

## ***Interactive comment on “Residential building stock modelling for mainland China” by Danhua Xin et al.***

### **Anonymous Referee #1**

Received and published: 12 June 2020

To whom it may concern,

I greatly appreciate the authors' efforts to address the issue of “grid-scale building stock modeling”, which is very helpful for the understanding of spatial distributions of seismic risk and beyond. However, I feel the manuscript contains some major problems, rendering it inappropriate for publication at the current version. Normally, a manuscript in the state and quality like this one should be rejected. But, I feel the overall modeling idea and process presented in the manuscript is generally reasonable and acceptable, and part of the problems might be due to language usages or writing experience. Thus, I think the paper has the potential and the possibility to be a valuable contribution to the disaster research community if it undergoes a very careful and intensive revision process. I am therefore still willing to suggest the Journal rating it as major revision.

C1

I strongly recommend the authors making a most careful, detailed, and thorough revision to make sure the manuscript's scientific soundness and expression readability. Below, please see my general and detailed comments:

General comments:

1) Some key elements in the modeling process (including validation section) haven't been clearly explained or presented, making readers have to guess (some of them I can figure out, others not) what they are specifically, how they are produced, and why using them is reasonable, such as F2, F3, 9 equations and others in the modeling process, and UCC, “1.32” in the validation section. I will try to specify them one by one in the next part “specific comments part”.

2) It is difficult for me to believe that the validation section is convincing. (a) First, the central theme of this effort is to disaggregate administrative unit resolution building stock data into “km grid” resolution counterpart. Thus for me, the logically nature and common validation should be verifying to what extent “this study' s result: km grid building stock distribution” is in line with “the actual building stock in those km grids”. However, the paper used two kinds of comparisons in provincial and district levels respectively to validate the reliability of the modeling. For me, this seems something like “self validates self” in “scale” sense, in other words, one “large scale” things verifies another things with approximately same large-scale, which seems to not meet with the paper's central theme: getting large scale things smaller ones. Here, it is worthy to be considered further what is the nature or the essence of the two comparisons/validations the paper provided? (b) Second, in section “3.2.1: Provincial-level based comparison between...”, there is “the value of (A)/(C) varies within the range of 0.31-0.65, which indicates the high consistency...” why and how you concluded this? This is really confused me. If we calculate another kind of ratios, for example, the ratios between this study's results (after getting them provincial level) with each province's population numbers in the current yearbooks, we perhaps could also get a series of ratios with small fluctuations (I am not sure, but worthy of try)... (c) Third, for section “3.2.2:

C2

District-level based comparison between. . .” First, how about other provinces? Second, and more importantly, I feel it might be better to put the two adjustment steps into the whole modeling process (i.e., making them as part of the modeling itself, namely part of Fig1, not regarding/treating them as validations), please consider.

3) Large part of the introduction (esp. the first half part) and the limitation discussion section are both too general. The overarching and/or specific objectives (e.g., why making efforts to address building stock’s financial/economic value), significance, application prospects, and if the planned objectives were delivered haven’t been clearly communicated, which should be relevant in part to the lack of the necessary introduction of other methods that also address building-stock, such as ATC-series (esp.13) of USA, EMS98 of Europe, ATC series-based variants in China, various remote sensing based or associated methods, and so on.

4) Please change the term “construction material” into “structure type” in the corresponding places of the whole manuscript, so as to meet the relevant conventional concept/term of the earthquake engineering field. Please make it clear what does the phrase “residential building stock value” specifically refer to, in other words, please make this value-focused phrase more accurate (i.e., what value? floor area or others?). It should be the financial or economic value of residential building stock in this paper, so please use a proper wording of this meaning consistently in the whole manuscript.

5) there are a good number of statement accuracy, nuance, or language wording/expression problems in the whole manuscript, lowering greatly its readability and making its contents sometimes very hard to follow. I will try to specify them in the next section “specific comments section”, but not all. I strongly recommend authors paying enough attention to this aspect.

Specific comments:

Line 1 (of the PDF doc. of the original manuscript, hereinafter): the paper’s title is too general, which cannot convey clearly the central theme of the work.

C3

Line 11: regarding “. . . . .especially in developing countries”. I don’t think there is a need to supplement this general but vague emphasis, which might make readers wonder "are there other stories (i.e., building damages/collapses are not the leading cause.....) in the developed world? Actually, for the long-established saying “Earthquakes Don’t Kill People, Buildings Do”, it is the same everywhere.

Line11-12: regarding “. . . . .targeted at near-real time post-earthquake mitigation”. I cannot understand why this info. is emphasized (esp., in the abstract), I also cannot figure out well how specifically risk analysis can contribute to this stage. Especially, the whole manuscript doesn’t contain any specific explanation, discussion or connections on this at all.

Line 14: for me, “using population density profile as the proxy” reads awkward, “using population density profile as a bridge” might be better (for the whole manuscript, the same or similar below/hereinafter). In addition, for clarity, accuracy, and information completeness (e.g., using what to do what), “. . .by disaggregating relevant urbanity level data in the 2010- census of each province into km grid scale and using population density profile provided in 2015 GHS as a bridge”, or similar expression like this might be preferable. Please check and consider.

Line 10-24: I feel the whole abstract need to be re-generalized after a thorough major revision of the whole manuscript.

Between 24 and 25: commonly, keywords should be provided.

Line 26-28: (1) “being” vs. “Target B” reads awkward; changing “being” into “including” might be better. (2) “IDDR 2018” and “over years” are inconsistent in meaning (i.e., from 2018 to 2019, there is only 1 year. . . . . (3)further, what is the relationship between the main and the subordinate clauses here?

Line 44: “As such” reads awkward, please check.

Line 48: There are provincial level, prefecture-level, district-level, and grid-level things

C4

that were addressed or mentioned in the whole paper. My wonder is what does the term “country-level” here specifically mean? You mean addressing/treating something in a country as a whole or addressing them/it in a whole country (country-wide)? Please check, and use precise expressions. Do please avoid using the general or vague wording as such in the whole manuscript; and there are quite many of them. Do please pay enough attention to this.

Line 50: (1) please change “construction age and material” into “age and structure type” (for the whole manuscript, the same or similar below/hereinafter). (2) “are used” reads awkward, please check. “can be used”?

Line 64: “in turn” reads awkward. Please check.

Line 25-87(i.e., the whole introduction part): (1) commonly, the specific objective of the work should be clearly communicated at the very end of the introduction. Unfortunately, the current manuscript doesn’t provide this. Do please add. (2) a large part of the introduction (esp. the first half part) is too general, and even not directly relevant to the central theme of the paper. Especially, I don’t think there is a need to communicate those kinds of basic knowledge in an academic paper, such as what is hazard, exposure, and vulnerability (and frankly, I feel some of the existing expressions of these terms read not that accurate). Instead, I feel other methods that also address building-stock, and advantages and disadvantages of them should be succinctly discussed, including ATC-series (esp.13) of USA, EMS98 of Europe, ATC series-based variants in China, various remote sensing-based or associated methods, and so on, which should be a great help for the authors to refine the specific objectives of their current work. Do please consider this.

Line 97: Please change “construction material” into “structure type”.

Line 105-106: the current sentence “one advantage of the 2010-census data is its further categorization of data into three urbanity levels, which better reflects the regional difference within each province” is inaccurate and even not reasonable; because, for a

C5

given province in the 2010-census, the associated urbanity levels were only provided for the province as a whole, there aren’t spatial distribution info.

Line 124-125: Please make the meaning of the sentence “. . .before disaggregating the urbanity-level based census data into each grid” clearer, i.e., make it clear that you will disaggregate which set of data; and it might be good to revise the sentence as “before disaggregating the urbanity-level based data in 2010 census into each grid”. Do please avoid such kind of vague or incomplete expressions in the whole manuscript, so as to make the text easy to follow.

Line 128-129: I really cannot understand what is the relationship between “Aubrecht et al. (2015) and Gunasekera et al. (2015)’s approach (although I did read these two papers)” and “the urban/township/rural population proportions of each province in 2010-census data set”; aren’t these proportions provided in the 2010-census directly? Or cannot these proportions be easily calculated from info. in the 2010-census directly? Or, did I misunderstand your original intended meaning here? i.e., when talking about “the urban/township/rural population proportion of each province” here, you refer to those of the 2015 GHS data set, right? So, do please try to make the sentence of such kind as accurate as possible, so as to avoid getting readers lost.

Line 130-131: “the population proportion of urban/township/rural urbanity level is 76.64%, 12.66% and 10.7%, respectively”, which means that the population proportion of urban/township/rural urbanity levels “in the 2010-census” are 76.64%, 12.66% and 10.7%, respectively”, right? If so, please make this info. complete and clear.

Line 131-132: The sentence “Then the grids (1km×1km) in 2015 GHS population density file of Shanghai are sorted from the largest to the smallest” reads awkward. Is it better if changing it into “Then the grids (1km×1km) of Shanghai in 2015 GHS file are sorted from the largest to the smallest in population density”? Please check.

Line 128-144: I guess there is an important assumption here, namely, the larger the population density, the higher the urbanized extent. If so, please write this out clearly

C6

to avoid making readers have to guess. Please check and revise.

Line 163: "up to now" reads awkward, please check.

Line 165: "from the 2010-census" reads awkward, please check.

Line 169-170: I don't think "F2" has been explained clearly, including how they were calculated and how they were used subsequently in the amplification of the 2010-census data. I can guess these. But I think it is necessary to introduce them clearly in the text (e.g., using one example), so as to avoid making the readers have to make that guess.

Line 179-180: Similarly to "F2", the wording for "F3" is also vague and incomplete. Please check and revise.

Line 183-185: there is "the population in each grid living in building types grouped by number of storey (1, 2-3, 4-6, 7-9,  $\geq 10$ ) or by construction material (steel/RC, mixed, other, brick/wood) can be derived". It is hard to guess how you achieve this. I guess there is another very important assumption here. Specifically, from the 2010-census, we can get a series of provincial level percentages of the population living in buildings with different floors or with different structure types with one urbanity level (urban, township, or rural); then it is assumed that all the grids with this same urbanity level are all evenly/uniformly have these same percentages. My guess may be correct, may not. But the author should make this highly generalized statement clear enough, so as to make this key modeling step easier to be understood.

Line 186: "the number of buildings" reads awkward relative to the main topic of this paper. Please check and revise.

Line 201: "currently" and "for instance" both read awkward. Please check and revise.

Line 205-206: there is "9 equations". My wonder includes what are they and how they function specifically? Please explain them in detail. There is the phrase of "linear problem". why you suddenly say this, why the problem is linear, how this linear problem looks like?

C7

Line 210-223: this part is very hard to follow. Please check and provide necessary details and explanations. For example, (1) in Line 216-219, there is "the remaining steel/RC buildings are proportioned to other storey classes from highest to lowest", and the like. Please specify, how you get these proportions? (2) regarding step 6 here, I think it is simply that the remaining buildings in each storey class are all belonging to "mixed" buildings.

Line 226: this sub-title cannot meet with the contents below.

Line 253-368: i.e., the two validations. Specific comments, please see the second point in the general comments section above.

Line 306-307: "on the other hand" reads awkward. Please check and revise. However?

Line 330: I feel more information regarding "UCC" should be provided, so that, it is easier for readers to understand why using UCC can make that adjustment. (existing studies of other researchers might have discussed this, but in the interest of the common requirement that a single paper had better be self-standing, key info. should be introduced)

Line 347-350: I feel the explanation of this "de-amplification factor" is not clear enough, more info. is needed. For example, I found that "1.32" is exactly the arithmetic mean of "1.33, 1.34 and 1.29". Is this just a coincidence? Please check and revise.

Line 414-420: I feel the current limitation discussion is too general. Instead, I feel the most relevant and direct limitation discussion (disadvantages and future improvement directions) should focus on those assumptions and "factors" that this modeling process used.

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2019-385>, 2020.

C8