Review of Manuscript egusphere-2023-1740

Testing the 2020 European Seismic Hazard and Risk Models using data from the 2019 Le Teil (France) earthquake

This manuscript presents a comparison between building damage states as observed in the field after the 2019 Le Teil earthquake and those calculated by means of combining different components of existing risk models from different sources (not just the 2020 European Seismic Hazard and Risk Models). The damage survey has been processed by the authors according to expert judgment to obtain damage in terms of the EMS-98 scale. Different rupture models from the literature, as well as the USGS ShakeMap for this earthquake, are used to generate several realisations of ground motion fields in terms of peak ground and spectral acceleration (PGA, SA). The PGA values are then converted to macroseismic intensities using conversion equations, and these macroseismic intensities results are compared against the 7.5 value obtained in existing literature from field surveys with the purpose of selecting one rupture model to be used for the subsequent damage calculations carried out using the OpenQuake engine. Three main comparisons in terms of damage are carried out, combining different components (e.g., exposure, fragility, site effects) of different risk models as well as different risk calculation methods/software, and contrasting them against the results of the processed damage survey of the 2019 Le Teil earthquake.

While the work presented in this manuscript is of interest to the research community to understand how different existing models and modelling choices affect the calculated damage and, most importantly, how the calculated damage compares against observations from a real earthquake, the manuscript has many significant shortcomings that would need to be addressed before it can be published in NHESS. I thus recommend that the manuscript be reconsidered for publication after major revisions.

Main Comments

1. In my view, the title of the paper does not accurately describe its contents, due to three main reasons:

   I. The word “testing” is being used loosely throughout the manuscript (see point 2 below).

   II. The paper makes comparisons using a variety of sources of model components (exposure, fragility, ruptures) that are not just from the 2020 European Seismic Hazard and Risk Models (ESHM20, ESRM20). The ground motion model used and labelled as being the ESHM20 one does not seem to be the model actually implemented in ESHM20 but a previous version. When using the ESRM20 exposure model the building classes are “simplified”, effectively changing the ESRM20 exposure model. To my understanding (as such an outline is missing in the introduction), three comparisons in terms of damage are carried out:

      1) Section 3.3.1: Comparison between (a) damage calculated with the Armagedom software, using the vulnerability index approach, EMS-98 vulnerability classes, and an in-house exposure model, and (b) damage calculated with OpenQuake, using fragility models from the European Seismic Risk Model 2020 (ESRM20) selected to be equivalent to the EMS-98 vulnerability classes, and the in-house exposure model converted onto ESRM20 building classes.

      2) Section 3.3.3: Comparison between (a) damage processed from the field survey, (b) damage calculated using the USGS ShakeMap, (c) damage calculated with OpenQuake, (seemingly) using the Kotha et al. (2020) GMPE (not the version used in ESHM20/ESRM20), and the BRGM V_{S30} model (which I infer is the ESRM20 V_{S30} model derived from geology, not used in ESRM20), and (d) the same as (c) but using the ESRM20 V_{S30} model derived from topography (used in ESRM20 for cratonic and subduction areas, but not for shallow crustal areas, which is the case of France). All cases use the same exposure model, a building-by-building model based on the individual buildings from the damage survey to which ESRM20 building classes were assigned by the authors. All cases use the ESRM20 fragility models.

3) Section 3.3.4: Comparison between (a) damage processed from the field survey and (b through g) six combinations of the following components:

   i. Exposure models: (i) the ESRM20 aggregated exposure model defined by administrative unit (one administrative unit), but with a large modification to the building classes that makes it different from the ESRM20 exposure model, and (ii) an in-house model derived from statistical data (8 or 9 centroids), to which ESRM20 building classes were assigned.

   ii. Site models: (i) the BRGM V_{S30} model (which I infer is the ESRM20 V_{S30} model derived from geology, not used in ESRM20), values retrieved for the centroid of the administrative unit or 8-9 points of the exposure models, and (ii) the ESRM20 V_{S30} model derived from topography (used in ESRM20 for cratonic and subduction areas, but not for shallow crustal areas, which is the case of France), with the value for the ESRM20 exposure being a population-weighted average of the whole administrative unit and the values for the in-house exposure model being retrieved from the 30 arc-sec cell that contains each of the 8-9 points.

   iii. Ground motions: (i) the USGS ShakeMap, and (ii) calculated with OpenQuake using the Kotha et al. (2020) GMPE (not the version used in ESHM20/ESRM20).

   As can be seen, no "pure" components of ESHM20/ESRM20 appear to have been being used ("pure" = exactly as they have been used in the ESHM20/ESRM20 models) and several components from other sources are being used as well. The title should reflect that the models being compared come from a variety of sources and decisions from the authors.

III. Finally, "testing […] hazard and risk models" may be misleading, as it can be easily interpreted as testing the full probabilistic seismic hazard and risk models (i.e., probabilities of exceedance of ground motion, average annual losses, etc.), which is not what is done in the paper (and, furthermore, cannot be done using data from one single earthquake).

To sum up, the paper shows comparisons (no statistical tests) of observed damage against damage calculated using components of risk models from different sources.

I believe it is fundamental that a new title be assigned to the manuscript, taking into consideration the comments above.

2) I have found the word "testing" being used loosely throughout the manuscript as a synonym of "comparing", "validating", "verifying", "carrying out quality assurance", etc. The word "testing" usually implies a formal statistical procedure using statistical indicators of goodness of fit, similarity between distributions, etc., which are not what is presented in the paper. The paper mostly carries out comparisons, without quantifying differences across different models/components. Please avoid over-using and over-stretching in meaning the word "testing", rewording where necessary. Some outstanding examples:

   a. The title in itself. The European Seismic Hazard and Risk Models are probabilistic models. The paper uses some of their components to carry out ground motion and damage calculations that are compared against damage observations from one earthquake. One earthquake cannot test or validate a probabilistic model, only its components.

   b. Line 34: Bommer et al. (2013) call their work "quality assurance" and not "testing". Throughout the paper they use the word "check" far more than they use the word "test".

   c. Sections 3.1 and 3.2: These sections are not testing ground motions or macroseismic intensities, they are comparing ground motions and macroseismic intensities calculated with different rupture models (against one value of macroseismic intensity) with the purpose of selecting one rupture to use in the remaining comparisons of the paper. The PGA and SA values are not compared against instrumental measurements at all (values of PGA are mentioned in lines 64-66 but not marked on the plots or mentioned again in Section 3.1). The sections are presented as "tests" when, in reality, they are an intermediate comparative step to select rupture parameters.

3) In line with the first point above, and with the purpose of aiding the reader to navigate comparisons carried out across so many different options, please re-phrase the last paragraph of the introduction to describe more accurately the work contained in the paper:

   a. Lines 46-47: This sentence states that the work is done "to test components of the ESHM20 and the ESRM20" models, giving the impression that only ESHM20 and ESRM20 components will be used, but components from other models are used as well, and these are not mentioned at all here. Please mention the other models used.

   b. Line 48: I suggest not using the expression "scenario simulations" to refer to ground motion scenarios calculated by means of ground motion models, as the word "simulations" is usually used to refer to physics-based ground motion simulations (this is not critical).

   c. Lines 49-50: This sentence may give the impression that "the most compatible scenario simulation" is selected in terms of the one that gives the results closest to the USGS ShakeMap, but this is not what is stated in lines 50-52 or in Sections 3.1/3.2 (and further along in the paper), which show comparisons of all rupture models with respect to each other (including the USGS ShakeMap) and finally comparing intensities against the value reported by Schlupp et al. (2022).

   d. Lines 49-52: The meaning of "the most compatible scenario simulation" and "the most plausible scenario simulation" is not clear. After reading the paper, I believe the authors mean "the most compatible earthquake rupture", or "the earthquake rupture that leads to the most compatible macroseismic intensities".

   e. Lines 46-54: While several sentences are dedicated to explaining the comparison of ground motions and macroseismic intensities (which is only a preliminary step to select a suitable rupture to carry out the damage comparisons), very little is said about the core of the work. Please consider delineating the content of the three damage comparisons in a similar fashion to what I have written above under point (1), or perhaps with a figure. This is relevant to help the user navigate the paper, as so many different considerations/decisions are being made in each case.

4) The authors state (lines 113 and 315) that they are using the Kotha et al. (2020) ground motion prediction equation (GMPE) in the form of its **KothaEtAl2020Site** implementation in OpenQuake. However, all ESHM20/ESRM20 sources indicate that this is not the final GMPE used in ESHM20 and ESRM20. This being the case, the **KothaEtAl2020Site** GMPE should not be labelled as "ESHM20 GMF" (e.g., line 314), as this can be misleading for the reader. A more fundamental implication is that, with this GMPE being used, it is not the ESHM20 ground motion model that is being "tested", as implied in the title. Weatherill et al. (2020) and the ESHM20 report (Danciu et al., 2021) explain that a series of modifications were introduced to the Kotha et al. (2020) GMPE for the implementation in ESHM20 and ESRM20. Fundamentally, and given that the authors of the present manuscript emphasise the comparison of different V_{S30} models, **KothaEtAl2020Site** has a different amplification function for site effects, and the site-to-site variability of the GMPE was calibrated only on measured V_{S30}, which means that an incompatibility arises when using it with inferred values of V_{S30}. As explained in the OpenQuake documentation[1]:

   a. **KothaEtAl2020Site** is a "preliminary adaptation of the Kotha et al. (2020) GMPE using a polynomial site amplification function dependent on Vs30 (m/s)".

   b. **KothaEtAl2020ESHM20** is an "adaptation of the Kotha et al. (2020) GMPE for application to the 2020 European Seismic Hazard Model, as described in Weatherill et al. (2020)". Page 89 of the ESHM20 report (Danciu et al., 2021) explains that **KothaEtAl2020ESHM20** is the GMPE used in ESHM20. Site effects in this implementation depend on V_{S30} and whether that V_{S30} is a measured quantity or inferred from proxies (e.g., slope), so as to account for the uncertainty associated with using inferred values. Page 69 of Danciu et al. (2021) specifies that ESHM20 refers to ground

[1] https://docs.openquake.org/oq-engine/master/reference/openquake.hazardlib.gsim.html#openquake.hazardlib.gsim

motions on the "reference rock" (V_{S30} of 800 m/s everywhere). The ESHM20 logic tree input file[2] also shows that **KothaEtAl2020ESHM20** is being used for the calculations.

c. **KothaEtAl2020ESHM20SlopeGeology** is an "adaptation of the ESHM20-implemented Kotha et al. (2020) model for use when defining site amplification based on slope and geology rather than inferred/measured Vs30". The ESRM20 logic tree input file[3] and its "cut" version used for shallow-crustal areas when comparing against past earthquakes[4] indicate that this is the GMPE used in ESRM20 to calculate losses. Site effects in this implementation depend on slope and geology, not V_{S30} (e.g., second paragraph of Section 3.2 of the ESRM20 report, page 16). ESRM20 uses this model together with the slope and geology of the ESRM20 model, which can be retrieved with the "exposure-to-site" tools cited in the present manuscript.

As a consequence, reference to the **KothaEtAl2020Site** GMPE should be modified so that it is not named as "the ESHM20 GMPE" or "the ESRM20 GMPE". Alternatively, the analyses could be re-done using the **KothaEtAl2020ESHM20SlopeGeology** GMPE and associated ESRM20 site model (slope and geology, not V_{S30}), as in ESRM20. One should also note that using **KothaEtAl2020ESHM20** with V_{S30} values other than 800 m/s would not necessarily be representative of either the ESHM20 or ESRM20 models.

5) Associated with the previous point, I believe it is very important that clarity is added with respect to the site models used in the comparisons. When comparing against Weatherill et al. (2023) (cited by the authors) and the ESHRM20 documentation, the explanations (e.g., lines 266-272) in the paper lack from some clarity:

a. It is not fully clear what the "BRGM's V_{S30} database" refers to, as there are two V_{S30} models in the cited reference Weatherill et al. (2023): one based on topography alone, and another based on geology alone. The ESRM20 exposure-to-site tools (which the authors use and cite in the present manuscript) return the V_{S30} values from the topography-based model, as the comparisons in Weatherill et al. (2023) showed that it performed better than the geology-based one. As Table 3-5 (line 310) shows different V_{S30} values for the two (and quite round values for the BRGM case), I infer that the "BRGM's V_{S30} database" refers to the geology-based V_{S30} model presented in Weatherill et al. (2023). Please clarify in the manuscript.

b. The manuscript would benefit from adding some sentences regarding the resolution of each of the two models, as this is relevant for the reader to understand what is being compared (e.g., in lines 267-272). From Fig. 7 of Weatherill et al. (2023) it looks like in the "BRGM's V_{S30} database" there are three geologic units, associated with three ranges of V_{S30} values (is the uncertainty being sampled to assign values in the paper?). The "point" workflow of the ESRM20 exposure-to-site tool returns the values associated with the 30-arcsec cell to which the target point belongs, as 30-arcsec is the resolution of the model.

c. It is noted that the V_{S30} values returned by the exposure-to-site tool are not used in ESRM20 in France (non-cratonic shallow seismicity). These V_{S30} values are used with the craton and subduction GMPEs selected for the areas of Europe where the shallow-crustal ESHM20 GMPE (i.e., **KothaEtAl2020ESHM20SlopeGeology**) is not applicable (e.g., see page 16 of the ESRM20 report, Crowley et al., 2021). The GMPE used for ESRM20 (i.e., **KothaEtAl2020ESHM20SlopeGeology** in OpenQuake) calculates site amplification based on slope and geology directly, not V_{S30}. Please clarify in the manuscript that the V_{S30} values labelled as ESHM20 are actually not used in ESHM20/ESRM20 in France.

[2] https://gitlab.seismo.ethz.ch/efehr/eshm20/-/blob/master/oq_computational/oq_configuration_eshm20_v12e_region_main/gmpe_complete_logic_tree_5br.xml
[3] https://gitlab.seismo.ethz.ch/efehr/esrm20/-/blob/main/Hazard/gmpe_logic_tree_5br_slope_geology.xml
[4] https://gitlab.seismo.ethz.ch/efehr/esrm20_scenario_tests/-/blob/main/models/esrm20/GMPE/gmpe_logic_tree_5br_shallow_default.xml

d. From my understanding, the site amplification model and V_{S30} maps are part of ESRM20 and not ESHM20, as ESHM20 focused on hazard on the reference rock. Please name them as ESRM20, not ESHM20.

6) In my view, it is necessary to add a map that shows the resolution/locations of the different exposure models and site models, the spatial extent of the municipality of Le Teil, the location of the selected rupture plane, etc. This is important for the reader to be able to understand the different models that are being compared and interpret the differences observed.

7) The conclusions section is too short and does not discuss the results with depth. It only focuses on marginal observations. It consists of three paragraphs, the first (and longest) of which focuses extensively on the comparison of macroseismic intensities (which is not the core of this work), the second of which briefly mentions that the exposure model was a key difference-maker in the results, without elaborating on reasons, and the third paragraph discusses potential improvements to the analysis by changing the criteria used to post-process the field damage survey, highlights the need for more standardised field survey practices, and comments about the importance of accounting for buildings not included in the survey, which has not been discussed in the paper and for which explanations are not given. Please re-write the conclusions focusing on the large number of different model components that have been compared, to reflect the work done.

   I have found the statement about the effect of the exposure model (lines 359-362) quite hard to see in Fig. 5, which shows so many different models. Moreover, lines 323-333 focus on the differences due to the V_{S30} model, not the exposure. I strongly recommend to find alternative ways to show and compare these results (perhaps several plots "grouping" results according to exposure, or V_{S30}), and potentially even to quantify the differences between models, so that it becomes clearer to the reader whether exposure or site effects have had a greater influence in the discrepancies with observed values.

   The importance of including in the calculations buildings that were not part of the damage survey is mentioned in the conclusions (lines 368-369), but I cannot find it discussed before. Please explain why it is important to include those buildings and comment on why the damage survey seems to cover such a small proportion of the buildings of the municipality of Le Teil. Did they only survey buildings on-demand from the owner? Can it be assumed that the rest of the buildings were undamaged? This is important as well to interpret the plots in Fig. 5.

   Apart from this, the first paragraph (lines 350-357) talks extensively about macroseismic intensities calculated with the AS2000 model. The acronym AS2000 is not defined at all within the text. Line 354 suggests the AS2000 has been used to convert from SA(1 s) to macroseismic intensity, and , lines 355-357 highlight that SA(1 s) is not representative of the buildings in Le Teil, but Section 3.2 discusses two models that convert from PGA/PGV (not SA) to macroseismic intensity. I thus infer AS2000 stands for Atkinson and Sonley (2000), one of the conversion models used by the Armagedom software, according to Sedan et al. (2013). However, no macroseismic intensity values calculated using the Atkinson and Sonley (2000) conversion equation are presented in the paper. Please revise and correct as needed.

8) Similarly to the conclusions, the abstract would need a revision to include mention of all other models that have been used, as per my previous comments. Please revise the last sentence of the abstract (lines 17-19), which vaguely hints on conclusions that do not match the conclusions section or the content of the work.

Other Comments on Content

1. Line 56: Please remove "and risk" from the title, as the section does not describe seismic risk in the area.

2. Lines 70-74: While this statement can be generally valid, it is noted that the ground motion model used in ESHM20 is a backbone model whose central tendency is derived from European data that may be lacking representation of such shallow earthquakes with a relatively large stress drop, but whose different branches account for the possibility of having more "unusual" stress parameters (i.e., uncertainty in the

stress drop is treated as an epistemic uncertainty). Please see Kotha et al. (2020) and Weatherill et al. (2020) and consider rephrasing (otherwise it suggests that the authors agree with Causse et al. 2021 in this particular case and believe a priori that the ESHM20 ground motion model cannot be able to represent this earthquake).

3. Line 101, Table 2-1: There are some aspects of the table that would benefit from clarification in the text:

    a. How should the reader interpret the first four columns that contain "R" and empty spaces? Does it mean that while a certain parameter is red, the EMS-98 damage grade is as indicated, irrespective of the other parameters? Are the four components ordered as per a hierarchy? I.e. if both vertical and horizontal structural elements are red, then it is damage grade 5, but if the horizontal structural elements are red and the vertical ones are yellow or green, then it is 4?

    b. The far right column shows all components in green and the damage grade resulting in 1. Is this because all entries in the survey have some sort of damage and thus "green" is to be interpreted as "damaged, but usable" and not include "undamaged"? It calls the reader's attention that everything is green and the damage grade is not zero. Please comment in the paper.

4. Line 106, Table 2-2:

    a. In the caption, please clarify this is the buildings' "final" tag (as opposed of tags by components). "… as a function of the *buildings' final tags* for the entire dataset".

    b. It calls my attention that several green buildings end up classified as ESM-98 damage grade 3, which corresponds to moderate structural damage and heavy non-structural damage. I would expect moderate structural damage to lead to the need of further inspection and repair before the building can be used, while "green" means that the building can be used again immediately. This could be the reason why in Fig. 4 the "observation based" probabilities for damage grade 2 are notably low when compared against damage grades 1 and 3 (the distribution has an unusual "valley" in damage grade 2). Can it be that several of the green buildings that ended up classified as damage grade 3 are, actually, damage grade 2? Moreover, Table 3-6 suggests the authors also believe green should map only to damage grade 1 or 2.

5. Associated with the previous point, there seem to be different probabilities of damage and numbers of damaged buildings from observations presented in different plots and the text, which I have found confusing. I have found/observed:

    a. The probabilities of damage from observations differ in Fig. 4 with respect to Fig. 5.

    b. The numbers of buildings from observations in Fig. 5b are much larger than the 327 buildings included in the damage survey. Why is this the case?

    c. At the same time, the plots in Fig. 5 have two separate categories, "Exp. judg.-based" and "Observation-based", but I have found no explanation regarding what this means, as lines 324-326 only say "*Two of the sources consist of probabilities based on expert judgement ("Exp. judg.-based"), and probabilities based on our conversion of the damage observations to damage grades ("Observation-based")*", but the meaning of "based on expert judgement" is not explained. It is noted as well that "*our conversion of the damage observations to damage grades*" is also "*expert judgment*", and thus the difference between the two requires a more detailed clarification.

    d. The above makes me wonder if one of the two "observation" labels in the plots in Fig. 5 has been created using Table 3-6. I have been unable to find any reference to Table 3-6. Please clarify if Table 3-6 is being used and reference it within the text if this is the case.

    e. If more than one method has been used to obtain damage grades from the survey data (apart from the one described in Section 2.2), all methods need to be specified (and given distinct names/labels) in Section 2.2.

f. The conclusions state "*The proposed testing procedure based on the observed damages could be improved by introducing a probabilistic rule for the conversion of damage observations on the three-level colour tag (red, yellow, green) scale to the EMS-98 damage scale*" (lines 364-365). To my understanding, this is exactly what Table 3-6 is showing. If this is the case, and it has been used, then please adjust the conclusions.

g. I cannot find any reason for Table 3-6 not to be used. Showing and discussing "observed" damage results obtained using both strategies (Table 2-1 and Table 3-6), which is potentially what is shown in Fig. 5 but not sufficiently explained, would convey to the reader the inherent uncertainty involved in the comparison between the models and the observations (i.e., "observations" are not a ground truth), which is fundamental in any comparison between models and data (i.e., the uncertainties do not only exist in the models).

6. Associated with the previous point, please explain in the paper how the ESRM20 damage scale (associated with the ESRM20 fragility models) was converted into the EMS-98 scale, as this is another source of uncertainty in the comparison.

7. Lines 110 and 161: The titles of Sections 3.1 and 3.2 need to be changed, as they do not reflect the content of these sections. Neither section presents a test. They are both a procedure to select a rupture model to carry out the damage comparisons. The first sentence of Section 3.1 needs to be changed as well, as the section does not present a comparison against macroseismic intensities.

8. Line 111 (and other instances): Although the citation of the Wald et al. (2022) paper indicates that it is the USGS ShakeMap that is being used, it would be good to be explicit (by saying "USGS ShakeMap"), as the USGS ShakeMap software is also used by other organisations with their own configuration (e.g., the European ShakeMap, the Italian ShakeMap).

9. Line 114: Which site model was used for the ground motion comparisons?

10. Line 139 states that the ground motions were "aggregated over all exposure centroids", but it is not specified whether the values shown are means or medians (of all points). Please specify.

11. Line 139: It is stated that ground motions are calculated at the exposure centroids. However:

    a. To my understanding, OpenQuake does not calculate the ground motions at the exposure points themselves but at the points of the site model that are closest neighbours to the exposure points (and assigns the ground motions to the exposure points by closest neighbours, not interpolation). This can be checked by looking at the sitemesh_XXX.csv output by OpenQuake, as this shows the locations at which ground motions were calculated. If this is the case, it would be relevant to know what site model is being used and its resolution with respect to the resolution of the exposure points.

    b. At this stage, the exposure model has not been described, and different exposure models are used later on in the paper. Please indicate if the "exposure centroids" refer to the building-by-building data of the post-earthquake damage survey or other locations.

12. Lines 149-150: It would be relevant to comment on whether the USGS ShakeMap for this earthquake was constrained with direct ground motion measurements (from stations) and/or Did You Feel It macroseismic intensity observations. For reproducibility, please include as well the version of the USGS ShakeMap used, as the USGS recalculates ShakeMaps when new data or new algorithms become available.

13. Line 151, Fig. 1: It would help the reader if the vertical axis contained the non-logarithmic values of the IM (potentially side by side with the logarithmic ones, or as a scale on the right side of the plot).

14. Line 181, Table 3-2: Is it relevant to show the parameters for the CA2015 model and not the FM2010 model?

15. Lines 193 and 197 use the acronym "KO2020", which has not been defined.

16. Lines 210-226: There are some aspects of the comparison shown in Section 3.3.1 that are not explained and are relevant for interpreting the results. Please specify in the paper:

    a. Lines 212-213 state that the "ESHM20 ground motion logic tree" was used, but so far there has been no reference to the ESHM20 ground motion logic tree, only to the **KothaEtAl2020Site** implementation of the Kotha et al. (2020) GMPE, which, as explained earlier, is not the one used in ESHM20. Please clarify which logic tree is being used.

    b. Lines 214-215: If "equivalent" exposure and fragility models are being used "*so as to limit the effect of these two factors on the differences between the two estimations*", what is the purpose of this comparison? Comparing a model in Armagedom against a model in OpenQuake? Is the equivalence between the models fully guaranteed? Please clarify the purpose of the comparison presented in Section 3.3.1.

    c. Lines 215-216: Please clarify in the paper the meaning of "the exposure model in Armagedom". I am not familiar with the software, but the paper of Sedan et al. (2013) gives the impression that Armagedom is a software and the user can input any exposure model as desired. Please clarify in the paper how this exposure model was defined.

    d. Lines 215-221: Does the exposure model used in OpenQuake maintain the 9 centroids mentioned in line 217?

    e. Please comment in the paper (a paragraph would suffice) about the details of the damage calculation in Armagedom: use of conversion models to transform PGA into macroseismic intensity, calculation of a mean damage grade as a function of macroseismic intensity, distribution into damage grades under the assumption of a Beta distribution, etc. This method is fundamentally different from the calculation carried out in OpenQuake in terms of PGA/SA, with damage grades directly retrieved from the fragility model, conversion of ESRM20 damage grades into ESM-98 damage grades, etc. Without these details and comparisons, it may not be fully evident to the reader what the purpose of this section is.

    f. Lines 224-225: These sentences compare the values obtained against observations, but the percentages of "heavy" and "very heavy" damage observed are not reported. Please add them in the text. It is also not clear why the observed values are not shown in Fig. 3, given that they are shown later in Figs. 4 and 5 (converting number of buildings into proportions, as in the other plots, or using a right-hand axis with a different scale on the same plot).

    g. Do the OpenQuake damage results correspond to the average damage resulting from all 1,000 ground motion realisations (only mentioned in Section 2.1) and all logic tree branches (if a ground motion logic tree was indeed used)? Please specify.

    h. Does Armagedom calculate different ground motion fields (1,000 as well?) to account for ground motion uncertainty?

17. Line 240: To my knowledge, the most recent reference of GED4ALL is Silva et al. (2022), and the preferred name for this building taxonomy is "GEM Building Taxonomy v3.0":

    Silva V, Brzev S, Scawthorn C, Yepes C, Dabbeek J, Crowley H (2022) A building classification system for multi-hazard risk assessment. International Journal of Disaster Risk Science 13:161–177. https://www.doi.org/10.1007/s13753-022-00400-x

18. Line 240: I would suggest to re-phrase "we selected a GED4ALL building class based on…" as "we defined building classes in terms of the GEM Building Taxonomy v3.0 (Silva et al., 2022), based on the building materials and the number of storeys". The current phrasing may erroneously convey that the taxonomy consists of a pre-defined list of building classes to choose from, instead of a classification system of attributes to be concatenated.

19. Line 245, Table 3-4: It is interesting that fragility models for infilled frames ("CR_LFINF") were selected for dual frame-wall systems ("CR/LDUAL"), instead of using the "CR_LDUAL" fragility models directly (one of which is mentioned in Table 3-3). Please comment in the paper on this choice. Moreover, the reinforced concrete ESRM20 classes selected correspond to different values of the lateral force coefficient, and it is not clear how this could be selected from the damage dataset. Please comment.

20. Lines 249-254: Please specify the GMPE used.

21. Lines 254-256: The label "SM – brgm V_{S30}" suggests that the BRGM model was used together with the USGS ShakeMap. How was this site model incorporated to the ShakeMap? Does this mean the ShakeMap used in the paper is not the one downloaded from the USGS but the authors have run the ShakeMap software themselves? Please clarify in the manuscript.

22. Line 283 (Fig. 4) and Line 341 (Fig. 5): Please clarify if the proportions of buildings in each damage grade stemming from the calculations have been calculated with respect to the total number of buildings (including undamaged ones) or only the number of damaged buildings (which I understand is the case for the observation values).

23. Line 284 (caption of Fig. 4), and Table A3: Please clarify what the acronym "BRGM/CCR" refers to. I find it confusing that it is named in Fig. 4, which corresponds to analyses carried out using the building-by-building exposure based on the 327 surveyed buildings, and then in Table A3, which lists 2,778 buildings, which is the number reported in both Sections 3.3.1 (line 216, "the exposure model in Armagedom, which includes 2778 buildings") and 3.3.4 (lines 293-294, "the second exposure model ("brgm exp.") is based on national statistical data, and includes 9 centroids with 2778 buildings"). Please clarify the relation between the exposure models used in Sections 3.3.1 and 3.3.4: are they the same? Please add reference to Table A3 within the text.

24. Lines 291-293, and Tables A1 and A2: It is not clear why the ESRM20 exposure model is not being used directly as it is, including its exposure-to-vulnerability mapping. The changes introduced by the authors mean that the calculations carried out with this model may not necessarily reflect what would have been obtained with the "original" ESRM20 model. Moreover, the choice of fragility classes for each exposure class shown in Table A1 appears as contradictory. In the screenshot of Table A1 below, I have marked the differences in the classes and annotated the classes used in ESRM20, which can be consulted in the *esrm20_exposure_vulnerability_mapping.csv* file of the ESRM20 v1.0 repository[5]. The differences are associated with the number of storeys (e.g., a 4-storey class has been selected for a 6-and-above-storey class, first row) and the lateral force coefficient and/or design code level (e.g., a low code class with 15% lateral force coefficient has been selected for a no-code class, seventh row). Please justify the need to use a "simplified" version of the exposure model (instead of the original ESRM20 exposure) and explain the criteria used to assign new classes in Table A1 (in the main body of the paper).

Table A1 Selected ESRM20 fragility classes based on the building types in Le Teil according to the ESRM20

Original ESRM20 type	N. buildings	Selected ESRM20 frag. class	# class
CR/LDUAL+CDL+LFC:4.0/HBET:6-	3	CR_LDUAL-DUL_H4 H6	1
CR/LDUAL+CDL+LFC:4.0/HBET:3-5	7	CR_LDUAL-DUL_H4	1
CR/LDUAL+CDN/HBET:6-	2	CR_LDUAL-DUL_H4 H6	1
CR/LFINF+CDL+LFC:4.0/H:2	67	CR_LFINF-CDL-10_H2 CDL-5	2
CR/LFINF+CDM+LFC:4.0/H:1	42	CR_LFINF-CDM-10_H2 CDM-5_H1	3
CR/LDUAL+CDM+LFC:4.0/HBET:6-	1	CR_LDUAL-DUL_H4 H6	1
CR/LFLS+CDN/HBET:6-	9	CR_LFINF-CDL-15_H4 CDN-0_H6	4
CR/LFINF+CDL+LFC:4.0/H:1	76	CR_LFINF-CDM-10_H2 CDL-5_H1	2
CR/LDUAL+CDM+LFC:4.0/HBET:3-5	3	CR_LDUAL-DUL_H4	1
MUR+CL/LWAL+CDN/H:2	378	MUR-CL99_LWAL-DNO_H2	5
CR/LFINF+CDM+LFC:4.0/H:2	37	CR_LFINF-CDM-10_H2 CDM-5	3
MUR+CL/LWAL+CDN/H:1	690	MUR-CL99_LWAL-DNO_H1	6
MUR+ST/LWAL+CDN/H:2	130	MUR-CL99_LWAL-DNO_H2 STDRE	5
CR+PC/LWAL+CDN/HBET:3-5	53	CR_LDUAL-DUL_H4	1
W/LWAL+CDN/H:1	100	W_LFM-DUL_H2 H1	7
W/LWAL+CDN/H:2	43	W_LFM-DUL_H2	7
CR+PC/LWAL+CDN/HBET:6-	1	CR_LDUAL-DUL_H4 H6	1
CR/LFINF+CDN/HBET:3-5	38	CR_LFINF-CDL-15_H4 CDN-0	4

[5] https://gitlab.seismo.ethz.ch/efehr/esrm20/-/blob/v1.0/Vulnerability/esrm20_exposure_vulnerability_mapping.csv

25. Lines 291, 294: Please clarify in the manuscript that only residential buildings from the ESRM20 exposure model are being included in the calculation (I have deduced this from looking at the ESRM20 exposure model for France). Please clarify as well if the BRGM exposure considers only residential buildings as well, and whether it covers the same spatial extent (even better if using a map). Please clarify if the damage observations only cover residential buildings as well.

26. Lines 300-304: By using a weighting scheme for the so-called "ESHM V_{S30}" model but not for the BRGM model, this comparison becomes not just about the V_{S30} models but the different ways of assigning values to an aggregated area. It would be useful to highlight this further in the text.

27. Line 310, Table 3-5: The table shows 8 locations but the text (line 294) says "9 centroids". Please correct where needed.

Language Use, Typos

Please make the following changes.

1. What do the authors mean with "ShakeMap analyses"? It seems to me that, in most cases, the authors simply mean "ShakeMaps". Please revise and re-phrase all instances along the paper. Examples:
    a. Line 14: Just "ShakeMaps in order to…".
    b. Line 49: Just "to distributions given by ShakeMaps".
2. Line 10: "validated individually, *although* testing and validating".
3. Line 12: "*damage from* past earthquakes".
4. Line 15: "components of the 2020 *European Seismic Hazard Model*" (not "Euro-Mediterranean").
5. Line 16: "the *degree* of damage" or "the *damage grade*".
6. Line 22: "insured and uninsured *direct economic* losses". I assume this was the intention, as only indirect economic losses are mentioned otherwise.
7. Line 23: "(PSHA, *PSRA* are…" (not "PSHR").
8. Line 53: Please define V_{S30} in its first appearance (this line).
9. Line 77: "*v*ulnerability *c*lasses" (small letters).
10. Line 93: "data in the *forms that we used are*" (no commas).
11. Line 101, Table 2-1: "Vertical load-bearing" and "Horizontal load-bearing" (not "loads").
12. Line 115: "the ruptures in the *ShakeMap* as well as".
13. Line 121: "scaling *relation*".
14. Line 123: "we assume that its *geometric centroid* is located at the hypocentre".
15. Line 131, Table 3-1: In the caption, "Rupture *parameters associated with* the five source models".
16. Line 156, Fig. 1, caption: "ground motion intensity measures *aggregated from* all exposure centroids".
17. Line 164: "to identify the *ruptures* leading to".
18. Line 168: The equation starts "MCS =" but the subscript of the standard deviation says "MMI". Is this correct? (See line 170 as well).
19. Line 177: "The *CA2015* model".
20. Line 201: "(FM2010), *and* b) the macroseismic intensity".

21. Line 203: “PGA given by ~~*and*~~ the ground motion-to-intensity”.

22. Line 210, caption: “at the exposure centroids *of the BRGM exposure* in the site models…” (or appropriate name for the exposure model).

23. Line 350: “closer to the estimation *of EMS-98 macroseismic intensity* by Schlupp et al. (2022)”. The text before that statement had not yet mentioned macroseismic intensity.

Issues with References

1. Line 384: There are numbers at the end of “Munson” and “Stamatakos”.

2. Lines 396-397: The citation of Crowley et al. (2021) is incomplete (no initials of first names, no DOI, mention of EFEHR Technical Report 002 missing). Please cite as (apply journal formatting style):

   Crowley, H., Dabbeek, J., Despotaki, V., Rodrigues, D., Martins, L., Silva, V., Romão, X., Pereira, N., Weatherill, G. and Danciu, L., 2021. *European Seismic Risk Model (ESRM20)*, EFEHR Technical Report 002, V1.0.1, 84 pp, https://doi.org/10.7414/EUC-EFEHR-TR002-ESRM20

3. Lines 408-411: The citation of Danciu et al. (2021) is not fully correct. Please cite as :

   Danciu, L., Nandan, S., Reyes, C., Basili, R., Weatherill, G., Beauval, C., Rovida, A., Vilanova, S., Sesetyan, K., Bard, P.-Y., Cotton, F., Wiemer, S., and Giardini, D.: *The 2020 update of the European Seismic Hazard Model: Model Overview*, EFEHR Technical Report 001, V1.0.0, https://doi.org/10.12686/A15, 2021.