Gridded Population – new data sets for an improved disaggregation approach

Steinnocher K.*, Kaminger I.°, Köstl M.* and Weichselbaum J.§

* AIT Austrian Institute of Technology
° Statistik Austria
§ GeoVille Information Systems

Abstract

There is a demand for population data that are independent from administrative areas. Raster representations meet this demand but are not yet available for all European countries. Spatial disaggregation of population data can overcome this gap and has been performed on a Europeen scale based on CORINE land cover data (CLC). The drawback of this approach is the limited spatial resolution of the CLC data set that leads to over-/underestimation of sparsely populated areas depending on the applied method.

With the recently published EEA Fast Track Service Precursor on Land Monitoring a new data set is now available that provides the degree of soil sealing for EEA-38 countries. Applying this data set as a proxy for population density the spatial disaggregation can be improved significantly. Since the soil sealing is not directly corresponding to residential building density a number of pre-processing steps are required beforehand. These steps are defined as simple rules and require CLC data as additional input.

The paper presents the results of the disaggregation for a European transect and detailed accuracy assessment for the territory of Austria. A comparison between the CLC based approaches and the new soil sealing based technique is presented and validated against the Austrian population grid provided by Statistik Austria. Remaining discrepancies are discussed and potential improvements proposed. Finally a proposal for cooperation is made in order to validate the product in other European countries.

Introduction

There is a demand for population data that are independent from administrative areas. Raster representations meet this demand but are not yet available for all European countries. Spatial disaggregation of population data can overcome this gap and has been performed on a Europeen scale based on CORINE land cover data (CLC). Gallego & Peedell (2001) applied a probabilistic disaggregation method, estimating population density weights for most CLC classes, while Steinnocher et al. (2006) used only residential classes of CLC for disaggregation[1]. The drawback of both approaches is the limited spatial resolution of the CLC data set that leads to over-/underestimation of sparsely populated areas respectively.

With the recently published EEA Fast Track Service Precursor on Land Monitoring a new data set is now available that provides the degree of soil sealing for EU27+ countries. Applying this data set as a proxy for population density the spatial disaggregation can be improved significantly. Since the degree of soil sealing is not directly corresponding to residential building density a number of pre-processing steps are required beforehand. These steps are defined as simple rules and require CLC data as additional input.

---

[1] Performed within FP6 IP geoland - Observatory Spatial Planning; Project No. FP-6-502871
http://www.fp6.gmes-geoland.info/OS/OSP/index.php

Degree of Soil sealing

The EEA Fast Track Service Precursor on Land Monitoring[2] is a raster dataset for built-up areas including continuous degree of imperviousness ranging from 0-100% in spatial resolution of 20x20m. The data set is based on orthorectified high resolution satellite imagery (Image2006), acquired primarily in the reference year 2006 (+/- 1 year). Supervised classification techniques were used to automatically map built-up areas, followed by visual improvement of the classification results. The degree of soil sealing for the classified built-up was derived from calibrated NDVI (normalised difference vegetation index). The data set covers EU27 and neighbouring countries, in total 38 countries (Kopecky & Kahabka, 2009), and is currently being updated in the frame of geoland2 for the reference year 2009. In the following we will address the data set as HR (high resolution) soil sealing layer.

Population disaggregation

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 re-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, available 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:

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

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

where Pdens and Hdens are the population and housing density respectively, Pop is the total population of the region and Ai 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 resides outside housing areas, and
- dependency between population and housing density is constant within a region.

Soil sealing to building density

The housing density required for the disaggregation approach is derived from the HR soil sealing layer, assuming that the degree of soil sealing is proportional to housing density. Since this assumption does not hold for all cases the soil sealing layer requires further processing.

In order to get a representation of housing densities it is necessary to mask out all sealed surface areas with non-residential function. These include the road and rail network, as well as industrial and

[2] http://etc-lusi.eionet.europa.eu/CLC2006/FTSP/built-up_areas

commercial areas. Masking the transport network is based on linear road and rail data, which are rasterized and slightly expanded, in order to cover associated areas as well. Non-residential built-up areas are derived from CLC classes 1.2 (industrial, commercial and transport units), 1.3 (mine, dump and construction sites) and 1.4 (parks, sport and leisure facilities). Due to the large minimal mapping unit of CLC masking is limited to larger areas of this kind. Thus smaller non-residential areas are still represented in the soil sealing layer that will cause systematic errors in the disaggregation approach (as will be shown in the validation chapter).

100% sealed surfaces outside urban areas usually indicate industrial or commercial complexes or gravel pits and therefore are masked out as well. The remaining HR soil sealing layer is assumed to represent residential building densities and is used as input to the disaggregation approach.

Population grid

The disaggregation was performed for a north-south transect of Europe, covering southern Sweden, Denmark, Germany, Poland, Czech Republic, Slovak Republic, Austria, Hungary and Italy. Population input data are provided on NUTS 3 level dated 2006, thus temporally corresponding with the HR sealing layer. Disaggregation is performed per NUTS 3 region, applying a spatial resolution of 500m. The resulting grid is shown in Fig. 1 and will be referred to as HR-POP grid.

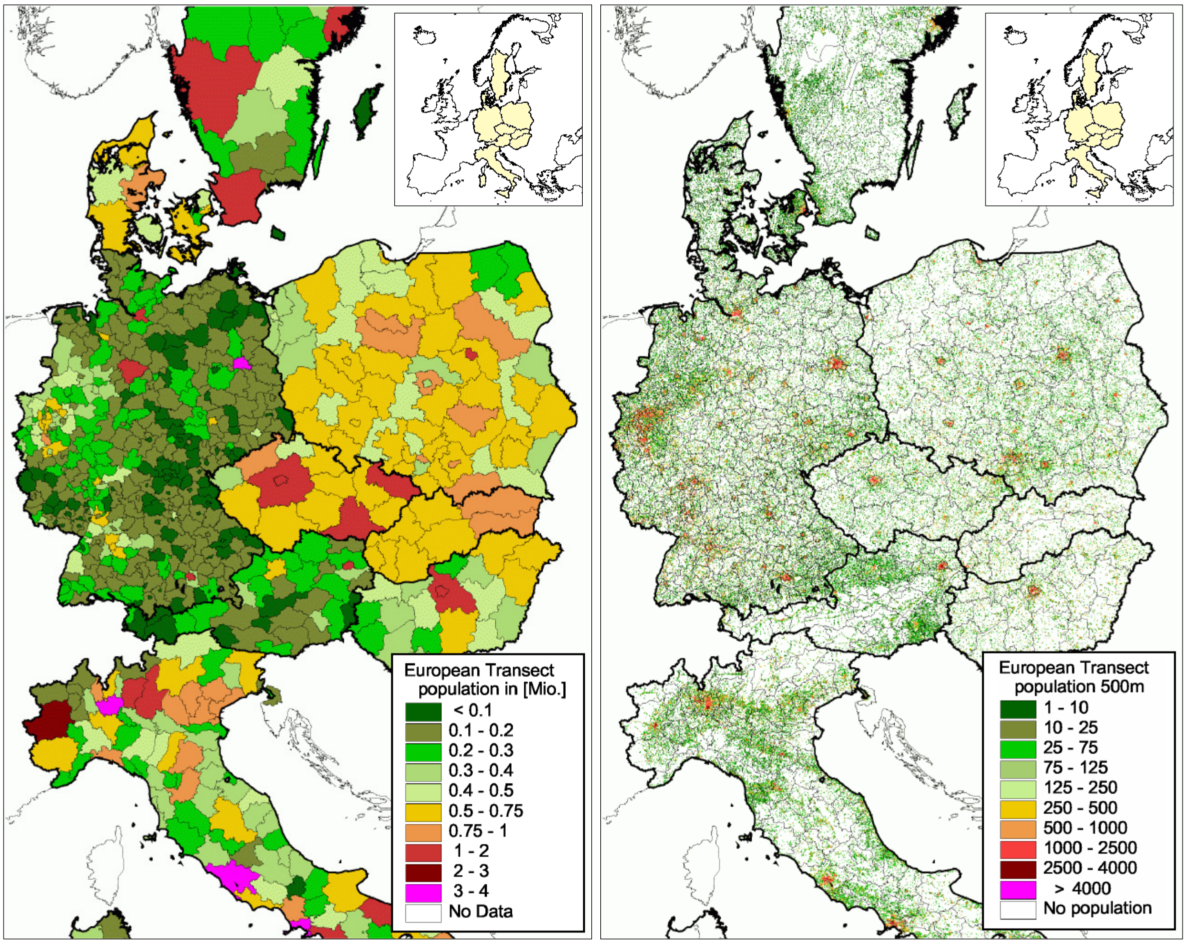

Fig. 1: European transect with popuplation on NUTS 3 level (left) and disaggregated to a 500m grid

Evaluation

For evaluating the disaggregation results a reference population grid is used. The reference grid, provided by Statistics Austria, is based on the registration census and aggregated to 500m cell size in order to correspond with the geometry of the disaggregation results. It will be referred to as StatAut grid.

In addition to the evaluation we also compare our results with the work of Gallego & Peedell (2001) and of Steinnocher et al. (2006). The former estimated population density weights for all potentially populated CLC classes and performed a probabilistic disaggregation resulting in a 1km grid covering EU 27. As it is has been produced by JRC and is now available from EEA we will refer to it as EEA-JRC grid. The latter applied the linear disaggregation method used in this study but was limited to residential classes of CLC, covering several central European countries in 500m resolution. It will be refered to as CLC-RES grid.

Fig. 2 shows the population grids for the central and eastern part of Austria. The upper images show the two CLC based disaggregation results. While the CLC-RES clearly underestimates the sparsely populated areas, the EEA-JRC populates almost all areas (except for high alpine areas and water bodies). In terms of spatial distribution the HR-POP grid is by far closest to the StatAut reference grid.

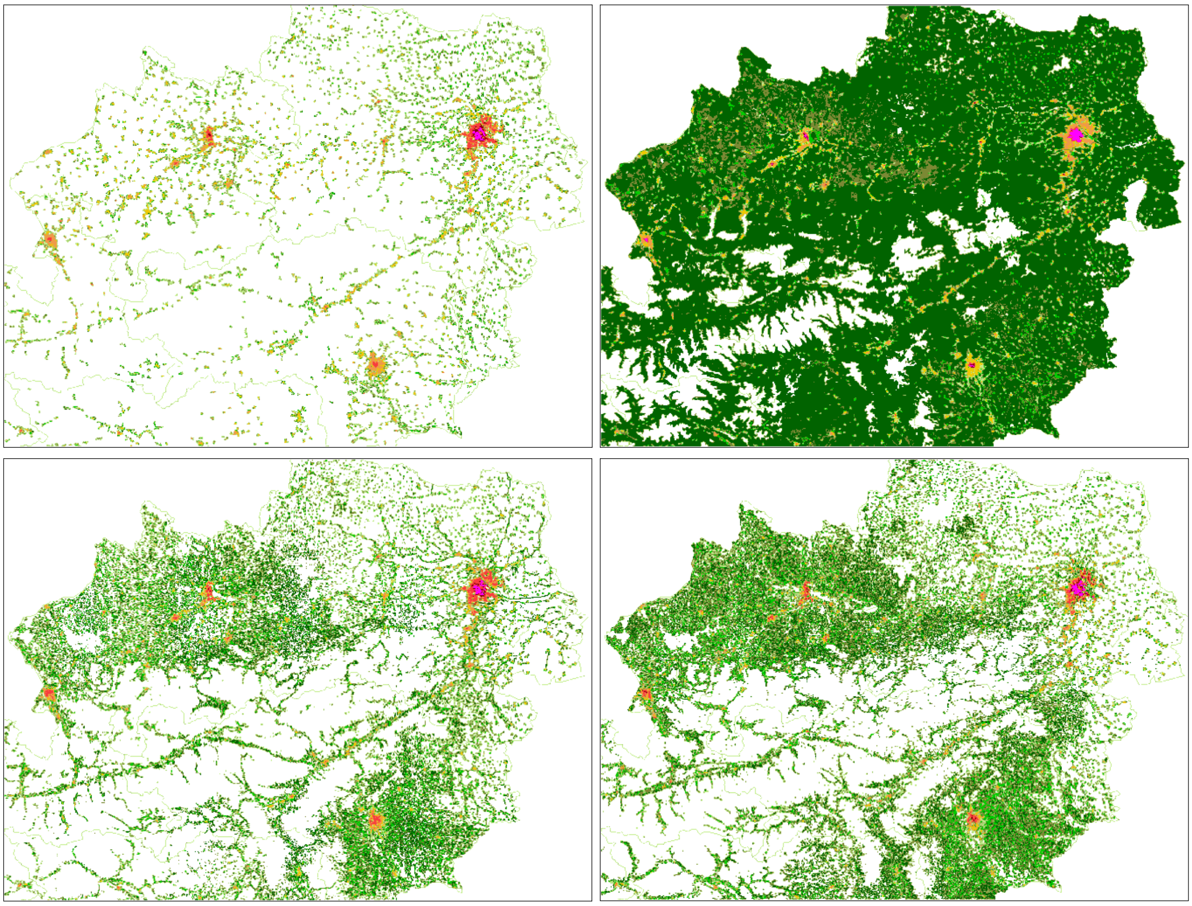

Fig. 2: Comparison of population grids: CLC-RES (u.l.), EEA-JRC (u.r.), HR-POP (l.l.) and StatAut (l.r.)

For a numeric evaluation the population cell values were grouped into classes and differences per grid cell from the StatAut reference grid were calculated. Fig. 3 shows the relative deviation of EEA-JRC, HR-POP and CLC-RES for the nine classes. There is a general tendency of overestimating lower populated grid cells and underestimating highly populated ones. One reason for this might be that with increasing building densities also the occurence of higher buildings increases. This leads to a non-linear relation between housing and population density that is not considered in the disaggregation models. The largest overall deviations occur with the CLC-RES grid, which is due to the missing representation of smaller settlements in the CLC data set.

The largest single deviation occurs for non-populated areas (class 0) whereto EEA-JRC assigns 8.6 % of the total population. This confirmes the visual impression of Fig. 2 and represents a systematic error of this data set. For the remaining classes the deviations of EEA-JRC and HR-POP lie in the same range, with HR-POP reaching better results for most classes. However, for the very highly populated classes which occur only in the center of urban agglomerations the EEA-JRC grid yields extremly good results.

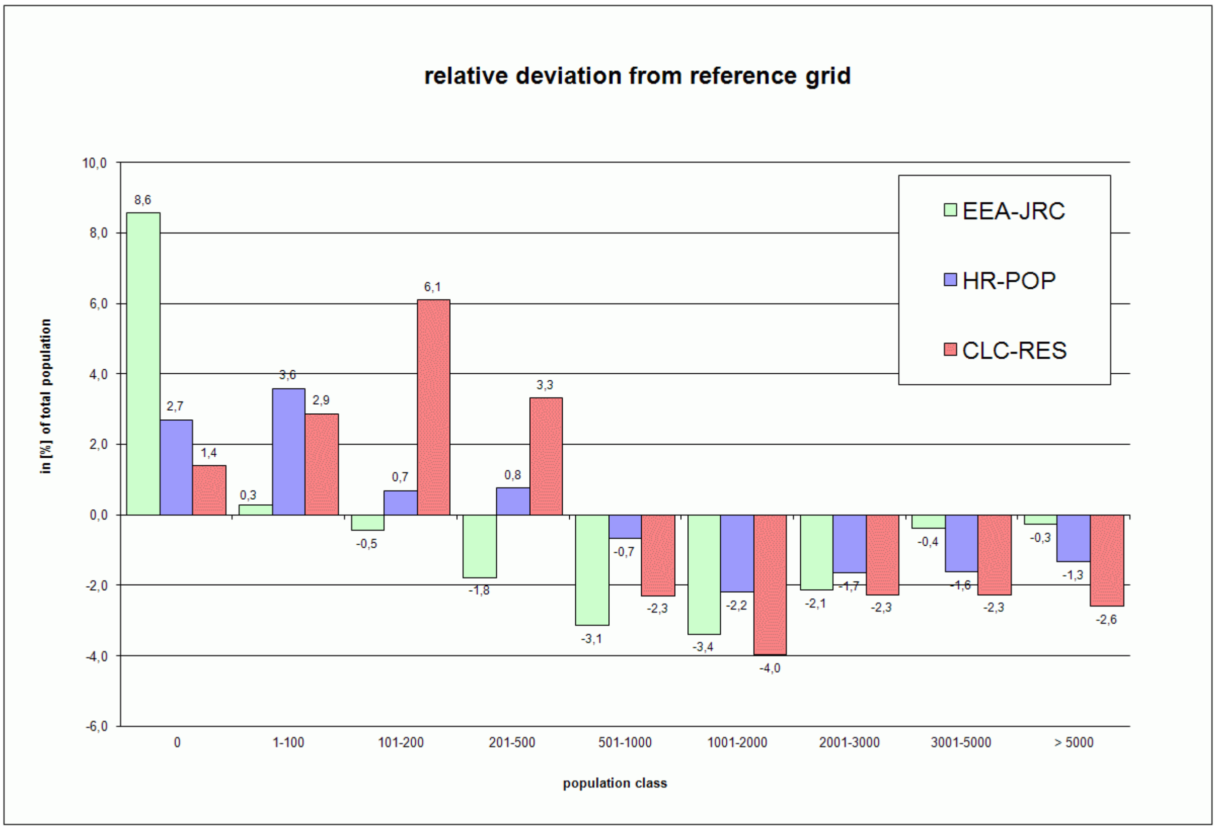

Fig. 3: Deviation from StatAut reference grid (in population classes)

In addition to the comparison of classes scatterplots give a good indication of how well numerical grids correlate. Fig. 4 shows the scatterplot between the StatAut reference grid and the HR-POP and EEA-JRC grids respectively for the provinces of Salzburg and Upper Austria. While there is a positive correlation between HR-POP and StatAut, the pattern in the right scatterplot indicates no clear correlation between EEA-JRC and StatAut. The latter results from the fact that population densities are constant within one CLC class and similar between neighbouring municipalities. This leads to the flat distribution on a low population level for rural areas. Settlements within these areas are too small for the CLC map and their population is underestimated. The randomly distributed points on the other hand represent cities that are represented in the CLC map. Their population is likely to be overestimated as a compensation for the underestimated population of smaller settlements.

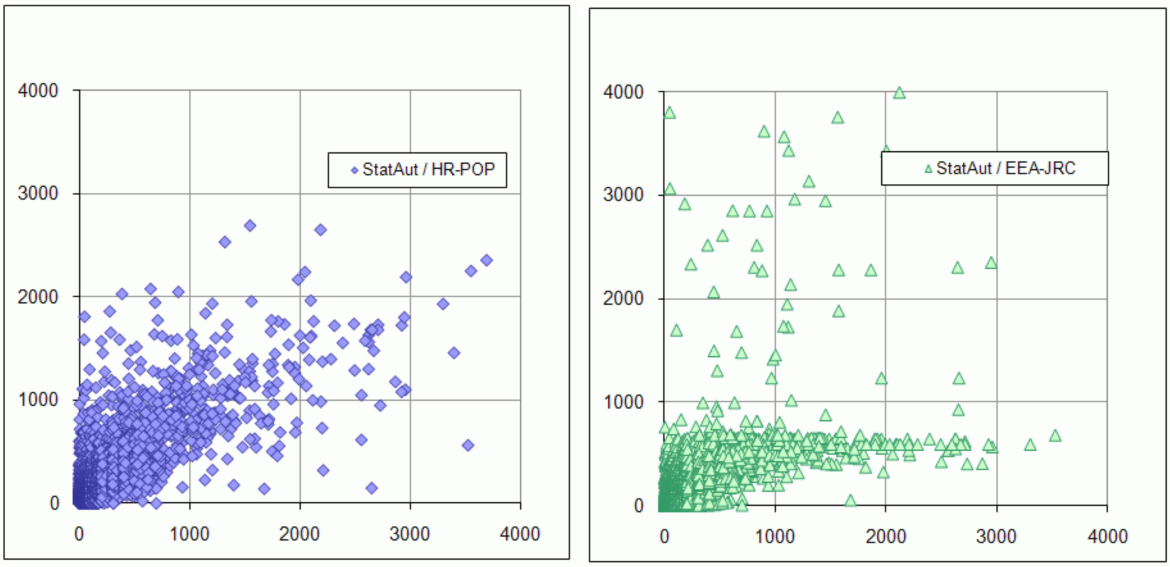

Fig. 4: Scatterplots of HR-POP and EEA-JRC grids versus StatAut reference grid (provinces of Salzburg and Upper Austria)

Besides the comparison on a regional level selected municipalities were analysed on a local level in order to better understand systematic errors in the HR soil sealing layer. One problem encountered is the already mentioned missing of the third dimension, leading to a systematic underestimation of the population in urban centers. In contrast to that town centers have a tendency of being overestimated as the population density is lower than indicated by the degree of sealing (streets and other impervious areas are not masked out). Masking of industrial and commercial areas is limited to large complexes whereas smaller areas with industrial or commercial function cannot be identified. As these areas usually have a high degree of soil sealing a large population number will be assigned to these areas. A similar effect occurs if gravel pits or quarries appear in the HR soil sealing layer.

In rural areas roads are masked out based on linear road network data. If the positional accuracy of these network data is low masking will be incomplete and roads will be populated in the course of the disaggregation process. On the other hand single buildings such as farms or scattered settlements are likely to be missed in the HR soil sealing layer, leading to an underestimation of population.

Conclusion

With the recently published EEA Fast Track Service Precursor on Land Monitoring a new land cover data set is now available that provides the degree of soil sealing for EEA 38 countries. This data set was applied for spatial disaggregation of population resulting in a 500m population grid for a European transect. The results were evaluated against a reference grid of Austria derived from the registration census of Statistics Austria.

The results of the evaluation are encouraging, although there are some limitations in the use of the soil sealing layer as a proxy for housing density. (1) Improvements are needed for the separation of residential versus industrial or commercial land use, (2) the third dimension of built-up areas cannot be considered in urban centers, and (3) road and rail networks cannot be eliminated entirely. In order to overcome these limitations the use of Urban Atlas data for population disaggregation is currently analysed. The Urban Atlas is providing pan-European comparable land use and land cover data for Large Urban Zones with more than 100.000 inhabitants as defined by the Urban Audit (Urban Atlas, 2010).

So far the evaluation of the HR-POP grid was limited to the territory of Austria due to the availability of the bottom-up reference grid of Statistics Austria. For a European evaluation there is a need for further reference data sets. We therefore like to finish the paper with a proposal for cooperation with statistical offices in European countries who are interested to support our effort in producing a European wide population grid.

Acknowledgements

The study was performed within the Core Information Service for Spatial Planning of the project geoland2 (http://www.gmes-geoland.info/service-portfolio/core-information-services/spatial-planning.html) in the frame of the GMES (Global Monitoring for Environment and Security) initiative. The project geoland2 is a Collaborative Project (2008-2012) funded by the European Union under the 7th Framework Programme (project number FP-7-218795).

References

Gallego J., Peedell S. (2001) Using CORINE Land Cover to map population density. Towards Agri-environmental indicators, Topic report 6/2001 EEA, Copenhagen, pp. 92-103. http://www.eea.europa.eu/publications/topic_report_2001_06 (last visited 10/09/2010)

Kopecky M., Kahabka H. (2009): Updated Delivery Report - European Mosaic. http://www.eea.europa.eu/data-and-maps/data/eea-fast-track-service-precursor-on-land-monitoring-degree-of-soil-sealing-100m-1 (last visited 10/09/2010)

Steinnocher K., Weichselbaum J., Köstl M. (2006): Linking remote sensing and demographic analysis in urbanised areas. In (P. Hostert, A. Damm, S. Schiefer Eds.): First Workshop of the EARSeL SIG on Urban Remote Sensing "Challenges and Solutions", March 2-3, 2006, Berlin, CD-ROM.

Urban Atlas (2010): GMES - Mapping Guide for a European Urban Atlas. Document Version 1.1 dated 26/08/2010. http://www.eea.europa.eu/data-and-maps/data/urban-atlas (last visited 21/09/2010)