LINKING REMOTE SENSING AND DEMOGRAPHIC ANALYSIS IN URBANISED AREAS

Klaus Steinnocher[1], Jürgen Weichselbaum[2] and Mario Köstl[1]

1. ARC systems research, Vienna, Austria, klaus.steinnocher@arcs.ac.at
2. GeoVille Information Systems, Innsbruck, Austria; weichselbaum@geoville.com

ABSTRACT

During the last decades urban areas have experienced an enormous growth in terms of human population and physical size. Applying remote sensing techniques has become a standard approach for monitoring the physical growth of urban areas. However, when it comes to social studies statistical data are used for analysis in a first line. Only a few studies have been undertaken that try to combine these complementary data sources. In this paper we will show how land cover information from remote sensing can be combined with population data in order to derive refined information products.

Demographic data is usually derived from census and represented in administrative units such as districts or municipalities. Spatial analysis on a finer level is not possible due to this restriction. Introducing remote sensing can partially overcome this problem by indicating where people actually live within the administrative units. The total population of one unit can then be allocated to the built-up areas within the unit leading to a spatial refinement of the statistical data.

The paper will present a number of examples for spatial disaggregation of population data based on remote sensing derived information on urbanised areas – ranging from binary settlement masks to housing densities – for regional and European applications. A special focus will be set on to the assessment of accuracies of the presented method. Detailed demographic data available for selected regions will be used as reference for the spatial disaggregation results. These comparisons show that the approach is reliable and the quality of the final information products can be controlled via the level of detail of the remote sensing analysis.

INTRODUCTION

During the last decades urban areas have experienced an enormous growth in terms of human population and physical size. Applying remote sensing techniques has become a standard approach for monitoring physical parameters and growth of urban areas (i, ii, iii, iv). However, when it comes to social studies statistical data are used for analysis in a first line. Only a few studies have been undertaken that try to combine these complementary data sources.

The benefits that could be derived from the combination of remote sensing and social science were discussed in (v) providing a number of potential applications. The derivation of socio-economic attributes from remote sensing in urban and suburban areas is reviewed in (vi) including the estimation of population and quality of life indicators. Regression models to estimate population from Landsat TM imagery in Australia were applied in (vii). The correlation between census dwelling data and residential densities derived from remote sensing were analysed in (viii). The relationship between population density and vegetation cover in urban areas was investigated in (ix) in order to improve differentiation of urban land use classes.

The usefulness of remote sensing data for disaggregation of areal census data is discussed in (viii). In particular the relation to the modifiable areal unit problem (x) is emphasized. The research presented here provides two applications of spatial disaggregation based on remote sensing and population data of different scale levels. The disaggregation results are validated against detailed reference data from the census.

METHODOLOGY

The core method applied in this study is spatial disaggregration. It is based on the assumption that data, provided globally for an entire region, can be distributed within the region by means of local parameters. The spatial distribution is normally performed by a weighted sum. A clear dependency between the global and the local parameter is a prerequisite for this approach.

We will use population data, usually available from the census in administrative units, and spatial information on housing, derived from remote sensing. In terms of spatial disaggregation the global parameter is the total population of the region while the local parameter is the housing density derived from EO. Applying housing density as a proxy for population density allows estimating the local population distribution. This approach can be formalised as follows:

$$Pdens = k * Hdens \quad (1)$$

$$Pop = \sum_i A_i * k * Hdens_i \quad (2)$$

where *Pdens* and *Hdens* are the population and housing density respectively, *Pop* is the total population of the region and A_i corresponds to the area of the housing density *i*. The factor *k*, representing the relationship between population and housing density, can be derived by solving equation (2). The local population density is then calculated from equation (1). The following assumptions were made when applying this approach:

- the population density is proportional to housing density,
- no population occurs outside housing areas, and
- dependency between population and housing density is constant within a region.

A straight forward application for the described method is the combination of a binary settlement mask (built-up versus non built-up) derived from remote sensing and population data on municipality level for analysis on a regional level. However there might be regions where population data is only available as a total for the entire region. Spatial disaggregation of population should then be based on housing densities rather than on built-up / non built-up information. In this case remote sensing provides crucial information for deriving detailed data on the spatial distribution of population within a region.

APPLICATION

European case study

The European application was performed on a transnational test region located in central Europe including the Czech Republic, Austria, Slovenia, and parts of Germany, Slovakia, Hungary and Italy (Figure 1). The exercise was based on CORINE Land Cover data and demographic data provided by Eurostat. Both data sources provide harmonised data sets over large parts of Europe, thus allowing an extension of the application to the entire continent (xi).

CORINE Land Cover (CLC) is a compilation of national land cover inventories which are integrated into a seamless land cover map of Europe. The resulting European database is based on a standard methodology and nomenclature (xii). The scale of the cartographic representation is 1:100.000 with a minimum mapping unit of 25ha, i.e. single land cover objects smaller than 25ha are not represented in the data set. At European level the data base is available in a 100m grid format, representing the first two levels of the three level nomenclature. In spring 2005 an update (CLC 2000) of the first data base (CLC 1990) was completed by the participating countries.

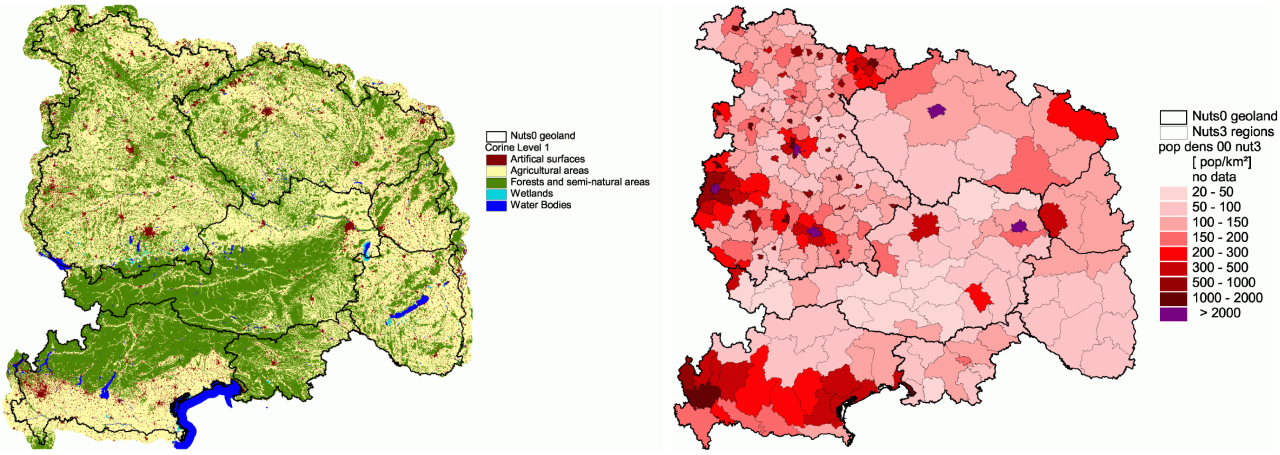

Figure 1: European test region. Left: CORINE land cover level 1 (2000), right: Population density per NUTS 3 area (2001)

The statistical data at European level are derived from the REGIO database, Eurostat's harmonised regional statistical database. It contains 14 different collections, including agricultural statistics, demographic statistics, economic accounts, education, environment statistics, community labour force survey, migration statistics, science and technology, structural business statistics, health statistics, tourism statistics, transport and energy statistics, unemployment, and environment statistics (xiii). These data are available at NUTS 3 level which is used as spatial reference on the European scale.

For the presented study the land cover class *urban fabric* from CLC 2000 was used that represents mainly residential areas including buildings, gardens, streets and related surfaces. No differentiation on housing density was made. From the REGIO data base population data were extracted and mapped onto the corresponding NUTS 3 areas. Figure 1 shows the CLC 2000 map and the population density 2001 per NUTS 3 area for the European test region. While the CLC data set gives a kind of “natural” impression of the test region, the NUTS 3 structure of the statistical data emphasizes its administrative nature, including a tremendous variation in size of the NUTS 3 areas (e.g. comparing Germany with the Czech Republic).

Spatial disaggregation of the population data was performed by assigning the population of each NUTS 3 area to the *urban fabric* areas within the corresponding NUTS 3 area. For reasons of presentation the results were intersected by a regular 3x3km grid allowing to calculate population density per grid cell. Figure 2 shows the *urban fabric* and the disaggregated population in the grid representation.

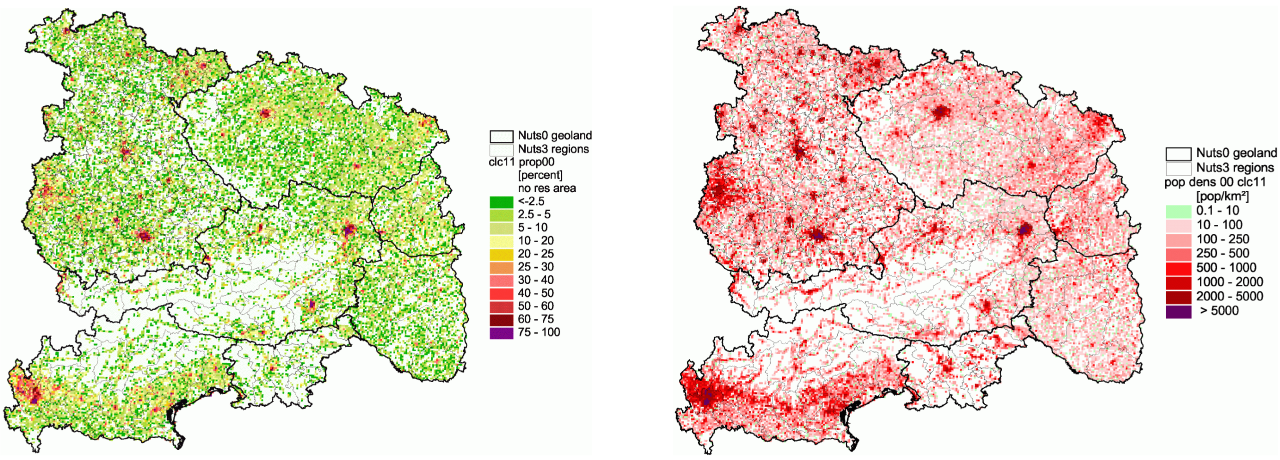

Figure 2: European test region. Left: urban fabric (2000), right: disaggregated population (2001)

On a European scale the grid based approach can help to tackle the modifiable area unit problem of the NUTS 3 areas. While harmonised statistical information for Europe is only available for NUTS 3 areas, the size of these areas varies tremendously. By disaggregating the population of a NUTS 3 area to the urbanised areas the polygonal representation can be replaced by a regular grid and a spatially balanced representation can be generated. This is of particular relevance when it comes to trans-boundary analyses.

Regional case study

The second application was performed for the most western state of Austria, called Vorarlberg. It covers a heterogeneous landscape including the densely populated Rhine valley in the west and less populated mountainous areas in the centre and east (Figure 5). As the settlement structure shows strong variations in the region this test site represents a challenge for the spatial disaggregation approach.

Residential areas were derived from multispectral SPOT 5 data in a mapping scale of 1:25.000 with a minimum mapping unit of 0,25ha. Three density classes were mapped within these residential areas: low density representing 0-50% artificial surfaces, medium density representing 50-75% artificial surfaces, and high density representing 80-100% artificial surfaces. Population data were provided by the Austrian census from 2001. For the spatial disaggregation only the total population for the state of Vorarlberg was used.

According to equation (2) the factor k is estimated from the total population and the areal proportions of the three density classes. The final population densities are then derived from equation (1). For the representation of the results a 250m regular grid is used. Figure 5 shows the result for the state of Vorarlberg.

RESULTS AND QUALITY ASSESSMENT

For the quality assessment of the case studies the estimated population from the disaggregation is compared to independent reference data from the Austrian statistical office (Statistik Austria). They present population from the census 2001 in a regular grid covering the entire area of Austria. This product is based on address-geocoded socio-economic data – including population, places of work, number of dwellings and number of buildings – aggregated to raster cells of 250x250m for reasons of data privacy (xiv).

Due to the level of detail on one side and the large coverage on the other side the population grid is an ideal reference data set both for the European and for the regional application. For comparison on the European scale the population grid was aggregated to the 3x3km grid in order to match the representation of the disaggregated population data. For the assessment of the regional application the disaggregation results were aggregated to the 250m population grid, thus allowing a direct comparison with the reference data set.

European case study

Figure 3 shows the disaggregation of the NUTS 3 population for Austria (a subset of Figure 2 right), Figure 4 presents the population grid derived from the census. A visual comparison of the two data sets allows analysing their major differences. While in the disaggregation result only about 50% of the grid cells are “populated”, in the reference grid almost 80% of the grid cells are occupied. This difference results from the generalisation effects of the CORINE land cover map. Small villages and dispersed settlements, that are found typically in alpine and rural environments, are not mapped due to the minimum mapping unit of 25ha. If no *urban fabric* exists in a grid cell, no population will be assigned to it in the disaggregation process and the grid cell will stay empty.

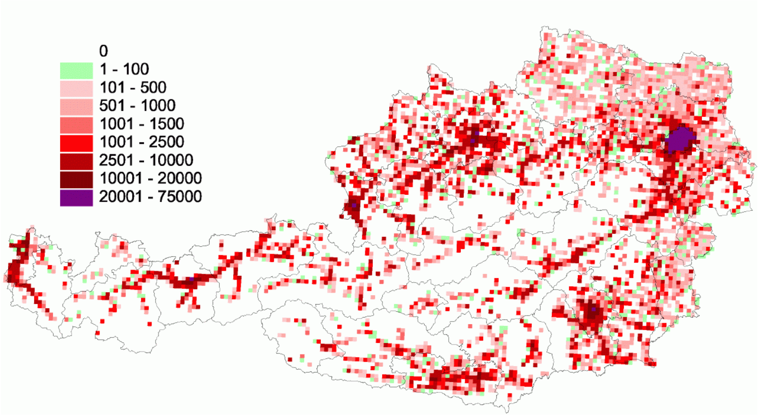

Figure 3: Population Austria 2001: disaggregation result of NUTS 3 population data (Eurostat)

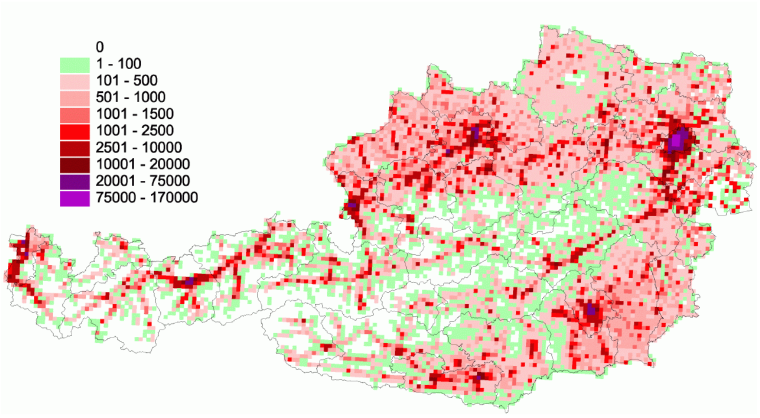

Figure 4: Population Austria 2001: reference grid (Statistik Austria)

On the other hand most of the empty grid cells show only low population in the reference grid. Intersecting the occupied grid cells of the disaggregation result with the reference grid and summing up the population of the respective grid cells adds up to about 90% of the total population; i.e. the occupied grid cells of the disaggregation results, although covering only about 60% of the actually populated area, include about 90% of the population. Since the total population of both data sets is the same, one can conclude that an overestimation of population of about 10% can be expected in the occupied cells of the disaggregation result.

A second systematic difference can be found when comparing urban centers, such as Vienna in the north east of Austria. While in the disaggregation result the population is equally distributed over the entire city, the reference grid shows significant differences between the center and the outer districts of Vienna. This results from the fact that no housing densities were considered in the disaggregation process. Thus, the population in urban centers is systematically underestimated while it will be overestimated in the surroundings.

In addition to the evaluation on grid cell level population data were aggregated to the 121 districts of Austria and comparisons were performed on this level. Table 1 shows the accuracies for the comparison of all districts and of a subset of 91 districts, where the urban districts were excluded. Again we face the problem of the urban centres, where the population is significantly underestimated, resulting in a clearly lower accuracy (21,4%) compared to the non urban districts (13,8%).

Table 1: Accuracies on district level for spatial disaggregation of population on European scale

	all districts	urban districts excluded
Relative average error	21,4%	13,8%
Absolute average error	14.505	8.278

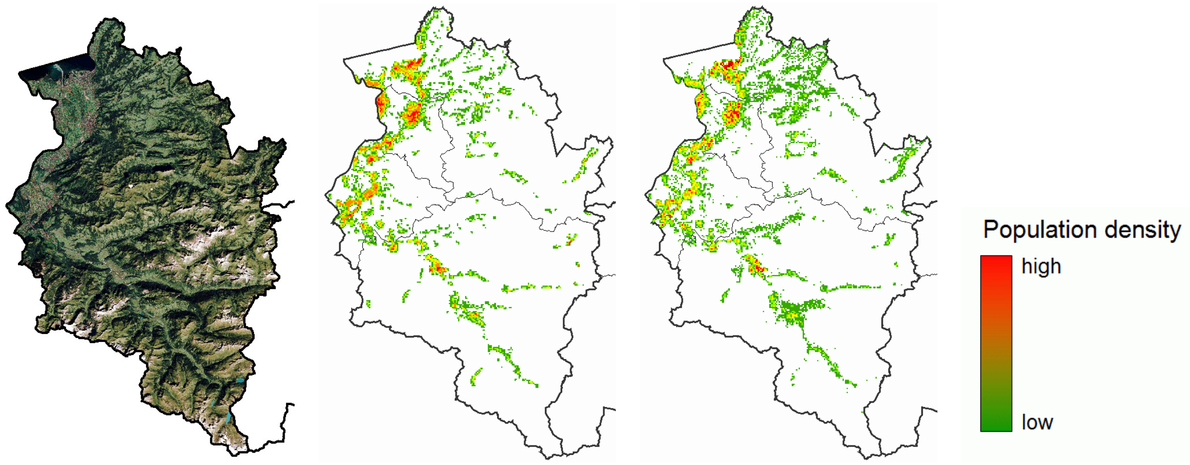

Figure 5: Regional test site Vorarlberg. Left: satellite image, centre: spatial disaggregation of population, right: population reference grid

Regional case study

Figure 5 shows the disaggregation results and the reference grid for Vorarlberg. Both maps present similar patterns of population distribution with some apparent differences. The major deviation is found in the eastern part of the region that is dominated by mountains. This area has a low population density with dispersed settlements that are not entirely covered in the land use map. Therefore the population is underestimated in the disaggregation results. The western part of the region shows a higher population density due to a number of larger settlements located in the Rhine Valley. Here the major differences occur in the town centres where the disaggregation slightly underestimates the population density. This results from the fact that – although different housing densities were included – different building heights in high density areas were not taken into consideration.

In addition to the evaluation on grid cell level, population data were aggregated to the 96 municipalities and 4 districts to assess the accuracy on these levels. Furthermore the disaggregation was also performed based on the CLC class *urban fabric* (as it was used on the European level) and compared to the reference data. Finally a simple population distribution proportional to district areas was calculated to demonstrate the improvement gained from the disaggregation approach.

Table 2 shows the accuracies on district and municipality level for the disaggregation based on housing densities, the disaggregation based on the CLC class *urban fabric*, and the proportional distribution. It is obvious that the spatial disaggregation based on housing densities yields the best results, but also the CLC *urban fabric* based approach is acceptable, considering the different scale of the data set. The proportional distribution – assuming a linear distribution over the entire region – leads to arbitrary results. The transition from district to municipality level leads to higher relative errors – due to the higher level of detail – but at the same time to reduced absolute errors.

Table 2: Accuracies on district and municipality level for spatial disaggregation of population in Vorarlberg

Average error	housing density	urban fabric (CLC)	proportional
District level – relative error	8,4%	12,7%	94,7%
Municipality level – relative error	12,8%	26,9%	113,3%
District level – absolute error	5.904	8.874	66.325
Municipality level – absolute error	467	981	4.136

CONCLUSIONS

The method of spatial disaggregration was applied to population data on a European and regional level. Starting with one global population number for a region the local distribution of population was estimated by means of residential areas and housing densities respectively derived from earth observation. The results were compared to local demographic data from the census, i.e. they were validated by means of an independent information source. This exercise was performed for a part of central Europe and the province of Vorarlberg in Austria.

Quality assessment of the disaggregation results was based on demographic data from the Austrian census. These data are available on a 250m regular grid, on municipality and on district level. For evaluation of the European case study the reference grid and district data of the entire area of Austria were applied. The regional case study was evaluated by means of the reference grid, municipality and district data of the state of Vorarlberg.

The reference units for Austria were 121 districts, for Vorarlberg 4 districts and 96 municipalities. The accuracies on district level excluding urban districts for the European disaggregation (13,8%) match those for the Vorarlberg test site applying the CLC *urban fabric* as basis (12,7%). This is due to the fact that Vorarlberg has no big urban centres; therefore the effect of underestimating the population in such centres is negligible on that scale. The remaining difference results from the different grid sizes used for the comparison – 3km for the European and 250m for the regional level.

Introducing different levels of detail in the land cover data – as done in Vorarlberg by applying one residential class as well as different housing densities – shows the benefit that can be gained from density analysis. While the application of a single residential class results in an error of 26,9% on the municipality level and 12,7% on the district level, the introduction of housing densities improves the results to error values of 12,8% and 8,4% respectively. In addition to the introduction of housing densities also the change of scale – minimum mapping unit of 25ha for CLC versus 0,25ha for the regional application – has an impact on the accuracy.

It is important to note that these accuracies are to be seen in contrast to the accuracies of the proportional distribution of the global population values. The errors calculated for this distribution add up to more than 100% for Vorarlberg. Thus the spatial disaggregration method results in information gain between three and ten times.

In addition to the statistical accuracy analysis described above a spatial analysis was performed by means of a grid based population distribution. This approach allowed analysing the difference in spatial patterns and locating systematic errors of the disaggregration method. The results of this analysis showed that there is a tendency of underestimating the population in urban centres and in dispersed settlements. As a consequence the population of the remaining settlement types – ranging from villages to medium sized cities – is slightly overestimated.

The underestimation in urban centres can be corrected partly by introducing housing densities, but on a regional level one requires additional information on building heights. Mapping of dispersed settlements is depending on the minimum mapping unit and thus on the resolution of the satellite images used. If the resolution is too coarse, single houses will not be mapped and the population will be underestimated. This effect can be seen in the alpine parts of both the European and regional evaluation sites. However, the error that results from this effect is comparably small as the absolute population of dispersed settlements is usually very small.

The study has shown that the disaggregation approach is reliable and the quality of the final information products can be controlled via the level of detail of the remote sensing analysis. Further research should concentrate on the problem of the urban centres in order to improve the overall quality of the disaggregation results.

ACKNOWLEDGEMENTS

The study was performed within the Observatory Spatial Planning of the Integrated Project geoland. The project geoland is funded by the European Commission in the frame of the GMES (Global Monitoring for Environment and Security) initiative of the 6th framework programme (Contract No SIP3-CT-2003-502871) and is co-ordinated by Infoterra GmbH, Friedrichshafen, and Medias France, Toulouse.

REFERENCES

i Donnay J P, M Barnsley & P Longley (Eds.), 2001. Remote Sensing and Urban Analysis (London: Taylor & Francis), 268 pp.

ii Lavalle C, L Demicheli, M Turchini, P Casals-Carrasco & M Niederhuber, 2001. Monitoring mega-cities: the MURBANDY/MOLAND approach. Development in Practice, 11 (2-3): 350-357.

iii Carlson T (Ed.) 2003. Urban remote sensing. Remote Sensing of Environment, 86(3), 273-432.

iv Yang X (Ed), 2003. Remote Sensing and GIS for Urban Analysis. Photogrammetric Engeneering and Remote Sensing, 69(9): 937-1036.

v D Livermann, E F Moran, R R Rindfuss & P C Stern (Eds.), 1998. People and Pixels – Linking Remote Sensing and Social Science (National Academy Press, Washington D.C.), 244 pp.

vi Jensen J R & D C Cowen, 1999. Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes. Photogrammetric Engeneering and Remote Sensing, 65(5): 611-622.

vii Harvey J T, 2002. Estimating census district populations form satellite imagery: some approaches and limitations. Int. J. Remote Sensing, 23(10): 2071-2095.

viii Chen K, 2002. An approach to linking remotely sensed data and areal census data. Int. J. Remote Sensing, 23(1): 37-48.

ix Pozzi F & C Small, 2005. Analysis of Urban Land Cover and Population Density in the United States. Photogrammetric Engeneering and Remote Sensing, 71(6): 719-726.

x Wrigley N, T Holt, D Steel & M Tranmer, 1996. Analysing, modelling, and resolving the ecological fallacy. In: Spatial Analysis: Modelling in a GIS Environment, edited by P Longley & M Batty (Geoinformation International, Cambridge), 23-40.

xi Steinnocher K, U Gangkofner, C Hoffmann, M Köstl, F Petrini-Monteferri & T Tötzer, 2005. Spatial planning indicators – the geoland approach. In: 10th International Conference on Information & Communication Technologies (ICT) in urban Planning and Spatial Development and Impacts of ICT on Physical Space (CORP2005), edited by M Schrenk, 185-191.

xii European Commission, 1994. CORINE land Cover Technical Guide (EUR 12585 EN, OPOCE Luxembourg), 136 pp.

xiii Eurostat, 2004. European Regional Statistics Reference Guide: 2004 Edition (KS-BD-04-001-EN-N; OPOCE Luxembourg), 189 pp.

xiv Kaminger I & E Wonka E, 2004. Von einer Österreichgliederung nach Gemeinden zu Planquadraten: Statistik Austria erweitert sein regionalstatistisches Angebot. In: 9th International Conference on Information & Communication Technologies (ICT) in urban Planning and Spatial Development and Impacts of ICT on Physical Space (CORP2004), edited by M Schrenk, pp. 549-554.