

Review report for

Towards a global database on building architecture and construction materials for urban climate models

Submitted to Earth System Science Data (2026-201)
by **Lorena de Carvalho Araujo et al.**

The manuscript presents a methodological framework for standardizing global building typology data, addressing an important gap by providing detailed information on building materials and systems that complements existing datasets focused on morphological parameters. The approach tackles data heterogeneity through protocols for survey response assessment, data correction, and validation against external databases such as Tabula and the World Housing Encyclopedia, with reliability assessed using Cohen's and Fleiss' Kappa. While the focus on residential buildings provides reasonable global coverage, the inclusion of non-residential typologies remains limited. More importantly, the reliance on survey data introduces significant coverage gaps, which are addressed through various imputation methods — raising questions about the proportion of data that is empirically grounded versus modelled. The paper would also benefit from clearer structuring to help readers follow the methodological steps more easily.

General Comments

- **Urban Heat Islands versus Urban overheating:** though the urban heat island is the most well known phenomenon in the field of urban climate, there is a general tendency to move away from this concept - especially in the context of heat mitigation, and focus more on urban overheating, e.g. following the paper of Martilli et al (2020, <http://doi.org/10.1016/j.uclim.2019.100541>) and Nazarian et al. (2022, <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022EF002682>). I would suggest the authors to reflect on this change in terminology, and put more emphasis on urban overheating instead of urban heat island.
- **Section 2 Objectives:** By identifying the critical gaps in the introduction, the objectives are implicitly already mentioned. The current section on Objectives in my opinion is more about the workflow used to come up with nation-specific maps/databases of urban properties other than morphological ones. I am not sure this Section is needed here. I do understand that the global LCZ map / Urban Atlas database and their built-up classes are an interesting vehicle to use in combination with the databases. However, I don't think this deserves a separate section. I would suggest to integrate this section with Section 3, and perhaps use a more general Section title, e.g. "Database Design Framework".
- **Survey:** In lines 148 and following, the survey is mentioned at various occasions, with some details. As a reader, I believe it would be relevant - at this point in the paper - to obtain more context on the survey, how it was designed, to whom it was sent etc ... However, the survey is now a separate section 4, even though many

parts of it are already addressed in the current Section 3. I believe it would be clearer for the reader to re-structure this, in combination with the suggestion on a new integral section “2 - Database Design Framework”. Finally, the Survey section is rather long, and might be distracting the reader a bit. Perhaps it is better to reduce the length of this topic to key elements, and put additional information into a supplementary section.

- **Data accuracy and reliability:** Section 5 discusses the data quality and completeness at length, and also provides a detailed - at times a bit lengthy - description of how missing data is treated. However, what is missing in my opinion, is an overview of the % of data per data source, across nations and variables. For example, how many values in the final database are actually sourced from the original survey results? And what about the other data sources? How does this look globally? This type of information can then also be used to identify the strengths and weaknesses of the product, and what should be improved in follow-up research.

Technical suggestions

- Line 40: Add reference to the work of Cathy Li on air-conditioning adoption rates: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023MS004107>
- Lines 60-66: I agree with the narrative here, and the identified critical gaps. However, it might be good to provide some additional background information on the wealth of (mostly morphological) building data that has become available in the past few years, such as (non-exhaustive):
 - Urban land cover data, see review by Chakraborty et al (2024, <https://www.nature.com/articles/s41467-024-52241-5>)
 - Section 6 in Lean et al (2024, <https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4858>)
 - Building footprints via Overture
 - Global Building Atlas: <https://essd.copernicus.org/articles/17/6647/2025/>
 - GHS-OBAT R2024A, <https://data.jrc.ec.europa.eu/dataset/f41a22f1-5741-4c41-86eb-6384654f6927>
 - Open Building Map: <https://www.nature.com/articles/s41597-025-06132-z>
 - EUBUCCO, Europe only: <https://eubucco.com/files/v0.2>
 - UT-Globus: <https://www.nature.com/articles/s41597-024-03719-w>
- Line 67: Clarify this paragraph is about existing datasets on construction materials, wall and roof albedo, window types, and HVAC systems. Note that there is now a “successor” of the Jackson et al. dataset, in the form of U-SURF from Cheng et al (2025, <https://doi.org/10.5194/essd-17-2147-2025>)
- Line 80: in an open-source ...
- Line 87: LCZ (Sect. 2), (add comma)
- Lines 89-91: Section numbering is off, there is two times section 6. This happens throughout the manuscript ... make sure all numbers are consistent.
- Line 96: is there no formal reference for Urban Atlas?
- Line 97: “These land cover classes provide ...” is positioned directly after Urban Atlas, whilst the sentence is about Local Climate Zones?
- Line 108: building materials, architectural practices, ~~practices~~ and systems ... Not sure what “systems” refers to here (and in other parts of the manuscript)?

- Line 127: the maps are rather small? Perhaps you can provide the figure as portrait instead of landscape, to make the illustrative maps larger, to better see the spatial variability?
- Line 167: did the authors integrate nation- or region-specific database, as a complement to the information received via the surveys?
- Line 210: does this statement also hold for many of the developing regions?
- Line 229: there are various schemes to identify household income levels, such as from the World Bank, United Nations, OECD, ... Can you clarify what scheme / definition is adopted here?
- Line 365: Section references in the intro paragraph are wrong (referring to Section 4)
- Line 379, rephrase: "Both measures quantify the extent to which agreement between survey responses exceeds what would be expected by chance."
- Line 396: add reference to Appendix after "Many Windows" category.
- Lines 426-429: in these cases, is there also a metadata field in the GIS database that indicates the origin of the final parameter choice? The same question also applies to what is described in Lines 415-424.
- Line 430: Since U-Surf ([Cheng et al., 2025](#)) is a globally explicit database on a 1km² resolution, developed for (urban) climate models, and partly based on LCZs. I would suggest that the authors explore whether an additional comparison is useful here.
- Line 432: space missing between "the survey.The Tabula"
- Line 466: out of curiosity: is building greening also part of the database, such as various types of green roofs, vertical green, ...
- Line 475: this sentence is unclear to me: "...contain non-exhaustive building lists focused on typologies relevant to their original contexts"
- Line 495: Meaning that the database contains no information on these topics? In that case, perhaps suggestions can be made for alternative sources of information, such as this one: <https://zenodo.org/records/10633387>
- Line 500 and others: I am not sure the term "lighting" is appropriate to describe the variables window share and the types of solar shading. "Lighting" typically refers to artificial light sources or illuminance levels. I would suggest to use either "Daylighting" or "Fenestration" variables.
- Line 535: Reference for "GDUBP database regions"? Write the acronym in full.
- Line 547: Where do the "default" values come from?
- Line 629: Norway pops-out on the map, with "metal" as the dominant wall material in MR-buildings. This surprises me. Has this been verified?
- Line 659: or because of lack of (financial) resources?
- Lines 690 and following: since non-residential typologies are global, how are these then distinct from the information obtained e.g. from the LCZ look-up tables for corresponding building uses?
- Table 6: should LLR be LR?
- Line 698: Make the discussion subsection a separate section
- Line 708: This sentence is unclear? "only the main typological variables (country, use and form) were prioritised over the auxiliary ones"
- Line 716: This was the case **for** some African nations

- Line 726: I think it would be good to also describe the fact that LCZ 7 is missing as a limitation. A typology that is often ignored, even though these areas are host to the most vulnerable populations. Please add this, and add some ideas - based on recent literature - on how future efforts can address this shortcoming.
- Line 742: Note that there is also a CONUS LCZ map available (Demuzere et al., 2020, <https://doi.org/10.1038/s41597-020-00605-z>) as well as a longitudinal multitemporal version, from Qi et al (2024, <https://www.nature.com/articles/s41597-024-03042-4>). Given the latter product, does this new database on building materials also allow the construction of multi-temporal properties, based on the era of construction?
- Line 750: “This database *should* address a crucial data gap ...” does it address this issue or not?
- Line 748: Conclusion and outlook. Some of the outlook sections are in my opinion linked to the limitations described in the previous Section. I would revise both sections, and merge them together. Or, move outlook to the limitation section, and keep the conclusion short, only describing what is actually presented in the paper.
- Line 810: Also reference the supplementary tables in figures in the main text. I e.g. do not see a reference to Fig. B1, B2
- Line 838 - Tables 1 and 2:
 - I am not sure if it is useful to have a “yes” and “no” answer indicating agreement. If it’s no, would a user not want to know what the difference is?
 - What is the hashtag number in the last column?
- Tables B3 and B4: I don’t understand what these Tables are about?
- Line 990: Sentence is repeated 3 times
- Appendix D: What are the references for the physical parameters in Tables D1 - D5
- Line 948: “Table D2” is missing in the title.