Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/nhess-2022-198-RC2, 2022

Reviewer Comment on nhess-2022-198

Jocelyn West (Referee)

Referee comment on "Using machine learning algorithms to identify predictors of social vulnerability in the event of a hazard: Istanbul case study" by Oya Kalaycıoğlu et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-198-RC2, 2022

Overall Comments:

I am very grateful for the opportunity to review this manuscript and learn from this research. This manuscript uses various machine learning (ML) models to classify and analyze households in Istanbul based on social vulnerability, with particular concern for earthquake hazards. The social vulnerability index (SVI) measure used as the outcome variable is based on a novel survey dataset of more than 41,000 households in Istanbul. I believe the analysis of social vulnerability at the household level is a valuable contribution to the fields of vulnerability studies and risk assessment because data availability typically limits vulnerability analysis to a larger geographic unit, such as neighborhoods or municipalities. This study is also among a small-yet-growing number of social vulnerability analyses that incorporate machine learning models, for which I commend the research team. The study uses ML to assess the contributions of individual vulnerability indicators from the SVI to a household's likelihood of being among the top 20% most vulnerable. In doing so, it has the potential to reveal which indicators matter most for the highest levels of vulnerability in this context. Finally, I appreciate the figures and data visualizations provided in the manuscript and associated web application to aid in the understanding and use of these results.

I have several specific suggestions that I would like the authors to address to help improve the manuscript before publication.

Specific Suggestions:

- An overarching question I would like the authors to address is whether, and how, the ML analyses might be used to improve upon the original SVI measure for Istanbul. What do the ML models add to the understanding of social vulnerability in Istanbul that

was not clear previously? I think the manuscript will be strengthened if the authors can better connect these dots for readers.

- I recommend building upon the discussion of the *scale* at which data were analyzed as a strength of this research. Social vulnerability is not often able to be analyzed at the household level, so that is a significant potential contribution of this research worth discussing.
- The current description of the data and methods used to construct the Istanbul SVI are not yet sufficiently complete and accurate to allow their reproduction by fellow scientists. As I understand, there is a vulnerability index for Istanbul that was created as a Phase 1 of this study. However, its description is not available in English (the language of this journal), and there seems to be no peer reviewed publication associated with the SVI. In light of this, the authors need to describe the construction of the index in detail before using it as the outcome variable in the ML models. Thus, I recommend adding a section to the manuscript describing the construction of the social vulnerability index. I encourage the authors to acknowledge the limitations of various approaches to index construction, including those raised by Spielman et al. (2020) and others. I include some recommended publications on this topic at the end.
- Building upon the explanation of the SVI construction, please also explain why the decision was made to evaluate the vulnerability index as a binary variable that refers to the top 20% of the vulnerability index. For instance, why not use the continuous index as the outcome variable? Why not use the top 25%? Finally, discuss any potential strengths, weaknesses, or consequences of defining vulnerability in this way.
- It was also my impression that the variables in the Istanbul vulnerability index are not specific to *earthquake* vulnerability in particular. Instead, they seem to refer to social vulnerability more generally and not in relation to any single hazard. This is not necessarily a problem with the data. However, the claim to understand *earthquake* vulnerability in particular needs to be more fully substantiated, or removed, because the SVI data used in this study do not appear to refer to earthquake-related vulnerability, even if the original household survey did focus on earthquakes.
- I am curious to know how the "risk of job loss" was assessed, and I would be interested to see more explanation of why job loss would be specific to a post-earthquake context, or whether this is a general measure of job insecurity. If possible, please describe briefly how that question was asked on the original survey and whether it was a self-assessment of potential job loss. This will hopefully help readers better understand that variable.
- The last sentence of the abstract suggests that "The machine learning methodology and the findings that we present in this paper can serve as a guidance for decision makers…" I would like to know more about how specifically the machine learning methodology could be used by decision makers. I do understand how the SVI data can be a tool for decision makers, but it is not yet clear to me how the ML component could practically be used by decision makers to reduce vulnerability. Would you argue that this ML analysis can be used to improve the SVI as a decision making tool? If so, explaining how and providing an example of a use case might be helpful.

Grammatical/proofreading:

- In the abstract, please define what the outcome variable is when mentioning it.
- Change "dept" to "debt" in the last line of Table 1
- I recommend avoiding the term "natural disasters" and instead simply using "disasters" or being more specific. (line 63, 103, 453, 529)
- Page 4, Line 121: Avoid using the word "intrinsic" to describe social vulnerability

because social science research specifies that vulnerability is *not* intrinsic to people; it is instead a condition borne by some people under certain conditions. People are also not vulnerable at all times or in all contexts. In other words, social vulnerability does not emerge from the characteristics in the SVI. Rather these variables can sometimes indicate or signal who is more likely to bear vulnerability created by structural forces and power imbalances. It is important that the language used is clear about this.

- Page 25, Line 497: I recommend rephrasing the sentence, "We have found that socially, economically, and environmentally vulnerable communities are more likely to suffer disproportionately from disasters," because this is not a finding of the current study as it does not evaluate impacts of disasters. Instead, this is a finding of many previous studies that you could perhaps cite here. Simply removing the words "we have found that…" may be sufficient.
- I find the phrase "social vulnerability risk" to be a bit confusing and inconsistent with most literature on this topic. Social vulnerability is typically considered a sub-component of risk, so these phrases should not be combined. Instead, use the phrase "social vulnerability" without the word "risk." For similar reasons, the phrase "social risk" should be replaced with either "social vulnerability" or "disaster risk," as appropriate.

Suggested readings on social vulnerability measurement and validation:

- Szczyrba, L., Zhang, Y., Pamukcu, D., Eroglu, D. I., & Weiss, R. (2021). Quantifying the Role of Vulnerability in Hurricane Damage via a Machine Learning Case Study. Natural Hazards Review, 22(3), 04021028. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000460
- Bakkensen, L. A., Fox-Lent, C., Read, L. K., & Linkov, I. (2017). Validating Resilience and Vulnerability Indices in the Context of Natural Disasters. Risk Analysis, 37(5), 982–1004. https://doi.org/10.1111/risa.12677
- Rufat, S., Tate, E., Emrich, C. T., & Antolini, F. (2019). How Valid Are Social Vulnerability Models? Annals of the American Association of Geographers, 109(4), 1131–1153. https://doi.org/10.1080/24694452.2018.1535887
- Spielman, S. E., Tuccillo, J., Folch, D. C., Schweikert, A., Davies, R., Wood, N., & Tate, E. (2020). Evaluating social vulnerability indicators: Criteria and their application to the Social Vulnerability Index. Natural Hazards, 100(1), 417–436. https://doi.org/10.1007/s11069-019-03820-z
- Tate, E. (2012). Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. Natural Hazards, 63(2), 325–347. https://doi.org/10.1007/s11069-012-0152-2