Dear NHESS Editors,

First of all, we would like to express our sincere appreciation for your very detailed and constructive comments and suggestions.

Next, in a sequence, we would like to respond to your comments in a point to point manner so that hopefully all the questions can be answered or clarified. All the answers and responses are in red.

GENERAL COMMENTS

The manuscript deals with the effect of rainfall and its role on landslide deformation and failure. The authors carried out a feature extraction method for a rainfall data set and it was categorized by a cluster analysis. Rainfall indexes were computed for rainfall characteristics such as quantity, duration, and the number of raining days in a given period of time. The results were later applied and validated to a data set acquired at a cleavage-parallel landslide in the Three-Gorges Reservoir area. The landslide displacement prediction using neural networks for the rainfall input in the form of raw data, monthly rainfall, and feature extracted rainfall were benchmarked. The authors concluded that using the feature extracted rainfall method is best at predicting landslide displacement compared to the other methods and at the same time the computational stress has been reduced significantly.

Although the topic is very interesting from a scientific and practical point of view, the manuscript presents some limitations, conceptual mistakes, technical errors and is sometimes confusing to read. Consequently, it is not suitable for in the present form. The paper must undergo major revisions for publication in NHESS.

The authors are strongly encouraged to review the paper in accordance to the high international standards of the NHESS Journal.

General comments: - The authors should re-organize the paper to have a coherent scheme regarding the presentation of the work carried out. At the moment the manuscript contains a lot of relevant information but it is scattered and spread around the paper in a disorganized manner. The authors are encouraged to consolidate this information inside the relevant sections of the manuscript and to avoid unnecessary repetitions.

- The manuscript lacks the relevant references in the topic and in addition only 13 references are cited inside the document from the 44 listed in the reference list. It is highly recommended that the authors should carefully look into this.

- Regarding also references, the paper is lacking of an analysis of the important results and issues raised by other studies, in particular in context of the submitted paper. Discussion of the results obtained in the submitted manuscript should be made by comparing qualitatively and if possible quantitatively with the results obtained in referenced studies.

- Basic descriptions and concepts are not defined inside the manuscript, such as: cleavage-parallel landslide and BP neural networks (to name a few).

- It is recommended that the authors revise the manuscript all over again and find the suitable words, phrases, technical terms and definitions in proper English. It becomes even more critical when the authors pretend to describe the methodology. Detailed comments:

A: We sincerely appreciate your constructive, detailed comments to the manuscript.

1) We have thoroughly revised the manuscript to make it more precise, consistent, and clear.
2) We have thoroughly checked all the concepts are clearly defined.

3) All the references to make sure all the cited papers are listed, and all the listed have been cited. The abstract and conclusion are re-written.

We have the one-to-one responses to the comments as follows.

1) The abstract should be improved. Some statements like the one in Line 14 and Line 16 are misleading and should be rephrased.

A: The abstract and conclusion are re-written as:

"In this paper, a feature extraction method using a cluster analysis (CA) is employed for the analysis of rainfall data. With this approach we effectively revealed the most significant features contained in a rainfall sequence and greatly reduced the burden for processing large amount of rainfall data. Meanwhile, it greatly improves the spectrum of usefulness of rainfall data."

2) The introduction is lacking relevant references. This introductory part should be rearranged in a way that the references are supporting the stated comments (i.e. Line 31).

3) The second paragraph of the Introduction should be fully rephrased and a better summary of the past studies should be carried out by the authors.

4) In Line 33, the authors should explain in detail how this is difficult (with supporting references) and how their method is an improvement for this.

A: The Introduction has been substantially re-written, and all the comments from both Referees are incorporated, and more references are added.

5) In Line 52, Kurtz et al. 2014 is incorrectly referenced in the manuscript. This work is not relevant to this paper and is based on other approach for feature extraction. The authors should use another reference or delete this one.

A: Thanks for pointing out this mistake. Kurtz et al 2014 has been deleted from the text and the reference list.

6) In Line 55, the authors should use the correct references to support the sentence.

A: The following references are added and citied here to support the claim:

Wang MJ, Shen, JH. Rainfall Landslide and Debris Flow Intergrowth Relationship in Jiangjia Ravine [J]. Journal of Mountain Science, 2011, 8(4):603-610.

Finlay PJ. The relationship between the probability of landslide occurrence and rainfall [J]. Canadian Geotechnical Journal, 1997, 34(6):811-824.

Gariano SL, Brunetti MT, Iovine G, Melillo M, Peruccacci O, Terranova C, Vennari F, Guzzetti F. Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy[J]. Geomorphology, 2015, 228(1):653-665.

7) Line 70 should describe briefly what a BP neural network is.

A: The short introduction of BP neural network method is added as the second paragraph of the subsection 3.4:

"The back propagation (BP) network is a kind of multilayer feedforward neural network. It is a widely tested and validated error back propagation algorithm. The network consists of an input layer, a number of hidden (middle) layers and an output layer. Based on Kolmogorov's theorem, a three layer BP

neural network can achieve approximation for any arbitrary nonlinear functions, so that we choose BP neural network to carry out this quantitative examination."

8) In Line 74, a description of this type of landslide should be included.

A: Cleavage-parallel landslide is defined at the start of Methodology.

9) The Methodology section should be fixed and improved. A better description should be included in order to make the reader understand better the approach used.

A: The Methodology part has been revises. This is similar to the comments from Referee #1.

10) Line 85 should include references.

A: The reference of Wu is added and cited:

Wu H. Monitoring and theoretical analysis of rainfall infiltration of Huangtupo landslide in the three gorges reservoir. China University of Geosciences for the Master Degree of Engineering, 2014.

11) Line 89 to Line 92, the authors should explain in detail the reason to use an average daily evaporation? Is it because of lack of data or is it a common practice?

A: This is the same comments from Referee #1, and we have modified the text to explain our approach.

12) Line 94 to Line 96, the authors should rephrase the statement and add the relevant reference.

A: This opening sentence has been modified, and the relevant reference are added:

Bui, D., Pradhan, B., Lofman, O., Revhaug, I. and Dick, Ø. B.: Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam. Natural hazards, 66(2), 707-730, 2013.

Du J, Yin K and Lacasse S.: Displacement prediction in Colluvial landslides, three Gorges reservoir, China [J]. Landslides, 10(2): 203-218, 2013.

13) Line 109 to 111 should explain, why the authors use this approach and why.

A: The selection of daily average, duration and the contiguous days of rainfall is based on searching through a large quantity of previous studies. And the analysis results our supported support our selection by displaying significant correlation between the selected rainfall parameters and the deformation behavior.

14) Line 121 to 124 needs to be fixed and rephrased. It is not understandable, what the authors mean with high cohesion and low coupling in this context.

A: This has been modified. Thanks for the comments.

15) In line 128, the authors mention, that the K-means algorithm is the most used clustering algorithm. In what context and explain the purposes.

A: The applicability and maturity of the K-means method is emphasized. The whole paragraph has been re-written.

16) Line 128 to 135 needs relevant references

A: The citations are added:

Steinley D: K-means clustering: A half-century synthesis [J]. British Journal of Mathematical and Statistical Psychology, 59(1): 1-34, 2006.

Hartigan J. A., and Wong M. A.: A K-means clustering algorithm, Applied Statistics, 28, 100-108, doi: 10.2307/2346830, 2013.

17) Line 136. Also the sample selection affects the final results. The authors should elaborate in this respect also.

A: In our study the sampling selection is irrelevant since we have used all the 211 samples of the rainfall events.

18) Line 148: What does it means that it is shaped like stairs to the Yangtze River?

A: Terrace is the right geological term to be used. Now this part is changed to:

"It is a single, north-facing, inclined cleavage-parallel slope on the Yangtze River terrace."

19) Line 153 to 158: the authors should explain in detail, why are they using that control point and why is that significant. They also should include the figure of the profile and where the point is located, inside figure 1.

A: This is the same comments from Referee #1, has been answers and revised in the text.

20) Line 158: the authors should reflect on adding a new point or several others as the landslide movement is not uniform and it will add value to the validation. This will definitely have a big influence on the final results.

A: The primary purpose of this paper is to examine the effectiveness of the proposed clustering analysis approach. We will continue the work to include more cases by using data in more observation points. Thanks for your valuable suggestion.

21) Line 160: the authors should include the Zigui County in Hubei Province inside a map to understand the actual location and to understand if this will affect the final results as rainfall is spatially correlated.

A: The map has been modified with an inset to indicate the location of Zigui County.

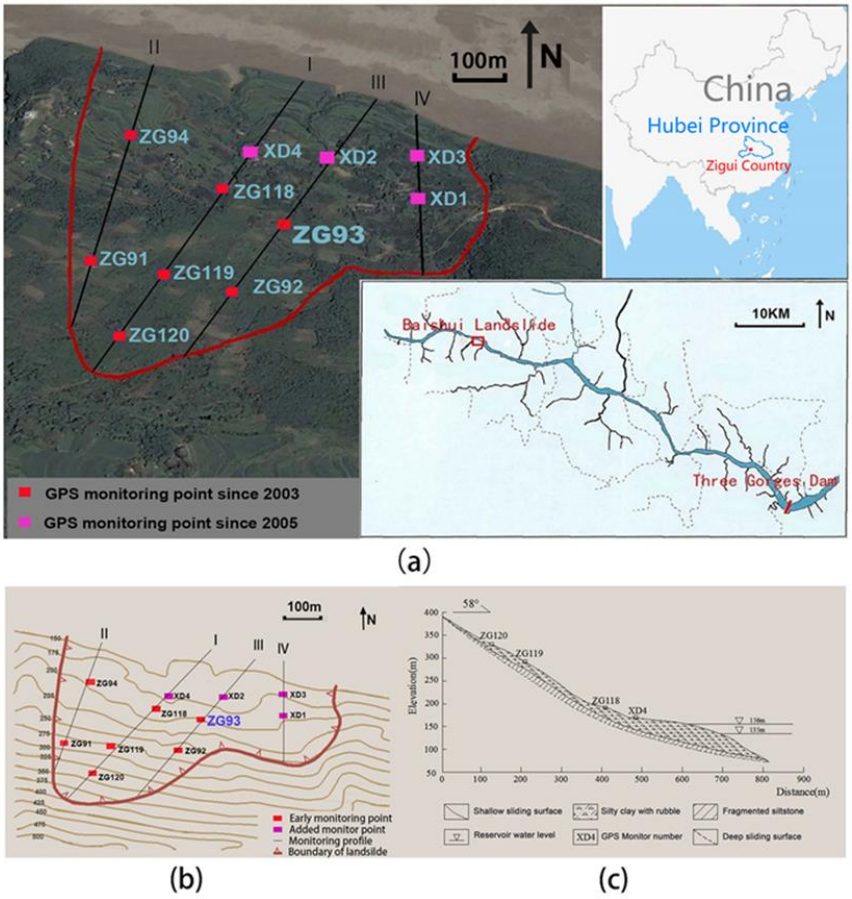

Figure 2. (a) The location of the Baishuihe Landslide (the west most red open square) in the Three Gorges Reservoir area; (b) The locations of the GPS benchmarks (the red and magenta solid squares) for displacement monitoring in the Baishuihe Landslide; (c) The vertical geological cross-section of the Baishuihe Landslide along Profile I.

22) Line 171, in figure 1, the letters should go first, then the legend.

A: This has been revised in the figure 2.

23) Line 175. Using average interpolation for a long period can be misleading as the warmer days without rain will tend to influence the sample. Can the authors elaborate why not only using the evaporation related to the rainy days?

A: This is the same comments from Referee #1, has been answers and revised in the text.

24) Figure 1: the red square in the smaller map is barely visible.

A: This is the same comments from Referee #1, has been answers and revised in the text and the Fig. 2.

25)

Line 196: the authors should explain how this coefficient is derived.

Line 203: the authors should explain how this optimal point was achieved.

Line 190 to 203: It is unclear inside the manuscript, why the aforementioned coefficients were used and its relevance to the results. Can similar results be achieved using other techniques of normalization?

A: This is answered as the reply to Referee #1's Comments # 7).

26) Figure 3: Check on the r-axis. Point C3 is out of the range.

A: The figure has been modified to correct this imperfection.

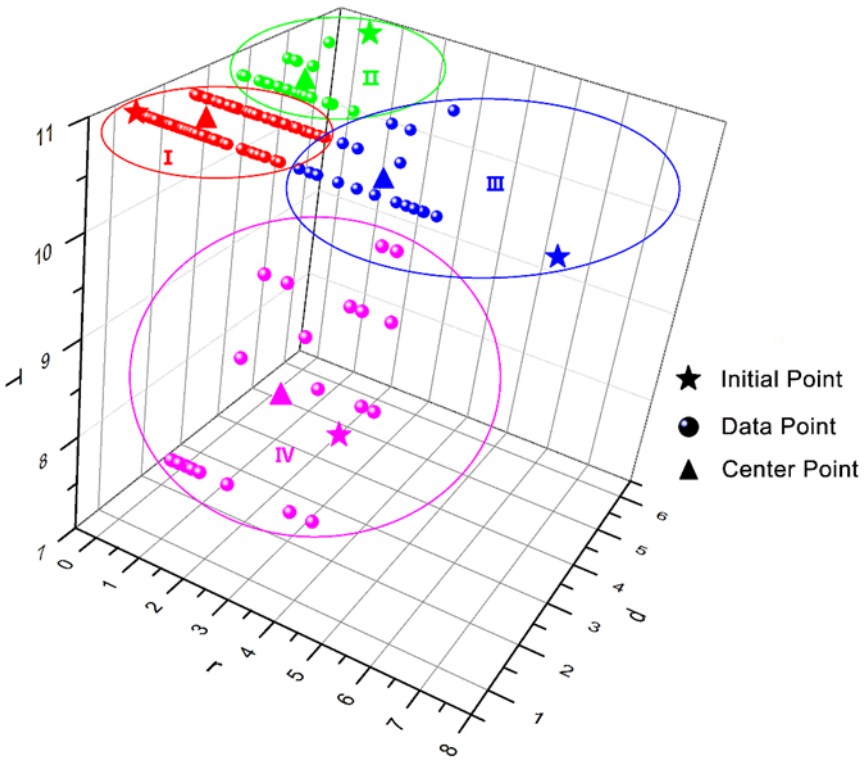

Figure 7. The classified rainfall types based on cluster analysis: I: sporadic rainfall; II: long-duration rainfall; III: short-duration storms; and IV: long-duration intermittent rainfall.

27) Line 211: explain briefly the principle of maximum dissimilarity.

A: This is best illustrated by using the Eq. 2 shown in

$$D = \sum_{i=1}^{k} \sum_{j=1}^{k} (x_i - x_j)^2 \quad (2)$$

In Eq. 2, D is the total distance (dissimilarity) the k data points share. The larger the D is, the higher probability that these k data points are in the different classes.

28) Line 219: What is the stability criterion?

A: In the revised version we have clearly defined the criterion for stopping further cluster classification.

29) Line 228: The authors should explain how they were coming to this conclusion.

This part has been modified as:

"Rainfall volume is the most important factor in causing the variations of displacement in the

cleavage-parallel landslide (Gariano et al., 2015). Therefore, after categorization, we use rainfall volume as the feature for extraction, taking the total rainfall volume in the same category as the feature of that particular category."

30) Line 248 to 249: Use relevant reference.

A: The number of nodes for the hidden layers in BP neural networks is determined empirically. To make it clear, the description of this part is revised as:

"The number of nodes for the hidden layer n1 in BP neural networks is determined empirically with $n_1 = \sqrt{n+m} + a$

where n is the number of nodes for the input layer; m is the number of nodes for the output layer; and a is a constant, which is set to be 2 in this work. The results of n_1 are 9, 4, and 5 in the BP neural networks for the three rainfall categories. The information of the number of nodes in different layers, along with the errors in prediction, are shown in Table 2."

31) Figure 5 and 6: Both daily forecasts and cluster analysis seem to give reasonable results. Is there a better way to show the improvement of the cluster analysis (a success rate approach)?

A: The selection of which clustering method to be used is not the theme of this paper. However, the K-means method is a matured clustering method. This is the reason we have chosen it as the method to be used in this paper.

32) Figure 5 and 6: One of the main advantages of the cluster analysis is based on the improvement of the computational effort. Can the authors further elaborate; is this improvement really worth it (as there will be also computation involved on the K-means algorithm)?

A: To reduce the computational burden is only one of the advantages of using clustering analysis. Besides the factor of reducing computational intensity, it also improves the accuracy of the prediction. It provides another way to preprocess the rainfall data.

33) The authors should discuss in the conclusion section, if this approach can be improved by selecting other type of clusters. Is there a possibility to perform a sensitivity analysis based on this?

A: The selection of which clustering method to be used is not the theme of this paper. However, the K-means method is a matured clustering method. This is the reason we have chosen it as the method to be used in this paper. This is the similar answer for the question of the above regarding Figs 9 and 10.

It is hard to conduct any sensitivity analysis, since we only have the monthly deformation data. It is hard to discuss which clustering approach of the rainfall is more sensitive against the relatively coarse in temporal monthly deformation.

34) The conclusions are mostly a repetition of the text inside the manuscript. The conclusion section should be fixed and rephrased.

A: This is the same comments from Referee #1, has been answers and revised in the text.

Sincerely thanks for your detailed and constructive comments and suggestions.