

# Building-level exposed asset value modelling for Germany: an Ahrweiler case study

Response to reviewers,

Aaron Buhrmann, Nivedita Sairam, Cecilia I. Nieves, James E. Daniell, Heidi Kreibich, Seth Bryant

## June 10th, 2026

We thank the reviewers for their valuable time and insights. Below is the point-by-point response, unchanged from the June 4th version. In general, the changes to the manuscript can be summarized as:

- Cleaning and consistency of terminology (e.g., 'model' vs. 'workflow'), syntax, and british spelling
- Addition of *Table 3: Method summary of study exposure model workflows and benchmark dataset* to provide the reader a quick side-by-side reference to understand the workflow methods.
- Change of treatment for EHRE "MIX" category building evaluations to be better anchored in EHRE classifications rather than benchmark (which is more consistent with how the other models were treated)
- Change of methods for LoD1+Eurostat residential to employ "usable floor area" estimates rather than simple building footprints. While this adds complexity and reduces the attributability of the workflow, it brings the method more in-line with Paprotny 2020b which, upon reflection, is likely more useful for readers.

## RC1: Anonymous Referee #1, 22 Jan 2026

*The present paper sets out to compare different datasets and methods for estimating the economic value and usage types (exposure modelling) of buildings in Germany at the object level (building-based). Conventional risk models characteristically present data in aggregate form for extensive regions (e.g. neighbourhoods or cities). In contrast, the*

*present article aims to determine the most transparent, sustainable and accurate building-based exposure model for Germany. The development of a hand-labelled benchmark dataset represents a significant contribution to the field. The study is suitable for publication, but several issues require further clarification and discussion.*

Thank you for the positive evaluation of our manuscript. The responses to individual comments are provided below.

## RC1 - Comment 1

*Introduction: The introduction offers a strong and comprehensive overview of exposure modelling, clearly motivating the need for object-level approaches in impact forecasting and local risk management. The discussion of asset value concepts such as replacement cost, depreciated cost, and net asset value is appropriate, but their relevance for different modelling applications could be stated more explicitly. Clarifying early on that the study intentionally compares models with different cost bases would help frame later results and avoid potential misinterpretation.*

Thank you for this comment. We appreciate the point raised. We found it challenging to balance conciseness with providing sufficient context for readers to understand the cost basis. This paragraph in the introduction has been extended to provide the suggested context.

## RC1 - Comment 2

*Data and Methods: The selection and description of datasets are thorough and well justified, and the modelling workflows are described with commendable transparency. The use of LoD1 cadastral data, Eurostat accounts, BEAM, and EHRE reflects realistic choices faced by exposure modellers. The benchmark dataset is a major strength of the study, but its limitations should be more clearly emphasised. In particular, the reliance on a single region with a predominantly residential building stock raises questions about representativeness, especially for industrial and service buildings. A clearer discussion of potential benchmark uncertainty and regional bias would strengthen the methodological credibility of the evaluation.*

Thank you for this feedback. The following sentence has been added: "Because this benchmark is drawn from a single, predominantly residential region, its representativeness is uncertain, especially for industrial and service buildings."

## RC1 - Comment 3

*Results: The sector classification results are clearly presented and reveal important structural patterns. The strong performance of LoD1-based models for residential buildings contrasts sharply with their near-complete failure to identify industrial assets, highlighting a fundamental limitation of cadastral building function categories rather than a modelling error. This finding is important and should be more explicitly framed as a warning against uncritical use of authoritative building function data for economic sector classification.*

This is an excellent suggestion. The following sentence has been added to the discussion: “This contrast between strong residential performance and near-complete failure for industrial assets reflects a limitation of cadastral function categories rather than a modelling error, warning against uncritical use of authoritative building function data for economic sector classification.”

## RC1 - Comment 4

*The comparison of regional asset values shows substantial divergence between models, underlining that exposure-related uncertainty can be of similar magnitude to uncertainty in vulnerability modelling. While the manuscript correctly attributes much of this divergence to differences in cost basis, the interpretation would benefit from more clearly separating effects driven by accounting concepts from those driven by spatial or sectoral disaggregation. The per-asset comparison against BKI construction costs is informative and convincingly demonstrates systematic underestimation, but it should be stated more explicitly that this reflects conceptual alignment with stock-average or depreciated values rather than an inherent model deficiency.*

Thank you for raising this point. With the information available, we are not able to attribute the error sources. The following sentence has been added to the results and discussions to reiterate the limitation: “Thus, these differences should be read primarily as conceptual alignment with stock-average or depreciated values rather than inherent model deficiency; with the available data, we cannot isolate the specific mechanisms driving each discrepancy.”

## RC1 - Comment 5

*Conclusions and Limitations: The conclusions are well aligned with the results and appropriately cautious. The emphasis on transparency, maintainability, and local validation is well supported and constitutes an important message for the community. However, the limits of generalising the findings beyond the Ahrweiler region should be stated more clearly. Strengthening this point would enhance the credibility of the study without diminishing its contribution.*

Thank you for this suggestion. The following sentence has been added: “While these findings are limited to one specific region, the frameworks and methods developed here provide the basis for broader evaluation across regions with different building stocks, economic structures, and data quality.”

## RC2: Anonymous Referee #2, 26 Jan 2026

*I would like to thank the authors for sharing this interesting piece of work. Overall, the piece is well written and interesting. I do have (perhaps obviously as a reviewer) a couple of comments and remarks.*

### RC2 - Comment 1

*I would consider rephrasing the title a little. The authors are indicate in the discussion that the results cannot directly be generalized to the rest of Germany, because it was such a specific use case. Of course, I do agree that certain results could be generalized across Germany, but it would require a couple of more case studies to validate that. As such, perhaps it would be better to emphasize the Ahrweiler region as well in the title.*

Thank you for this suggestion. Reviewer 1 felt similarly. The title has been changed to: “Building-level exposure asset value modelling for Germany: an Ahrweiler case study”.

### RC2 - Comment 2

*In the methods, I miss a brief section that explains how the different models/datasets are being compared. Given that you compare them based on precision and recall, it is generally common practice to already explain this in the methods section.*

Thank you for this suggestion. The following sentence has been added: “Following the supplement workflow (Sect.~\ref{s:sec:classification comparison}), we calculated one-vs-rest true positives, false positives, false negatives, true negatives, precision, recall, and F1 for each benchmark sector, then benchmark-weighted mean precision, recall, and F1 for each model.”

### RC2 - Comment 3

*For the OSM classifications, what is the ratio of CLC vs OSM land use categories. And for consistency, why not fully base this in CLC, instead of only filling the gaps. Especially given the OSM results (and that OSM is also used a lot, so many people are interested in how OSM performs), it would be nice to go into a bit more detail on this. I would be curious to see how OSM performs when fully using CLC land cover classes.*

Thank you for this feedback. We desired to replicate a 'best available' data approach that an exposure modeller might use, hence the use of hierarchical CLC + OSM imputation. The methods section has been expanded to better explain this, and the following sentence added: "This results in most assets (96.73\%) sourcing land use information from the OSM polygons, but in some areas where these are missing the CORINE value is taken".

## RC2 - Comment 4

*To better understand the results, how do the different geometries compare of the buildings? That is not entirely clear yet. LoD1-based and OSM-landuse have all 844 objects, but EHRE has only 699 but it is based on OSM (and some additional 100x100 polygons added?). Line 311 to 329 also reflect a bit on this, but I find it a little bit hard to follow (especially the bit from 321 to 329).*

Thank you for this feedback. We agree these discrepancies are not intuitive and we recognise our lack of clarity. To address this, the aforementioned paragraph has been revised. Also, a new figure (S2) has been added to compare the different geometries of the datasets.

## RC2 - Comment 5

*I also do not fully understand why OSM was omitted for the total asset value Many people use OSM, so it would actually be interesting to see how this compares. I mean, the other datasets also have a very skewed distribution of unit values, which are also slightly hard to follow sometimes on why that is precisely the case. I am not sure if the wrongly mapped service section would be a sole reason to not estimate total asset value. Or at least it would be interesting to see how that would be different. Especially also because industrial was actually very well mapped for OSM? And given that in many countries (outside of Europe) Lod1 type of data is typically not available, I would still proceed on including an OSM-based approach to the end, also to "warn" users on the potential pitfalls. Could be a good source for many people to cite*

Thank you for this comment. An OSM land-use-based asset value model was excluded from Section 3.2 and 3.3 to improve the clarity of the manuscript. The following has been added to the discussion to elaborate on our chosen approach: " Including it in the regional-total and per-asset value comparisons would require joining the OSM labels to an asset-value source such as Eurostat, producing an OSM+Eurostat hybrid that would mix label performance with the same Eurostat value basis already represented by LoD1+Eurostat. This would add redundancy and confusion rather than new asset-value information, while EHRE already provides an OSM-based asset estimate. "

## RC2 - Comment 6

*I propose to remove Table 6. Next to the subjectivity of the weights and scores, they also differ so little (6.9 vs 6 vs 6.3) that it still does not really say much. I think lines 416 to 436 already explain in enough detail how the models/datasets vary.*

Thank you for this comment, we assume you mean *Table 4: Scores and weights for the asset value models*. (not Table 6). While we agree that these scores have a strong subjective component, we assigned them through a critical review of the different models, a method that is not uncommon in complex multi-criteria decision making contexts. Any loss or risk modeller would need to apply a similarly-subjective process to be able to select an input exposure model to work with. In fact, the evaluation of individual attributes increases the objectivity of the process, by forcing the modeller to focus on each aspect at a time instead of thinking of each model as a “bulk”. We thus use the scores to more clearly communicate our evaluation of the models and encourage readers/modellers to follow such an approach. While the ‘Total scores’ vary little as you point out, individual categories have greater variation, which can be useful for readers in their model selection. E.g., imagine a reader who is building an exposure model where sustainability is of the most importance (i.e., they weigh 1.0 rather than the 0.10 we assign). Therefore, this table does provide useful information despite the similar total scores. Alternatively, the table could be moved to the supplement; however, as the discussion in 3.4 makes frequent (implicit) reference to the table, such a move could make the section more difficult for the reader.

## RC2 - Comment 7

*In line 444, one could also argue that risk models, especially for flooding, often do not include this rich detail in exposure characteristics. A more coarse but well-validated unit value through a “simple” sector-classification would already be a very good baseline to aim for.*

Thank you for raising this. The following sentence has been added: “Further, these additional variables are rarely ingested by loss models, and therefore their inclusion in our evaluation would be of limited use to most readers.”

## RC3: Guilherme Samprogna Mohor

### RC3 - Comment 1

*I would suggest a complete restructuring of the manuscript. Instead of focusing on candidate “models”, I suggest presenting how each aspect or component was tackled, that is:*

*(1 – objects) how different building function, land use classes, etc. were harmonised;  
(2 – economic values) how the economic sectors were harmonised;  
(3) how the sectors and building functions were matched; and  
(4) how replacement costs and depreciated costs etc. can be compared (and if not, what were the assumptions to make the comparison in this manuscript). In the current form, these questions are presented per candidate model, but very unclear, and above all, when comparing them it is also unclear.*

Thank you for this suggestion. Your comment that the manuscript is “unclear” is well-taken. To improve readability, we have revised the manuscript as follows:

- Included “Table 3: Method summary of candidate models considered”:
- Added more context sentences to the results sections to better guide the reader through the disparate result metrics (e.g., RC vs. DC)

During manuscript prep, we considered a similar restructuring, and agree such a framing would make the manuscript easier for the reader; however, this is not possible with the datasets we consider because the data we ingest bundles each stage (i.e., aspect/component) differently. E.g., EHRE is a fully formed exposure model whereas LoD1+Eurostat and LoD1+BEAM we build from components. Therefore, to provide an evaluation of each stage as you suggest, would require back-calculating components (e.g., from EHRE), requiring large assumptions to produce an intermediate product (i.e., not something all exposure modellers require). Comparing these intermediate back-calculated products would require major caveats and discussion around the assumptions we took, and maybe numerous ‘scenarios’ to compare the different possible assumptions for back-calculation of intermediate products. In summary, while such a per-component harmonization structure seems simpler upon first consideration, because of the datasets are fundamentally different in their nature, this structure actually introduces complication – and to intermediate phases not of interest to all readers. Therefore, we opted instead for a data-first minimally processed evaluation allowing the reader to apply their own interpretation for all but the final ‘weighted scoring model’ section where we layer on our biases/expertise. We feel this separation provides for a more useful and transparent manuscript, rather than an assumption-heavy harmonized modelling approach.

## RC3 - Comment 2

*With that suggestion I bring another point: over reliance of supplementary material. This work definitely needs supporting material (the suggested restructuring would not avoid that). But in many places an average or range of values could be given in the main script, providing the reader a quick and valuable reference without constantly flipping to the supplementary tables.*

Thank you for this comment. We carried out several iterations to balance the level of detail required in the main manuscript for the reader to follow along and appreciate your feedback that

some details should be moved back into the main manuscript. Towards this, we have made the following changes::

- In the main body, results section “3.1 Sector Classification” added a sentence summarizing Table S12.
- In the main body, methods section 2.4 Benchmark Dataset, added a sentence describing the ranges of values from the supplement table.

### RC3 - Comment 3

*When the candidate models are compared, mostly on the object-level, another very unclear part is about the mismatching building functions, what happened to them, how they were dismissed or counted separately (?).*

We assume this refers to the Sector Classification (Section 3.1) when comparing EHRE (osm-indexed) to the benchmark (LoD1-indexed). In the main manuscript, these mismatches are simply shown as-is on Figure 7, with a brief explanation in the figure caption and a geometry-focused discussion in a following paragraph (Line 311). A more precise treatment is only required when calculating the quantitative metrics, which is wholly provided in the Supplement (S5).

Regardless, your comment is well received and we agree that this is a confusing topic. To make the manuscript more clear, the below figure a description of each case was added to the supplement.



### RC3 - Comment 4

*Before any classes are reclassified or adjusted or partially dismissed, do the original datasets agree on the TOTAL assets within the pilot region?*

This ‘total building count’ comparison is shown on Table 3. The following sentence has been added to the results and discussion to highlight this: “Looking first at the building counts, the two LoD1-based models share the same inventory of  $\{61756\}$  buildings, whereas EHRE includes  $\{54490\}$  explicit buildings and  $\{65056.9\}$  buildings when its remainder layers are included (Table~\ref{tab:model\_results\_full\_extent}).”

### RC3 - Comment 5

*It is a pity that the final published data (the Sanitized data) includes the chosen economic sector, but not the original building function, which could make transparent the matching procedure adopted.*

Agreed. These have been included in the revision.

## RC3 - Specific Comments 1

*Title: “exposure” instead of “exposed”*

Done

*Line 48: “e.g. by federal state in 1946” seems to be a very German-specific information. Please clarify.*

Done

*Line 63: specify better what is meant by “procedure for estimating exposed assets”.*

Done

*Lines 83-85, 98-101, and maybe elsewhere: In the Introduction much is said about loss modelling, giving context and objective for this development. However, exposure and loss should be kept separate. Some references given (and some other points in the Introduction) seem to be only about loss modelling and not the exposure.*

*Added the following sentence: Although these and most related studies focus on loss modelling rather than exposure modelling alone, many of their findings translate directly because not only is exposure modelling embedded within the broader loss-modelling workflow, it is also a precondition for it.*

*Lines 90-95: Would it rather be “per asset type” than “per-asset” or “per-building” replacement cost?*

You are correct, ESRM20 is an aggregated exposure model. It does not provide per-building (or per-asset) values in the sense of assigning individual values to individual footprints, because it does not work with individual footprints. In the context of ESRM20 “per-building” replacement costs refers instead to the cost of replacing a building of a certain “type”. “Type” is defined as a combination of: occupancy case (residential, commercial, industrial, geographic location (not as a building footprint but as a country, region, admin area in general, level of urbanisation), material of the structural system, and expected average area (m<sup>2</sup>). This last sentence is what Table S8 explains, except for the average area part, which is implicit, given that the table reports values/m<sup>2</sup>. The “per-building” replacement cost stems from multiplying the cost/m<sup>2</sup> from Table S8 with the expected average area of a particular type of building (type = occupancy, structure, location). These “per-building” values are the ones then used by EHRE. Thus EHRE ignores the surface area of the OSM footprints.

However, we understand how using the expression “per-building” in the context of asset-level modelling can be confusing for the reader, and we have thus changed the manuscript to read “per-asset-type”.

*Fig 1.: Provide sources of rivers and administrative limits datasets.*

These are shown at the bottom of the respective panels. Added a transparent background so the text is more noticeable.

*Line 122: The gathered data are not all “freely available”, as in Tab. 1.*

This sentence refers to the candidate models, which are all freely available per-table 1 (LoD1 has a free license). Revised the text so this is more clear.

*Tab. 1: could be separated per type of information, similar to Tab S2.*

Thank you for noticing this. Tab S2 is misleading, as the groups are not mutually exclusive. E.g., EHRE provides both building data and asset value data. Dividers have been removed.

*Line 152: BEAM was developed “for” the JRC, not “by”.*

Sentence revised.

*Ch 2.2.2 + Tab2: BEAM already explicitly separates “building” and “content” values. Was there a need to do that again?*

You are correct that BEAM provides these separately for residential, but NOT for non-residential. Therefore, the “per-sector multipliers” described on line 250 were required.

*Lines 171-174: If EHRE values are based on the Census 2011, are the values “referenced to 2020” or corrected to 2020?*

EHRE uses exposure quantities and classifications inherited from ESRM20, which for Germany are based on 2011 census data, but the replacement-cost unit values are expressed on the ESRM20 2020 cost basis. Thus, “referenced to 2020” refers to the monetary replacement-cost assumptions, not to a full correction of the underlying 2011 census-derived building and population data to 2020.

*What is meant by “Supplementary Data Table 1”? It appears many times at different points. Should they be one of the suppl tables?*

No. this is a separate xlsx file that should have been provided for review. We will ensure the editors provide this for the next round. File can also be found here:  
<https://drive.google.com/drive/folders/1OdvGlbOrO0LtHJttsPwDIUH-p4IAKML8?usp=sharing>

## RC3 - Specific Comments 2 - LoD1+Eurostat

*I understand that there are different pathways for Residential and Non-Residential assets. Fig 3, however, does not make that clear. The text also goes back and forth about it.*

Figure 3 has been revised so these separate paths are more clear.

*Tables S3-S6 could be better condensed (which is only 1 step before comparing to the other datasets in the other candidate models; reinforcing what I suggest above about restructuring the manuscript).*

These tables express distinct and disparate categories, concepts, and values; therefore, there are no ways to condense/combine these without losing information.

*Would it be more correct to say this candidate is LoD1 + Eurostat + Insee?*

Your understanding is correct that the candidate model we call “LoD1+Eurostat” uses data from Insee to derive the GFC ratios (along with the national residential values from Paprotny); however, the model name is meant to be a minimal distinguishing description, not a comprehensive one.

## RC3 - Specific Comments 3 - LoD1+BEAM

*BEAM values are NUTS2-level averages of €/ per “land” m<sup>2</sup>, per land-use type; that is, not per m<sup>2</sup> of building, but of land use. That means, the residential building assets €/ per m<sup>2</sup> of e.g. dense urban fabric will be the same in all dense urban fabric within the NUTS2. It does not differentiate per building, but per land use class, regionally at the NUTS2 level. This is not wrongly explained in the manuscript, but also not clear.*

Thank you for raising this point. This is an important distinction that was not properly articulated. Added a sentence to clarify this.

*Moreover, why did you sum the total value in Ahrweiler to redistribute per building instead of keeping them per polygon (land use type, differentiating for example low dense and dense urban areas)?*

Because our manuscript is focused on building-level assets, and BEAM is land-use polygon scale as you mention above. Added a small clarification for this. Furthermore, if we had calculated the total of BEAM polygons only (multiply value by polygon area), unrealistically high (individual) building values would result in polygons with only a few buildings intersecting.

## RC3 - Specific Comments 4 - EHRE

*Line 255-258: Do I understand correctly, the EHRE-class “other” includes the OBM buildings for which there is no equivalence in ESRM20 asset values and are therefore dismissed. The EHRE-“mixed” do have asset values in ESRM20, and are matched to the benchmark-dataset “ambiguous”, is that correct? Fig 5 does not list the mixed-ambiguous class.*

Thank you for raising these questions, this was not clarified well in the manuscript and our re-mapping of “mixed” and “ambiguous” was confusing and has been scrapped.

Yes, “other” assets are dismissed/dropped (considered out of domain for the classification calculations and excluded from value calculations). This category has also been added to Fig 5 for clarity.

“Mixed” is a set of sub-categories – with values – that we ~~were~~ remapping from their raw EHRE occupancies (mostly residential and industrial) to the benchmark ‘ambiguous’ to better align with the semantics of the benchmark labels. However, as your comment highlights, this adds some complication and confusion to the metric calculation that, upon reflection, is not worth it. We added a sentence to this effect. Therefore, we have revised Figure 7 and the quantitative metric calculations to remove special treatment for the ‘MIX’ sub categories. Further, we have added the EHRE categories to the “Supplementary Data Table 1”.

## RC3 - Specific Comments 5 - OSM land use classification

*Why give preference at the land-use level to OSM over CORINE?*

Thank you for raising this as it was not stated. Added the following sentence: We prioritize OSM land use polygons over CORINE gridded land use data as the former provides more granular information.

## RC3 - Specific Comments 6 - Benchmark Dataset

*Line 277: such a mixed used is an existing class in ALKIS/LoD1.*

We understand your comment refers to our use of the “ambiguous” category as a catch-all, rather than employing a “mixed” category. The reason we chose “ambiguous” is that the category includes use types that are clearly not mixed (as shown in the supplementary data table):

- Single, multiple and multi-storey garages
- Warehouse buildings, without mixed use

While we could create a new ‘mixed’ category to hold the single use ‘Commercial buildings, with apartments’, we felt this would add complication without much value as only ¼ of the candidate models have a comparable category. In summary, Fig 7 is meant to provide a high level summary of how the sector classifications compare with a minimal number of categories. Text has been revised to clarify this.

*Line 283: where does this regional factor comes from?*

Clarified.

*Fig7: I understand that Agriculture is not considered in the Benchmark. But what happened to Eurostat “Production” and BEAM “Industrial” sectors?*

No buildings with these labels were included in the sample. Added a sentence to the caption to highlight this.

*Line 298: The “many industrial or ambiguous” are exactly 66 industrial and 68 ambiguous, correct (based on Fig 7)?*

Correct.

*Lines 298-300: What is meant by “insufficient resolution”?*

Clarified that this means ‘label groups are too-broad’

*Line 311-312: Why? Chapter 2.3.3 does not mention LoD1 at all.*

Correct, EHRE is osm-indexed, so chapter 2.3.3. does not mention LoD1.

*Line 316: EHRE-unmatched are still shown in Fig 7. What was discarded from other models?*

Added a sentence clarifying that only assets from EHRE were subject to exclusion (as the remainder are LoD1-indexed)

*Line 333: "over-estimation"?*

Clarified that this refers to over-assignment of the residential label class.

*Line 339-340 – As far as I know, building functions in LoD1 follow the ALKIS standard. If they do, there are "industrial" labels in it, as well as mixed uses (e.g. residence with commerce).*

Yes this is correct, LoD1 follows the ALKIS standard and generally there are labels such as 31001\_1130 = Residential building with trade and industry, 31001\_2110 = production building, or 31001\_2111 = factory. However, these were not present in our case study area Ahrweiler. The same is true for mixed uses (e.g. 31001\_1100 = Mixed-use building with residential use, 31001\_1123 = Residential and commercial building, or 31001\_1220 = Agricultural and forestry residential and operational building.

To aid the reader in following the categories between the different models, legend entries for the empty categories have been added to Fig 7.

*Line 355-356: confusing sentence, please rephrase.*

Done.

*Area BEAM-Industrial and Eurostat-Production equivalent?*

Yes, these are equivalent, meaning the same LoD1 functions are assigned consistently to these two sectors.

## RC3 - Specific Comments 7

*Tab 3:*

*-I suggest presenting all values at Mi € or all at 10<sup>9</sup>*

Done.

*-Market services with 23€/m seems too low*

We assume you are referring to the 21 €/m<sup>3</sup> value for LoD1+Eurostat market services on Table 3. And yes, this observation is what motivated the sentence on line 365: "Looking at the unit costs, both LoD1-based models produced some **implausible values** as a result of disaggregating regional values onto sectors poorly represented by LoD1."

*-Are residential m<sup>2</sup> of footprint or floor area? (please, clarify this everywhere in the manuscript)*

Extended Table 1 to clarify the geometry/area basis of the different costing sources. Extended model descriptions and the limitations section to better discuss this.

*-What are "remainder" in the EHRE? How are there decimal-count of buildings?*

Clarification added to the caption.

*Lines 361-362: Rewrite in a form not to mix LoD1+BEAM €/ building m<sup>3</sup> with BEAM's DC, which is per land area m<sup>2</sup>.*

Good catch. Clarified that this value is from LoD1+BEAM (not BEAM directly)

*Lines 367-368: Both have only 19 buildings.*

Fixed.

*Tab 4: I would expect Benchmark classification to be objective and follow Table S12.*

Good catch. Updated the table and incorporated a more precise metric conversion (described in the supplement).

*All criteria here were given subjectively? By one author or as average of all authors?  
With small or large divergences?*

Added a sentence to clarify.

*Data availability: Not all data sources are freely available.*

Fixed.

*[optional] Lines 45-58 + Lines 376-378: In these paragraphs, the different types of values in loss modelling are discussed. I would suggest taking into consideration the paper of Molinari et al. 2020, which has loss models using different asset values as well.*

Yes, this is an excellent paper that was pruned during one of our editing rounds. We have extended the introduction to include this reference.