STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Working Paper No. 11
English only

Work Session on Geographical Information Systems
(Ottawa, Canada, 5-7 October 1998)

Item 6 of the provisional agenda

RURAL VILLAGES AS STATISTICAL AREAS

Submitted by Statistics Finland [1]

CONTRIBUTED PAPER

1 Prepared by Marja Tammilehto-Luode.

I. NEED FOR STATISTICS ON RURAL AREAS

1. Statistics Finland has delimitated localities in every population census since 1960. Localities refer to clusters of buildings housing at least 200 residents, where the area covered by buildings normally does not exceed 200 meters. The definition of a locality takes into account the sparse population in Finland - or does it? The latest statistics from 1995 show that 81.1 % of the population in Finland live in localities, while nearly one million still live outside them.

2. Rural areas often refer to areas outside localities. The land area of rural Finland is 93.5 % of the total land area of Finland. Rural inhabitants are a group of people whose population structure is not very well known. There are statistics on rural areas as a whole, but not much regional data with which areas eligible for support or subsidies, for example, could be defined.

3. Studies based on grid data (Rusanen *et al.*) provide a more detailed picture of rural areas. With a grid of 1 km x1 km, the country is artificially divided into small parts. This type of data has its advantages. However, the grids themselves show neither functional villages nor do they refer directly to morphological villages. The GIS technology offers better ways for defining large rural clusters.

4. The huge area of countryside in Finland plays a very important role in the agricultural and rural policy of the EU. Structural Fund reform with a new objective programme, inter-regional co-operation operation, national initiative programmes, rural development package, community initiative, peripheral regions programmes - these are all words very much in the air. There is not a lot of data which could be applied in the planning, developing, auditing or evaluating of these programmes.

5. Ever-growing migration to large population centres is an alarming trend in Finland. Why this is happening and how we could stop or slow down the trend are questions for which appropriate information must be made available before they can be answered.

6. Ultimately, there is a need for more detailed statistics on the Finnish rural areas. Furthermore, new technology and new tools are available for producing statistics of this kind.

II. PREPARATION OF THE PROJECT

7. A project with a preliminary plan was presented to several authorities interested in statistics on rural areas. Statistics Finland suggested joint funding by all ministries responsible for regional policy plans. After several discussions it was concluded that, in the end, rural Finland falls under the mandate of the Ministry of Agriculture, as has been the tradition. There was no willingness to invest, but much interest in the power of influence instead, from the other ministries. The final solution was joint implementation of the project by the Ministry of Agriculture and Statistics Finland. Statistics Finland saw the project as a development project with major effects on statistical production in the future.

8. For Statistics Finland, the main goal of the project was to develop an "automated" method for delimitating small rural villages. A further goal was extensive utilisation for the first time of Finnish longitudinal census data in the preparation of time series of small area statistics. It was envisaged that rural statistics and statistical analyses could be offered flexibly and mainly according to specific demand. A summary of the used methods and rural statistics would be produced, but major data would only be published upon request.

9. The most crucial part of the work was defining a method with specific variables in the spatial algorithm for delimitating a rural village. The definition phase entailed interviewing several rural researchers, local authorities and other specialists.

10. There is no common or harmonised definition of a rural area, either at the international or national level in Finland. The definition varies depending on the criteria used (e.g. Gilg, A 1985, Malinen *et al.* 1994). A rural area can be descriptively defined by the common images that people have. A rural area can also be defined politically or administratively, as well as by its typical functions, distances to centres of population or by population density figures.

11. Although there are many research reports concerning rural areas, they do not provide concrete material for methods to delimitate rural villages. The reports use the statistical data already available, resulting in different figures depending on the criteria used.

12. It looked as if the only way to proceed with the dilemma was to try out different criteria. In any case, using the new method would provide much more detailed rural area statistics than was previously the case.

III. THE DELIMITATION METHOD

13. There were 8 major criteria for the delimitation method:

i) it should be capable of being processed similarly for the whole country;
ii) it should be applicable to rural areas and different types of communities;
iii) it should be as simple as possible;
iv) it should be efficient in terms of working time and other costs;
v) it should identify real as well as possible rural villages;
vi) it should provide tools for a continuous system of statistics on built-up areas from rural villages to town centres;
vii) it should be capable of identifying by name all villages from 50 inhabitants up;
viii) it should be capable of easily producing statistics by villages and defined group of villages.

14. In the Finnish register-based statistical system, buildings are the basic statistical entities. All buildings can be located by map co-ordinates. Distances between buildings can be calculated, as well as the number of inhabitants in the buildings. Delimitation of different types of areas can be made independent of administrative or other redefined areas according to the characteristics of the buildings (or dwellings, or families, inhabitants and enterprises in the buildings).

15. It was obvious that the buildings and their locations played the most important role in the project. It was decided that the variables derived from the characteristics inside buildings should be later used for statistics and the classification of the villages, but, first, the morphology of the villages had to be defined. The morphological delimitation of villages can be defined simply by using information on the locations of the buildings and the distances between them (Taajamat 1995, Localities, SVT, Population Census 1995, Volume 4. Statistics Finland). As for defining the localities, one of the most crucial factors in the algorithm for delimitating an agglomeration "automatically" is establishing the correct distance between buildings so that a building can be included in an agglomeration.

16. Initially, there were three different types of methods that were to be tested. The first method, known as the Thiessen Method (e.g. Tomlin 1990), is based on a Triangular Irregular Network (TIN). The method produces a network which covers the whole area in question. Each area has at least one point that is a building. The disadvantage of the method is that the net has large polygons

in rural areas, where distances between buildings are large. The sizes of the polygons do not indicate directly the distances between buildings, either. The network also covers the whole area and does not exclude uninhabited areas.

17. The second tested method could be called the Grid-and-Buffer Method (e.g. Cell-based Modelling with GRID 1991.). It creates a regular grid network for the whole area. Only grids with buildings are included to the first round. Adjacent grids are merged together to create clusters of buildings. Each cluster is then buffered to smooth out boundaries on one hand and on the other hand to merge two close clusters together if necessary. The method needs very high computing capacity and must be implemented in several stages. One disadvantage of this method is that the grid has to be located over a map. It does not necessary give the optimal locations in the sense of the buildings' spatial deviation. Another disadvantage is that a grid is a square so the distance between grids can be calculated using either the side or the diameter of the square. The effects of the square can be minimised by choosing an optimally sized grid.

18. The third tested method is the Buffer Method (e.g. Tomlin 1990). With this method, each point, i.e. each building, is buffered by a certain distance. Touching buffers are merged together. The method is somewhat similar to the Grid-and-Buffer Method, but is more precise and needs even more computer capacity than the second method.

19. The first method was rejected after testing because of the disadvantages mentioned above. The third method was too time-consuming and even critical in the sense of workstation capacity. The second method, the Grid-and-Buffer Method, was chosen for further testing. The method was cost-efficient, yet provided a good tool for delimitating the villages. The critical aspect of the method is that the right grid and buffer sizes are chosen to combine the effect of the distances between the buildings.

IV. PRODUCING THE BOUNDARIES OF VILLAGES

20. Basic data were needed on full-time residential buildings and other buildings and their locations, excluding, for example, leisure homes and buildings used only in agricultural production. Five pilot sites were chosen representing different community types in rural areas. One pilot site represented the southern agricultural areas, one the river valleys in western Finland, one the northern border areas and one the eastern rural hills. Prototypes were made by using the building data on the pilot sites. With these prototypes, different grid sizes and distances within the algorithm were tested.

21. During this testing period, statisticians interviewed local authorities showing alternative delimitations on the maps. The role of the local authorities was to find the best and the most realistic delimitation for rural villages – the best way of identifying the existing agglomerations. After this testing period, the best method was selected and the final production test was applied to the whole of the Päijät-Häme region.

22. When the final test area delimitation was completed, the panel of selected specialists verified the method to be used as standard for the whole of Finland. At the end of May 1998, when this paper is written, the project is at this stage.

23. The final production of boundaries will be completed in pre-designed and tested stages and by combining several parts of the whole database. The results will include a database of the boundaries of all agglomerations (ArcInfo-format) from one building to large urban areas. The number of inhabitants is entered afterwards and the final definitions of rural villages are approved with the

help of these inhabitant figures (or by combining them with other variables). All the agglomerations of more than 50 inhabitants will be given standardised codes and names.

24. Printouts of the final delimitations will be prepared for each region in Finland. The authorities of the regions (20 in all) will verify the delimitations and correct the names of the agglomerations where necessary. Preliminary names for the agglomerations are input as a digital namebase of a map to the scale of 1: 200 000.

V. REPORTING AND DATA SERVICE

25. The project will have two final reports. The first one will include the description of the method and the production process used, together with a summary of the types of statistics that can be produced with the databases available. The second report will be a research report, including a major analysis of the rural areas of Finland. The database will be available by the end of 1998. The reports are planned for publication during 1999. The main use of the database is planned for customised services. The data provide tools for better decision-making and allocation of resources for the planning, developing, auditing and evaluating of rural areas. The base will provide local people with real data to assist in making better applications of regional funds for the development of economic or social activity in rural villages. Regional authorities will also be able to justify eligible areas and allocate resources on the basis of better information. The data service concept will be adjusted to Statistics Finland's other small area statistical services with similar technical, substantial, legal and marketing solutions.

VI. EXAMPLES OF OUTPUT FROM THE DATABASE

26. Regional deviation of different sizes of villages is the very basic output from the database. The table may look as follows:

Table 1 : Sizes of rural villages by region

Number of inhabitants/ Name of the Region	10-50 inh.	50-99 inh.	100-149 inh.	150-199 inh.	200-499 inh.
Uusimaa					
Itä-Uusimaa					
...					
Lappi					
Åland					

27. Distances from large service centres are very important to the villages. The distances can be calculated using digital road data. It is also possible to take into account the road speed limits, for example, in estimating the time distance in question. Regional cross-tabulation may look as follows:

Table 2: Distances of villages from a service centre by size of village in the Region of Varsinais-Suomi

Distance	km						min		
	0-2	3-5	6-10	11-15	more than 15	less than 15	15-30	30-60	more than 60
Size of village									
1-9									
10-49									
50-99									
100-149									
150-199									
200-499									
...									

28. The functional areas can be defined by different criteria. For example, a study of commuting directions can show the dependency of the villages on certain centres (see an attached map). The characteristics of villages can be shown by different statistics. The following table is an example

Table 3 : Characteristics of village of Päijät-Häme region

	Number of inh.	% of inh. under 7	% of inh. over 55	Number of jobs	Major industry	Unemp. rate	School	Doctor's surgery
Kalliola								
Paimela								
Vesivehmaa								
...								
Whole region								

VII. CONCLUSIONS

29. Finland is a very sparsely populated country. If a 1 km x 1 km grid is placed over the country, less than a third, that is about 100,000, of the grids are inhabited. Of all the inhabited grids, more than 60% have less than 10 inhabitants. Statistics by grid squares (1 km x 1 km) contain serious confidentiality problems in rural areas. Grids by themselves are very abstract areas and do not refer to any commonly known population clusters.

30. Because of georeferenced buildings and good quality register-based data, Finland has good opportunities for making all kinds of GIS analyses at lower costs than those entailed by countries needing field surveys. Geography at a very detailed level is implemented, or can be implemented, in all register-based statistical entities.

31. Finland has a relatively large number of small villages, but what is considered small in Finland may be regarded as very small in other countries. The importance of small villages may look like an exaggeration to more populated countries. Small villages have an important role in the Finnish identity, innovativity and landscape. With high technology and good databases, creating a database of delimitated rural villages is not a pointless undertaking.

REFERENCES

Cell-based Modeling with GRID. Arc/Info User's guide. Environmental Systems Research Institute, INC.
Gilg, Andrew (1985). An Inroduction to Rural Geography. Edward Arnold.
Malinen, Pentti, Reijo keränen Heikki keränen (1994). Rural Area Topology in Finland - Tool for Rural Policy. University of Oulu, Research Institute of Northern Finland. Research reports 123.
Taajamat 1995. SVT, Volume 4. Population Census 1995. Statistics Finland.
Tomlin, Dana (1990). Geographic Information Systems and Cartographic Modeling. Prentice Hall, N.J.