Dear Topical Editor and Reviewers:

On behalf of my co-authors, we thank you very much for reviewing our manuscript and giving us the opportunity to revise the manuscript. We appreciate the comments on our manuscript entitled "Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform" (essd-2019-200).
We have revised the manuscript carefully according to the comments. All the changes were high-lighted (red color) in the manuscript. And the point-by-point response to the comments of the reviewers is also listed below.
Looking forward to hearing from you soon.

Best regards,

Prof. Liangyun Liu
liuly@radi.ac.cn
Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences
No.9 Dengzhuang South Road, Haidian District, Beijing 100094, China

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.

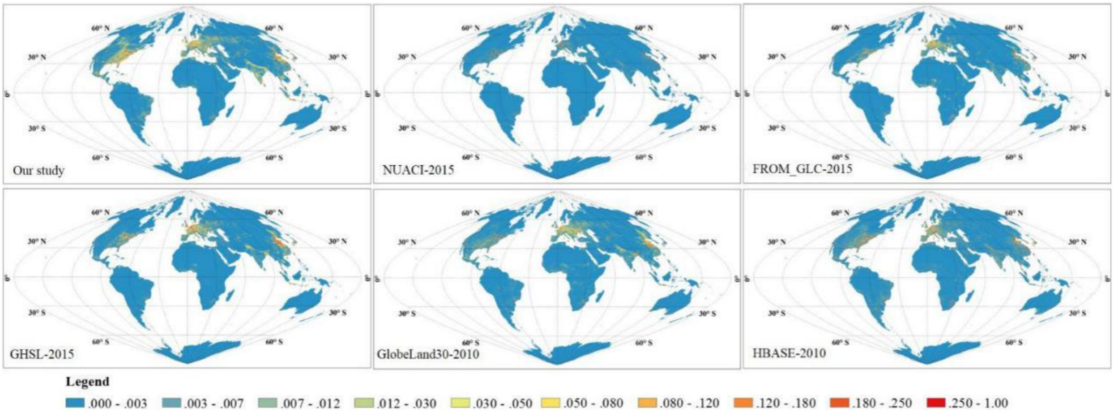


Figure 7: The spatial patterns of six global 30-m impervious products after aggregating to the resolution of 0.05°.

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.

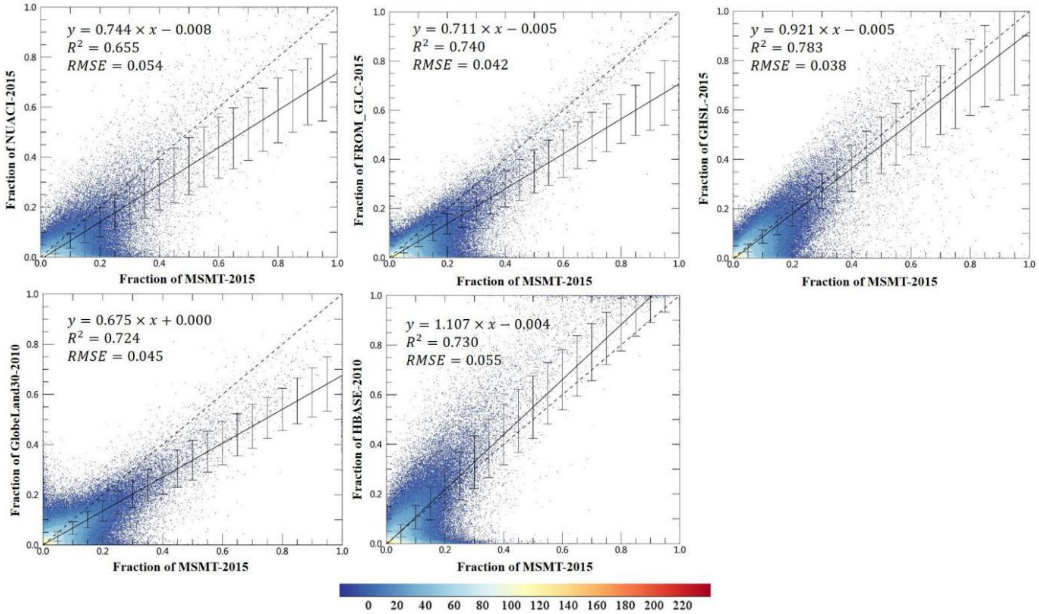


Figure 8: Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial grid of 0.05°×0.05°. The error bars were the standard deviation between reference datasets with fitted results.

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%.

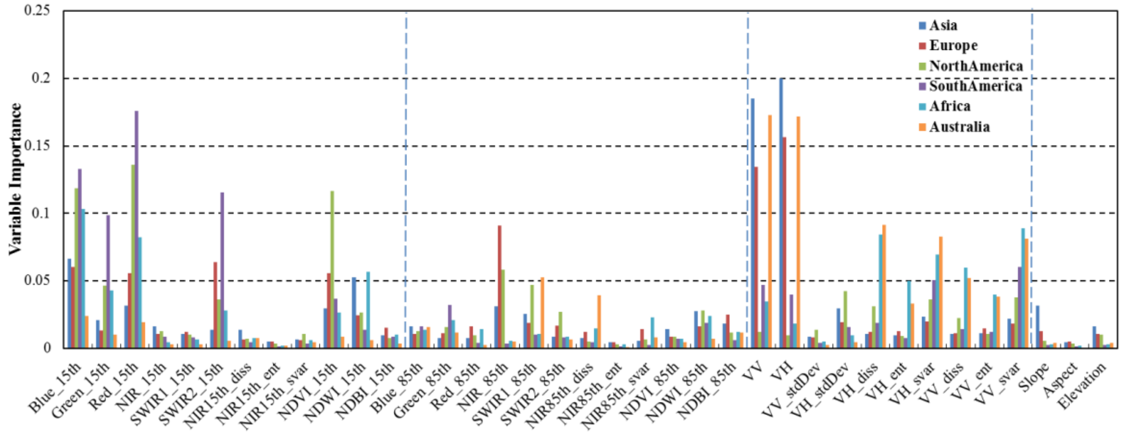


Figure 3: The importance of the input features derived from the random forest model using the training samples in six continental regions.

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.

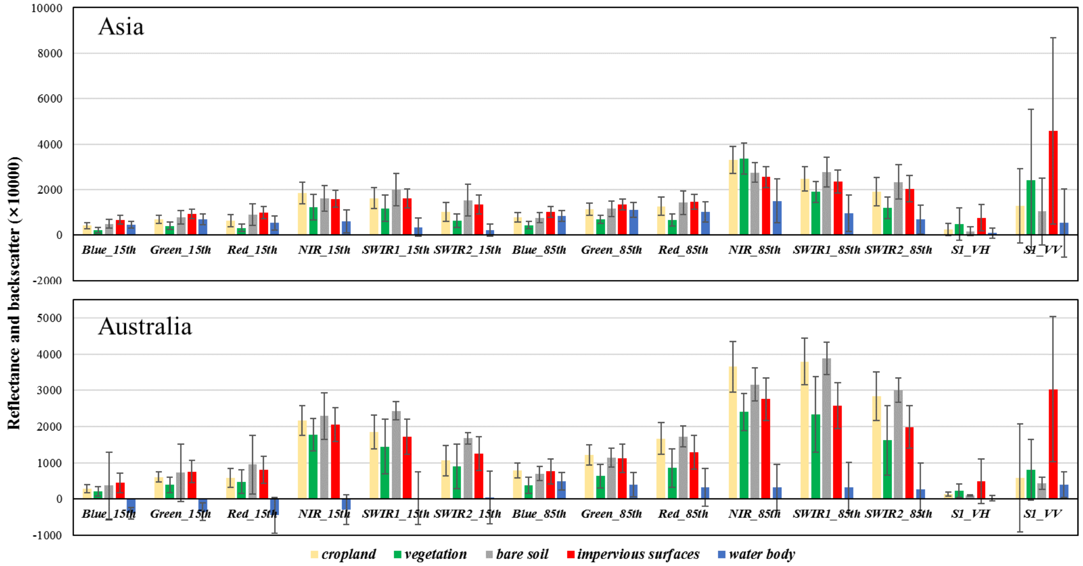


Figure 4: The reflectance/backscatter characteristics of different land-cover types over Landsat optical and Sentinel-1 SAR imagery in the Asia and Australia regions.

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

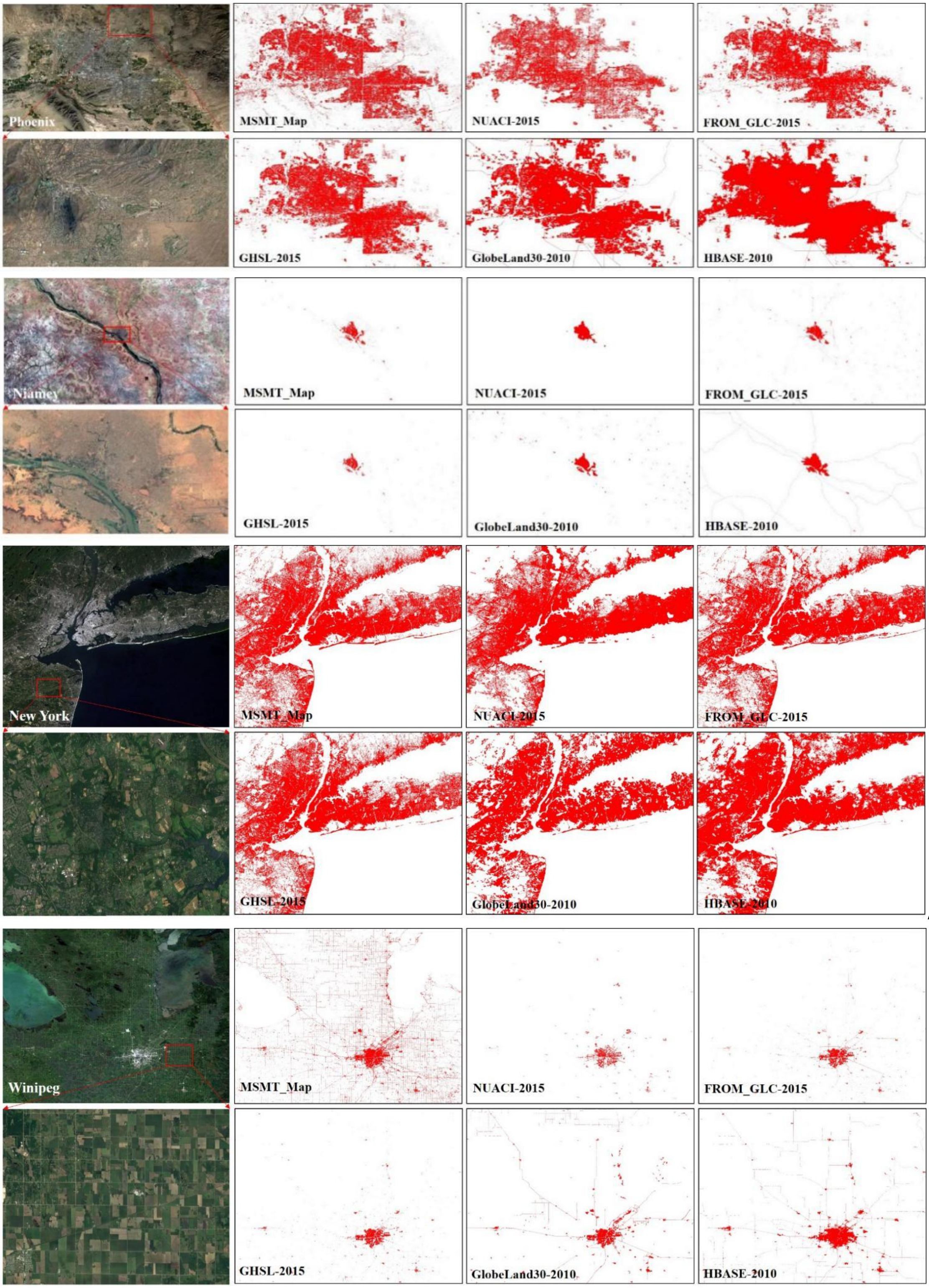
	U.N.	0.985	0.800	0.998	0.963	0.975	0.970	0.972	0.788	0.995	0.965	0.975	0.933	0.947	0.971	0.943	0.941
	P.N.	0.757	0.932	0.686	0.835	0.966	0.868	0.919	0.884	0.882	0.785	0.806	0.862	0.959	0.907	0.624	0.834
	O.A.	0.838	0.829	0.689	0.833	0.961	0.880	0.950	0.911	0.893	0.818	0.870	0.883	0.943	0.911	0.728	0.856
	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.


Phoenix
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010
Niamey
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010
New York
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010
Winipeg
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010

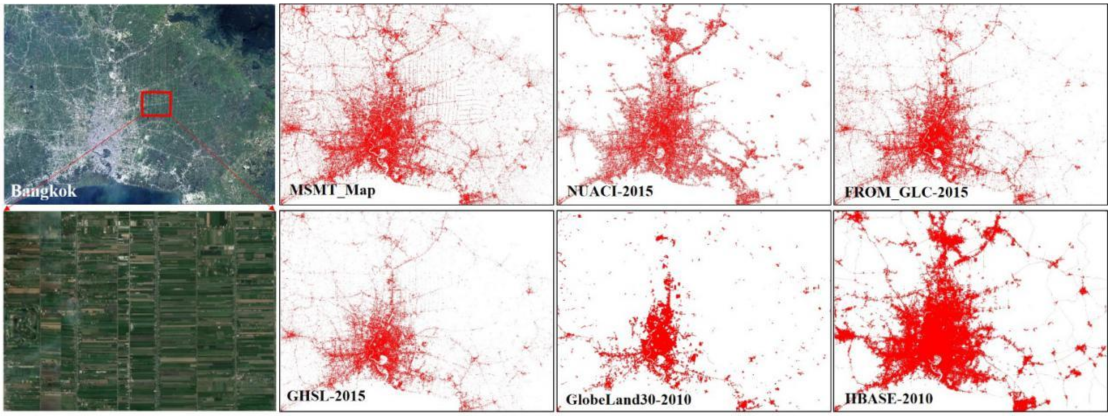


Figure 9: Comparisons between the MSMT_RF-based impervious surface maps and other products (corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019), the GlobeLand30 products developed 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)

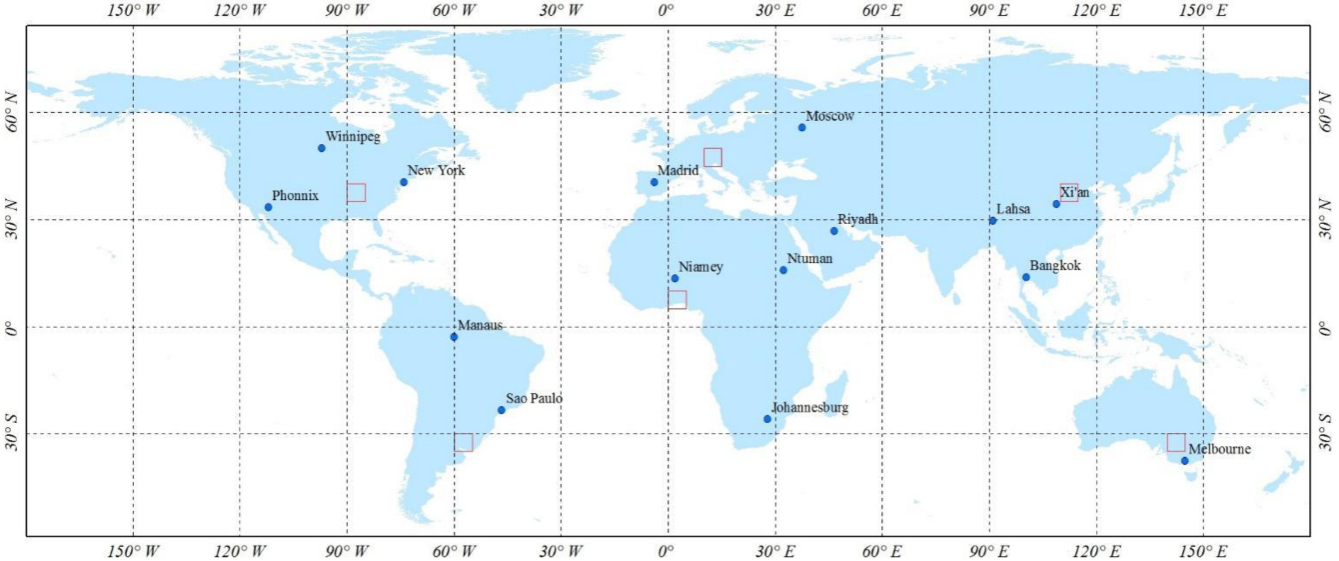


Figure 1: The spatial distribution of the fifteen validation regions (blue) corresponding to regions of different impervious landscapes on different continents, together with the six 5 °×5 ° validation regions (red) used to measure the variable importance.

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:

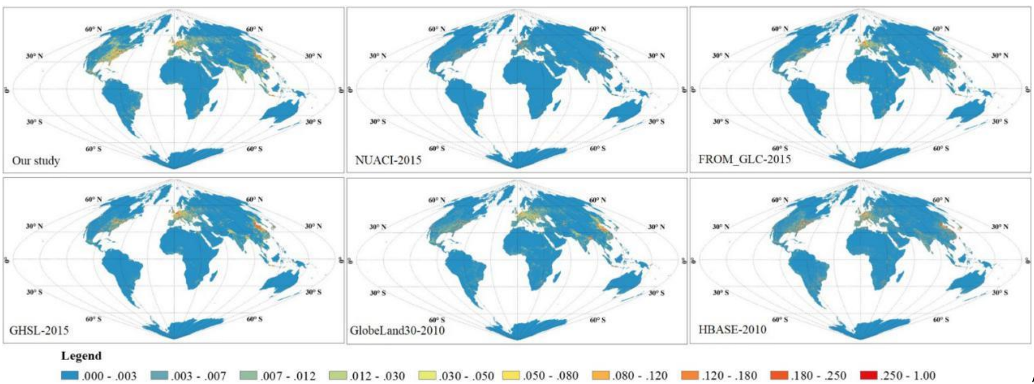


Figure 7. The spatial variations of six global 30-m impervious products after aggregating to the resolution of 0.05 °.

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

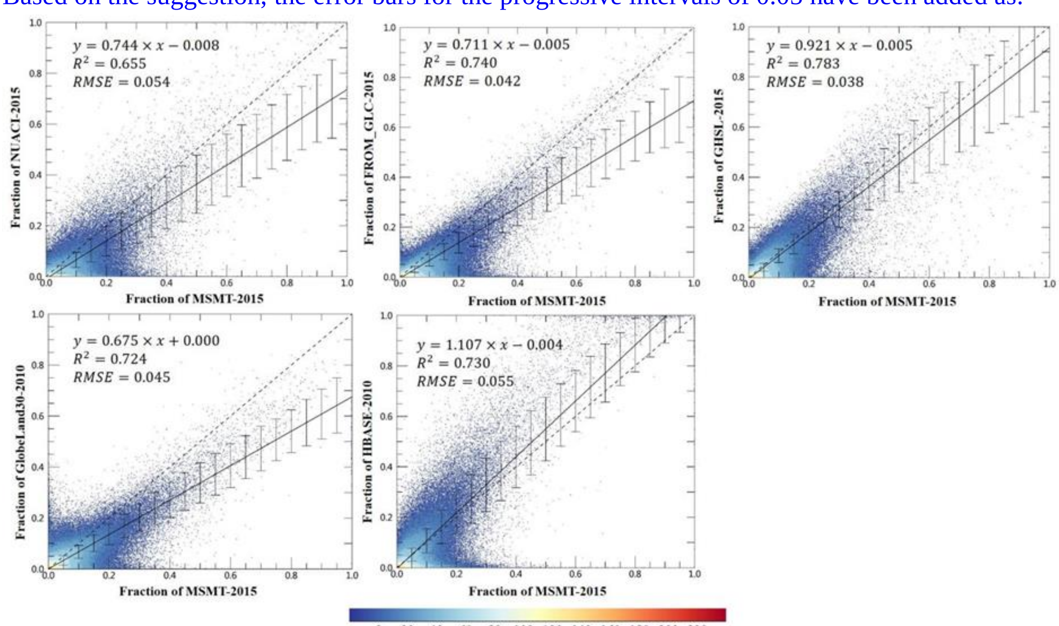


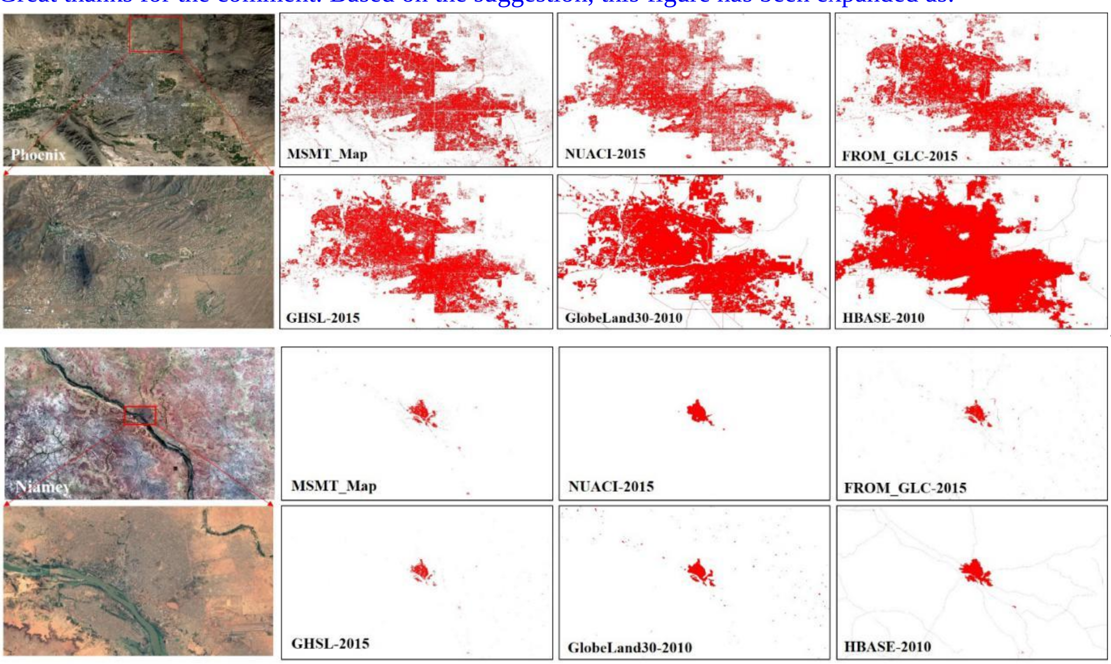
Figure 8: Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial grid of 0.05°×0.05°.

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:



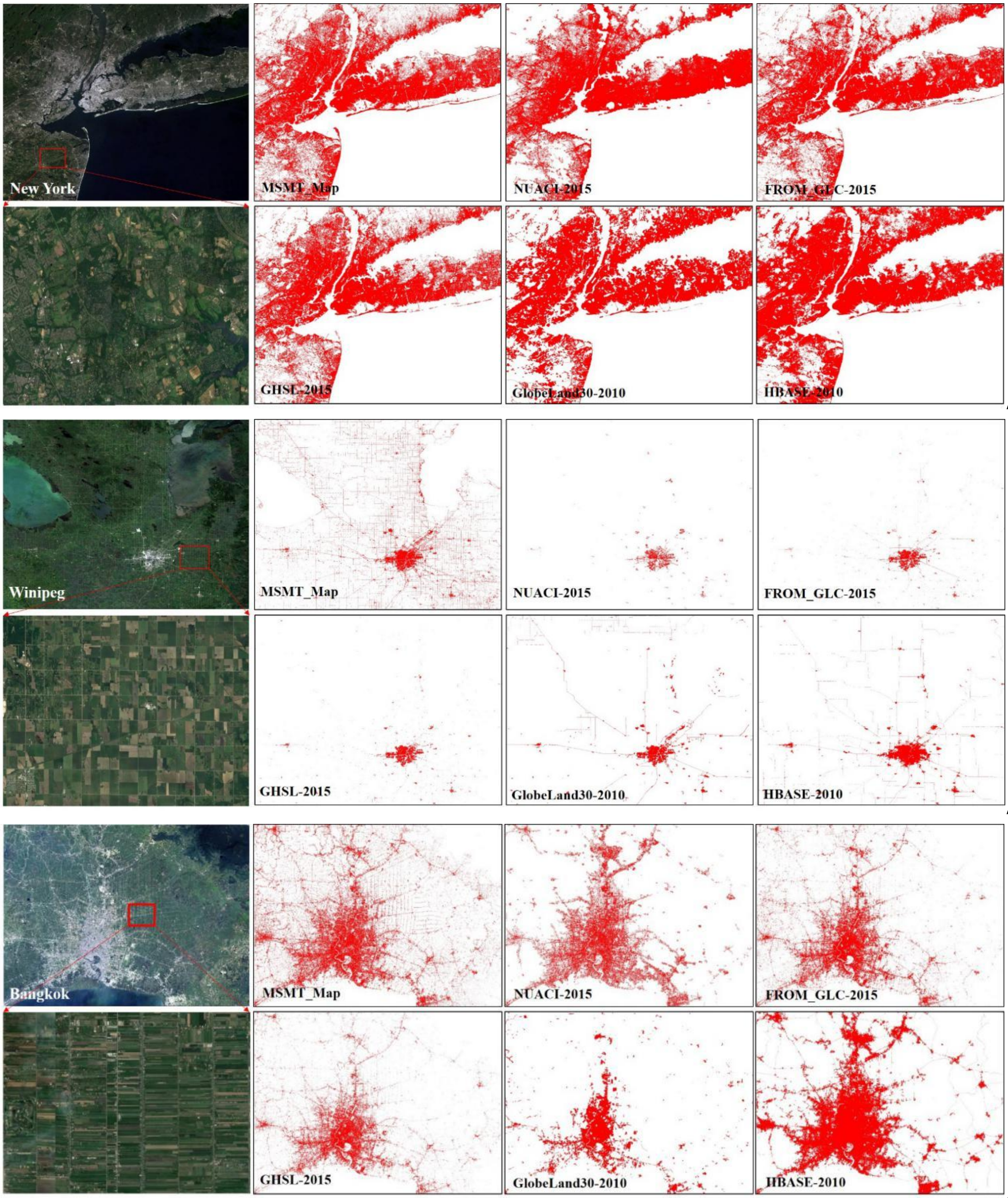


Figure 9: Comparisons between the MSMT_RF-based maps and other impervious surface products (corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019), the GlobeLand30 products developed 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.

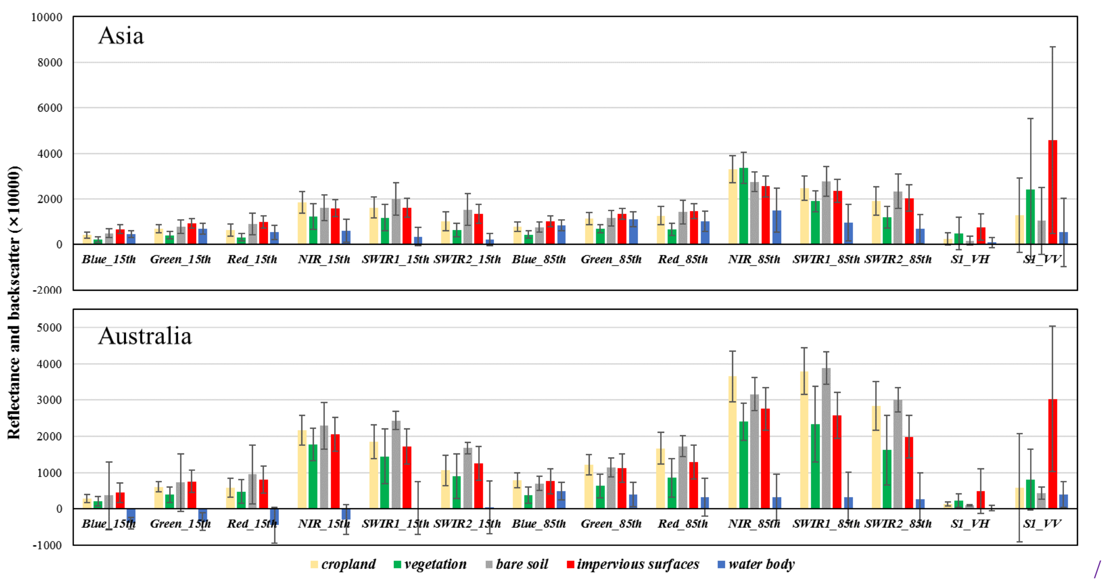


Figure 4: The reflectance/backscatter characteristics of different land-cover types over Landsat optical and Sentinel-1 SAR imagery in the Asia and Australia regions.

"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.

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.

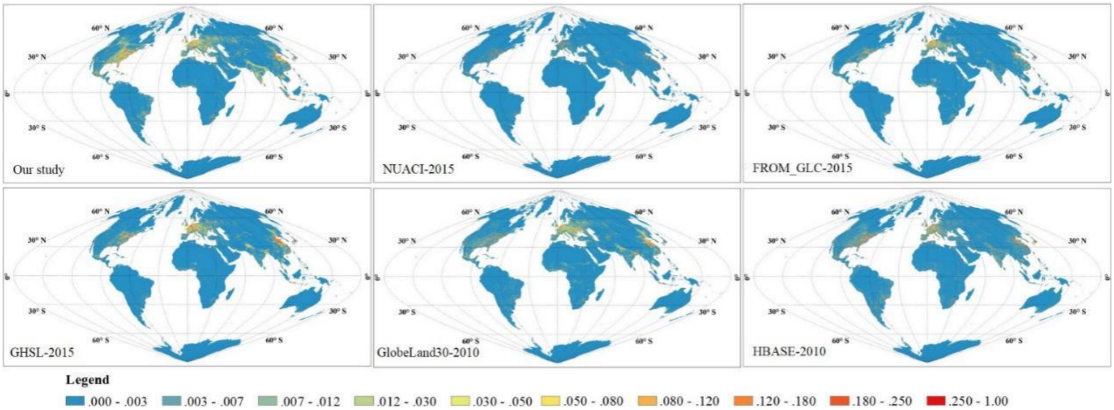


Figure 7: The spatial patterns of six global 30-m impervious products after aggregating to the resolution of 0.05 °.

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)."

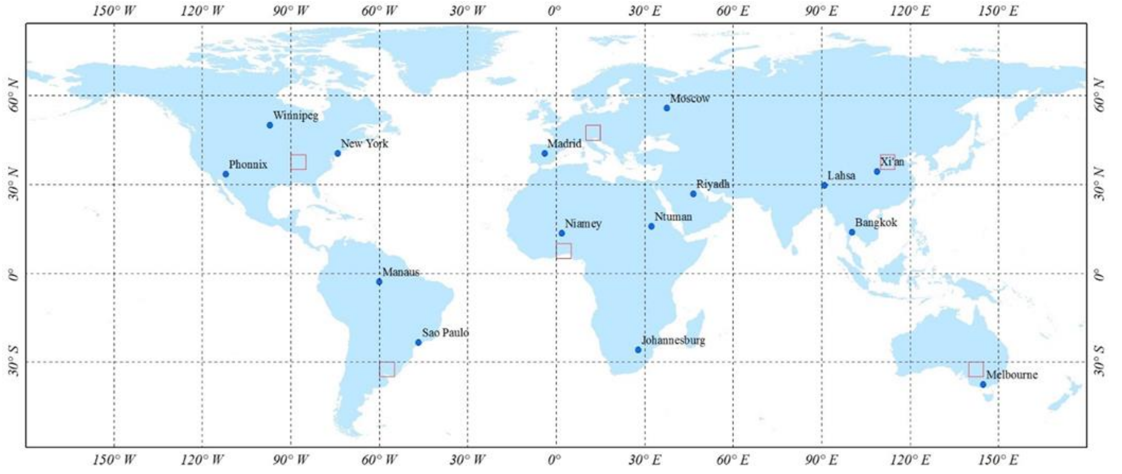


Figure 1: The spatial distribution of the fifteen validation regions (blue) corresponding to regions of different impervious landscapes on different continents, together with the six 5 °×5 °validation regions (red) used to measure the variable importance.

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...”

Short Comment #2

This paper presents a new global 30m impervious surface map produced with multi-source and multi-temporal remote sensing datasets and random forest (MSMT_RF). Compared with the currently available impervious products (i.e., GlobeLand30, FROM_GLC and NUACI), this MSMT_RF-based product has higher overall accuracy and kappa coefficient, which are 96.6% and 0.90, respectively. The superiority of the MSMT_RF-based product stems from two significant innovations of the method proposed in this study. First, multi-source and multi-temporal remote sensing data are combined to produce the impervious surface map. The comprehensive information provided by the combined data is useful in classifying land cover types, so the superiority of the MSMT_RF-based product in comparison with the other products is convincing. Second, a novelty method is proposed for selecting training samples based on the available impervious product and VIIRS NTL and MODIS EVI imagery. This method allows for the fully automatic selection of training samples to avoid manual training sample selection, which is time-consuming and laborious, especially at a global scale. This method has significant implications for producing more perfect global data products based on existing data products. I believe this study is a breakthrough over previous works in impervious surface mapping and will appeal to a broad readership. However, there are still some minor issues that should be addressed before final publication.
Great thanks for the positive comment. The manuscript has been improved according to your and other reviewers' comment.

Line 35, "urban the environment" should be "urban environment"
Great thanks for the comment. It has been corrected.

Figure 1, I cannot see the blue rectangles but only black points, which are supposed to be the blue rectangles. The authors should figure out how to make blue rectangles clear.
Great thanks for the comment. As we re-selected the validation regions based on the impervious landscapes, the new spatial distribution figure was changed as:

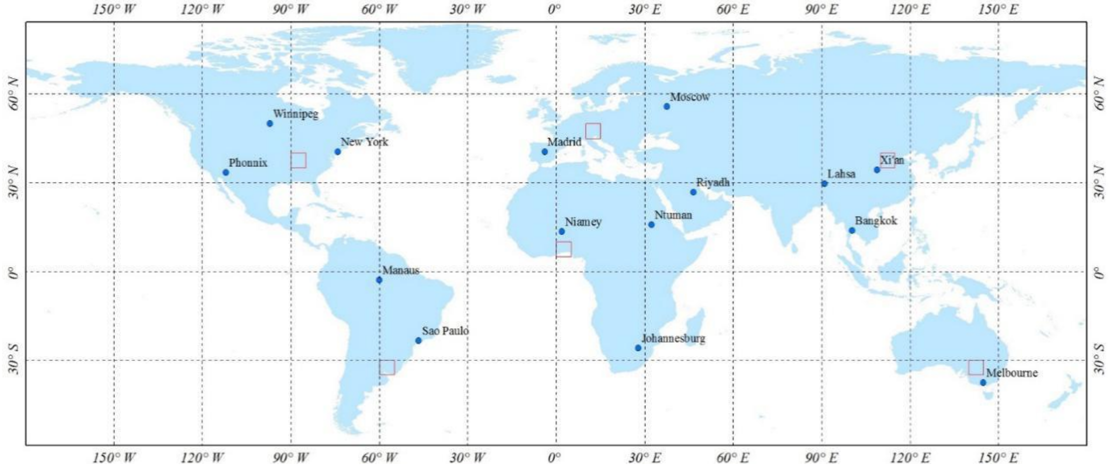


Figure 1: The spatial distribution of the fifteen validation regions (blue) corresponding to regions of different impervious landscapes on different continents together with the six 5 °×5 °validation regions (red) used to measure the variable importance.

Why did the authors select training samples based on Globe30 product but not FORM_GLC, which is also a 2015 product and seems to be more appropriate? Please elaborate.
Great thanks for the comment. The reasons why we chose the GlobeLand30 instead of FROM_GLC have been added 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."

Figure 5, please provide the label of axes.

Great thanks for the comment. The label of axes was added as:

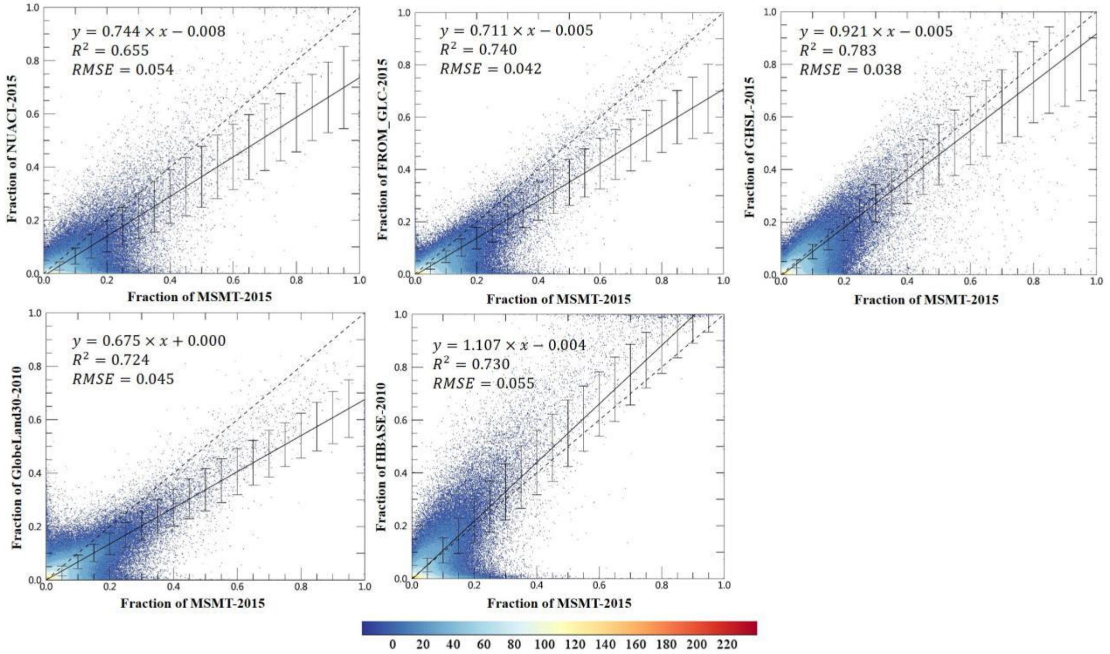


Figure 8: Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial grid of 0.05°×0.05°. The error bars were the standard deviation between reference datasets with fitted results.

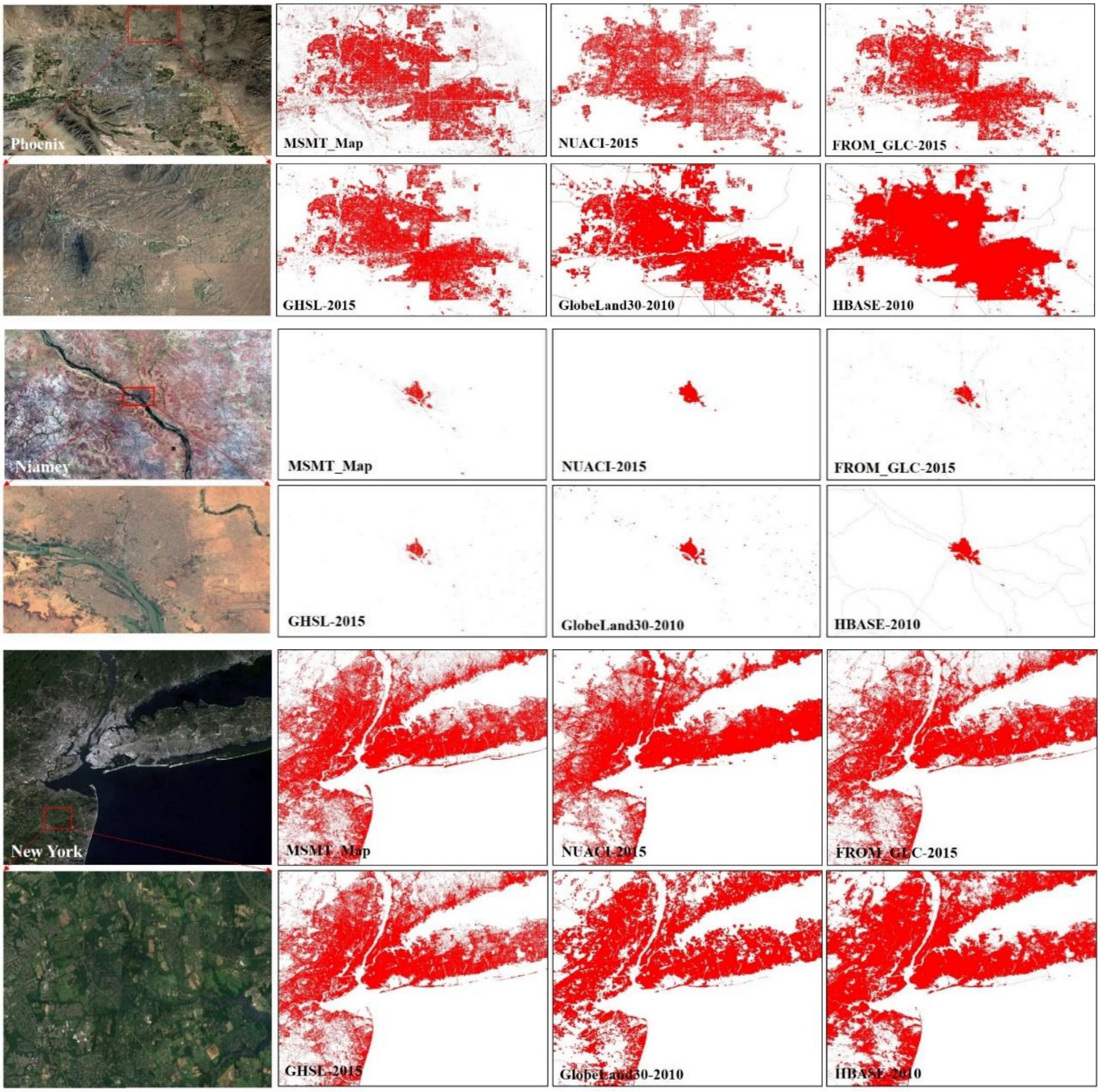
Table 2. How the different categories, e.g., high, low, medium, are defined? Are they defined quantitatively or subjectively? Please elaborate.

Great thanks for the comment. The impervious surface density (low, medium and high) was defined by combining the histogram of impervious areas at 0.05°×0.05°. In the revised manuscript, we re-selected the validation regions through the land-cover landscapes according to the suggestion of Reviewer 1. Specifically:

"To quantitatively assess the performance of the global impervious surface map, 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."

Figure 6. I suggest the authors to provide the location information (e.g., city name or latitude-longitude grid) of these areas. It will allow readers to check ground truth in Google Earth.

Great thanks for the comment. The city names were added as:



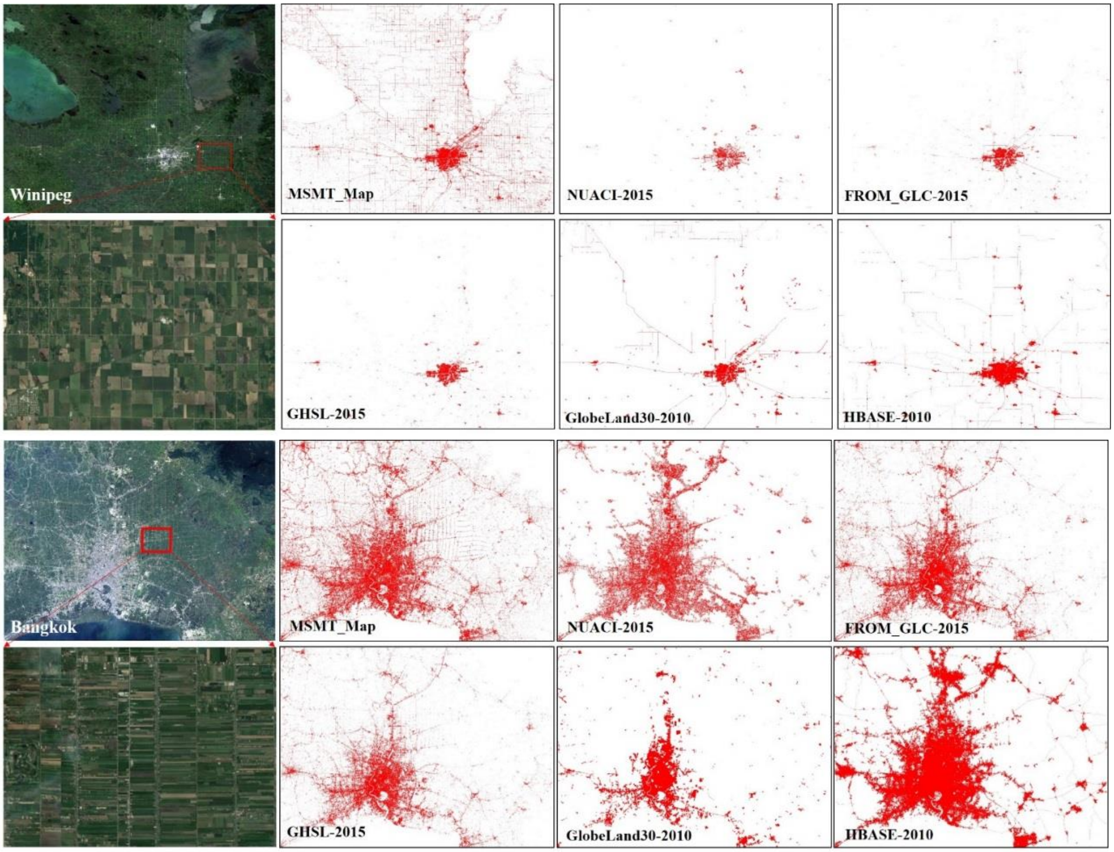


Figure 9: Comparisons between the MSMT_RF-based maps and other impervious surface products (corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019), the GlobeLand30 products developed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

Page 20, the authors found that the importance of Landsat textural features is low, whereas previous studies confirmed the contribution of textural features to impervious mapping. More explanations can be given on this contradiction. One possible explanation may be the different data sets. Many studies have indicated that textural information is helpful in land cover classification, especially in high-resolution images. Shaban and Dikshit (2001) used the textural information in SPOT images, while the authors used that in Landsat-8 images. The difference in spatial resolution may cause the different contribution of textural features in impervious surface mapping.

Great thanks for the comment. Actually, as the SAR backscatter and texture features also had ability to provide information on the structure and variability properties of surface materials, the importance of Landsat textural features was low. If only considering the optical Landsat imagery, the importance of Landsat textual features were significantly improved. This reasons have been added as:

"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."

Page 20, I agree with that the improvement made by this study is mainly due to the combination of the multi-source and multi-temporal information, but it may be misleading to state that the classification-based method performed better than spectral index-based method since they are performed based on the different data sets. I do not think the classification-based method can achieve a high accuracy only with Landsat data.

Great thanks for the comment. Yes, it was misleading to state that classification-based method performed better than the spectral index-based method, the improvement of mapping accuracy was mainly due to the combination of the multi-source and multi-temporal information. Therefore, we removed these misleading paragraph in the revised manuscript.

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

Xiao Zhang[1, 2], Liangyun Liu[1, 2], Changshan Wu[3], Xidong Chen[1, 2], Yuan Gao[1, 4], Shuai Xie[1, 2] and Bing Zhang[1, 2]

[1] State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Department of Geography, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[4] College of Geometrics, Xi'an University of Science and Technology, Xi'an 710054, China

Correspondence to: Liangyun Liu (liuly@radi.ac.cn)

Abstract. The amount of impervious surface is an important indicator in the monitoring of the intensity of human activity and environmental change. The use of remote sensing techniques is the only means of accurately carrying out global mapping of impervious surfaces covering large areas. Optical imagery can capture surface reflectance characteristics, while synthetic aperture radar (SAR) images can be used to provide information on the structure and dielectric properties of surface materials. In addition, night-time light (NTL) imagery can detect the intensity of human activity and thus provide important a priori probabilities of the occurrence of impervious surfaces. In this study, we aimed to generate an accurate global impervious surface map at a resolution of 30-m for 2015 by combining Landsat-8 OLI optical images, Sentinel-1 SAR images and VIIRS NTL images based on the Google Earth Engine (GEE) platform. First, the global impervious and non-impervious training samples were automatically derived by combining the GlobeLand30 land-cover product with VIIRS NTL and MODIS enhanced vegetation index (EVI) imagery. Then, based on global training samples and multi-source and multi-temporal imagery, a random forest classifier was trained and used to generate corresponding impervious surface maps for each 5°×5° cell of a geographical grid. Finally, a global impervious surface map, produced by mosaicking numerous 5°×5° regional maps, was validated by interpretation samples and then compared with ~~three~~ five existing impervious products (GlobeLand30, FROM_GLC, ~~and~~ NUACI, HBASE and GHSL). The results indicated that the global impervious surface map produced using the proposed multi-source, multi-temporal random forest classification (MSMT_RF) method was the most accurate of the maps, having an overall accuracy of ~~96~~95.1~~6~~% and kappa coefficient of 0.~~903~~ 898 as against 85.6% and 0.695 for NUACI, ~~92.5~~89.6% and 0.~~769~~ 780 for FROM_GLC, 90.3% and 0.794 for GHSL, 88.4~~91.1~~% and 0.~~717~~ 753 for GlobeLand30, and 88.0% and 0.745 for HBASE using all 15 regional validation data ~~87.43% and 0.585 for NUACI~~. Therefore, it is concluded that a global 30-m impervious surface map can accurately and efficiently be generated by the proposed MSMT_RF method based on the GEE platform. The global impervious surface map generated in this paper are available at https://doi.org/10.5281/zenodo.3505079 (Zhang and Liu, 2019).

1 Introduction

Impervious surfaces are usually covered by anthropogenic materials which prevent water penetrating into the soil (Weng, 2012)~~. Impervious surfaces include~~, which are primarily composited by asphalts, sand and stone, concrete, bricks, glasses, etc. (Chen et al., 2015). Due to the rapid growth in the area covered by impervious surfaces, a series of climate, environmental and social problems are emerging, including the urban heat island, traffic congestion, waterlogging and the deterioration of urban ~~the~~ environment (Fu and Weng, 2016; Gao et al., 2012; Weng, 2001; Zhou et al., 2017; Zhuo et al., 2018). Furthermore, as an important indicator in the monitoring of the intensity of human activity and of ecological and environmental changes, the mapping of impervious surfaces is of great interest in many disciplines (Xie and Weng, 2017). Accurate large-area impervious surface mapping is, therefore, urgent and necessary.

Due to the frequent and large-area coverage that it provides, increasing attention has been paid to the use of remote sensing technology for impervious surface mapping. In recent years, a lot of effort has gone into mapping impervious surfaces at different spatial resolutions (Elvidge et al., 2007; Schneider et al., 2010; Schneider et al., 2009). For example, Schneider et al. (2010) used multi-temporal MODIS data to produce a 500-m global urban land map, achieving an overall accuracy of 93% and kappa coefficient of 0.65. Elvidge et al. (2007) combined the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) and LandScan population count data to produce a 1-km global impervious surface area map. 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).

Recently, with the advent of free medium-resolution satellite data (e.g. Landsat and Sentinel-2), combined with rapidly-increasing data-storage and computation capabilities, many regional or global fine-resolution impervious surface maps have been produced using Landsat and Sentinel-2 images (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). Specifically, the National Land Cover Dataset (NLCD) produced the first 30-m map of the United States including impervious surface as ~~a~~ 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 ~~and GlobeLand30 products also~~ 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~~, focus on the overall accuracy of the mapping of all land-cover types rather than that of impervious surfaces alone.~~ the impervious surface layers usually suffered low accuracy except for GlobeLand30 (which includes manual interpretation). ~~Latterly~~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 ~~for~~ to ~~producing~~ 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.

There are three critical challenges for global impervious surface mapping at medium spatial resolution. These include finding an adequate image identification method, image selection scheme and image processing platform (Liu et al., 2018).

First, although a wide range of methods have already been presented for impervious surface mapping, it is still hard to generate an operational and accurate global impervious surface map at 30-m resolution. The methods used so far can be divided into three main groups: spectral mixture analysis methods (Ridd, 1995; Wetherley et al., 2017; Wu, 2004; Wu and Murray, 2003; Yang and He, 2017; Zhuo et al., 2018), spectral index-based methods (Deng and Wu, 2012; Liu et al., 2018; Xu, 2010), and image classification methods (Chen et al., 2015; Okujeni et al., 2013; Zhang et al., 2018a; Zhang et al., 2012; Zhang and Weng, 2016). The spectral mixture analysis methods have great advantages in terms of the repeatable and accurate extraction of quantitative sub-pixel information (Weng, 2012). However~~,~~, these spectral mixture methods can produce underestimates in areas with high density impervious surfaces and overestimates in areas with low density impervious surfaces~~these spectral mixture methods can produce underestimates in areas whether the density of impervious surfaces is high and overestimates in areas of low density~~, and may have great difficulties to identify one suitable endmember to represent all types of impervious surfaces (Sun et al., 2017; Weng, 2012). The spectral index-based methods have been widely applied in regional impervious surface mapping due to their simplicity, flexibility and convenience (Liu et al., 2018; Sun et al., 2019b; Xu, 2010). However, the spectral index-based methods have great difficulty in finding the optimal threshold for separating the impervious pixels from bare areas and vegetation pixels (Sun et al., 2017). The image classification methods can efficiently combine remote sensing datasets from multiple sources (Zhang et al., 2016; Zhang et al., 2018a; Zhou et al., 2017) and have great capabilities in spectrally complex impervious surface mapping (Okujeni et al., 2013), which has been an area of great interest in recent years (Goldblatt et al., 2018; Zhang et al., 2018b). However, it is very hard to select training samples for large-area impervious surface mapping using these methods (Weng, 2012).

Second, although individual optical data sets have been successfully employed for regional or global impervious surface mapping, accurate estimation of impervious surfaces remains challenging due to the diversity of urban land-cover types, which leads to difficulties in separating different land-cover types with similar spectral signatures (Zhang et al., 2014b). The incorporation of multi-source and multi-temporal remote sensing imagery has been demonstrated to improve impervious surface mapping accuracy (Weng, 2012; Zhu et al., 2012). For example, optical imagery is only able to capture surface reflectance characteristics, while synthetic aperture radar (SAR) data can provide details of the structure and dielectric properties of the surface materials (Sun et al., 2019b; Zhang et al., 2014b; Zhu et al., 2012). Zhang et al. (2016) found that the addition of dual-polarimetric SAR features resulted in an accuracy improvement of 3.5% compared with using optical SPOT-5 imagery only and also that dual-polarimetric SAR data had a superior performance to single polarimetric SAR data for impervious mapping. Similarly, Shao et al. (2016) explained that the combination of GaoFen-1 optical imagery with Sentinel-1 SAR imagery efficiently reduced the confusion between impervious surfaces and water and bare areas. Furthermore, Zhu et al. (2012) found that the inclusion of multi-seasonal imagery increased the mapping accuracy from 77.96% to 86.86% and that the further addition of texture variables increased the mapping accuracy to 92.69% for urban and peri-urban land-cover classification. The reasons for the accuracy increase were that the texture imagery could capture the local spatial structure and the variability in land cover categories and also that the temporal information could describe the phenological variability. Schug et al. (2018) also used the multi-seasonal Landsat imagery to successfully map impervious extent and land cover fractions. In addition, as an important data source for the measurement of socioeconomic activities, DMSP-OLS night-time light (NTL) imagery have been widely used in many impervious-related applications (Li and Zhou, 2017). For example, Elvidge et al. (2007) successfully produced a global 1-km impervious map using DMSP-OLS NTL imagery, Goldblatt et al. (2018) combined DMSP-OLS NTL and Landsat-8 imagery to accurately produce 30-m impervious surface maps at a national scale. Therefore, the integration of multi-source and multi-temporal datasets is necessary and crucial to the production of accurate global impervious surface maps.

Lastly, the mapping of impervious surfaces at the global scale usually requires huge amounts of computation and large storage capabilities. Fortunately, the Google Earth Engine (GEE) cloud-based platform consists of a multi-petabyte analysis-ready data catalog co-located with a high-performance, intrinsically parallel computation service (Gorelick et al., 2017), meaning that the requirements for large-area image collection and very large computational resources can easily be met by using the free-access GEE cloud-computation platform. For example, Liu et al. (2018) produced multi-temporal global impervious surface maps and Pekel et al. (2016) developed global high-resolution surface water maps and analyzed long-term changes using the GEE cloud-computation platform. Recently, Massey et al. (2018) produced a continental-scale cropland extent map for North America at 30 m spatial resolution for the nominal year 2010 based on the GEE platform. It can be seen, therefore, that the GEE is an efficient and useful computation platform for regional/global applications.

So far, due to the limitations of data collection and computation capability, impervious surface mapping has mainly focused on using a single type of remote sensing data or on case studies made at the regional scale. Although the GEE platform provides multi-petabyte analysis-ready data and efficient data-processing capabilities, an efficient method that can fully integrate these multi-source and multi-temporal datasets and produce accurate impervious surface maps at a spatial resolution of 30-m for the whole world is still lacking. The aims of this study, therefore, were (1) to produce a global 30-m impervious surface map from multi-source multi-temporal remote sensing datasets including Landsat-8 OLI, Sentinel-1 SAR, VIIRS NTL and MODIS imagery using the GEE platform; (2) to investigate the accuracy of the global 30-m impervious surface mapping using validation samples and cross-comparison with ~~three~~ five existing impervious surface products (GlobeLand30 (Chen et al., 2015), FROM_GLC (Gong et al., 2013), ~~and~~ NUACI (Liu et al., 2018), GHSL (Florczyk et al., 2019) and HBASE (Wang et al., 2017a)). The results indicate that the global impervious surface map produced by the proposed method is accurate and is suitable for regional or global impervious surface applications.

2 Datasets

2.1 Remote sensing datasets

In this study, ~~five~~ 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 ~~imagery~~products was used to derive the set of global impervious surface and non-impervious surfaces training data. ~~, as well as STRM/ASTER DEM topographical variables, were selected and collected for the mapping of impervious surfaces across the world using the GEE platform.~~

All available Landsat-8 surface reflectance (SR) imagery from 2015 and 2016(USGS, 2015), which had been archived on the GEE platform, were used in this study for the nominal year 2015 because of the frequent cloud contamination in the tropic areas. All the SR images were radiometrically corrected by the Landsat Surface Reflectance Code (LaSRC) atmospheric correction method (Hu et al., 2014; Vermote et al., 2016), and bad pixels including clouds, cloud shadow, and saturated pixels were identified by the CFMask algorithm (Guide, 2018).

The Sentinel-1 satellite provides C-band SAR imagery at a variety of polarizations and resolutions~~, and the repeat cycle of the polar-orbiting two-satellite constellation is 6 days~~ (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, all available Sentinel-1 imagery from 2015 and 2016, which had already been calibrated and ortho-corrected then archived on the GEE platform, were also used for the nominal year 2015. In addition, each Sentinel-1 image on the GEE had been pre-processed with the Sentinel-1 Toolbox, including thermal noise removal, radiometric calibration and terrain correction (https://developers.google.com/earth-engine/sentinel1). Also, as HH- and HV-polarized Sentinel-1 SAR imagery does not cover the whole world (Sun et al., 2019a), a combination of dual-band cross-polarized (VV and VH) Interferometric Wide Swath (IW) mode imagery in both 'ascending' and 'descending' orbits was used. The spatial resolution of this imagery was 10-m and the repeat cycle of the polar-orbiting two-satellite constellation is 6 days.

The ~~last data source used was 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). This dataset has undergone a void-filling process using other open-source data (ASTER GDEM2, GMTED2010 and NED) in the GEE platform. As for the high-latitude areas that lacked the SRTM data, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) (Tachikawa et al., 2011) was used instead.

The VIIRS NTL, collected by NASA/NOAA’s Suomi National Polar-orbiting Partnership satellite (https://maps.ngdc.noaa.gov/viewers/VIIRS_DNB_nighttime_imagery/index.html), has the unique ability to record emitted visible and near-infrared (VNIR) radiation at night with a spatial resolution of 15 arc seconds (equivalent to 0.5 km at the equator) (Elvidge et al., 2017). Compared to the DMSP-OLS NTL data, the VIIRS NTL data provide higher spatial resolution, and finer radiometric resolution, which allows weaker surface radiation to be detected (Bennett and Smith, 2017). It is also the main data source used for studying the expansion of impervious surfaces and related sociodemographic issues (Elvidge et al., 2017). In this study, a combination of VIIRS NTL, MODIS EVI imagery and GlobeLand30 land-cover products was used to derive the set of global training samples.

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~~. The imagery~~ 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 EVI imagery for 2015 in the GEE used the blue band to remove residual atmospheric contamination caused by smoke and sub-pixel thin clouds (https://developers.google.com/earth-engine/datasets/catalog/ MODIS_006_MYD13Q1).

~~The last data source used was 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 . This dataset has undergone a void-filling process using other open-source data (ASTER GDEM2, GMTED2010 and NED) in the GEE platform. As for the high-latitude areas that lacked the SRTM data, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model Version 2 (GDEM V2) was used instead.~~

2.2 Global ~~Global i~~impervious surface products

In this study, ~~three~~ five global impervious surface products ~~−~~ (GlobeLand30, FROM_GLC, ~~and~~ NUACI, HBASE and GHSL) − were used to validate the global impervious surface map produced using the MSMT_RF method. The GlobeLand30 data were also used to automatically derive the global impervious and non-impervious training samples.

GlobeLand30 is an operational 30-m global land-cover dataset produced using the Pixel-Object-Knowledge-based method (POK-based) approach in 2000 and 2010 (Chen et al., 2015). In this study, the global impervious product derived from GlobeLand30 in 2010 (GlobeLand30-2010, http://www.globallandcover.com/GLC30Download/index.aspx) was ~~selected~~ produced by combining pixel-based classification, multi-scale segmentation and manual editing based on the high resolution imagery~~(http://www.globallandcover.com/ GLC30Download/index.aspx). The impervious surface land-cover type in GlobeLand30-2010 has~~and had been validated as having a user's accuracy of 86.7%.

FROM_GLC, first produced in 2010, was the first 30-m resolution global land-cover dataset and was produced by supervised classification of 8,900 Landsat images (Gong et al., 2013). In this study, the second generation of FROM_GLC from 2015 (FROM_GLC-2015) (http://data.ess.tsinghua.edu.cn/) was used. This dataset was produced by using multi-seasonal Landsat imagery acquired between 2013 and 2015 and incorporates the day of year, geographical coordinates and elevation data (Li et al., 2017).

The NUACI-based maps, developed using the spectral index-based method applied to Landsat and DMSP-OLS NTL imagery, are multi-temporal global 30-m impervious surface datasets (Liu et al., 2018). In this study, the NUACI impervious map from 2015 (NUACI~~–~~-2015) was used (http://www.geosimulation.cn/ GlobalUrbanLand.html). This map has been validated as having an overall accuracy of 0.81–0.84 and kappa coefficient of 0.43–0.50 at the global level (Liu et al., 2018).

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).

2.3 Validation samples

To quantitatively assess the performance of the global impervious surface ~~map~~datasets, ~~twelve~~ fifteen~~1 ° × 1 °~~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)),~~ (blue rectangles), including five high-density, four medium-density and three low-density regions,~~ were ~~randomly~~ selected ~~based on the density and distribution of the impervious surfaces~~ (Fig. 1). For each validation region, 600-1000 samples were randomly generated using the stratified random sampling strategy (Bai et al., 2015). As there were significant advantages to using Google Earth for validation sample selection (Zhang et al., 2018c), each sample was labeled either as "non-impervious surface" or "impervious surface" based on visual interpretation of the available high-resolution remote sensing imagery in Google Earth. ~~In addition, T~~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. Like the work of Sun et al. (2019b), if the impervious area in the 30-m ×30-m validation window was more than a predefined threshold of 50%, we will consider this validation point as impervious surface, otherwise, it would be labeled as non-impervious surface. After careful interpretation, a total of ~~10~~11,~~142~~ 942 samples including ~~2381~~ 4952 impervious samples and 69~~7761~~ 90 non-impervious samples were obtained. In order to minimize the subjective influence of interpretation, the validation samples were collected independently by three different scientists. If there was dispute between the interpretation results of three scientists, the validation point was discarded.

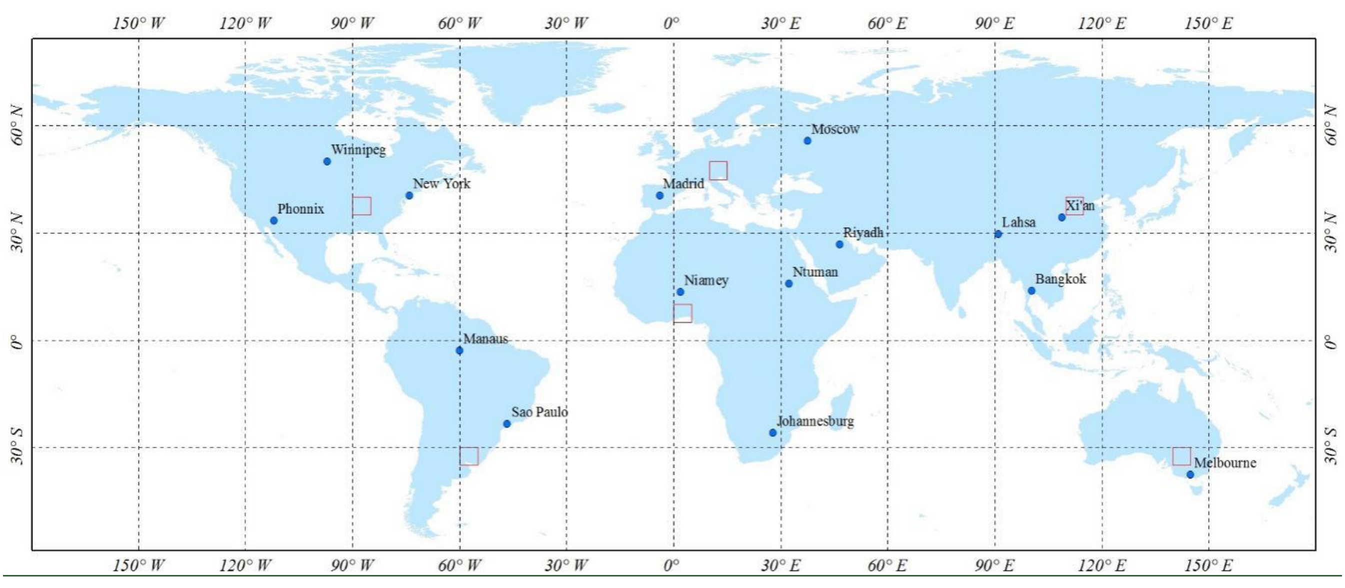


Figure 1: The spatial distribution of the ~~twelve~~ fifteen~~1 °×1 °~~validation regions (blue) corresponding to regions of different~~ impervious surface density~~ impervious landscapes on different continents together with the six 5 °×5 °validation regions (red) used to measure the variable importance.

~~3 Methods~~

~~**To develop the global 30-m impervious surface map for 2015, the multi-source and multi-temporal random forest classification (MSMT_RF) method was proposed. The method is illustrated in Fig. 2. First, time series of Landsat-8 SR and Sentinel-1 SAR imagery archived on the GEE platform were collected. Secondly, the temporal–spectral–textural features and temporal–SAR features were derived from the Landsat-8 and Sentinel-1 imagery using image compositing methods. Thirdly, based on the GlobeLand30-2010 impervious surface products, and the VIIRS NTL and MODIS EVI imagery, the global impervious and non-impervious training samples were automatically generated. The random forest classifier was trained at each 5°×5° geographical grid cell using the temporal–spectral–textural–SAR-topographical features and the global training samples. Finally, the global impervious surface map was compared with three existing impervious surface products and further validated using the visual interpretation samples.**~~

3.~~1~~ ~~Derivation~~ Collection of global training samples ~~from GlobeLand30-2010~~

As the reliability and representativeness of the training samples would affect the classification accuracy directly (Foody and Mathur, 2004), we proposed ~~using the multi-source datasets~~combining GlobeLand30, VIIRS NTL and MODIS EVI data to derive accurate impervious and non-impervious samples. 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. However, as these was temporal interval of 5 years between GlobeLand30 and our study, i~~I~~t was 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.

~~First~~Specifically, as GlobeLand30 used an object-based labeling method to remove the "salt-and-pepper effect" caused by the pixel-based classification method (Chen et al., 2015), the impervious surfaces consisted of independent blocks. Usually, a large number of mixed pixels and misclassifications occur at the boundary of image blocks or objects, and Yang et al. (2017) also found that GlobeLand30 exhibited higher accuracy in homogeneous areas. The land-cover heterogeneity was calculated as the number of land-cover types occurring in a local window (Jokar Arsanjani et al., 2016). According to the statistics of Chen et al. (2015), there were a little commission and omission errors in each scene when the area of impervious surface block was less than 8×8. In this study, the local window size was set to 9×9 after balancing the sample reliability and completeness because the higher window size would cause the candidate samples miss those small and broken impervious objects (such as: rural villages). Therefore, if the land-cover heterogeneity in the 9×9 local window was greater than 1 (meaning that the land-cover types within the window consisted of both impervious and non-impervious types), the center pixel was removed from the candidate training point set (CanTPS_Imp).

Secondly, to minimize the effects of mapping error in GlobeLand30-2010 and temporal interval between GlobeLand30-2010 and the input imagery for training samples in CanTPS_Imp, the VIIRS NTL data, revealed the intensity of socioeconomic activities, was imported to refine each training point in 2015. However, as the coarse spatial resolution of VIIRS NTL imagery might cause a 'blooming effect' in suburban areas, the EANTLI proposed by Zhuo et al. (2018) was applied to reduce the blooming effects:

$$EANTLI = \frac{1+(NTL_{norm}-EVI)}{1-(NTL_{norm}-EVI)} \times NTL \,, \qquad (1)$$

where NTL_{norm} is the normalized NTL value, EVI is the annual mean value of the time-series MODIS EVI products and NTL is the actual value of the VIIRS NTL data.

The EANTLI measured the likelihood of the pixel corresponding to an impervious surface, so it was reasonable to assume that the pixels where EANTLI exceeded a certain threshold were impervious surface pixels. In this study, as the candidate training points in CanTPS_Imp were collected from homogenous 9×9 pixel areas (270 m×270 m), the EANTLI image in 2015 (EANTLI-2015) was first resampled to the 270 m to match with these candidate points. The GlobeLand30-2010 impervious surface map had a user's accuracy of 86.7%, and we assumed that the process of transforming non-impervious surfaces into impervious surfaces was irreversible during the period 2010 to 2015, so the impervious segmentation threshold was selected as being the lowest 15th quantile of the cumulative probability of all candidate impervious points for EANTLI-2015; namely, if the cumulative probability of the impervious point in CanTPS_Imp was lower than the threshold, the candidate point was removed from CanTPS_Imp. 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 ~~reflect the~~have high EANTLI~~-2015~~ 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.

Lastly, although the candidate training points were refined using the GlobeLand30~~-2010~~ land-cover product and EANTLI-2015 imagery, the volume of candidate training points was still huge and so it was necessary to further resample the CanTPS_Imp. ~~Furthermore,~~ As the non-impervious surfaces consisted many land-cover types (water, vegetation, cropland and bare soil)~~.~~ 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, ~~Based on the work of , who investigated the impact of training sample selection on the impervious classification accuracy, the~~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~~ proportional~~ 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). In addition, as the land-cover distribution varied with geographical region, the stratified random sampling strategy was applied at every 5°×5° geographical grids to ensure the training samples locally adaptive.~~ resampling method was chosen for use in this study. Furthermore, the non-impervious surfaces consisted of many land-cover types (water, vegetation, cropland and bare area). As the bare land was easier to be confused with the impervious surface compared with the water and vegetation types especially in the cities with rapid urbanization , and the suburban areas or rural villages were also easy to confused with croplands , the non-impervious surfaces were spited into three independent groups including: bare area, cropland and other land-cover types. Then, in order to guarantee the rationality of training samples, the stratified random sampling strategy was applied at every 5°×5° geographical grids. Finally, the approximate value of 1:3 was used to represent the proportion of impervious to non-impervious surfaces (bare area, cropland and other land-cover types).~~ Using~~ the proportional resampling method and~~ the stratified random sampling strategy with the uniform distribution, a total of 4,483,000 training samples, including 3,499,000 non-impervious samples and 984,000 impervious samples, were collected over the land areas across the globe.

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

~~3.2~~4 Multi-source and multi-temporal impervious classification method

To develop the global 30-m impervious surface map for 2015, the multi-source and multi-temporal random forest classification (MSMT_RF) method was proposed. The method is illustrated in Fig. 2. First, time series of Landsat-8 SR and Sentinel-1 SAR imagery archived on the GEE platform were collected. Secondly, the temporal–spectral–textural features and temporal–SAR features were derived from the Landsat-8 and Sentinel-1 imagery using image compositing methods. Thirdly, based on the global training samples derived from GlobeLand30-2010, VIIRS NTL and MODIS EVI imagery, the random forest classifier was trained at each 5 °×5 ° geographical grid cell using the temporal–spectral–textural–SAR-topographical features. Finally, the global impervious surface map was compared with existing impervious surface products and further validated using the visual interpretation samples.

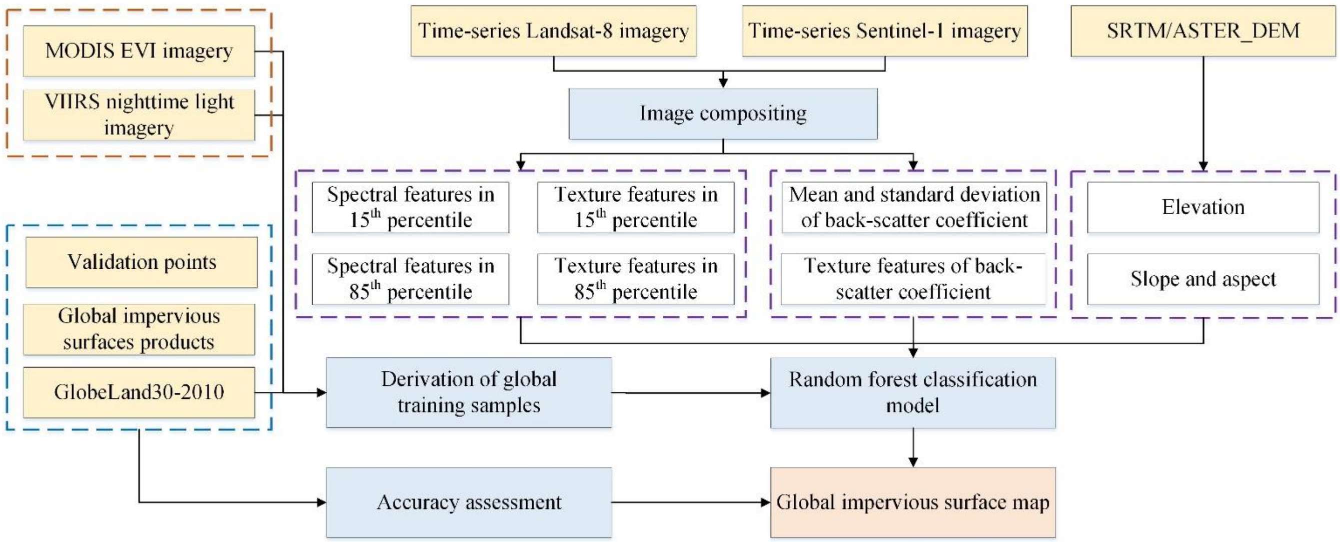


Figure 2: Flowchart illustrating the MSMT_RF method.

~~3.2~~4.1 Multi-source and multi-temporal feature selection

As mentioned above, the datasets used in this study had been acquired from various satellite sensors and had distinctive features. Also the incorporation of multi-source and multi-temporal remote sensing data has been demonstrated to improve the accuracy of the mapping of impervious surfaces. In this study, three kinds of satellite imagery, including Landsat-8 SR, Sentinel-1 SAR and SRTM/ASTER DEM imagery, were collected for the global classification of impervious surfaces.

After masking out the bad pixels (cloud, shadow and saturated pixels), the time-series Landsat SR imagery were needed to reduce the number of dimensions of the temporal–spectral features to guard against the Hughes phenomenon (Zhang et al., 2019). Similar to what Hansen et al. (2014) and Zhang and Roy (2017) introduced to capture phenology, the 15th and 85th percentiles of Landsat SR were used instead of the minimum and maximum values to minimize the effects of residual shadows and cloud caused by the errors in the CFMask method (Massey et al., 2018). In addition, as ~~the~~ 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 combination of 15th and 85th percentiles of Landsat SR was used to efficiently capture intra-annual variation information of various land-cover types. It should be noted that only the six optical bands (Blue, Green, Red, NIR, SWIR1 and SWIR2) were selected because the Coastal band was sensitive to the atmospheric scattering (Wang et al., 2016). Liu et al. (2018) found that the Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI) and Normalized

Difference Built-up Index (NDBI) were of great help in impervious surface identification; therefore, these three spectral indexes were added to the spectral features, giving a total of 18 features for the two-epoch imagery. Furthermore, as the texture information contributed to the classification performance (Weng, 2012), the local textural measures based on the Gray Level Co-occurrence Matrix (GLCM) were adopted; however, because of the redundancy and similarity between texture features (Rodriguez-Galiano et al., 2012), only the variance, dissimilarity and entropy of the NIR band were selected from the 7×7 local window for the two-epoch imagery (Chen et al., 2016; Zhang et al., 2014b). The optimal window size for texture measurements was highly dependent on the image spatial resolution and the land cover characteristics (Zhu et al., 2012) and Shaban and Dikshit (2001) computed texture measurements with different window sizes as inputs for urban area classification and suggested window sizes of 7 ×7 pixels perform best.

As the Sentinel-1 SAR imagery had been pre-processed in the GEE platform, the annual mean and standard deviation of the VV and VH imagery were directly derived from the time-series of Sentinel-1 SAR imagery. Zhang et al. (2014b) found that SAR texture features were also relevant to impervious surfaces and the dissimilarity, variance and entropy features of the VV and VH imagery were identified as effective indicators for the texture description of different urban land cover types. As Zhang et al. (2014b) explained the window size for calculating GLCM should be smaller as terrains are smaller under coarser resolution, the window size was chose as 9×9 pixels at 10-m spatial resolution, equivalent to 3×3 pixels in 30-m. Moreover, as the spatial resolution of the Landsat SR (30-m) was three times that of the Sentinel-1 imagery (10-m), the SAR data were resampled to 30-m for integration with the Landsat SR data.

Lastly, as Sentinel-1 SAR imagery usually had high backscatter similar to the impervious surface over mountainous areas, terrain information were useful auxiliary for removing these false positive at these areas (Ban et al., 2015). Similarly, Clarke et al. (1997) found that terrain variables were of great help in identifying impervious surfaces because they usually located in the flat areas.~~,~~ In this study, the elevation, slope and aspect, calculated from the SRTM/ASTER DEM data, were added to the feature vector. This gave a total of 37 features for each pixel location, including 18 spectral features and 6 texture features from the Landsat imagery, 10 SAR features and 3 topographical variables. The features are listed in Table 1.

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)

~~3.2~~4.2 Random forest classification model

As in the work of Zhang and Roy (2017), there were two options for models to use in the global impervious surface classification: global and local models. The global model is a single classifier, trained using the global training samples, and then used to classify the entire global data set. The local model, is trained using regional samples; the regional classification results are then mosaicked to produce the global map. Zhang and Roy (2017) confirmed that locally adaptive models achieve a higher classification accuracy than a single global model. Therefore, the global land surfaces were split into approximately 1000 5°×5° geographical grid cells after considering the data volume and amount of computation needed for the regional mapping. In addition, to ensure the classification agreement ~~consistency~~ across the cell boundaries, as in the work of Zhang et al. (2018c) and Zhang and Roy (2017), the training samples from the adjacent 3×3 geographical cells were also imported to train the classifier in classifying the central geographical cell.

As for the specific techniques used in classifiers, according to our previous investigations (Zhang et al., 2019), the Random Forest (RF) classifier is more capable of handling high-dimensional multicollinearity data. It is also less affected by noise and feature selection as well as being more accurate and efficient than other widely used classifiers such as the SVM (Support Vector Machine), CART (Classification And Regression Tree) and ANN (Artificial Neural Network) classifiers. Therefore, the RF classifier was selected for the development of the global impervious surface map.

The RF classifier has only two parameters: the number of classification trees (Ntree) and the number of selected predication features (Mtry). Furthermore, many researchers have demonstrated that there is almost no correlation between these two parameters and the classification accuracy (Belgiu and Drăguț, 2016; Du et al., 2015; Gislason et al., 2006); therefore, the default values of 500 for Ntree and the square root of the total number of training features for Mtry were selected.

~~3~~4.3 Accuracy assessment

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~~To quantitatively assess the global consistency between the MSMT_RF-based impervious surface map and the three existing products (GlobeLand30-2010, FROM_GLC2015 and NUACI_2015)~~, ~~four~~ 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 ~~support~~ provide the quantitative ~~analysis~~metrics of the agreement~~. Two evaluation indicators~~ including~~–~~ ~~the~~ coefficient of determination (R^2) and ~~the~~ root mean square error (RMSE)~~— were calculated to measure the consistency~~.

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

~~4~~ 5 results

5.1 The importance of multi-source and multi-temporal features

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%.

~~Zhang et al. (2014b)~~

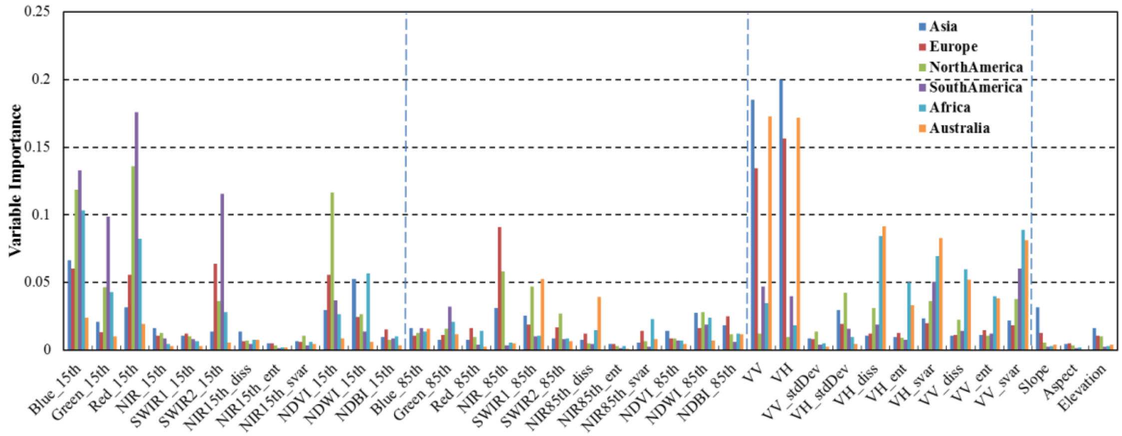


Figure 3: The importance of the input features derived from the random forest model using the training samples in six continental regions.

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 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.

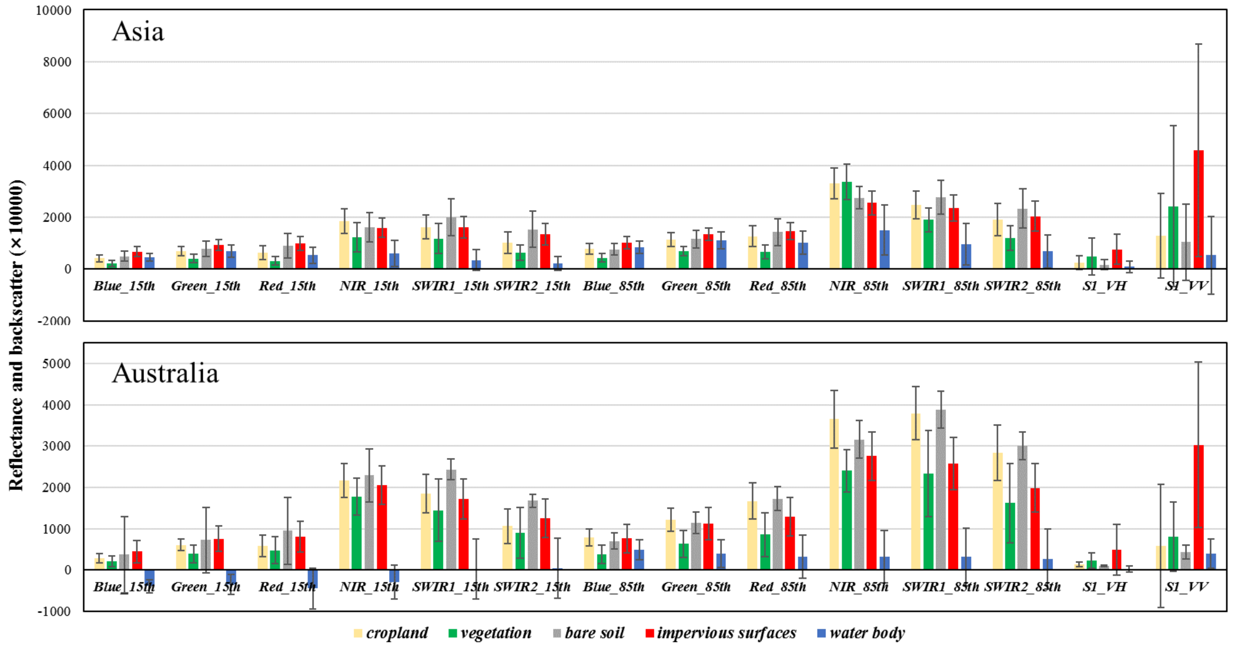


Figure 4: The reflectance/backscatter characteristics of different land-cover types over Landsat optical and Sentinel-1 SAR imagery in the Asia and Australia regions.

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.~~Zhu et al. (2012)Schug et al. (2018)~~ 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.

~~4~~5.~~1~~ 2 Global impervious surface map ~~and cross-comparison~~

The global distribution of the fraction of impervious area (FIA) at a spatial resolution of 0.05° is illustrated in Fig. ~~3~~5, whilst the meridional and zonal total FIA for each 0.05° longitude and latitude bin are shown at the top and left of Fig. 3. From an intuitive and statistical perspective, globally, impervious surfaces are mainly concentrated in three continents: Asia (34.43%), North America (28.04%) and Europe (24.98%), followed by South America (5.89%), Africa (5.63%) and Australia (1.06%). In addition, the zonal statistics indicate that 70% of the impervious surfaces are distributed between 30°N and 60°N because these regions contain the key areas of Asia, North America and Europe, which are the locations of the most developed countries and highest population densities. The meridional results illustrate that there are four peak intervals: 100°W to 70°W (United States), 10°W to 40°E (European Union), 60°E to 90°E (India) and 100°E to 130°E (China and southeastern Asia). The two peak values in the meridional direction are located in the centers of the United States and China.

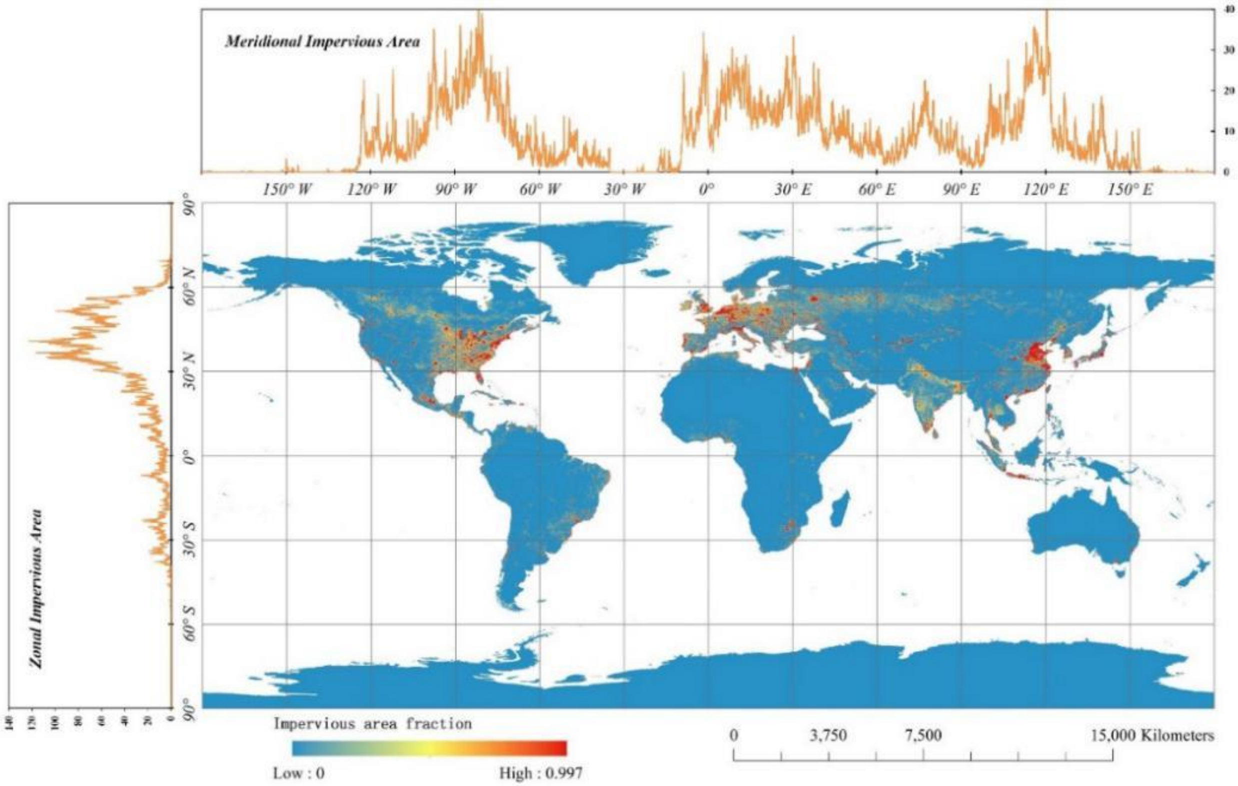


Figure ~~3~~5:~~.~~ Global fraction of impervious cover derived from multi-source and multi-temporal Landsat-8 SR and Sentinel-1 SAR imagery acquired in 2015 to 2016. The spatial resolution of the map is 0.05°.

Summaries of the impervious surface areas at a national scale were also produced. The statistical results indicated that the total impervious surface areas of the top 20 countries account for 75.96% of the total global area. Fig. ~~4~~ 6 presents the top 20 counties in terms of impervious surface area and corresponding fractions of the world total. Overall, there is a positive correlation between these statistical fractions and the land area, population and degree of economic development of these nations. Specifically, it was found that the U.S. has the biggest impervious surface area, accounting for more than 20% of the global total, and only the top 3 countries (U.S., China and Russia) exceed 5% of the total global area. The ranking is also basically consistent with the statistics produced by the Organization for Economic Co-operation and Development (OECD) for built-up areas in 2014 (https://stats.oecd.org/Index.aspx?DataSetCode=BUILT_UP).

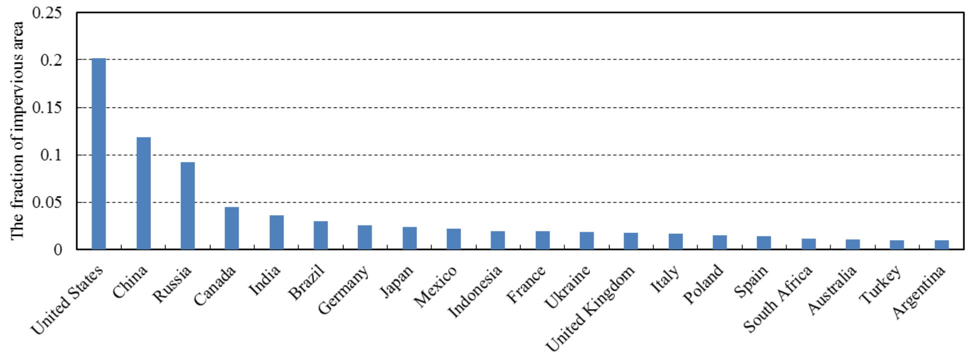


Figure ~~4~~6:~~.~~ The top 20 countries in terms of impervious surface area and corresponding fractions of the global total.

5.3 Spatial variations of global impervious products

To quantitatively ~~evaluate~~ analyse the spatial ~~consistency~~ agreement between the MSMT_RF-based impervious surface map and the ~~three~~ five existing products (GlobeLand30-2010, FROM_FLC-2015, ~~and~~ NUACI~~–~~-2015, GHSL-2015 and HBASE-2010), ~~four~~ 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.

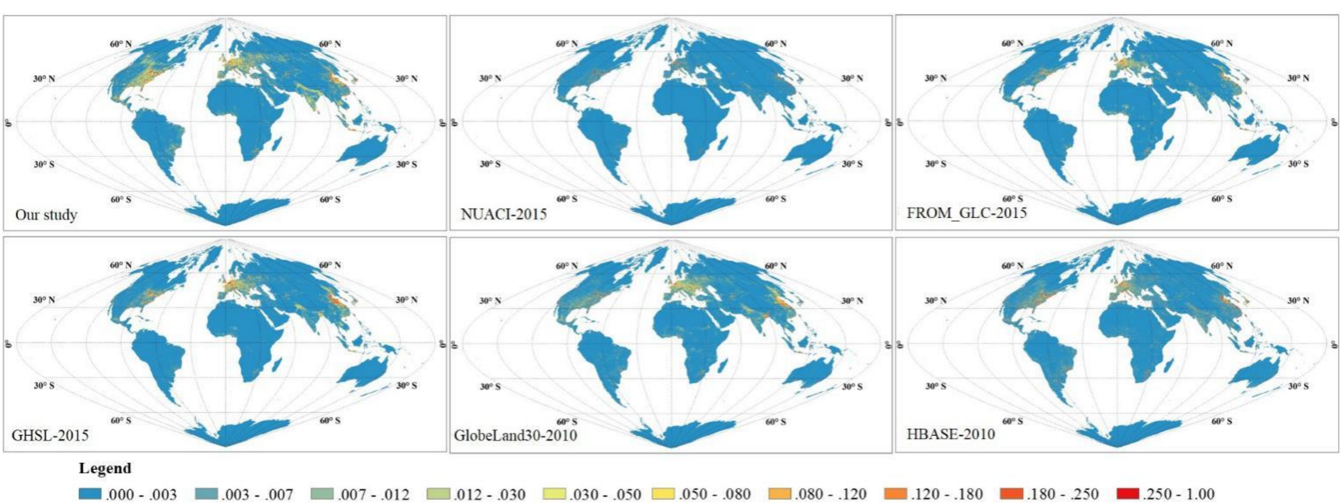
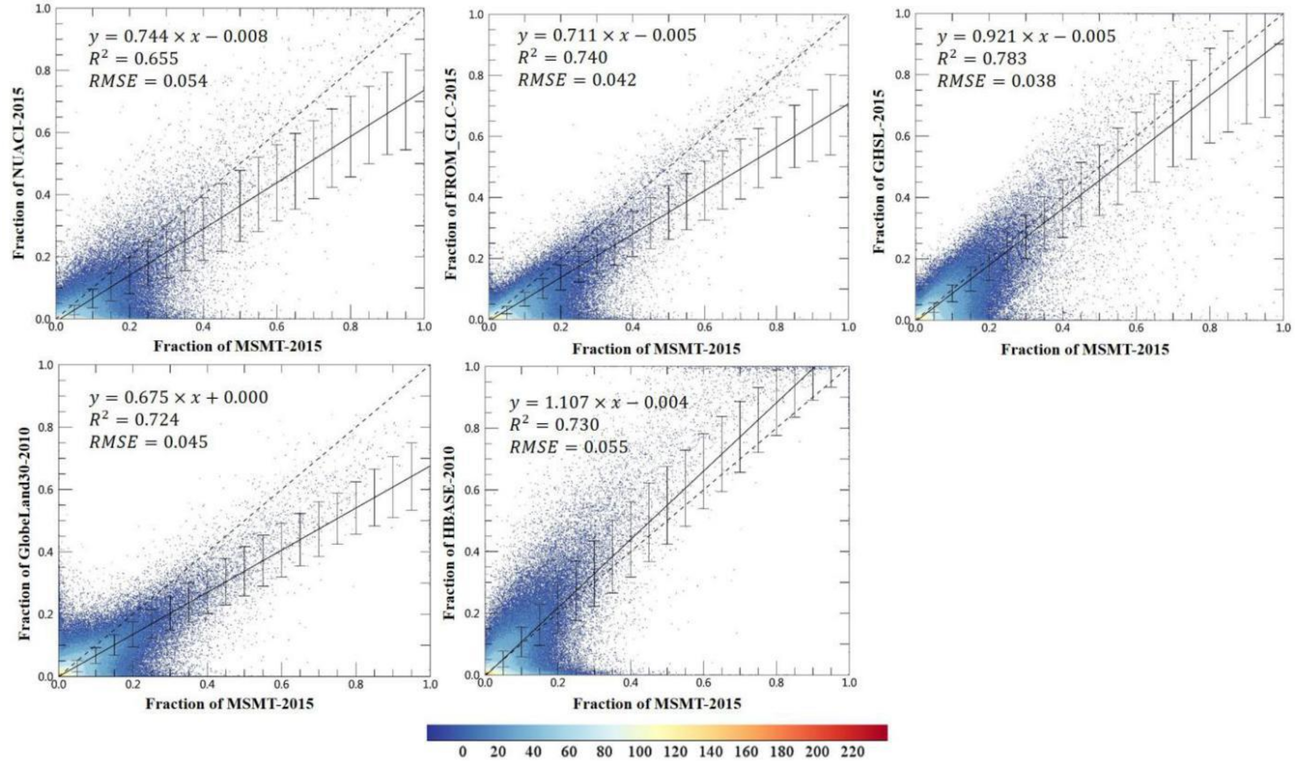


Figure 7: The spatial patterns of six global 30-m impervious products after aggregating to the resolution of 0.05 °.

Scatter plots of the ~~three~~ five products against the MSMT~~_RF-based~~-2015 impervious map were then made, as illustrated ~~as~~ in Fig. 8~~5~~. The results indicate that there ~~is~~ was a greater ~~consistency~~ agreement between the MSMT~~_RF-based~~-2015 map and ~~FROM_GLC~~GHSL-2015 (R^2 ~~of~~ =0.~~7696~~783, ~~and~~ RMSE ~~of~~ = 0.~~0546~~038 and slope=0.921)~~,~~ than for ~~GlobeLand30-2010 (R2 of 0.7416 and RMSE of 0.0604) or NUACI_2015 (R2 of 0.6631 and RMSE of 0.0996)~~other products. Specifically, ~~The main differences between the GlobeLand30 and the MSMT_RF-based maps are 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). As~~ 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~~ in this case~~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 ~~are~~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. ~~Lastly,~~ 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~~subsequent validation (Section 4.2) indicated that there was a high degree of consistency (close to 1:1) between FROM_GLC2015 and the MSMT_RF-based results in high-density regions but that the product suffered from the problem of underestimation in low and medium-density regions. The slope of the regression line for these results is also less than 1~~.

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.



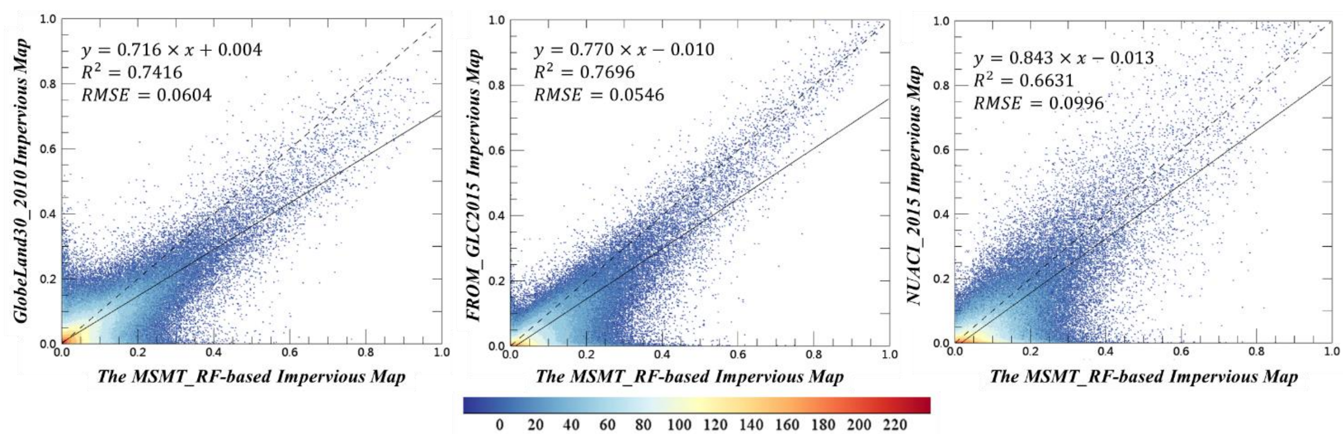


Figure ~~5~~8:~~.~~ Scatter plots between the MSMT_RF-based impervious map and the GlobeLand30-2010, FROM_GLC-2015, ~~and~~ NUACI~~-~~-2015, GHSL-2015 and HBASE-2010 global impervious surface products at a spatial ~~resolution~~ grid of 0.05 °×0.05 °. The error bars were the standard deviation between reference datasets with fitted results.

~~4~~5.~~2~~ 4 Accuracy assessment using validation samples

The accuracy of the ~~four~~ five global impervious surface maps over ~~12~~ 15 validation regions with different impervious ~~densities~~ 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.~~966~~ 951 and kappa coefficient of 0.~~903~~ 898 compared with 0.~~925~~ 896 and 0.~~769~~ 780 for FROM_GLC-2015, 0.~~873~~ 856 and 0.~~585~~ 695 for NUACI~~_~~-2015, 0.903 and 0.794 for GHSL-2015, ~~and~~ 0.~~911~~884 and 0.7~~17~~53 for GlobeLand30-2010, and 0.880 and 0.754 for HBASE-2010 ~~-~~using all ~~12~~ 15 regional validation data.

From the perspective of the value of the ~~producer's~~ user's accuracy for impervious surfaces, the MSMT_RF method performed ~~significantly~~ better than the other impervious surface products (meaning lower omission error) achieving the accuracy of 0.932, especially in the ~~low- and medium-density~~cropland-prevalent and vegetation-prevalent impervious ~~regions~~landscapes (such as: Bangkok, Winnipeg, Xi'an…). Specifically, ~~as the minimum mapping unit of GlobeLand30 was a 4×4-pixel area, many rural impervious surfaces were ignored in these low and medium-density regions, which caused large omission errors that ranged from 73.6% to 12.4%.~~ NUACI-~~_~~2015 had the lowest user's ~~producer'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 ~~to~~ with the MSMT_RF method for ~~high-density~~ ~~regions~~big cities (such as New York, Moscow and Johannesburg), but its accuracy decreased sharply over ~~low- and medium-density~~'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%.~~Moreover, we could found that the producer's accuracy for impervious surfaces and kappa coefficient varied with the impervious density, specifically, the higher accuracies were achieved at high impervious density regions followed by medium and low impervious density regions. 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 ×30-m validation window and the impervious areas exceed the predefined threshold of 50%) which were most difficult to identify for impervious surface mapping.~~

As for the ~~user's~~ producer's accuracy for impervious surface (measuring the commission error)~~s, and the producer's accuracy and user's accuracy for non-impervious surfaces~~, the ~~MSMT_RF~~GHSL-2015 products performed best and ~~method~~ 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. ~~an accuracy similar to that achieved by FROM_GLC2015 and GlobeLand30-2010 and higher than that of NUACI_2015. In contrast to the results for the producer's accuracy for impervious surfaces, the three other products mostly performed very well as measured by these three metrics, especially in low- and medium-density regions.~~

~~Table 2. Accuracy of the four impervious surface maps over 12 validation regions~~

		a	b	c	d	e	f	g	h	i	J	k	l	Overall
	I.D.	high	low	low	high	medium	medium	low	high	high	medium	high	medium	
The MSMT_RF map	P.I.	0.885	0.822	0.847	0.938	0.757	0.782	0.765	0.905	0.990	0.941	0.969	0.845	0. 900
	U.I.	0.955	0.961	1.000	0.944	0.858	0.991	1.000	0.987	0.940	0.923	0.948	0.990	0. 952
	P.N.	0.979	0.994	1.000	0.971	0.974	0.999	1.000	0.996	0.985	0.980	0.962	0.999	0. 986
	U.N.	0.944	0.969	0.975	0.969	0.951	0.958	0.965	0.970	0.998	0.985	0.977	0.974	0. 970
	O.A.	0.948	0.968	0.978	0.968	0.937	0.960	0.968	0.974	0.986	0.972	0.965	0.976	0. 966
	Kappa	0.880	0.868	0.905	0.910	0.767	0.852	0.850	0.927	0.955	0.914	0.928	0.898	0. 903
FROM_GLC2015	P.I.	0.965	0.100	0.824	0.787	0.272	0.489	0.395	0.819	0.974	0.525	0.942	0.112	0. 715
	U.I.	0.910	0.900	1.000	0.967	0.925	1.000	1.000	0.939	0.873	0.991	0.980	1.000	0. 948
	P.N.	0.955	0.998	1.000	0.986	0.996	1.000	1.000	0.983	0.966	0.999	0.986	1.000	0. 988
	U.N.	0.983	0.862	0.972	0.900	0.869	0.907	0.914	0.944	0.994	0.891	0.959	0.869	0. 920
	O.A.	0.958	0.862	0.975	0.919	0.872	0.915	0.918	0.943	0.968	0.901	0.968	0.871	0. 925
	Kappa	0.906	0.155	0.889	0.810	0.372	0.615	0.531	0.838	0.900	0.634	0.933	0.178	0. 769
NUACI_2015	P.I.	0.309	0.211	0.141	0.680	0.162	0.579	0.247	0.626	0.758	0.539	0.843	0.042	0. 527
	U.I.	0.669	1.000	1.000	0.871	0.786	0.939	0.952	0.788	0.900	1.000	0.933	0.556	0. 873
	P.N.	0.928	1.000	1.000	0.948	0.991	0.993	0.998	0.946	0.980	1.000	0.957	0.994	0. 977
	U.N.	0.741	0.877	0.876	0.852	0.852	0.922	0.895	0.887	0.944	0.894	0.894	0.855	0. 874
	O.A.	0.730	0.881	0.879	0.857	0.849	0.924	0.897	0.868	0.937	0.905	0.909	0.852	0. 873
	Kappa	0.277	0.312	0.220	0.664	0.223	0.675	0.356	0.614	0.785	0.650	0.811	0.058	0. 585
GlobeLand30-2010	P.I.	0.876	0.244	0.671	0.500	0.346	0.263	0.593	0.774	0.895	0.642	0.785	0.310	0. 641
	U.I.	0.926	1.000	1.000	0.944	0.904	1.000	1.000	0.940	0.971	0.985	0.982	1.000	0. 960
	P.N.	0.967	1.000	1.000	0.985	0.992	1.000	1.000	0.984	0.994	0.997	0.990	1.000	0. 992
	U.N.	0.943	0.882	0.949	0.793	0.881	0.872	0.940	0.931	0.975	0.915	0.865	0.895	0. 902
	O.A.	0.938	0.886	0.953	0.820	0.882	0.878	0.945	0.933	0.975	0.925	0.904	0.900	0. 911
	Kappa	0.855	0.355	0.778	0.547	0.448	0.373	0.716	0.806	0.916	0.734	0.797	0.435	0. 717

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

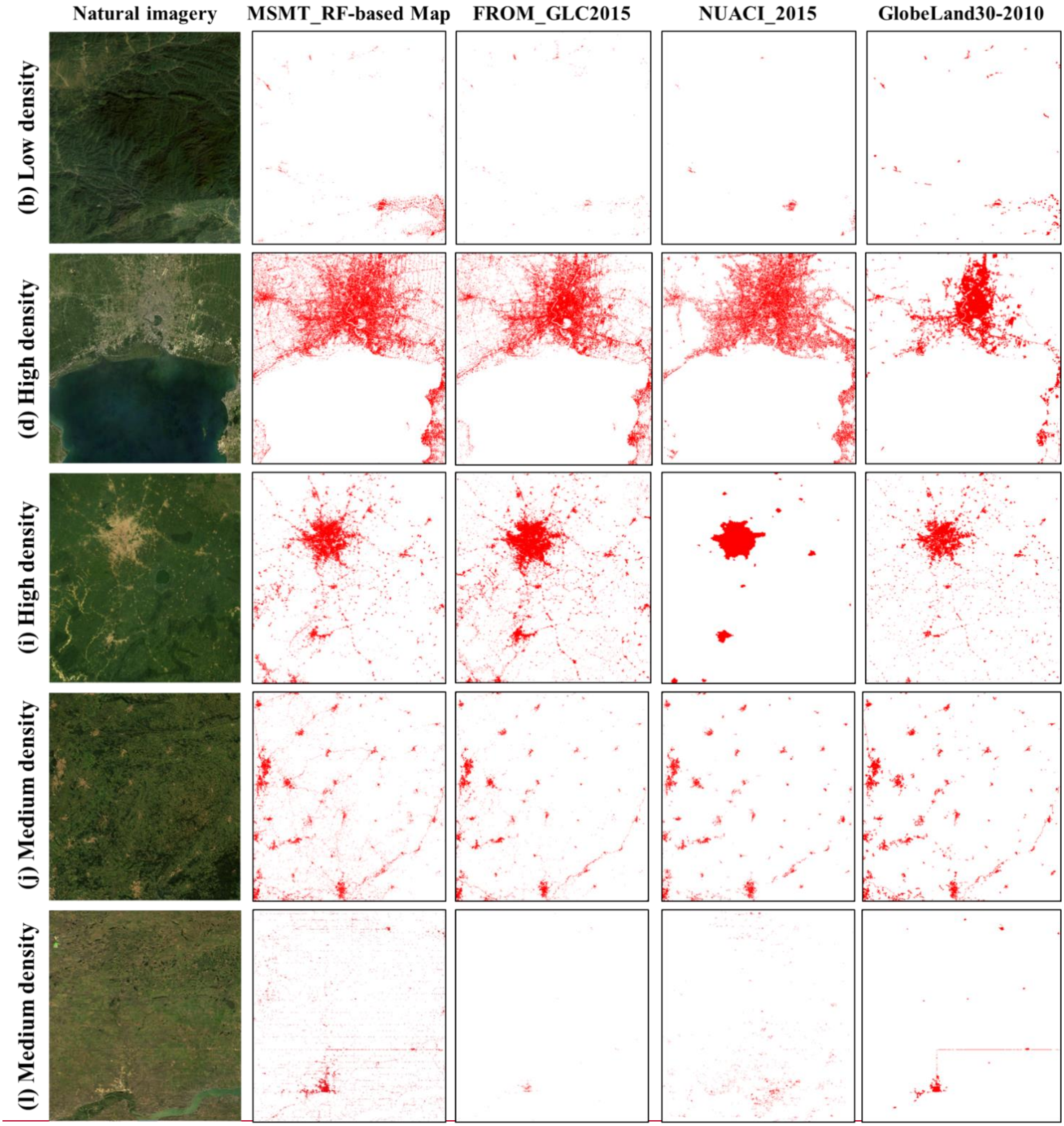
	U.N.	0.985	0.800	0.998	0.963	0.975	0.970	0.972	0.788	0.995	0.965	0.975	0.933	0.947	0.971	0.943	0.941
	P.N.	0.757	0.932	0.686	0.835	0.966	0.868	0.919	0.884	0.882	0.785	0.806	0.862	0.959	0.907	0.624	0.834
	O.A.	0.838	0.829	0.689	0.833	0.961	0.880	0.950	0.911	0.893	0.818	0.870	0.883	0.943	0.911	0.728	0.856
	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

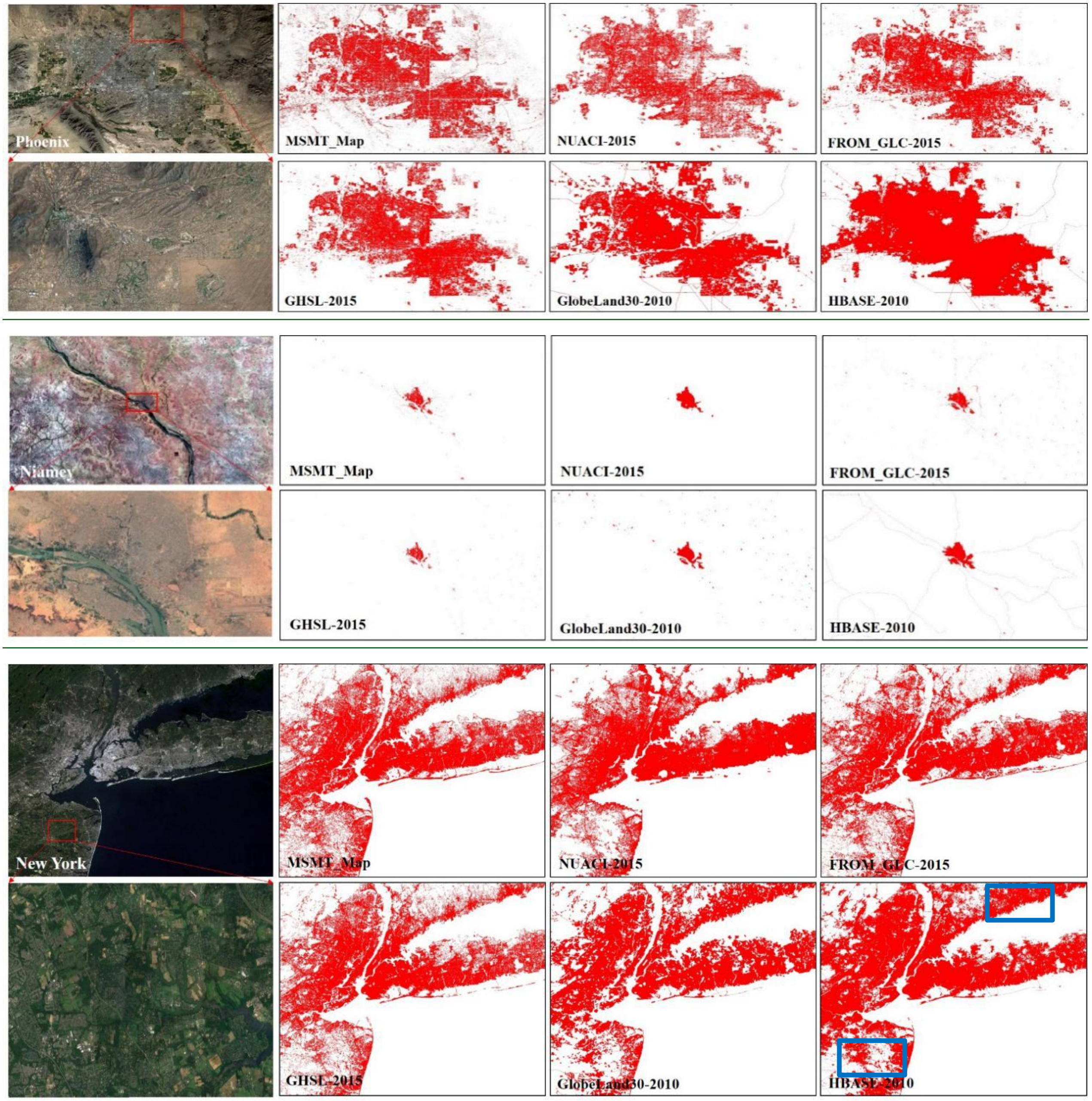
Note: I.~~D~~L., impervious ~~density~~landscape, CR, cropland-prevalent impervious landscape, BS, bare soil-prevalent impervious landscape, VG, vegetation-prevalent impervious landscape, P.I., producer's accuracy of impervious surfaces, U.I., user's accuracy of impervious surfaces, P.N., producer's 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 ~~four~~ six impervious products, five validation regions, including two ~~high-density~~bare soil-prevalent regions (Phoenix and Niamey), one ~~low-density~~vegetation-prevalent city (New York) and two ~~medium-density~~cropland-prevalent regions (Winnipeg and Bangkok), were selected in Fig~~gure~~. ~~6~~9. ~~First~~Specifically, in the first ~~low-density~~bare soil prevalent region of Phoenix ~~(Table 2b and Figure 6b)~~, 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 ~~had the most obvious underestimation problem, except for the urban centers, almost all impervious objects (peripheral urban and rural villages and roads) were missed.~~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. ~~Th~~This~~e~~ ~~series~~ 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). ~~The NUACI_2015, only identifying the urban areas and some obvious impervious objects, also missed fragmented and small rural villages. The causes of omission maybe came from the threshold method used by the NUACI_2015, specifically,~~ Liu et al. (2018) ~~developed a novel NUACI index to enhance the impervious surfaces and suppress 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 obvious impervious surfaces but bad performance for fragmented impervious surfaces.~~ As for the GlobeLand30-2010, ~~the omission errors were mainly~~there was little omission for the fragmented impervious objects over peripheral urban ~~due to~~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 ~~first high-density region (Table 2d)~~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. ~~GlobeLand30-2010 produced underestimates in many peripheral urban areas, which was mainly due to the temporal interval of 5 years. FROM_GLC2015 and NUACI_2015 still omitted some rural roads and villages (top-right and center-left of Figure 6d) compared with the MSMT_RF-based results. In the second high-density region (Table 2i), NUACI_2015 identified large and medium-sized cities at the cost of missing all of the fragmented villages. The MSMT_RF-based results and FROM_GLC2015 accurately delineated the spatial distributions of urban and rural roads and settlements. GlobeLand30-2010 still suffered from a few omission errors caused by the temporal difference between data sets and the limitation of minimum mapping unit.~~

Lastly, in the two cropland-prevalent cities of Bangkok and Winnipeg, the MSMT-2015 had greater advantages~~first medium-density region (Table 2j and Fig. 6j), the situation was similar to that for Figure 6d – the omission errors in FROM_GLC2015, NUACI_2015 and GlobeLand30-2010 were caused by the omission of small and fragmented villages and roads. As for the last validation region, the medium-density region (Table 2l and Figure 6l) consisted of typical rural areas of North America containing a large number of small and rural roads. As Figure 6l illustrates, almost all roads and villages were missed in FROM_GLC2015 and NUACI_2015 whereas GlobeLand30-2010 could identify the main roads while still missing the minor roads. In summary, as the MSMT_RF method could accurately and comprehensively identify fragmented and small villages and roads, it gave a higher producer's accuracy~~ 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. ~~.~~


Natural imagery
MSMT_RF-based Map
FROM_GLC2015
NUACI_2015
GlobeLand30-2010
(b) Low density
(d) High density
(i) High density
(j) Medium density
(l) Medium density


Phoenix
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010
Niamey
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010
New York
MSMT_Map
NUACI-2015
FROM_GLC-2015
GHSL-2015
GlobeLand30-2010
HBASE-2010

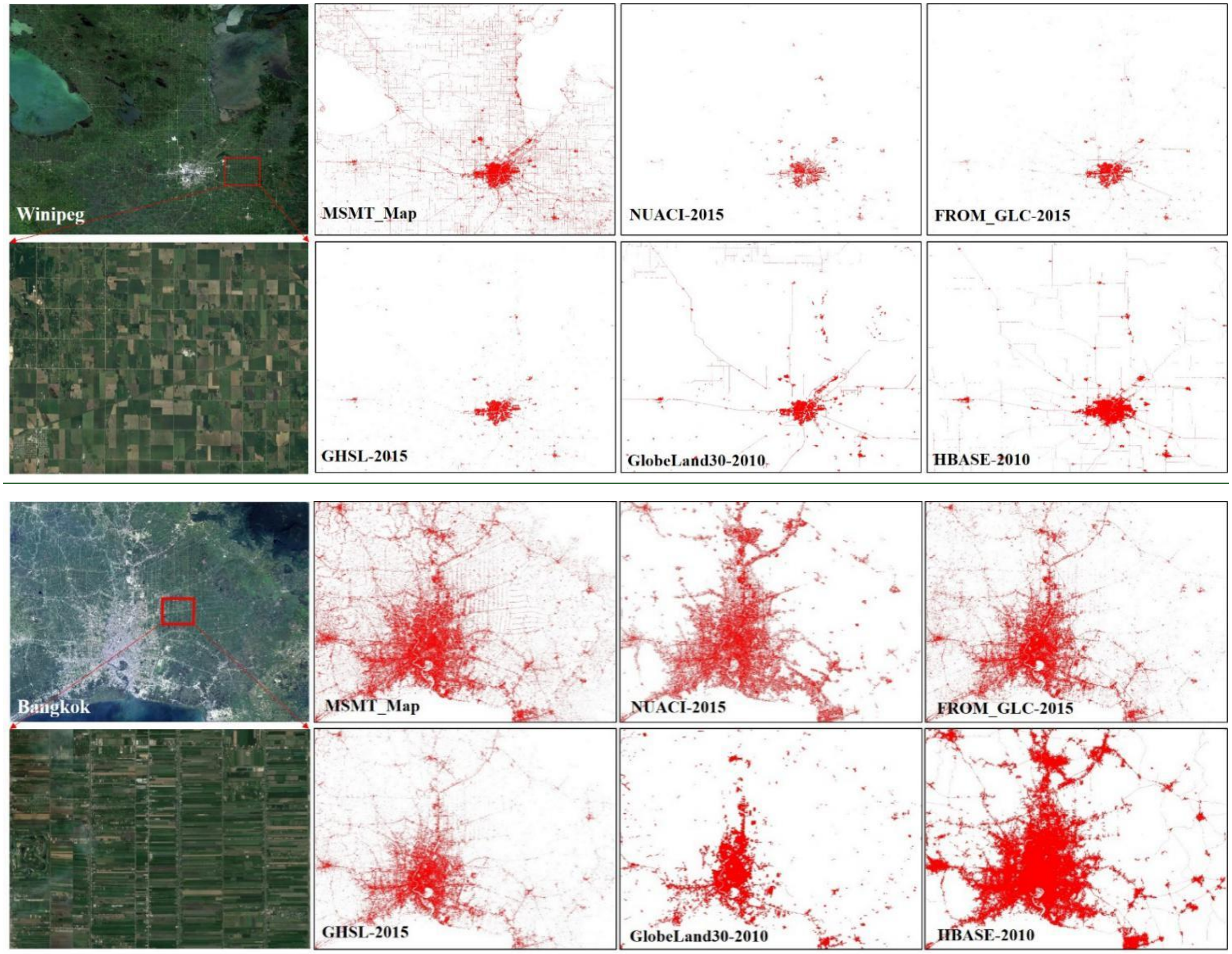


Figure ~~6~~9:~~.~~ Comparisons between the MSMT_RF-based impervious surface maps ~~(the second column)~~ and other ~~impervious surface~~ products (~~(the 3rd to 5th columns~~ corresponded to the NUACI products developed by Liu et al. (2018), the FROM_GLC products developed by Gong et al. (2013), the GHSL products developed by Florczyk et al. (2019),~~ the NUACI products developed by Liu et al. (2018) and~~ 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 ~~surface densities~~landscapes~~. The first column listed the corresponding natural-composited Landsat images in the growing season of 2015~~.

~~5~~ 6 Discussion

~~5.1 The advantages of multi-source and multi-temporal classification~~

~~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 some studies have demonstrated that the integration of multi-source and multi-temporal information can improve the mapping accuracy, these studies mainly focused on the regional scale and regions of high impervious surface density (Zhang et al., 2014b; Zhu et al., 2012). At present, global impervious~~

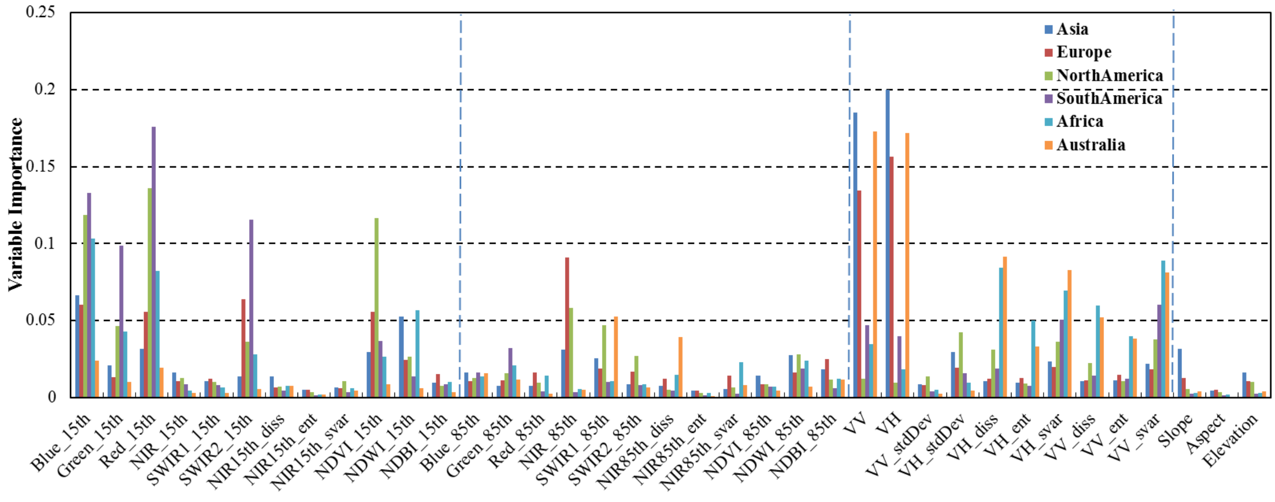
~~surface maps are still produced by optical imagery alone or by using a combination of optical and DMSP-OLS or VIIRS NTL imagery (Huang et al., 2016; Liu et al., 2018; Schneider et al., 2010). In this study, we first developed the global 30-m impervious surface map using multi-source and multi-temporal imagery. To quantitatively demonstrate the need for using multi-source, multi-temporal information and the results of using it, 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 at the Python environment. 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. 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. 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. Following these were the 85th percentile of Landsat SR in the NIR and SWIR1 bands as well as the SAR texture features, whose mean importance was approximately 3%. In summary, only the integration of multi-source training features could guarantee the classification accuracy across different impervious landscapes. Similarly, Zhang et al. (2014b) also concluded that the combination of optical and SAR imagery could significantly improve the land-cover classification and impervious surface area estimation.~~

~~Secondly, as the intra-annual variability could increase the separability of impervious and non-impervious surfaces (Zhang and Weng, 2016), the importance of multi-temporal optical features was also investigated. 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: for example, the importance of the 15th percentile in the NIR and SWIR1 bands was low while that of 85th percentile was high. Therefore, the total importance of the bi-seasonal spectral features exceeded 70% in some cases. 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.~~

~~Lastly, although the importance of Landsat texture features and the topographical variables was lower than 5% in these six regions, some scientists have demonstrated that these features contribute a lot to the improvement of impervious mapping accuracy. For example, Shaban and Dikshit (2001) found that the combination of texture and spectral features improved the classification accuracy by 9% to 17% compared with the use of pure spectral features. Zhu et al. (2012) also 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. Subsequently, the cumulative importance of topographical variables over the region in Asia exceeded 5%, the topographical variables were necessary for impervious surface mapping in mountain areas. 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.~~



~~Figure 7. The importance of the input features derived from the random forest model using the training samples in six continental regions.~~

~~From the perspective of the impervious mapping method, the comparison between our MSMT_RF product and NUACI_2015 also demonstrates that the classification-based method performed better than the spectral index-based method (Section 4.2). We concluded that this improvement was mainly due to the combination of the multi-source and multi-temporal information in the classification method.~~

~~5~~6.~~2~~ 1 Reliability and sensitivity of the global training samples

In contrast to other classification-related studies that require manual efforts to collect training samples (Gao et al., 2012; Im et al., 2012; Zhang et al., 2016), we overcame the expensive cost of collecting accurate and sufficient training samples at a global scale. To ensure the accuracy and reliability of the training samples, a combination of the GlobeLand30-2010 land-cover product, which had been validated to have a ~~user's~~ producer's accuracy (which measures the commission error) of 9~~7.1~~4.7% for impervious surfaces (see Section ~~4~~5.~~2~~4), and DMSP-OLS NTL imagery was adopted to guarantee the reliability of each sample. As it was difficult and challenging to evaluate the accuracy of all the training samples, we randomly selected 1% of the total training samples (in Section 3~~.1~~) including 34,990 non-impervious and 9,840 impervious points to measure the reliability of the global training samples. After careful checking, we found that these training samples achieved accuracies of 91.9% and 99.5% for impervious and non-impervious surfaces, respectively.

Meanwhile, even if the training samples still contained a small number of erroneous points, the random forest model has been demonstrated to be resistant to noise and presence of erroneous samples (Belgiu and Drăguţ, 2016). In this study, we randomly changed the category of a certain percentage of the 34,990 samples and used the ‘‘noisy” samples to train the random forest classifier. Fig. ~~8~~ 10 illustrates the overall accuracy and impervious producer’s accuracy decreased for the increased percentage of erroneous samples. It was found that the overall and impervious producer’s accuracy remained stable when the percentage of erroneous samples increased from 1% to 20% while it rapidly decreased when the percentage of erroneous samples was higher than 20%. Similarly, Gong et al. (2019) also found that the decrease in overall accuracy was less than 1% when the error in the training samples was less than 20%.

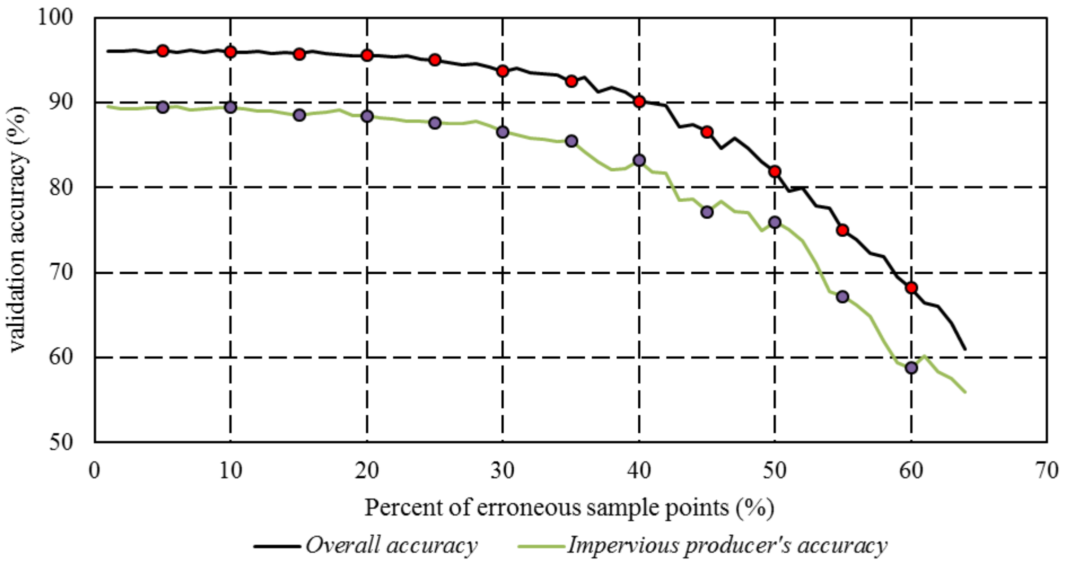


Figure 10~~8~~:~~.~~ Sensitivity analysis showing the relation between the classification accuracy and the percentage of erroneous samples points.

Therefore, the reliability and sensitivity analysis indicated that: (1) the random forest model is resistant to noisy training samples and performs well if the percentage of erroneous samples is lower than 20%; and (2) the training samples derived from the GlobeLand30 and DMSP-OLS NTL imagery were accurate enough for use in global impervious surface mapping.

~~5~~6.~~3~~ 2 Limitations of the proposed method

Although the proposed MSMT_RF method has been demonstrated to have the ability to produce the accurate impervious surface products, there are still some limitations to the method. First, as the training samples derived from the GlobeLand30-2010 are restricted to a 9×9-pixel local window and further refined by the integration of MODIS EVI and VIIRS NTL imagery, low-density impervious samples might be omitted and cause further omission of low-density impervious surfaces (rural villages, small roads and so on). Although, in this study, spatially adjacent training samples from the surrounding 3×3 areas were imported to reduce the omission of low-density samples, according to the accuracy assessment, higher omission errors were found in low and medium-density regions than in high-density regions. Therefore, our future work will pay more attention to the omission of low-density impervious surfaces.

Secondly, as Weng (2012) pointed out, mixed pixels are common in medium-resolution imagery due to the limitations of the spatial resolution and spectral heterogeneity of the landscape. The effectiveness of 'hard' classifiers is easily affected by these mixed pixels (low-density impervious pixels also constitute mixed pixels). Due to the proportion of impervious surfaces within a pixel, impervious surface areas are often overestimated in urban areas or underestimated in rural areas when using medium-resolution images (Lu and Weng, 2006). Therefore, our future work will focus on simultaneously producing the likelihood ('soft' probability) of each pixel being an impervious surface. At present, some scientists have produced continuous impervious fractions at a regional scale: for example, Okujeni et al. (2018) used the support vector regression method to estimate the fraction of impervious surfaces at the pixel scale.

~~6~~ 7 Data availability and user guidelines

The global impervious surface map data set generated in this paper are available on Zenodo: https://doi.org/10.5281/zenodo.3505079 (Zhang and Liu, 2019).

To facilitate the readers to reproduce this work, Table 3 gives the details of the datasource and platform information of the datasets and processes in this study. The input remote sensing datasets and products came from three parts including: GEE platform, free access websites and our group. Specifically, five kinds of basic datasets in section 2.1 were available at GEE platform. The ~~three~~ five ~~global~~ impervious surface products in section 2.2 were downloaded from the free access websites from National Geomatics Center of China, Tsinghua University, ~~and~~ Sun Yat-sen University, National Aeronautics and Space Administration (NASA), and Joint Research Centre (JRC). The validation samples were produced by our group using visual interpretation.

Further, the process of derivation of global training samples was implemented by using the multi-source datasets at localhost computation platform, and the random forest classification at each 5°×5° regional grid was developed by our group on the GEE platform using JavaScript language. The importance of multi-source and multi-temporal features and reliability and sensitivity of global training samples were analyzed at the localhost Python computation environment.

Table 3. The detailed information of the datasets and processes in this study

	Datasource and platform	Detailed datasets and processing steps
Datasets	Google Earth Engine platform	Landsat-8 optical, Sentinel-1 SAR, VIIRS NTL, MODIS EVI and STRM/ASTER DEM topographical imagery
	Free download websites	GlobeLand30-2010, FROM_GLC-2015, NUACI-2015, HBASE-2010 and GHSL-2015~~-~~ products

	Our Group	Validation samples
Processes	Google Earth Engine platform (JavaScript language)	The random forest classification at each 5 °×5 °regional grid
	Localhost platform (Python environment)	Derivation of global training samples The importance of multi-source and multi-temporal features reliability and sensitivity of global training samples

~~7~~ 8 Conclusions

Due to the spectral heterogeneity and complicated make-up of impervious surfaces, large-area impervious mapping is challenging and difficult. In this study, a global 30-m impervious surface map was developed by using multi-source, multi-temporal remote sensing data based on the Google Earth Engine platform. First, the global training samples were automatically derived from the GlobeLand30-2010 land-cover product together with VIIRS NTL and MODIS EVI imagery. Then, a local adaptive random forest model was trained using the training samples and multi-source and multi-temporal datasets for each 5 °×5 °geographical grid. Following that, the global impervious map produced by mosaicking a large number of 5 °×5 °regional impervious surface maps was validated by comparing it with ~~three~~ several existing products (GlobeLand30-2010, FROM_GLC-2015, ~~and~~ NUACI~~_~~-2015, HBASE-2010 and GHSL-2015) using approximately 1~~0,~~1,~~142~~ 942 interpretation samples. The results indicated that the MSMT_RF-based impervious surface map ~~had~~ achieved ~~the~~ the highest overall accuracy 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~~highest overall accuracy of 96.7% and a kappa coefficient of 0.888, followed by FROM_GLC2015 (92.5% and 0.769), GlobeLand30-2010 (91.1% and 0.717) and NUACI_2015 (87.43% and 0.585)~~. Therefore, it can be concluded that the global 30-m impervious surface map produced by the proposed MSMT_RF method is accurate and reliable for use in global impervious surface mapping.

Author contributions. Conceptualization, Liangyun Liu; Investigation, Xiao Zhang; Methodology, Liangyun Liu and Xiao Zhang; Software, Xiao Zhang and Xidong Chen; Validation, Xiao Zhang, Shuai Xie, Xidong Chen and Yuan Gao; Writing – original draft preparation, Xiao Zhang; writing—review and editing, Shuai Xie, Bing Zhang and Changshan Wu.

Competing interests. The authors declare that they have no conflict of interest.

Financial support. This research was funded by the Key Research Program of the Chinese Academy of Sciences (ZDRW-ZS-2019-~~1~~2), the National Natural Science Foundation of China (41825002), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19080304).

Acknowledgments. We gratefully acknowledge the free access of GlobeLand30 land-cover products provided by National Geomatics Center of China, the FROM_GLC land-cover products provided by Tsinghua University, ~~and~~ the NUACI impervious surface products provided by Professor Xiaoping Liu at Sun Yat-sen University, the GHSL impervious surface products produced by National Aeronautics and Space Administration, and HBASE produced by the Joint Research Centre.

References

Bai, Y., Feng, M., Jiang, H., Wang, J., and Liu, Y.: Validation of Land Cover Maps in China Using a Sampling-Based Labeling Approach, Remote Sensing, 7, 10589-10606, https://doi.org/10.3390/rs70810589, 2015.

Ban, Y., Jacob, A., and Gamba, P.: Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor, ISPRS Journal of Photogrammetry and Remote Sensing, 103, 28-37, https://doi.org/10.1016/j.isprsjprs.2014.08.004, 2015.

Belgiu, M. and Drăguţ, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31, https://doi.org/10.1016/j.isprsjprs.2016.01.011, 2016.

Bennett, M. M. and Smith, L. C.: Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics, Remote Sensing of Environment, 192, 176-197, https://doi.org/10.1016/j.rse.2017.01.005, 2017.

Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., and Hanssen, R. F.: ESA's sentinel missions in support of Earth system science, Remote Sensing of Environment, 120, 84-90, https://doi.org/10.1016/j.rse.2011.07.023, 2012.

Brown de Colstoun, E. C., Huang, C., Wang, P., Tilton, J. C., Tan, B., Phillips, J., Niemczura, S., Ling, P.-Y., and Wolfe, R. E.: Global Man-made Impervious Surface (GMIS) Dataset From Landsat. NASA Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, https://doi.org/10.7927/H4P55KKF, 2017.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J.: Global land cover mapping at 30m resolution: A POK-based operational approach, ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27, https://doi.org/10.1016/j.isprsjprs.2014.09.002, 2015.

Chen, X., Cao, X., Liao, A., Chen, L., Peng, S., Lu, M., Chen, J., Zhang, W., Zhang, H., and Han, G.: Global mapping of artificial surfaces at 30-m resolution, Science China Earth Sciences, 59, 2295-2306, https://doi.org/10.1007/s11430-016-5291-y, 2016.

Clarke, K. C., Hoppen, S., and Gaydos, L.: A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area, Environment and planning B: Planning and design, 24, 247-261, https://doi.org/10.1068/b240247, 1997.

Deng, C. and Wu, C.: BCI: A biophysical composition index for remote sensing of urban environments, Remote Sensing of Environment, 127, 247-259, https://doi.org/10.1016/j.rse.2012.09.009, 2012.

Didan, K., Munoz, A. B., Solano, R., and Huete, A.: MODIS vegetation index user's guide (MOD13 series), Vegetation Index and Phenology Lab, The University of Arizona. , 1-38, https://doi.org/10.5067/MODIS/MYD13Q1.006, 2015.

Du, P., Samat, A., Waske, B., Liu, S., and Li, Z.: Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features, Isprs Journal of Photogrammetry & Remote Sensing, 105, 38-53, https://doi.org/10.1016/j.isprsjprs.2015.03.002, 2015.

Elvidge, C. D., Baugh, K., Zhizhin, M., Feng, C. H., and Ghosh, T.: VIIRS night-time lights, International Journal of Remote Sensing, 1-20, https://doi.org/10.1080/01431161.2017.1342050, 2017.

Elvidge, C. D., Tuttle, B. T., Sutton, P. C., Baugh, K. E., Howard, A. T., Milesi, C., Bhaduri, B. L., and Nemani, R.: Global Distribution and Density of Constructed Impervious Surfaces, Sensors, 7, 1962-1979, https://doi.org/10.3390/s7091962, 2007.

ESA: Sentinel-1 SAR User Guide Introduction. Availabe online: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar (accessed on 26 December 2019), 2016.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Reviews of Geophysics, 45, https://doi.org/10.1029/2005rg000183, 2007.

Florczyk, A., Corbane, C., Ehrlich, D., Freire, S., Kemper, T., Maffenini, L., Melchiorri, M., Pesaresi, M., Politis, P., and Schiavina, M.: GHSL Data Package 2019, Luxembourg. EUR, 29788, https://doi.org/10.2760/290498, 2019.

Foody, G. M. and Mathur, A.: Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification, Remote Sensing of Environment, 93, 107-117, https://doi.org/10.1016/j.rse.2004.06.017, 2004.

Fu, P. and Weng, Q.: A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery, Remote Sensing of Environment, 175, 205-214, https://doi.org/10.1016/j.rse.2015.12.040, 2016.

Gao, F., Colstoun, E. B. d., Ma, R., Weng, Q., Masek, J. G., Chen, J., Pan, Y., and Song, C.: Mapping impervious surface expansion using medium-resolution satellite image time series: a case study in the Yangtze River Delta, China, International Journal of Remote Sensing, 33, 7609-7628, https://doi.org/10.1080/01431161.2012.700424, 2012.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R.: Random Forests for land cover classification, Pattern Recognition Letters, 27, 294-300, https://doi.org/10.1016/j.patrec.2005.08.011, 2006.

Goldblatt, R., Stuhlmacher, M. F., Tellman, B., Clinton, N., Hanson, G., Georgescu, M., Wang, C., Serrano-Candela, F., Khandelwal, A. K., Cheng, W.-H., and Balling, R. C.: Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover, Remote Sensing of Environment, 205, 253-275, https://doi.org/10.1016/j.rse.2017.11.026, 2018.

Gong, P., Liu, H., Zhang, M., Li, C., Wang, J., Huang, H., Clinton, N., Ji, L., Li, W., Bai, Y., Chen, B., Xu, B., Zhu, Z., Yuan, C., Ping Suen, H., Guo, J., Xu, N., Li, W., Zhao, Y., Yang, J., Yu, C., Wang, X., Fu, H., Yu, L., Dronova, I., Hui, F., Cheng, X., Shi, X., Xiao, F., Liu, Q., and Song, L.: Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017, Science Bulletin, https://doi.org/10.1016/j.scib.2019.03.002, 2019.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., and Chen, J.: Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data, International Journal of Remote Sensing, 34, 2607-2654, https://doi.org/10.1080/01431161.2012.748992, 2013.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202, 18-27, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Guide, P.: Landsat 8 surface reflectance code (LaSRC) product, Availabe online: https://landsat.usgs.gov/sites/default/files/documents/lasrc_product_guide. pdf (accessed on 26 December 2018), 2018. 2018.

Hansen, M. C., Egorov, A., Potapov, P. V., Stehman, S. V., Tyukavina, A., Turubanova, S. A., Roy, D. P., Goetz, S. J., Loveland, T. R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., and Bents, T.: Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD), Remote Sensing of Environment, 140, 466-484, https://doi.org/10.1016/j.rse.2013.08.014, 2014.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., and Megown, K.: Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information, Photogrammetric Engineering & Remote Sensing, 81, 345-354, https://doi.org/10.1016/S0099-1112(15)30100-2, 2015.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M.: Development of a 2001 national land-cover database for the United States, Photogrammetric Engineering & Remote Sensing, 70, 829-840, https://doi.org/10.14358/PERS.70.7.829, 2004.

Hu, Y., Liu, L., Liu, L., Peng, D., Jiao, Q., and Zhang, H.: A Landsat-5 atmospheric correction based on MODIS atmosphere products and 6S model, IEEE Journal of Selected Topics In Applied Earth Observations And Remote Sensing, 7, 1609-1615, https://doi.org/10.1109/JSTARS.2013.2290028, 2014.

Huang, X., Schneider, A., and Friedl, M. A.: Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights, Remote Sensing of Environment, 175, 92-108, https://doi.org/10.1016/j.rse.2015.12.042, 2016.

Im, J., Lu, Z., Rhee, J., and Quackenbush, L. J.: Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data, Remote Sensing of Environment, 117, 102-113, https://doi.org/10.1016/j.rse.2011.06.024, 2012.

Jokar Arsanjani, J., Tayyebi, A., and Vaz, E.: GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries, Habitat International, 55, 25-31, https://doi.org/10.1016/j.habitatint.2016.02.003, 2016.

Langanke, T.: Copernicus Land Monitoring Service–High Resolution Layer Imperviousness Product Specifications Document, Copernicus team at EEA, Availabe online: https://land.copernicus.eu/user-corner/technical-library/hrl-imperviousness-technical-document-prod-2015 (accessed on 2 May 2020), 2016.

Li, C., Peng, G., Wang, J., Zhu, Z., Biging, G. S., Yuan, C., Hu, T., Zhang, H., Wang, Q., and Li, X.: The first all-season sample set for mapping global land cover with Landsat-8 data, Science Bulletin, 62, 508-515, https://doi.org/10.1016/j.scib.2017.03.011, 2017.

Li, X., Gong, P., and Liang, L.: A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data, Remote Sensing of Environment, 166, 78-90, https://doi.org/10.1016/j.rse.2015.06.007, 2015.

Li, X. and Zhou, Y.: Urban mapping using DMSP/OLS stable night-time light: a review, International Journal of Remote Sensing, 38, 6030-6046, https://doi.org/10.1080/01431161.2016.1274451, 2017.

Li, X., Zhou, Y., Zhu, Z., Liang, L., Yu, B., and Cao, W.: Mapping annual urban dynamics (1985–2015) using time series of Landsat data, Remote Sensing of Environment, 216, 674-683, https://doi.org/10.1016/j.rse.2018.07.030, 2018.

Liu, X., Hu, G., Chen, Y., Li, X., Xu, X., Li, S., Pei, F., and Wang, S.: High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform, Remote Sensing of Environment, 209, 227-239, https://doi.org/10.1016/j.rse.2018.02.055, 2018.

Lu, D. and Weng, Q.: Use of impervious surface in urban land-use classification, Remote Sensing of Environment, 102, 146-160, https://doi.org/10.1016/j.rse.2006.02.010, 2006.

Massey, R., Sankey, T. T., Yadav, K., Congalton, R. G., and Tilton, J. C.: Integrating cloud-based workflows in continental-scale cropland extent classification, Remote Sensing of Environment, 219, 162-179, https://doi.org/10.1016/j.rse.2018.10.013, 2018.

Okujeni, A., Canters, F., Cooper, S. D., Degerickx, J., Heiden, U., Hostert, P., Priem, F., Roberts, D. A., Somers, B., and van der Linden, S.: Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-impervious-soil fractions across multiple cities, Remote Sensing of Environment, 216, 482-496, https://doi.org/10.1016/j.rse.2018.07.011, 2018.

Okujeni, A., van der Linden, S., Tits, L., Somers, B., and Hostert, P.: Support vector regression and synthetically mixed training data for quantifying urban land cover, Remote Sensing of Environment, 137, 184-197, https://doi.org/10.1016/j.rse.2013.06.007, 2013.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment, 148, 42-57, https://doi.org/10.1016/j.rse.2014.02.015, 2014.

Pekel, J. F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418-422, https://doi.org/10.1038/nature20584, 2016.

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., and Syrris, V.: Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014, Publications Office of the European Union, https://doi.org/10.2788/253582, 2016.

Pflugmacher, D., Cohen, W. B., Kennedy, R. E., and Yang, Z.: Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics, Remote Sensing of Environment, 151, 124-137, https://doi.org/10.1016/j.rse.2013.05.033, 2014.

Ridd, M. K.: Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities†, International Journal of Remote Sensing, 16, 2165-2185, https://doi.org/10.1080/01431169508954549, 1995.

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., and Jeganathan, C.: Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture, Remote Sensing of Environment, 121, 93-107, https://doi.org/10.1016/j.rse.2011.12.003, 2012.

Schneider, A., Friedl, M. A., and Potere, D.: Mapping global urban areas using MODIS 500-m data: New methods and datasets based on ‘urban ecoregions’, Remote Sensing of Environment, 114, 1733-1746, https://doi.org/10.1016/j.rse.2010.03.003, 2010.

Schneider, A., Friedl, M. A., and Potere, D.: A new map of global urban extent from MODIS satellite data, Environmental Research Letters, 4, 044003, https://doi.org/10.1088/1748-9326/4/4/044003, 2009.

Schug, F., Okujeni, A., Hauer, J., Hostert, P., Nielsen, J. Ø., and van der Linden, S.: Mapping patterns of urban development in Ouagadougou, Burkina Faso, using machine learning regression modeling with bi-seasonal Landsat time series, Remote Sensing of Environment, 210, 217-228, https://doi.org/10.1016/j.rse.2018.03.022, 2018.

Shaban, M. and Dikshit, O.: Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh, International Journal of Remote Sensing, 22, 565-593, https://doi.org/10.1080/01431160050505865, 2001.

Shao, Z., Fu, H., Fu, P., and Yin, L.: Mapping Urban Impervious Surface by Fusing Optical and SAR Data at the Decision Level, Remote Sensing, 8, 945, https://doi.org/10.3390/rs8110945, 2016.

Sun, G., Kong, Y., Jia, X., Zhang, A., Rong, J., and Ma, H.: Synergistic Use of Optical and Dual-Polarized SAR Data With Multiple Kernel Learning for Urban Impervious Surface Mapping, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12, 223-236, https://doi.org/10.1109/jstars.2018.2883654, 2019a.

Sun, Z., Wang, C., Guo, H., and Shang, R.: A Modified Normalized Difference Impervious Surface Index (MNDISI) for Automatic Urban Mapping from Landsat Imagery, Remote Sensing, 9, 942, https://doi.org/10.3390/rs9090942, 2017.

Sun, Z., Xu, R., Du, W., Wang, L., and Lu, D.: High-Resolution Urban Land Mapping in China from Sentinel 1A/2 Imagery Based on Google Earth Engine, Remote Sensing, 11, 752, https://doi.org/10.3390/rs11070752, 2019b.

Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D. B., Oimoen, M. J., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., and Haase, J.: ASTER Global Digital Elevation Model Version 2 – Summary of validation results, Kim Fakultas Sastra Dan Budaya, https://doi.org/10.1093/oxfordjournals.pubmed.a024792, 2011.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., and Rostan, F.: GMES Sentinel-1 mission, Remote Sensing of Environment, 120, 9-24, https://doi.org/10.1016/j.rse.2011.05.028, 2012.

USGS: Landsat surface reflectance data, Reston, VA, Report 2015-3034, 1 pp., 2015.

Vermote, E., Justice, C., Claverie, M., and Franch, B.: Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product, Remote Sensing of Environment, 185, 46-56, https://doi.org/10.1016/j.rse.2016.04.008, 2016.

Wang, P., Huang, C., Brown de Colstoun, E., Tilton, J., and Tan, B.: Global human built-up and settlement extent (HBASE) dataset from Landsat, NASA Socioeconomic Data and Applications Center (SEDAC): Palisades, NY, USA, https://doi.org/10.7927/H4DN434S, 2017a.

Wang, P., Huang, C., Tilton, J., Tan, B., and Brown de Colstoun, E.: HOTEX: An approach for global mapping of human built-up and settlement extent, 1562-1565, https://doi.org/10.1109/IGARSS.2017.8127268,2017b.

Wang, Y., Liu, L., Hu, Y., Li, D., and Li, Z.: Development and validation of the Landsat-8 surface reflectance products using a MODIS-based per-pixel atmospheric correction method, International Journal of Remote Sensing, 37, 1291-1314, https://doi.org/10.1080/01431161.2015.1104742, 2016.

Weng, Q.: A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China, International journal of remote sensing, 22, 1999-2014, https://doi.org/10.1080/713860788, 2001.

Weng, Q.: Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends, Remote Sensing of Environment, 117, 34-49, https://doi.org/10.1016/j.rse.2011.02.030, 2012.

Wetherley, E. B., Roberts, D. A., and McFadden, J. P.: Mapping spectrally similar urban materials at sub-pixel scales, Remote Sensing of Environment, 195, 170-183, https://doi.org/10.1016/j.rse.2017.04.013, 2017.

Wu, C.: Normalized spectral mixture analysis for monitoring urban composition using ETM+ imagery, Remote Sensing of Environment, 93, 480-492, https://doi.org/10.1016/j.rse.2004.08.003, 2004.

Wu, C. and Murray, A. T.: Estimating impervious surface distribution by spectral mixture analysis, Remote Sensing of Environment, 84, 493-505, https://doi.org/10.1016/s0034-4257(02)00136-0, 2003.

Xie, Y. and Weng, Q.: Spatiotemporally enhancing time-series DMSP/OLS nighttime light imagery for assessing large-scale urban dynamics, ISPRS Journal of Photogrammetry and Remote Sensing, 128, 1-15, https://doi.org/10.1016/j.isprsjprs.2017.03.003, 2017.

Xu, H.: Analysis of Impervious Surface and its Impact on Urban Heat Environment using the Normalized Difference Impervious Surface Index (NDISI), Photogrammetric Engineering & Remote Sensing, 76, 557-565, https://doi.org/10.14358/pers.76.5.557, 2010.

Yang, J. and He, Y.: Automated mapping of impervious surfaces in urban and suburban areas: Linear spectral unmixing of high spatial resolution imagery, International Journal of Applied Earth Observation and Geoinformation, 54, 53-64, https://doi.org/10.1016/j.jag.2016.09.006, 2017.

Yang, Y., Xiao, P., Feng, X., and Li, H.: Accuracy assessment of seven global land cover datasets over China, ISPRS Journal of Photogrammetry and Remote Sensing, 125, 156-173, https://doi.org/10.1016/j.isprsjprs.2017.01.016, 2017.

Zhang, H., Lin, H., Li, Y., Zhang, Y., and Fang, C.: Mapping urban impervious surface with dual-polarimetric SAR data: An improved method, Landscape and Urban Planning, 151, 55-63, https://doi.org/10.1016/j.landurbplan.2016.03.009, 2016.

Zhang, H., Lin, H., and Wang, Y.: A new scheme for urban impervious surface classification from SAR images, ISPRS Journal of Photogrammetry and Remote Sensing, 139, 103-118, https://doi.org/10.1016/j.isprsjprs.2018.03.007, 2018a.

Zhang, H., Zhang, Y., and Hui, L.: Seasonal effects of impervious surface estimation in subtropical monsoon regions, International Journal of Digital Earth, 7, 746-760, https://doi.org/10.1080/17538947.2013.781241, 2014a.

Zhang, H., Zhang, Y., and Lin, H.: A comparison study of impervious surfaces estimation using optical and SAR remote sensing images, International Journal of Applied Earth Observation and Geoinformation, 18, 148-156, https://doi.org/10.1016/j.jag.2011.12.015, 2012.

Zhang, H. K. and Roy, D. P.: Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification, Remote Sensing of Environment, 197, 15-34, https://doi.org/10.1016/j.rse.2017.05.024, 2017.

Zhang, L. and Weng, Q.: Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 113, 86-96, https://doi.org/10.1016/j.isprsjprs.2016.01.003, 2016.

Zhang, L., Zhang, M., and Yao, Y.: Mapping seasonal impervious surface dynamics in Wuhan urban agglomeration, China from 2000 to 2016, International Journal of Applied Earth Observation and Geoinformation, 70, 51-61, https://doi.org/10.1016/j.jag.2018.04.005, 2018b.

Zhang, X. and Liu, L.: Development of a global 30-m impervious surface map using multi-source and multi-temporal remote sensing datasets with the Google Earth Engine platform, https://doi.org/10.5281/zenodo.3505079, 2019.

Zhang, X., Liu, L., Chen, X., Xie, S., and Gao, Y.: Fine Land-Cover Mapping in China Using Landsat Datacube and an Operational SPECLib-Based Approach, Remote Sensing, 11, 1056, https://doi.org/10.3390/rs11091056, 2019.

Zhang, X., Liu, L., Wang, Y., Hu, Y., and Zhang, B.: A SPECLib-based operational classification approach: A preliminary test on China land cover mapping at 30 m, International Journal of Applied Earth Observation and Geoinformation, 71, 83-94, https://doi.org/10.1016/j.jag.2018.05.006, 2018c.

Zhang, Y., Zhang, H., and Lin, H.: Improving the impervious surface estimation with combined use of optical and SAR remote sensing images, Remote Sensing of Environment, 141, 155-167, https://doi.org/10.1016/j.rse.2013.10.028, 2014b.

Zhou, T., Zhao, M., Sun, C., and Pan, J.: Exploring the Impact of Seasonality on Urban Land-Cover Mapping Using Multi-Season Sentinel-1A and GF-1 WFV Images in a Subtropical Monsoon-Climate Region, ISPRS International Journal of Geo-Information, 7, 3, https://doi.org/10.3390/ijgi7010003, 2017.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., Jin, S., Dahal, D., Yang, L., and Auch, R. F.: Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative, ISPRS Journal of Photogrammetry and Remote Sensing, 122, 206-221, https://doi.org/10.1016/j.isprsjprs.2016.11.004, 2016.

Zhu, Z., Woodcock, C. E., Rogan, J., and Kellndorfer, J.: Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data, Remote Sensing of Environment, 117, 72-82, https://doi.org/10.1016/j.rse.2011.07.020, 2012.

Zhuo, L., Shi, Q., Tao, H., Zheng, J., and Li, Q.: An improved temporal mixture analysis unmixing method for estimating impervious surface area based on MODIS and DMSP-OLS data, ISPRS Journal of Photogrammetry and Remote Sensing, 142, 64-77, https://doi.org/10.1016/j.isprsjprs.2018.05.016, 2018.