Manuscript essd-2022-92, First Revision
Submitted to: Earth System Science Data

A global map of Local Climate Zones to support earth system modelling and urban scale environmental science

Demuzere, M., Kittner, J., Martilli, A., Mills, G., Moede, C.,
Stewart, I. D., van Vliet, J., and Bechtel, B.

Response to Referee #2

July 14 2022

We thank the chief editor and the reviewers for their appreciation of our work and the valuable comments on the manuscript. Please find the point-by-point responses below, indicated in blue.

General comments

This work describes a new dataset of land cover types (10 urban – 7 natural) using Local Climate Zones at the global scale. Work is clearly described, evaluated, and presented. The associated dataset is of high quality, and I expect will become a landmark data source for the community.

Thanks for reviewing our research paper, and supporting its publication after addressing the clarifications described below.

The discussion on "accuracy" vs "robustness" could be improved (see specific comments). Additionally, there is no acknowledgment that LCZ training polygons are susceptible to human errors (again see specific comments).

See responses below for details on "accuracy versus robustness".

On the last point. In Section 2.1 we state that "*While the LCZ maps created by individuals are often of poor to moderate quality, The Human Influence Experiment (HUMINEX) (Bechtel et al., 2017; Verdonck et al., 2019a) demonstrated large accuracy improvements (up to 20%) when multiple (poor to moderate quality) training datasets were used together to create a single LCZ map*".

To clarify, we have extended the first part of this sentence to: "*While the training area polygons and corresponding LCZ maps created by individuals are often of poor to moderate quality …*". We believe this statement is appropriate, not only referring to potential errors introduced by the subjective interpretation of the human operator, but also the consequences of perception, interpretation, experience and prior knowledge (see HUMINEX for more details).

Section 3.2 and Figure 10 show that the correlation R2 for building height is only ~0.5, however this is only very briefly mentioned in results, and not mentioned elsewhere (e.g. discussion/conclusion/abstract). So, while 2D information like lambda_B appears to be very well captured, 3D information remains a significant limitation. This is a key result and its implications should be discussed more thoroughly.

Thanks for this suggestion. We have added some notes on the 2D versus 3D information in the discussion (last paragraph), as follows:

" … *More in general, the results of the thematic benchmark reveal that two-dimensional informa-*

tion (urban land cover and building surface fractions) is well represented, but that the corresponding three-dimensional (3D) information requires more attention. Ongoing developments such as the work on the Digital Synthetic City Ching et al. (2019), tailored towards providing more detailed information on the urban landscape (WUDAPT Levels 1 and 2), or global 3D building information (Li et al., 2020a; Esch et al., 2022; Kamath et al., 2022) might contribute to improve the quality of future LCZ map release…. "

A lower reliance on acronyms would assist the casual reader. For example Figures 9 and 10 are not decipherable without referring to other sections of the text.

Thank you for this suggestion. We have revised the captions of all figures and tables in the main manuscript, and adjusted the text where needed, to make sure its meaning is clear without the need to refer to other sections of the text.

However, overall, an impressive body of work.

Specific comments

Line 41: "Earth System Models (ESMs) have only recently evolved to accommodate urban-scale landscapes, even though the parameters that are used by ESMs to these landscapes are limited in scope"

Some global climate models have had integrated urban canyon models for over a decade (e.g. CLMU in CESM). I'm not sure if these are ESMs (ESM relates to the carbon cycle, not the global scale, some readers may misinterpret this). I think safer/clearer to say many global-scale models ignore urban landscapes or represent them simply.

The context of this statement is that in CMIP5, only one of the GCM/ESM models was dealing with urban surfaces (CESM with the Community Land Model (CLM) as land surface model (LSM)). In the most recent CMIP6 archive, there are a few more, all of which use CLM as LSM. So in that sense we believe that the statement on the fact that models only recently (given the long history of global-scale climate models and their LSMs, Fisher and Koven (2020)) evolved to accommodate urban landscapes is accurate. Yet we agree the distinction between GCMs and ESMs might be misinterpreted by some of the readers, so we have adjusted this sentence accordingly.

Fisher RA, Koven CD. Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. J Adv Model Earth Syst. 2020;12(4). doi:10.1029/2018MS001453

Line 120: suggest removing "well-trained" as subjective.

From previous works (e.g. Bechtel et al., 2015, 2017; Verdonck et al., 2019) we know that training of the human operator can improve the classification. This training includes an in-depth understanding of the LCZ scheme, its context in terms of urban climate and thermal characterization of the urban environment, and an understanding of the guidelines that describe the optimal shape, size, frequency, siting, … of the training area (TA) polygons.

All of this context was discussed thoroughly with the RUB students. Afterwards they practiced with the LCZ driving test and by creating TA sets for a large number of known cities that were reviewed by local experts. At the same time, the LCZ Generator capabilities were also used, allowing the students to revise the training areas based on the internal quality control, and subsequently re-submitting the revised training area files until an OA of at least 70% was reached. In this respect note that contributions from RUB members are not visible on the LCZ Generator platform, as we also use the system for internal procedures and tests.

To summarize, we would like to keep the “well-trained” formulation in the text, as we believe this is an important part of the LCZ mapping workflow. Since the way of training our students is generally inspired by the outcomes of HUMINEX, this is indicated accordingly in the revised manuscript.

Bechtel B, Alexander P, Böhner J, et al. Mapping Local Climate Zones for a Worldwide Database of the Form and Function of Cities. ISPRS Int J Geo-Information. 2015;4(1):199-219. doi:10.3390/ijgi4010199

Bechtel B, Demuzere M, Sismanidis P, et al. Quality of Crowdsourced Data on Urban Morphology—The Human Influence Experiment (HUMINEX). Urban Sci. 2017;1(2):15. doi:10.3390/urbansci1020015

Verdonck M, Demuzere M, Bechtel B, et al. The Human Influence Experiment (Part 2): Guidelines for Improved Mapping of Local Climate Zones Using a Supervised Classification. Urban Sci. 2019;3(1):27. doi:10.3390/urbansci3010027

Line 127: “only the best submission is retained” what distinguishes a “best” submission?

With “best” we refer to the submission of the same city with the highest overall accuracy. This is clarified in the text accordingly.

Line 128: How is accuracy determined?

During the filtering process, we mostly use the overall accuracy (OA) metric provided by the LCZ Generator (Demuzere et al., 2021) as a guideline to select good submissions, which is partly in line with the recommendation of Bechtel et al. (2019). But as also written on the FAQ of the LCZ Generator (see here), we do acknowledge the fact that high overall accuracies do not automatically mean that the map is correct, or that all TA polygons are a correct representation of the landscape. Yet dealing with such big data (80000+ polygons before filtering) requires automation procedures, meaning that compromises have to be made to come up with a workable solution.

Bechtel B, Alexander PJ, Beck C, et al. Generating WUDAPT Level 0 data – Current status of production and evaluation. Urban Clim. 2019;27:24-45. doi:10.1016/j.uclim.2018.10.001

Demuzere M, Kittner J, Bechtel B. LCZ Generator: A Web Application to Create Local Climate Zone Maps. Front Environ Sci. 2021;9. doi:10.3389/fenvs.2021.637455

Section 2.4.1: I would describe this as a test of robustness, not accuracy, as this does not test whether the classifications are correct, just whether they change with different inputs. This method also assumes that training areas are accurate, but TAs are classified subjectively by humans. True accuracy can be tested with building resolving spatial datasets. However, I accept this “accuracy” terminology has been established elsewhere in the literature, but a comment to clarify accuracy vs robustness would assist readers.

It is a general problem in LCZ mapping that there is typically no independent testing data available and also there is not even necessarily one true class for each pixel as discussed in Bechtel et al. (2015). Thus we use two independent approaches to test the quality of the product - a comprehensive cross-validation scheme and a thematic benchmark. Thus the accuracy measures always reflect a comparison between the classification result and independent samples. We agree that the metrics are well established and we thus prefer to keep this terminology. We do however provide some more context on the meaning of the accuracy metrics, by adding the following statement in the revised manuscript:

“*It is important to note that these accuracy metrics reflect the consistency of the TA samples, but do not guarantee that the TA polygons are semantically correct. However, since a huge TA database from various sources and cities was used, this gives much more confidence than using a TA set for a single city.*”

Line 200: “The overall accuracy denotes the percentage of correctly classified pixels.” As described above, the method does not assess whether pixels are classified correctly, only how often they are unchanged (and potentially remain incorrect). With poor training data, the overall “accuracy” could approach 100% but be completely wrong. Please rephrase.

100 % accuracy could only be achieved using a single class as training data, which is excluded in the given procedure using balanced training samples. Yet it is true that the measure is based on independent sample data which also contains errors. This was added to the description:

“*The overall accuracy denotes the percentage of independent test pixels that were assigned the same class as the test label. OA_u reflects this percentage for the urban LCZ classes only, and OA_{bu} is the overall accuracy for the built versus natural LCZ classes only, ignoring their internal differentiation.*“

Line 422: While the use of “Global South” and “Global North” is quite common, some see these terms as problematic as they are geographically inaccurate, deterministic, and paternalistic. If authors mean “lower wealth” they could just say that.

This is a valid concern, but we basically just adopted the terminology of the paper that performed this research (Nagendra et al., 2018). We have changed this now to low, middle and high income countries, a terminology used by the Worldbank.

Nagendra H, Bai X, Brondizio ES, Lwasa S. The urban south and the predicament of global sustainability. Nat Sustain. 2018;1(7):341-349. doi:10.1038/s41893-018-0101-5

Technical corrections

None