An automatic extraction method of 250 meter global impervious surface

Jiongyi Huang (1), Xin Huang (1)(2)*, Jiayi Li (1)

[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, China
[2]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, China
Email: xhuang@whu.edu.cn; zjjerica@163.com

KEY WORDS: Global mapping, Impervious surface, Automatic sample collection

ABSTRACT: Impervious surface is a physical characteristic of urbanization. Delineating impervious surface at global scale helps to evaluate inter-regional urban development and provide supporting data for government decision-making. However, how to obtain accurate and widely distributed training samples is a great challenge in global mapping. In this paper, a training sample collection method based on existing thematic maps is used to automatically depict global impervious surface in 2019 with a spatial resolution of 250 m. The method consists of the following steps. First, we construct the training sample pool based on spatial and timing information from the current thematic maps. To improve the mapping performance in arid regions, some manually collected training samples are added to the sample pool as well. Then, an adaptive extraction process is implemented on local training sample pool by measuring the Kullback-Leibler divergence between the extracted samples and samples in the impervious surface pool. Next, we divide the world into $5^o \times 5^o$ grids and RF classification is conducted in each grid. Finally, we manually collect 38,496 test sample points worldwide for the validation. The F-score of the global mapping is 0.87, which indicates that our method overcomes the difficulty of obtaining training samples in global mapping, and obtains accurate and reliable mapping result. The validation for different continent and city level are also conducted. According to statistics, the area of global impervious surface in 2019 is 742,936 km^2. Nevertheless, the current global impervious surface is unevenly distributed. Among all continents, Asia has the largest impervious surface, accounting for 42% of the world impervious surface. North American ranks the second, with a percentage of 23%. The third is Europe which possesses 20% of the global impervious surface. The percentage in Africa is 7%, which is closed to the impervious surface in North America (6%). Oceania has the smallest impervious area, accounting for only 2%.

1. INTRODUCTION

Impervious surface mainly refers to lands covered by artificial construction, e.g. roads, buildings, parking lots, and so on (Arnold and Gibbons,1996). It describes the process of urbanization in physical characteristics and can be used as an important indicator for monitoring urban development. With the aggravation of urbanization, the impervious surface expands continuously. As a result, the urban surface runoff changes, which affects water cycle, surface temperature, urban ecosystem, and flood prevention (Li et al., 2016). Therefore, timely and accurate monitoring of impervious surface can provide necessary basic data for urbanization and urban environment research.

In recent years, several studies have attempted to map the global impervious surface. The main contribution includes: the 998m global mapping product GLC 2000 (Bartholomé and Belward, 2005); the 500m global mapping product MCD12Q1 Collecton6 (Sulla-Menashe et al., 2019); the 300m surface cover product CCI-LC (ESA, 2017); 30m global impervious layer product GAIA (Gong et al., 2020). With the progressive research, spatial and temporal resolution of mapping products have been greatly improved. However, there are great differences when estimating the area of global impervious surface in different studies, and there is no consensus on this issue. Therefore, it is of great significance to explore accurate and effective global mapping methods.

Compared with local scale, global mapping faces great challenges. Different climates, cultures and ecology between regions lead to discrepancy in the composition and patterns of impervious surface. The sub-regional supervised classification method effectively alleviates this problem by learning the characteristics of the local impervious surface. Therefore, how to obtain widely distributed and representative training samples is the key to global mapping. To obtain training samples by visual interpretation can ensure the accuracy of the samples, but consumes huge manpower and material resources. Thus, that is not suitable for global mapping. The historical thematic map contains a lot of land cover information, which is a potential source of training samples (Radoux et al., 2014).

Therefore, in this paper, we use a sample cleaning method that automatically extracts and retains reliable information from historical thematic maps. Additionally, it is the first time to map the 2019 global impervious surface at 250-m resolution. The method mainly consists of the following three parts:

1. Construct a training sample pool based on the historical thematic maps and enhance samples in arid areas.
2. Adaptively extract local training samples according to KL divergence.
3. Implement RF classification in 5°×5° grids.

Finally, the mapping result is evaluated by manually collected test samples. Analysis of global impervious surface distribution is also given.

2. DATA & METHOD

2.1 Construction of training sample pool

2.1.1 **Impervious surface sample pool:** With similar definition of impervious surface, four products, i.e. GlobCover, CCI-LC,MCD12Q1.v5 and MCD12Q1.v6, are selected as the sources

to construct training sample pool. All products are uniformly converted to 250m resolution in WGS84. Detailed information of the four products is shown in Table 1.

Table 1. The data source of training samples.

Usage	Data	Resolution	Available year	Source
Construction of training sample pool	The MODIS Land Cover Type product Collection5（MCD12Q1.v005）	500 m	2001-2013	(Friedl et al., 2010)
	The MODIS Land Cover Type product Collection6（MCD12Q1.v006）	500 m	2001-2018	(Sulla-Menashe et al., 2019)
	GlobCover 2005 / GlobCover 2009（GlobCover）	300 m	2005, 2009	(Defourny et al., 2006) ; (Bontemps et al., 2011)
	ESA Climate Change Initiative land cover maps（CCI-LC）	300 m	1992-2018	(ESA, 2017)
Auxiliary data	GHS_POP_GPW4_GLOBE_R2015A (GHS-POP)	250m	1975, 1990, 2000,2015	https://ghsl.jrc.ec.europa.eu/ghs_pop.php
	VIIRS Nighttime Day/Night Band Composites Version 1	15 arc sec	2012-2019	https://ngdc.noaa.gov

Considering of the data availability, only historical thematic maps from 2001-2013 is selected as the data source for training sample pool. Sample cleaning process depends on the intra-annual statistical information to screen out the potential training samples of impervious surface. On this basis, morphological information and statistical information in time series are used to select reliable samples. Figure 1 shows the general route for constructing training sample pool of impervious surface.

With the urbanization process, impervious surface is expanding. In order to keep adequate widely-distributed samples, the probability P_{imp} of impervious surface in each pixel is calculated based on the 2013 historical thematic maps. Only those pixels with P_{imp}>50% were reserved as potential samples, on the basis of which training sample pool is generated. Additionally, similar operation is implemented on other years to attain potential samples in time series, which assists sample cleaning.

Population is an important indicator of urbanization, which is positively correlated with the density of artificial surface. The consistency concentrates in impervious surface with large population. On the contrary, the inconsistency concentrates in impervious surface with small population. To ensure the diversity of training samples, we divide the potential training samples into two parts according to population density derived from GHS-POP. Then, sample cleaning process for each part is conducted separately.

Each potential training sample can be preserved if it satisfies either of the neighborhood or time series constraint. The neighborhood constraint is that all pixels in 3×3 neighborhood around are potential training samples. The time series constraint refers that during 2001 to 2013, most of the time the pixel is potential training sample. The samples preserved compose the initial training sample pool of impervious surface.

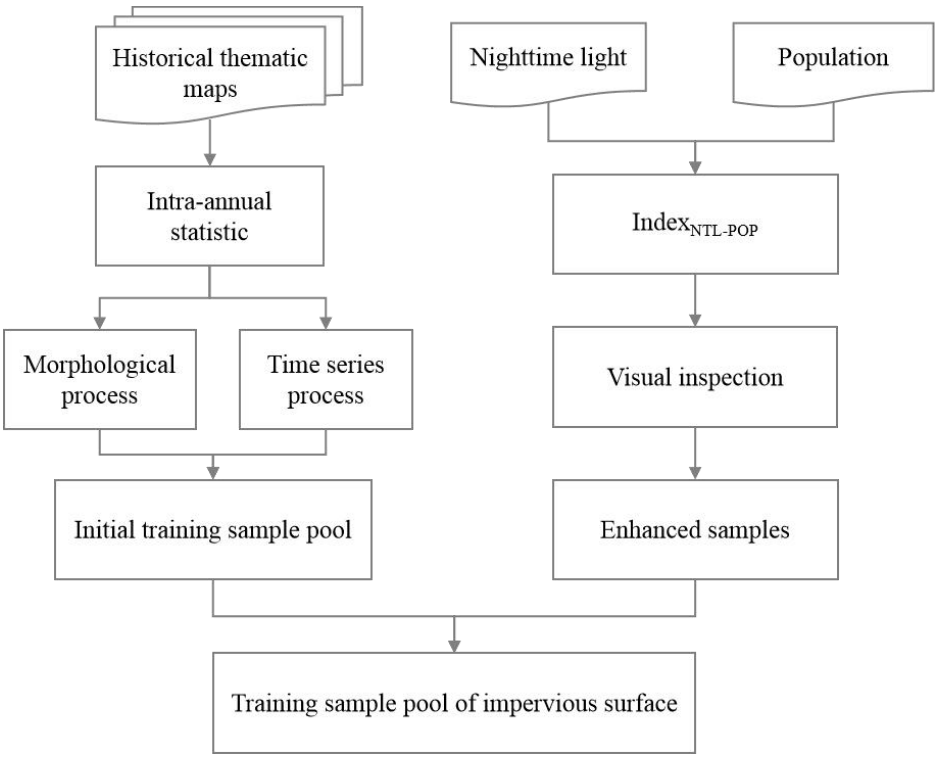

Figure 1. The construction of impervious surface training sample pool

2.1.2 **Sample enhancement in semi-arid and arid area:** The spectral information between the impervious surface and the surrounding bare land is highly confused. Thus, the mapping performance in arid and semi-arid areas is relatively poor (Gong et al., 2020). It means that the quantity of training samples obtained in arid and semi-arid areas by method in 2.1.1 is inadequate. In order to improve the mapping performance, it is necessary to supplement and enhance the training samples in those area. Population and night light data reflect the process of urbanization from the social and economic aspects, which effectively reveal the distribution of impervious surface. The 250m GHS-POP in 2015 is used as the population data and the VIIRS Nighttime Day/Night Band Composites Version 1 in 2019 is selected as the nighttime light data. All data is converted to 250m resolution in WGS84, and then fused to calculate the index:

$$Index_{NTL-POP} = 0.5 \times (NTL + POP) \tag{1}$$

where $Index_{NTL-POP}$ refers to the nighttime light and population index. NTL is the linearly normalized nighttime light data after removing the outliers. POP is the linearly normalized population data after removing the outliers.

The enhanced samples are attained through visual interpretation based on Landsat with the assist of nighttime light and population index. Finally, the enhanced samples are added to the training sample pool.

2.1.3 **The construction of pervious surface sample pool:** Compared with impervious surface, pervious surface is widely distributed. Therefore, through simple morphological corrosion, the training sample pool of pervious surface is constructed.

Firstly, the pixels that are always labeled as pervious surface in all historical thematic maps from 2001 to 2013, are preserved as potential samples. Then, the potential samples are morphologically corroded with a radius of 2 km to obtain the training sample pool of pervious surface.

2.2 Adaptive extraction of local training samples

The mapping performance of supervised classification heavily depends on the quality of training samples. The construction of training sample pool eliminates errors and obtain reliable samples. However, those samples are not completely independent in spatial distribution and lead to redundant information. It is very important to determine the optimal extraction ratio of samples to obtain representative and low redundancy training samples. To solve this problem, we propose a local adaptive automatic sample extraction method.

Firstly, the world is divided into 5º×5º local grids. In each grid, the similarity between the impervious surface samples randomly extracted at different ratio and the samples in the training sample pool is calculated. The KL divergence (Kullback and Leibler, 1951) is used to measure the similarity:

$$D_{KL} = \sum_{i=1}^{N} p(x_i) \ln \frac{p(x_i)}{q(x_i)} \tag{2}$$

where D_{KL} refers to the KL divergence; N is the number of unique feature values calculated from the impervious sample pool after PCA dimension reduction; $i \in [1, N]$; x_i is the feature value; $p(x_i)$ refers to the frequency of x_i in the extracted impervious samples; and $q(x_i)$ refers to the frequency of x_i in the impervious sample pool.

Then, nonlinear function is used to fit the KL divergence at different sampling ratio:

$$y = ax^b + c \tag{3}$$

where x is the sampling ratio; y is the KL divergence; a, b, c are coefficients.

When the first derivative of the fitting function is -1, the sampling ratio is adopted as the best ratio. Random extraction from training pool is conducted at this ratio to attain the training samples in the grid.

2.3 RF classification

Classification features are derived from MODIS surface reflectance data and surface temperature data (https://ladsweb.modaps.eosdis.nasa.gov). See Table 2 for details. The data are uniformly converted to WGS84 and resampled to 250m resolution.

The 40th quantile is used to composite seasonal spectral features. Additionally, the normalized spectral indexes are calculated between any two spectral bands in each season. In total, there are 21 indexes per season. For nighttime surface temperature, the average value for each season are

adopted. In total, there are 116 features.
RF classification is conducted in each 5º×5º grid to attain 2019 global impervious surface map.

Table 2. The source of features.

Data	Description	Usage
MOD09Q1	Terra Surface Reflectance 8-Day Global 250m	To composite seasonal spectral band B1 and B2
MOD09A1	Terra Surface Reflectance 8-Day Global 500m	To composite seasonal spectral band B3 to B7
MOD11A2	Terra Land Surface Temperature and Emissivity 8-Day Global 1km	To composite seasonal average nighttime surface temperature

3. RESULT

3.1 Mapping result

The extraction result of 2019 impervious surface is shown at regional and urban scales in Fig.2. At regional scale, four districts with dense impervious surface, i.e. North America, South America, Europe and East Asia, are presented. At urban scale, the impervious surface of Beijing, New Delhi, New York, São Paulo, Cairo and Melbourne are superimposed on Google high resolution images. It can be found that the extracted impervious surface is highly consistent with Google images.

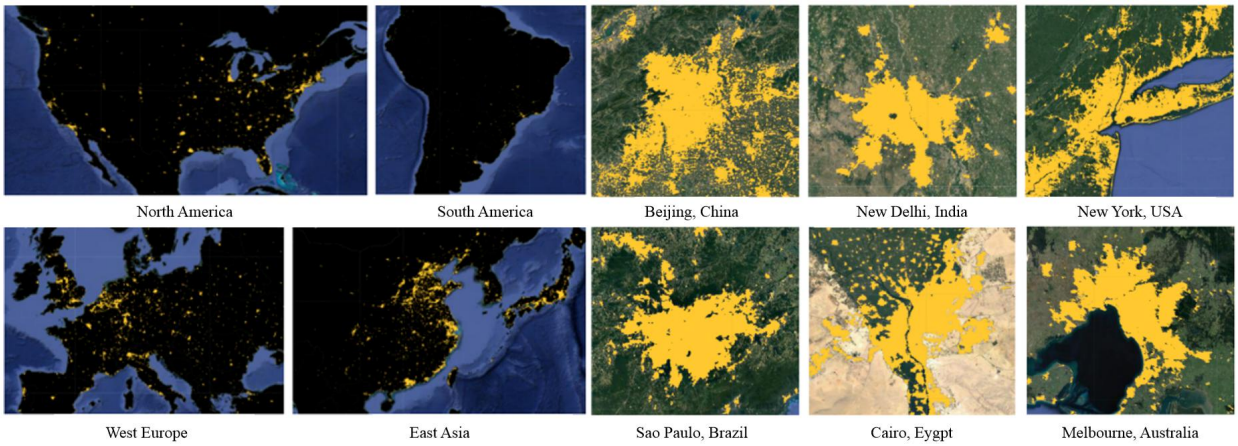

Figure 2. Global impervious surface

Impervious surface is closely related to urbanization process. In order to validate the extraction result, we divide the global cities into three level, i.e. L1, L2 and L3 according to their population. Then, ten cities are randomly selected from each level and in total, there are 30 cities to generate test samples. The distribution information of validation cities is shown in Table 3. Test samples with the size of 38,496 are labeled through visual interpretation based on the 30m Landsat images.
Producer's accuracy (*PA*), user's accuracy (*UA*), F-score and overall accuracy (*OA*) are regarded as accuracy indicators in this paper. The accuracy assessment is conducted at global, city level and continent scales respectively and shown in Figure 3.

The F-score of the global result is 0.87, which indicates a reliable mapping performance. At city level, L2 has the highest F-score, followed by L1 and L3. The *UA* of L3 is obviously higher than its *PA*, which caused by omission. At continent scale, Oceania has the highest F-score (0.93), while Europe has the lowest (0.81). The F-score of North America, South America and Africa are similar.

Table 3. Cities for validation.

City Level	L1		L2			L3		
Population	≥10 million		1~10 million			<1 million		
City	Country	Continent	City	Country	Continent	City	Country	Contine
1 Delhi	India	Asia	1 Shiraz	Iran	Asia	1 Vancouver	Canada	North America
2 Beijing	China	Asia	2 Pretoria	South Africa	Africa	2 Gold coast	Australia	Oceania
3 Lagos	Nigeria	Africa	3 Cincinnati	United States	North America	3 Itajai	Brazil	South America
4 Lima	Peru	South America	4 Sydney	Australia	Oceania	4 Tyre	Lebanon	Asia
5 Jakarta	Indonesia	Asia	5 Milan	Italy	Europe	5 Kindia	Guinea	Africa
6 Moscow	Russia	Europe	6 Singapore	Singapore	Asia	6 Kansas	United States	North America
7 Istanbul	Turkey	Asia	7 Kampala	Uganda	Africa	12 Oral	Kazakhstan	Asia
8 Buenos Aires	Argentina	South America	8 Chicago	United States	North America	13 Hanzhong	China	Asia
5 Mumbai	India	Asia	9 Suwon	Korea	Asia	14 Villaverde	Spain	Europe
5 Bogota	Colombia	South America	10 Ufa	Russia	Europe	15 Iquitos	Peru	South America

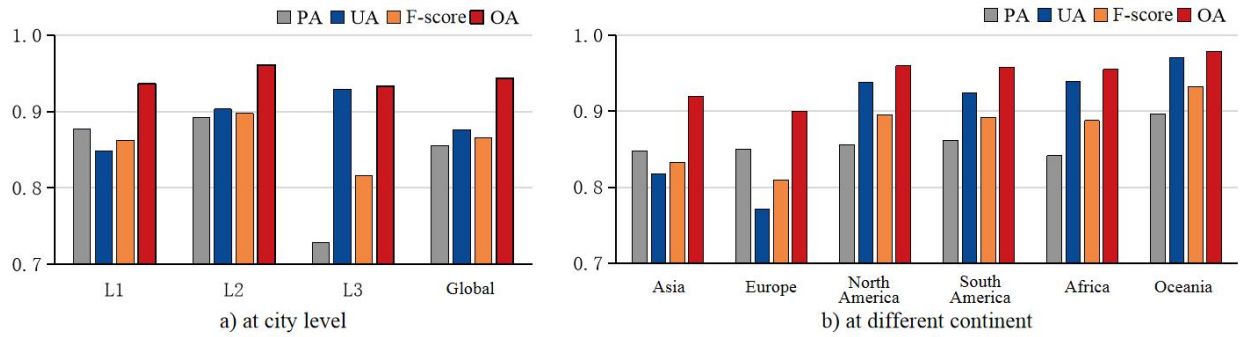

Figure 3. Accuracy assessment.

3.2 The distribution of global impervious surface

According to statistics, the area of impervious surface of 2019 is 742,936 km^2. However, the current global impervious surface is unevenly distributed. Among all continents, Asia has the largest impervious surface, accounting for 42% of the world impervious surface. North American ranks the second, with a percentage of 23%. The third is Europe which possesses 20% of the global impervious surface. The percentage in Africa is 7%, which is closed to the impervious surface in North America (6%). Oceania has the smallest impervious area, accounting for only 2%.

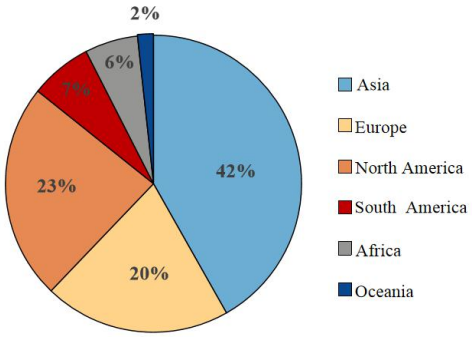

Figure 4. The distribution of impervious area

4. CONCLUSION

In this paper, a training sample collection method based on existing thematic maps is used to automatically depict global impervious surface of 2019 with a spatial resolution of 250 m.

To validate the mapping result, visual inspection and accuracy assessment are conducted. Through visual inspection, it is found that the mapping result is high consistent with Landsat images. In accuracy assessment, 38,496 test samples collected from 30 cities through visual interpretation are used. The F-score of the global mapping is 0.87, which indicates that our method overcomes the difficulty of obtaining training samples in global mapping, and obtains accurate and reliable mapping result. The validation for city level and different continent are also conducted. At city level, L2 has the highest F-score, followed by L1 and L3. At continent scale, Oceania has the highest F-score (0.93), while Europe has the lowest (0.81).

Furthermore, we analyse the distribution of global impervious surface. According to our statistics, the area of impervious surface in 2019 is 742,936 km^2. However, the current global impervious surface is unevenly distributed. The top three continents that possess the most impervious surface are Asia (42%), North American (23%) and Europe (20%).

Reference:

Arnold Jr, C. L., & Gibbons, C. J. (1996). Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American planning Association*, *62*(2), pp. 243-258.

Bartholomé, E., Belward, A. S. 2005. GLC2000: a new approach to global land cover mapping from Earth observatio data. *International Journal of Remote Sensing*, *26(9)*, pp. 1959-1977.

Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., Perez, J., R. 2011. GLOBCOVER 2009 Products description and validation report. *Retrieved Sep 30, 2020, from: https://epic.awi.de/id/eprint/31014/16/GLOBCOVER2009_Validation_Report_2-2.pdf.*

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., Leroy, M. 2006. Globcover: a 300 m global land cover product for 2005 using ENVISAT MERIS time series. *In: Proceedings of the Recent Advances in Quantitative Remote Sensing Symposium: 25–29 September 2006. Valencia,* pp. 538-542.

ESA. 2017. Land Cover CCI Product User Guide Version 2.0. *Retrieved Sep 30, 2020,from:http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.*

Friedl, M., A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, *114(1)*, pp. 168-182.

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., Zhou, Y. 2020. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment, 236(1)*, pp. 111510.

Kullback, S., Leibler, R., A. 1951. On information and sufficiency. *The annals of mathematical statistics, 22(1)*, pp. 79-86.

Li, D., Luo, H., Shao, Z. 2016. Review of Impervious Surface Mapping Using Remote Sensing Technology and Its Application. *Geomatics and Information Science of Wuhan University, 41(5)*, pp.569-577.

Radoux, J., Lamarche, C., Van Bogaert, E., Bontemps, S., Brockmann, C., Defourny, P. 2014. Automated training sample extraction for global land cover mapping. *Remote Sensing, 6(5)*, pp. 3965-3987.

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., & Friedl, M. A. 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, *222(1)*, pp. 183-194.