Response to comments

Paper #: essd-2019-200

Title: Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform

Journal: Earth System Science Data

Reviewer #2

The paper describes a new impervious surface dataset developed by combining several remote sensing instrumentation at 30m resolution. As described in the introduction, several datasets describing impervious studies exist at a global scale. The strength of this paper is in my opinion the use of multi-sensor information and the use of an open-source platform the generate these maps (Google Earth Engine). Furthermore, a relatively good accuracy of the map is achieved compared to three other impervious surface products. The paper is very well written and is easy to follow. The introduction also gives a very good overview of current existing literature. The paper is very mature and contains all information one would expect for this kind of work. Most of the comments that popped in my mind while reading the paper were assessed later in the manuscript. As such, for me only minor revisions are necessary. I describe some comments below.

Great thanks for the comment. The manuscript has been improved according to your and other reviewers' comments.

General comments

- Training points are achieved from Globeland30 and are not independent based on independent experts (which is done for the validation data). Several checks are done on the training data, but you are still using a derived product with errors to train your model. In the discussion, this problem is assessed (section 5.2). However, I would state this more clear that the training sample can contain errors in the material and methods section and potentially move the discussion to the material and methods section or refer in the material and methods section that this problem will be assessed later.

Great thanks for the comment. According to the suggestion, we added a paragraph to explain that the training sample may contain some error, because they were collected from Globeland30. The detailed explanations are listed in the Method Section "3 Collection of global training samples" as: "Although a series of rules were applied to guarantee the high confidence of global training samples, due to the classification error in GlobeLand30 and the temporal interval between GlobeLand30 and input imagery, the global training dataset inevitably contained some erroneous samples. The relationship between the percentage of the erroneous samples and the mapping accuracy of impervious surface was analyzed in the Discussion section 6.1, and the results indicated that the error in the training samples had little effect on the mapping accuracy."

- Only homogeneous training points from Globeland30 are included. Therefore, the training points are always clear impervious surfaces leading to only clear impervious surfaces to be classified later. Don't you underestimate the total amount of impervious surfaces then in your final product? How does the total % of impervious surface compare to Globeland30, GLC and NAUCI for the globe? This can maybe be compared to the results presented in figure 5.

Great thanks for the comment. Although only homogeneous training points from GlobeLand30 are included, the accuracy assessment in the section 5.4 has demonstrated that the proposed method achieved lower omission error than other products (NUACI-2015, FROM_GLC-2015, GHSL-2015, GlobeLand30-2010 and HBASE-2010).

From the perspective of the value of the user's accuracy for impervious surfaces, the MSMT_RF method performed better than the other impervious surface products (meaning lower omission error) achieving the accuracy of 0.932, especially in the cropland-prevalent and vegetation-prevalent impervious landscapes (such as: Bangkok, Winnipeg, Xi'an…). Specifically, NUACI-2015 had the lowest user's accuracy of 0.562 and this might be due to its poor performance over small impervious surfaces (Sun et al., 2019b). FROM_GLC-2015 had a similar performance with the MSMT_RF method for big cities (such as New York, Moscow and Johannesburg), but its accuracy decreased sharply over 'small-city' regions (such as Lhasa, Winnipeg). The performance of GHSL-2015 was closest to the MSMT-2015 over most validation regions, but it also missed the fragmented objects (villages and roads) over cropland-prevalent city (such as Bangkok and Winnipeg). As the minimum mapping unit of GlobeLand30 was a 4×4-pixel area, many rural impervious surfaces were ignored in these validation regions, which caused large omission errors of 23.9%. Finally, partly due to the 5 years' interval between the HBASE-2010 and validation samples, HBASE-2010 also suffered the omission error of 12.5%.

In addition, we analyzed the spatial variations of six global impervious products at the spatial resolution of 0.05°, the figure also indicated the proposed method gave consistent mapping with other datasets.

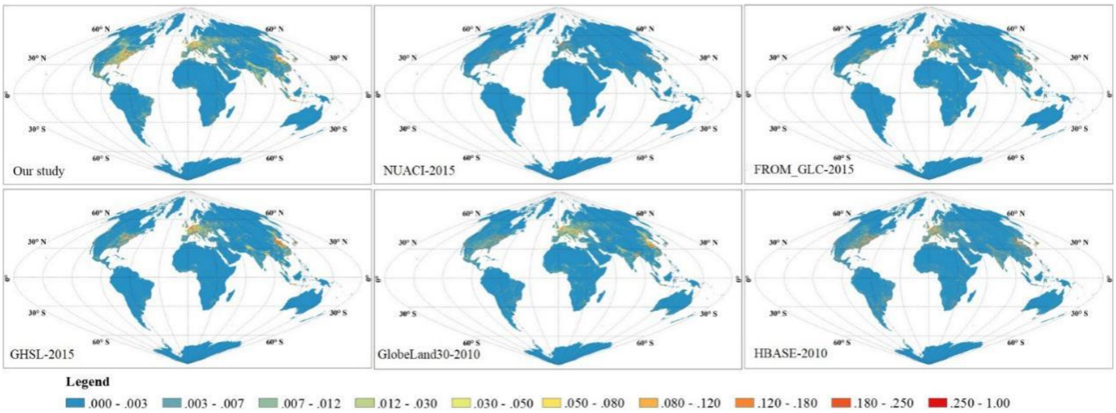

Figure 7: The spatial patterns of six global 30-m impervious products after aggregating to the resolution of 0.05°.

Finally, the total impervious areas of NUACI-2015, FROM_GLC-2015, GHSL-2015, GlobeLand30-2010 and HBASE-2010 were 49.53%, 54.67%, 78.55%, 67.76% and 97.24% that of the MSMT-2015 (our study), respectively.

- The validation points are retrieved from 12 regions. How representative are these regions for the globe? Since impervious areas might have very different characteristics depending on the region. For Africa for example, the validation points are achieved for two big cities only.

Great thanks for the comment. Yes, we agree that impervious areas might have very different characteristics depending on the region. However, it was a time-consuming task to collect validation samples over the globe. According to the comment and suggestion from your and other reviewers, to make the validation regions more representative, we re-selected these regions by combining the impervious landscapes, for example, desert landscape urban areas such as Phoenix city, vegetation prevalent cities such as New York City. Specifically, the section 2.3 "validation samples" was changed as:

"To quantitatively assess the performance of the global impervious surface datasets, fifteen validation regions, covering different continents and various urban landscapes (the bare soil prevalent cities: Phoenix (PNX), Madrid (MDR), Riyadh (RYH), Niamey (NIM), Johannesburg (JHB), Ntuman (NTU) and Lhasa (LHS), vegetation prevalent cities: New York (NYK), Manaus (MNS), Moscow (MSC), San Paulo (SPL) and Melbourne (MBN), as well as cropland prevalent cities: Winnipeg (WIP), Bangkok (BGK) and Xi'an (XAN)), were selected (Fig. 1). For each validation region, 600-1000 samples were randomly generated using the stratified random sampling strategy (Bai et al., 2015)."

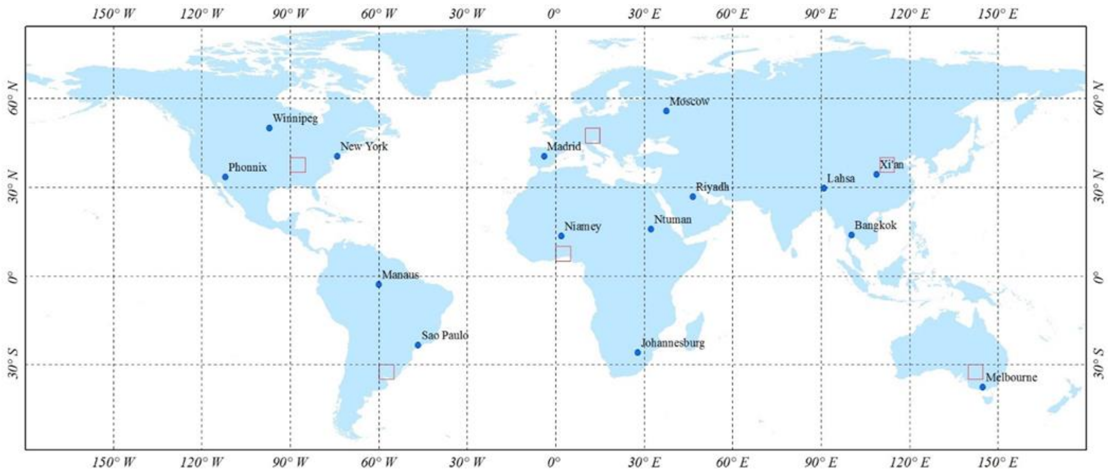

Figure 1: The spatial distribution of the fifteen validation regions (blue) corresponding to regions of different impervious landscapes on different continents, together with the six 5°×5° validation regions (red) used to measure the variable importance.

Specific comments

- Globeland30 data from 2010 is used as training data. How do you account for changes in urban areas between 2010-2015? You state that there is an irreversible state from non-impervious to impervious surfaces, but this could mean that some non-impervious surfaces in 2010 have now changed to impervious in 2015.

Great thanks for the comment. As there was temporal interval of 5years between GlobaLand30 and input imagery, we assumed that the process of transforming non-impervious surfaces into impervious surfaces was irreversible during the period 2010 to 2015, meaning that the global impervious training samples derived from GlobeLand30-2010 could also be used to represent the situation in 2015.

However, it was possible that the non-impervious pixels in 2010 were transformed into impervious surfaces in 2015. Therefore, some non-impervious training samples in Globeland30-2010 would be impervious surface in 2015. In order to mitigate the problem, the EANTLI light data was used to remove these changed samples. For example, the non-impervious training samples with high EANTIL value ($EANTLI = \frac{1+(NTL_{norm}-EVI)}{1-(NTL_{norm}-EVI)} \times NTL$) in 2015 would be removed. Detailed explanation was revised in the section of "3. Collection of global training samples” as:

“As for the non-impervious pixels, there was usually a negative correlation between non-impervious surfaces and EANTLI values, and the non-impervious surface samples turned into impervious surface would have high EANTLI values in 2015, so if the cumulative probability of a candidate non-impervious point in CanTPS_Imp was greater than the top 20th percentile of the cumulative probability of all candidate non-impervious points (the threshold being based on the overall accuracy of 80.33% for GlobeLand30-2010 and a little potential conversion samples), the candidate non-impervious point was also removed.”

- Add to table 1 that the 15 + 85 percentile are used for the Landsat bands and vegetation indices.

Great thanks for the comment. It has been added as:

Table 1. Training features for global impervious surface mapping.

Data	**Features**	**References**
LandSat-8 OLI	Reflectance: 15th and 85th percentiles of Blue, Green, Red, NIR, SWIR1 and SWIR2	Liu et al. (2018)
	Normalized indices: 15th and 85th percentiles of NDVI, NDWI and NDBI	
	Textural variables: variance, dissimilarity and entropy of the NIR	Chen et al. (2016)
Sentinel-1 SAR	Annual statistics: mean and standard deviation of VV and VH	Sun et al. (2019b)
	Textural features: dissimilarity, variance and entropy of VV and VH	Zhang et al. (2014b)
DEM	Elevation, slope and aspect	Clarke et al. (1997)

- Line 265, remove ‘the’

Great thanks for the comment. It has been removed as:

“In addition, as Sun et al. (2017) explained that the growing season was the best time for impervious surface mapping over temperate continental climate zones...”