Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-189-RC2, 2019


Open Access Earth System Science Data Discussions



# *Interactive comment on* "Exposure data for global physical risk assessment" *by* Samuel Eberenz et al.

**Anonymous Referee #2**

Received and published: 26 November 2019

The authors present an open-source method that can be used to downscale low-resolution economic predictors to high-resolution gridded data by using nightlight intensity and gridded population data. The method and required data are described and a validation of the methodology is conducted for 14 selected countries. A global high-resolution dataset for 227 countries is created using this method and openly available for download. The documentation of the method in the open-source archive CLIMADA and the dataset are state-of-the-art and easily assessable to users. The presentation of the method and the dataset within the present manuscript needs major improvements. In general, the method and the dataset are described incompletely, the validation exercise and the subsequent consequences appear ad-hoc and unmotivated. In particular, the manuscript lacks a clear and precise writing style in various locations that make it



Printer-friendly version

Discussion paper

difficult for the reader to follow. Important information is missing, appears in different locations or is poorly referenced. This is a data description paper so all the relevant information concerning the data (including the input data) should be assembled here. Besides the specific locations noted below I ask the authors to critically revise the full manuscript to improve readability and understanding. I had to (re-)read many parts of the manuscript several times to finally get the full picture.

Major points:

1. The manuscript is about a global exposure dataset (for asset and/or GDP exposure?) for 227 countries. However, most of the manuscript deals with validation of the 14 test countries and some metropolitan areas. I expect the authors to include a description of the full dataset in section 3. This should include a clear statement of the countries and time periods included, missing countries or regions with low coverage in the available dataset and maybe even a worldmap figure. The reader should not download the huge dataset or consult Worldbank data in order to obtain this information himself.

2. The name of the method 'LitPop' and the function 'LitPop' (sometimes in italic) are used as synonyms. This is VERY confusing for the reader. To avoid confusion I would strongly encourage the authors to use LitnPopm (with m and n in the exponent) every time you talk about the function, even in the case when the exponent is one you should write Lit1Pop1 (with ones in the exponent).

3. Although I am not an expert on nightlight data I have the impression that there are some subtleties involved the user should know about. I quick google search tells me that usually an exponent >1 for nightlight data is used when deriving economic proxies to partially deal with the saturation issue. (This is somehow also apparent from your results in Figure 3). What about latitude-dependence of light intensity and the influence on your global dataset? I think the discussion in section 2 on input data needs to be advanced so the reader really gets to know the dataset and its subtleties.

4. In line 144 the authors state: 'While the absolute value of LitPop in itself does not bear any interpretable meaning, its relative value in comparison to the national or subnational sum determines how much of a macroeconomic indicator each pixel receives'. Saying that the authors do exactly the opposite in their validation exercise: they use aggregated absolute values from their method to compare it to observed quantities. Even more, this point is very difficult to extract from the manuscript. Only after jumping back and forth in the manuscript I understood that they actually calibrate their method with national GDP data and then compare subnational estimates. This needs to be stated much clearer.

5. The functionality used in Eq. 1 seems rather ad-hoc and only motivated by a study used in China. Have the authors used different approaches, different functional dependencies? What were their findings? Why is the exponential scaling beneficial? Mathematically, only the relative weighting between Lit and Pop is changed by the two exponents. Therefore, the approach could be simplified by using only one exponent that reflects the relative difference between both contributions. Have the authors looked into this direction?

6. The authors say they use two skill scores in line 178. Later they widen their analysis to three skill scores (e.g. Fig 3), which they interchangeably call methods as well. The authors should adjust their manuscript accordingly and stick to one naming convention.

7. The relevance of skill score 'beta' remains obscure (line 185). First, it is fully unclear about what slope the authors are talking. Second, the concept of linear regression in this context is fully unclear. Third, skill score 'beta' basically contains the same information as 'rho' (eq. 4), it is just a different scaling with respect to the standard deviations. I therefore do not understand why beta is needed in the first place and would ask the authors to remove one of them (beta or rho) as they are based on the same information.

8. The range of exponents m and n explored in the validation seems random and bares

any motivation. What is the motivation for the range of exponents explored? I strongly encourage the authors to motivate their validation and to conduct a more stringent validation accordingly.

9. The validation section (3.2) needs to clarified and amended. On first read (see my point raised before) I had the understanding that the analysis around Fig. 3 is based on 14 data points only (E.g. the caption of Fig. 3 points at this as well). Only later I understood that the authors use many data points (14 x subnational regions). The number of data points is never mentioned, however. How is the interquartile range defined? Please be much more precise and proactive.

10. Based on the redundancy of either beta or rho (stated above) and your diverging findings for different skill scores within your validation, I find the final decision to use the downscaling with m=n=1 (line 240) very ill-founded. At this stage I would expect a more thorough and stringent assessment of the different exponents and functionalities (see comment above). Otherwise, the full validation exercise seems redundant.

11. Line 187 (and others in the following): the notion of economically strong (or large) and weak regions is not very well defined. The reader can sort of understand what the authors hint at but it remains very unclear. How do they distinguish strong from weak regions? What is the precise criterion? Does this hold nationally or internationally?

12. The sentence 'There is probably a lot of housing and infrastructure in suburban México that is used by a population that works in the city and thus contributes to the GRP of Mexico City' (strange comparison of stocks and flows) and the following discussion is very difficult to digest for the non-expert reader. I find this discussion very relevant and think it should be extended here or at some other point in the manuscript as it directly links to many relevant issues: a) What does nightlight intensity actually capture? Assets or GDP? b) What is the highest downscaling resolution one should aim at when population is most likely a better proxy of the location of assets but nightlight also captures economic activity (e.g. driving cars)? Also in the light of above

sentence when GDP and assets seem to be separated by municipal boundaries. c) how can the interaction of both data sources most efficiently be combined? How does the present methodology add to this discussion? How can the different exponents be interpreted in this respect?

13. The discussion in line 284-292 is very vague as it is very hard to judge for the reader when to apply the authors' recommendation: high-resolution vs. coarsely resolved?, use a higher exponent of nightlights instead... instead to what? Why use exponent n=3 when this was never a potentially recommended value in the validation before? The discussion on auxiliary data should be placed somewhere else.

14. It is very unfortunate that the validation was (or could) only be conducted for 14 countries and no low-income country. The subsequent application of this method to all countries globally has to be treated with caution. In the present manuscript I am missing a detailed discussion of the reliability of the dataset for specific regions and/or income groups and a discussion of potential workarounds. What is the result of the authors' validation in terms of income groups? Is there any information (e.g. trends with income) that could be valuable for low income countries not treated here? What about very small countries, islands, etc? How could other data sources (e.g. household survey data from the Worldbank) be used to improve the data? What has been conducted with this respect in the literature so far (c.f. following paper and the references cited there: Gunasekera, R., et al. (2015). "Developing an adaptive global exposure model to support the generation of country disaster risk profiles." Earth-Science Reviews 150: 594-608.)?

15. The concept of intermediate downscaling appears in line 257 very ad-hoc and is used thereafter without further explanation.

16. LitPop as a top-down approach is first introduced in line 302. It would make much more sense to make this statement much earlier otherwise one should avoid this notion in general.

17. The term ‘exposure’ is used differently throughout the manuscript. It seems that the authors use it for ‘asset exposure’ but this is not fully clear. Exposure is very general and could be understood as population or GDP exposure as well. Therefore, I encourage the authors to be more precise and use the expression ‘asset exposure’ every time they mean it.

18. All abbreviations (e.g. GDP, GRP), all variables, and all subscripts (e.g. pix) need to be explained at first use, even if the authors think that they are self-explanatory. Thereafter another redefinition should be avoided and the authors should stick to their abbreviations.

19. Figure 4: The usage of Mexico (country) and México (region) is very confusing for the reader. Clearly state this difference and maybe use ‘México region’ to underline the difference.

Minor points:

20. The discussion in line 31 should include another freely available gridded GDP dataset: Kummu, M., et al. (2018). "Gridded global datasets for Gross Domestic Product and Human Development Index over 1990-2015." Sci Data 5: 180004.

21. The reference Murakami et al is outdated. Please update to: Murakami, D. and Y. Yamagata (2019). "Estimation of Gridded Population and GDP Scenarios with Spatially Explicit Statistical Downscaling." Sustainability 11(7).

22. Line 34: The statement on high-resolution GDP data availability for academic purposes only is not true. Upon checking the reference I found that the data is freely available. The corresponding reference should be included in the manuscript: Geiger, Tobias; Daisuke, Murakami; Frieler, Katja; Yamagata, Yoshiki (2017): Spatially-explicit Gross Cell Product (GCP) time series: past observations (1850-2000) harmonized with future projections according to the Shared Socioeconomic Pathways (2010-2100). GFZ Data Services. http://doi.org/10.5880/pik.2017.007

23. Line 55 (and others): the reference to Zhao et al. cannot be found in the list of references.

24. Line 177: What does the exponent '5' stand for in nGRP_i? Looks like a footnote which I am unable to locate. Same issue in line 189 and 228.

25. Line 182: Seems like the separated equation for rho got lost and appears inline now. The enumeration eq. 4 is also missing.

26. Figure 2: I do not understand what do you mean by log-normal colorbar? I would appreciate the colorbar to have a label. What kind of USD do you use here? PPP-adjusted, current or real? This applies similarly for Fig A1.

27. Line 219: replace top -> bottom

28. Line 326-328: The information on RMSF is repeating what the authors mentioned earlier around line 190.

29. Line 240: remove 'an'

30. Line 243: A reference to the data in the appendix would be very helpful here as the reader is unable to extract the information for Mexico from section 3.2.

31. Line 264: the reference for Pittore et al cannot be found in the list of references.

32. Line 334: replace get > become

33. Caption figure A1: replace 'the Mexico and USA' > 'Mexico and the USA'

---

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-189, 2019.



Printer-friendly version

Discussion paper