

Interactive comment on “Fine-grained, spatio-temporal datasets measuring 200 years of land development in the United States” by Johannes H. Uhl et al.

Jonathan Holt (Referee)

jonathan.holt@duke.edu

Received and published: 19 October 2020

In the manuscript, “Fine-grained, spatio-temporal datasets measuring 200 years of land development in the United States”, the authors seek to construct spatially-explicit settlement data for the United States extending back to the early nineteenth century. This is an important contribution to the study of settlement and development in the United States. Insights from these data will help researchers study historical trends as well as predict future changes to the settled landscape.

The essence of this manuscript is the introduction of two improvements to an existing dataset, the HISDAC-US: 1) individually owned buildings or units and 2) unique built-

C1

up property locations. The authors make these improvements by ingesting data from Zillow’s ZTRAX and translating it into an accessible format (geoTIFF files).

In Section 2 the authors describe and showcase each of the three main data products (BUPR, BUPL, BUA). The figures should be applauded for their clarity and detail.

In Sections 3-4, the authors discuss data processing and validation. The extensive three-component validation is impressive and suggests reasonable accuracy in most urban areas during most years. The positional accuracy of mobile home parks and pseudo-locations is troubling, but as the authors note this is a small fraction of the data. Figure 9e and 9f are particularly helpful for visualizing the likely sources of false positive and false negative signals. Presumably, users concerned with location accuracy will chose to aggregate the data to a lower spatial resolution; furthermore, users may wish to use data from the year 1900 onwards to ensure sufficient accuracy for their analyses.

Overall, this data description is very well presented. The procedure is clear, logical, and methodologically sound. The authors discuss sources of error and take care to quantify different types of error across time and space. The authors provide supplementary datasets that allow users to further quantify error for their own analyses.

I believe this manuscript is worthy of publication. However, I have some minor concerns and/or suggestions:

- 1) While it is nice to look at BUA in figure 3, It isn’t immediately clear to me that BUA is worthy of its own layer (i.e., are BUPR and BUPL sufficient?). If BUA is simply $BUPR > 0$, then any practitioner could generate BUA with the click of a mouse or with one line of code.
- 2) Is it the case that all publicly owned buildings are omitted from ZTRAX (line 230)? If so, please provide some statistics to help the reader understand the magnitude of this omission. For example, what percentage of buildings in a major US city like Boston are publicly owned?
- 3) My biggest concern is the issue of removed buildings (line 95). For example, imagine that the “contemporary” BUPL detects 10 buildings within a single grid cell. In one scenario, there could be 10 buildings in the

C2

grid cell. In another scenario, there could be one building standing, while the other nine have been removed. Worse yet, there could be zero buildings because all 10 have been removed. The issue of removed buildings perhaps warrants more attention.

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-217>, 2020.