Preliminary Assessment of the Global Urban Footprint and the Global Human Settlement Layer for the city of Milan

Abstract

Two new global urban products have recently appeared: the Global Urban Footprint (GUF) and the Global Human Settlement Layer (GHSL). This paper evaluates the GUF and GHSL for the city of Milan, Italy through comparison with two European Union (EU) land use/cover reference products, namely the Urban Atlas and LUCAS. The results demonstrate that the GUF and GHSL are very similar to each other and, with some exceptions, show overall good agreement with the reference datasets. This study will be extended to other European cities in the future.

Keywords: Assessment, Global Urban Footprint, Global Human Settlement Layer, Land use/cover maps, LUCAS, Urban Atlas.

1 Introduction

Better baseline information on urban areas is needed. Early efforts at the delineation of urban areas were included as part of global land cover mapping from remote sensing, e.g. urban land cover (or built-up or artificial surfaces) is one of the land cover classes in numerous global land cover maps, e.g. GlobeLand30 (Chen et al., 2015). Population-based methods have also been used to delineate urban from rural areas, e.g. through the GRUMP products developed at Columbia University (CIESIN, 2004). However, when these products have been compared spatially, there are large disagreements in urban extent globally (See et al., 2013).

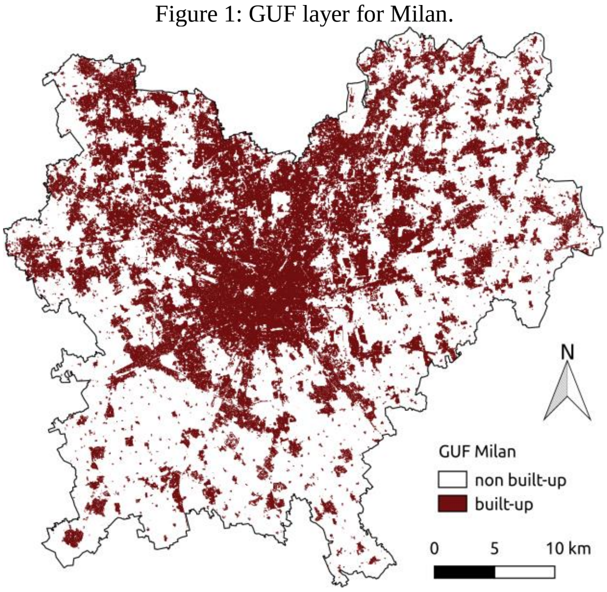
More recently, a new generation of global urban products has appeared, i.e. the Global Urban Footprint (GUF) developed by the German Aerospace Center (DLR) (DLR, 2016), which provides a global urban mask, and the Global Human Settlement Layer (GHSL) produced by the Joint Research Center (JRC) of the EU (Pesaresi et al., 2016), which provides a more detailed characterization of urban landscapes. Although these products have been validated by the map producers, they also need to be evaluated from a user perspective in terms of how well they capture urban areas around the world. Thus, the aim of this paper is to assess the GUF and GHSL for Milan municipality, Italy (hereafter Milan), based on a comparison with the latest Urban Atlas product for Milan (EEA, 2011) and LUCAS data from 2015 (Eurostat, 2015a), both of which are authoritative sources of reference information. In the next section (Section 2) these four data sources are described, followed by the methodology for assessment of the two global urban products for Milan (Section 3), the results (Section 4), conclusions and future work (Section 5).

2 Datasets

2.1 Global Urban Footprint (GUF)

The Global Urban Footprint (GUF) is a global mask of built-up areas at a resolution of 12 m (at the equator) in WGS84 reference system (EPSG: 4326), where built-up areas are defined as areas that contain man-made buildings with a vertical or height component. The mask contains three values: 0 for non-built up, 128 for missing data and 255 for built-up areas. The product was created using around 180K scenes from TerraSAR-X and TanDEM-X for the reference year 2011. The GUF was then post-processed using a number of different ancillary data, e.g. road information and settlement data from OpenStreetMap (OSM), among many others. Figure 1 shows the GUF mask for Milan.

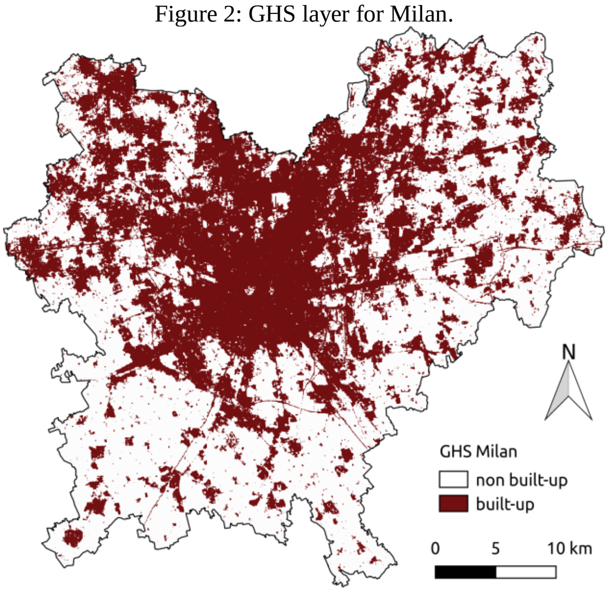
Figure 1: GUF layer for Milan.



2.2 Global Human Settlement Layer (GHSL)

Developed by the Joint Research Centre (JRC) of the European Commission, the Global Human Settlement Layer (GHSL) was produced using a fully automatic classification workflow using multi-resolution (0.5m-75m), multi-platform (e.g. SPOT, Landsat, Sentinel), multi-sensor and multi-temporal imagery (Pesaresi et al., 2016). Similar to GUF, the main GHSL product (named GHS BUILT UP GRID, hereafter simply GHS) is a global mask showing built-up areas (defined by the presence of constructed, man-made objects with a vertical dimension) at a 38m resolution in Google Mercator projection (EPSG: 3857). The 2014 map for Milan is in Figure 2.

Figure 2: GHS layer for Milan.



2.3 Urban Atlas (UA)

The GMESUA (Global Monitoring for Environment and Security Urban Atlas), referred to hereafter as the Urban Atlas (UA), is a detailed spatial characterization of urban areas in EU member states produced by the European Environment Agency (EEA, 2011). This product, available in ETRS89 with a Lambert Azimuthal Equal Area projection (EPSG: 3035), was first produced for the reference year 2006 using a detailed land use land cover (LULC) nomenclature. The UA LULC nomenclature is shown in Table 1 up to level 3. Data for Milan for 2006 were downloaded from the EEA website and are shown in Figure 3.

2.4 LUCAS

Every three years, Eurostat (2015a) undertakes a LUCAS (Land Use Cover Area frame Sample) survey to record the LULC at points systematically located across EU members states. The most recent survey took place in 2015 in which 273,401 samples were collected by 750 surveyors. LUCAS is the only official in-situ dataset available for EU wide validation exercises, e.g. to validate CORINE land cover or other land cover data sets (EEA, 2006). There are published protocols for data collection that each surveyor must follow (Eurostat, 2015a).

Table 2 shows the LUCAS nomenclature at level 1; level 2 is also reported only for the artificial land class (Eurostat, 2015b). The LUCAS data for 2015, available in WGS84 reference system (EPSG: 4326), were downloaded from the Eurostat website for Italy and then extracted for Milan (see Figure 4). The total number of points in Milan is 112.

Table 1: UA LULC nomenclature.

Level 1	Level 2	Level 3
1 Artificial surfaces	1.1 Urban Fabric	1.1.1 Continuous urban fabric
		1.1.2 Discontinuous urban fabric
		1.1.3 Isolated Structures
	1.2 Industrial, commercial, public, military, private and transport units	1.2.1 Industrial, commercial, public, military and private units
		1.2.2 Road and rail network and associated land
		1.2.3 Port areas
		1.2.4 Airports
	1.3 Mine, dump and construction sites	1.3.1 Mineral extraction and dump sites
		1.3.3 Construction sites
		1.3.4 Land without current use
	1.4 Artificial non-agricultural vegetated areas	1.4.1 Green urban areas
		1.4.2 Sports and leisure facilities
2 Agricultural areas, semi-natural areas and wetlands		
3 Forests		
5 Water		

Table 2: LUCAS LULC nomenclature.

Level 1	Level 2
A00 Artificial land	A10 Roofed built-up areas
	A20 Artificial non built-up areas
	A30 Other artificial areas
B00 Cropland	
C00 Woodland	
D00 Shrubland	
E00 Grassland	
F00 Bare land and lichens/moss	
G00 Water areas	
H00 Wetlands	

3 Methodology

3.1 Harmonization of Nomenclatures

To compare the GUF and GHS layers with the reference UA and LUCAS data sources, the nomenclatures of these datasets must be harmonized. The rule is to create binary raster maps with pixel values equal to 0 for non built-up areas and 1 for built-up areas. While GUF and GHS are already classified this way, UA and LUCAS must be reclassified to match the two classes. The reclassification is performed according to the rules defined in Table 3, where the codes refer to the class names available in the previous Table 1 (for UA) and Table 2 (for LUCAS).

Figure 3: UA layer for Milan.

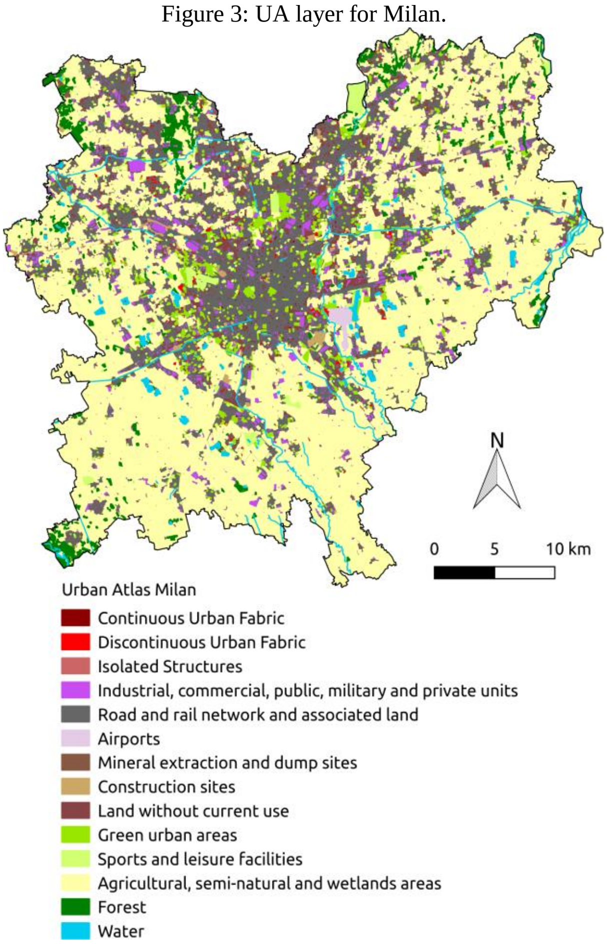
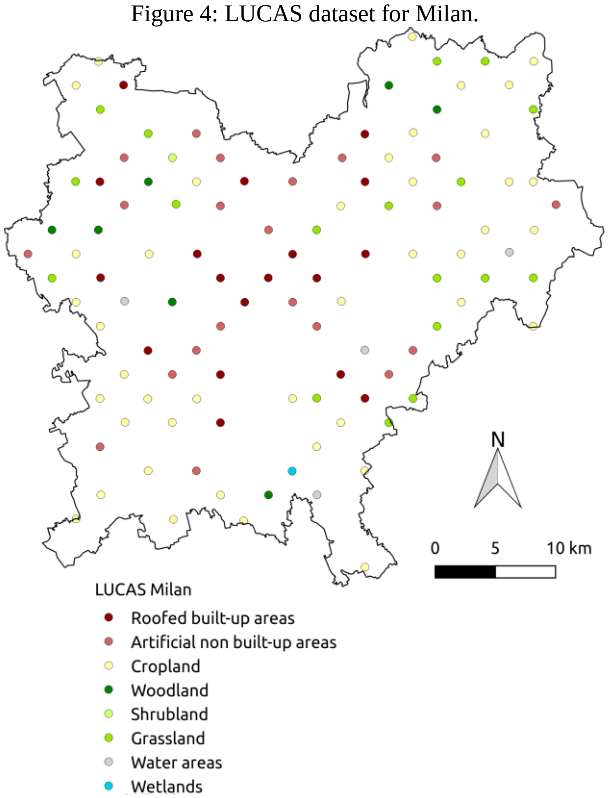


Table 3: Reclassification of UA and LUCAS classes into non built-up and built-up classes.

	non built-up (0)	built-up (1)
UA	1.1.3, 1.2.2, 1.2.3, 1.2.4, 1.3.1, 1.3.3, 1.3.4, 1.4.1, 1.4.2, 2, 3, 5	1.1.1, 1.1.2, 1.2.1
LUCAS	A20, A30, B00, C00, D00, E00, F00, G00, H00	A10

Note that a perfect match between the UA / LUCAS classes and the built-up/non built-up classes is not possible. The correspondence between classes in Table 3 is defined by determining whether or not, for each UA and LUCAS class, the presence of buildings and man-made structures with a vertical component was prevalent. Some classes classified as artificial in UA (e.g. 1.1.3 Isolated structures and 1.2.4 Airports) and in LUCAS (e.g. A30 Other artificial areas) are considered to correspond to non built-up areas, although they clearly include some presence of features with a vertical component.

Figure 4: LUCAS dataset for Milan.



3.2 Comparison of Products

The following comparisons are made:

- comparison of GUF and GHS with each other;
- calculation of overall agreement between GUF/UA, GHS/UA and GUF/GHS/UA;
- calculation of overall agreement between GUF/LUCAS and GHS/LUCAS.

4 Results

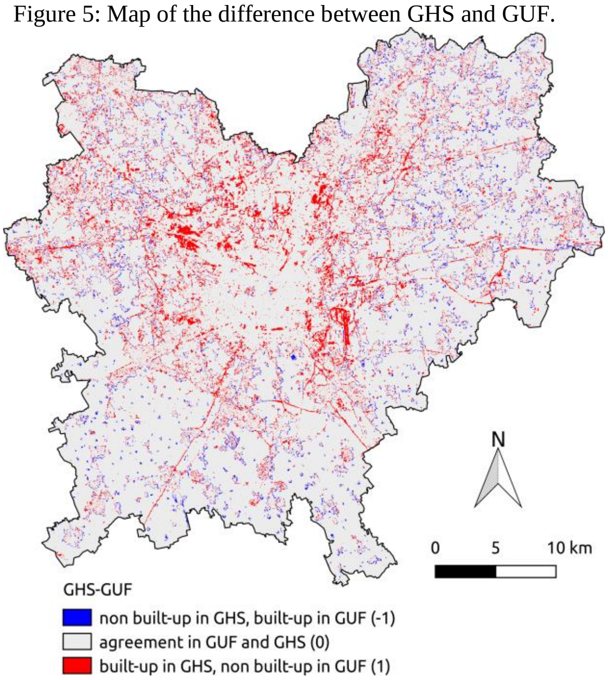
4.1 Comparison of GUF and GHS

This first comparison looks at the intrinsic similarity between GUF and GHS. GUF is first resampled to the same pixel size of GHS (38m). The confusion matrix, showing the number of pixels of agreement and disagreement between the two maps, the marginal agreements (MA) and the overall agreement (OA, in the bottom right cell) is shown in Table 4. Figure 5 represents a raster map showing the difference between GHS and GUF.

Table 4: Confusion matrix between GHS and GUF.

		GUF		
		non built-up (0)	**built-up (1)**	**MA**
GHS	**non built-up (0)**	1075015	51961	95.4%
	built-up (1)	162634	697598	81.1%
	MA	86.9%	93.1%	89.2%

Figure 5: Map of the difference between GHS and GUF.



There is good OA between the datasets, i.e. 89.2%. Most of the disagreement occurs in areas (colored in red in Figure 5) mapped as built-up in GHS and non built-up in GUF (difference equal to 1). Comparison with the official vector cartography of Milan reveals that 10.3% of these pixels, which can be also visually detected, correspond to roads (mainly located outside Milan city center and corresponding to the main roads), which means they are often wrongly mapped as built-up areas in GHS. The remaining 2.6% of the pixels correspond to the opposite situation, i.e. areas (colored in blue) mapped as non built-up in GHS and built-up in GUF (difference equal to -1), where there is no discernible pattern.

4.2 Comparison of GUF and GHS with UA

A similar procedure is undertaken to compare GUF and GHS with the UA reference dataset, reclassified according to the rules shown in Table 3 and rasterized at the same resolution of the two datasets. The confusion matrices between GUF and UA and GHS and UA are shown in Tables 5 and 6, respectively. In addition, Table 7 shows the proportions of pixel values for the differences UA-GUF and UA-GHS (-1, 0, 1).

The OAs with the UA reference source shown in Table 7 are quite satisfactory at 86.6% and 83.4% for GUF and GHS, respectively. It should be noted that for both GUF and GHS, most of the disagreement occurs in areas classified as built-up, which are instead considered non built-up in UA (difference value equal to -1). The UA non built-up classes that are most responsible for this disagreement with GUF and GHS are shown in Table 8; for each class, the difference in the disagreement between GUF and GHS is also shown.

Table 5: Confusion matrix between UA and GUF.

		UA		
		non built-up (0)	**built-up (1)**	**MA**
GUF	**non built-up (0)**	5476631	344643	94.1%
	built-up (1)	903861	2616220	74.3%
	MA	85.8%	88.4%	86.6%

Table 6: Confusion matrix between UA and GHS.

		UA		
		non built-up (0)	**built-up (1)**	**MA**
GHS	**non built-up (0)**	1077319	50018	95.6%
	built-up (1)	279706	580587	67.5%
	MA	79.4%	92.1%	83.4%

Table 7: Percentage of pixels resulting from the differences UA-GUF and UA-GHS.

	GUF/GHS built-up only	**agreement**	**UA built-up only**
UA-GUF	9.7%	86.6%	3.7%
UA-GHS	14.1%	83.4%	2.5%

Table 8: Percentages of UA non built-up classes generating the disagreement with GUF and GHS, and the difference between the GUF and GHS disagreement.

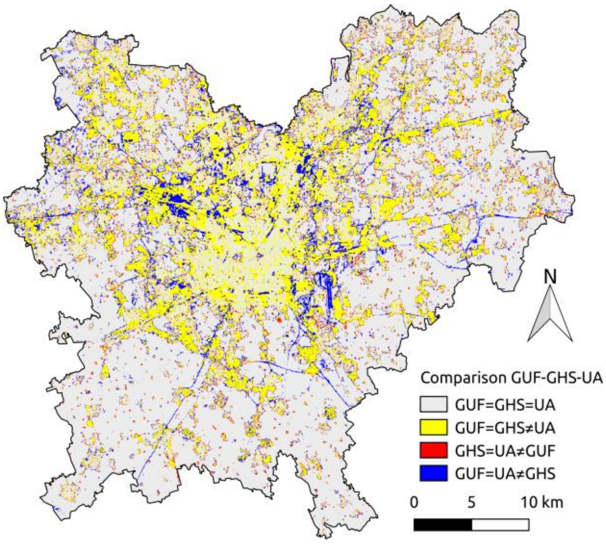
UA class	**GUF**	**GHS**	**GUF-GHS**
1.1.3	1.66%	0.84%	0.82%
1.2.2	49.64%	40.52%	9.12%
1.2.4	0.63%	1.37%	-0.74%
1.3.1	1.84%	2.04%	-0.20%
1.3.3	4.24%	3.87%	0.37%
1.3.4	4.66%	5.82%	-1.16%
1.4.1	6.30%	7.90%	-1.60%
1.4.2	9.82%	8.30%	1.52%
2	20.06%	28.15%	-8.09%
3	0.31%	0.38%	-0.07%
5	0.84%	0.83%	0.01%

For most UA non built-up classes, the percentages shown in Table 8 are very similar for GUF and GHS (differences smaller than 1.6%), which confirms the intrinsic similarity between the two. The UA classes responsible for most of the disagreement are 1.2.2 Road and rail network and associated land (49.64% for GUF and 40.52% for GHS), and 2 Agricultural areas, semi-natural areas and wetlands (20.06% for GUF and 28.15% for GHS). The disagreement related to class 1.2.2 is mainly due to the fact that in both GUF and GHS, roads and railways are often wrongly mapped as built-up areas, while the disagreement related to class 2 is mainly due to the transformation of agricultural areas into urban since the UA for Milan is only available for 2006 while GUF and GHS relate to 2012 and 2014, respectively. However, it can also be seen that for these two classes the disagreement between GUF and GHS is also larger (9% for class 1.2.2 and 8% for class 2).

Finally, Figure 6 shows an agreement/disagreement map combining GUF, GHS and UA. To perform this comparison, GUF and UA are resampled at the same resolution of GHS.

69.7% of pixels (colored in grey in Figure 6) have the same value in the three maps. For the pixels where there is disagreement in the three datasets (30.3%), most (19.5%, colored in yellow) corresponds to the case when GUF is equal to GHS but they differ from UA for the reasons explained above. The remaining disagreement, when UA agrees with only one of either GUF or GHS (represented by the red and blue areas in Figure 6), corresponds to only 10.8% of the pixels.

Figure 6: Agreement/disagreement map between GUF/GHS/UA.



4.3 Comparison of GUF and GHS with LUCAS

The GUF and GHS layers are then compared with the LUCAS reference dataset, which consists of 112 points (see Figure 4) and is reclassified according to the rules shown in Table 3. The corresponding confusion matrices are shown in Tables 9 and 10.

Table 9: Confusion matrix between LUCAS and GUF.

		LUCAS		
		non built-up (0)	**built-up (1)**	**MA**
GUF	**non built-up (0)**	72	1	98.6%
	built-up (1)	22	17	43.6%
	MA	76.6%	94.4%	79.5%

Table 10: Confusion matrix between LUCAS and GHS.

		LUCAS		
		non built-up (0)	**built-up (1)**	**MA**
GHS	**non built-up (0)**	63	0	100%
	built-up (1)	31	18	36.7%
	MA	67.0%	100%	72.3%

The OA with LUCAS data (79.5% and 72.3% for GUF and GHS, respectively) is still satisfactory, although lower than the UA. Almost all disagreement derives yet again from a misclassification of non built-up areas. In particular, most misclassified points (65.2% and 58.1% for GUF and GHS, respectively) belong to LUCAS class A20 Artificial non built-up areas, which are urban areas but not directly on buildings or man-made structures (e.g. they fall on roads, gardens or parking areas). In addition to the misclassification of roads in GHS, these errors are also caused by the resolution of GUF and GHS (12m and 38m, respectively), which is too coarse to capture this LULC change.

5 Conclusions

This study represents a preliminary assessment of two new global urban products for the city of Milan. The overall agreement with EU reference datasets ranged between 72.3% (for GHS and LUCAS) to 86.6% (for GUF and UA). GUF was shown to have higher agreement with both UA and LUCAS datasets. Some of the disagreement between GUF, GHS and UA can be attributed to the road network, which is often wrongly classified in GHS since built-up areas are defined as the presence of buildings with a vertical component. This study will be extended in the future to examine how well built-up areas are represented in GUF and GHS in other cities using additional reference datasets and methods, e.g. building footprints and high resolution digital elevation models.

Acknowledgements

The authors would like to thank DLR for providing the GUF data upon request.

References

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X. and Mills, J. (2015) Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7–27.

CIESIN - Center for International Earth Science Information Network (2004) Global Rural-Urban Mapping Project (GRUMP): Urban Extents. [Online] Available from: http://sedac.ciesin.columbia.edu/data/collection/grump-v1 [Accessed 2 February 2017].

EEA (2006) *The Thematic Accuracy of Corine Land Cover 2000 - Assessment using LUCAS*. European Environment Agency: Denmark, Copenhagen.

EEA (2011) *Mapping guide for a European Urban Atlas*. Available from: http://www.eea.europa.eu/data-and-maps/data/urban-atlas [Accessed 2 February 2017].

Eurostat (2015a) *LUCAS 2015 (Land Use / Cover Area Frame Survey)*. Technical reference document C1 Instructions for Surveyors. Eurostat: Strasbourg, France.

Eurostat (2015b) *LUCAS 2015 (Land Use / Cover Area Frame Survey)*. Technical reference document C3 Classification (Land cover & Land use). Eurostat: Strasbourg, France.

DLR. (2016) *Global Urban Footprint.* [Online] Available from: http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-9628/16557_read-40454 [Accessed 2 February 2017].

Pesaresi, M., Ferri, S., Ehrlich, D., Florczyk, A.J., Freire, S., Halkia, M., Julena, A., Kemper, T., Soille, P., and Syrris, V. (2016) *Operating procedure for the production of the global human settlement layer from Landsat data of the epochs 1975, 1990, 2000, and 2014*. JRC Technical Report.

See, L., Fritz, S., Perger, C., Van der Velde, M., Albrecht, F., Schill, C., McCallum, I., Schepaschenko, D., and Obersteiner, M. (2013) Urban Geo-wiki: A Crowdsourcing Tool for Improving Urban Land Cover. In C. S. editor, *Citizen E-Participation in Urban Governance: Crowdsourcing and Collaborative Creativity*, pp. 119-143. Hershey, PA: IGI Global.