Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-65-AC2, 2019




# *Interactive comment on* "A 30-meter resolution national urban land-cover dataset of China, 2000–2015" *by* Wenhui Kuang et al.

**Wenhui Kuang et al.**

kuangwh@igsnrr.ac.cn

Received and published: 8 November 2019

**General Comment:**
Kuang et al. refined an existing land use and land cover data set (China's Land Use/cover Dataset) specifically for generating fractions of impervious surface area (ISA) and vegetation (within cities) at national level. While I find the manuscript and the dataset of general interest, I still have some concerns that I think the authors need further consideration. I will comment on this manuscript following the guidelines from the publisher's website.
**Response:**
Thank you very much for your constructive comments. We revised the manuscript according to your comments and suggestions. Detailed response and changes in



Printer-friendly version

Discussion paper

manuscript were listed in the PDF file in the supplement.

**Section 1**

**Section 1, Paragraph 1**

**Section 1, Paragraph 1, Point 1:** The data set related to ISA fraction and urban vegetation fraction at national level presented in the manuscript is new but not the method used to estimate them. The use of NDVI and other auxiliary data including reflectance to estimate ISA fraction has been done previously (e.g., Sexton et al., 2013, Remote Sensing of Environment). Obviously, these types of citations are neglected in the manuscript.

**Response:**

Thank you for your comments. We found that previous studies mainly focused on the analysis of urban land covers at individual city scale. For example, Sexton et al. used a single regression model to retrieve ISA in Washington, D.C.-Baltimore MD (Sexton et al., 2013, RSE). Our case focused on mapping of intra-urban land-cover at a national extent with the support of GEE, which is much complex. In this revised manuscript, we added more text to describe the methods and added more key references.

**Changes in manuscript:**

We revised the related sentences in the introduction part and cited more references, including Sexton et al. (2013) in page 3, lines 1-10.

**Section 1, Paragraph 1, Point 2:** Another factor that may lead to the judgement of the data set presented in this manuscript not as useful as it claimed by the authors is the mapping interval (5-year). A quick search of the current literatures would tell you that the scientific community now advocates for urban land cover datasets at a higher temporal frequency (e.g., annual mapping), particularly for urban environmental and climate studies. However, the authors did not even identify/mention possible use of their datasets of a five-year interval. For example, how does your dataset contribute to "world urban database" that may eventually help studies in urban climate using earth system models (e.g., weather research and forecasting, WRF)? I am not advocating for a case study or specifically linking your dataset to "world urban database", but a potential linkage between this presented dataset and environmental studies/applications would help us evaluate the contribution of your dataset to the scientific community or beyond. Based on the current literatures in mapping urban land use and land cover change at annual interval, I think the presented dataset may be of limited use to characterize duration, change magnitude, and timing of urbanization.

**Response:**

Thanks for your comments. Yes, we agree entirely with you that annual datasets have higher values than five or ten-year datasets. Currently ESA- and MODIS-based annual land cover products and Landsat-based urban datasets were generated. However, the ESA- and MODIS-based datasets cannot effectively capture urban spatial patterns due to coarse spatial resolution, while Landsat-based urban datasets have relatively low accuracy (for example, producer's accuracy and user's accuracy are 0.50–0.60 and 0.49–0.61, Liu et al. 2018, RSE) that cannot meet requirements of real applications. In order to produce high-spatial resolution and high accurate urban datasets, we integrated different data sources and approaches to produce China's urban datasets that can meet real applications. Because of time-consuming and intensive labor, it is challenging to generate annual datasets. We think five-year urban dataset is suitable considering the following reasons: (1) urban expansion often occurred dispersedly and in relatively small patch sizes in a year, thus Landsat images with 30 m spatial resolution cannot effectively capture this kind of changes in annual time interval; (2) urban datasets are often related to socioeconomic data, which they are often surveyed at five or ten-year interval; (3) most of national land cover products such as US National Land Cover Database (NLCD) released at five-year or ten-year interval (1992, 2001, 2006, 2011, 2016) (Yang et al., 2018, ISPRS Journal of Photogrammetry and Remote Sensing) and China's Land Use/Cover Dataset (CLUD) at five-year interval (Liu et al.,

2005, RSE; Zhang et al., 2016, RSE), considering time and labor.
In the original manuscript, we did not discuss the use of this product. Thanks for your suggestion, we added texts to indicate the potential uses of this product in different fields such as environments, urban climate, human settlements management, socioeconomic analysis, and future planning of national urban development.

**Changes in manuscript:**

We revised the abstract of the manuscript. We also added texts in the introduction section to indicate the potential use of this product in different fields.

**Section 1, Paragraph 2**

**Section 1, Paragraph 2, Point 1:** Additionally, the methods presented in this manuscript are not in "best practices". For example, the reference impervious surface fractions used to build regression models in this study are extracted from spectral mixture analysis (obviously extracted manually). This seems to be against what the authors claimed in the Introduction section that manual extraction of endmembers may lead to biased estimations of ISA and vegetation fractions (it should have biased estimations). At least, I think the authors should provide an assessment of the reference ISA fractions (similar to what you did for final datasets) used to build the model at each city and how uncertainties/errors from this subjective reference dataset can eventually propagate to the final ISA and vegetation fraction dataset. Anyway, I think the authors should provide estimates of errors and uncertainties associated with this dataset (which is related to data quality in question 2).

**Response:**

Thank you for your comments and suggestions. The ISA dataset was generated using the same approach that was detailed in our previous publication (Kuang et al., 2014, Landscape and Urban Planning). The results were validated using reference data and an overall accuracy of 91.1% was obtained. More texts were added to indicate the accuracy issue.

**Changes in manuscript:**
We added the accuracy of ISA classification in the revised manuscript, see page 6, line 5.

**Section 1, Paragraph 2, Point 2:** It is worth noting that spectral mixture analysis is recently standardized at global scale and can be used to estimate ISA and vegetation fractions at an annual interval (e.g., Small 2013 in Remote Sensing of Environment).
**Response:**
Yes, spectral mixture analysis is a powerful tool for decomposing multispectral imagery into different fractional images. As Small indicated that globally standardized spectral mixture analysis can effectively extract substrate, dark and vegetation. However, ISA cannot be accurately and directly extracted from multispectral image using spectral mixture analysis considering the wide spectral variation of ISA, that is, similar spectral signatures between ISA and other non-vegetation types, such as bare soils and water. Also the meaning of substrate and dark used in Small (2013) is different with ISA.

**Section 1, Paragraph 3**
**Section 1, Paragraph 3, Point 1:** I do not quite agree with the authors that the dataset provides metrics for urban structure. This is confusing since urban structure may more refer to its landscape patterns, where shopping malls are located and where residential areas are located. The dataset only refers to the landscape composition in urban areas.
**Response:**
There is a little confusion about the urban structure in this manuscript. In the revised version, we replaced “urban structure” with “intra-urban land-cover”. Thanks for your comments.
**Changes in manuscript:**
We revised the manuscript and replaced “urban structure” with “intra-urban land-cover”,

which can be found in page 1 line 11, page 2 line 25, page 2 line 30, page 3 line 2, page 3 line 27, page 3 line 30, page 6 line 1, page 9 line 13, page 10 line 27, page 11 line 2.

**Section 2**

**Section 2, Paragraph 1**

**Section 2, Paragraph 1, Point 1:** The dataset is accessible and complete as described in the manuscript. As the authors refined the existing dataset for generating ISA and vegetation fractions, the accuracy of the presented dataset should be also dependent on the accuracy of the previous dataset. Thus, the final reported accuracy should be the product of the accuracy of the previous dataset and the newly generated dataset. Further accuracy assessment of this dataset should be reported.

**Response:**

Yes, we agree entirely with you that the accuracy in the previous dataset will affect the accuracy of the final results. Although the accuracy of previous dataset is high enough in the view of pixel level, the 30 m spatial resolution of pixel-level ISA data still contains a mixture of ISA and greenness (or even bare soil, water) because of the complex urban landscape. Therefore, we used the logistic regression approach to modify the pixel-level ISA data, then to produce fractional ISA dataset in order to improve the area statistics. We added the accuracy assessment results in the revised manuscript.

**Changes in manuscript:**

We added the accuracy assessment results of this dataset, see page 6, line 5.

**Section 2, Paragraph 2**

**Section 2, Paragraph 2, Point 1:** In comparison with other global urban datasets as shown in Fig. 8, I think the dataset from this manuscript is not as accurate as ESA land cover dataset. It seems that the new dataset sets a hard boundary for urban areas and discard neighboring regions beyond the boundary. This dataset is then may be of further limited use for studies in climate modeling (e.g., in WRF) that requires continuous land cover datasets in both spatial and temporal domains.

**Response:**

Thank you for your comments. In our research, we focused on urban area and excluded the area without a sufficient population size. Therefore, we have the clear boundary of urban extent. For other global ISA datasets such as the ESA land cover dataset, they are valuable for global environmental studies, but these datasets have some shortcomings such as coarse spatial resolution resulting in poor spatial patterns of urban land covers (ISA, greenness, water) and relatively low accuracy in the urban landscape. Our objective is to provide accurate urban ISA and greenness datasets with much higher spatial resolution (30 m in our study). In order to compare different datasets, we summarize the current urban land products to delineate different among them (Tabel 1, shown in Fig. 1 below) and provides an example figure to show the area statistics based on Beijing city. Because other products can't effectively distinguish urban and rural lands, their urban areas were considerably overestimated (Figure 1, shown in Fig. 2 below). Based on accuracy assessment of our results, we obtained accuracy range between 92.0% and 98.9%, much higher accuracies than other existing products.
In addition, our dataset is developed from CLUD. The rural area is presented in CLUD (Figure 2, shown in Fig. 3 below). They can be integrated into our results if needed.

**Section 3**

**Section 3, Paragraph 1**

**Section 3, Paragraph 1, Point 1:** The spatial resolution is not consistent. The manuscript claimed it at 30 m, but what I see from the dataset is 250 m.

**Response:**

Yes, we developed the 30 m spatial resolution products, but the uploaded dataset was resampled to 250 m spatial resolution, considering the data size. The

30 m resolution dataset will be available by contacting the corresponding author (kuangwh@igsnrr.ac.cn)

**Section 3, Paragraph 2**

**Section 3, Paragraph 2, Point 1:** The only comparison I can think of, which the author can do, is a comparison between your dataset and other existing global dataset in terms of changes in urban areas over time (rather than just simple visual comparisons of maps). Specific numbers from each dataset for selected cities can help us further evaluate the performance of the method and the dataset. But this is a minor comment.

**Response:**

Good suggestion, thanks. We conducted a comparison of different products based on Beijing city, as replied in Section 2, Paragraph 2, Point 1.

**Changes in manuscript:**

We added the figure in section 4.2 of the manuscript. The manuscript was revised to provide detailed explanation.

**Section 4**

**Section 4, Paragraph 1**

**Section 4, Paragraph 1, Point 1:** I would suggest the authors add more metadata to describe the dataset in the downloaded documents: spatial resolution, extent, cities included, accuracy for each city, legend. The dataset I downloaded from the website does not include that information although brief information is available on the website.

**Response:**

Thanks for your suggestion. We revised the metadata of the website so that readers can obtain detailed information about this product.

**Changes in manuscript:**

We resubmitted the metadata file on the website.

**Section 4, Paragraph 2**
**Section 4, Paragraph 2, Point 1:** Figure 4 can be improved. I do not really understand what Fig. 4 tells us: is the logistic regression is the right method to use? Maybe random forest regression is better?
**Response:**
Figure 4 showed the logistic regression model of impervious surface estimation based on four cities – Dalian, Jinan, Wuhan, and Xi'an. The left (orange) and middle (green) histograms showed the frequency distribution of NDVImax value for ISA and non-ISA sample points, respectively. The right figure (blue curve) showed the logistic regression model fitted with the sample points in the left and medium figures. For different regions or cities, the regression models vary. We built different models in China for ISA estimation. For example, the regression curve of Xi'an showed a steep slope and Dalian a relatively smooth slope.
Random Forest (RF) is a commonly used machine learning method for urban land classification and ISA estimation when multiple variables were used. However, when only one variable was used, RF does not have the advantage over other methods. In particular, when training samples are only located some specific regions, RF-based model cannot be effectively transferred to other regions without training samples. Considering that our study is to establish a model based on one variable and this model will be used to estimate ISA at national scale, we selected the logistic regression approach to estimate ISA in order to effectively use this model to estimate different cities. Based on our exploration in limited number of cities, the logistic regression model provided accurate estimation with RMSE of 0.1.

**Section 4, Paragraph 3**
**Section 4, Paragraph 3, Point 1:** I am not clear of what criteria you used to apply your built models to other cities. Based on locations? How practical for this method to be applied at broad scale or national scale?

**Response:**
Good question. In our original version, we did not clearly describe this issue. So in the revised one, we added more texts to explain this issue. In China, population density and economic condition have wide variation, resulting in considerably different ISA distribution across the country. In order to solve this problem, we developed different models according to specific economic regions. For example, we choose the Chinese economic geographic zones and assumed a consistent logistic regression relationship within each partition. A certain number of cities were selected and the logistic regression parameters of each city were calculated. The average value of the parameters in each economic and geographic zone is obtained as a regression parameter for all cities in the same zone. Based on this method, we calculated the preliminary ISA value for cities in each zone.
**Changes in manuscript:**
More texts were added to provide the explanation, see page 6, lines 7-18.

**Section 4, Paragraph 4, Point 2:** This approach can be easily improved with more automatic methods for example using globally standardized spectral mixture analysis (Small et al. 2013 in Remote Sensing of Environment. Thus, the method you used does not fit in the “uniqueness” point as identified on the publisher’s website, see the reviewer guidelines).
**Response:**
As replied in Section 1, Paragraph 2, Point 2, the globally standardized spectral mixture analysis is a valuable tool to provide a standard method for land use classification, but it cannot effectively and directly extract ISA datasets without intensive post-processing. Therefore, we proposed the integrated approach to provide accurate ISA and greenness datasets, although this approach takes much time and labor to produce the product.

Please also note the supplement to this comment:
https://www.earth-syst-sci-data-discuss.net/essd-2019-65/essd-2019-65-AC2-supplement.pdf

---

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-65, 2019.

| Name | Spatial resolution | Abbreviation | Method | Reference |
|---|---|---|---|---|
| China's Urban Land use/cover Dataset | 30 m | CLUD-Urban | Visual interpretation | - |
| Land Cover from Moderate-resolution Imaging Spectroradiometer | 500 m | MODIS LC | Decision tree classification | (Friedl et al., 2010) |
| European Space Agency global land-cover data | 300 m | ESA LC | Unsupervised classification and change detection | (Bontemps et al., 2011) |
| Global Land Cover at 30 m resolution | 30 m | GlobaLand 30 | Pixel-Object Knowledge (POK)-based classification | (Chen et al., 2015) |
| Built-up grid of the Global Human Settlement Layer | 30 m | GHS Built | Symbolic machine learning | (Pesaresi et al., 2013, 2016) |
| Multi-temporal Global Impervious Surface | 30 m | MGIS | Normalized urban areas composite index | (Liu et al., 2018) |

**Fig. 1.** Table 1: List of urban land products for comparison.



Printer-friendly version

Discussion paper

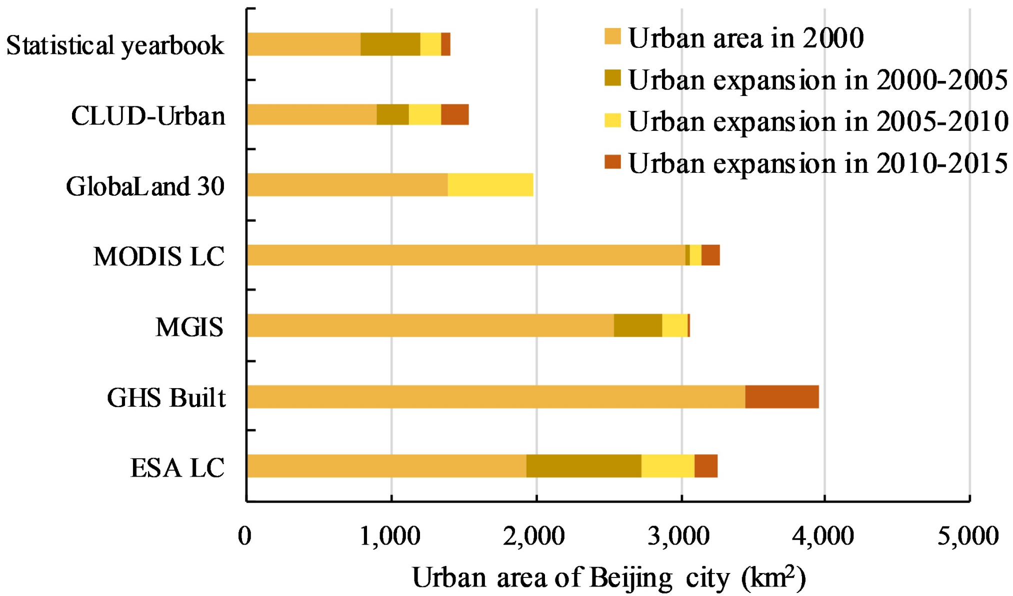


**Fig. 2.** Figure 1: Comparison of urban land area and change in Beijing city based on different urban land-use products.



Printer-friendly version

Discussion paper

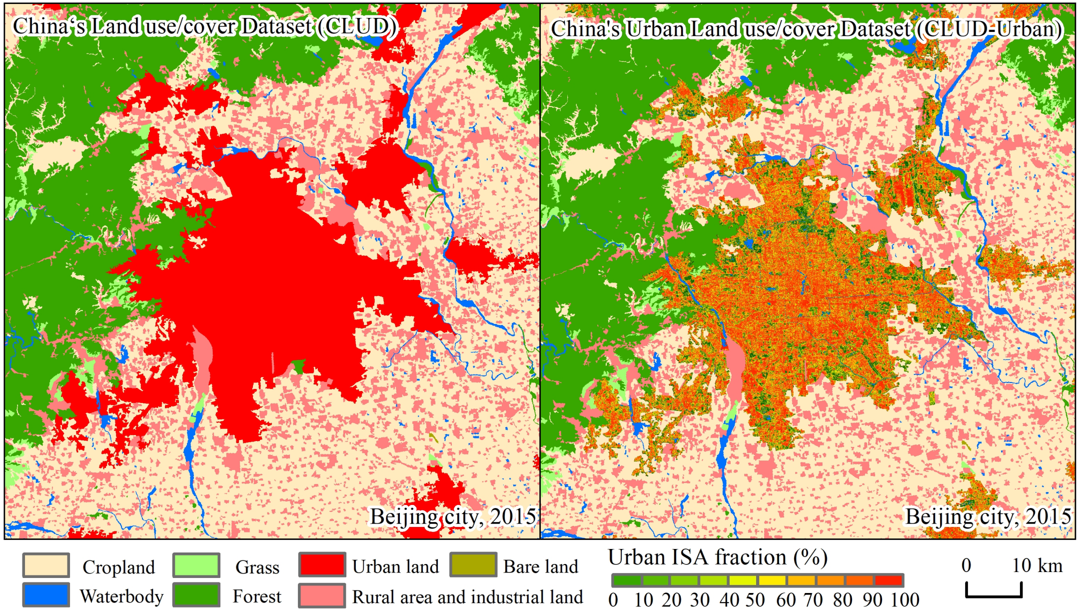


**Fig. 3.** Figure 2: Comparison of former CLUD (left figure) and the newly developed CLUD-Urban (right figure, delineating detailed intra-urban land-cover).



Printer-friendly version

Discussion paper