Dear NHESS Editors and Referees,

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 Authors present and employ cluster analysis (CA)-based feature analysis to rainfall data for rainfall feature extraction. This method extracts the most significant features of a rainfall sequence and greatly reduced rainfall data quantities. This approach is applied and validated to a data set acquired at a cleavage-parallel landslide in the Three-Gorges Reservoir area. The topic address scientific questions within the scope of NHESS.

A: Thanks for your encouraging words.

The theoretical background is well-argued. Review of literature seems complete. The description of study area is sufficiently complete. The description of methodology and successive parts of paper are not well organized. Results and discussion sections are very short compared to amount of work done. They should be widely increased. The readability of the whole paper is sufficient with a quite good English, which however can be improved. Overall, the work presents some carelessness and incompleteness. It can be published on NHESS journal only after a major revision.

A: We have closely followed your detailed comments, by:

1) The description of methodology are re-organized;
2) Results and discussion sections are expanded to include more discussions.
3) We have carefully checked the grammar and spellings, include the numerical values; we tried to eliminate the incompleteness to our best knowledge.

SPECIFIC COMMENTS

I have some specific comments that should be addressed before the manuscript can be accepted for publication.

1) Section 2.1, "The relationship between rainfall and evaporation" (Page 3, Lines 82-92).

Authors should better explain if they are dealing with real or potential evapotranspiration. Moreover, some details on the calculation of evapotranspiration should be needed. Furthermore, how they passed from monthly to daily ET?

A: The evapotranspiration is real observation data; but they are in monthly value. The daily ET value has been gotten simply by using the monthly value divided by number of days in that month. Due to the lack of other supplementary observations such as the temperature, moisture, wind speed, albedo etc., this is the best way we can estimate the daily ET. We have added the sentence in Page 3 Lines 95 – 97 to point out how the daily ET is calculated, as

"Nevertheless, we would like to point out that the daily evaporation value is calculated by simple division of the monthly value with the number of days in that month. This is the most practical way we can do, due to lack of more detailed supplementary meteorological observations in this area."

2) Section 2.2, “Statistics of rainfall by times” (Page 3, Lines 93-101).

Authors stated that “most studies carry out statistics analysis of rainfall based on precipitation per month or per day”. But many papers are present in the literature in which statistical analysis are carried out using daily rainfall data. Authors should considerer these works.

A: Thanks for Your constructive suggestion to make the studies more comprehensively reviewed. We have added the followings references (also added in the reference list) on statistical analysis using daily data, in Section 2.2, after the first sentence:

“For example, Crozier and Eyles (1980) used daily rainfall and established thresholds to compare terrain sensitivity and to assess the occurrence probability of landslide. Using daily rainfall data from Kuala Kenderong and Kg. Jeli along the Gerik-Jeli Highway, Lateh et al. (2013) analysed the correlation of landslide events and rainfall precipitation. The rainfall induced landslides was investigated by applying the cumulative rainfall method which comprises the reconstruction of absolute antecedent rainfall for 20 landslide events.”

3) Section 2.3, “The features of the rainfall data: rainfall volume, rainfall duration and rainfall time” (Page 4, Lines 102-126).

For a better clarity and understanding of the text, Authors should specify in detail how the values of the parameters used to evaluate the three indexes were chosen.

A: The values we chose are based on numerical sensitivity test for getting high coherence and low coupling of the rainfall-slide events. We have clarify it in Section 2.3.

4) Section 2.4, “Clustering Analysis using the K-means clustering algorithm” (Page 4, Lines 127-135, Page 5, Lines 136-143).

The main methodology of the paper is represented by the application of the K-means clustering algorithm. All the used variables are shortly introduced and this leads to some misunderstandings. A figure with a flow chart would be very useful to understand the variables and the all the process.

A: Section 2.4 has been re-written, a new figure to show the flow chart of the data processing procedures (also show below) is added.

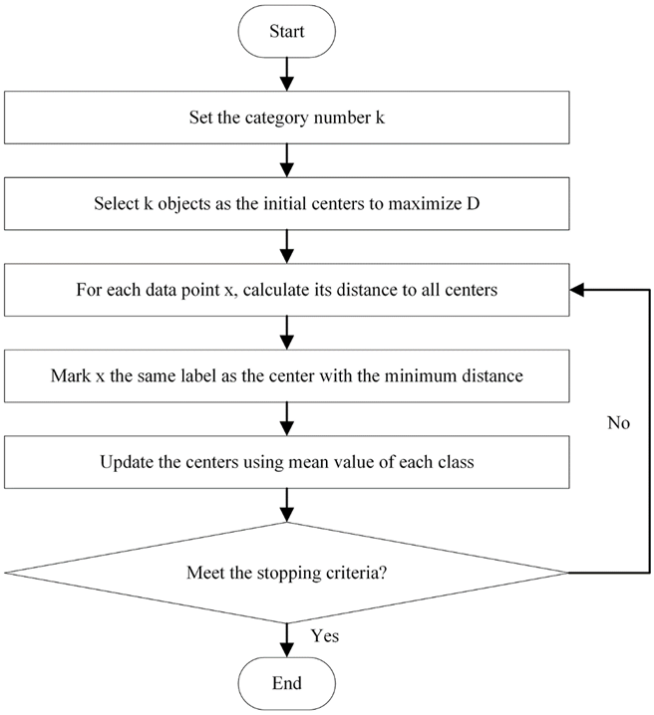

Figure 1: The flow chart of the K-means algorithm.

5) Section 3.1, "Geological background and data collection" (Page 5, Lines 145-158).

Authors should better explain why they have chosen the ZG93 point. Is it representative for all the landslide body?

A: The rationale, along with the new Figure, is added as:

"The selection of ZG93 is based on: 1) It is roughly located at the center of the Baishuihe Landslide so that it is the most unlikely point to be contaminated by false alarm or local signals generated by boundary effect in those monitoring points close to landslide flanks; 2) Observational facts, as shown as the red curve and triangles in Figure 3 below, support our selection for the fact that it is sensitive enough to catch the subtle displacement in the early stage of the monitoring period (prior to the end of 2007) on one hand; and behaved as the average of all the point after rapid change occurred in May 2007 on the other hand."

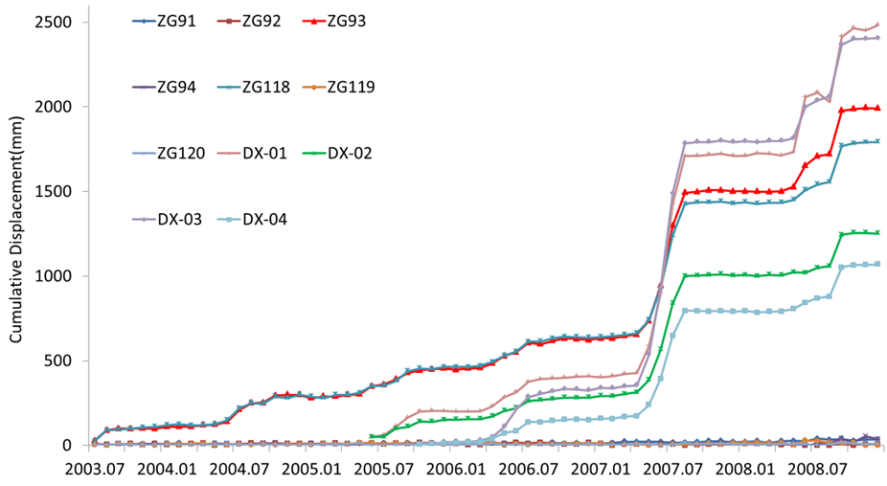

Figure 3: The cumulative displacement of monitoring points in the Baishuihe Landslide.

6) Section 3.2, “Feature analysis of rainfall data” (Page 5, Lines 159-169, Page 6, Lines 170-182, Page 7, Lines 184-187).

A column chart with the average monthly rainfall would be needed. Moreover, also an ECDF graph for duration and cumulated rainfall would be useful for analyzing differences.

A: The suggested column chart with the average monthly rainfall is added as Fig. 5. and the ECDF graphs for duration and cumulative rainfall are added as Fig. 6. Thanks for Your constructive comments.

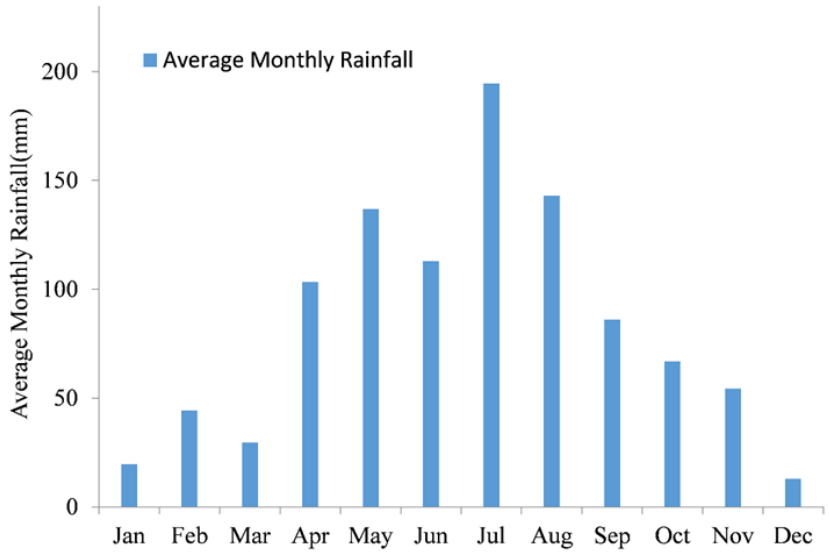

Figure 5. The average monthly rainfall column chart for the period of 2003-2008.

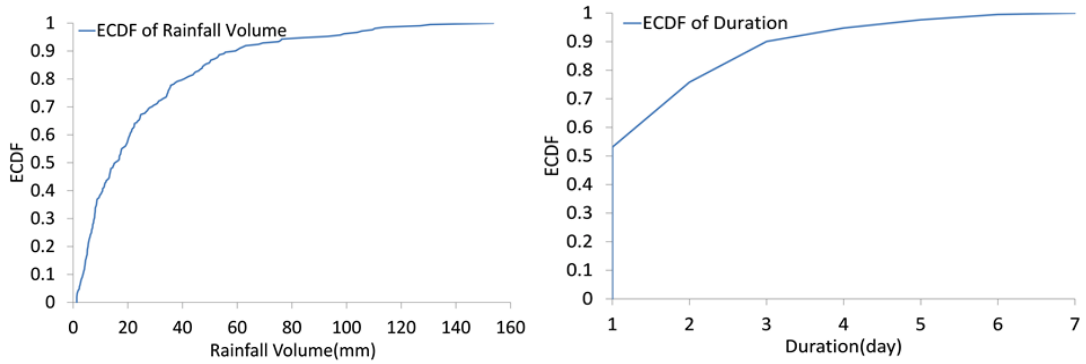

Figure 6. The ECDF plots of the cumulative rainfall and the duration for rainfall events.

7) Section 3.3, “Feature extraction of Rainfall data and Categorization results” (Page 7, Lines 189-203, Page 8, Lines 205-218, Page 9, Lines 219-237).

This paragraph is very confusing. The definition of the three indices are unclear. How Authors obtained the values for p1 and p2? What “scaling coefficient” means?

A: This section has been re-written. The 3 indices are definite first (r, d, and T), before they are discussed. The scaling coefficients are simply for best separation of the classes in cluster analysis, so that they could be any value. They are simply for better visual effect when plot out.

8) Section 3.4, “Prediction of landslide displacement with BP neural network” (Page 9, Lines 238-250, Page 10, Lines 253-265, Page 11, Lines 266-270, Page 12, Lines 271-282).

Several variables are introduced but no longer used in the following.

A: We have taken the advice. These variables are further illustrated and used in Table 2.

A sensitivity analysis, considering several validation periods (in addition to the one used in the work: 2006-2008) would be needed in order to evaluate the performance of the analysis.

A: Actually, the selection of train data and test data is not strict. However, a good performance of BP neural network requires enough train data. In this paper, less than half of data is selected as train data and good results are obtained.

9) Section 4, "Result Discussion" (Page 12, Lines 283-295)

This section is very short. Authors should better argue and comment the obtained results.

A: The conclusion is substantially re-written and expanded.

10) Section 5 "Conclusion" (Page 12, Lines 296-303, Page 13, Lines 304-305)

Poor conclusions. Authors should better explain the main findings and implications of their work.

A: The conclusion part has been re-written and expanded to re-cap the main findings and point out the direction of future study along this line of thinking.

TECHNICAL CORRECTIONS

1. Page 1, Line 30: Please rewrite better the following sentence "At present time". I suggest to use "At the present". (changed)
2. Page 1, Line 31: I suggest to change "is" with "are". (changed)
3. Page 2, Line 69: I suggest to change "Land slide" with "Landslide". (changed)
4. Page 3, Line 75: I suggest to define a variable for the cumulative rainfall. Please insert "E (mm)" and rewrite "cumulative rainfall E (mm)". (changed)
5. Page 3, Line 75: I suggest to define a variable for the average annual rainfall. Please insert "MAP (mm)" and rewrite "average annual rainfall MAP (mm)". (changed)
6. Page 3, Line 75: I suggest to define a variable for the monthly average of evaporation. Please insert "MME (mm)" and rewrite "monthly average of evaporation MME (mm). (changed)
7. Page 3, Line 85: I suggest to replace "mm/d" with "mmd-1". (changed)
8. Page 3, Line 86: I suggest to change "day" with "days". (changed)
9. Page 4, Line 108: I suggest to define better the name of variables for the rainfall volume, rainfall duration and rainfall time (not changed, since they are defined clear enough)
10. Page 5, Line 149: I suggest to change "140-m" with "140 m". (changed)
11. Page 5, Line 150: I suggest to change "600-m" with "600 m". (changed)
12. Page 5, Line 151: I suggest to change "700-m" with "700 m". (changed)
13. Page 5, Line 169: I suggest to replace "mm/d" with "mmd-1" and please use the same number of decimal places. Please correct "4" with "4.0" and "6.26" with "6.3". (changed)
14. Page 6, Line 175: I suggest to replace "mm/d" with "mmd-1" and please use the same number of decimal places. Please correct "1.28" with "1.3". (changed)
15. Page 5, Line 169: I suggest to replace "mm/d" with "mmd-1" and please use the same number of decimal places. Please correct "4" with "4.0" and "6.26" with "6.3". (changed)
16. Page 6, Lines from 180 to 182: Please use the same format for the text. (changed)

17. Page 6, Line 180:I suggest to replace “N equals to 2” with “N = 2”. （changed）
18. Page 6, Figure 1: Please use the same graphic element for represent the horizontal scale and North indicator symbol.（changed）
19. Page 7, Figure 2: Please use an appropriate format for the x-axes, please remove the ticks on the upper x-axes. Please use a better representation for the legend.（changed）
20. Page 7, Figure 2: I suggest to separate the values of Year/cumulated rainfall from graph with a new table. changed）
21. Page 7, Figure 2: Please use the same number of decimal places.（changed）
22. Page 8, Figure 3: I suggest to use a 2D graph for represent the r, d variables, and a different scale of colours for represent the T value.

    A: We still believe that the 3D plot is a better way to express the relationships among the parameters. If a 2D plot is adopted, it will result in a wrong impression that T is the function of r and D; actually, it is not and is independent of r and d.
23. Page 8, Lines 211-212: I suggest to use a subscript index. Please change “C1” with “C1”, “C2” with “C2”, “C3” with “C3” and “C4” with “C4” （changed）
24. Page 8, Line 212: Numbers reported in the text “C4 = (2.45, 4, 7.33)” do not always meet them reported in Figure 3. Please check.

    A: Actually it might be caused by the visual distortion when it is projected into a transparent 3D coordinate.

    Page 10, Figure 4: Please use the same format for all the graphs. （changed）
25. Page 10, Table 1: Please use a variables to report in table the three types of rainfall input data. Please use the same number of decimal places.

    A: The decimal places have been unified.
26. Page 11, Figure5, 6: Please use the same format for all the graphs. In particular, the authors use the same colors to represent the values of displacement and value of the prediction error relative to the three types of rainfall input data.

    A: The may be caused by the pdf file has not truly reflect the figure expression in MS Word. See the revised draft for a better clarification.

27. Page 11, Figure5, 6: I suggest to use a two q-q plots representation. The quantile-quantile or q-q plot is an exploratory graphical device used to check the validity of a distributional assumption for a data set.

    A: A q-q plot for landslide displacement prediction based on 3 types of rainfall input and corresponding analysis has been added as follows.

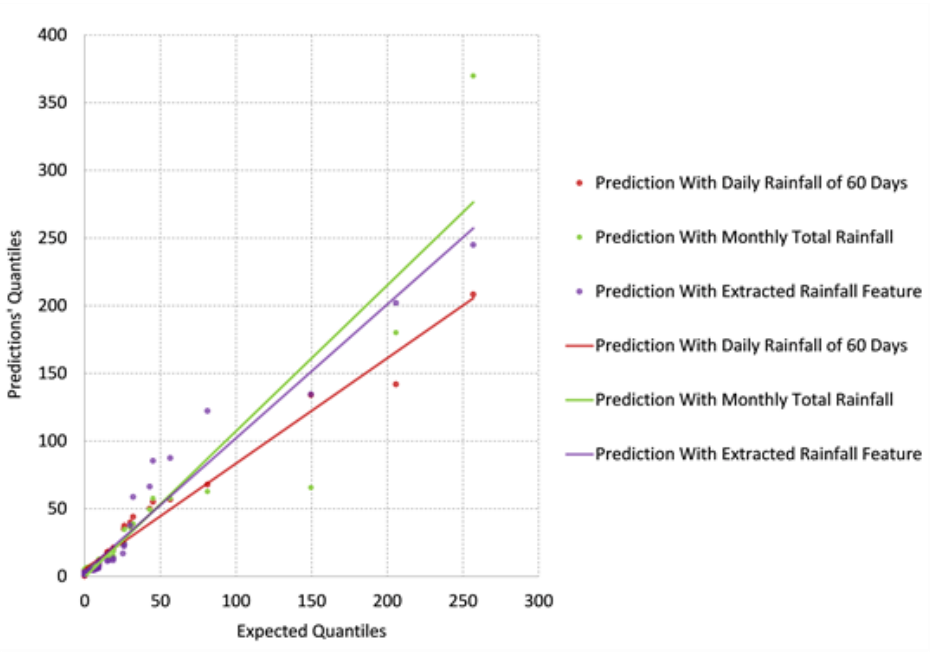

Figure 11: The q-q plot for landslide displacement prediction based on 3 types of rainfall input.

"The q-q plot shown in Fig. 11 is an exploratory graphical expression used to check the validity of a distributional assumption for data sets. It is employed for analyzing the relationship between observed displacement data and the predictions with three types of rainfall input. If the observed and the predicted data sets have the same distribution, the fitted line in the q-q plot will approach y=x. As can be seen from Fig. 11, the fitted curve of the data points from the prediction with extracted rainfall feature is closer to the line y=x with slope of 1; while the prediction with monthly total rainfall is overestimated and the prediction with daily rainfall of 60 days is underestimated. It indicates that the extracted rainfall feature represents real rainfall better than daily rainfall of 60 days and monthly rainfall in landslide displacement prediction."

We sincerely appreciate the detailed and constructive comments and suggestions from Referee #1.