Geostatistikk og bruk av grid i Statistisk sentralbyrå

Vilni Verner Holst Bloch

Vilni Verner Holst Bloch: Geostatistics and use of grids in Statistics Norway

KART OG PLAN, Vol. 68, pp 181–191. P.O.B. 5003, NO-1432 Ås, ISSN 0047-3278

Geographical information has traditionally been about what can be «seen from above»: physical features that can be drawn on a map. But planners, researchers, managers and politicians also need spatial information on societal topics. To meet these demands, Statistics Norway will deliver geographical statistics on a range of different sosio-economic topics to the national spatial data infrastructure. The statistics will be based on standardized grids.

Key words: Geospatial data, grids, Public enterprises, Private sector, Inspire

Vilni Verner Holst Bloch, Adviser and MSc Landscape ecology and resource geography, national contact point European forum for geostatistics, Statistics Norway, NO-2225 Kongsvinger.
E-mail: vvh@ssb.no

Background

Statistics Norway presented population statistics on 250 x 250 meter grids in the municipal reports of The Population and Housing Census 2001. However, the first use of grids in Statistics Norway was in 1875 (Lie and Roll-Hansen 2001, p 284-285)

Today it is possible to geo-reference (geographically locate) people, enterprises, buildings, etc., with their respective attributes. Examples of attributes could be gender, age, education and income for people, number of employees or type of economic activity for enterprises, and size and type of use for buildings.

Connection of personal id-number, id-number for enterprises, id-number for buildings etc. to a numerical address with coordinates, allows a range of socio-economic variables to be presented geographically. One of the easiest ways of presenting these variables is by using a square grid. This allows the statistics to be presented in a new way, and different types of data can be combined that lack a logical/numerical link but do have a geographical link.

At present, the only statistics produced regularly for delivery as grid data are total population on 1x1 km and 250x250 m grids. In addition, a number of registers contain entities that are identified by coordinates and are updated regularly. Statistics Norway can therefore produce cost-effective grid maps on many topics. Confidentiality and quality issues must be considered carefully for each topic and for the given level of detail.

The grid and other regional units

The standard grids are defined in UTM zone 33 with datum WGS84. Standard grid cell sizes are 1 x 1 km, 250 x 250 meters and 100 x 100 meters. All grids are free and downloadable from www.gridclub.ssb.no. The grid will be downloadable as shape-files (shp), ArcInfo export format (e00) and sosi-files (Norwegian national exchange format for digital maps) by the end of 2008.

As mentioned above, the only standard digital grid data produced regularly each year are total population on 1x1 km grids. Grid maps of total agricultural area in use are also produced annually in the publication «Jordbruk og miljø» («Agriculture and environment»). Grid based statistics for other subjects are produced on a project basis and for presentation purposes only.

Grids are used both for sampling surveys and for presentation. The metadata attached to each grid can vary and should be harmo-

nized with international standards. This would also facilitate more advanced geographical analysis across disciplines.

Administrative and statistical divisions such as counties, municipalities and basic regional units vary in size and shape and may contain both large uninhabited areas and areas of dense population. The basic regional units are the smallest regional units for which Statistics Norway publish statistics, and they also give the most detailed picture of settlement patterns. Although one criterion for delimiting the basic regional units was homogeneity (for example of population structure), some of the basic regional units contain both rural and developed areas. Maps produced using the basic regional units may therefore not give a sufficiently correct and detailed picture of the settlement pattern for some uses.

Statistics Norway has assigned an address coordinate to each registered resident of Norway. All statistics which can be related to the birth number can in theory be geo-referenced. Using GIS, statistics can be aggregated in other ways than those traditionally used in statistics presentations. For example, statistics can be aggregated onto a new, uniform framework such as square grid cells, in which all units have the same size and shape. Figure 1 shows the population distribution in Norway using grid cells. Figure 2 shows population distribution in all member countries of *The European forum for geostatistics* who have population data at the address level.

Grid cells do not replace other types of statistics divisions such as basic regional units. Grid cells are rather a supplement that will allow new uses of statistics. The users of grid

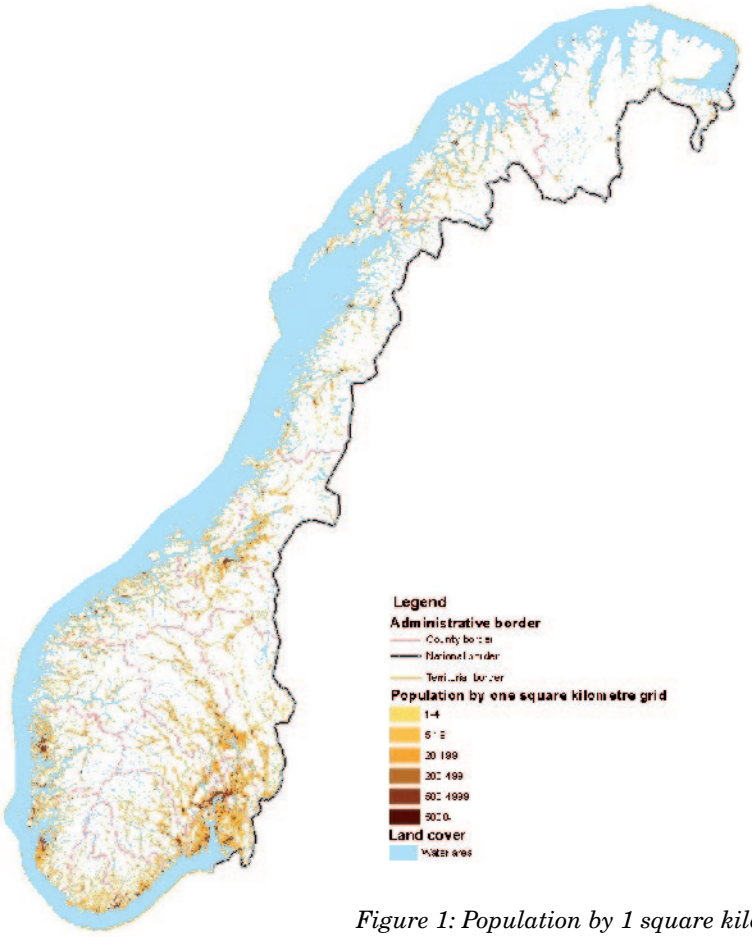

Figure 1: Population by 1 square kilometre grid. 1.1.2007

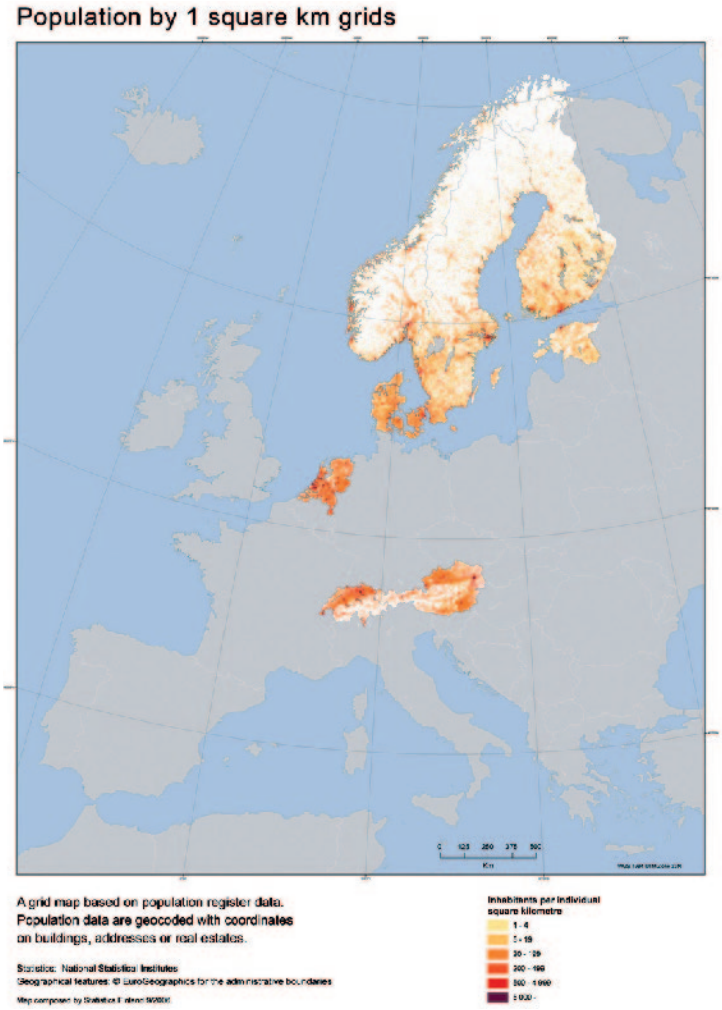

Figure 2: Population by 1 square kilometre grid. Members of European forum for geostatistics

cells have indicated to Statistics Norway that there is a growing demand for this kind of product.

Confidentiality

Due to the Statistical Act § 2-6, Statistics Norway as a main rule does not publish tables containing fewer than 3 units in a group (table cell) in which the sampling method can allow identification of individuals.

Grid cells are *table cells* in the sense used in the Statistical Act. In a grid map of population using 1x1 square km grid cells, a number of grid cells will have one or two observations. In Norway, electing this grid cell size will result in about 3 400 and 3 800 cells with 1 or 2 persons respectively (6-7 per cent of the approximately 55 500 inhabited grid cells in Norway). In comparison there are about 13 400 inhabited basic regional units (grunnkretser). The 7 200 cells with either 1 or 2 persons constitutes 12.8 per cent of the inhabited grid cells.

These cells with 1 or 2 observations are unevenly distributed between urban and rural areas. In the sparsely populated northernmost county of Finnmark (see figure 3), almost every fifth inhabited cell (20%) has fewer than three persons. Only 7% of the cells have fewer than three inhabitants in the more densely populated county of Akershus.

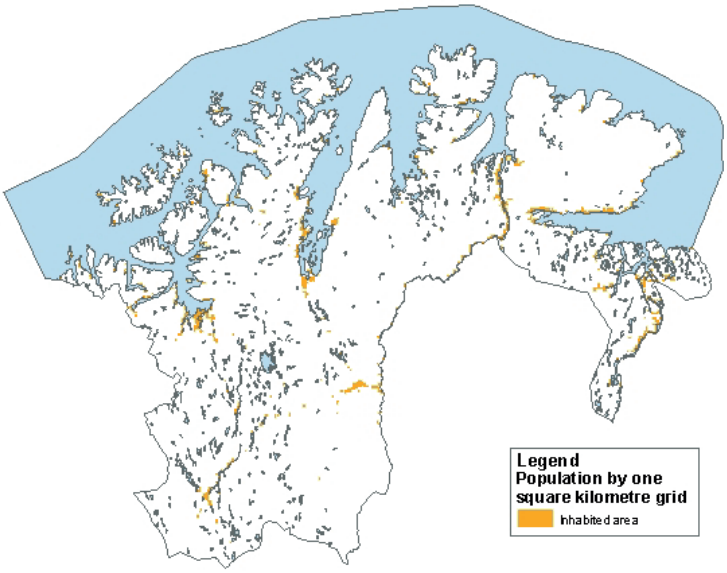

Figure 3: Population by 1 square kilometre grid. Finnmark county. 1.1.2007.

Confidentiality issues also arise when other geographical information is combined with the grid net. For example, combining the grid with digital municipal borders or a digital road network will give information on where the cells are located. If these grid cells are to be anonymous, several issues need to be considered. Must the sum of values from the grid give the correct population total for the country, or is it sufficient to display all inhabited cells without necessarily giving the exact number for every cell? Is three observations per grid cell an acceptable lowest value in terms of maintaining anonymity, or should the threshold value be increased?

The board of confidentiality at Statistics Norway was asked (Ottestad, 2006) to consider various methods for handling confidentiality and to set criteria for disclosure control. The following methods were discussed:

– Suppression method
– Heldal's method
– Least value method
– Larger grid cells
– Clustering method
– Average method

The board of confidentiality recognized that all of these methods might be used, but that they would produce different results. Methods which do not display all inhabited grid cells were considered to be poor solutions from a user perspective. Changing the grid size and/or shapes would also reduce user friendliness. Setting limits or cut-off values for individuals presents challenges with regard to households. When variables other than residents are used (for example households) one should consider setting a threshold for the number of inhabited addresses in each grid cell. At present we lack a satisfactory overview of households in Norway, and hence cannot set limits for the number of households per grid cell. An alternative would be to set the threshold number of individuals per cell so high, that it normally would include more than one household.

Viewed as an isolated piece of information, presentation of the fact that there is one resident in a square kilometre is not in conflict with confidentiality rules. However, later publication of other socio-economic variables or information about people in grid cells where the number of persons is low would be in conflict with the confidentiality rules.

The board of confidentiality at Statistics Norway concluded that publication of population statistics (number of persons) on a one square kilometre grid should not include exact values for grid cells containing fewer than 10 persons. Statistics Norway will therefore use the following values for grid population statistics: 0, 1–9, 10, 11, 12 and so on. For grid cells with 1–9 persons the value is set to 5.

Examples of use of grids

A grid product is useful for many different kinds of users and applications, for example: public management and planners, research, marketing and planning of private services (telecommunications, establishment of enterprises, banks, building projects etc.). The product gives general information about settlement and other patterns in the country, and also allows for additional kinds of analysis.

The examples and figures below present different uses of grids and highlight the possibilities offered by the grid approach. Note, however, that methods of representing and distributing geostatistics are likely to change in the future due to changes in use of variables, classes, confidentiality level etc.

Geostatistics on population

The central register for the population (DSF) includes all inhabitants in Norway. The register does not include coordinates but can be linked to the official address register (GAB – see next chapter) by a numerical address identifier. Because of a lack of concurrence in numerical address between the registers, not all of the population is identified by coordinates. Also, a few addresses in GAB lack coordinates or have wrong coordinates. As of January 2008, 99.7 per cent of the population was identified by coordinates.

Statistics derived from population on grid (figure 1) show that 55 545 km^2 of Norway's total 323 824 km^2 have at least one resident. The average number of residents per inhabited km^2 is 84, while the average population density in Norway is 15. Extensive areas are without residents and the population is concentrated along the coast, fjords and valleys. The density varies among counties. The smallest and most densely populated county is Oslo with 254 inhabited km^2 and an average of 2137 residents per inhabited km^2. The county of Sogn og Fjordane has the fewest residents per inhabited square kilometre.

A simplified model like the one shown in figure 4 provides information with adequate

Figure 4: Example: Identification of residents within a distance from a hospital.

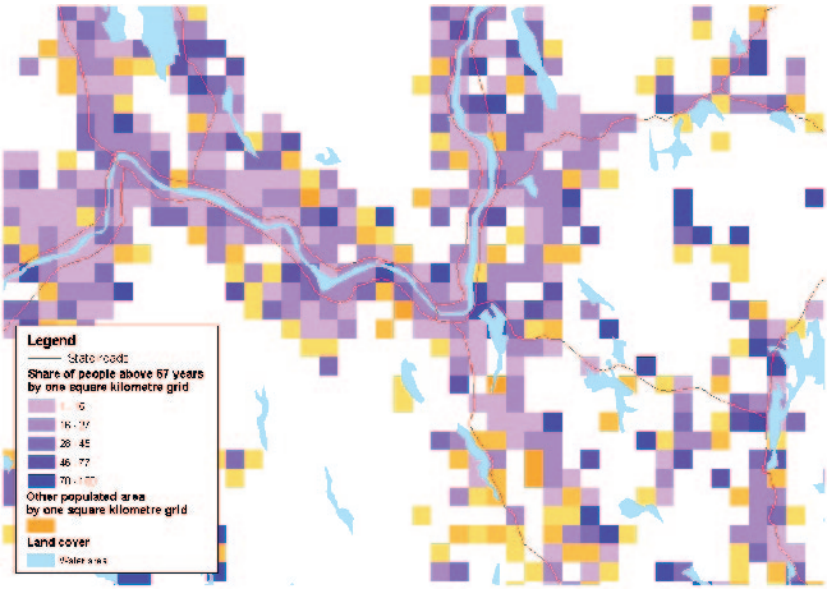

Figure 5: Example: Percent of elderly people (67 and above) in each grid cell.

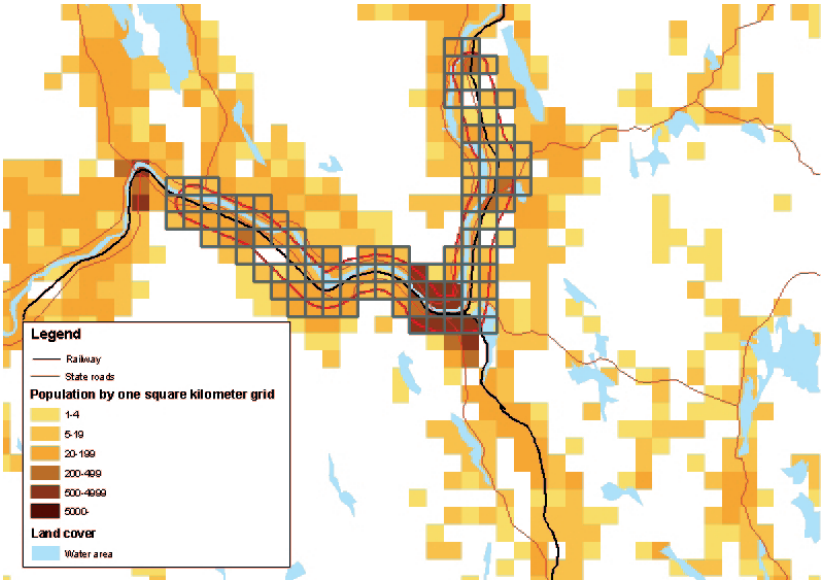

Figure 6: Number of residents within a distance (ex. 1 km) from a road

accuracy for many purposes. In this case one can easily count the number of people within reach of an ambulance helicopter, given some time/spatial limits. Aggregating geodata on grids with a given cell size will also reduce runtime for searches within a radius.

In figure 5, information about the age of residents at the address level has been used to create statistics about percentage of elderly people within each grid cell. This could serve different purposes; it could for instance indicate vulnerability to certain diseases or the

potential need for day care within an area. Intervals can be adjusted to the needs of users.

Figure 6 shows an example of using «radiation» from a linear element to count the number of people affected within a buffer zone. In this case the linear element is a road which «emits» noise or affects aesthetics in the surroundings.

Building activity

The official registers for Ground-properties, Addresses and Buildings (GAB), consists of three mutually linked registers where the A- and G-part comprise all addresses and ground-properties. The B-part comprises information on all buildings larger than 15 m^2 including their co-ordinates and information on building type, area of largest floor, etc. The G-part comprises information about all ground-property size etc. Each address, building and ground property is located by a point coordinate. The Norwegian Mapping Authority is responsible for maintaining the register.

Figure 7 presents statistics on dominating building type on a square kilometre grid. 116 753 km^2 of Norway's total 323 824 km^2 have at least one building. 29 414 square kilometres have buildings but no residents. In

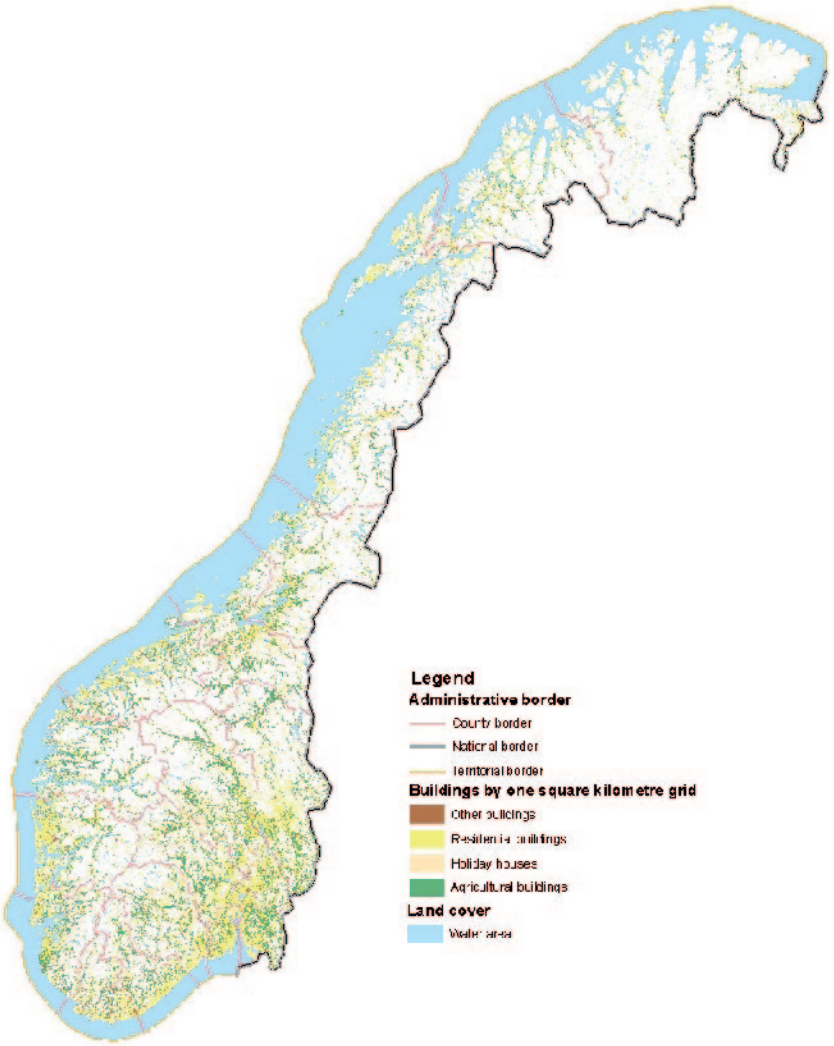

Figure 7: Building type by 1 square kilometre grid. 1.1.2007

the grid squares with at least one building, the mean number of buildings is 32 and the median is 9. The square kilometre with highest building density (number of buildings) has 1828 buildings.

The total number of square kilometre grids with at least one new building in the period from 1.1.1985 – 1.1.2007 was 65 508. 75 per cent of these grids had fewer than 8 new buildings.

Economic activity

The Central Register of Establishments and Enterprises (CRE) is a statistical register administered by Statistics Norway. Each enterprise is identified by coordinates. The CRE include all relevant enterprises and establishments in both the private and public sector. The register includes information on organisation number, name, address, code of economic activity (NACE), number of employees, turnover, status/ activity, etc. In the initial geo-referencing project 75.4 per cent of the enterprises were identified by coordinates (Myro and Torp 2002).

Figure 8 show statistics on number of employees from the register of establishments and enterprises, based on the enterprises

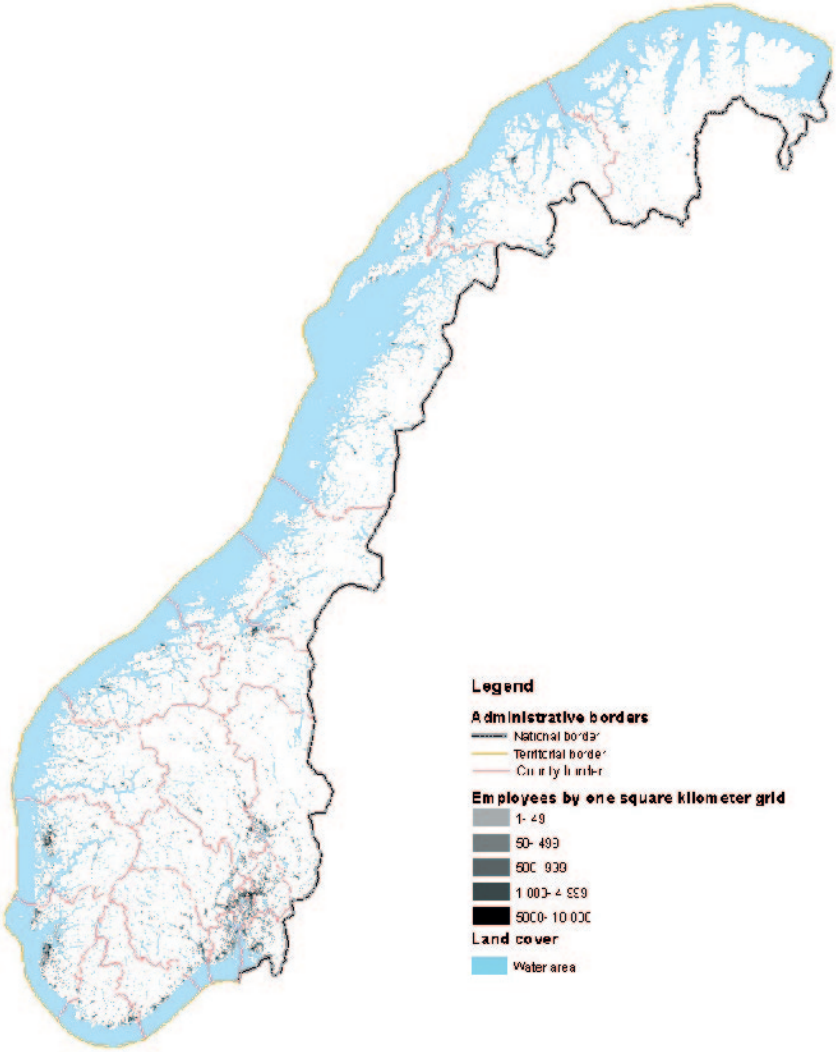

Figure 8: Employees by 1 square kilometre grid. 1.1.2007

which are geo-referenced by coordinates. 12 776 square kilometre grids contain enterprises which have employees in Norway and 50 per cent of these grid cells have fewer than 5 employees.

Distribution of agricultural areas

The administrative farms register comprise all ground properties with at least 5 000 m^2 of land for agriculture and / or at least 25 000 m^2 of forestry land, regardless of activity. The administrative farms register is a basic unit register including key identification to other administrative registers on activity and subsidy arrangements. The administrative farms register is geo-referenced by the coordinates of the central farm building or the geographically central point on the ground property.

Figure 9 shows the distribution of agricultural area in use on a 1x1 km grid (from Bye, Sandmo and Berge, 2006). This information about agricultural land «in use» could be analysed together with road networks to find time or distance to markets, or with population data to display areas under pressure. It could also be combined with data on agricultural maps to find areas of high value for potential agricultural use.

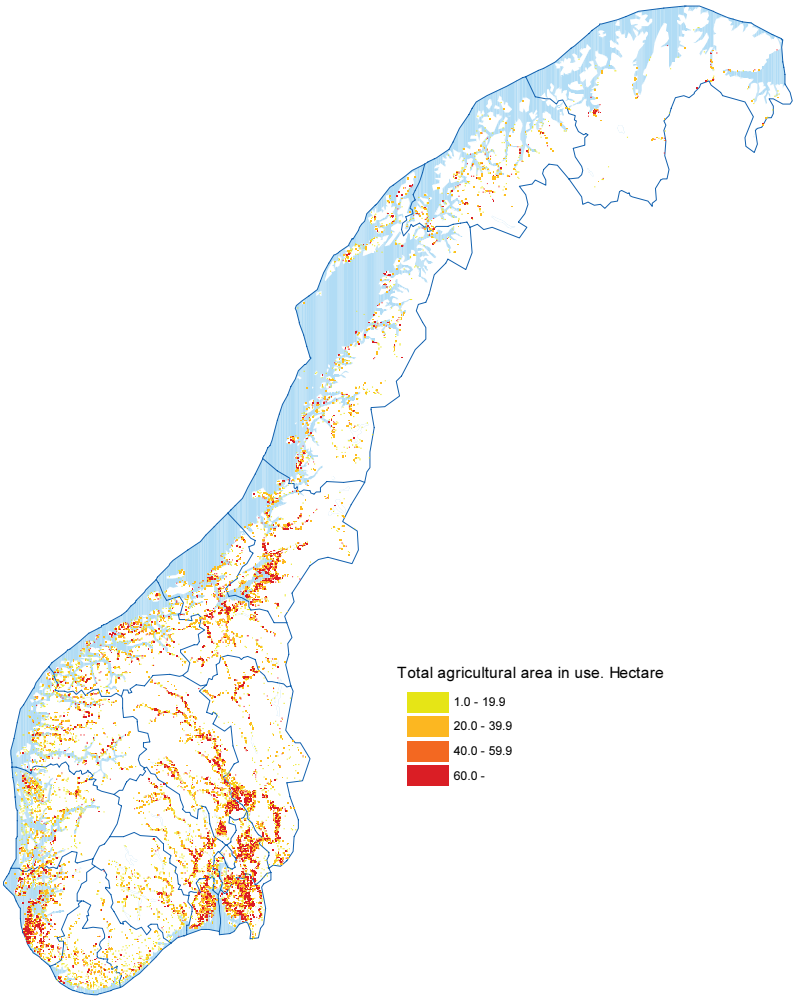

Figure 9: Total agricultural area in use. 2005. Hectare per km²*

** Preliminary figures*

Literature and further reading

Bye, Anne Snellingen, Sandmo, Trond og Berge, Gisle (2006): *Jordbruk og miljø. Resultatkontroll jordbruk 2006*. Rapport 2006/ 37. Statistisk sentralbyrå 2006.

GridClub of Europe (2007). http://www.gridclub.ssb.no/

Heldal, Johan and Fosen, Johan. (2001): *«Statistisk konfidensialitet i SSB. Et diskusjonsnotat»*, Notater 2001/28, Statistisk sentralbyrå. (Statistics and confidentiality in Statistics Norway. A discussion note)

Lie, Einar and Roll-Hansen (2001). *Faktisk talt. Statistikkens historie i Norge. ISBN: 9788215001500*. Universitetsforlaget 2001.

Myro, Svein and Torp, Camilla (2002): *Stedfesting av bedrifter i Bedrifts- og foretaksregisteret. Hovedprosjekt*. Notater 2002/ 27. Statistisk sentralbyrå 2002.

Nordic Forum for Geo- statistics (2006). Norway, Kongsvinger, Skansgården 12th – 14th September 2006. http://www.ssb.no/omssb/kurs_seminar/nordic/

Ottestad, Arne Knut (2006). *Metode for presentasjon av statistikk på rutenett. (Method for presentation of statistics on grid)*. Note to Board of confidentiality at Statistics Norway. 22.09.2006.

Statistics Finland (1995): *Differences in the spatial structure of the population between Finland and Sweden in 1995*. Research report 221.

Takle, Mona (2002): *Befolkningsstatistikk på rutenett. Dokumentasjon*. Notater 2002/ 22. Statistisk sentralbyrå 2002. (Population statistics on grids.)

Valen, Thorleiv (2006). *Minutes from meeting in Board of confidentiality at Statistics Norway.* 20.10.2006.