

I have been nominated to act as reviewer of the manuscript “Influencing factors and development patterns of cracking–sliding failure of loess in China” by Jiarui Mao, Xiqiong Xiang, and Yanrong Li. The manuscript aims to characterize the landslides occurring on loess in a quite vast area of China, with the ambitious target of (i) preventing landslides occurrence, (ii) reduce losses due to this type of failures. Unfortunately, the paper appears to be not scientifically sound, rather simple in its explanation of phenomena, and data and analyses appear inadequate to fulfill the proposed targets mentioned above. Therefore, the paper is not ready for publication in *Natural Hazards and Earth System Science*. In the following, I provide some more specific details:

1. Research is poorly described, as well as the available datasets. No description on how data were collected has been provided, nor the data types (landslides are points or polygons, vector or raster?). As it is, the paper does not contain either novel data or new ideas/insights based on data collected by previous researchers. The only statistical analyses applied to data presented consist in a percentage comparison, which is rather simplistic compared to the many possible approaches applicable to analyse the influence of a set of variables on landslides occurrence, which is evident from the specific literature worldwide.
2. It is not clear to me where do the landslide data come from, what is their spatial, temporal and thematic accuracy, and, above all, their degree of homogeneity within all these accuracies. It is not clear what type of inventory is used (event, archive...) and which sources were used to get data. Even though the information comes from scientific literature, the authors appear not to fully handle the possible inhomogeneity of their data, and, therefore, the quality of the dataset.
3. The study area is 6.4×10^5 km², twice the surface of Italy, for example. In a landslide prone area such as the loess plateau, I am expecting a much larger dataset than a few hundreds (I take this number from a rough estimation of red dots in Figure 2). In such large areas, it is just not correct to simply draw some considerations from a small subsample of the landslides data available and extending them to areas where climatic, morphologic, anthropic, geologic conditions are (sometimes sensibly) different. In other words, to what extent the authors can extend to more than half a million square kilometre area the monthly ground temperature variations (Figure 8), or the daily temperature variations? For instance, daily temperature variations are considered for a small sample of 32 landslides. It is not stated where the sample is located, or if that location may be thought representative of all the Loess plateau. The same applies for the rainfall induced events. Comparing plots of figure 7, it is not clear why the Shaanxi has so many more failures of the other places. Is it just a larger area or it is an expression, for example, of the climatic variability inside the Loess plateau? I would have liked that the authors had pointed that out, and properly commented. If any, what areas can be represented by the Shaanxi? Why is it so different from the others in terms of rainfall thresholds? In the slope classification (Section 2), again data are taken by a local study of Shaanxi and then applied to the whole Loess plateau. To what extent is it representative of an area that crosses two climatic zones?

Furthermore, also the temporal factor is important. For instance, temperature data come from observations that were obtained from a monitoring activity that last one year (from Nov 2004 to Oct 2005). Authors do not mention or deal with possible problems of temperature trends. Was it an average

year, or it was an exceptionally cold/hot one? If available, a plot of the average temperature of the last (or even following) 10 years would have helped.

4. Geography of area is not characterized. Elevations, slopes, aspects, hypsometry, should be statistically described. A little geomorphology and geology is described, but I expect something more clear and structured. There is no section on study area, with subsections dealing with geography, geomorphology, climate, and geology. In such broad studies, a reader expects these sections to be a frame to the analyses.
5. Language is often too generic. For example, words like “large dataset”, “poor stability”, “fair stability”, “good stability” shouldn’t be used in academic English of scientific papers. I also disagree with the term “cracking-sliding” as a type of failure. None of the existing landslides classifications encompass that term. Authors should find terms compliant with published nomenclature. If they are proposing a new nomenclature term, they should say that explicitly, and justify it in detail, it should be a totally different paper.
6. The manuscript does not follow a structure accepted for a scientific paper. (i.e. Introduction, study area and available data, methods, results, discussion, conclusion). In particular, I disagree with the idea of not writing a Discussion section. Instead, the authors have chosen to add very simple considerations while presenting some (unclear) data, which is scientifically questionable. Furthermore, the factors hypothesized to be influent for the development of “cracking-sliding” failures are presented singularly, therefore the interaction of these factors, and their specific role or possible chains of processes inducing landsliding remains hidden and unclear.
7. References are only from China, whereas loess research is not only produced in China, but also in all the other countries where loess deposit is present (as the authors point out).

I recommend rejection. If the authors wanted to re-submit the paper, they should address all the above mentioned issues, provide new insights based on the data/evidences available, rewrite the paper describing a clearly reproducible research, based also on international scientific literature. Furthermore, I would suggest that the authors focus on much smaller test areas, where they have good quality datasets on different variables, and where they can possibly carry out a much more detailed landslide inventory, also using satellite or aerial images. This approach would enlarge the time window of the observations. The test areas could even sum up to just a few hundreds of square kilometres, but should be accurately chosen to be representative of larger morpho-litho-climatic sectors. The advantage would be to handle more easily the data and draw well data-supported hypotheses.

Best regards.