GENERAL COMMENTS

The contribution “Rainfall feature extraction using cluster analysis and its application on displacement prediction for a cleavage-parallel landslide in the Three-Gorges Reservoir area” by Y. Liu M. and co-Authors is good and potentially publishable. 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.

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.

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?

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.

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.

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.

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?

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.

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?

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 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.

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

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

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.

TECHNICAL CORRECTIONS

Page 1, Line 30: Please rewrite better the following sentence “At present time”.I suggest to use “At the present”.

Page 1, Line 31: I suggest to change “is” with “are”.

Page 2, Line 69: I suggest to change “Land slide” with “Landslide”.

Page 3, Line 75: I suggest to define a variable for the cumulative rainfall. Please insert “*E (mm)*” and rewrite “cumulative rainfall *E* (mm)”.

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)”.

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).

Page 3, Line 85: I suggest to replace “mm/d” with “mmd^{-1}”.

Page 3, Line 86: I suggest to change “day” with “days”.

Page 4, Line 108: I suggest to define better the name of variables for the rainfall volume, rainfall duration and rainfall time

Page 5, Line 149: I suggest to change “140-m” with “140 m”.

Page 5, Line 150: I suggest to change “600-m” with “600 m”.

Page 5, Line 151: I suggest to change “700-m” with “700 m”.

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”.

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”.

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”.

Page 6, Lines from 180 to 182: Please use the same format for the text.

Page 6, Line 180:I suggest to replace “N equals to 2” with “N = 2”.

Page 6, Figure 1: Please use the same graphic element for represent the horizontal scale and North indicator symbol.

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.

Page 7, Figure 2: I suggest to separate the values of Year/cumulated rainfall from graph with a new table.

Page 7, Figure 2: Please use the same number of decimal places.

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.

Page 8, Lines 211-212: I suggest to use a subscript index. Please change “C1” with “C_1”, “C2” with “C_2”, “C3” with “C_3” and “C4” with “C_4”

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.

Page 10, Figure 4: Please use the same format for all the graphs.

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.

Page 11, Figure5, 6: Please use the same format for all the graphs. In particular, the authors use the same colours to represent the values of displacement and value of the prediction error relative to the three types of rainfall input data.

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.