Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-270-AC3, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on “Landslide risk zoning in Ruijin, Jiangxi, China” *by* Xiaoting Zhou et al.

Xiaoting Zhou et al.

wuwch@ecut.edu.cn

Received and published: 14 December 2020

We sincerely thank you for the overall feedback and the constructive comments on the manuscript. Below are our responses to what you commented.

The manuscript under review presents a well-structured and clearly readable application of a machine learning (Random Forest)-approach to spatially predict landslide occurence. The illustrations are instructive and well-elaborated.

Reply: Thank you so much for your positive comments.

However, title and scope of the paper are completely misleading. The study just resembles a classification of terrain units (30 m × 30 m pixels) for the probability of landslide occurrence based on several geo-environmental factors and does not consider the temporal probability of such events to occur in the context of a hazard assessment,

possibly serving as a basis for landslide risk zonations.

Reply: Thanks for having raised this issue. This study was aimed at using the Random Forest algorithm to analyze the probability of landslide occurrence and map landslide risk zone of the study area based on a comprehensive consideration of the influences of hazard factors by field investigation and remote sensing technology. Hence, after a careful consideration we decided to use the title of the paper “Landslide risk zoning in Ruijin, Jiangxi, China”. As for the temporal probability of such events to occur in the context of a hazard assessment you mentioned, we think it is well worth further investigations. The difficulty encountered was to know the exact occurrence dates and time of the historical landslides in the study area. Even during the field investigation, local people could just tell you “this landslide event took place on day of June or July... ” without further concrete information, especially, for those occurred more than 4-5 years ago. Thus, it appears impossible for the time being to analyze the temporal probability of landslide occurrence with high accuracy. Yet, your point will be taken into account in future work when we have more investment to purchase equipment to monitor such events or employ local people to record such information.

The presented analyses have nothing to do with any kind of a risk analysis since no (spatial) vulnerability assessments of potential objects at risk are presented or incorporated in any kind of (spatio-temporal) risk analysis. In such, the paper only resembles the application of a common machine learning approach for landslide susceptibility classification.

Reply: In the study area, the damage caused by landslides mainly happened to the road systems and built up areas around and on slopes (e.g., residential buildings, school and temple), which have been carefully taken into account. The road buffering procedure and propensity weight assignment, were actually something related to the spatial risk analysis. The polygons of residential buildings extracted from Google Earth were superimposed on the landslide risk zoning map to assess the landslide risk of the buildings within 60m and 120 m of spatial extent. It is true that this part was not well discussed in the original version, but it is added in the revision. Thank you!

Besides this, I am not sure if prediction of landslide susceptibility using any kind of inventory-based analysis is really admissible for such a large territory with only 155 landslides.

Reply: The approach used in our paper is the Random Forest (RF) algorithm which is able to handle the high- and hyper-dimensional datasets with reliable prediction accuracy but requires few training samples (Breiman 2001). This is the big advantage of RF over other algorithms. The accuracy of the RF model versus the verification set (VS), as well as the success rate and prediction rate curve showed good risk prediction results, where Kappa Coefficient (KC) is 0.8299 or 82.99% and the overall accuracy (OA) 91.49%. According to Cohen (1960) and Landis and Koch (1977), this prediction reaches “almost perfect” level. Therefore, despite of the limited amount of landslide points (e.g., 155) for training and modeling, we believe that the risk assessment is reliable. It may hence provide useful reference for risk management prevention in the study area, and the approach be extendable to similar area for landslide prediction and risk assessment. Please check our revision. Thank you!

The landslides are not described at all regarding their typology or triggering mechanisms and their spatial relation to the geo-environmental factors used for susceptibility modelling.

Reply: Thanks for your comments. We are sorry for having overlooked this in the first version. It has been added in the revision, respectively in “2.2.2 Field survey data” a separate subsection “2.3 Distribution of landslide density in each geo-environmental factor” and a new figure 7.

The sampling of negatives for modelling is questionable since it is trivial that on shallower slopes landslide susceptibility is low. With such a small landslide data set, negative sampling should be conducted with much greater care on steeper non-landslide slopes to investigate the ability of the method to correctly predict the landslides.

Reply: Thanks for having raised this issue. Actually, no-risk (or negative as you mentioned) sampling was not conducted in the same way as you had thought. We did this in flat (not shallow) areas, e.g., urban, waters and croplands where slopes are lower than 1-5°, actually, 1-3° supposing that landslide is very unlikely to take places there. For risk modeling, or rather, probability analysis, we have to take samples from two extreme ends, that is, landslide hazard occurred areas (where the probability is 1.0, meaning that the hazards have truly taken places) and no-risk stable areas (whose probability is 0.0). To take no-risk samples on steeper slopes would be a risky issue itself as we were not sure whether such samples were really no-risk ones. Theoretically, any risk modeling shall not involve such uncertainty but just be based on what is sure. Thank you!

To conclude, the paper adds nothing scientifiically new to what is already known from the literature and just represents a case study application that would need much more work to be publishable.

Thank you for your general comments, which have revealed the overlooked points in the original version of the paper, and provided us an opportunity to improve it. Some points that need to be clarified are listed here. From a large view, you are right, the paper involved a known machine learning approach, i.e., RF algorithm, and known geo-environmental factors, and appears to have nothing new. But we developed a complete digitization and weight assignment scheme so that no-digital data such as geological map can be digitized with risk propensity indication for risk modeling. Among the tens of documented publications on landslide modeling and prediction, it is rarely seen such a complete and innovative procedure. Secondly, it was after a comparison that we decided to employ RF algorithm as it can process huge volume of hyper-dimensional data for accurate classification and prediction but requires only few samples. This advantage is superior to other machine learning approaches. Thirdly, besides the technical aspect, scientific paper shall provide practical value to our society. In our case, the results can serve as useful reference for local government to implement disaster reduction and prevention measures. We believe that our revision has been improved and can meet what you had expected. Thanks.

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2020-270/nhess-2020-270-AC3-supplement.pdf

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-270, 2020.