Statistical Commision and
Economic Commision for Europe

Working Paper No.24
English only

Conference of European Statisticians
UN/ECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography,
Neuchâtel, Switzerland, 10-12 April 2000

Topic (iv) Spatial analysis in a statistical context and disclosure control procedures

GRID data and area delimitation by definition

Towards a better European territorial statistical system

Submitted by Statistics Finland, Statistics Norway and Statistics Sweden[1]

1. Introduction

Regional statistics of the EU are managed on the basis of the NUTS (Nomenclature des Unites Territoriales Statistiques) system. These units were established by Eurostat to provide a single, uniform breakdown to territorial units for the production of regional statistics for the European Union. For many applications, the NUTS framework is far from ideal. The system is proving insufficient mainly due to differences in the sizes of units in the same category and changes in the units over time.

Generally, the NUTS regions are the administrative regions of a country. A further problem with the NUTS system arises from the fact that in many cases the administrative division does not coincide with the division by the natural, or other, phenomena of interest. For example, using administrative areas for defining urban areas does not quite work, at least not in the Nordic Countries. Municipalities are not only different in their geographical sizes, but also in their numbers of residents and their spatial distribution. We have large municipalities with one or two big urban agglomerations, but several rural agglomerations and also small municipalities with only urban agglomerations.

Based on their own experiences, the Nordic Countries have promoted the use of grid type statistics for the whole of Europe and, especially, for Eurostat. It seems that most, if not all, national statistical institutes already use a system of grids for producing professional geo-statistics. The question, therefore, is whether a need exists for a common solution for all national statistical institutes, in the same way there is a common solution for producing statistics on administrative areas (NUTS). It has also been argued that standardised grid data are more comparable across countries than data based on different administrative areas (polygon data) are, even if the grids are estimated from base data obtained using polygons.

[1] Prepared by Marja Tammilehto-Luode, Lars Rogstadt and Lars Backer

2. Register-based systems and geo-coded points as basic statistical units

Finland, Norway and Sweden belong to the group of countries in which statistics production is essentially based on comprehensive registers. In addition, the approach to the geo-coding of statistical units is point-based instead of area-based in these three countries.

In the Nordic Countries, the prime sources for geo-coding are the centroids of buildings in Finland, centroids of real estate properties in Sweden and standardised addresses in Norway. In the three countries, digital road and street networks with addresses are also available for linking statistics by their address location to the geography.

The geo-coded points and their links to statistical units make the production of small area statistics quite flexible in the Nordic Countries. This may be the main reason why statistics by relatively small areas, such as grid squares of 1km x 1km, or even smaller, have been produced and used quite widely in Sweden and Finland and also being increasingly used in Norway.

3. Co-operation with GIS since 1995

Between 1995 and 1997, Statistics Finland and Statistics Sweden ran the first co-operation project on the usage of grid data and GIS. This first study involved a small number of variables from both countries' data records for the purpose of determining the viability of joint use of these data. As a matter of research interest, the objective was to analyse the similarities and differences in the spatial structure of the population between the two countries (Rusanen et al. 1997). The study showed the functional distribution of the population in large areas containing both urban and rural districts.

The first co-operation project concerned the "night-time" population; in other words, the population was geo-coded by their residences. The data also covered fully both countries and comparisons were made in respect of both urban and rural areas.

To gain more experience in the usage of grid type data, Statistics Finland and Statistics Sweden decided to continue the co-operation in 1998. The second project was aimed toward expanding the database to the daytime population and workplaces and toward testing smaller grid sizes for the analysing of urban areas. Examples were produced using different GIS methods.

The examples were chosen to represent the potential of the data in both the government and private sectors. The examples related to changes in the spatial structure of industries and to a market analysis of an industry. The examples produced with empirical data made a major contribution to the understanding of the processes and practices involved in constructing and displaying grid data for both countries (Tammilehto-Luode, Backer, 1999).

The experiences from the joint projects on the usage of grid data have been so promising that a Nordic initiative for a European project to implement grids as complementary to the NUTS hierarchy was made last year in Eurostat's Working Party on Statistics and GIS. It was proposed that the project should mainly cover standardisation of the spatial units for statistics and their definitions.

This Nordic co-operation entered its third phase this year with the first efforts aimed toward constructing a joint map of the population density in the Nordic Countries by using grid data from the four countries. At the same time, active co-operation is pending in many fields, including studies into the harmonisation and standardisation of grid type of data, harmonisation of the methods for delimiting urban settlements and co-operation for the development of the data dissemination processes and tools.

Thus far, a common population data set based on 1km x 1km grid squares has been compiled using data from the three national statistical institutes. The geo-coded grid data are presented in a common Universal Transversal Mercator projection, zone 33, WGS84 datum. They are presented as a map, with additional data describing the national borders, sea areas, major lakes and the NUTS3 and NUTS5 units. Table 1 shows some results from the statistical analyses of the data, including some particularly interesting figures on differences in the spatial distribution of population in the three countries.

Table 1: Background data and different population density figures in Finland, Sweden and Norway

Country	**Finland**	**Norway**	**Sweden**
	31.12.98	01.01.99	01.01.97
Total area - km2	338 145	323 758	449 964
Total land area - km2	304 593	306 253	410 934
Population - Statistical Yearbook	5 159 646	4 445 329	8 847 625
Population in grid data set	5086 018	4 416 527	8 474 007
Difference (%) between statistics and grid data set	-1.43	-0.65	-4.22
Inhabited grid cells	103 798	57 325	105 044
Inhabited grid cells – proportion of land area	34.1	18.7	25.6
Inhabitants per km2 land area	17	15	22
Inhabitants per inhabited km2	49	77	81
Population in urban settlement areas *	81	75	84
Maximum population in one grid cell	20 948	14 347	21 209

* For Norway, population in agglomerations with at least 200 inhabitants and 50 metres or less distance between houses.
Data for Finland and Sweden from 1995; for Norway from 1999.

4. Nordic Forum for Geo-statistics

The first drafts for a Nordic Forum for Geo-statistic were laid down during the ISI conference in August 1999. Statistics Sweden volunteered to organise the first Nordic GIS meeting for the statistical institutes, and this took place on 20-21 March 2000. The items on the agenda included harmonisation of the Nordic data sets, mapping with grids, delimitation techniques, dissemination projects and future challenges.

To make the results of these Nordic projects visible and also provide a “blackboard” for future distribution of information, a web site entitled “Nordic Forum for Geo-statistics” is due to be opened.

An embryo of a plan for future co-operation was also discussed during the Nordic meeting. It was agreed that the plan should include a list of scheduled project events and key milestones and deliverables. The key milestones should cover such items as:

- Understanding customer needs
- Partnership agreements
- Project development over the web
- Codes and geo-coding
- Nordic statistical databases
- A system of maps
- Dissemination products

A summary of the proceeding of the Nordic meeting will be published on the Web site to be opened.

5. Examples of Nordic experiences

In the UN/ECE Work Session on methodological issues in Neuchâtel, the Nordic Countries will present examples of their work with geo-coded harmonised statistics. The joint map displaying population density in the three countries with the help of precision statistics by 1km x 1km grid square will also be made public. In addition, there will be presentation concerning the pertinent ongoing work in these three countries.

5.1. How grid data have been applied

The Finnish examples will show the different types of applications in which grid data have been used in Finland.

Grid data generally refer to those statistics on population, labour force and employment in which a regular grid square constitutes the statistical area. The location of the squares is pinpointed with map co-ordinates. The most frequently used square size is 1km x 1km.

Data by grid square have been available from Statistics Finland since the 1970 population census. Basing on orders from customers, co-ordinates have been added to the pertinent register-based statistical data using the centre point co-ordinates of buildings. The unit-specific data that are definable by building co-ordinate have been summed up to grid squares. The customer has received for his/her disposal statistics in which the grid square co-ordinate (bottom left hand corner) represents the area code.

Grid square data are regarded as the most flexible statistical geographical data. The square is a spatial stable. It does not move from one year to the next, as administrative areas may do. Thus, it is independent of any regional changes. Data by small squares can easily be summed up to form larger areas.

Exploitation of data tied to time and place on individuals and their activity produces findings that cannot be observed with analyses based on administrative regional units. Besides studies on depopulation, applications of grid square data have in Finland included, for example, studies concerning regional incidence of diseases (Karvonen et el. 1997), unemployment (Muilu et el. 1998), changes in rural industrial structure (Räisänen et el. 1998), regional segregation (Vaattovaara 1998) and migration (Kauppinen et al. 1997).

Due to the availability of geo-coded register-based data, the opportunities for exploiting socio-economic geographical data are exceptionally good in Finland. However, the use of geographical data produces several major problems, or challenges, relating to such issues as data security, data quality, know-how of thematic mapping/visualisation and copyright.

5.2. Constructing a hierarchical grid database

5.2.1. Introduction

In order to improve the geo-statistical system for northern Europe we are confronted with mapping an area as big as Germany, France and Italy combined but with merely 1/10 of their population. The population is not much concentrated but still leaving large sparsely populated areas. Nuts regions in Sweden are on most levels huge compared to their counterparts in Germany. When comparing the total population in these regions the difference is much less marked.

Swedish Statistical Yearbook for 1998:

Country	Area	Population	Density
Sweden	450	8587	20
Denmark	43	5146	121
Finland	338	4998	15
Iceland	103	207	3
Norway	324	4248	13
Nordic Countries	1258	23186	18
Germany	357	81539	228
France	552	56634	105
Italy	301	59104	51
GFI-Combined	1210	197277	163
Canada	9971	27297	3
USA	9364	248710	27

In our efforts to prepare the ground for a harmonised geo-statistical system for Northern Europe, we seem generally to be satisfied with the Nuts system as it is. The problem appear when comparing statistics with the other parts of Europe. Within our region, Nuts 5 and larger systems of regions are useful for most purposes where administrative areas are concerned. But, on a smaller scale it is difficult to suggest a common building-block (an irregular tesselation). Compared to the relative size of municipalities here we therefore use 1 km x 1 km grids as the next basic building block below this level. From this module all other (smaller and larger) grids are constructed.

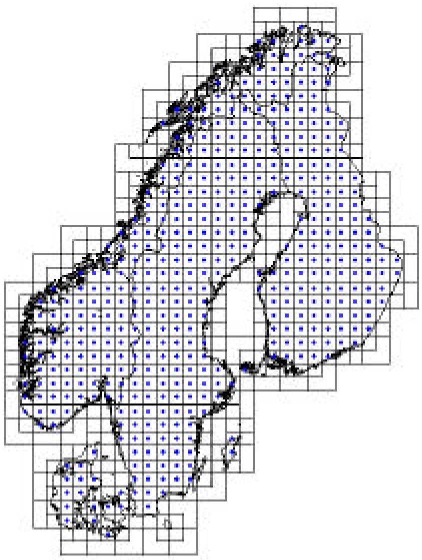
5.2.2. Data and data management

The data set used for testing consisted originally of population statistics only. The next step will be to expand the database and introduce information on business and industrial enterprises and employment on grids. The aim being to start the building of tested methods for the delimitation of all types of (land use) concentrations from urban cores to not inhabited regions, and then later perhaps to describe “Nature” from the mountains down to urban parks. The NSI’s of the Nordic countries have in co-operation produced a CD-ROM product (“Statistics beyond borders”) with harmonised statistics (on administrative areas) for this region. GIS experts from these countries are involved in both these efforts, and will naturally try to dissolve any contradiction.

In order to cope effectively with the relatively large databases needed to store and aggregate data for the whole region from micro level to large regions on an ordinary PC, a very effective DBMS system is required. For this purpose we have imported the Swedish data set into a SuperStar database[2]. This software will be used for the dissemination of the results of our next census 2005. This DBMS is very fast but has a simple mapping tool that does not handle grids. We have

[2] Superstar,SuperCross and Supermap are registered trademarks belonging to Space- Time Research Pty Ltd.

therefore developed a coding system and a registration solution that fools the mapping module to do that too.
On the mapping side this solution has posed some technical problems that we have to solve. First we need to agree on a common projection as well as a co-ordinate system that may be used for the presentation of administrative regions and natural features as well as a suitable grid structure in a way that is natural for us. We have as a preliminary solution adopted a UTM projection (zone 33) using a EUREF89(Finland and Norway) /WGS84(Sweden) projection. Area and distance units are meters, km and km2. The use of a compiled database as implemented by Supercross demand a clean hierarchical system of regions (regular or irregular Nuts codes for Nuts regions work well but have to device a code system for the quadtree solution. The code system must be applicable to most different types of fast DBMS in addition and not be tied to this special case. The solution is very simple: In the ordinary Swedish national co-ordinate system 7 digits are used to get down to meter level. Thus 7+7 digits are needed to denote a km grid using the lower left corner co-ordinates. By combining the first digit from the x co-ordinate with the first digit from the y co-ordinate we get the (2-digit) code for a 1000km grid. By adding the second x co-ordinate and the second y co-ordinate we get a 100km grid inside the last grid. 10 digits are needed for the 1km grid. To divide this further we add a new system that divides a 1000m grid into 4 500m grids. The first lower left is called 00 the next to the right is 01, the one on top of 00 is 10 and the last 11. The 250m grids have the codes 0000,0001,0010,0011, 0100,0101,0110,0111, 1000,1001,1010,1011, and 1100,1101,1110,1111. With the same logic we might (if needed) divide this grid into smaller units.



It is essential that the integrity of individual information is defended in agreement with the laws protecting individual integrity. The first defence is that statistics are not stored with individual (personal) codes, and secondly they are stored aggregated up to the lowest grid level (here 250m). In addition to this the databases are compiled and not accessible without breaking the code. Still this might not suffice to protect micro-data from 3 independent national states (2 of which are members of the European Union and one not (Norway))
The performance of this database will be demonstrated at the ECE conference in Neuchatel in April 2000.

5.3. Delimitation of urban settlement areas based on the use of administrative registers and GIS - a Norwegian approach[3]

5.3.1. Introduction

Statistics Norway has started to work with automatic methods concerning the delimitation of urban areas. In addition to clearly documented user needs, there were at least three other factors that gave momentum to the work which started in 1997: co-ordinates were assigned to buildings and addresses in the official Ground property, Addresses and Buildings register (GAB-register); digital, vectored basic maps to proper scale became available and geographical information systems (GIS) were introduced at Statistics Norway.

Statistics Norway have now developed a register-based and automatic methodology for the delimitation of urban settlements, which has been incorporated in Statistics Norway's operations since 1 January 1999.

5.3.2. Delimitation of urban settlement areas

Since 1960, urban settlements have been delimited at ten-year intervals in connection with Population and Housing Censuses. The boundaries were originally designed on analogous maps by the local staff of each of our now 435 municipalities and digitised later on by the Norwegian Mapping Authority. All in all, this was a very costly process that produced unharmonised descriptions for urban settlement areas.

5.3.3. Statistics Norway's definition of urban settlement 1998

A hub of buildings shall be registered as an urban settlement if it is inhabited by at least 200 persons (60 - 70 dwellings).

The distance between the buildings shall normally not exceed 50 metres. Distances of more than 50 metres are allowed in areas that cannot, or may not, be built up. These can be, for example, green parks, facilities for sports, industrial areas and natural barriers, such as rivers or arable land. Smaller hubs of buildings that naturally belong to an urban settlement should be included if situated at a distance of up to 400 metres from the main urban settlement.

Computerised and automatic delimitation of urban areas is done in two separate steps. First, the number of residents is geographically distributed to the co-ordinates of addresses or buildings. Second, a geographical information system (GIS) is used to aggregate polygons of urban settlements according to an agreed set of criteria. In January-February 2000 we improved the method by increasing the used data sources – the most important aspect being that we created a digital road database in order to link hubs of buildings in a way that is geographically more accurate. Exhaustive documentation on this is available (*SSB 1999/17*).

[3] Based on a paper by Per Schøning, Division for environmental statistics, Statistics Norway, per@ssb.no

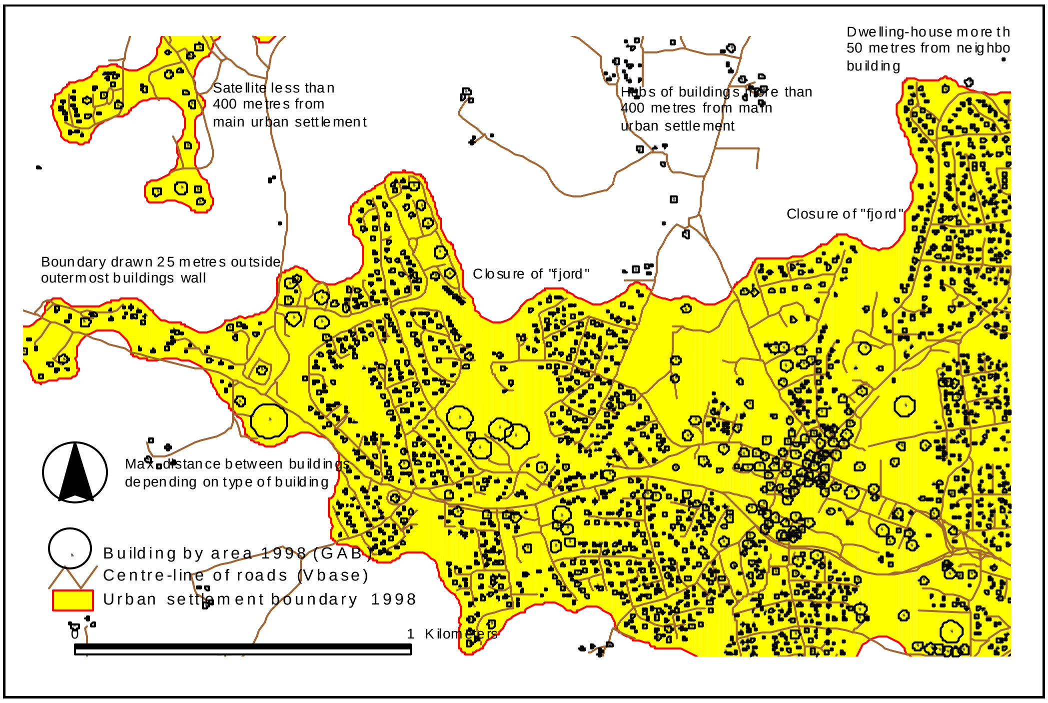


Figure showing the principles of the automatic GIS method

Urban settlements have been delimited for 1998 and 1999 using the new methodology. The preliminary figures for 1999 show that we have 970 urban settlements in which 74.6 per cent of our resident population live and which make up 0.7 per cent of our total mainland area (further details at: *www.ssb.no*).

Statistics Norway has also used this method to aggregate centre areas within the urban areas, basing on a preliminary set of criteria for the localisation of commercial and administrative centre areas in urban settlements (a pilot project in co-operation with the county administration in the capital area - Oslo and Akershus).

References

Backer, L., Rogstad, L., Tammilehto-Luode, M. (1999). Use of grids to improve the comparability of statistical data. Working document E/GIS/17. Meeting of the Working Party. “Geographical Information System for Statistics”. Luxembourg, 21-22 October 1999.

Harala, R., Tammilehto-Luode, M. (1999). GIS and Register-based Population Census. Statistics, Registers, Science. Experiences from Finland. Edited by Juha Alho. Statistics Finland.

Karvonen, M., Rusanen, J., Sundberg, M., Colpaert, A., Naukkarinen, A.,Tuomilehto, J., for the DiMe Study Group (1997). Regional differences in the incidence of insulin-dependent diabetes mellitus (IDDM) in Finland during 1987-1991. Annales Medicin, 29, 297-304.

Kauppinen, J., Rissanen, E., Rusanen, J., Naukkarinen, A., Muilu, T., Colpaert, A. (1997). Migration as a function of population. Nordia Geographical Publications, 26, 17-27.

Muilu, T., Rusanen, J., Naukkarinen, A., Colpaert, A. (1999). Local Poverty in Finland 1995. Paper presented at the IGU Commission meeting in Albuquerque, New Mexico, USA, 21-22 June, 1999.

Rusanen, J., Naukkarinen, A., Colpaert, A., Muilu, T. (1997). Differences in the spatial structure of the population between Finland and Sweden in 1995. Research Reports 221. Statistics Finland.

Räisänen, S., Rusanen, J., Naukkarinen, A. (1996). Socio-economic grid data and GIS for analysing the changes in the Finnish countryside. In Second Joint European Conference & Exhibition on Geographical Information, 27-29 March 1996, Barcelona, Spain. Geographic Information from Research to Application through Co-operation, Vol. 1. 651-660.

SSB (1999/17): Computerised delimitation of urban settlements - a method based on the use of registers and digital maps. Statistics Norway. Department of Economic Statistics. Documents 1999/17.

SSB (1999/21): Land use statistics for urban settlements - Method based on the use of registers and digital maps. Statistics Norway. Department of Economic Statistics. Documents 1999/21.

Tammilehto-Luode, M., Backer, L. 1999. GIS and Grid Squares in the Use of Register-based Socio-economic Data. Bulletin of the International Statistical Institute. ISI 99. 52nd Session. Proceedings. Book 1.

Vaattovaara, M. (1998). Pääkaupungin sosiaalinen erilaistuminen. (Social differentiation within the Helsinki metropolitan area, in Finnish with Abstract in English). Tutkimuksia (Studies) 1998:7. Helsingin kaupungin tietokeskus (City of Helsinki Information Centre). Helsinki.