Statistical grids as tool for small area analysis

Ingrid Kaminger, Gernot Katzlberger and Norbert Rainer, Statistics Austria

Summary

In recent years geo-referenced as well as point-based data of official statistics have been established in various countries and are today becoming an integral part of the data dissemination system of regional and spatial statistics. Thus, the increasing demand of spatial statistics and, especially, of small area statistics can now be better met.

Within this context of new ways of establishing and disseminating spatial statistics, statistical grids play an important role. Grids represent a hierarchical system of square grid nets covering the whole territory of a country or even beyond, with a fixed origin. Grid sizes can vary from 100m by 100m up to 10km by 10km or more, thus being quite flexible in the scale of data resolution.

Statistical grids can therefore be viewed as a spatial reference system that can serve as the smallest statistical area unit for which – respecting statistical confidentiality – data may be provided to the user. Individual point-based data within an individual grid cell are aggregated, in the same manner as individual statistical data are aggregated for dissemination purposes. Grids data are thus grid-based statistical data.

Statistical grids are independent from administrative areas or boundaries; they are evenly distributed over the territory; they may be manipulated by standard GIS tools, they are easy to generate from point-based data and grid data can be combined with other (statistical or non-statistical) data, such as land use or other spatial data.

Statistical grids are therefore a flexible tool for data analysis. The main strengths of statistical grids are certainly the spatial flexibility to aggregate data according to administrative or other functional areas, and the independence from administrative boundaries. This paper aims to provide an overview of the main strengths of statistical grids, with special reference to small area analysis. The paper is based on the experience gained in Austria in the past few years. Therefore also the examples and illustrations refer to Austria.

1. Concepts of statistical grids

Statistical grids are defined as a region-wide spatial reference system for geodata. The grids are independent of administrative boundaries and therefore allow a more enhanced delineation of areas. Because of their large scale, spatial distributions can be more easily identified. To establish a temporal comparable system, a standardized and coordinate based grid is required. The grid is defined by the coordinate system, so different grid systems can be used for regional, nationwide or continent wide data.

There are various advantages when using grids. Technically, the grids are easy to handle with today’s geo information systems. It is possible to aggregate to higher grid levels very efficiently and automatic mapping can also be done easily. From an analytical point of view a comparison between different regions is easy to achieve because of the equal area of the grid cells. Also temporal comparison is possible because the grids are not changing over time like administrative boundaries. For cartographic visualization it is also important that, due to the constant area of the grid cells, absolute values can be shown without doubts.

As grids are usually much smaller than regional administrative units, the subject of data confidentiality is a big issue. The methods of providing the necessary confidentiality range from methods of suppressing values in grid cells which are sparsely populated to contamination of the values by adding or subtracting random values. Other techniques aggregate the values into classes or merge neighbouring cells to hierarchically higher grid cells (mixed grids). In each case information is lost, either the sum of the whole region is not correct or the regional depth is lost.
2. Grids in Austria

2.1. Statistical grid systems in Austria

The first steps towards assigning statistical data to grids at Statistics Austria go back to the 1980s (WONKA, 1983). However it was not until 2003, that the first dataset covering all of Austria was published. This was made possible through the establishment of a national register of buildings and dwellings providing the necessary geo-coordinates. New methods of spatial analysis were hence made possible and customer requests could satisfactorily be met.

A national grid was defined by Statistics Austria to suit Austrian geodata and the Austrian base map (ÖK500). This is based on the Lambert Conformal Projection (EPSG 31287), choosing as centre the meridian at 13°20' and 47°30' latitude. The gridnet was constructed parallel to both this meridian and the latitude. To suit various customer demands a series of 7 grid sizes was introduced from 125m x 125m up to 10km x 10km.

Figure 1: Austrian Grid Map based on the Lambert Conformal Conic Projection

Statistical grids are placed over the entire territory of Austria. However, since Austria is a mountainous country, a lot of grid cells are not inhabited. The smaller the grid size, the more obvious this becomes, see Table 1.

Table 1: Number of Grid cells in Austria

<table>
<thead>
<tr>
<th>Grid size</th>
<th>No of grid cells in AT</th>
<th>No of grid cells with buildings total</th>
<th>%</th>
<th>No of grid cells with population total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>125m</td>
<td>5.533.169</td>
<td>514.506</td>
<td>9,3</td>
<td>445.925</td>
<td>8,1</td>
</tr>
<tr>
<td>250m</td>
<td>1.374.615</td>
<td>267.345</td>
<td>19,4</td>
<td>233.240</td>
<td>17,0</td>
</tr>
<tr>
<td>500m</td>
<td>338.753</td>
<td>127.323</td>
<td>37,6</td>
<td>112.114</td>
<td>33,1</td>
</tr>
<tr>
<td>1000m</td>
<td>85.428</td>
<td>49.406</td>
<td>57,8</td>
<td>43.157</td>
<td>50,5</td>
</tr>
<tr>
<td>2500m</td>
<td>14.169</td>
<td>11.690</td>
<td>82,5</td>
<td>10.138</td>
<td>71,6</td>
</tr>
<tr>
<td>5000m</td>
<td>3.741</td>
<td>3.423</td>
<td>91,5</td>
<td>3.128</td>
<td>83,6</td>
</tr>
<tr>
<td>10000m</td>
<td>1.017</td>
<td>929</td>
<td>91,3</td>
<td>897</td>
<td>88,2</td>
</tr>
</tbody>
</table>
In addition to the grids based on the Lambert Conformal Projection, Statistics Austria provides a second grid system: a pan-European grid based on the ETRS-LAEA projection. This pan-European grid is driven by the data specification of the EU directive INSPIRE (http://inspire.jrc.ec.europa.eu/). A standardized European projection system is particularly advantageous for exchanging geodata in the EU since these data therefore no longer need to go through a complex conversion process. It also facilitates the small-scale presentation and evaluation of cross-border data in Europe. Austrian data on the basis of the ETRS-LAEA grid was first published in 2009.

Furthermore, in addition to the grid sizes which have to be provided for INSPIRE (100m, 1km, 10km, 100km) further grid sizes (250m, 500m, 2 000m, 5 000m) will be made available by Statistics Austria. The Federal Ministry of Agriculture, Forestry, Environment and Water Management will also provide data based on this grid system.

Both grid systems are currently being run in parallel, but due to confidentiality reasons a complete change to the European grid is in progress. In the future, all new data made available will be based on the ETRS-LAEA grid only.

Due to the two different projections the ETRS-LAEA grid is distorted in comparison to the MGI-Grid. Therefore it is not possible to use both systems combined in one calculation or analysis. The main challenge is now to convince existing customers (users of the MGI-Grid) to change to the ETRS-LAEA grid.

**Figure 2: Comparison between ETRS-LAEA and MGI-Lambert Grid when displayed in the MGI-Lambert projection**

![Comparison between ETRS-LAEA and MGI-Lambert Grid](image)

### 2.2. Data for statistical grids

The datasets collected at the censuses as well as those kept in the registers nowadays have a unique building-ID. This can be used as link to the register of buildings and dwellings from which geocoordinates are assigned to the data. Therefore, Statistics Austria is able to provide precise aggregated grid data.

In the earlier censuses a great variety of data on different thematic topics has been collected, which were aggregated to grids. Certain datasets were combined into thematic data packages, such as data on population, buildings and dwellings, employment and local units as well as heating and fuel.

- Population package: Population by marital status, age classes, gender, educational attainment, citizenship, country of birth, etc.
- Buildings and dwellings package: Buildings by building use, number of rooms, floor space, energy source for heating, etc.
Employment and local units package: Number of local units, by category of number of employees, number of employees, by economic sector and gender

Heating and fuel package: type of energy and heating facilities

The main packages are published based on the census. So far, smaller packages on recent data are available yearly, like population or buildings (starting 2010).

Datasets for the smallest grid sizes (100m/125m and 250m) have been greatly demanded, as these show the distribution of spatial phenomena best. Statistics Austria is also very flexibly and customer oriented. Customers can order other characteristics to be aggregated to grids. This can be e.g. more detailed age groups, a combination of characteristics or other available characteristics which are not included in the standard packages.

2.3. Confidentiality

For confidentiality reasons data in sparsely populated grid cells has to be suppressed. The general principles of disclosure control as has been applied for the 2001 Census are also applied for the grid data. The absolute numbers of the respective units (number of people with main residence, people with secondary residence, households, buildings, residential buildings, dwellings, local units and employed persons) are provided to the customer without any restriction, even for the smallest grid sizes. However, the characteristics for the units are suppressed if the grid cells are not sufficiently populated. The data protection threshold depends on the type of the unit. Characteristics relating to people with their main residence (e.g. age structure) are only passed on if there are at least 31 people with their main residence in a grid cell. If the data refers to buildings (e.g. building usage), there must be at least 4 buildings in the grid cell in order to disclose characteristics of these buildings.

Complying with this data protection rules only the more populated areas can be passed to customers. This can cause problems when rural areas are in the focus of the analysis. The customer will have to decide, whether a regional statistical analysis based on small grid cells suppresses too many features for data protection reasons, or whether it is better to use a larger grid size which less suppressed data. This concerns mainly the 250m grid cells. Figure 3 gives an example of how many grid cells a customer will get when using 250m or 500m grid sizes.

Figure 3: This map shows as an example for one rural area which grids are falling under the data protection threshold (31 inhabitants, 4 buildings or 4 dwellings).
The number of grid cells which are falling under the data protection threshold is relatively high, but the sum of the characteristics which the customer will receive is also quite high. The reason is that in more densely populated areas like cities very few features are lost due to the confidentiality. Table 2 shows the number of data which cannot be passed to customers because of data protection. Since the privacy threshold concerning the population data is significantly higher than concerning the building and dwelling characteristics, the proportion of the population characteristics without data is accordingly high.

**Table 2: Effect of confidentiality in Austria**

<table>
<thead>
<tr>
<th>Grid size</th>
<th>Percentage of units for deleted data which cannot be passed to customers because of data confidentiality (entire country)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>inhabitants</strong></td>
</tr>
<tr>
<td>250m</td>
<td>20.4%</td>
</tr>
<tr>
<td>500m</td>
<td>11.2%</td>
</tr>
<tr>
<td>1.000m</td>
<td>4.0%</td>
</tr>
<tr>
<td>2.500m</td>
<td>0.9%</td>
</tr>
<tr>
<td>5.000m</td>
<td>0.7%</td>
</tr>
<tr>
<td>10.000m</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

This table shows that data protection affects mostly 250m grids and regarding the datasets for inhabitants even the 500m grids.

### 3. Administrative units vs. regional statistical grid units

Traditional official statistics dealt with the descriptive analysis of administrative units such as provinces, municipalities or census districts. These administrative units have been evolved historically (provinces or municipalities) or have been created to fulfil a special purpose like equally populated areas (census districts). Administrative territorial divisions are political administrative units, and therefore important for certain tasks (e.g. in the field of regional planning). This applies particularly to the municipality data. But when it comes to represent the exact regional statistical distributions, geometric administrative territorial divisions are preferred. Similarly this applies for spatial analysis such as the planning of infrastructure.

#### 3.4. Territorial divisions in uneven populated areas

Particularly problematic are density value representations on the basis of municipalities in uneven populated areas like mountainous regions or regions with large areas of agriculture or forestry. This is illustrated by the example of an alpine area, the Zillertal in Tyrol. As the population distribution shown in Figure 4a, the valley areas are populated fairly evenly. This is also reflected when creating a map where the density values refer to equal and uniform territorial units, such as the grid cells (see Figure 4b). When using data based on the administrative municipalities the map shows different density values as result (see Figure 4c). If the border of the municipalities and the populated area are nearly identical, then the results are very high population densities. In this example, the municipalities Zell am Ziller and Fügen even reach densities of 100-500 inhabitants per km², which corresponds to the criteria of EUROSTAT as an area of high population density. However, if these municipalities contain large alpine areas, the density value is always levelled. Such density figures do not state anything useful about the actual population density, as the data are primarily influenced by the fact whether the share of the populated part of the municipality area is large or not. Furthermore, the exact boundary of the municipalities is pretending a precision that does not exist. In contrast, the population density values based on the coarse 2.5 km grid displays the characteristics of the valley much better.
3.5. Territorial divisions in regards of temporal comparison of the data

The two cartograms in Figure 5 show the boundaries of the city of Villach and its surroundings at two different points in time. Because of the municipality boundary changes between 1961 and 2001 a comparison over time is not possible. But not only the communities but also the statistical census districts are subject to boundary changes, so that temporal comparisons are hard to achieve. The changes can be aggregation of two or more regions or a subdivision of one region into more parts or other forms. In case of boundary changes, a comparison is only possible if also the old statistical datasets are adapted to the new administrative boundaries. However grid datasets are not affected by such changes because of their immutable reference.
3.6. Grid based data in Regional Planning

The preparation of maps and plans for the purpose of regional planning usually requires the combination of information from different sources. Thematic maps with regional statistical content perform an important task, in particular, when referring to statistical data on basis of grids. With the help of a regional statistical database which is built on grids, it is possible to create maps, to display the statistical distributions of objects (e.g. population distribution) or reflect structural characteristics (e.g. age structure of the population).

It should be mentioned that for many purposes it would be ideal to use data on the basis of individual buildings. For confidentiality reasons this data must be aggregated, but data aggregation on grid level is a good compromise. As mentioned above, grid data, as they are a stable, are very well suited for statistical comparisons over time.

3.6.1. State wide Regional Planning

A first, if only very rough overview of the distribution of the population of a country is possible to achieve when showing statistical maps on the basis of political districts. When creating a map on the basis of municipalities, the smallest local government units, it even provides a better regional view. The maps in Figure 6 and Figure 7 are showing what result is obtained when taking municipalities and 2.5 km grid cells. It is hard to identify the regional statistical differences when using administrative areas.

Figure 6: Population density in Austria by municipalities

*Population density by Municipalities*

![Population density in Austria by municipalities](image-url)
Figure 7: Population density in Austria by 2.5km grid

Population density 2.5km grid

<table>
<thead>
<tr>
<th>Population per km² by 2.5km grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 50</td>
</tr>
<tr>
<td>50 - &lt; 250</td>
</tr>
<tr>
<td>250 - &lt; 500</td>
</tr>
<tr>
<td>500 - &lt; 1000</td>
</tr>
<tr>
<td>1,000 -</td>
</tr>
</tbody>
</table>

It is possible to enlarge the grid size (e.g. using 5km grid cells) when needed for different planning tasks. This scaling of the grids cannot be done with administrative areas. Comparing the population density map on the basis of municipalities with 2.5 km grid cells, it is obvious that only the grid shows the densely populated alpine valleys.

3.6.2. Local Regional Planning

In Austria the local regional planning is the task of the municipalities. That means not only the zoning plan but also concepts for the local development have to be established. Large scale maps with statistical information can inform about the spatial potential of a specific region within a town or municipality. It is about the question of how the needs and problems within a region or community can be solved and can be covered such as the recreational, supply and disposal functions in a city. In order to answer these questions, regional statistical grids are suited best. On this level of planning, statistical data and maps are only meaningful if they relate to possible small statistical grid cells such as the 100m/125m size.

The following two Figures show two examples from the field of local regional planning.
Figure 8: Zoning Classification and Population (125m grid)

The purpose of this figure is to detect the total population (persons with primary and secondary residence) in the small garden areas, which lie within the flood drainage area of the Danube. This map illustrates the conflict between natural hazards and human settlements.

Figure 9: Catchment areas of public transport

This map shows how many residents live within or outside a radius from a bus station. It is possible to depict the areas with good or poor accessibility to the public transport system. At the same time one can see the bus frequency at peak times.
3.7. Cross Border analysis

How well a cross-border statistical task can be solved, depends not only on whether the statistical indicators are comparable, but also on the way in which statistical data is regionalized. So far, the definition of regional statistical indicators generally based on more or less spatially low-resolution parameters on administrative units. Planning decisions at the regional level, however, require an approach as flexible as possible, based on geographical grid units.

Regional statistical studies observe far too little, that the borders of administrative territorial divisions such as the NUTS and LAU 2 units (in Austria, these are the municipalities) were created under the terms of administration. The geographic boundaries often have nothing to do with the investigated statistical facts. As the administrative territorial units are given from the outset, it is not possible to know if the values are affected by the delineation of the boundary. For example, the average size of local units (municipalities) in France is about 15 km² with an average population of 1,500; in Denmark, however it is 156km² with a population of about 19,000.

When using the term cross-border as the border between a city and surrounding municipalities, the same problems occur. Within a city the administrative regions are very small, even down to city blocks. However, the neighbouring municipalities may not be divided into city blocks, so that the comparison is hard to achieve.

In order to allow meaningful cross-border analysis, regularly divisible territorial units with same size are needed. These conditions are only fulfilled by geometric entities such as statistical grids. (See as example Figure 10)

Figure 10: Population density in Austria and Slovenia

3.8. Combination of different data sources

Using a standardized grid for all kind of data enables the user to combine different datasets for small area statistics. In addition to population or building data offered by statistical institutes on the basis of the census or registers, other sources from other data providers can be integrated. For example this can be agricultural data like size of farms, forested land or the number of cattle or the digital soil map which consists of multiple map layers. These layers contain data of soil types, water conditions, permeability and soil depth. Other data contain damage by hail, frost or drought or climate data like precipitation, humidity or sunshine duration. In particular information about the wind and sunshine duration are important for the development of energy concepts.
This map shows the extent of reduction of high-quality soil due to settlement. The greater the number of soil climate, the more is the crop yield (max. value = 100).

It is also possible to combine data with a digital elevation model (DEM). It is not just the link of the elevation model with the statistical data on the basis of regional statistical grid units, but also the visualization of a digital elevation model, the slope of the terrain and exposure. This can be useful to distinct preferred farmland (e.g. exposition to south) and disadvantaged areas (e.g. elevation above 2,000m).

When using line based data, and performing a distance analysis using grids, it is possible to achieve very exact values for the specific task. This figure shows the population within 150m of the railway tracks (in red). The results with the grid data are by far better than utilizing census districts for this sort of analysis.
3.9. Structuring functional regions

A different utilization of statistical grids is to create functional regions. For example, this can be a regional structure type of rural areas, urban regions or permanent habitable area. For these analyses grid datasets from different sources have to be combined. This can be population data, land cover, tourism or agricultural data. Data which is not available on grids can be assigned to grid cells with different disaggregation methods.

One example is the delineation of urban regions on basis of grids. Since 1971 Statistics Austria is delineating the urban regions of Austria. As for the previous years (1971, 1981, 1991) the model resulted in a two zones approach - a core zone with a high population density and an outer zone with a high percentage of commuters into the core zone. However, the method of delineation has changed over time. While the method utilized in 1991 used the settlement areas as spatial reference basis, the new method (2001) is based on population and employment figures by 500m grids. The delineation of the outer zones, which was done as before on the bases of LAU2, now also considers increasingly the topographic situation as well as settlement structures.

Figure 13: Delineation Town Regions 1991 and 2001
4. Conclusion

It was the aim of this paper to provide an overview of the main uses of statistical grids for spatial analysis. This overview is certainly not exhaustive, the examples given are only of illustrative character. Nevertheless, it is hoped that the power of grids based data and grids-based analysis could be illustrated to a large degree.

From the point of official statistics, grids are viewed as the spatial reference system that can be used as the smallest statistical area unit for which data may be provided to the customer. Thus, the main advantage of a grid system lies in the combination of both facts: the analytical power of grids and the possibility to provide statistical data.

However, it is not yet common state of the art that geo-referenced data are available in the official statistics data systems in many countries and that the various data sets can appropriately be combined. Austria is in the fortunate position to establish grids-based data since the development of the register of buildings and dwellings a couple of years ago. In this register each building has a point-based geo-reference. All data on residents are linked to the building in which they live as well as all local units are linked to the building where they perform their economic activity. Therefore, data on buildings and dwellings, data on population as well as characteristics of local units can be entered into a grid system.

References


Statistics Austria: MGI-Lambert grid: 

Statistics Austria: ETRS-LAEA grid for Austria, 

Statistics Finland: Production and dissemination of grid data since the 1970 Census in Finland, Conference of European Statisticians, Fifty-eighth plenary session, Paris, 2010


