

Comment on **essd-2021-62**

Anonymous Referee #1

Referee comment on "The dataset of walled cities and urban extent in late imperial China in the 15th–19th centuries" by Qiaofeng Xue et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-62-RC1>, 2021

The Authors present in the data descriptor, the dataset on the urban extent in late imperial China. Although creating such datasets was definitely a time-consuming process for such a long period and large spatial extent, I have some comments, doubts, and suggestions that could help in future usage of the database, by other scholars. Please find them below.

- a) The Authors write that 'most cities in late imperial China built city walls, and these walls were usually built around the urban built-up area (Yannis et al., 2017).' The paper cited here, by Yannis et al., 2017, goes much further starting with (maybe too strong – I think you can clarify that, as the specialists) quote of Osvald Sirén (1924), that 'There is no real city in Northern China without a surrounding wall, a condition which, indeed, is expressed by the fact that the Chinese use the same word ch'eng for a city and a city-wall: for there is no such thing as a city without a wall. It is just as inconceivable as a house without a roof.'

I think better clarification and support of the thesis on city walls' importance in China is needed in your paper. You state that city walls in China are the only reliable data on the urbanization processes (e.g. due to lack of other data, as stated in the paper). But it needs to be clarified because the role of city walls in China differed in the past when compared to the role of the city walls in other parts of the world at the time. Otherwise, one can assume that the dataset presented here has limited value for the scientific community.

- b) The Authors cite the works by Skinner, but do not refer broadly to the fact that Skinner produced his own database of more than 150 attributes for all cities incl. information on the city walls. Although in Chinese scientific literature there has been a discussion with Skinner approach (Cao 2001), I think it would be good to discuss it

briefly here and refer to the main differences between the datasets, so that it is clear for the readers what is new and why it was changed, regarding previous works.

- c) Please clarify if the city walls extent was understood as administrative boundaries or where there some exceptions?
- d) In the title you have the 15th-19th centuries, although many cities from the 14th century can be also found in the database. Could you clarify that?

Comments regarding datasets:

- a) Paragraph 6 on the data accuracy is very general and mostly based on the 'accuracy ranking', but in my opinion, it is not fully clear how did you rank the accuracy in detail? Much more information (incl. examples) is needed here since this is now the only uncertainty attribute in your dataset. Please explain the ranking rules, so that the readers have no doubts about how it was exactly done.
- b) Also, there is not much on the geometric accuracy of your data (I assume this is because it was done on recent detailed remote sensing data), but there might happened that some cities from the database were completely destroyed and are not detectable on the current remote sensing data – how was the extent of the walls assessed then? Are those cities present in the database or not?
- c) The datasets' attributes are explained in the Data Records files, but there are discrepancies between the attributes presented there and in the shapefiles. E.g. in CEUD dataset, we can find the attribute 'TYPE' which is not explained in the Data Records file. Similarly, in CCWAD attributes we can find 'area', but Data Records do not clarify the unit it is presented in. Attribute 'References' should be fully explained – e.g. in the form of the full list of options which can be found there. Preferably the list of attributes should be explained both in the Data Descriptor (e.g. in the Data availability section or in the form of a respective table) and in the Data repository.
- d) It is not clear why in some cases there are differences among the databases and cities covered. For instance, in the CCWAD there is a city of Yijinai which is attributed by the end year 1372, while in CUED it is covered only for the representative year 1400, but as Weiyuan. None of the later periods cover it. Is it because it changed the name, the extent, and did not survive till the next, 1537 (or any later) representative years?
- e) The above-mentioned example shows an important issue of the database – in the Data Records CCWAD file you write: 'Due to the ancient Chinese cities often have several names at the same time and they are always change, here we provide the most common name for them.' -it is not clear what 'most common' mean. I would recommend reconsidering that issue – would it be possible to either present all the names as the attributes (I suppose you have it for representative years) or use e.g. clear rule like, the last name, the name which was used for the longest period. You can also explain better 'most common', as now it is not fully clear.
- f) The Authors write that: 'In addition, we also need some remote sensing images for auxiliary work. The 1970s China remote sensing image from the U.S. Geological Survey (USGS) website (<https://earthexplorer.usgs.gov/>) was the most important'. It is not explained in the text or in Figure 3b (or l. 200), which sensors exactly were used here? We assume it was not Landsat, so please state clearly which images were used and add respective details. This is important because actually, remote sensing data are the main source of the extent of the walls in your work.
- g) Is there a possibility to verify somehow, e.g. for selected cities, if your extent is assessed correctly, e.g. by comparing the area to the reliable statistical records from the respective period?

Other comments:

- a) 40-42 – how this work can help in Chinese sustainable urbanisation currently? It is somehow unclear (similarly the reference to sustainable development goals in the Abstract).
- b) 125-126 – the same title appears twice – was it a different publication or edition? Please clarify
- c) There are some minor English spelling and grammar issues (e.g. line 100 – ‘lat’ instead of ‘late’) and repetitions (line 110-114 – word ‘region’ is repeated 5 times).
- d) Style should also be corrected – e.g. lines 240-241, l. 251 – ‘It should to analysis the time series...’, l. 339-340
- e) The Authors use the names of the regions as Region I, II, etc – did you consider replacing it with proper geonames?
- f) In the Conclusion the Authors explain a very interesting relation between city walls and war – could you elaborate it? It might be very important in terms of your dataset usage.
- g) In point 3 (l. 379) of the Conclusions the Authors mention the cities without the walls – could you explain, how common the phenomena was? When and why it happened?