Do we need a global human settlement analysis system based on satellite imagery?

Daniele Ehrlich, Martino Pesaresi
Institute for the Protection and Security of the Citizen
European Commissions, Joint Research Centre
Ispra, Italy
daniele.ehrlich@jrc.ec.europa.eu

Abstract— **Urbanization figures that document the rapid urbanization of this century are derived from census data and available from population figures aggregated at the administrative level. These invaluable figures for reporting purposes are unsuited for the operational physical analysis of the global built environment. The rigorous quantification of the built up environment is required for city and regional planning, national comparative built up analysis within cities and rural areas, global built up assessment for climate change models and for environmental change adaptation measures especially in coastal areas, as well as proxy to exposure in disaster early warnings systems and disaster loss models. The paper briefly analyses the potential alternative sources for physical built up and identifies the advantages of aerial and High Resolution satellite imagery for this task. The paper provides an analysis of the required data needed to monitor urbanization. This paper introduces the GHSL used to process imager for the analysis of the built environment. It then focuses one processing products that derives human settlement information from the processing of Very High Resolution imagery. The paper shows with an example, the potential of such systems and shows the strategies to measure the extent of built up areas and also support the criteria to validate a future global human settlement layer.**

I. INTRODUCTION

There are no shortage of sources that report of our urbanization century, the massive population growth and the unprecedented expansion of urban areas. These arguments are supported, by and large, by official population statistics available globally. In fact, urbanization is considered above all a social phenomenon related to migration of people from area of low concentration of people in rural settings to higher concentration of people typical of urban centers.

Global statistics on urbanization originate from censuses and statistical survey that focus, by and large, on the enumeration of people [1]. The figures are made available at the geographical units that coincide with the smallest administrative subdivision of that country. The urban areas are typically defined by at least two parameters or the combination of the two that include the total amount of population in the district and the density of population. In addition to population total and density there are other parameters related to that may be included in the definitions of urbanization based on the dominant economic function. Because of the diversity of definitions between countries, urbanization figures are difficult to compare. Low density population may be classed rural in one country and urban in a different country that uses lower population thresholds to define urban areas. Also, these global statistics o lack the spatial detail needed in an operational context.

The census based global population statistics are unable to provide answers to many urbanization questions raised by national governments, international organizations, donors, and the scientific community. Un-answered questions include: Is the often reported statistics that only 3% of the land masses are occupied by built up [2] still valid? Shouldn't this statistics be updated given today's world population of 7 billion? What is the rate of sprawl/urbanization in different regions of the world? How does the urbanization process compare within a country and across countries? What are the new patterns of urban sprawl? How do the new urbanization patterns change transport, energy consumption and sustainability? How does city extent and urban sprawl affect disaster risk and crisis management? Can we produce improved global fine scale physical exposure and population datasets to support global early warning systems and disaster loss models?

Urbanization is not only defined based on population statistics. The term urbanization also refers to the physical transformation of Earth land surface into constructed land or built-up. The built up is the land occupied by buildings used for dwellings and economic activities. That built-up occurs in both urban and non-urban settings. At fine scale, built up areas can be mapped by delineating clusters of contiguous buildings. Built up areas are also referred within the geographical sciences and the remote sensing community as human settlements. The aggregation of contiguous built up areas/human settlements define the spatial extent of villages, towns and cities. Monitoring the size of human settlements through repeated measurements would allow to monitoring urbanization. To date a global human settlement information database and monitoring system for quantifying urbanization globally does not exist.

Information on the built up environment and the spatial extent of cities and villages is typically available within city management and city regional planning authorities but not for all parts of the world. Measures of built up are carried out using aerial photography, and - increasingly - with Very High Resolution (VHR) imagery. Built up information are also contained in urban land use and urban land cover classes. Urban land use, urban land cover classes and human settlement maps are not available for all the populated places, and when available not always compiled in national fine scale human settlement databases.

978-1-4799-0211-8/13/$31.00 ©2013 IEEE

The cadaster, or land parcel system also contains information on the built up. A large majority of countries, especially high income countries, maintain an inventory of all buildings for property registration and taxation purposes. Due to privacy and sensitivity issues these data are available within the country almost exclusively for the authorities that in charge. Rarely are these data released. Should the data be available for compiling settlement information there would be considerable technical challenges that relate to the characteristics of the data. In fact, the data may be still available in different geographic projections, collected with different standards, showing different semantics, and in formats not immediately accessible. In practice, access to this data for global human settlement analysis or global urbanization studies is hardly an option.

In summary, there are three sources of data that could be used to measure the built up and urbanization. First, globally available census statistics are based primarily on population figures. However, censuses are not released in a standardized way and with the spatial detail required for analyzing urbanization globally. Second, the information form planning authorities available would be suited. However, many countries do not update their plans regularly and others and also within high income countries the low density populated regions may lack such plans. And simply the tasks to ingest city plans or land use maps into a single global database may be daunting. Third, cadaster or land registry parcel systems cannot be considered due to data accessibility issues. Also, cadasters are not implemented in all countries.

This paper aims to show that global human settlement information can be obtained from high resolution optical satellite data. This paper first introduces the principles used in this study. This paper then summarizes the features of the global human settlement layer (GHSL) [3] a new system to map the built up globally and monitor urbanization based on High Resolution (HR) optical satellite data. It then presents examples of settlement maps for different parts of the world that are produced from preliminary processing within the GHSL.

II. BACKGROUND

This work focuses on quantifying the built up environment; and the process of urbanization as the physical process that converts natural into constructed land. Within this work we 1) define built up and human settlements, 2) identify the appropriate measurement systems; 3) introduces the processing infrastructure of GHSL, 4) and propose criteria for validation based on the definitions used.

A. What is built up?

This work identifies the building as the characterizing element of the built up environment. Buildings are intended as the man-made construction with walls and roof used as dwellings and for other societal functions. The requested information product is the built up area map that can be defined quantitatively based on the location of the building and the spatial rule that expands the built up area beyond the building footprint [4]. The built up area map is also referred herein as fine scale human settlement map and is the objective of the information extraction.

This work makes a distinction between fine scale human settlement maps and coarse scale settlement map. Fine scale human settlement maps are derived from HR imagery where buildings can be identified and mapped, and are then validated based on presence of buildings using rigorous spatial criteria. Coarse scale human settlement maps are derived from coarser satellite imagery that are unable to resolve isolated building and sparsely scattered built up areas. The objective of the information extraction includes buildings as well as other image objects and image features. In fact, settlement maps have been derived also from proxy variable to the built up, and there is no accepted semantics or definition that can be used to validate human settlement maps derived from coarse resolution satellite imagery.

HR optical remote sensing is the measurement system of choice. Aerial photography has been the preferred datum for the analysis for detecting buildings, built up areas and the built up environment. That is used within local city administration for planning and territorial management. However, aerial photography is typically not available globally. Also, aerial photography when accessible is not always available with technical parameters that would facilitate automatic massive data processing.

The commercially available HR satellite imagery suites the purpose global fine scale human settlement mapping and monitoring. 1) The datum is adequate to map buildings and the built up. 2) The datum is commercially available and thus accessible across the globe with restriction on very few geographical areas. Also, the number of system/sensors that will provide HR data in the future is increasing allowing for improved urbanization monitoring. 3) The imagery can provide information that can be calibrated and standardized for global processing and thus comparison in space (geographical comparisons) and time (monitoring urbanization).

Extracting built up area information for mapping and monitoring the extent of human settlements globally is a daunting tasks. The operational mapping of built up that is still conducted using visual analysis techniques, is not an option for global operational processing. This paper introduces the Global Human Settlement Layer a methodology and information technology platform hosting a range of information extraction algorithm customized to different image types.

B. Image processing within GHSL

The information products presented herein are processed within the GHSL methodology designed to process satellite imagery to map human settlements globally. It hosts information extraction algorithms customized to process satellite imagery at different resolution and that exploits the morphological, texture and, spectral characteristics of the imagery. Each processing can produce therefore intermediate products that that can then be combined to maximize the extraction of information. The outputs include among others built up area, average buildings size, building counts, and built up densities. The outputs scale and information content is related to the input data and the type of processing applied. Processed output are typically available at 1:10 000, 1:50 000 and 1:500 000 scale [3]. The GHLS relies on a processing infrastructure constructed around two modules on HPC machine with 128 GM of RAM and a 4 quad-core Intel(R), Xeon(R), CPUs E7420 2.13GHz. The infrastructure typically processes 1 Gbyte every minute. Every QucikBird panchromatic band was thus processed in less than two minutes.

This paper discusses the output of only one intermediate product available within the GHSL, which is related to the first processing of over 600 VHR images [3]. This paper evaluates only the subset of maps derived from processing the panchromatic channel of QuickBird imagery with spatial resolution of 0.6 x 0.6 m. This preliminary assessment will guide more elaborated information products within GHSL that will combine information from the multispectral bands of QuickBird and bands from other sensors as well. This test used algorithms that exploit the textural [5] and morphological properties [6] of the imagery to derive a built up area index – the fine scale human settlement map - available as raster file with spatial unit of 10 x 10 m and RMS of 5 m.

To evaluate the results we recoded the built up area index into binary built up area maps. Figure 1-3 show examples of three 6 x 6 km large study sites as listed in Table 1. The figure shows the image, with overlaid the built up area map and the 1 km UTM grid. The three cities and location within a city were selected to to represent different built up patterns and geographical setting.

For the first site, Guadalupe (Figure 1), we also provide basic statistics of agreements based on a reference dataset of building footprints. At visual inspection compared the building footprint dataset with the QuickBird imagery. The reference dataset was deemed adequate for comparison purposes. The agreement measures were based on a map comparison approach. For comparison purposes, both the reference building footprint dataset and the built up area index where rescaled at 30 x 30 cells large grids as proposed in Tenerelli and Ehrlich [7]. The map comparison exercise produced agreement for built up and non-built up as well as omission and Commission error as listed in Table II and visualized in Figure 1. Forthcoming work will use the concept and methodology used for the Guadalupe case study to derive accuracy assessment methods using samples of building footprints rather than a complete building footprint map for comparison.

For the two other case studies, Guadalajara (Figure 2), Dhaka (Figure 3), we provide selected 6x6 km in the outskirts of the city. The images borders follow the 1 km UTM partitioning that with the geographic coordinates are listed in Table 1. The images are part of a larger selection of image chips that are used for validation purposes. Reference to validation can be found in [3].

TABLE I. STUDY AREAS SHOWN IN FIGURE 1-3

Shown in figure	Sensor, location and UTM coordinates		
	Sensor ***Acquisition date***	***Location*** ***UTM zone***	***Long Lat*** ***Upper left***
1	QB 2007-03-29	Guadaloupe 20 N	648000 1800000
2	QB 2007-02-24	Guadalajara 13 N	661000 2295000
3	QB2- 2002-02-19	Dakha, 46 N	235000 2648000

III. RESULTS

The agreement statistics for the Guadalupe site are summarized in Table 2. Figure 1 shows the agreement (yellow), the omission (blue) and the commission (light brown). This map comparison is not an accuracy assessment, for that a “certified” ground truth data should be available. For example, in Figure 1 the UTM grid cell corresponding to col 3 and row 4 shows that the classifier properly classifies built up that is not reported in the reference dataset. That accounts for commission. Cell in col 6 and raw 3 shows the image is contaminated by clouds and cloud shadows that prevent the image from being classified. That results in omission. The automatically produced built up area map fails to map scattered building in complex built environment as seen in cells corresponding to col 1 and row 2 and 3 of Figure 1. The overall agreement of statistics for the nearly 32 878 cells (water is excluded from analysis) shows that this preliminary output computed automatically with no human intervention can already be considered for operational processing. The processing of the spectral band of the same images is likely to improve the results. Testing is currently underway to combine outputs from different processing strategies.

Figure 1. Thirty six km^2 large area centered on a populated area of Guadaloupe.

The image section shown in Figure 2 is centered on the South West periphery of the city of Guadalajara. The built up is clearly visible under the transparency as a regular and planned structures. The image section shown in Figure 3 Dhaka is centered on the North West of the city of Dhaka (table 1, Figure 3) and shows irregularly spaced settlements of the built up. The two cities show completely different built up patterns and both are successfully mapped automatically.

TABLE II. MAP COMPARISON STATISTICS

	Sensor, location and UTM coordinates				
	BU auto	N_BU auto	**Total Cells 32878**	Agreement	85
BU Ref	8473	3100		Comission	9.4
N_BU Ref	1916	19389		Omission	5.0

IV. CONCLUSIONS

This paper argues for a human settlement information monitoring system based on high resolution satellite imagery. The arguments is supported by the demand for built up information at fine scale put forward by different user

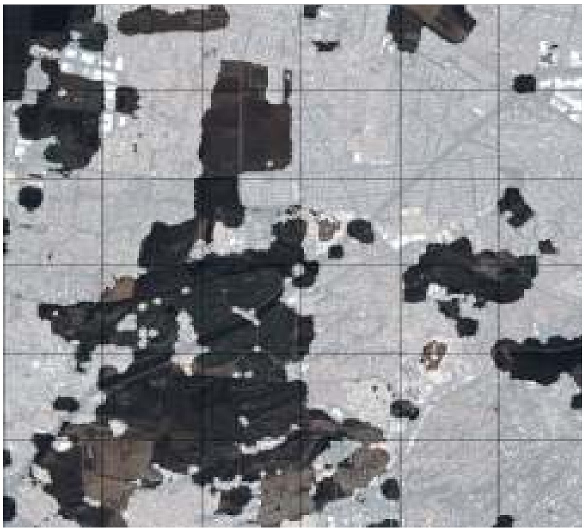

Figure 2. Thirty six km^2 for the South-West of Guadalajara (Mexico). The built up map (white) is overlayed and showed in trasparency. The UTM grid of 1 x 1 km is superimposed on the imagery. The light are

communities. In fact, the current statistics are based on population rather built up information and available at coarse resolution. The potential alternative sources of information on the built up which includes urban land use maps, urban cover classes, built up area maps are not standardized products, fragmented and not available for the entire Earth's land surface. The cadaster information provides very detailed information but largely not accessible for privacy issues. The paper provides definitions of what a human settlement information layer should be and how it is derived. It introduces a processing infrastructure capable of processing imagery at global level and provides examples of one of the potential outputs. The paper also illustrates validation concepts that can be applied for validation human settlement maps globally.

This paper shows that the datasets, the tools, the information processing infrastructure are now available. The testing of the GHSL infrastructure will continue. The next challenge is to make the data available and to dedicate resources to operate such an urbanization monitoring system, and this is an institutional issue.

Figure 3. Thirty six km^2 coverin ghe North East of Dhaka (Bangladesh). The built up map (white) is overlayed and showed in trasparency. UTM grid of 1x 1km is superimposed on the image

ACKNOWLEDGMENT

The authors would like to thank the many JRC colleagues that contribute to develop the GHSL and have made this work possible.

REFERENCES

[1] United Nations, Economic and Social Affairs ,2012, "World Urbanization Prospects The 2011 Revisions" available <http://esa.un.org/unup/pdf/FINAL-FINAL_REPORT%20WUP2011_Annextables_01Aug2012_Final.pdf>

[2] P. Gamba and M. Herold. "Preface to Global Mapping of Human Settlment" in Global Mapping of Human Settlment, Ed. P. Gamb and M. Herold, 2009, pp. xi-xii.

[3] M. Pesaresi, M; Ehrlich, D; Ferri, SL; Halkia, , M; Kemper, T; G; Scavazzon, M; Soille, P; Syrris, L. Zanchetta. "Global Human Settlement Layer: Methodology and Implementation", 2012 EUR report.

[4] D. Ehrlich and P Tenerelli. "Optical Satellite imagery for quantifying spatio-temporal dimensio of physical exposure in disaster rsk assessments". Natural Hazards, 2012. In press

[5] M Pesaresi, A. Gerhardinger, F. Taytakire, "A roust built-up area presence index by anisotrpic rotation-invariant textural measure" IEEE J. Sel. Toics Appl. Earth Observ. Remote Sens. Vol 1. No. 3, 2008, pp. 180-192.

[6] C. Ouzounis, M. Pesaresi and P. Soille, "Differential area profiles: decomposition properties and efficient computation". Pattern Analysis and Machine Intelligence, IEEE Transaction on, 2012, Vol. 34 no. 8, aug 2012., pp. 1533-1548.

[7] P. Tenerelli, D. Ehrlich, "Analysis of built-up spatial pattern at different sales: can scattering affect map accuracy?" Int J Digital Earth, 2011 4:107–116.