

## Response to comments of Reviewer 2:

We sincerely thank the reviewer for taking time to read our work and give us helpful and detailed comments. Here we would like to address the reviewer's concerns as follows:

1. *"The manuscript deals with the analysis of flood fatalities in Vietnam using a national disaster database (DANA) and tree based methods. Despite the topic could be interesting, I would not recommend this paper for publication in NHESS. The fundamental problem relates to the objective of examining damage-influencing variables on flood fatalities using the DANA dataset, which only includes some damaged quantities for different sectors (e.g. number of collapsed houses, schools, hospitals, etc.), without any further information on hazard and vulnerability parameters. As a consequence, without such information, the results of the study are rather obvious (the number of fatalities is related to number of collapsed houses and roads), so it is hard to me to see the scientific contribution of the work."*

Response: We support the reviewer's assertion that "The fundamental problem relates to the objective of examining damage-influencing variables on flood fatalities using the DANA dataset, which only includes some damaged quantities for different sectors (e.g. number of collapsed houses, schools, hospitals, etc.), without any further information on hazard and vulnerability parameters". Some of the limitations that are highlighted relate to the available data in the national disaster database of Vietnam, DANA. Our paper was limited to the analysis of records of direct damage data, as the reviewer has pointed out.

The DANA data was previously available at <http://www.cfsc.gov.vn>. However, the website changed to <http://phongchongthientai.vn/default.asp> since 2014. A part of the database can be found in this address <http://www.desinventar.net/DesInventar/profiletab.jsp?countrycode=vn&continue=y>. We collected the data from the Central Steering Committee for Natural Disaster Prevention and Control of Vietnam with over 200 data cards from 1989 to 2015. Due to the lack of studies using DANA and indeed the lack of studies on flood fatalities in Vietnam, we responded these gaps and assessed the relationship of flood impact attributes to flood fatalities in Vietnam.

We agree with the reviewer that "As a consequence, without such information, the results of the study are rather obvious (the number of fatalities is related to number of collapsed houses and roads), so it is hard to me to see the scientific contribution of the work". However, we still think that the study can have some small contributions with the objective of analysing the national disaster database of Vietnam (DANA). The results can be a validation for the variable important analysis using tree-based methods in this study. In addition, the analysis using boosting technique can provide a measurement of variable importance (of flood impact attributes on flood fatalities) and partial dependence plots for the variables as in Figure 12 and Figure 13.

### Specific comments

2. *"In the abstract the authors say that "The findings allow us to make recommendations for government policies on improving housing quality for the poor in flood-prone areas in Vietnam", however a discussion on this point is lacking in the paper (except few lines in the conclusions)"*

Response: We agree that it is an important aspect. We had some discussion on the housing factor and housing quality in Discussion section in the last paragraph on page 10 and the first paragraph on page 11 with the illustration of Figure 14 and Figure 15.

3. *"The authors state that "The study contributes a method to analyse the national disaster database, provides a substantial insight in flood-related fatalities in Vietnam and offers a valuable application for other Asian countries" (page 3, lines 12-13). The use of tree based methods is not new and many*

*papers examining disaster damage data exist in the literature. In addition, considering the type of data analyzed in the paper, which are the “substantial insights” provided?”*

Response: We agree with the reviewer that “The use of tree based methods is not new and many papers examining disaster damage data exist in the literature”. We would like to mean that it is the first time this method is used to analyse a national disaster database. Besides Vietnam, many other countries also have their national disaster database, such as 94 national disaster databases are stored and available at <http://www.desinventar.net/DesInventar/index.jsp>. To our best knowledge, there have been no long-term empirical studies on the application of machine learning algorithms in investigating the relative importance of variables on flood fatalities.

4. *“Section 2 “Disaster damage data”. This section should be improved providing more details regarding the DANA database.”*

Response: We agree that it is an important point. We will revise this section thoroughly.

5. *“I would suggest to entirely rewrite sections 4, 5 and 6, avoiding repetition of concepts throughout the paper. In the present form, the paper lacks of a real discussion section, which is now just a repetition of the results presented in a previous section.”*

Response: We thank the reviewer for this suggestion. We will revise these sections according to the reviewer’s comment.

6. *“The conclusion section is vague and it should be improved. The authors state that “This study can contribute to the body of flood hazard knowledge by analysing and reporting on flood fatalities in Vietnam” (page 11, lines 27-28). What is this contribution? In addition, your analysis did not include any information on flood hazard.”*

Response: We acknowledge that the conclusion section needs improving. We would like to mean that this study can have some small contributions in analysing the national disaster database of Vietnam (DANA). Due to the lack of studies using DANA and indeed the lack of studies on flood fatalities in Vietnam, we responded these gaps and assessed the relationship of flood impact attributes to flood fatalities in Vietnam.

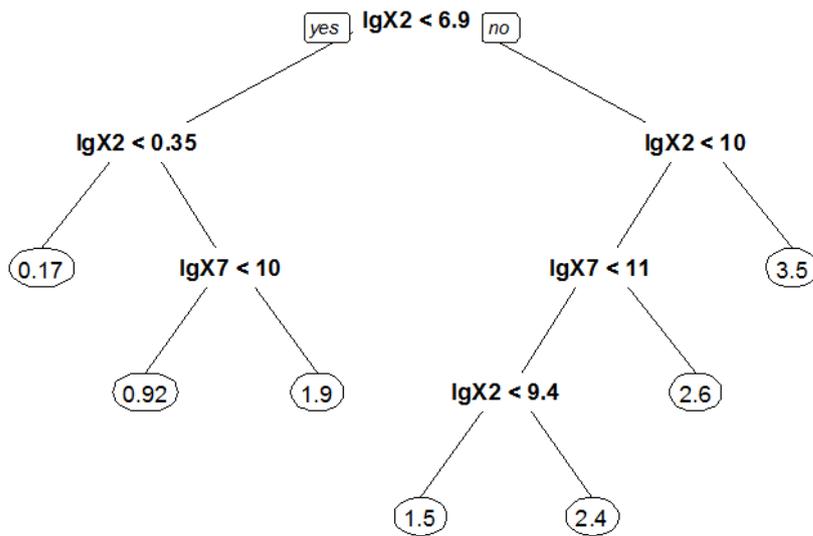
The reviewer is correct to point out that our analysis did not include any information on flood hazard. We only analysed flood damage attributes that are available in the national disaster database.

7. *“The results of the grown trees seem to be not compatible with the results of Figure 2, where I can see other parameters having correlation coefficients comparable to the one observed for lgX7.”*

Response: It is true that Figure 2 showed that the predictors are highly correlated. Bi (2012) mentioned about advantages of Breiman’s random forests to determine the variable importance of correlated predictors; and Strobl et al. (2008) said that conditional variable importance for random forests is suitable for high correlated predictors. We will add the analysis using conditional variable importance for random forest in the paper.

8. *“Caption of Figure 9 should be improved”*

Response: We thank the reviewer's comment. We would like to demonstrate in another tree as follows:



**Minor comments:**

9. "Page 2, line 13: [. . .] "and to a lesser extent IN developing ones""

Response: We thank reviewer's comment. We will fix this error.

10. "Page 6, line 19: [. . .] "was performed to validATE""

Response: We thank reviewer's comment. We will fix this error.

11. "Page 11, line 23: "This paper proposed an approach [. . .]": the approach is not new, you should write instead that "This paper used an approach [. . .]"

Response: We will rewrite as the reviewer's comment.