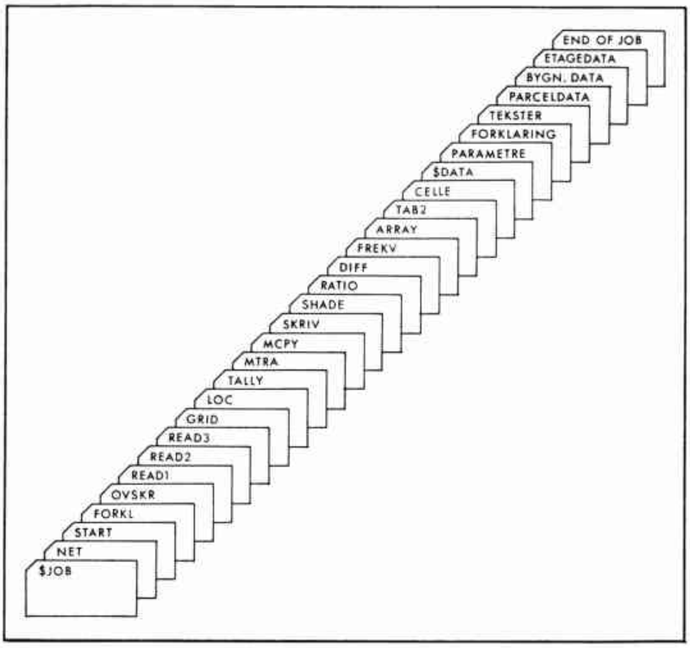

Fig. 3. Deck-set-up.

number of cards for each category must not exceed 1000. If the necessity should arise, however, the program can be split up and put on tape (overlay), allowing for larger buffers, and the number of cards increased accordingly. Finally, additional room may be provided by running the program under ALTIO (alternative input- output system) though with a prolongation of the running time. A program up to about 5000 cards per category is being tested.

Scarcity of storage has made it impossible to include a subroutine for the drawing of isoline maps of the contents of the different matrices. Consequently this drawing must be run separately by the routine PLTTRI, an operation which has also been written and tested (unpublished material).

The program is punched on approximately 4000 cards, cf. fig. 3.

Uncertainties

The *numerical* uncertainties which may appear are identical with those known from manual calculations; i.e. due to rounding or truncation. All errors of this type are not greater than 1 (one) unit, as all computations are carried out with approximately 8 significant

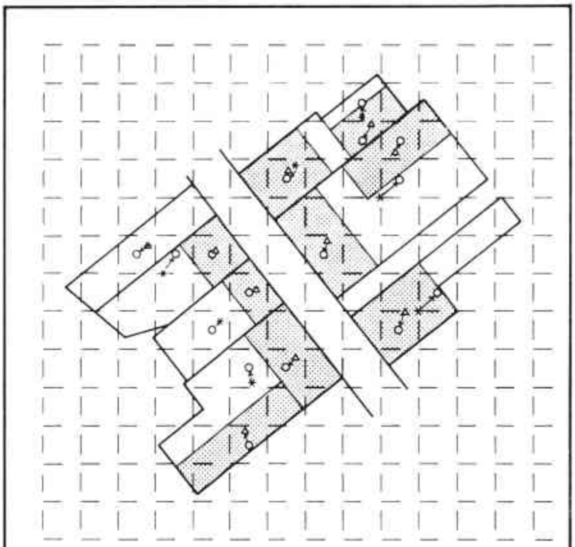

Fig. 4. Detail of map and grid with central points.
Fig. 4. Detalje af kort og rudenet med centralpunkter.

digits provided that the values remain within the range of 10^{38} to 10^{-38} for IBM7094; and rounding and truncation of decimals is the last operation prior to printing out.

Fundamental errors do occur but are only of a type due to the grid method. This type of error is illustrated in the simplest way in fig. 4.

The broken lines belong to the grid, whereas the unbroken ones represent site and building limits. The asterisks indicate site centre, the symbol Δ, building and floor centres, and the symbol O, the cell centres to which the areas in question belong. The site areas are nearly always larger than their respective cells, which fundamentally cannot be considered an error, but may occasion the question whether the chosen cell size gives a reasonable representation of these areas, or if the problem actually is whether grids can represent areas at all. Presumably, an answer of universal validity cannot be given, the following should only hint at some general lines.

If the cell is very big in relation to the areas referred to it the situation may approximate to an analysis of distribution of points. Consequently, cells of relative large size will prove worthless in investigations of small towns and agricultural areas, as a grid with cell size 1 km² or more would be of no value. *Gunnar Olsson* (1967) states that the "ideal" cell size is one for which the average occurrence of points per cell is 1 (one), a rule-of-thumb advanced by Gunnar Olssons as useful in many cases. A similar rule concerning cell size in analyses is that length of cell side should be twice the average area of points in the grid (*Curtis,* 1950). If the cell is not big in relation to the areas to be investigated it is often found convenient to apply a cell size representing an average of the investigated areas. No reasons are given in support of this rule, but it deserves notice that it remains uncontradicted. It should also be taken into account when estimating the value of these rules that the cell size decides the degree of generalization obtained. That the rules tend to decide the cell size which will give the lowest degree of generalization is evident as far as Gunnar Olsson is concerned, uncertain with Curtis and McIntosh (*Miller and Kahn*), and in case of the last rule apparently a "suitable" level of generalization is aimed at. The resemblance between rule 1 and 3 should not be ignored, however. In both, a unit (point and area respectively) is referred to each cell.

These fundamental errors have more consequences than those already mentioned for some of the results. If the hatched areas represent the actual buildings it appears that their centre sometimes falls into the same cell as the site centre does, whereas in other cases it falls into a different cell. This means that expressions such as unbuilt area, built-up and floor space ratios (coincidence of floor and building coordinates) may assume values which are not meaningful. As examples may be mentioned the occurrence of negative unbuilt areas, built-up ratios of more than 1000 thousandths, or cells for which the ratio cannot be computed because divisor (the site area) is zero and similar absurdities as to floor space ratio. When divisor is zero in a computation the result printed out will appear as –1.0. If no areas are referred to a cell –0. is printed out. This rule has been adopted to distinguish e.g. an unbuilt area of 0.0 m² from a cell without site or building data (–0.).

In the present material it has not been possible to remove the disadvantage that buildings and sites are not cell-registered coherently. In a forthcoming version this drawback will be eliminated by assigning a site identification to buildings and floors, which, however, necessitates a minor change of the program.

Supplementary Comments

The basic construction is very broad and the only justification of the title: "An elementary program on urban description" are the headings.

The heading cards, which have to be changed when revising the program, are grouped in a separate data deck and therefore within easy reach for user changes. A change of the program is being developed to make it applicable to land-use data of agricultural areas and it will be run on the data collected on the lower part of the Varde Å Valley (south-western Denmark) during the HFK68 students' course.

An extended version with a maximum number of 5000 punched cards is being tested for use in an investigation on Ras Beirut conducted by *Per Kongstad,* assistant professor at the University of Copenhagen.

The urban aspect of the present program has been developed in cooperation with Per Kongstad and participants in various courses for students majoring in geography.

The time of programming – including testing – amounted to approximately 250-300 hours.

SUMMARY

For the purpose of obtaining practice in field work for students majoring in geography the courses arranged also comprise urban surveys. Assigned to coordinates the results appear as area data at different levels.

Preliminary, a material of this kind is worked up in elementary totals, averages, standard deviations, frequencies, simple residuals, and ratios. These dull, but very time-consuming processes are performed by the program by assigning the whole amount of data to a grid with variable specifications. Simple budgets on the respective areas are tabulated for each single cell as well as for the total grid, and the interrelations between the different area categories are established, in the present material expressed by way of unbuilt areas, floor space and built-up ratios as well as height index.

Only the separately stored heading cards refer the program described in this paper to the concept of town. Fed with other cards the program is very easily changed for application on any other material.

In innumerable discussions *Niels Kingo Jacobsen* and *Per Kongstad,* professor and assistant professor respectively at the University of Copenhagen, as well as students joining the HFK66 and HFK67 courses, have all contributed to the formulation of problems.

REFERENCES:

Bengtson, B.-E. and Nordbeck, S. (1964): Construction of Isarithms and Isaritmic Maps by Computers, BIT, Bind 4, Hefte Nr. 2.

Curtis, J. T. and McIntosh (1950): The interrelation of certain analytic and synthetic phytosociological characters. Ecology 31, pag. 434-455.

Evans, F. C. (1952): The influence of size of quadrat on the distributional patterns of plant populations. Contr. Lab. Vert. Biol., No. 54, Ann Arbor.

Greig-Smith, P. (1964): Quantitative Plant Ecology, London 1964, pag. 1-19, 131-157.

Miller, R. L. and Kahn, S. L. (1962): Statistical Analysis in the Geological Science, New York 1962.

Nordbeck, Stig (1962): Location of Areal Data for Computer Processing. Lund Studies in Geography, Ser. c., No. 2.

Nordbeck, Stig (1964): Framställning af kartor med hjälp av siffermaskiner, Geografiska Institutionen, Lund. Avhandlingar nr. 40.

Olsson, Gunnar (1967): Lokaliseringsteori och stokastiska processer. Regionale Analysemetoder, Oslo 1967. RSA-NS, Rapport 7., kap. II.

Pelto, C. R. (1954): Mapping of multicomponent systems, Journ. Geol., 62: 5, pag. 501-511.

Skellam, J. G. (1952): Studies in Statistical Egology, (I. Spatial patterns), Biometrika Vol. 39. Pts. 3, 4.

Et elementært bybeskrivende program

Af Ole Hebin

Indledning og general beskrivelse

Til brug for hovedfagskursus 1967 (HFK67) udvikledes i løbet af efteråret 1967 og foråret 1968 et EDB-program, der foretager en mængde elementære, men meget tidskrævende sammentællinger og beregninger på koordinatsat karteringsmateriale fra byrekogniseceringen.

Alle beregninger og sammentællinger foretages i et sædvanligt, retvinklet, kvadratisk rudenet med variabel rude- og netside.

Resultaterne fremkommer som sædvanlige arealregnskabsoplysninger i form af arealer for bestemte karakteristika, forskellige kvoter samt observationsfrekvenser.

Input

Karteringen i marken søges gjort så grundigt som mulig, hvorfor det findes arbejdsmæssigt lettest at inddele det indsamlede materiale i tre kategorier allerede under feltarbejdet: