The study deals with the analyses of various seismic risk scenarios for the Sabana Centro region in Colombia, located in the northern region of the capital city Bogotá, which concentrates important industrial facilities, educational facilities. It is an interesting study that follows the state-of-the-art procedures of scenario risk analyses (at least up to the computation of direct losses), for a region that hasn't been studied before, and thus is a good contribution to the scientific literature that communicates the seismic risk in the country.

Having said this, there are some comments about important issues in the study, that hopefully will help improve its clarity, coherence, and thoroughness. It must be said that after doing the review, the reviewer saw that many of the comments and limitations of the study were included in the discussion section as further developments, however there are many that should be included to make the results sound and representative of the region of study, otherwise many of the presented results could be very misleading.

Dear reviewer

You kindly spent time delving into our manuscript, and we are grateful. Thank you for acknowledging the value of our study. We will do our best to provide a complete explanation of the limitation of the study to make the results sound and representative, as you request in your comments.

Section 2

1. It would be interesting to include why the Sabana Centro region is of particular interest. In previous studies of the major cities, generated GDP or % of the population in comparison with the whole country were presented as reasons for the study of a particular city or region.

R/: Thank you for your comment. This is an important region in the department of Cundinamarca and for the authors is an interesting area for its economic and social growth rate in recent years. We state this in the manuscript, but to improve the statement, we added the following sentence to the updated manuscript:

Cundinamarca is one of the four most populated regions of the country and Sabana Centro is the province that contributes the most population (18%) to the department and also contributes 32% of the department's GDP.

2. Additionally, it is mentioned that it concentrates many economic and industrial activities, but at the end the analysis only deals with the residential building stock.

R/: The previous sentence was mentioned to highlight the fact about the province and its contribution within the region. As the reviewer says, we only deal with buildings for residential use. This is the predominant use in the study area. In addition, there is not enough information in the region to characterize commercial and industrial buildings.

3. Line 105: "The majority of the building stock of the region is comprised of one- and two-story houses" It would be good to show a reference with the numbers based on the 2018

Census for this. It is interesting that this is mentioned and still no two-story houses are considered in the analyses.

R/: Unfortunately, there is no information in the census related to the building height and the information in the national cadastral database does not include Sabana Centro. The phrase was based on the authors' fieldwork observations in the region based on 9000 remote and field surveys. Based on your comment we will include this information in the section that deals with the exposure model.

Section 3.1 Seismic hazard

4. Line 171: "eighteen crustal events were selected from the catalogue to be used in this study" Does this mean that only 'historical' events included in the catalogue were included? No new possible events from the event-based tables from the PSHA model?

R/: We appreciate your comment because our sentence was not clear. We indeed used eighteen events from the PSHA model we develop. Aware that our explanation is unclear, we updated it in the revised manuscript to clearly state our procedure.

5. Table 2: Include column with the distance to the population centroid taken as reference point in the disaggregation to be able to compare this selection of scenarios with the disaggregation graphs presented in Figure 4 and their representativeness in the overall 475-years return period hazard in the region. Given the proximity of the events based on the disaggregation, directivity effects should be considered for some of these events. As further seen in the discussion this was not considered but it would be good to mention it and not leaving it till the discussion as a further development.

R/: We appreciate this comment. Based on your suggestion and other reviewer's comments, we will include a figure instead of Table 2, allowing to visualize the location of the events. We will include a discussion about the directivity effects on this section.

Section 3.2 Exposure model for the residential building stock

6. Was the replacement cost updated to 2021? 2022? In which way was this done if indeed it has been updated? If not, it should be done and explained. Were the new inflated exposure building numbers (based on population as proxy) in any way compared to the dwellings or building numbers reported in the 2018 Census for these regions?

R/: Thank you for your questions and recommendations. The replacement cost was not updated in the original manuscript. Based on your comment, we decided to update the replacement costs using information of current prices for housing in the region. We compared the number of dwellings with those reported by the 2018 census and the difference is 28%. This difference is mainly because the census considered dwellings whose wall material is poured concrete and we did not include it since it was not possible to assign a percentage of construction. Clarifications in this regard are included in the manuscript and we decided to update the exposure model and use the information of the dwellings in the 2018 census instead of the one based on population. Nonetheless, we think if worth highlighting that in the absence of census information, inferring based on population can provide a result with a moderate degree of approximation.

7. Table 4 only considers unreinforced masonry of 1 storey, which is known to be less vulnerable than the unreinforced masonry of 2 stories, which is actually more common in many urban areas. This typology should be included (assuming something probably based on census data or the surveys), as in the region it is very common to find 2-storey, in some cases more than single storey houses (as previously mentioned in the study also). In the current version, the study may be underestimating the losses.

R/: Thank you for this important comment. Other reviewers also raised their attention to the need of updating the exposure model, further discretizing between building heights. We agree specially that the difference between one- and two-story houses is significant. Based on this comment and the other reviewers' suggestions, we decided to update the exposure model using information of the 2018 national census, with further differentiation between building heights. To discretize the building height, we used information collected using remote and field surveys of 9000 houses in the Sabana Centro region, conducted by students of Universidad de La Sabana. We found that for houses between 1 and 3 stories, 34% are one-story houses, 48% are two-story houses and 18% are three-story houses. Accordingly, we updated our exposure model.

In addition, we decided to use the fragility functions calculated by Martins and Silva. (2021), which all use the same modeling approach, and they account for the differences in height. Furthermore, we also updated the discussion section.

Section 3.3 Physical vulnerability of residential building stock to seismic ground shaking

8. Chilean wood structures are known to be in better shape than those in Colombia, and they consider a different type of construction technique. The same goes to the curves used in HAZUS, which are not as representative of the local conditions and may be underestimating the risk. If they are going to be used a more thorough explanation of the limitation of using them should be included.

R/: Regarding the selection of fragility functions, we acknowledge the reviewer is correct about the representativeness of fragility functions.

After updating the exposure model accounting for differences between one- and two-story houses, for consistency in the revised version of the manuscript, we decided to use only analytical fragility functions. Most of them were those calculated by Martins and Silva. (2021) and Villar-Vega et al. (2017). These functions are analytical and all of them use the same modelling approach. We used these functions for several buildings, with some exceptions such as the thin reinforced concrete buildings, for which we kept the fragility functions developed by Arroyo et al. (2021), because these are also analytical and use a more accurate model for these types of structures. We kept the comments about the issue that the fragility functions by Martins and Silva. (2021) were not developed accounting for the particularities of Colombian construction.

In the case of wood houses, we decided to keep the fragility curves calculated for Chilean buildings. These curves were developed by the authors using detailed drawings of Chilean houses, which are similar in configuration to the wood houses found in Sabana Centro. Regarding the method, we used a single DOF oscillator, like those used by Martins and Silva. (2021) and Villar-Vega et al. (2017). In general, the wood houses found in the field surveys in the region have good quality materials and adequate construction, thus we consider that the fragility curves used are suitable.

9. Table 5 needs a clarification of what each curve considers in each of the damage states. If the vulnerability model is considering a unique consequence model, there may be incompatibility between the loss ratios of the derived vulnerabilities, as each one considers each damage state in a specific different way. This is one of the main issues when combining vulnerability functions from different sources. This is particularly true given the damage results of the studies are shown considering these categories of the damage states. The reviewer saw this mentioned in the further developments of the discussion, but is should be included in the computation of the vulnerability curves here in some way, for the results to be coherent.

R/: We appreciate this important comment. Indeed, we agree with your argument and as we mentioned in our previous response, we decided to use a set of fragility functions derived using the same methodology.

10. Given the exposure is not considering separately the 2-storey housing, there are no vulnerabilities for 2-storeys considered, even when it is more common in the urban environment than the single storey houses. This should be included.

R/: Again, thank you for pointing us to this fact. As we mentioned in response to comment # 7, we updated our exposure model accounting for one, two and three-story houses.

Section 3.4. Social Vulnerability (SV)

11. Major comment: One of the main criticisms of the paper is the consideration of the social vulnerability index as a percentage increase using the expression (1+SVI). As stated in the study “The min-max normalization was used to standardize the SV indicators from zero to one to estimate the SVI per municipality. Higher score indicate more socially vulnerable municipalities, and lower scores reflects less vulnerable ones. Then, the indicators were integrated by summing them with equal weight, as followed in Contreras et al. (2020c). The resulting SVI index is therefore used to adjust the percentage of economic losses with respect to the cost presented by the building inventory, i.e. multiplying them by (1+SVI) (Carrenio et al., 2007).”. The problem with this is that there is no analysis done on the significance of the variables included within the study and no way to know if there are variables that shouldn’t be included and if anything is counted double.

Thank you for your comment. The reviewer is correct; in the submitted manuscript version, there was no analysis done on the significance of the variables included within the study and no way to know if there are variables that shouldn't be included and if anything is counted double. Therefore, considering this observation which is the same written by Reviewer 1, to avoid problems interpreting the model and overfitting, we checked the multicollinearity by looking at the variance inflation factor (VIF) of each variable and indicator. The VIF was identified in a linear regression that included collinearity diagnostics produced in SPSS (Field, 2005). We excluded those variables and indicators that were potentially correlated with others and those that did not add significant information according to the collinearity diagnostics (Table 2). Eventually, the model included only independent and relevant indicators to estimate the SV in the case study area (Tables 3 and 4).

Coefficent[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	0.728	0.000					
	Indigenous population	1.178	0.000	0.124			0.047	21.282
	Population density (inhabitants/km2)	-3.327	0.000	-0.047			0.030	33.453
	Number of people per household	-3.378	0.000	-0.012			0.029	34.994
	Population unemployed	10.076	0.000	0.134			0.191	5.222
	Population with unsatisfied basic needs	-8.921	0.000	-0.099			0.023	43.108
	Total population in poverty	7.205	0.000	0.129			0.238	4.203
	Households with no electric energy access	4.234	0.000	0.153			0.078	12.894
	No sewage system	1.160	0.000	0.123			0.351	2.848
	Illiteracy rate	-1.619	0.000	-0.027			0.209	4.785
	Deceased due to COVID-19	11.379	0.000	0.710			0.044	22.967

a. Dependent Variable: SV

Table 1. Variance inflation factor (VIF) detected.

Excluded Variables/indicators[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Female population	.[b]				0.000		0.000
	Age dependance	.[b]				0.000		0.000
	Total population	.[b]				0.000		0.000
	Number of households	.[b]				0.000		0.000

Excluded Variables/indicators[a]							
Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
					Tolerance	VIF	Minimum Tolerance
Population looking for employment	.[b]				0.000		0.000
Household with computer and internet	.[b]				0.000		0.000
Households with access to improved water source	.[b]				0.000		0.000
Education level completed primary	.[b]				0.000		0.000
Education level secondary	.[b]				0.000		0.000
Population enrolled in education institution	.[b]				0.000		0.000
Hospital , clinics per 1000 population	.[b]				0.000		0.000
Population with no healthcare	.[b]				0.000		0.000
Population registered to national healthcare	.[b]				0.000		0.000
COVID-19 cases confirmed	.[b]				0.000		0.000
COVID-19 cases active	.[b]				0.000		0.000
People recovered from COVID-19	.[b]				0.000		0.000

a. Dependent Variable: SV

b. Predictors in the Model: (Constant), People dead due to COVID-19, Total population in poverty, No sewage system, Number of people per household, Native indigenous population, Population unemployed, Illiteracy rate, Population density (inhabitants/km2), Households with no electric energy access, population with unsatisfied basic needs

Table 2. Excluded variables/indicators.

Collinearity Diagnostics[a]														
				Variance Proportions										
Model		Eigenvalue	Condition Index	(Constant)	indigenous population	Population density (inhabit-ants/km2)	Number of people per household	Population unemployed	Population with unsatisfied basic needs	Total population in poverty	Households with no access to electricity	No Sewage system	Illiteracy rate	People dead due to COVID-19
1	1	8.495	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	1.004	2.909	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00
	3	0.740	3.389	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.01
	4	0.517	4.053	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.02
	5	0.102	9.134	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.09	0.00	0.01	0.04
	6	0.066	11.307	0.00	0.00	0.00	0.00	0.03	0.00	0.15	0.04	0.02	0.02	0.00
	7	0.044	13.893	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.01	0.01	0.16	0.00
	8	0.017	22.300	0.00	0.00	0.01	0.00	0.21	0.01	0.01	0.08	0.07	0.01	0.11
	9	0.011	27.201	0.00	0.00	0.00	0.00	0.00	0.02	0.12	0.00	0.03	0.32	0.03
	10	0.004	48.357	0.00	0.04	0.24	0.00	0.75	0.00	0.45	0.73	0.08	0.02	0.08
	11	4.453E-5	436.754	1.00	0.89	0.74	1.00	0.00	0.96	0.16	0.05	0.42	0.47	0.71

a. Dependent Variable: SV

Table. 3. Selected variables/indicators.

Composite indicators	Indicators
Population	Native Indigenous population Population density (inhabitants/km2) Number of people per household
Economy	Population unemployed Population with unsatisfied basic needs (UBN) Total population in poverty
Infrastructure	Households with no electric energy access No sewage system
Education	Illiteracy rate
Health	Deceased due to COVID-19

Table. 4. Selected composite indicators and indicators.

12. Additionally, considering this index as a "percentage increase" is extremely misleading. If there was a way to correlate the SVI of each variable in economic terms to the direct economic loss, then this could be done. But this is not done and there is no parametric study or anything else to validate any of the assumptions.

R/: Thank you very much to the reviewer for raising this interesting question. Little research has tested the correlation between social vulnerability (SV) and losses. To our best knowledge, the relationship between SV and modeled losses has been so far informative rather than indicating that total losses (measured as dollar losses or debris generated) increase with SV (Schmidtlein et al, 2011). However, it was found that only relative losses (dollar losses per average family income) tend to increase with SV. Case study areas with a low SV tend to have more material goods with significant monetary value (dollar) exposed to risk, than areas with high SV. Therefore, we should expect a negative correlation between property losses and SV (Cutter and Finch, 2008). It is important to understand that while total loss (dollar) in case study areas with high SV is lower, the impact of those losses in their communities is high (Schmidtlein et al., 2011).

We will include a discussion of this in the updated manuscript.

13. This SVI cannot be considered a percentage unless there is backup data validating this. This has been done also in fatality modelling where the models that are presented in any publication are previously calibrated and validated with data from historic events. Moreover, considering previous events reporting post-loss amplification that include costs from the response and recovery stages in some disasters, it has been shown that numbers over 30-40% are almost non-existent (what is demand surge? Olsend and Porter 2010), while this study mention cases with increases of up to 60%. There may be a problem with the explanation of the methodology, but as it is right now it is very difficult understand how it can relate to economic losses, especially direct physical losses after an event.

R/: Thanks again for this question. Social vulnerability assessment considers variables and indicators with different units, e.g., number, percentages, number of people/m2, etc. This is the main reason to use normalization to construct the social vulnerability index (SVI), which is why there is no unit, and the levels of SV are expressed in ordinal categories, e.g., high, medium, and low, according to ranges defined by authors. The reason to use percentage is that this value is integrated with the percentage of economic losses with respect to the cost presented by the building inventory.

14. (These limitations are also afterwards mentioned by the authors in the discussion, but it is a MAJOR limitation of the inclusion of the SVI methodology in the results in this study, as there is no validation or calibration of any kind for the methodology)

R/: Data related to the damages caused by natural hazards are usually low quality. It is difficult to compare the severity of the event's characteristics in different zones of any case study area (Schmidtlein et al, 2011), because it is wrong to assume that losses are uniformly distributed (Cutter and Finch, 2008). Unfortunately, SV is estimated at the national or regional scale, making it difficult to calibrate or validate any SV model based on damages after earthquakes, which will be an ideal procedure, as the reviewer suggested.

Overall, considering all your valuable comments about the SV, we decided to improve the calculations of these indexes, as indicated in the response to previous comments. We did, however, decided to keep the estimation of the losses based on this index, but we were careful to clearly state that we did so to provide an approximation of the impact of social vulnerability on economic losses, and that there is a need of a more accurate procedure to do this. In addition to commenting this on the losses section, we also added a paragraph to the discussion.

15. Table 9 numbers are misleading as a direct non-weighted average of the 18 scenarios is not probabilistically and statistically sound. It should consider the contribution of each event, otherwise the less probable events are counted in the same way as the more probable ones. In this way, as when computing AAL from a probabilistic analysis, the contributions should consider the probability of occurrence of each scenario. After saying this, it is advised not to present this table and instead present one with the analysis of each scenario done separately as in a deterministic approach, unless it is possible to demonstrate that the 18 scenarios included account for the 100% of the 475 years return period loss and a weighted average is calculated based on the contribution of each.

R/: We appreciate this comment. We agree that 18 scenarios are not enough to represent 100% of the hazard. We should though mention that the selected scenarios do represent a significant percent of the contribution to the seismic hazard, as the examination of figure 4 can reveal, particularly for the PGA. We added this comment in the corresponding section of the revised manuscript.

In the light of the comment, we decided to remove table 9 and instead include a figure that shows the statistical variability of each damage state for the eighteen scenarios.

Section 4.4:

16. As stated previously, to present these analyses, it would be important to show the contribution of the 18 scenarios to the total hazard in the region (based on the disaggregation results). If not Figure 13 is misleading, considering that it says: “Within the municipalities, the mean percentage of losses is presented with respect to the total expected losses in the region”.

“The economic losses experienced by a province due to an earthquake depends on the event’s epicenter as it is depicted in Figure 12.” Figure 12 does not show in any way anything regarding the epicenter. It may not be only the epicenter but also the Mw for each event the cause of the differences, so this statement is not provable from the Figure. Delete it.

R/: We understand the reviewer comment and agree that each scenario has a different contribution to the hazard. Accordingly, we decided to remove Figure 13. We kept figure 12 but we improve our analysis of this figure. The main point here is that Sabana Centro has a uneven distribution of the building stock, therefore, similar earthquakes with the same magnitude and similar depth such as the Mw 6.25 Sopó, Mw 6.25 Tabio and Mw 6.25 Nemocón have different consequences in terms of economic losses.

Discussion:

17. Just until this section this is stated: “The simulations of eighteen seismic scenarios with a return period of 475 years show that half of the building stock will experience some degree of damage”. How was this ‘475 years’ return period calculated? Even when the disaggregation was done for the ‘475 years’ return period, how can it be confirmed that the 18 scenarios add up to the 100% contribution for the hazard for this return period? Either way this statement should be included in some way in previous sections and not only until the discussion.

R/: Thank you for this observation. As we mentioned in answer to question 15, we will add this to the corresponding section in the manuscript, before the discussion.

Effects of SV

18. Major comment: The main criticism for this approach is also stated by the authors in this sentence: “First, as not all social aspects exert equel effect after an earthquake, it is necessary to develop a weighted approach to best estimate a more realistice SVI for earthquake events.

R/: Thank you for bringing this statement to our attention. We have decided to eliminate it because the allocation of a weighted approach not only will not capture all social aspects after an earthquake but allocating weights to the composite indicators will add subjectivity to the analysis.

19. A second improvement required is to devise a better way of estimating the economic impact of social vulnerability. One potential approach is to generate a database of past earthquakes with different consequences that include the economic costs”. There are not needed future improvements but major limitations of the proposed approach. In this kind analyses, as when performing a linear regression, it is important to avoid counting double and establish the significance of each variable within the analysis, if not this could be overestimating the vulnerability and losses in the region considerably.

R/: Thank you again for this observation. We agree with the reviewer that to avoid problems interpreting the model and overfitting, we checked the multicollinearity by looking at the variance inflation factor (VIF) of each variable and indicator. The VIF was identified in a linear regression that included collinearity diagnostics produced in SPSS (Field, 2005). We excluded those variables and indicators that were potentially correlated with others and those that did not add significant information according to the collinearity diagnostics (Table 2). Eventually, the model included only independent and relevant indicators to estimate the SV in the case study area (Tables 3 and 4 shown above in the answer to comment 11).

However, in this sentence, we were referring to the economic impact of the SV after an earthquake, not the method for the vulnerability assessment. Considering that this sentence can be misunderstood, we also decided to eliminate it.

20. Also, using the min and max approach is very subjective as many variables as unemployment and poverty are tempered with by local organism. This kind of indicators are good to compare and prioritize actions withing regions but cannot be used in the way they are presented in this study to increase direct physical losses.

Thanks for this comment. However, it is necessary to differentiate the method, the indicators and data sources in the assessment of SV, the operationalization of the analysis, and the actions. Regarding the method, as we explained before, SV assessment considers variables and indicators with different units. This is the main reason to use normalization to construct the social vulnerability index (SVI). With respect to the indicators, we agree with the reviewer that governments can temper data regarding unemployment and poverty in the most vulnerable countries, this data does not exist, or it is not accessible, but any analysis has an uncertainty level, and we decided to accept it using official numbers. As we explained before based on the literature review, integrating the level of SV to the physical losses will not produce a significant increase in the last one, considering that they are negatively correlated (Cutter and Finch, 2008). Integrating the SV analysis makes the risk assessment more holistic and useful to prioritize actions at the scale of the assessment because it can be national, regional, or municipal.

Minor modifications:

21. Line 118: Typo: This GROWTH counts for 64% of the total population of the region

22. Line 248: Incomplete sentence: "The number of masonry buildings represents the 88.61% of the total buildings in Sabana Centro, whereas those of concrete and wood represent 3.16% and 8.24%, respectively, with."

23. Line 261: Repeated to "In the absence of specific curves locally developed for the Sabana Centro province, fragility curves were selected to to represent these structures"

R/: We appreciate the reviewer pointing to these typos and we fixed them in the updated manuscript.

References

Arroyo, O., Feliciano, D., Carrillo, J., and Hube, M. A.: Seismic performance of mid-rise thin concrete wall buildings lightly reinforced with deformed bars or welded wire mesh, Engineering Structures, 241, https://doi.org/10.1016/j.engstruct.2021.112455, 2021.

Cutter, S. L., & Finch, C.: Temporal and spatial changes in social vulnerability to natural hazards. Proceedings of the National Academy of Sciences, 105(7), 2301-2306. doi:doi:10.1073/pnas.0710375105, 2008.

Field, A.: Discovering statistics using SPSS. Chenai: Sage publications, 2005.

Martins, L. and Silva, V.: Development of a fragility and vulnerability model for global seismic risk analyses, Bulletin of Earthquake Engineering, 19, 6719–6745, https://doi.org/10.1007/s10518-020-00885-1, 2021.

Schmidtlein, M. C., Shafer, J. M., Berry, M., & Cutter, S. L.: Modeled earthquake losses and social vulnerability in Charleston, South Carolina. Applied Geography, 31(1), 269-281. doi:https://doi.org/10.1016/j.apgeog.2010.06.001, 2011.

Villar-Vega, M., Silva, V., Crowley, H., Yepes, C., Tarque, N., Acevedo, A. B., Hube, M. A., Gustavo, C. D., and Santa María, H.: Development of a fragility model for the residential building stock in South America, Earthquake Spectra, 33, 581–604, https://doi.org/10.1193/010716eqs005m, 2017.