Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-82-RC2, 2019


Open Access Earth System Science Data Discussions



# *Interactive comment on* "Allocating people to pixels: A review of large-scale gridded population data products and their fitness for use" *by* Stefan Leyk et al.

**Stephen Matthews (Referee)**

sxm27@psu.edu

Received and published: 8 July 2019

Thank you for the opportunity to read and comment on the manuscript "Allocating People to Pixels: A Review of large-scale Gridded Population Data Products and their Fitness for Use."

First and foremost this is a valuable resource/summary and is most welcome. I haven't used many of the products in my own research but in my instructional role I have referred to these products and encouraged students to explore them. This comparative summary provides a useful overview to the data products, a set of spatial-temporal and modeling issues, and introduces guidelines to help determine fitness of use. That is, I



Printer-friendly version

Discussion paper

can envision promoting this “review” in instruction and reference to colleagues. I liked the paper.

My comments are general and are provided to help potentially improve the experience for the reader. They are not listed in any priority order but just things that occurred to me as I read the material.

1: Why use the word “pixels” in the title? . . . and not “grids”? While the journal readership will be familiar with both terms (and I know there are publications on people and pixels) but all of the products highlighted are gridded population data sets. Further, the abstract does not contain the word pixel(s) and the word is rarely used in the manuscript (just in a few subheadings). Even in subheadings I would prefer to see ‘grids.’ Then why heading #4 “people in places”?

2: In section #3 the reader is introduced to the POPGRID website at www.popgrid.org. This is great. Many of the tables in the manuscript are from this website but curiously they are either selected extracts or restructured. One would have thought that consistency between the two would be more useful and that the paper should follow the format of the tables that are online. This is especially so as in Table 1 where there is no obvious structure to the listing of data resources but the same table online organizes these same data resources/products by whether the data are “unmodeled,” “lightly modeled,” or “highly modeled.” I would suggest that the latter helps the reader and this is especially so when it comes to later sections of the paper that introduce the various methods for population redistribution (and also section #5.2)

3: I would much prefer that Table #1 comes before the start of section 3.1. It has been introduced but we wait for until after the short description of all component data resources before we encounter it.

3b: Table #1 should be titled “Detailed characteristics and availability . . . “ as it covers “availability” as well.

3c: some columns in Table #1 are closely related to “fitness for use” expanded on in section #6 and alludes to at other points in the paper, so perhaps highlight them as such in the table and/or more explicitly refer back to them in the text (perhaps both when first introduced and in sections #5 and #6). This would help tie the two parts of the paper together (i.e., the “review” and the “fitness of use” sections).

4: I agree that HYDE is a very interesting data set and perhaps of import to the readers of the journal but these data are unique historically and also are available at a fairly crude level compared to all other gridded data products. That is, is this product sufficiently different to include as say supplemental material rather than list with the others?

5: I wondered if the ancillary data section (#4.1) and Figure 1 should come after section 4.2. Perhaps a relatively minor issue.

5b: Figure #1 has a lot of information that might be better explained. Ditto Figure #2.

6: I know the readership of the journal is likely to be very sophisticated in this area but I still think that a glossary of key terms is necessary and should be presented early in the paper. Maybe I am thinking about my role as an instructor but I suspect in assigning this reading I would have to prepare the novice/intermediate user to several key terms.

7: I very much liked Figures 3 and 4 and their brief description.

8: I wondered if lines 11-15 could be bolded/italicized or perhaps just a separate paragraph to give emphasis to the interrelatedness of the determinants of fitness for use.

9: Fitness of use #2 focuses on urban population analysis. This seemed like an opportunity to link to Table 4 at www.popgrid.org/compare-data on “Global and Continental Urban Extent / Settlement Layers: Summary Characteristics.”

10: Not sure there is, but if there was, an expansion of the theme “How have these data sets been used previously” would be useful. Some good examples, just wondering about classic papers or studies that could be used as exemplars.

11: I recognize that there are pieces of the next comment scattered in the paper but for the user I wondered if more could be made of time-constant geographies and issues related to the embeddedness of units of analysis and scale. Some readers will be interested in trajectories of place as well as multilevel modeling (perhaps) but the number of temporal data collection points and the embeddedness of levels is not always made explicit. I agree that this can be complex but alluding to these issues in the context of research questions can help in the fitness of use.

12: While not the main purpose I wasn't sure about all the referencing to social media. I wonder if the trap of thinking about the future has compelled the authors to discuss this but it all reads fairly superficially. At least to me. I do think an expanded section on data challenges (e.g., non-representative samples) and fast changing data environments (social network data, real-time data) are worth discussing but if so, then in more depth.

13: Minor – gridded data sets are not particularly recent and they are part of the history and formation of remote sensing, raster GIS, and map algebra tools/perspectives in the spatial sciences (in disciplines spanning the environmental, geographic, and social sciences). An early citation to time-constant geographies and socioeconomic applications in GIS that focused on grid-based data products include David Martin (1991, 1995), now at Southampton.

---

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-82, 2019.



Printer-friendly version

Discussion paper