Spatial Accuracy Evaluation of Population Density Grid Disaggregations with Corine Landcover

Johannes Scholz, Michael Andorfer and Manfred Mittlboeck

Abstract The article elaborates on the spatial disaggregation approach of the 1 km population density grid created by the European Forum for Geostatistics in a defined study area where accurate population reference data are available. The chapter presents an approach to disaggregate the population grid to target resolution of 100 and 500 m respectively and describes the evaluation methodology. The resulting population grids are evaluated with respect to the reference population dataset of the Austrian Bureau of Statistics. In addition, the results are evaluated regarding their correlation to the reference or a random population dataset. The results indicate that there is evidence that the disaggregated population grid with 500 m resolution is more accurate than the 100 m population grid. In addition, the 100 m disaggregated population raster shows more correlation with the random population grid. Furthermore, the chapter shows that densely populated zones are estimated with higher accuracy than medium and sparsely populated areas.

1 Introduction

The European Union provides population data, which is comprised of the national census data sets, of which a few are available to the public. The available population dataset created by the GEOSTAT 1A project of the European Forum for Geostatistics (EFGS) is a 1 km population grid that is hosted by EUROSTAT. Due to the fact that the GEOSTAT 1A population grid is very coarse for detailed simulation activities, a downscaling or disaggregation process is necessary in order to obtain population density data on a finer granularity level.

J. Scholz (✉) · M. Andorfer · M. Mittlboeck
Studio iSPACE, Research Studios Austria, Schillerstrasse 25, Salzburg, Austria
e-mail: johannes.scholz@researchstudio.at

D. Vandenbroucke et al. (eds.), *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography,
DOI: 10.1007/978-3-319-00615-4_15,

Downscaling is a well-known term in environmental studies, which describes the process of generating fine granular data from coarse base data (Bierkens et al. 2000; Reibel and Agrawal 2007). In order to generate a dataset of finer spatial granularity, auxiliary data are necessary that provide additional information on the spatial phenomena to be disaggregated. In this chapter, Corine Landcover data are employed to detect population densities as complementary source to the GEOSTAT 1A population grid. As 1 km population grid cells may contain parts with varying population density—e.g. dense urban zones or urban areas with parks having no population. By population grid cells which span over different population "density zones" the representation of those zones are blurred.

Corine Landcover data are chosen as auxiliary data due to their availability over Europe and the consistent semantics over Europe. This is of interest for the research project that forms the organizational frame for this chapter and research work. The research project aims at modeling and simulating socio-economical phenomena on a detailed level. The results of the study should be applicable in all member states of the European Union. Due to the fact that fine granular population data are not available for all European countries in a consistent manner, the population raster with 1 km resolution is employed as harmonized population dataset.

Application fields of fine granular population data are found in the assessment of natural disasters like floods or hailstorms (Tralli et al. 2005; Chen et al. 2004). Thiecken et al. (2006) used disaggregated population data in order to evaluate the population affected by the flooding in Germany of 1999 and 2002. Such population grids help to estimate the impact of noise on people living around airports (Vinkx and Visée 2008).

In literature downscaling methods have been discussed in depth. A number of publications elaborate on the disaggregation of data from a zonal system—i.e. districts, communes—to smaller zones. The process is supported by ancillary data, usually land cover data. These approaches assign a population density to each land cover type in a certain zone of the study area. A number of methods belonging to the zonal family use a regression model to improve and obtain the population density for each land cover class (Yuan et al. 1997; Briggs et al. 2007). Mennis (2009) replaces the regression with average densities determined from a sample of zones having a single land cover type. Gallego (2010) evaluates the Expectation-Maximum likelihood algorithm (Flowerdew et al. 1991) that is able to substitute the regression step.

Eicher and Brewer (2001) report three different downscaling approaches:

- Binary method (Langford and Unwin 1994): This method assigns the population to a single land cover class.
- Three-class method: This method allocates some population density to forestry and agricultural classes.
- Limiting variable method: This approach starts with a homogeneous population density for all land cover classes per administrative unit. The population density is then refined through thresholds applied to each land cover class and a redistribution of the "leftover" population to other land cover classes.

Eicher and Brewer (2001) conclude in their chapter, that the limiting variable method performs best. Other publications related to the group of limit-based methods described by Eicher and Brewer (2001) are e.g. Reibel and Bufalino (2005) or Mrozinsky and Cromley (1999).

Gallego (2010) describes four methods to generate dasymetric population density grids based on population data on a commune level and Corine Landcover. This chapter evaluates the disaggregation of four different methods, namely CLC-iterative, CLC-Lucas, CLC-Lucas logit and the EM algorithm. The results of the disaggregation processes are compared with the GEOSTAT 1A population grid. In this chapter the CLC-Lucas logit method performed best, but according to Gallego (2010) are the differences between the approaches moderate.

An approach to disaggregate population data to a resolution of 100 m is presented in Gallego et al. (2011). The chapter evaluates six methods to disaggregate population data based on the commune population data and Corine Landcover. Other ancillary data sources are EUROSTAT point survey and the land use/cover frame survey. The disaggregated data were evaluated using parts of the GEOSTAT 1A population grid, and best results were obtained with a modified version of the limiting variable method.

The aim of this chapter is to evaluate a spatial disaggregation approach of the GEOSTAT 1A population density grid in a defined study area with accurate reference data. The article elaborates on the disaggregation of the GEOSTAT 1A population grid having 1 km resolution with two distinct target granularities—100 and 500 m. An evaluation of the correlation of the resulting population grids with reference data and a weighted random population grid, results in the "similarity" of the disaggregated, reference and random population grid.

This chapter is organized as follows. In Sect. 2 the study area and data used in this chapter are explained. The disaggregation methodology and evaluation approach is described in Sect. 3. The results are given in Sect. 4. Section 5 comments on the results obtained. A conclusion and outlook is given in Sect. 6.

2 Study Area and Data

The study area is located in the northern part of the province of Salzburg and some parts of Upper Austria, Austria. Special focus is set on the northern part of the province of Salzburg and the northern parts of the City of Salzburg and surrounding areas. Hence, the study area covers densely populated as well as rural areas. The data used in this chapter are originating from the Austrian Bureau of Statistics, EUROSTAT and European Environmental Agency. The following chapter elaborates on the study area followed by a description of the spatial data used in the experiment.

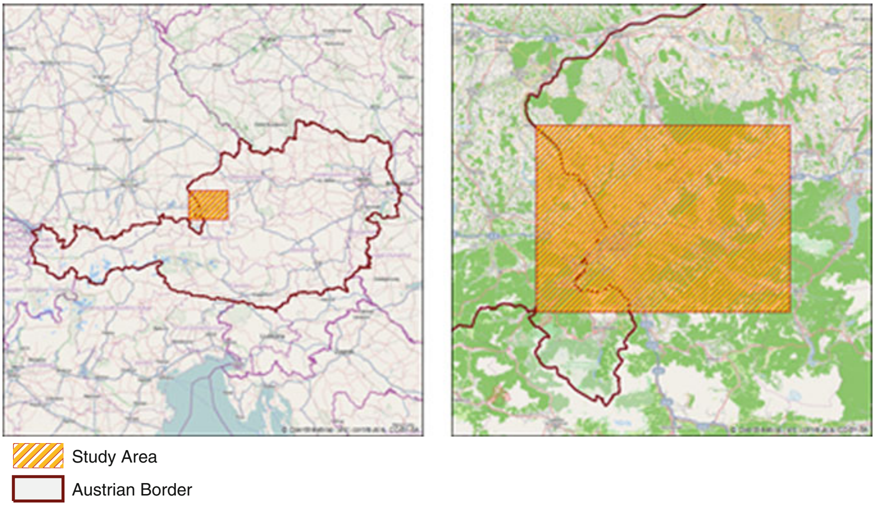

Fig. 1 Location of the study area in Austria (*left*), and the detailed map of the study are (*right*)

2.1 Study Area: Northern part of Salzburg

The study of this chapter is conducted in the northern part of the Province of Salzburg and the western parts of Upper Austria, Austria. In order to have areas with varying population density represented in the study, the area of interest comprises of urban and rural areas.

Figure 1 shows the location of the study area in this chapter. In addition, the base Corine Landcover (CLC 2006) classes are depicted in order to underpin the varying population density in the different land cover classes. Important for this chapter are densely populated areas of the City of Salzburg, that are covered by urban fabric and the outskirts of the city that show urban sprawl. To the north of the city of Salzburg large areas dominated by agriculture and forestry can be found, that are only sparsely populated.

2.1.1 Used Source Datasets

In order to conduct an evaluation of the quality of downscaling methods, several datasets are necessary that originate from official statistical sources. In this chapter data of the Austrian Bureau of Statistics are used for ground truth information, and population datasets of the EUROSTAT provide one ingredient for spatial downscaling. In addition, the European Environmental Agency provides data on the Corine Landcover Classification.

The Austrian Bureau of Statistics collects the population census and provides aggregated census data—i.e. population numbers in a regular grid—with a resolution of 100 and 500 m compatible to the European Reference raster in Lambert azimuthal equal area projection (ETRSA89-LAEA). Hence, the spatial resolutions of the population rasters of the Austrian Bureau of Statistics fit to each other and to the European Reference raster. The 100 m population raster dataset serves as "ground truth" for validating of the disaggregation results, due to the underlying accurate census data. The reference year of the Austrian census data is 2010.

The population raster to be disaggregated is provided by EUROSTAT, and was created by the GEOSTAT 1A project of the EFGS. The resolution of the population density grid is 1 km and the population data are based on the reference year 2006. The data sources used to generate the GEOSTAT 1A population raster are listed in EFGS (2012). Hence, they are not mentioned in this chapter due to the minor relevance for this work.

In order to spatially disaggregate population grids, ancillary data are necessary that provide additional information on where population lives. On a European level Corine Land Cover 2006 (CLC) is an appropriate dataset that maps the land cover in a 100 m resolution grid. CLC is produced by applying common interpretation rules to SPOT-4 and IRS P6 satellite images (EEA-ETC/TE 2002). The results of the CLC are land cover datasets representing the land cover in a 1 ha resolution raster with a minimum mapping unit of 25 ha. The CLC nomenclature consists of 44 classes, which are hierarchically organized. If a polygon cannot be clearly assigned to one dominant land cover type the area is denoted as "heterogeneous". Gallego (2010) reports that smaller urban areas are not represented due to their small patch size smaller than the minimum mapping unit of 25 ha. The CLC data for the study area is given in Fig. 2.

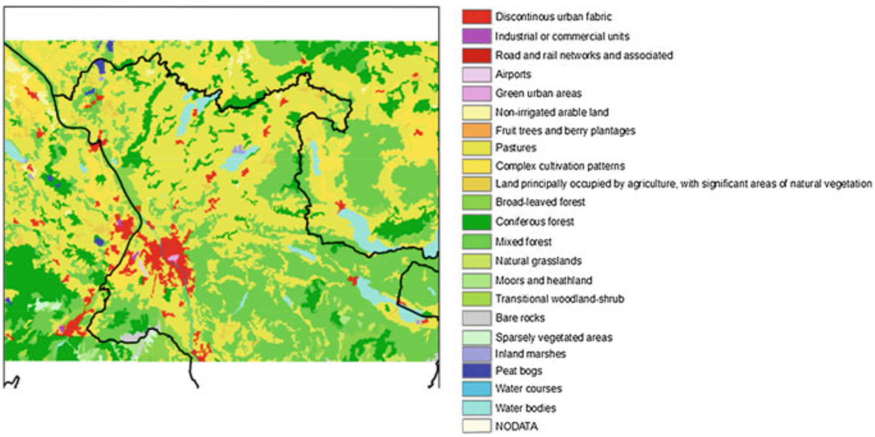

Fig. 2 Corine Landcover data of the study area in the northern part of Salzburg. The biggest continuous urban fabric agglomeration in the middle of the map is the city of Salzburg

3 Spatial Disaggregation of Geostat 1A Population Raster and Evaluation of Disaggregated Population Raster

This section describes the spatial disaggregation method used in this chapter to generate a 100 m population raster from the original 1 km Geostat 1A European population grid. The method used in this chapter is strongly related to the approach presented by Gallego and Peedell (2001) and Gallego (2010). In addition, this publication evaluates the results of the disaggregation of the Geostat 1A population raster, by comparing it with accurate census data of the Austrian Statistical Bureau. Besides, a detailed evaluation of the disaggregation accuracy of different CLC classes provides strengths and weaknesses of the disaggregation approach as such.

3.1 Spatial Disaggregation of Geostat 1A Population Raster

Spatial disaggregation of datasets refers to a process that creates high resolution datasets based on low resolution information with auxiliary data. For the case of population density, CLC data employed as ancillary information on where population is living. The disaggregation approach used here is similar to Gallego and Peedell (2001) and Gallego (2010). The spatial disaggregation methodology is described in detail in this section.

In order to spatially disaggregate the population data of the Geostat 1A raster with 1 km resolution CLC data are employed and integrated similar to the CLC-iterative method described by Gallego (2010), Gallego and Peedell (2001) and Thieke et al. (2006). Gallego (2010) as well as Gallego and Peedell (2001) describe downscaling of population data based on population data per commune (EU LAU 2 level) and CLC—which is an approach that is based on different spatial resolution. This is underpinned by the fact that the spatial extent of communes varies to a certain extent. This is reflected by the statistical evaluation of the surface area of LAU 2 entities based on EUROSTAT (2012) which is depicted in Table 1. In this table the mean surface area of a LAU 2 entity—i.e. a community—and the standard deviation and the skewness is given. The standard deviation given in Table 1 indicates that the spatial extent of the communities varies, which results in different spatial resolutions. In this chapter, we disaggregate based on one

Table 1 Statistical evaluation of LAU2 entity area data for the EU27 (except Denmark and Germany) based on EUROSTAT (2012)

Statistical metric	LAU2 entity area (km^2)
Mean	37,64
Median	145,20
Standard deviation	211,67
Skewness	47,56

homogeneous spatial resolution which is determined by the resolution of the Geostat 1A population grid—1 km.

The disaggregation of the population data—i.e. the Geostat 1A population raster is done using the CLC-iterative method. This approach assumes that base data to be disaggregated is available as polygons representing communes (Gallego 2011; Gallego 2010; Gallego and Peedell 2001). In this chapter this approach is altered, due to the fact that the dataset to be disaggregated is a raster data model. The methodology presented here does not follow the CLC-iterative method presented in Gallego and Peedell (2001) completely. Hence communes cannot be stratified within each NUTS2 region into dense, intermediate and sparse population communes. NUTS is an abbreviation for the Nomenclature of territorial units for statistics, a hierarchical system of territorial units in the European Union according to EC Regulation No. 1069/2003. Nevertheless, the model presumes a fixed-ratio:

$$Y_{cm} = U_c W_m \tag{1}$$

In Eq. 1 m denotes a raster cell in the original, coarse Geostat 1A grid. In this model Y_{cm} represents the population density for land cover class c in the raster cell m of the Geostat 1A grid. In addition, U_c denotes the relative population density for each land cover class c. W_m is a number that ensures the pycnophylactic constraint (Tobler 1979) for each raster cell m after the estimation of U_c. In order to calculate U_c and Y_{cm} the following equations (Eq. 2) are necessary:

$$\begin{aligned} X_m = \sum_c S_{cm} Y_{cm} \Rightarrow W_m &= \frac{X_m}{\sum_c S_{cm} U_c} \\ \Rightarrow Y_{cm} &= U_c \frac{X_m}{\sum_c S_{cm} U_c} \end{aligned} \tag{2}$$

X_m denotes the population in raster cell m, and S_{cm} is the space that is covered by land cover class c in raster cell m. The disaggregation process starts with a parameter U_c using Eq. 3. The population data for each target raster cell m'—where m' denotes a target raster cell having finer resolution than m—are disaggregated with the coefficients U_c and S_{cm} (Eq. 3). Furthermore, $X_{m'}$ denotes the population in raster cell m', and $S_{cm'}$ is the space that is covered by land cover class c in raster cell m'

$$Y_{cm'} = U_c \frac{X_{m'}}{\sum_c S_{cm'} U_c} \tag{3}$$

Consecutively, the population attributed to raster cell m is estimated by using Eq. 4. Furthermore, the known population in the raster cell X_m—the original coarse raster cell m—is compared with the estimated population X_m^* in order to calculate disagreement indicators given in Eq. 5.

$$X_m^* = \sum\nolimits_{m' \in m} \sum\nolimits_c S_{cm'} Y_{cm'} \tag{4}$$

$$\psi_m = \frac{X_m^*}{X_m} \quad \delta_m = \sum \left| X_m^* - X_m \right| \tag{5}$$

Due to the fact that the estimation of the disaggregated population density is dependent on the population resident in the coarse raster cell m denoted as X_m the authors omit the iterative calculation of U_c. In addition, through Eq. 5 an evaluation of the population density of Geostat 1A and the estimated population of the disaggregated population raster is possible.

3.2 Evaluation of Disaggregated Population Raster

The evaluation of the disaggregated population raster is done by utilizing aggregated accurate census data provided by Austrian Statistical Bureau, having a spatial resolution of 100 and 500 m. In addition, the disaggregated population raster data are compared with a guided—i.e. weighted—random distribution, in order to evaluate if the results of the disaggregation process show similarities to a random distribution. The evaluation approach is depicted in Fig. 3. In general, the Geostat 1A population raster is disaggregated and compared with the reference dataset of the Austrian Statistical Bureau and a weighted random population grid, resulting in several deviation numbers and statistics. The following sections elaborate on the evaluation method in detail.

In order to evaluate the accuracy of the disaggregated population grids the disagreement between the population of disaggregated raster cells X_m^* and the reference raster cells X_m^R is calculated by using Eq. 6. Due to the same extent and origin of the two grids a direct comparison is possible. This is done for two given spatial resolutions, 100 and 500 m respectively—resulting in disagreement $D_{Da,R,100}$ and $D_{Da,R,500}$. Similar to Eq. 6 the absolute disagreement between reference and weighted random data (see Eq. 7)—$D_{Da,Rnd,100}$ and $D_{Da,Rnd,500}$—as

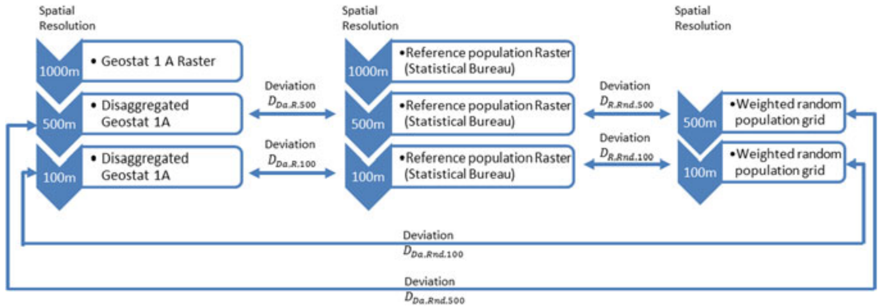

Fig. 3 Evaluation approach of the disaggregated population raster. The approach emphasizes on the deviations between the disaggregated population raster, reference population raster from Austrian Statistical Bureau and weighted random population grid

well as the disagreement between disaggregated and weighted random grid (see Eq. 8)—$D_{Da,Rnd,100}$ and $D_{Da,Rnd,500}$—is calculated.

$$D_{Da,R} = \sum \left|X^{*}_{m'} - X^{R}_{m'}\right| \tag{6}$$

$$D_{Da,Rnd} = \sum \left|X^{R}_{m'} - X^{Rnd}_{m'}\right| \tag{7}$$

$$D_{Da,Rnd} = \sum \left|X^{*}_{m'} - X^{Rnd}_{m'}\right| \tag{8}$$

In addition, the disagreements between the population grids depicted in Fig. 1 are represented as difference grids, which can be processed in any GIS. Hence, the evaluation of the disagreement grids contains a comparison of the statistical parameters for each disagreement grid. In detail, there are the following disagreement grids used in this chapter:

- $G_{Da,R,500}$: difference between disaggregated and reference population grid, 500 m resolution
- $G_{Da,R,100}$: difference between disaggregated and reference population grid, 100 m, resolution
- $G_{R,Rnd,500}$: difference between reference and random population grid, 500 m, resolution
- $G_{R,Rnd,100}$: difference between reference and random population grid, 100 m, resolution
- $G_{Da,Rnd,500}$: difference between disaggregated and random population grid, 500 m, resolution
- $G_{Da,Rnd,100}$: difference between disaggregated and random population grid, 100 m, resolution

The disagreement grids are analyzed regarding their statistics in order to analyze the "behavior". Thus, the comparison with a weighted random grid is done in order to evaluate if the disaggregation on the two examined granularity levels becomes more similar to a random distribution. Hence, the following statistical parameters are computed for each disagreement grid: maximum absolute difference, standard deviation σ, skewness, and kurtosis. In addition, the calculation of the correlation matrix between the population rasters is done in order to evaluate the similarity between them. Correlation matrices express the similarity of raster layers (Snedecor and Cochran 1968).

The weighted random distribution, used to evaluate the nature of the disaggregated population raster, is a function that creates random point distribution—i.e. population—with respect to a given probability. Due to the fact, that the basic correlation between CLC and population density is known in advance, we used that information in order to create a probability surface for the placement of random points. Hence, the random point generator placed the number of "humans" in the study area that is defined by the census data of the Austrian Statistical Bureau. The random points are placed such that raster cells with larger

Table 2 Population density classes and respective Corine Landcover classes (Gallego and Peedell 2001)

Population density	Corine Landcover classes
Dense	112
Medium	211, 222, 231, 242, 243
Sparse	121, 122, 123, 141, 311, 312, 313, 321, 322

values are more likely to have a point placed in them—which is defined by Corine Landcover. For that reason the CLC classes were divided into three categories (Gallego and Peedell 2001): densely populated, medium populated and sparsely populated. The respective Corine Landcover Classes are displayed in Table 2. Subsequently, the weighted random points are aggregated into regular grids having 100 and 500 m resolution sharing the extent of the reference grid of the Austrian Statistical Bureau.

In addition, an evaluation of the population grids for three population density classes, given in Table 2, is conducted. The population grids are divided into population density classes. The correlation coefficient between the population grids and the three population density classes is calculated and evaluated. Furthermore, the variance—as a sign of disagreement—is computed based on the density classes and the generated disagreement grids.

4 Results

This section elaborates on the numerical results of the 500 and 100 m disaggregation of the Geostat 1A population grid and the evaluation thereof. First, the chapter highlights the results of the disaggregation methodology and presents some graphical outcomes of the disaggregation process. Secondly, the evaluation of the disaggregated Geostat 1A raster with 100 and 500 m resolution is presented.

The disaggregation process as described in Sect. 3.1 results in population datasets that have finer granularity than the original Geostat 1A grid. The resulting disaggregated population raster with a resolution of 100 m is given in Fig. 4 *right*, denoted with *Disaggregated Population* 100. The reference dataset originating from the Austrian Statistical Bureau is depicted in Fig. 4 *left*, denoted with *Population* 100.

The resulting disaggregated population raster with a resolution of 500 m is given in Fig. 5 *right*, denoted with *Disaggregated Population* 500. The reference dataset originating from the Austrian Statistical Bureau is depicted in Fig. 5 *left*, denoted with *Population* 500.

The evaluation of the disaggregated datasets follows the approach described in Sect. 3.2. This comprises the calculation of the disagreements between population grids based on Eqs. 6–8 and Fig. 3. In addition, the disagreement grids are created

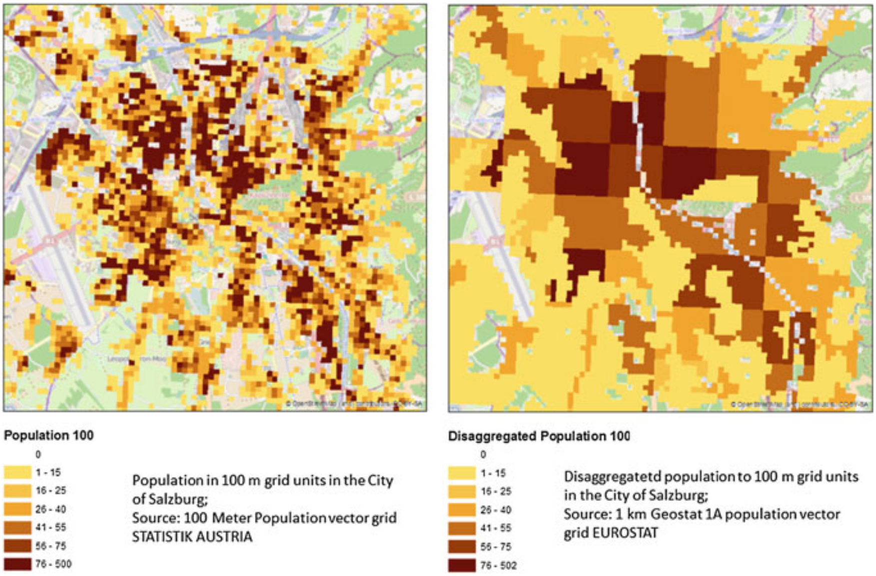

Fig. 4 Visual comparison of the reference population grid with 100 m resolution of the Austrian Statistical Bureau (*left*), and the disaggregated population raster with 100 m resolution (*right*). Both grid datasets show the city of Salzburg and their outskirts

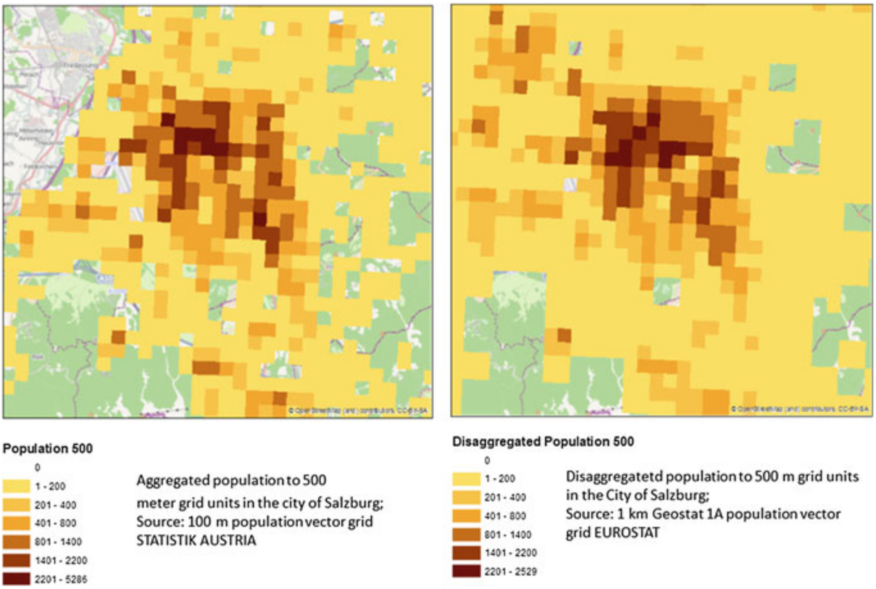

Fig. 5 Visual comparison of the reference population grid with 500 m resolution of the Austrian Statistical Bureau (*left*), and the disaggregated population raster with 500 m resolution (*right*). Both grid datasets show the city of Salzburg and their outskirts

Table 3 Statistical evaluation of the disagreement grid $G_{Da,R}$ for 100 and 500 m resolution

Reference grid versus disaggregated grid	100 m resolution $G_{Da,R,100}$	500 m resolution $G_{Da,R,500}$
$D_{Da,R}$	493239	210593
Maximum absolute difference	708	2828
Standard deviation σ	10,21	86,47
Skewness	15,63	13,8
Kurtosis	478,01	319,1

Table 4 Statistical evaluation of the disagreement grid $G_{R,Rnd}$ for 100 and 500 m resolution

Disaggregated grid versus random grid	100 m resolution $G_{Da,Rnd,100}$	500 m resolution $G_{Da,Rnd,500}$
$D_{Da,Rnd}$	353290	358170
Maximum absolute difference	143	2383
Standard deviation σ	5,6	120,05
Skewness	10,9	10,53
Kurtosis	160	141,8

Table 5 Statistical evaluation of the disagreement grid $G_{R,Rnd}$ for 100 and 500 m resolution

Reference grid versus random grid	100 m resolution $G_{R,Rnd,100}$	500 m resolution $G_{R,Rnd,500}$
$D_{R,Rnd}$	519523	406322
Maximum absolute difference	134	4730
Standard deviation σ	10,27	135,26
Skewness	18,5	13,4
Kurtosis	604	291,73

Table 6 Correlation coefficient of population grids

Correlation coefficient	100 m resolution	500 m resolution
Disaggregated population versus reference population	0,47	0,87
Random population versus reference population	0,34	0,74
Disaggregated population versus random population	0,56	0,80

with Map Algebra, and the statistics of these grids are calculated respectively. The results are presented in Tables 3, 4 and 5.

To evaluate the correlation between the population grids the correlation and covariance matrices are calculated in a pairwise manner. These results show the correlation between the reference, disaggregated and random population grids on two levels of detail (100 and 500 m). The results are presented in Table 6.

An evaluation of the standard deviation of the density classes of the disagreement grids gives completes the accuracy results of the population grids (see Table 7). In Table 8 the absolute deviation of the disaggregated population raster with respect to the population density classes is given, which is calculated based on Eqs. 6–8.

Table 7 Standard deviation of the disagreement grids and their population density classes

Standard deviation	Densely populated	Medium populated	Sparsely populated
Disaggregated population minus. reference population $G_{Da,R}$			
100 m resolution	44,2	7,5	2,4
500 m resolution	236	45,5	12,5
Random population minus reference population $G_{R,Rnd}$			
100 m resolution	49,1	7,2	2,3
500 m resolution	443,6	62,7	19,1
Disaggregated population minus. random population $G_{Da,Rnd}$			
100 m resolution	25,26	2,5	1,0
500 m resolution	376	45,5	16,1

Table 8 Absolute disagreement of the disaggregated population grid with respect to population density classes, and reference population numbers in density classes

	Densely populated	Medium populated	Sparsely populated
Absolute disagreement			
100 m resolution	161265	298394	29882
500 m resolution	50952	127526	23095
Relative disagreement			
100 m resolution	101 %	146 %	216 %
500 m resolution	37 %	61 %	83 %
Reference population			
100 m resolution	160112	204503	13833
500 m resolution	138571	208370	27955

5 Discussion of the Results

The results of the disaggregation process and the evaluation are given in Sect. 5. Based on the numerical results given, a discussion thereof is conducted. This section focuses on the interpretation of the results achieved, and comments critically on the numbers. The section highlights the disaggregated population grids and the comparison with the reference population raster data. In addition, the evaluation of the weighted random population distribution in combination with the reference and disaggregated grid should elaborate on the "difference" between disaggregation and randomly generated datasets.

First the chapter elaborates on the visual difference between disaggregated and reference population grid. The disaggregated and reference population grid with 100 m resolution are depicted in Fig. 4. Noticeable are the visual differences that are observable between the reference and the disaggregated population dataset. In comparison to the latter, the disaggregated and reference population grids with 500 m resolution (see Fig. 5) show less visual differences. This underpins the assumption that the disaggregated dataset with 100 m resolution has lower accuracy than the 500 m population raster.

A numerical evaluation of the disagreement between the disaggregated reference population grid for 100 and 500 m shows lower disagreement $D_{Da,R}$ at the 500 m level (see Table 3). In addition, the authors look at the "distance" between the reference grid to the disaggregated and random population grid. In order to have a metric for that, the authors look at the correlation coefficients of the grids and the standard deviation of the disagreement rasters. Tables 4, 5 show that the standard deviation of the disagreement grid $G_{Da,Rnd}$ and $G_{R,Rnd}$ for 100 and 500 m resolution respectively. For 100 m resolution $G_{R,Rnd}$ shows a standard deviation of 10,27 whereas the $G_{Da,Rnd}$ has a deviation of 5,6. For 500 m resolution $G_{R,Rnd}$ shows a standard deviation of 135,26 whereas the $G_{Da,Rnd}$ has a deviation of 120,05. This gives evidence, that the disaggregated population raster at 100 m is "closer" to the random data than the reference grid. For 500 m resolution the standard deviation of $G_{Da,Rnd}$ is lower than the one of $G_{R,Rnd}$, but the proportion between them is lower than at 100 m level. Hence, the authors assume that the 500 m disaggregation results differ from random population grid with a similar "distance" as the reference grid. In addition, the standard deviation of the $G_{Da,R}$ is lower than $G_{Da,Rnd}$ which shows that the disaggregated grid is "closer" to the reference grid, which is supported by the disagreement numbers $D_{Da,R,500} < D_{Da,Rnd,500}$. Furthermore, the disagreement of the disaggregated grid to the random grid $D_{Da,Rnd}$ at 100 m resolution is lower than $D_{Da,R,100}$, and the standard deviation shows a similar behavior. This indicates that the disaggregated population raster at 100 m level is closer to the random population grid. In general, the facts support the argument that the disaggregated 500 m population raster shows fewer inaccuracies than the one with 100 m resolution, which is closer to the weighted random population grid.

In order to evaluate on the similarity of the disaggregated population data with the reference population grid, the chapter highlights the correlation between the raster data sets at different levels of detail accordingly (see Table 6). Generally speaking, the correlation between disaggregated and reference grid shows a correlation coefficient of 0,87 at the 500 m level. Compared to the value of 0,47 at 100 m level the authors conclude that the 500 m disaggregation result is similar to the reference population, as the correlation of the disaggregated to the random population at 500 m resolution is slightly lower. For 100 m resolution the situation is different, as the correlation between disaggregated and random population is higher than the correlation between disaggregated and reference population. Nevertheless, the correlation coefficients at the 100 m level are low in comparison to the 500 m resolution. Generally, the correlation coefficient between random and reference population is lowest at both levels of detail, whereas the relative distance to the correlation coefficient of disaggregated and reference is lower at the 500 m resolution level. The evaluation of the correlation coefficients of the population density grids shows that at 100 m resolution level the disaggregated population raster shares most similarities with the withed random population grid, whereas at

500 m resolution the correlation coefficient between disaggregated and reference grid shows the highest value.

In addition, the standard deviation of the disagreement grids with respect to population density zones, given in Table 2, is analyzed. The results are given in Table 7, where $G_{Da,Rnd}$ shows the lowest standard deviation in all population density classes of the 100 m resolution. For 500 m resolution $G_{Da,R}$ shows the lowest standard deviation for all population density classes, except for the medium populated areas. For medium populated areas $G_{Da,R,500}$ and $G_{Da,Rnd,500}$ share a standard deviation of 45,5 which indicates that the distance between disaggregated to reference and disaggregated to random population grid are equal, and the disaggregation at the 500 m level for medium populated areas shows accuracy deficits. In addition, $G_{Da,R}$ of sparsely populated areas with 500 m resolution shows a slightly lower standard deviation than $G_{Da,Rnd}$. Hence, the distance between disaggregated and reference population grid and disaggregated and random population grid is comparable small when looking at the relative difference. Hence, the authors assume that the disaggregation for sparsely populated areas shows inaccuracies.

The absolute differences for the population density classes in different levels of detail, given in Table 8, indicate that for the 500 m resolution the absolute disagreement in densely populated areas is lowest in comparison to the reference population. For sparse and medium populated areas the absolute disagreement is generally higher in comparison to the reference population, whereas population density grid the 500 m resolution shows lower disagreement than at 100 m resolution.

6 Conclusion and Outlook

The article elaborates on the disaggregation of the GEOSTAT 1A population grid for a study area located in the northern part of the province of Salzburg and some parts of Upper Austria in Austria. The chapter describes an approach to downscale a population raster of 1 km resolution towards a target resolution of 500 and 100 m respectively, by using Corine Landcover as ancillary data. In order to evaluate the results achieved the authors employ an accurate reference dataset originating from the Austrian Bureau of Statistics representing the census in Austria.

The results achieved with the methodology described show that disaggregating of population grids with Corine Landcover is possible. The accuracy evaluation indicates that the results achieved at 100 m resolution show more correlation with a weighted random population distribution than with the reference population data. For 500 m resolution the disaggregated grid is slightly more correlated with the reference dataset. In addition, the total absolute disagreement is lower for the 500 m grid. These numerical results give evidence, that the 500 m disaggregation of the GEOSTAT 1A raster is more accurate than the 100 m disaggregation.

In detail, the results for population density zones show that densely populated areas can be estimated quite well, whereas disaggregation results in medium and sparsely populated areas tend to be more inaccurate. This fact is mentioned in Gallego et al. (2011) and Gallegeo (2010) as well.

Further research in this area could include the investigation of the effect of further ancillary data and other disaggregation approaches mentioned in literature. Interesting for the authors are crowd sourced auxiliary data, like open street map data, that have at least some additional inherent land cover information. In addition, an evaluation of the obtained accuracy in the context of the intended application area is pending.

Acknowledgments The present work has been funded by the European Commission under Framework Programme for RTD 7 through the project “Modeling and Simulation of the Impact of Public Policies on SMEs (MOSIPS)”—Grant agreement no.: 288833.

References

Bierkens MFP, Finke PA, de Willigen P (2000) Upscaling and downscaling methods for environmental research. Kluwer, Dordrecht, p 190

Briggs DJ, Gulliver J, Fecht D, Vienneau DM (2007) Dasymetric modelling of small-area population distribution using land cover and light emissions data. Remote Sens Environ 108:451–466

Chen K, McAneney J, Blong R, Leigh R, Hunter L, Magill C (2004) Defining area at risk and its effect in catastrophe loss estimation: a dasymetric mapping approach. Appl Geogr 24:97–117

Eicher C, Brewer C (2001) Dasymetric mapping and areal interpolation: implementation and evaluation. Cartography Geogr Inf Sci 28:125–138

EEA-ETC/TE (2002) Corine land cover update. Technical guidelines. Web: http://www.eea.europa.eu/publications/technical_report_2002_89

EUROSTAT (2012) Local administrative units. Web: http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/local_administrative_units. Accessed 29 Dec 2012

Flowerdew R, Green M, Kehris E (1991) Using areal interpolation methods in GIS. Pap Reg Sci 70(3):303–315

Gallego FJ (2010) A population density grid of the European Union. Popul Environ 31(6):460–473

Gallego FJ, Batista F, Rocha C, Mubareka S (2011) Disaggregating population density of the European Union with CORINE land cover. Int J Geogr Inf Sci 25(12):2051–2069

Gallego J, Peedell S (2001) Using CORINE land cover to map population density. Towards agri-environmental indicators, topic report 6/2001 European environment agency, Copenhagen, http://reports.eea.eu.int/topic_report_2001_06/en. pp 92–103

Langford M, Unwin DJ (1994) Generating and mapping population density surfaces within a geographical information system. Cartogr J 31(1):21–26

Mennis J (2009) Dasymetric mapping for estimating population in small areas. Geogr Compass 3(2):727–745

Mrozinski R, Cromley R (1999) Singly- and doubly-constrained methods of areal interpolation for vector-based GIS. Trans Geogr Inf Syst 3:285–301

Reibel M, Agrawal A (2007) Areal interpolation of population counts using pre-classified land cover data. Popul Res Policy Rev 26(5–6):619–633

Reibel M, Bufalino M (2005) Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. Environ Plann A 37(1):127–139

Snedecor GW, Cochran WG (1968) Statistical methods, 6th edn. The Iowa State University Press, Ames, Iowa

Thieken A, Mueller M, Kleist L, Seifert I, Borst D, Werner U (2006) Regionalisation of asset values for risk analyses. Nat Hazards Earth Syst Sci 6:167–178

Tobler WR (1979) Smooth pycnophylatic interpolation for geographical regions. J Am Stat Assoc 74:519–530

Tralli DM, Blom RG, Zlotnicki V, Donnellan A, Evans DL (2005) Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. ISPRS J Photogr Remote Sens 59(4):185–198

Vinkx K, Visee T (2008) Usefulness of population files for estimation of noise hindrance effects. In: Proceedings of ICAO committee on aviation environmental protection. CAEP/8 modelling and database task force (MODTF), 4th meeting. Sunnyvale, USA, February 2008, pp 20–22

Yuan Y, Smith RM, Limp WF (1997) Remodeling census population with spatial information from Landsat TM imagery. Comput Environ Urban Syst 21(3–4):245–258