

CONFERENCE OF EUROPEAN STATISTICIANS

**Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of
Statistics and Geography**
(Tallinn, Estonia, 25-28 September 2001)

Topic (iii): Spatial analysis

ESTABLISHING A GRID NET FOR THE WHOLE OF FINLAND

Submitted by Statistics Finland¹

Contributed paper

I. GOAL FOR USING A GRID NET

1. Since the 1970 population census, Statistics Finland has been aggregating statistical data to grids with the cell size of 1:1 km or smaller, using the coordinates of buildings. The grid data are mostly based on population, labour force or employment, in which a regular grid square constitutes the statistical area. Using GIS software, it has been possible to convert an aggregated grid data set to digital GIS format.

2. The grids have mostly been constructed using the coordinates contained in the data. The coordinates are the middle points of the buildings where people live, etc. Regional codes (municipality codes, locality codes) have also been taken from the data in this way. Only the cells containing actual data are included in the resulting data sets. The grids from different types of data (population, work places) or the grids from different time periods are therefore different and cannot be easily compared.

3. A grid net is a tool for analysing and comparing statistical grid data. It is a vector-based grid data set, where cells are regular square polygons. It covers the whole area of Finland, including the grid cells with no data. The regional codes in a grid net are defined spatially by using digital maps. Statistical data units can be aggregated using the coordinates of the statistical units and the aggregated data can then be joined to the grid net using an identification code based on e.g. the lower left-hand corner of a grid cell.

II. CONSTRUCTING A GRID NET

4. A grid net with regional codes can be established using different GIS software. We have used ArcInfo. Two different methods can be employed: in the first, the construction can be made using a specific point, e.g. the lower left-hand corner of a cell, and only the result is converted to polygon coverage. In the second method, cell polygons can be used from the beginning.

5. The first grid net with the size of 1:1 km was made for our TeemaCD product three years ago using points. The points were generated into point coverage using the lower left-hand corner of a cell. Municipal and postal area codes were added using the ArcInfo's Identity function with municipality and postal code coverage. The code was defined by the polygon of the region where the point was situated. Figure 1 shows points and municipal boundaries together. The cell indicated by an arrow is given the code 423, because the lower left-hand point of the cell is situated on the area of municipality 423.

¹ Prepared by Maija Leppänen.

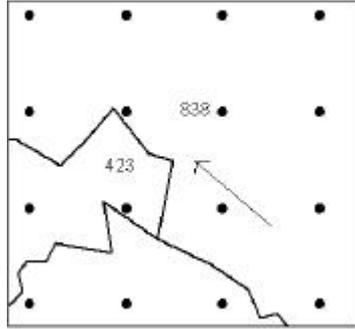


Figure 1

6. Another point coverage was also generated using the upper right-hand corner of a cell, and another Identity function was made for those cells that did not obtain a code in the first phase. The point coverages were then joined and converted into polygon coverage.

7. A grid net with a cell size of 250:250m was made this year in the context of a commission for the Ministry of the Interior for use in rescue services. The method used was based on polygons from the beginning. A grid net was generated with ArcInfo, using as the origo the lower left-hand corner of a rectangle covering Finland. The municipal boundaries were converted to ArcInfo grid format using the cell size of 250:250 m. A municipality code was used as a cell item. If a cell was on the area of two or more municipalities, it was assigned the code of the municipality in which its largest area fell. The resulting grid was then converted back to polygon coverage. The grid net and the polygon-based municipal coverage were then combined and the cells were given municipality codes as a result. In the same way, postal codes were added using postal code boundaries. The function used in the last phase could be intersection or union, depending on the result desired.

8. Figure 2 describes the method with municipal boundaries and grid net. As can be seen, the result in this case may be different from the one obtained with the point-based method. The indicated cell can now be assigned code 838, because its largest area falls on 838.

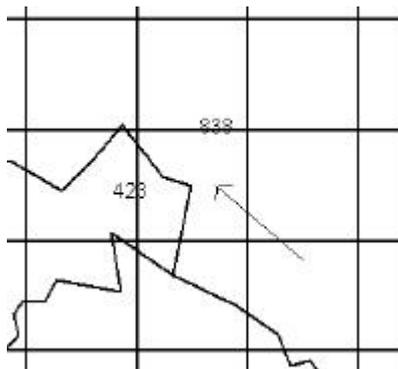


Figure 2

III. EXPERIENCES

9. Using points from the beginning does not demand great hardware resources, but the method itself and the results were not very good. Slightly better results could have been obtained using the middle point of a cell. There was also some manual work because not all the cells were given an area code.

10. Sufficient hardware is a problem with the polygon-based method. It requires plenty of continuous disk space (many gigabytes) because ArcInfo needs a lot of temporary disk space when generating topology. In practice, the work to produce a grid with the cell size of 250:250 had to be undertaken in parts. At first, there were over 12 million cells in the net before it cut along the boundary of Finland. The end result contains about 6 million cells. On the other hand, the method is much simpler and quite automatic, and the result much better, than the point-based method. The same method can also be used for defining other area codes (locality codes, etc.). The portion of water in one cell can be calculated using digital waterways.

11. The data can be aggregated in different ways, either traditionally via the mainframe or by using a relational database (SQL), or the aggregation can be made using GIS (e.g. ArcInfo's Frequency). As data are being aggregated, identification codes based on coordinates are also calculated for each cell. Different types of datasets or data from different time periods can then be joined to the same net.

12. Examples of gridnet application include comparisons of inhabited and uninhabited cells in different time periods or analyses of depopulating cells for the purpose of studying trends in regional population changes. It is also possible to compare different types of data, e.g. data on population and work places. If the grid net includes information about water, population density can be calculated quite accurately.

13. One important aspect of analysing grid data is to visualise them as thematic maps. With GIS software aggregated data can be joined to a grid net by using identification codes and it is then easy to extract cells from the net using area codes or coordinates, to make thematic maps, to change a thematic variable, etc. It is also possible to convert the grid net, or a part of it, into different GIS formats (e.g. ArcInfo's grid, ArcView's shapefile or Mapinfo format).

IV. WHAT NEXT?

14. Compared to administrative boundaries, a grid net is stable so, once made, it needs no updating. Only regional codes have to be updated when they change. If a coordinate system has to be changed, then the work has to be redone from the very beginning.

15. Using the grid net as ArcInfo coverage is somewhat cumbersome, as are conversions to other formats. We therefore have plans to save the grid net in a spatial database, which could solve these problems. It would then be possible to extract from the spatial database, and in the format of the used GIS software, only the desired part of the net.