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 #1

This manuscript introduced a global scale impervious surface map generated with multi-source remote sensing datasets, and comparative analysis suggested that the developed map outperformed the state-of-art land cover products. Despite producing a global impervious surface map using manifold datasets is an important contribution to the global land cover dataset, a major revision suggestion may be given from my side.

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

1. The review of impervious surface datasets should be further improved. Here I recommend a few examples: Global Man-made Impervious Surface (GMIS) Dataset, Copernicus land monitoring surface – high resolution layer imperviousness (although this dataset only covers Europe continent, it can be used as training and validation sample source), NLCD imperviousness products, Global Human Built-up And Settlement Extent (HBASE) Dataset.

Thanks for the comment. To make the cross-comparisons more comprehensive, two global impervious products (HBASE and GHSL) have been added in the Section 2 Datasets as (as GMIS and HBASE were companion datasets and GMIS provided continual imperviousness products, only HBASE was included):

The HBASE (Human Built-up and Settlement Extent) dataset was the first global 30-m dataset of man-made impervious cover derived from the Global Land Survey (GLS) Landsat data for 2010 (HBASE-2010) (https://sedac.ciesin.columbia.edu/data/set/ulandsat-hbase-v1). It was produced by combining meter-resolution training data (exceeding 20 millions), Open Street Map, VIIRS NTL, GLS Landsat SR and MODIS NDVI products, and achieved a kappa coefficient of 0.91 using scene-level cross validation in Europe (Wang et al., 2017a; Wang et al., 2017b).

The GHSL (Global Human Settlement Layer), a global information baseline describing the spatial evolution of the human settlements in the past 40 years, was developed by using symbolic machine learning model trained by the collected high-resolution samples, multi-temporal Landsat imagery in the epochs 1975, 1990, 2000, and 2015 (Florczyk et al., 2019). In this study, the GHSL impervious surface map at 30-m for 2015 (GHSL-2015) (https://ghsl.jrc.ec.europa.eu/download.php) was employed for comparison analysis, which achieved an overall accuracy of 96.28% and kappa coefficient of 0.3233 validated using Land Use/Cover Area frame Survey (LUCAS) reference data (Pesaresi et al., 2016).

Further, based on the suggestion, we combined two regional impervious products to interpret the validation samples as:

"**To ensure the reliability of each validation sample, two prior impervious products, including NLCD impervious products (Homer et al., 2015) and Copernicus land monitoring surface – high resolution layer imperviousness (Langanke et al., 2016) which were validated**

to achieve high overall, user's and product's accuracies exceeding 82% and 90% respectively, were overlaid to the high-resolution remote sensing imagery. In addition, the location of each sample was moved to the center of the relevant surface object (building, road, etc.) because of the greater spectral mixing effect and uncertainty at the boundary of the objects."

Lastly, the HBASE-2010 and GHSL-2015 global impervious products have been added in the Section 5.3 and 5.4 (accuracy assessment is detailedly answered in Question 3):

5.3 Spatial variations of global impervious products

To quantitatively analyse the spatial agreement between the MSMT_RF-based impervious surface map and the five existing products (GlobeLand30-2010, FROM_FLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010), all global 30-m impervious surface maps were first aggregated to a resolution of 0.05°. Fig. 7 illustrated the spatial patterns of six global impervious products, intuitively, the NUACI-2015 had lower impervious areas than other products especially in the North-America and Europe, and the GHSL-2015, GlobeLand30-2010 and our product (MSMT-2015 map) had greater spatial agreement because the impervious areas of FROM_GLC-2015 and HBASE-2010 in the China were obviously smaller. Further, our product had higher impervious areas over North-America especially over the Canada than other products because the proposed method had greater ability to identify small and fragmented impervious objects such as villages and roads which was been demonstrated in the following section 5.4 over Winnipeg region.

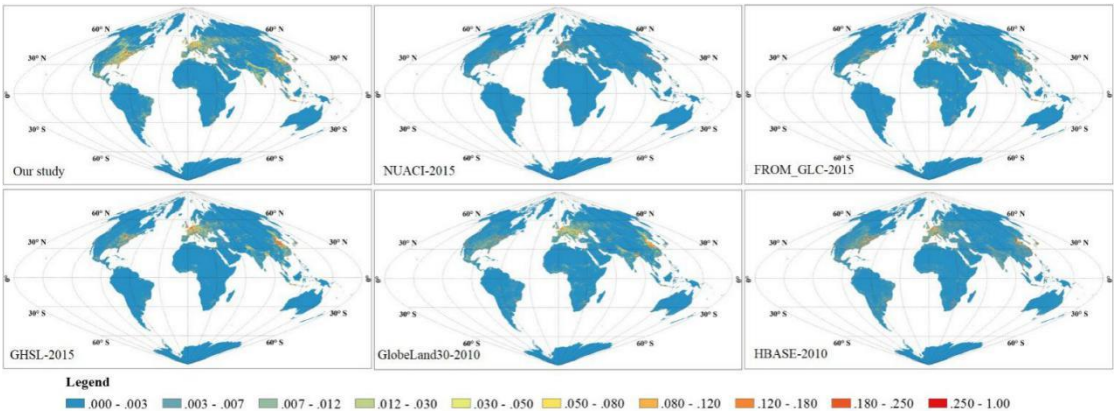


Figure 7: The spatial patterns of six global 30-m impervious products after aggregating to the resolution of 0.05 °.

Scatter plots of the five products against the MSMT-2015 impervious map were then made, as illustrated in Fig. 8. The results indicate that there was a greater agreement between the MSMT-2015 map and GHSL-2015 (R^2=0.783, RMSE= 0.038 and slope=0.921) than for other products. Specifically, as NUACI-2015 has been demonstrated to miss some small, fragmented villages and roads (Sun et al., 2019b), the slope of the regression line was less than 1.0 and R2 was the low value of 0.655 in this case. The scatter plot between FROM_GLC-2015 and MSMT-2015 indicated that there was a high degree of agreement between FROM_GLC-2015 and MSMT-2015 results in 'high-fraction' regions (close to 1:1) but FROM_GLC-2015 was obviously lower than MSMT-2015 over 'low fraction' regions, so the slope of the regression line for FROM_GLC-2015 was also less than 1. The main differences between the GlobeLand30 and the MSMT_RF-based maps were due to the temporal interval of 5 years and the limitations of the minimum 4×4 mapping unit for GlobeLand30-2010 (Chen et al., 2015), so the scatters were mainly concentrated below the 1:1 line. The HBASE-2010 had higher impervious areas than MSMT-2015 especially for the 'high-fraction' regions, but the following section demonstrated that it suffered the over-estimation problem, so the regression slope was higher than 1 and R^2 only reached the

value of 0.730. In addition, to intuitively understand the stability of regression model, the error bars, calculated as the standard deviation of reference data with the fitted results, were added to the scatter plots. It could be found that the error bars increased first and then stabilized as the impervious fraction increased.

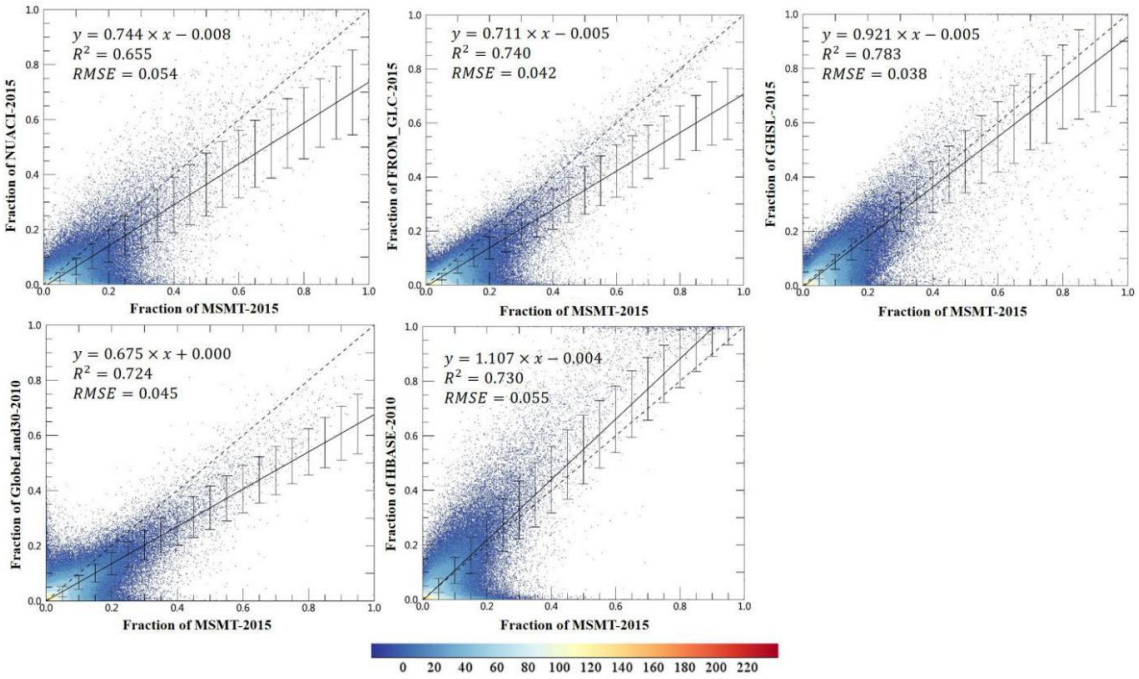


Figure 8: Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial grid of 0.05°×0.05°. The error bars were the standard deviation between reference datasets with fitted results.

2. The scientific importance of your dataset should be enhanced. Demonstration of multiple dataset contributions to land cover classification was not straightforward. Data (spectral or radar) characteristics on different land covers (e.g. vegetation, impervious surface and bare soil) should be revealed in detail.

Great thanks for the comment. Based on the comment and latter suggestion, the detailed responses of different land-cover types over optical and SAR imagery have been added in the section 5.1 "The importance of multi-source and multi-temporal features" as:

Because of the spectral heterogeneity of impervious surfaces, it is very difficult to accurately map impervious surfaces using only optical remote sensing imagery (Zhang et al., 2014b). Although a few studies have demonstrated that the integration of multi-source and multi-temporal information can improve the mapping accuracy, these studies mainly focused on regions with high impervious surface density (Zhang et al., 2014b; Zhu et al., 2012). At present, global impervious surface maps are still produced by optical imagery alone or by using a combination of optical and DMSP-OLS or VIIRS NTL imagery (Huang et al., 2016; Liu et al., 2018; Schneider et al., 2010). This is the first study that developed the global 30-m impervious surface map using multi-source and multi-temporal imagery. To quantitatively demonstrate the need for using multi-source, multi-temporal information, we randomly selected six 5°×5° regions (red rectangles in Fig. 1) from six different continents and then calculated the importance of the training features using the RF model. Specifically, the RF model computed the average

increase in the mean square error by permuting out-of-bag data for a variable while keeping all the other variables constant, thus measuring the variable's importance (Pflugmacher et al., 2014). Training features that had a high importance were the drivers of the model decision and their values had a significant impact on the output values.

The importance of all 37 training features for the six regions is illustrated in Fig.3. These results indicate that the Sentinel-1 SAR features (VV and VH) had the greatest contribution to the final decision in most regions because SAR images can provide information about the structure and dielectric properties of the surface materials. Next in importance were the 15th percentile of Landsat SR in the blue, green, red and SWIR2 bands and the corresponding NDVI and NDWI indices, as well as the texture variance and dissimilarity for Sentinel-1 SAR. The importance of these feature was close to or exceeded 5% in most cases. Then came the 85th percentile of Landsat SR in the NIR and SWIR1 bands as well as the SAR texture features, with a mean importance about 3%.

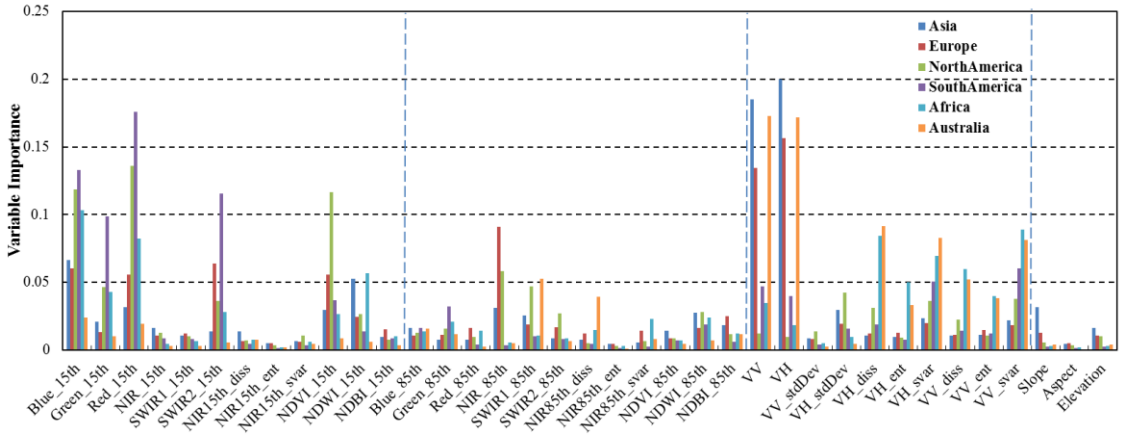


Figure 3: The importance of the input features derived from the random forest model using the training samples in six continental regions.

To intuitively understand the characteristics of different land-cover types on optical and SAR imagery, two regions (the vegetation-prevalent region of Asia and bare soil-prevalent semi-arid region of Australia) were selected for comparison analysis. Fig. 4 illustrated the reflectance and backscatter statistics (mean and standard deviation) of five typical land-cover types (cropland, vegetation, bare soil, impervious surfaces and water body). Obviously, impervious surfaces had highest backscattered signals in VV because of the high dielectric properties of the building materials, the unique geometry of manmade features, and the special radar echo properties of artificial structures, followed by the vegetation land-cover types. Further, since only a small part of the polarized signals (vertical turning horizontal) were returned to the sensor, the VH was significantly lower than VV but the ranking orders of different land-cover types in VH was similar to that of VV. Due to the complicated construction and heterogeneity of the impervious surfaces, the impervious surfaces also had highest standard deviation, for example, the urban central usually reflected higher VV and VH signals than the village buildings. If only Sentinel-1 SAR features were used to identify impervious surfaces, there would be serious confusion between the mountainous vegetation with low reflectance impervious surfaces (such as: villages and small cities), fortunately, the optical reflectance features performed well to distinguish them because of significant spectral differences. However, if only the multi-temporal optical reflectance images were used to detect the impervious surfaces, there would be obvious confusion between impervious surfaces with bare soils and croplands, for example, the spectral characteristics of

impervious surfaces, bare soils and croplands were overlapping in the Asia region (Fig. 4). In summary, only the combination of multi-source training features could guarantee the classification accuracy across different impervious landscapes.

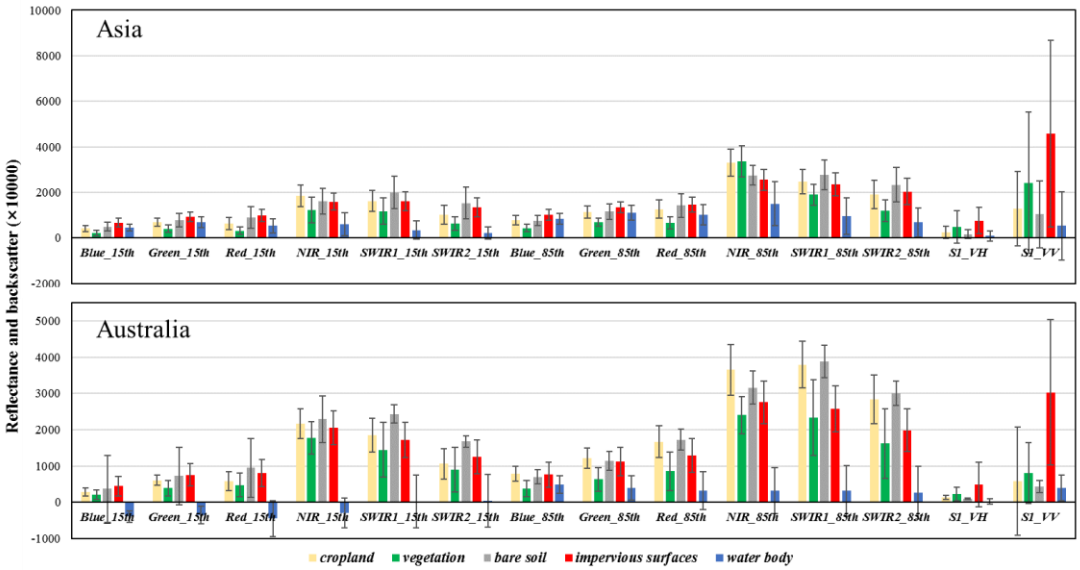


Figure 4: The reflectance/backscatter characteristics of different land-cover types over Landsat optical and Sentinel-1 SAR imagery in the Asia and Australia regions.

Secondly, although the 15th percentile had a higher importance than the 85th percentile in most of the spectral bands, we found that there was a large degree of complementarity between the images from two different seasons (Fig. 3). For example, the importance of the 15th percentile in the NIR and SWIR1 bands was low while that of 85th percentile was high, and the total importance of the bi-seasonal spectral features exceeded 70% in some cases. The reasons that the temporal information was important for accurately mapping of impervious surface included: (1) some land-cover types such as cropland had similar spectra with impervious surface at fallow season, but with the growing season imagery imported, this misclassification could be easily removed; (2) Sun et al. (2017) explained that the growing season was the best time for impervious surface mapping over temperate continental climate zones, and Zhang et al. (2014a) found that winter (dry season) is the best season to estimate impervious surface in subtropical monsoon regions. The multi-temporal information can address the problem of seasonal variability at different geographical zones. Fig. 4 (Australia region) also illustrated that the cropland and impervious surfaces were spectrally inseparable in the 15th percentile but the difference was obvious in the 85th percentile. Therefore, temporal variability can be considered an important contribution for accurate impervious surface mapping.

Thirdly, the importance of Landsat texture features was lower than 5% in these six regions, because the Sentinel-1 SAR backscatter and texture features were able to provide information on the surface material and its spatial structure and variation. Due to the complexity of land-surfaces and different mechanism of optical and SAR imagery, the optical textures could complement a lot to SAR features at mountainous and semiarid areas (Asia and Australia regions). Some studies demonstrated that these features contributed a lot to the improvement of impervious mapping accuracy. For example, Shaban and Dikshit (2001) emphasized that the integration of texture variables increased the accuracy from 86.86% to 92.69% because texture imagery could capture the local spatial structure and the variability of land-cover categories.

Lastly, since most regions are located in the flat areas, only the cumulative importance of topographical variables over the region in Asia exceeded 5%. The reasons why topographical information reached high importance over mountainous areas were because the impervious surfaces usually located in the flat areas (Ban et al., 2015) and Sentinel-1 SAR imagery had high backscatter signals over mountainous areas similar to the impervious surfaces, which increased the importance of topographical variables. Similarly, Clarke et al. (1997) explained that topographical variables (slope, aspect and DEM) contribute a lot to impervious surface mapping over mountainous areas. These features are, therefore, indispensable in the accurate mapping of impervious surfaces in mountainous regions.

3. The accuracy assessment experiment should be improved and expanded to multiple urban landscape types (e.g. globally selecting validation sites in more bare soil prevalent cities and vegetation prevalent cities), so that readers can clearly understand how multiple datasets work in land cover mapping under varying landscape conditions (the same reason as the 2nd comment).

Great thanks for the comment. Based on the suggestion, the accuracy assessment experiment has been expanded as:

The accuracy of the five global impervious surface maps over 15 validation regions with different impervious landscapes is presented in Table 2. Six evaluation metrics, including the producer's accuracy (which measures the commission error) and user's accuracy (which measures the omission error) of the impervious surface, the producer's and user's accuracy of non-impervious surfaces as well as the overall accuracy and kappa coefficient, were used to assess the accuracy. Overall, the MSMT_RF-based map achieved the highest overall accuracy of 0.951 and kappa coefficient of 0.898 compared with 0.896 and 0.780 for FROM_GLC-2015, 0.856 and 0.695 for NUACI-2015, 0.903 and 0.794 for GHSL-2015, 0.884 and 0.753 for GlobeLand30-2010, and 0.880 and 0.754 for HBASE-2010 using all 15 regional validation data.

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%.

As for the producer's accuracy for impervious surface (measuring the commission error), the GHSL-2015 products performed best and achieved the accuracy of 0.973, followed by the MSMT-2015 of 0.948, GlobeLand30-2010 of 0.947, FROM_GLC-2015 of 0.946, NUACI-2015 of 0.898 and HBASE-2010 of 0.841. Compared with user's accuracy of impervious surface, these reference products had better performance on this metric, which meant they had lower commission error.

Table 2. Accuracy of the six impervious surface maps over 15 validation regions

	NAME	BGK	JHB	LHS	MDR	MNS	MBN	MSC	NYK	NIM	NTU	PNX	RYH	SPL	WIP	XAN	O.A.
	I.L.	CR	BS	BS	BS	VG	VG	VG	VG	BS	BS	BS	BS	VG	CR	CR	
MSMT-2015	**U.I.**	0.951	0.963	0.691	0.929	0.993	0.957	0.987	0.995	0.869	0.750	0.988	0.918	0.984	1.000	0.929	0.932
	P.I.	0.997	0.922	0.989	0.961	0.938	0.972	0.961	0.981	0.987	0.951	0.975	0.944	0.965	0.915	0.940	0.948
	U.N.	0.997	0.958	0.996	0.986	0.966	0.987	0.949	0.952	0.997	0.975	0.975	0.954	0.978	0.958	0.922	0.964
	P.N.	0.951	0.981	0.873	0.975	0.996	0.980	0.982	0.987	0.964	0.859	0.987	0.932	0.990	1.000	0.909	0.953
	O.A.	0.974	0.960	0.899	0.971	0.975	0.978	0.970	0.983	0.969	0.888	0.981	0.938	0.980	0.971	0.926	0.951
	Kappa	0.948	0.912	0.747	0.925	0.945	0.948	0.939	0.957	0.904	0.754	0.963	0.874	0.958	0.934	0.850	0.898
NUACI-2015	**U.I.**	0.695	0.885	0.031	0.469	0.935	0.690	0.933	0.960	0.526	0.587	0.765	0.822	0.935	0.777	0.562	0.735
	P.I.	0.979	0.693	0.889	0.818	0.952	0.918	0.977	0.927	0.968	0.915	0.968	0.912	0.917	0.923	0.927	0.898
	U.N.	0.985	0.800	0.998	0.963	0.975	0.970	0.972	0.788	0.995	0.965	0.975	0.933	0.947	0.971	0.943	0.941
	P.N.	0.757	0.932	0.686	0.835	0.966	0.868	0.919	0.884	0.882	0.785	0.806	0.862	0.959	0.907	0.624	0.834
	O.A.	0.838	0.829	0.689	0.833	0.961	0.880	0.950	0.911	0.893	0.818	0.870	0.883	0.943	0.911	0.728	0.856
	Kappa	0.677	0.641	0.040	0.500	0.914	0.706	0.899	0.789	0.624	0.590	0.740	0.761	0.879	0.783	0.476	0.695
FROM_GLC-2015	**U.I.**	0.717	0.952	0.027	0.844	0.938	0.891	0.953	0.984	0.549	0.763	0.883	0.749	0.935	0.854	0.595	0.794
	P.I.	0.990	0.779	1.000	0.973	0.974	0.958	0.982	0.972	0.960	0.930	1.000	0.975	0.986	0.981	0.982	0.946
	U.N.	0.992	0.862	1.000	0.992	0.987	0.982	0.977	0.931	0.994	0.963	1.000	0.984	0.992	0.993	0.986	0.968
	P.N.	0.772	0.972	0.686	0.947	0.968	0.950	0.942	0.960	0.887	0.864	0.895	0.823	0.961	0.938	0.652	0.870
	O.A.	0.853	0.893	0.689	0.953	0.970	0.953	0.964	0.969	0.896	0.885	0.941	0.876	0.970	0.950	0.765	0.896
	Kappa	0.706	0.772	0.037	0.872	0.933	0.889	0.927	0.923	0.641	0.750	0.883	0.746	0.936	0.879	0.548	0.780
GHSL-2015	**U.I.**	0.619	0.752	0.453	0.815	0.880	0.849	0.958	0.991	0.451	0.619	0.940	0.672	0.925	0.899	0.741	0.787
	P.I.	1.000	0.949	1.000	0.989	0.996	0.978	0.982	1.000	1.000	1.000	0.995	0.996	0.996	0.991	0.968	0.973
	U.N.	1.000	0.979	1.000	0.997	0.998	0.991	0.977	1.000	1.000	1.000	0.995	0.998	0.998	0.996	0.968	0.985
	P.N.	0.717	0.886	0.795	0.938	0.941	0.932	0.948	0.979	0.867	0.804	0.943	0.783	0.955	0.957	0.742	0.868
	O.A.	0.806	0.903	0.825	0.949	0.958	0.945	0.966	0.994	0.880	0.851	0.968	0.849	0.970	0.966	0.840	0.903
	Kappa	0.615	0.770	0.530	0.860	0.903	0.870	0.932	0.985	0.563	0.664	0.935	0.687	0.936	0.919	0.686	0.794
GlobeLand30-2010	**U.I.**	0.310	0.704	0.410	0.825	0.804	0.744	0.908	0.981	0.537	0.779	0.923	0.831	0.902	0.749	0.750	0.761
	P.I.	0.992	0.950	0.991	0.978	0.961	0.975	0.962	0.954	1.000	0.968	0.966	0.905	0.972	0.954	0.874	0.947
	U.N.	0.997	0.981	0.998	0.993	0.983	0.991	0.955	0.901	1.000	0.984	0.968	0.926	0.984	0.984	0.859	0.970
	P.N.	0.582	0.867	0.782	0.941	0.905	0.891	0.891	0.955	0.885	0.874	0.926	0.866	0.942	0.898	0.726	0.852
	O.A.	0.648	0.888	0.810	0.949	0.921	0.911	0.929	0.936	0.899	0.904	0.945	0.883	0.953	0.911	0.798	0.884
	Kappa	0.303	0.731	0.483	0.861	0.818	0.783	0.857	0.917	0.645	0.790	0.890	0.762	0.898	0.779	0.597	0.753
HBASE-2010	**U.I.**	0.801	0.915	0.527	0.888	0.913	0.744	0.984	0.998	0.720	0.776	0.953	0.909	0.941	0.911	0.883	0.875
	P.I.	0.911	0.784	0.957	0.843	0.965	0.970	0.770	0.915	0.947	0.968	0.905	0.757	0.855	0.806	0.719	0.841
	U.N.	0.919	0.872	0.989	0.942	0.983	0.989	0.625	0.771	0.989	0.984	0.900	0.755	0.901	0.902	0.552	0.883
	P.N.	0.817	0.953	0.816	0.960	0.955	0.887	0.969	0.994	0.927	0.873	0.950	0.908	0.961	0.958	0.784	0.909
	O.A.	0.859	0.886	0.841	0.928	0.959	0.908	0.826	0.933	0.930	0.903	0.926	0.826	0.916	0.905	0.739	0.880
	Kappa	0.718	0.756	0.586	0.816	0.907	0.779	0.633	0.824	0.776	0.787	0.853	0.654	0.826	0.785	0.450	0.754

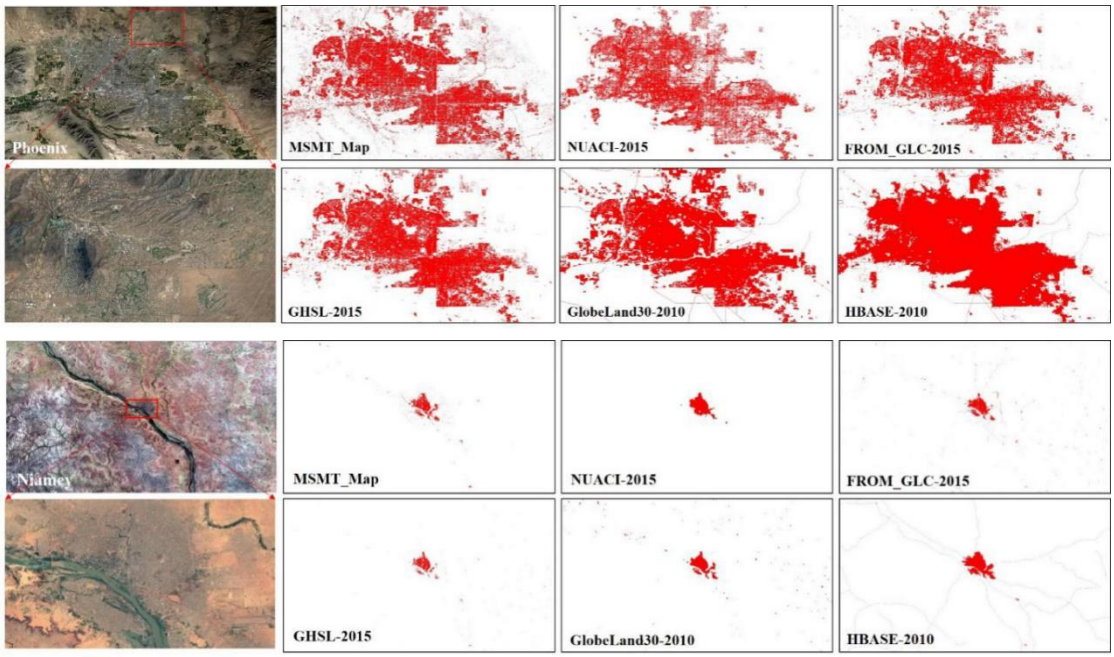
Note: I.L., impervious landscape, CR, cropland-prevalent impervious landscape, BS, bare soil-prevalent impervious landscape, VG, vegetation-prevalent impervious landscape, P.I., producer's accuracy of impervious surfaces, U.I., user's accuracy of impervious surfaces, P.N., producer's accuracy of non-impervious surfaces, U.N., user's accuracy of non-impervious surfaces, O.A., overall accuracy.

To intuitively compare the performance of these six impervious products, five validation regions, including two bare soil-prevalent regions (Phoenix and Niamey), one vegetation-prevalent city (New York) and two cropland-prevalent regions (Winnipeg and Bangkok), were selected in Fig. 9. Specifically, in the first bare soil prevalent region of Phoenix, the NUACI-2015 obviously under-estimated the impervious surfaces in the center of Phoenix city. The causes of omission maybe came from the threshold method used by the NUACI-2015. Liu et al. (2018) developed a novel NUACI index to enhance the impervious surfaces and suppressed the non-impervious surfaces and then found an optimal threshold for NUACI index to split the impervious and non-impervious surfaces. However, the NUACI values of rural villages and roads were usually located in the mixed areas of impervious and non-impervious surfaces, so the NUACI-2015 had great ability for large-size impervious surfaces but with poor performance for fragmented impervious surfaces. FROM_GLC-2015 performed well in the central city but missed impervious objects over peripheral urban. For example, the enlargement region (red rectangle), composited by sparse buildings and bare soils, was underestimated by the FROM_GLC-2015. This omission error maybe came from the sparse training samples (91,433 training samples in the globe) (Gong et al., 2013). The GHSL-2015, accurately capturing the central and peripheral impervious objects, had significant agreement with the MSMT-2015, it achieved the user's accuracy of 0.940 and producer's accuracy of 0.995 in this region (Table 2). As for the GlobeLand30-2010, there was little omission for the fragmented impervious objects over peripheral urban because of the temporal interval of 5 years and the minimum 4×4 mapping unit (Chen et al., 2015). The HBASE-2010 had biggest impervious areas among several global products but it misclassified the vegetation and bare soils into impervious surfaces in the urban central, so it had highest commission error of 9.5% in Table 2. As for the second bare soil prevalent city of Niamey, these products, except for the GHSL-2015 which had smaller impervious area than other products and missed the peripheral impervious objects, had similar performance with the Phoenix: the NUACI-2015 had high omission error especially for the fragmented objects, the HBASE-2010 lost the impervious details and achieved highest commission error of 5.3% in Table 2, the GlobeLand30-2010 missed some small objects (the limitation of minimum 4×4 mapping unit) and peripheral impervious objects caused by the temporal interval, and the FROM_GLC-2015 had great performance on the dense impervious areas but it was under-estimated over peripheral areas.

Next, in the vegetation-prevalent region of New York, six products generally had similar identification results and accurately captured the spatial distribution of New York city, so they achieved high mapping accuracy exceeding 90% in Table 2. However, from a detail perspective, there were still differences between these products. Specifically, NUACI-2015 performed well in the central of city but missed the sparse impervious objects over the peripheral city, for example, the enlargement region (red rectangle) illustrated the mixture of vegetation and sparse buildings over the peripheral city, the NUACI-2015 and GlobeLand30-2010 had smaller impervious areas than other products. The HBASE-2010 still suffered the highest commission error of 8.5% and had biggest impervious areas because it misclassified the bare soils and vegetation in the central city into impervious surfaces (blue

rectangles). The GHSL-2015, FROM_GLC-2015 and MSMT-2015 achieved higher mapping accuracy because they captured both dense and sparse impervious objects in the central and peripheral city.

Lastly, in the two cropland-prevalent cities of Bangkok and Winnipeg, the MSMT-2015 had greater advantages and achieved highest user's accuracy of 95.1% and 100% compared to the NUACI-2015 of 69.5% and 77.7%, the FROM_GLC-2015 of 71.7% and 85.4%, the GHSL-2015 of 61.9% and 89.9%, the GlobeLand30-2010 of 31.0% and 74.9%, and HBASE-2010 of 80.1% and 91.1% in Table 2. Fig.9 intuitively illustrated the performance of each product. GlobeLand30-2010 had smaller impervious areas in the central city because of the temporal interval and missed the road networks due to the minimum mapping unit of 4×4. As a result, the GlobeLand30-2010 achieved the lowest user's accuracy. NUACI-2015 captured impervious surfaces in the central city but missed the road networks and sparse village buildings in the peripheral cities. FROM_GLC-2015 and HBASE-2015 had similar performance in these two regions, which captured medium and large cities but missed the road networks and villages buildings. As HBASE-2010 contained the OpenStreetMap data to provide information on major road network (Wang et al., 2017a), the omission error of the HBASE-2010 was relatively low and only these village roads and buildings were missed, however, it still suffered serious over-estimation problem. Especially in the Bangkok city, the non-impervious pixels (bare soils, water, and vegetation) was misclassified as impervious surfaces. Therefore, the HBASE-2010 reached the highest commission error among these impervious products in Table 2.



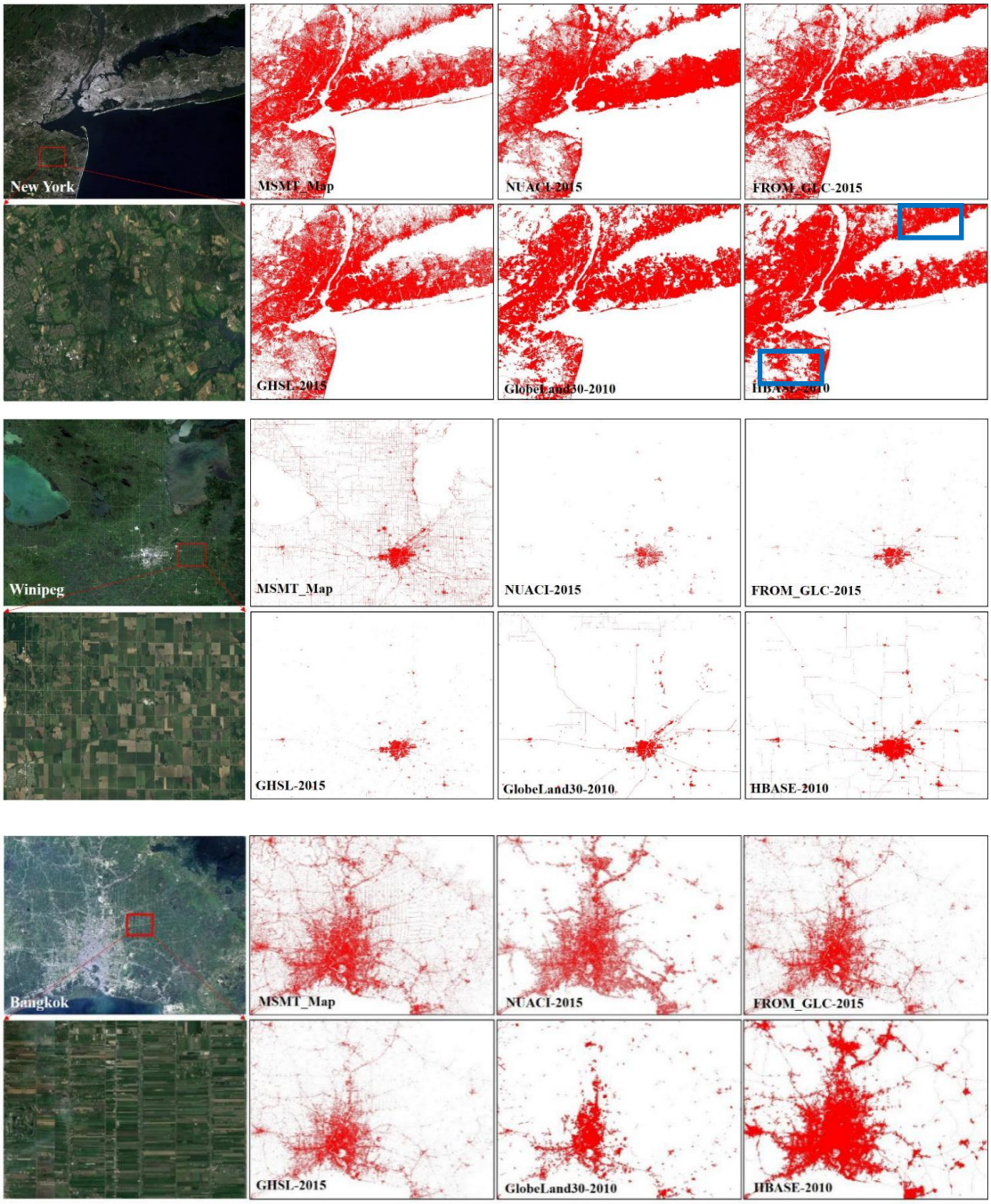


Figure 9: Comparisons between the MSMT_RF-based impervious surface maps and other products (corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019), the GlobeLand30 products developed by Chen et al. (2015), and the HBASE products developed by Wang et al. (2017a), respectively) for five regions with various impervious landscapes.

4. The training sample source/method may not be scientifically sound. GlobeLand30 was adopted as impervious surface training sample source, however, this global land cover product provides users with artificial layer but not impervious surface layer. The “impervious surface land cover” used in this study is actually a mixture land cover of vegetation, impervious surface and bare soil in urban area.

Great thanks for the comment. The reasons why we used the GlobeLand30 to derive the global training samples are included as follows.:

1) After carefully checking, there was great consistency for the definition of impervious surface between GlobeLand30 and our study. Specifically, the GlobeLand30 in (Chen et.al 2015) defined "Artificial surfaces mainly consists of urban areas, roads, rural cottages and mines, which are primarily based on asphalts, concrete, sand and stone, bricks, glasses, and other materials";

-our study: "Impervious surfaces are usually covered by anthropogenic materials which prevent water penetrating into the soil (Weng, 2012), which are primarily composited by asphalts, sand and stone, concrete, bricks, glasses, etc."

Similarly, the NUACI products also shared same definition with GlobeLand30 as "the term 'urban land' in this paper refers to 'impervious surface', i.e., artificial cover and structures such as pavement, concrete, brick, stone and other man-made impenetrable cover types (Liu et al. 2018)" .

2) The GlobeLand30 had several advantages over other impervious products (NUACI, FROM_GLC and GHSL..) including: it was developed by combining pixel-based classification, multi-scale segmentation and manual editing based on high-resolution imagery, so almost impervious objects in GlobeLand30 were checked by visual interpretation. In addition, the non-impervious training samples in this study included three sub-classes (cropland, bare soil and other non-impervious land-cover types), if we chose the NUACI or GHSL products, these non-impervious samples similar to impervious surface cannot be completely collected. The reasons have been added in the Section 3 - "Collection of global training samples" as:

"The GlobeLand30 land-cover product was used to derive global training samples because it had many advantages including: (1) the impervious surface layer in GlobeLand30 was accurately developed by combining the pixel-based classification, multi-scale segmentation and manual editing based on high resolution imagery and validated to achieve an user's accuracy of 86.7%; (2) it simultaneously contained the impervious surface and other land-cover types similar to impervious surface (such as cropland and bare land), so the global training samples including several non-impervious land-cover types could be easily collected to build the RF model for accurately mapping of impervious surface. "

5. More explanations may be required to arguments in this paper.

Here I am left with a limited number of questions about the specifics of implementation and implications of the method and results as well as clarity of this manuscript, which I note below.

Great thanks for the comment. These questions have been answered one by one in the following comments.

"However, Gao et al. (2012) explained that these coarse-resolution global impervious surface maps were not suitable for many applications and policy makers at local or regional scales (Line 35)." This part is not understandable, why the previous impervious surface maps are not suitable for certain applications and policy makers? Could you please explain it with more straightforward instances?

Thanks for the comment. To make the expression more straightforward, this sentence has been rewritten as:

"However, since the complex characteristics of impervious landscapes and inherent resolution of human activity, coarse-resolution global impervious surface maps were not suitable for many

applications and policy makers at local or regional scales, for example, the urban-rural pattern planning and road network monitoring usually required the fine spatial resolution impervious surface products (Gao et al., 2012)"

"(Chen et al., 2015; Gao et al., 2012; Goldblatt et al., 2018; Gong et al., 2019; Gong et al., 2013; Homer et al., 2015; Li et al., 2018; Liu et al., 2018; Sun et al., 2017). Line 50". Because of their similar works as you did in this paper (i.e. global or regional impervious surface maps), it is necessary to give more introduction to previous land cover datasets, and to present the importance of your work.

Thanks for the comment. Yes, it is necessary to give more details to previous land-cover datasets, this paragraph has been added as:

Specifically, the National Land Cover Dataset (NLCD) produced the first 30-m map of the United States including impervious surface as three separate land-cover types (Developed low, Developed medium, and Developed high intensity) using Landsat imagery, DMSP OLS and USGS National Elevation Dataset (NED) digital elevation data, and achieving the user' accuracy of 0.48~0.66 (Homer et al., 2004). Similarly, the FROM_GLC produced the global 30-m impervious surface map as an independent land cover type with the user' accuracy of 0.307 (Gong et al., 2013); GlobeLand30 combined the pixel-based classification, segmentation and manual editing based on the high resolution imagery to develop the 30-m impervious surface map as an independent layer with the user's accuracy of 0.867 (Chen et al., 2015). However, as sparse training samples of impervious surfaces cannot capture all relevant spectral heterogeneity when producing these land-cover products, the impervious surface layers usually suffered low accuracy except for GlobeLand30 (which includes manual interpretation). Therefore, a few studies proposed to independently produce the impervious surface products. For example, Liu et al. (2018) proposed the Normalized Urban Areas Composite Index (NUACI) method to produce a global 30-m impervious surface map and achieved an overall accuracy of 0.81-0.84 and a kappa values of 0.43-0.50. However, the NUACI product had a relatively poor performance in terms of producer's accuracy (0.50–0.60) and user's accuracy (0.49-0.61). Brown de Colstoun et al. (2017) combined the object-based segmentation, random forest classification and post-processing to develop the Global 30-m Man-made Impervious Surface (GMIS) and Human Built-up and Settlement Extent (HBASE) dataset in 2010 which achieved a kappa coefficient of 0.91 using scene-level cross validation in Europe (Wang et al., 2017b). Pesaresi et al. (2016) used the multi-temporal Landsat imagery and symbolic machine learning method to produce the Global 30-m Human Settlement Layer (GHSL) in 2014, and achieved a total accuracy of 96.28% and kappa coefficient of 0.3233 based on Land Use/Cover Area frame Survey (LUCAS) reference data. Therefore, an accurate impervious surface map at fine spatial resolution is still urgently needed using an efficient mapping method."-

"However, these land-cover products focus on the overall accuracy of the mapping of all land-cover types rather than that of impervious surfaces alone (Line 55)". It may be confusing here. It implies the land cover products that focus the overall accuracy deliver low quality impervious surface map. What difference existing between "focusing on overall accuracy" and "focusing on impervious surface alone"?

Thanks for the comment. This sentence has been rewritten as:

"However, as sparse training samples of impervious surfaces cannot capture all relevant spectral heterogeneity when producing these land-cover products, the impervious surface layers usually suffered

low accuracy except for GlobeLand30 (which includes manual interpretation). Therefore, a few studies proposed to independently produce the impervious surface products. For example..."

"However, the NUACI product had a relatively poor performance in terms of producer's accuracy (0.50–0.60) and user's accuracy (0.49-0.61). Therefore, an accurate impervious surface map at fine spatial resolution is still urgently needed (Line 55)". Why did you only mention accuracy of NUACI here? How about other land cover datasets?

Thanks for the comment. The accuracies of other datasets have been added as:

the National Land Cover Dataset (NLCD) produced the first 30-m map of the United States including impervious surface as three separate land-cover types (Developed low, Developed medium, and Developed high intensity) using Landsat imagery, DMSP OLS and USGS National Elevation Dataset (NED) digital elevation data, and achieving the user' accuracy of 0.48~0.66 (Homer et al., 2004). Similarly, the FROM_GLC produced the global 30-m impervious surface map as an independent land cover type with the user' accuracy of 0.307 (Gong et al., 2013); GlobeLand30 combined the pixel-based classification, segmentation and manual editing based on the high resolution imagery to develop the 30-m impervious surface map as an independent layer with the user's accuracy of 0.867 (Chen et al., 2015).

Liu et al. (2018) proposed the Normalized Urban Areas Composite Index (NUACI) method to produce a global 30-m impervious surface map and achieved an overall accuracy of 0.81-0.84 and a kappa values of 0.43-0.50. However, the NUACI product had a relatively poor performance in terms of producer's accuracy (0.50–0.60) and user's accuracy (0.49-0.61). Brown de Colstoun et al. (2017) combined the object-based segmentation, random forest classification and post-processing to develop the Global 30-m Man-made Impervious Surface (GMIS) and Human Built-up and Settlement Extent (HBASE) dataset in 2010 which achieved a kappa coefficient of 0.91 using scene-level cross validation in Europe (Wang et al., 2017b). Pesaresi et al. (2016) used the multi-temporal Landsat imagery and symbolic machine learning method to produce the Global 30-m Human Settlement Layer (GHSL) in 2014, and achieved a total accuracy of 96.28% and kappa coefficient of 0.3233 based on Land Use/Cover Area frame Survey (LUCAS) reference data (Pesaresi et al., 2016).

"However, these spectral mixture methods can produce underestimates in areas whether the density of impervious surfaces is high and overestimates in areas of low density (Sun et al., 2017; Weng, 2012) (Line70)". Spectral unmixing technique may have underestimate and overestimate issues, but how about its overall or average accuracy when comparing it with pixel-level mapping approaches?

Great thanks for the comment. The spectral unmixing techniques and pixel-level mapping methods represented the 'soft' and 'hard' classifications respectively, and had different advantages for impervious mapping. The accuracies of spectral unmixing techniques and pixel-level classification mainly depended on the reliability of endmember and training data respectively, so it was difficult to directly compare the performance of two methods. However, the spectral unmixing techniques had great difficulties to identify one suitable endmember to represent all types of impervious surfaces, so the pixel-level mapping approaches were more popular for impervious surface mapping. The disadvantages of spectral unmixing techniques have been added as:

"However, these spectral mixture methods can produce underestimates in areas with high density

impervious surfaces and overestimates in areas with low density impervious surfaces, **and may have great difficulties to identify one suitable endmember to represent all types of impervious surfaces** (Sun et al., 2017; Weng, 2012)"

The "data preprocessing" and "mapping approach" was mixed up, which makes readers difficult to capture the point of datasets and classification methods, so I may suggest splitting them into different sections.

Great thanks for the comment. Based on the suggestion, we split the data preprocessing of "Collection of global training samples" as an independent section 3.

Questions on remote sensing datasets for classification (Lines 120). Descriptions of purpose and necessity for different remote sensing datasets were not clear, better wording in "Datasets" may be required

Great thanks for the comment. Based on the suggestion, the descriptions of the remote sensing datasets have been added, the specific changes have been listed in the following comments.

Do five data sources contribute equally to classification? How do they theoretically work for differentiating different land covers? (Line 120)

Thanks for the comment. The functions of five datasets have been added as:

"In this study, three kinds of data sources including Landsat-8 optical imagery, Sentinel-1 SAR data and STRM/ASTER DEM topographical variables were selected and collected for the mapping of impervious surfaces across the world on the GEE platform. Furthermore, the combination of VIIRS NTL imagery and MODIS EVI products was used to derive the set of global impervious surface and non-impervious surfaces training data."

How does C-band SAR imagery contribute to differentiate impervious and pervious surfaces? How do artificial buildings, forests, grassland, and bare soil respond to SAR imagery? Please clarify it. (Line 130)

Thanks for the comment. The explanations why we imported the Sentinel-1 SAR imagery are added as:

"The Sentinel-1 satellite provides C-band SAR imagery at a variety of polarizations and resolutions. (Berger et al., 2012; ESA, 2016; Torres et al., 2012). **Due to the high dielectric properties of the building materials, the unique geometry of manmade features, and the special radar echo properties of artificial structures, the impervious surfaces usually had stronger backscattered signals than other land-cover types (such as: barren land, cropland and so on) in the SAR imagery**. In this study..."

How do EVI imagery work for classification procedure? (Line 145) Why do you involve DEM dataset in land cover mapping? How does this dataset work in differentiating impervious and pervious surfaces? (Line 150)

Thanks for the comment. The work of EVI imagery has been added as:

"The MODIS EVI imagery (MYD13Q1) from the MODIS V6 products contains the best available EVI data from among all the acquisitions obtained over a 16-day compositing period and has a spatial resolution of 250-m (Didan et al., 2015), **which was used to mitigate the NTL data's saturation**

problem and exclude false positive impervious samples (vegetated samples in the urban) when deriving the global training samples. In this study, ..."

The reasons why we import the DEM dataset in impervious surface mapping are added as:

"The Shuttle Radar Topography Mission digital elevation model (SRTM DEM), provided by the NASA JPL at a resolution of 1 arc-second (approximately 30 m) and covering the area between 60° north and 56° south (Farr et al., 2007), **was an useful auxiliary dataset for impervious surface mapping over mountainous areas because impervious surfaces mainly located in the flat areas and Sentinel-1 SAR data usually reflected high backscatter similar to the impervious surfaces in mountainous areas** (Ban et al., 2015)."

Questions on introducing state-of-art global impervious surface products. (Lines 205). GlobeLand30 actually does not provide impervious surface land cover, please adopt other global land cover dataset instead (Line 160). Furthermore, the review of the published global impervious surface datasets should be improved. For instance, three important global (or continental) impervious surface datasets – global man-made impervious surface (GMIS) dataset, NLCD impervious surface layer and global human built-up and settlement extent (HBASE) dataset- were not introduced.

Many thanks for the comment. After carefully checking the impervious surface definition of GlobeLand30 in Chen et al. (2015), we found there is consistency between GlobeLand30 and our study for defining impervious surface:

-GlobeLand30: "Artificial surfaces mainly consists of urban areas, roads, rural cottages and mines, which are primarily based on asphalts, concrete, sand and stone, bricks, glasses, and other materials";

-our study: "Impervious surfaces are usually covered by anthropogenic materials which prevent water penetrating into the soil (Weng, 2012), which are primarily composited by asphalts, sand and stone, concrete, bricks, glasses, etc. (Chen et al., 2015)."

Similarly, the NUACI products also shared same definition with GlobeLand30 as "the term 'urban land' in this paper refers to 'impervious surface', i.e., artificial cover and structures such as pavement, concrete, brick, stone and other man-made impenetrable cover types" (Liu et al. 2018)

Next, in order to improve the review of the published global impervious surfaces datasets, the Human Built-up and Settlement Extent (HBASE) and Global Human Settlement Layer (GHSL), have been added as: (Note: as HBASE and GMIS were companion datasets and GMIS was continues impervious fraction, we only selected the HBASE dataset):

The HBASE (Human Built-up and Settlement Extent) dataset was the first global 30-m dataset of man-made impervious cover derived from the Global Land Survey (GLS) Landsat data for 2010 (HBASE-2010) (https://sedac.ciesin.columbia.edu/data/set/ulandsat-hbase-v1). It was produced by combining meter-resolution training data (exceeding 20 millions), Open Street Map, VIIRS NTL, GLS Landsat SR and MODIS NDVI products, and achieved a kappa coefficient of 0.91 using scene-level cross validation in Europe (Wang et al., 2017a; Wang et al., 2017b).

The GHSL (Global Human Settlement Layer), a global information baseline describing the spatial evolution of the human settlements in the past 40 years, was developed by using symbolic machine learning model trained by the collected high-resolution samples, multi-temporal Landsat imagery in the epochs 1975, 1990, 2000, and 2015 (Florczyk et al., 2019). In this study, the GHSL impervious surface

map at 30-m for 2015 (GHSL-2015) (https://ghsl.jrc.ec.europa.eu/download.php) was employed for comparison analysis, which achieved an overall accuracy of 96.28% and kappa coefficient of 0.3233 validated using Land Use/Cover Area frame Survey (LUCAS) reference data (Pesaresi et al., 2016).

Questions on selecting training samples (Lines 205). The impervious surface training samples were selected based on GlobeLand30 map. However, GlobeLand30 only provides "artificial surface" which consists of impervious surfaces and small patch vegetation areas in urban area. Thus, the training samples of MSMT_RF could be no longer reliable although extra datasets were used for samples filtering. The training samples for classifier may be collected from other impervious surface datasets instead of GlobeLand30 map.

Great thanks for comment. The "artificial surface" in GlobeLand30 was defined as:"Artificial surfaces mainly consists of urban areas, roads, rural cottages and mines, which are primarily based on asphalts, concrete, sand and stone, bricks, glasses, and other materials", which is same as our definition for impervious surface: "Impervious surfaces are usually covered by anthropogenic materials which prevent water penetrating into the soil (Weng, 2012), which are primarily composited by asphalts, sand and stone, concrete, bricks, glasses, etc. (Chen et al., 2015)."

Next, the reasons why we chose the GlobeLand30 instead of other products (GMIS, GHSL, FROM_GLC and so on) were: (1) GlobeLand30 had the user' accuracy of 0.867 for impervious surface, and each impervious surface object was edited by manual interpretation, which greatly guaranteed the high confidence of impervious surface; (2) The training samples in this study contained impervious surface and non-impervious surfaces (barren land, cropland and other land-cover types), if we chose the GMIS or GHSL products, we cannot collect the training data of some non-impervious surfaces (barren land and cropland) which usually shared similar spectra with impervious surface. The reasons have been added in the Section 3 –"Collection of global training samples" as:

The GlobeLand30 land-cover product was used to derive global training samples because it had many advantages including: (1) the impervious surface layer in GlobeLand30 was accurately developed by combining the pixel-based classification, multi-scale segmentation and manual editing based on high resolution imagery and validated to achieve an user's accuracy of 86.7%; (2) it simultaneously contained the impervious surface and other land-cover types similar to impervious surface (such as cropland and bare land), so the global training samples including several non-impervious land-cover types could be easily collected to build the RF model for accurately mapping of impervious surface.

Finally, in order to guarantee the confidence of the training samples, we took two steps: (1) selecting the homogeneous areas as the candidate set; (2) using the EANTLI index to minimize the effects of classification error and the land-cover changes caused by the temporal interval.

It is not clear that how and why twelve sampling sites (i.e. high-density sites, medium density sites and low-density regions) were selected? How do spectral features vary among these sites? What features were exhibited by different density regions? This information should be updated.

Great thanks for the comment. The twelve sampling sites were randomly selected by combining the histogram of impervious fraction. However, these sites cannot demonstrate the importance of multi-source datasets. Therefore, based on the previous and latter suggestions, the sampling sites have been re-selected by combining the land-cover types.

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)

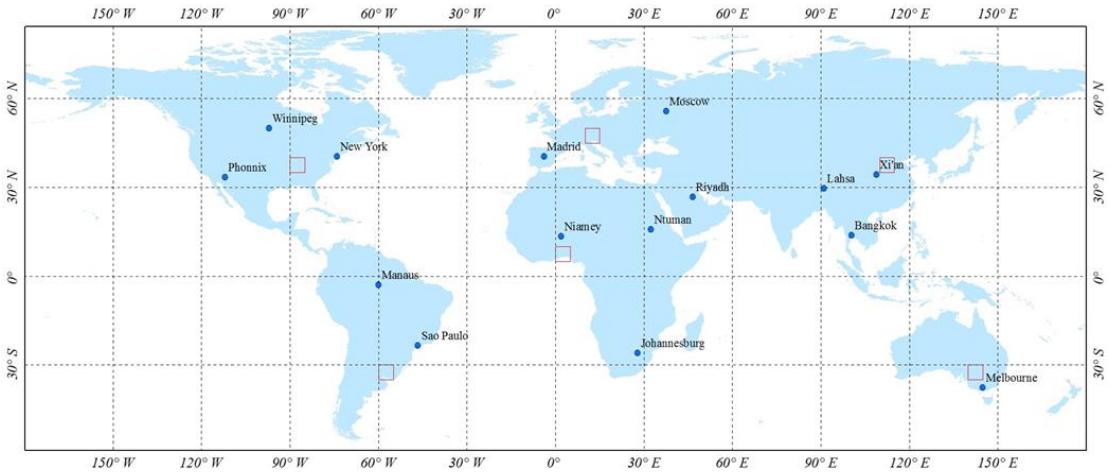


Figure 1: The spatial distribution of the fifteen validation regions (blue) corresponding to regions of different impervious landscapes on different continents, together with the six 5°×5° validation regions (red) used to measure the variable importance.

Moreover, data preprocessing procedures and mapping methods are mixed up, which makes manuscript confusing. I suggest separating them into two sections. In particular, more explanations of the preprocessing operations should be made, instead of only citing reference literature.

Great thanks for the comment. Based on the suggestion, the "Collection of global training samples" was split into an independent section 3. In addition, some arguments have been added more explanations

As the non-impervious surfaces consisted many land-cover types (water, vegetation, cropland and bare area) and some of them were spectrally similar to the impervious surface. For example, the bare soil and high reflectance impervious surfaces usually shared similar surface reflectance especially in arid and semi-arid areas with large areas of bare soils because the composition of impervious surfaces included rock material which was also found in bare areas (Sun et al., 2019b; Weng, 2012), the cropland showed similar reflectance to these low reflectance impervious surfaces (such as rural village, old cities) because they were usually composited of vegetation and high reflectance artificial materials or bare soils (Li et al., 2015). Therefore, the non-impervious training samples were split into three independent groups including: bare area, cropland and other non-impervious land-cover types. Furthermore, many studies had demonstrated that the distribution and balance of training samples had great influence on the mapping accuracy. For example, Zhu et al. (2016) found unbalanced training samples directly resulted in rare land-cover types under-represented relative to more abundant classes. Since the impervious surface was usually sparser than the non-impervious land-cover types (bare soil, cropland and so on), the training samples with uniform distribution were selected to ensure the rationality of training samples and capture all relevant spectral heterogeneity within impervious surfaces, namely, the approximate ratio of 1:3 was

used to represent the proportion of impervious to non-impervious surfaces (bare area, cropland and other non-impervious land-cover types).

Some parts of introduction to data preprocessing were also not understandable. Here is an example: "the suburban areas or rural villages were also easy to confused with croplands (Li et al., 2015)". It is not reasonable to compare a land use element (suburban area) with a land cover element (cropland). Explanations are always required for each of your arguments.

Great thanks for the comment. Yes, it was unreasonable to compare the land-use element with land-cover element, so this sentence has been rewritten as "the cropland showed similar reflectance to these low reflectance impervious surfaces (such as rural village, old cities) because they were usually composited of vegetation and high reflectance artificial materials or bare soils (Li et al., 2015)". The explanations of other arguments have been added according to the previous response.

Questions on accuracy assessment (Lines 315). Two accuracy assessment was conducted respectively in "fraction" way and "classified pixel" way. How much difference do the two accuracy assessment methods make? What special information can be provided by each method?

Great thanks for the comment. The detailed explanations of 'fraction-based validation' and 'sample-based validation' have been added as:

To completely analyze the performance of the MSMT_RF-based method, two validation methods including 'fraction-based' and 'pixel-based' were adopted. First, the 'fraction-based' validation method mainly illustrated the spatial agreement of impervious surfaces between the MSMT_RF-based impervious surface map and several existing products (GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, HBASE-2010 and GHSL-2015) from a global perspective. Specifically, all these global 30-m impervious surface maps were aggregated to a resolution of 0.05 °×0.05 ° and the fraction of impervious area was then calculated. Following that, scatter plots of the linear regression between the MSMT_RF-based results and the reference data were produced to provide the quantitative metrics of the agreement, including coefficient of determination (R^2) and root mean square error (RMSE).

In addition, a 'pixel-based' validation method, based on the visual interpretation samples over fifteen 1 °×1 ° regions covering different impervious landscapes and continents, was used to quantitatively analyze the accuracy metrics, including overall accuracy (O.A.), producer's accuracy (P.A.), user's accuracy (U.A.) and kappa coefficient (Olofsson et al., 2014) for assessing the performance of the MSMT_RF-based global impervious surface mapping.

Questions on Figure 5. As mentioned in previous questions, further review of currently available global impervious surface maps is needed. In the revised manuscript, I suggest adding error bars for progressive fraction intervals (e.g. 0.05, 0.1, 0.15, 0.2, : : :, 1.0)

Great thanks for the comment. Based on the previous and this comment, we added two global impervious products (GHSL and HBASE). Except for this scatter plots, we have added the spatial variations of six global impervious products as:

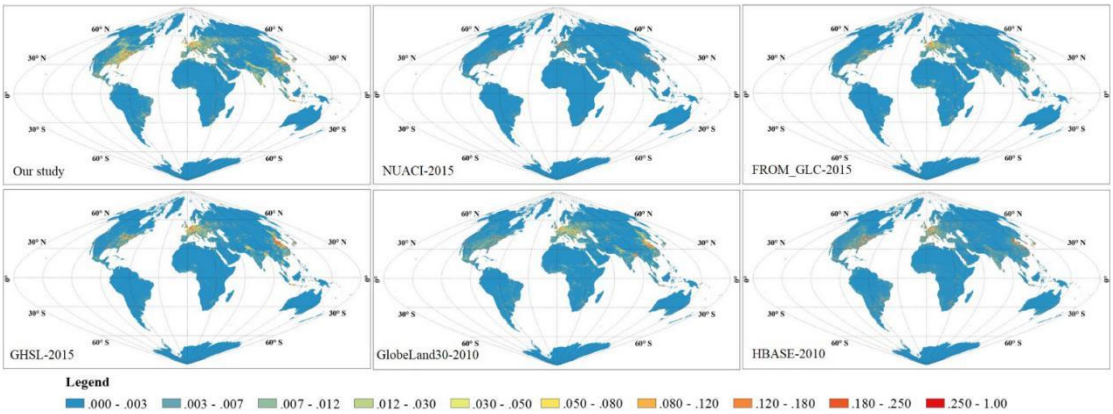


Figure 7. The spatial variations of six global 30-m impervious products after aggregating to the resolution of 0.05°.

Based on the suggestion, the error bars for the progressive intervals of 0.05 have been added as:

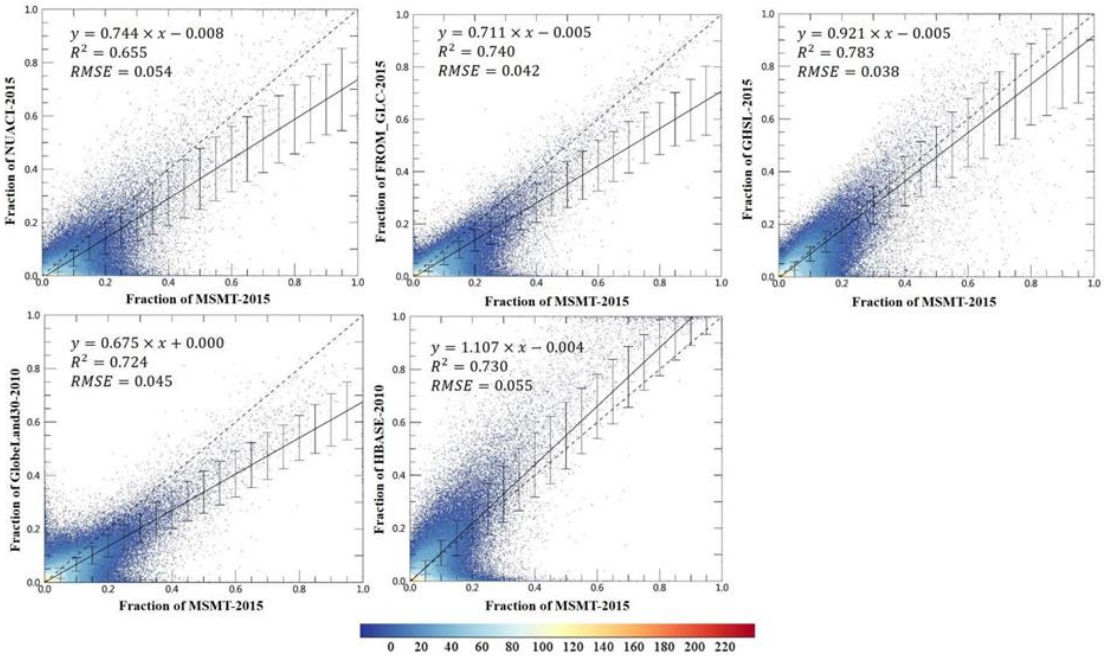


Figure 8: Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial grid of 0.05 °×0.05 °.

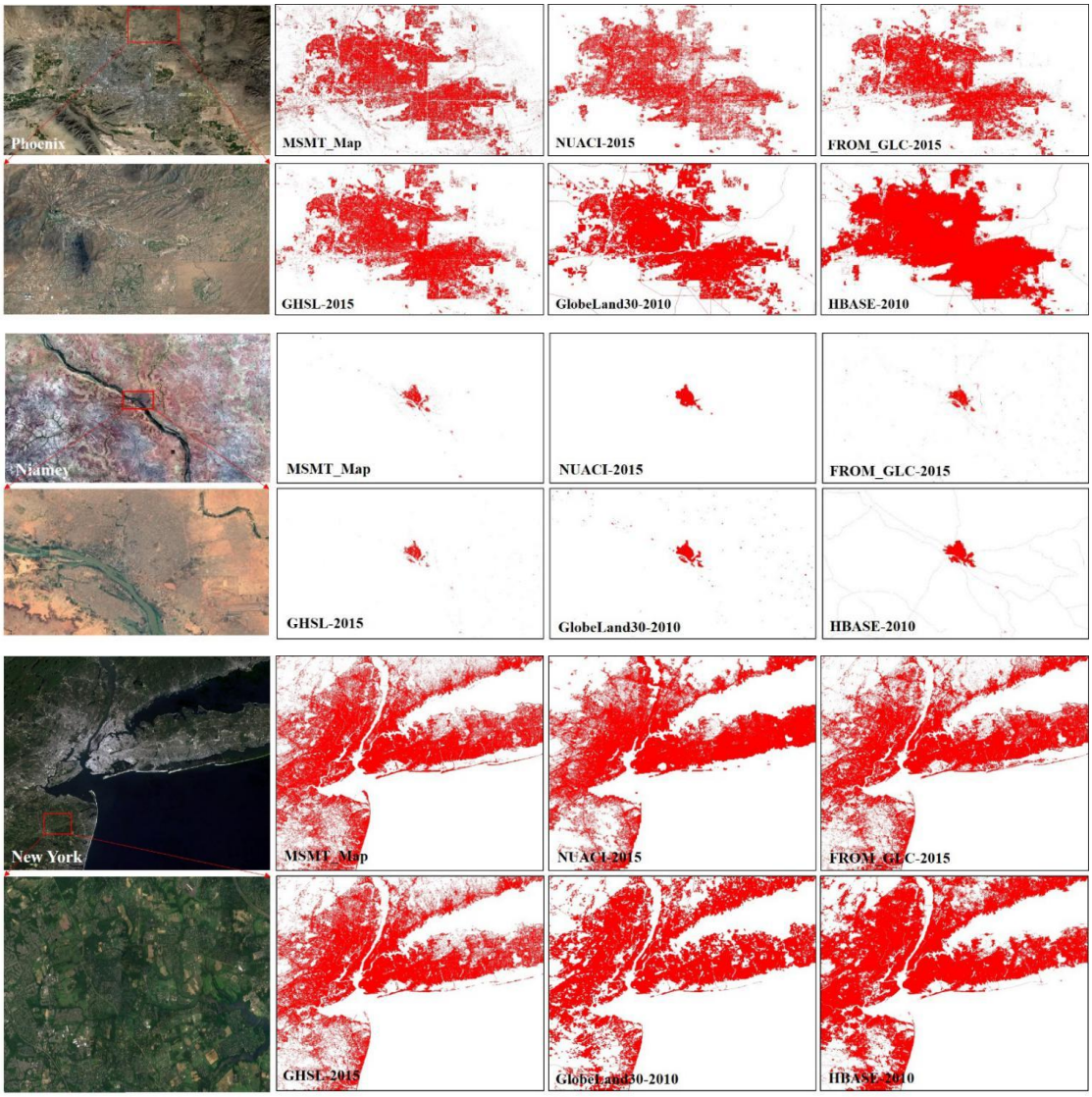
Confusing parts in Section 4.2. "As the stratified random sampling strategy was applied to each validation region independently, the low and medium density regions were easier to select these mixed impervious validation points (simultaneously containing the impervious and non-impervious surfaces in the 30-m validation window and the impervious areas exceed the predefined threshold of 50%) which were most difficult to identify for impervious surface mapping (Line 380)." What information did you want.

Great thanks for the comment. The meaning of this sentence was that the impervious surface mapping usually suffered relatively low accuracy over low and medium density regions where contained a higher proportions of mixed impervious surfaces (simultaneously containing the impervious and non-impervious surfaces in the Landsat pixel and the impervious areas exceed the

threshold of 50%). In the revised manuscript, these confusing sentences have been removed.

Questions for Figure 6. It is clear to show difference between impervious surface maps but not clear to visually compare RGB pixels with your maps. The RGB satellite images of macro areas may not be suitable to compare it with classified land cover map. Subset urban areas are preferred so that readers can clearly see how well the map is classified. Besides, I may not agree that "low, medium, high- density" areas are representative for comparison. To improve the figure, I suggest globally selecting urban areas with different landscapes (e.g. desert landscape urban areas such as Phoenix city, vegetation prevalent cities such as New York City). Furthermore, please do more works in reviewing global impervious surface datasets.

Great thanks for the comment. Based on the suggestion, this figure has been expanded as:



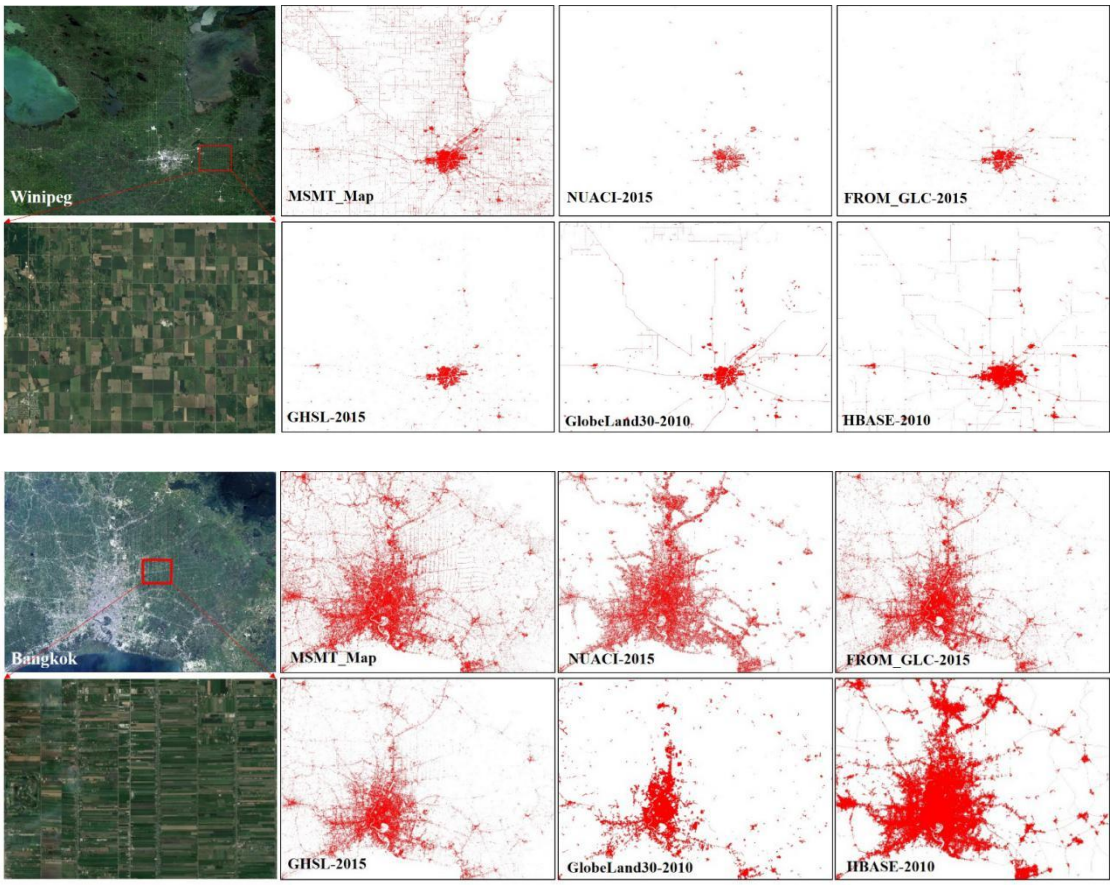


Figure 10: Comparisons between the MSMT_RF-based maps and other impervious surface products (corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019), the GlobeLand30 products developed by Chen et al. (2015), and the HBASE products developed by Wang et al. (2017a), respectively) for five regions with various impervious landscapes.

Figure 7 is an experiment result, and it should be moved to "Result"

Great thanks for the comment. The section has been moved to the "Result" as the Section 5.1 'The importance of multi-source and multi-temporal features'.

"The importance of all 37 training features for the six regions is illustrated in Fig. 7. These results indicate that the Sentinel-1 SAR features (VV and VH) had the greatest contribution to the final decision in most regions because SAR images can provide information about the structure and dielectric properties of the surface materials (Line 440)". What VV and VH feature difference is revealed between different land covers (e.g. impervious surface, forest, croplands, bare soil, water)?

Great thanks for the comment. Based on the suggestion, the response of different land-cover types over optical and SAR imagery have been added as:

To intuitively understand the characteristics of different land-cover types on optical and SAR imagery, two regions (the vegetation-prevalent region of Asia and bare soil-prevalent semi-arid region of Australia) were selected for comparison analysis. Fig. 4 illustrated the reflectance and backscatter statistics (mean and standard deviation) of five typical land-cover types (cropland, vegetation, bare soil,

impervious surfaces and water body). Obviously, impervious surfaces had highest backscatter signals in VV because of the high dielectric properties of the building materials, the unique geometry of manmade features, and the special radar echo properties of artificial structures, followed by the vegetation land-cover types. Further, since only a small part of the polarized signals (vertical turning horizontal) were returned to the sensor, the VH was significantly lower than VV but the ranking orders of different land-cover types in VH was similar to that of VV. Due to the complicated construction and heterogeneity of the impervious surfaces, the impervious surfaces also had highest standard deviation, for example, the urban central usually reflected higher VV and VH signals than the village buildings. If only Sentinel-1 SAR features were used to identify impervious surfaces, there would be serious confusion between the mountainous vegetation with low reflectance impervious surfaces (such as: villages and small cities), fortunately, the optical reflectance features performed well to distinguish them because of significant spectral differences. However, if only the multi-temporal optical imagery were used to detect the impervious surfaces, there would be obvious confusion between impervious surfaces with bare soils and croplands, for example, the spectral characteristics of impervious surfaces, bare soils and croplands were overlapping in the Asia region (Fig. 4). In summary, only the combination of multi-source training features could guarantee the classification accuracy across different impervious landscapes.

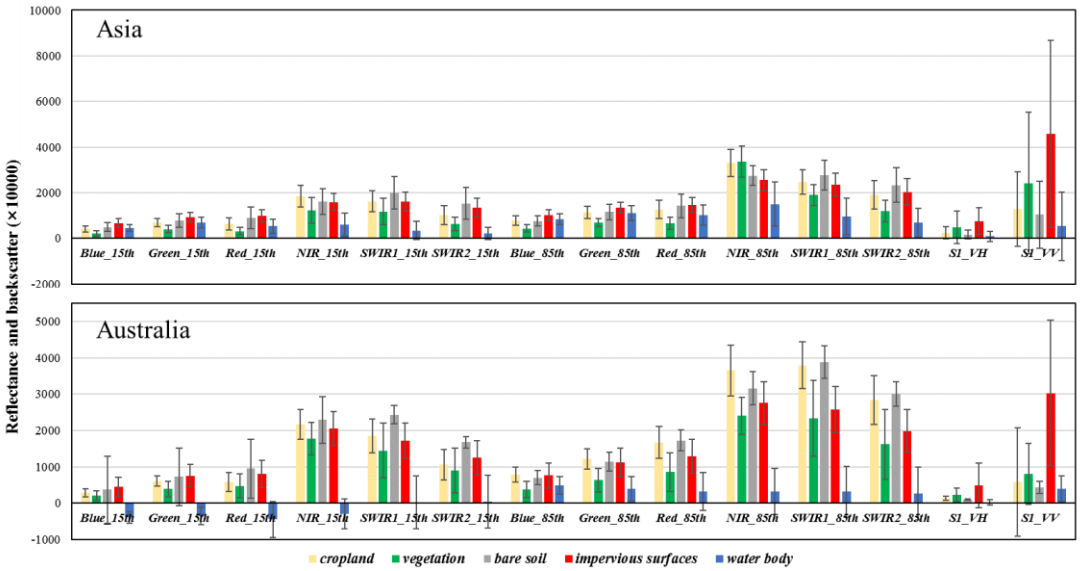


Figure 4: The reflectance/backscatter characteristics of different land-cover types over Landsat optical and Sentinel-1 SAR imagery in the Asia and Australia regions.

"Similarly, Zhu et al. (2012) demonstrated that the inclusion of multi-temporal imagery increased the accuracy by 8.9%. Schug et al. (2018) also found that bi-seasonal information could produce a more reliable performance than a single-year composited image. Therefore, temporal variability can be considered an important addition to accurate impervious surface mapping (Line 455)." Discussion and explanation should be made. Please exactly explain the theory in which how these datasets work for improving classification accuracies. Which land cover accuracy is improved by including these datasets?

Great thanks for the comment. The reasons why the temporal variability was important for impervious mapping have been added as:

The reasons that the temporal information was important for accurately mapping of impervious

surface included: (1) some land-cover types such as cropland had similar spectra with impervious surface at fallow season, but with the growing season imagery imported, this misclassification could be easily removed; (2) Sun et al. (2017) explained that the growing season was the best time for impervious surface mapping over temperate continental climate zones, and Zhang et al. (2014a) found that winter (dry season) is the best season to estimate impervious surface in subtropical monsoon regions. The multi-temporal information can address the problem of seasonal variability at different geographical zones. Fig. 4 (Australia region) also illustrated that the cropland and impervious surfaces were spectrally inseparable in the 15th percentile but the difference was obvious in the 85th percentile. Therefore, temporal variability can be considered an important contribution for accurate impervious surface mapping.

"Similarly, Clarke et al. (1997) explained that topographical variables (slope, aspect and DEM) contribute a lot to impervious surface mapping. These features are, therefore, indispensable in the accurate mapping of impervious surfaces in complex landscapes (Line 465)." Clarke et al. (1997) was cited without further explanation. Readers would like to know mechanism of topographical variable contributing to impervious surface mapping?

Great thanks for the comment. The mechanism of topographical variable contributing to impervious surface mapping has been added as:

Lastly, since most regions are located in the flat areas, only the cumulative importance of topographical variables over the region in Asia exceeded 5%. The reasons why topographical information reached high importance over mountainous areas were because the impervious surfaces usually located in the flat areas (Ban et al., 2015) and Sentinel-1 SAR imagery had high backscatter signals over mountainous areas similar to the impervious surfaces, which increased the importance of topographical variables. Similarly, Clarke et al. (1997) explained that topographical variables (slope, aspect and DEM) contribute a lot to impervious surface mapping over mountainous areas. These features are, therefore, indispensable in the accurate mapping of impervious surfaces in complex landscapes