

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

Anonymous Referee #1

Received and published: 13 September 2019

Summary and general comments

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.

1. Are the data and methods presented new? Is there any potential of the data being useful in the future? Are methods and materials described in sufficient detail? Are any references/citations to other data sets or articles missing or inappropriate?

C1

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. 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 dataset 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.

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

C2

uncertainties associated with this dataset (which is related to data quality in question 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).

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.

2. Is the data set accessible via the given identifier? Is the data set complete? Are error estimates and sources of errors given (and discussed in the article)? Are the accuracy, calibration, processing, etc. state of the art? Are common standards used for comparison?

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.

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.

3. Are there any inconsistencies within these, implausible assertions or data, or noticeable problems which would suggest the data are erroneous (or worse). If possible, apply tests (e.g. statistics). Unusual formats or other circumstances which impede

C3

such tests in your discipline may raise suspicion.

The spatial resolution is not consistent. The manuscript claimed it at 30 m, but what I see from the dataset is 250 m.

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.

4. Is the data set usable in its current format and size? Are the formal metadata appropriate? Check the publication: Is the length of the article appropriate? Is the overall structure of the article well structured and clear? Is the language consistent and precise? Are mathematical formulae, symbols, abbreviations, and units correctly defined and used? Are figures and tables correct and of high quality?

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.

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?

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? 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).

2019.

C5