Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-251-AC3, 2020

Natural Hazards and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Landslide susceptibility assessment based on different machine-learning methods in Zhaoping County of eastern Guangxi" *by* Chunfang Kong et al.

Chunfang Kong et al.

xukai@cug.edu.cn

Received and published: 17 October 2020

Printer-friendly version

Discussion paper

Response to RC2 from Anonymous Referee #2

October 17, 2020

Dear Anonymous Referee #2:

Thank you for your comments concerning our manuscript ID nhess-2020-251 (Landslide susceptibility assessment based on different machine-learning methods in Zhaoping County of eastern Guangxi). Those comments are all valuable and very helpful for revising and improving our paper, as well as of important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet the suggestions. Revised portion are marked in highlight in the paper. The main corrections in the paper and the responds to the reviewer's comments are as flowing:

1 General comments

1. **Two machine learning methods (SVM, RF) combined with particle swarm optimization (PSO) support, total four models were used to evaluate landslide susceptibility in this paper. The results show that the PSO algorithm has a good improvement on SVM and RF models. This paper has a clear**

Printer-friendly version

Discussion paper

research framework and the outcomes I believe it is beneficial to readers. However several issues are not clearly stated in the current article that need more clarifications.

We appreciate for reviewers' approval, and have same opinions with the reviewer. Once again, thank you very much for your approval.

2. **For example, the landslide inventory used in this study is very important since it directly affects the performance of susceptibility models. However the detail information of landslide inventory such as a single rainfall event inventory or a compiled inventory cover different events, material types (rock, earth, soil, mud, debris), landslide patterns (new landslides or old landslides), minimum mapping area, generation methods (satellite or aerial images interpretation, field investigation) etc. are not mentioned in the article.**

   The landslide inventory map in Zhaoping County was prepared from field investigation of Guangxi Geological Survey Bureau (Huang and He, 2018). Please see L120-121. And according to the field investigation report of the geological hazard project by Guangxi Geological Survey Bureau in 2018, there are 345 landslide disaster points in Zhaoping County (Huang and He, 2018). Please see L107-109. These landslide disaster points are used as landslide disaster training and testing samples without considering landslide properties, material types, patterns, minimum mapping area, and generation methods and so on.

3. **The second issue, multi-temporal satellite images have been widely used to extract the LULC and NDVI information under different time backgrounds. However, in this paper the only one date (2017/12/24) Landsat 8 OLI image was used to extract temporal information like the LULC and NDVI seeming insufficient to reflect the temporal variation of land covers.**

   We have same opinions with the reviewer. One date (2017/12/24) Landsat 8 OLI

image, which is close to the field investigation time, was selected to extract the LULC and NDVI index for the Zhaoping County.

When the LULC and NDVI index, as landslide related factors, are put into four ML models with the other eight factors, they only reflects the status of the study area and do not need to reflect the temporal variation. So one date Landsat 8 OLI image on is enough.

4. **The third issue, in order to more completely establish the relation of rainfall scale, natural environmental characteristics, and LULC change with the landslide occurrence potential through the susceptibility model. In my opinion, it's necessary to build the landslide susceptibility model by using the landslide inventories compiled from different scale of rainfall events. For the predisposing factors, except those geology and geomorphological factors, some temporal predisposing factors like rainfall intensity or rainfall accumulation of each event as well as environmental factors like land use, vegetation cover etc. are also suggested to consider.**

   A total of ten factors of high correlation with landslide disaster occurrence were chosen based on the field investigation report of the geological hazard project by Guangxi Geological Survey Bureau and the disaster factors correlation analysis in Zhaoping County: slope, aspect, curvature, annual rainfall, NDVI, stratum lithology, tectonic complexity, LULC, residential density, and road network density County (Huang and He, 2018). These factors reflect the geological environment characteristics for Zhaoping County, including geological characteristics, geomorphological characteristics, meteorological characteristics, ecological characteristics, environmental characteristics, and human activities characteristics, and so on. Please see L125-129.

5. **Finally, I can't find the results or discussions for part of the goals and conclusions (P5 and P29, please see specific comments).**

The questions of P5 and P29 have been discussed in part of specific comments. Please see comments 1 and 8.

6. **Overall speaking, the several main issues mentioned above are suggested to improve before this paper can be considered for publication.**

   We are very grateful for your comments, and those comments are all valuable and very helpful for revising and improving our paper, as well as of important guiding significance to our researches.

2 Specific comments

1. **P5, Ln79-80, "(1) determine the landslide susceptibility assessment factors by multi-source data fusion and correlation factor analysis", what do you mean "data fusion"? And I didn't see "correlation factor analysis" in the article.**

   "Data fusion" here refers to a series of data processing and analysis, and the main processes are as follows: Firstly, the collected multi-source data are preprocessed, including data screening and correlation factor analysis, in order to determine the landslide disaster assessment factors. Then, the selected landslide disaster assessment factors are standardized and classified. Finally, a standardized data set is obtained to run four ML models.

   "Correlation factor analysis" here refers to the analysis process of distinguishing the factors that have high correlation with the occurrence of landslide disaster from all factors provided by field investigation of Guangxi Geological Survey Bureau.

   Due to space limit, it is not carefully explained here.

2. **P7, Ln108, "...there are 345 hidden danger points of landslide...", please explain what "hidden danger points" means.**

   I am sorry. "Hidden danger points of landslide" should be replaced by "landslide disaster points", it has been corrected in the paper, thank you for your careful reviews, and please see the revised L108.

3. **P. 8, Table 1, the classification interval for aspect level 1 is 22.5°, for level 8 is 67.5°, for the remaining six levels the interval is 45°. Why the classification interval for level 1 and level 8 is different? Additionally, the level 1 and level 8 actually reflect similar aspect, however the extreme different grading number (1 and 8) could seriously affect landslide susceptibility. Please give more explanations.**

   I am sorry. Table 1 for aspect level 1 is "[0, 22.5)" should be replaced by "[337.5, 22.5)"; and the level 8 is "[292.5, 360)" should be replaced by "[292.5, 337.5)". It has been corrected in the paper, thank you for your careful reviews, and please see the revised Table 1.

4. **P15-16, P18-19, the description of basic theory for SVM and RF model can be simplified but need some reference citations.**

   The reference citations have been added in the paper, and please see the revised L210, L215-216, and L228 of P15-16, and L262, L267-269 of P18.

5. **P25, Ln 342-344, "...because of the sensitivity of the RF and PSO-RF models to the proportion of landslide samples, it is necessary to carry out sample screening before using RF and PSO-RF models to evaluate the susceptibility of landslide.", this description is unclear. Please explain more about the sensitivity results and how to carry out sample screening?**

   Our study indicated that the accuracy of landslide disaster prediction is low if the RF and PSO-RF models are used directly, but the prediction accuracy will be

greatly improved, if the RF and PSO-RF models are used for landslide samples selection and then the models prediction are carried out. Therefore, it is necessary to carry out sample screening before using RF and PSO-RF models to evaluate the susceptibility of landslide.

6. **P25, Table 5, how to calculate the percentage of landslide points in different susceptibility? The summation of the percentage number in each column should be 100?**

   Table 5 indicates the percentages of landslide points falling into different susceptibility levels, calculated by dividing the grid number of disaster points falling into different susceptibility levels by the total grid number of this grade, and the total number of this percentage is not 100.

7. **P26, Ln 360, please give more explanation for the field investigation results.**

   Field investigation refers to the onsite investigation carried out by the staff of Guangxi Geological Survey Bureau in 2018, with the purposes of obtaining the landslide inventory map of Zhaoping County (Huang and He, 2018). This has been explained in the 2.1 and 2.2. Please see L107-109 and L120-121.

8. **P29, Ln 398-400, "...our study also found that the selection of training samples will affect the susceptibility evaluation results during the process of landslide susceptibility evaluation using four ML methods.", I can't find the discussion in the manuscript?**

   Our study indicated that the accuracy of landslide disaster prediction is low if the RF and PSO-RF models are used directly, but the prediction accuracy will be greatly improved, if the RF and PSO-RF models are used for landslide samples selection and then the models prediction are carried out. Therefore, it is necessary to carry out sample screening before using RF and PSO-RF models to evaluate the susceptibility of landslide. Please see L353-355. At the same time,

The results also demonstrated that PSO-RF model has a better prediction performance than the PSO-SVM model, which is mainly due to the large number of factors selected in this study, the PSO-RF model, a type of ensemble learning, exhibited advantages over a traditional ML method by not only accounting for different types of factors but also evaluating the relative importance of the factors in terms of landslide stability (Zhang et al., 2017). Please see L341-346.

The above discussions show that the selection of training samples will affect the susceptibility evaluation results in the process of landslide susceptibility evaluation using four ML methods.

3 Technical corrections

1. **P2, Ln 25, the method "RF" was misspelled as "FR".**

   The "FR" has been replaced by "RF" and thanks for your careful reviews. Please see revised L26.

2. **P3, Ln 51, the author name of reference (Tien Bui et al., 2012) was repeated.**

   The author name of reference (Tien Bui et al., 2012) has been revised, and thanks for your careful reviews. Please see revised L51.

All editorial errors have been corrected accordingly as suggested by the reviewer throughout the manuscript, and we would appreciate the reviewer for those comments.

In summary, we are very grateful for your comments, and those comments are all valuable and very helpful for revising and improving our paper, as well as of important guiding significance to our researches.