Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-82-RC1, 2019

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.

Tracy Kugler (Referee)

takugler@umn.edu

Received and published: 5 July 2019

This article presents a valuable summary and comparison of the major gridded population data products currently available. While the datasets discussed are generally well-established and have each been individually described in previous publications, having a summary comparing the key properties of these datasets in one place will be very useful for researchers trying to determine which product is best suited for their particular application. Furthermore, the concept of "fitness for use" is a helpful approach in this context. Each of the datasets discussed incorporates different types of input information and applies different methodology, which results in the final data products

having noticeably different properties. In the absence of definitive means to validate and assess the accuracy of gridded population data products (as discussed in the article), researchers should thoughtfully consider how the properties of a given dataset will affect the results of their application. The guidelines presented in the final section of this article provide a helpful framework for thinking through such considerations.

In general, the article is well-written, and the concepts are presented clearly. I have just one suggestion about potential additional content. I have some suggestions about rearranging the structure of the article to clarify the flow of ideas. I also have some more specific requests for clarification, suggestions regarding the figures, and one technical issue.

__Content suggestion__ Could the authors comment on how total population grids may be combined with data (e.g. published census tables) containing a wider range of characteristics?

__Structural Suggestions__

* Rearrangement of paper sections I suggest moving the Review of current data products (sec. 3) to follow the discussion of methods. Having the methods descriptions first would give readers better context to understand the methods used in each data product. Additionally, within the Key methods and ancillary data section (sec. 4), I suggest swapping the order of the Ancillary data (4.1) and Methods for population redistribution (4.2) subsections. Discussing different types of ancillary data makes more sense once the user knows what the ancillary data are used for. Conversely, the specifics of ancillary data are not necessary for understanding the methods. With this arrangement, the discussion starting at line 5 on p. 15 could be set off as its own section (headed something like, "Multiple answers and uncertainty") and placed after the discussion of ancillary data. As a central concern of the article, this discussion seems to deserve its own subsection. Following these suggestions would result in the following structure for the paper: * Introduction (unaffected) * Background and historical development (unaf-

fected) * Key methods and ancillary data (moved and rearranged) * Methods for population redistribution * Ancillary data * Multiple answers and uncertainty * Current data products (moved) * A fitness for use perspective (unaffected) * Concluding remarks and future work (unaffected) This structure puts the discussion of Methods for population distribution immediately following the Background and historical development section, which seems like a natural progression.

* Review of Current data products Within this section, the description of each product should be as parallel as possible. e.g., resolution, coordinate system, and type of population mapped in first sentence. Then narrative overview of input data and methods. Concluding with availability in the last sentence. Currently, the descriptions of Esri WPE, WorldPop, HYDE, and Demobase stray from this structure. For national and regional/continental products, coverage (countries/continent/region included) should be stated early in the description (first or second sentence). If possible, it would also be helpful to have coverage information for these products in table 1.

__Specific Clarification Questions__

* p. 9, line 16: What is the basis for HYDE prior to 1950? * p. 14, lines 18-23: It is not clear whether or how the statistical modeling approaches that are the subject of this paragraph differ from statistical approaches to dasymetric allocation mentioned on line 14. Please clarify. * p. 14, line 24 - p. 15, line 4: Again, it is not clear to me how so-called hybrid approaches fundamentally differ from non-hybrid weighted dasymetric methods. This brief paragraph doesn't really help clear up my confusion. Is the difference primarily in the techniques used to generate the weights? If so, what exactly is it that differentiates a hybrid approach from a straight weighted dasymetric approach? * p. 19, lines 30-31: It is somewhat confusing to draw a connection between de-jure and night time and de-facto and daytime populations. In the context of census data, de-jure and de-facto typically refer to longer timescales than day vs. night. e.g., If a person typically lives at a given address but is away on vacation on the day of the census, they could be counted at that location in a de-jure census, but at the location of

their vacation in a de-facto census. If a person commutes work (their daytime location) on the day of the census, though, they would still be counted at their residence, even in a de-facto census. * p. 20, lines 16-19: Maybe include a note of caution that the statistically estimated error relates specifically to the estimated relationship between population and ancillary data, not directly to the final population estimate (which also may incorporate uncertainty due to uncertainty in the input population and ancillary data themselves). * p. 21, line 18: Other questions are phrased in relation to the analysis under consideration. To be consistent, question 2 could be phrased as something like, "Does the analysis focus on urban populations?"

__Suggestions on Figures__

* Figure 1: As presented, this figure is somewhat difficult to interpret. * The specific "Good pixel proportion" of each Landsat scene seems like more detail than is necessary for this figure. The cost of introducing the rather confusing vertical axis in the top part of the figure to convey this detail doesn't seem worthwhile. * It seems odd to tie OSM to a specific point in time, since it is continuously evolving. * If a main intention of the figure is to compare the time points of ancillary data to census data, the census markers should be drawn out to the edge of the figure, rather than buried in the middle.

* Figure 2 * Label the 4 quadrants of the figure as a,b,c,d, and refer to specific quadrant(s) in the text to clarify the link between the text description and the illustration of a particular method. (Specifically, are the three methods mentioned on lines 13 & 14 of p. 14 – binary, 'intelligent', and statistical – meant to relate to the three variations of dasymetric illustrated in the figure?) * Indicate within the figure that the 6 gray cells represent built-up area or a similar category as shown in ancillary data

* Figure 3: A more intuitive visual arrangement of this figure would be to have panel (a) (input population data) on its own row at the top, followed by panels (b-e) (ancillary data) on two rows, followed by panel (f) (output population grid) on its own row at the bottom.

__Technical issue__ In the PDF download of the manuscript, clicking on the link in the caption for Table 1 attempts to go to: https://www.popgrid.org/popgrid_files/popgrid-5, which doesn't exist. (The '5' is from the line number on the page.) The URL as written in the manuscript does work. It might be best just to link to the HTML version at https://www.popgrid.org/compare-data, which is consistent with the link on p. 12 (line 26)

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