Dear Anonymous Referee #2:

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

Next, in a sequence, we would like to respond to 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.

The authors propose a new data-driven approach to quantify various states of activity of landslides and support, in perspective, decision-making within early-warning systems. The topic is undoubtedly interesting and with a potential of providing better information on site-specific landslide activity.

Thanks very much for your encouraging words.

Nevertheless, I find that this paper does not really show whether the proposed approach gives a real advantage over other existing data-driven, empirical or physically based methods in quantifying landslide stability/instability. A comparison of several methods would be greatly helpful.

In introduction, the basic principles and main advantages and disadvantages of the existing methods (Saito's method, FEM, LEM) are detailed, hoping to highlight the starting point of this paper. In the case study, it is difficult to compare the results of this method with other methods whose results usually are presented with safety factor, because this paper indicates landslide instability with the proposed state fusion entropy index. As a supplement, more macroscopic phenomenon has been added as the evidence to validate the effectiveness of this method.

<span style="color:red">Saito's method is an empirical forecast model and is suitable for the prediction of sliding tendency and then the failure time. Based on homogeneous soil creep theory and displacement curve, it divides displacement creep curves into three stages: deceleration creep, stable creep and accelerating creep, and establishes a differential equation for accelerating creep. The physical basis of Saito's method helped it to successfully forecast a landslide that occurred in Japan in December 1960, but also makes it strongly dependent on field observations. LEM is a kind of calculation method to evaluate landslide stability based on mechanical balance principle. By assuming a potential sliding surface and slicing the sliding body on the potential sliding surface firstly, LEM calculates the shear resistance and the shear force of each slice along the potential sliding surface and defines their ratio as the safety factor to describe landslide stability. LEM is simple and can directly analyse landslide stability under limit condition without geotechnical constitutive analysis. However, this neglect of geotechnical constitutive characteristic also restricts it to a static mechanics evaluation model that is incapable to evaluate the changing regularities of landslide stability. In the meanwhile, LEM involves too many physical parameters such as cohesive strength and friction angle, which makes it greatly limited in landslide forecast and early warning. As a typical numerical simulation method, FEM subdivides a large problem into smaller, simpler parts that are called finite elements. The simple equations that model these finite elements are then assembled into a larger system of equations that models the entire problem. FEM then uses variational methods from the calculus of variations to approximate a solution by minimizing an associated error function. In landslide stability analysis, FEM can not only satisfy the static equilibrium condition and the geotechnical constitutive characteristic, but also adapt to the discontinuity and heterogeneity of the rock mass. However, FEM is quite sensitive to various involved parameters and the computation will increase

greatly to get more accurate results. If parameters and boundaries are precisely determined, LEM and FEM can provide results with high reliability. [Has been added in "Introduction"]

The macroscopic behaviors of Xintan landslide near historical maxima was investigated according to previous studies (Wang, 1996). In June 1982, some trees in the top area of Jiangjiapo were dumped. A small amount of north-west tensile cracks appeared on the steeper section of the east. Around August 1982, the front edge of Jiangjiapo went through a small collapse. In June 1983, the colluvial deposits between Guangjiaya and Jiangjiapo showed signs of resurrection. At the end of 1984, the trailing edge of the landslide showed an "armchair" shape and the leading edge was bulged out. Some collapse pits were found on the upper side while several new tensile cracks in the middle. Meanwhile, some small collapses which seem irrelevant to rainfall occurred. In May 1985, old cracks widened and new cracks appeared, forming a ladder-shaped landing ridge. Moreover, Jiangjiapo presented a clear trend of the overall slippage. These proofs suggest that the historical maximum index is highly consistent with landslide macroscopic deformation behaviors. [Has been added in "Introduction"]

Furthermore, there is no evidence that the method can be successfully used in an early-warning perspective, which is the goal set in the abstract. My main concern is that the entropy approach used by the authors is based solely on measurements of displacements, seemingly in a single point of a landslide. The authors show that the pattern of state fusion entropy is (not surprisingly!) consistent with that of displacements (input information). Thus, what does the entropy tell in addition to what is already obvious by looking at the displacement pattern and, perhaps, by setting displacement rate thresholds to provide early warning? This has not been clarified. In addition, can the performance of the model be improved by integrating several displacement measurements (and perhaps pore pressures, water level, water content, deep deformations, etc.)? This is an important topic to be addressed.

Thanks for your comments. This approach is proposed to analyse landslide stability changing regularities and further provide clues for landslide early warning. The cumulative state fusion entropy may be similar to cumulative displacement (Xintan landslide). However, they can also be very different which has been presented in the result interpretation of Baishuihe landslide. As for the data selection. Nowadays, one landslide may be monitored by multiple monitors with multiple sensors and various data can be obtained such as surface displacement, deep displacement, pore pressure, water content and so on. There is no doubt that all these monitoring data contain the information about landslide state and much more comprehensive landslide states can be obtained if all these monitoring data are utilized. However, this comprehensive monitoring data is not yet common. And thus a traditional operation, selecting one typical displacement data of GPS, is adopted for generality and simplicity. Research of multi-monitoring and multi-sensor data fusion has been carried.

It may be argued that the displacement rate thresholds are set arbitrarily in a displacement-based monitoring system. However, I see that even in this data-driven approach there are arbitrary site-specific decisions made by the authors (e.g. page 9 line 4), which perhaps can affect the model output. So, for a model to be truly data driven, I expect no arbitrary choices, or arbitrary choices to have little influence: the dataset should provide the answer itself.

Thanks for your advice. Theoretically, the k-means clustering method is based on the data distribution of input data. The cluster number K only determines the division roughness of clusters

and has little impact on the distribution of clusters which is the basis of the state fusion entropy approach. Therefore, the cluster number was empirically set to 3 in the case study. Now some strategies have been proposed to determine cluster number totally and automatically according to input data. And this can also be used as an improvement of the method.

Finally, the content of the work does not seem to match its title: monthly displacements are probably too far from a "real-time" landslide monitoring when incipient failure is concerned. I expected to see interpretation of daily, hourly or even more frequent observations of landslide displacements prior to failure.

Thanks for your constructive suggestion. While defining deformation states, deformation velocity and acceleration are selected because they are considered to represent the landslide deformation characteristics well on the assumption that displacement is monitored monthly. At this time scale, the monitoring error of GPS can be ignored compared to landslide actual deformation. However, as the time resolution of displacement monitoring data increases, the impact of monitoring errors will be greater. In this case, landslide deformation features may not be deformation velocity and acceleration but determined by some feature extraction methods. Neglecting the consideration of monitoring error, the method is capable to monitoring data with higher time resolution and corresponding feature extraction methods are under study.

Due to these concerns, I feel that this manuscript is not ready for publication in the present form. I recommend the authors update their work by addressing the above points and, in particular, by including evidence of good performance of their model in making usable predictions of landslide failure based on high-resolution displacement patterns, which could be used in an early warning system.

For now, the state fusion entropy is designed without the function of forecasting but it still offers helps for landslide stability analysis and further the early warning. Cumulative state fusion entropy reflects the overall instability of landslide and its changing forms (fluctuation around zero type and fluctuant increasing type) also do help to judge landslide evolutionary stages and deformation tendency. Besides, the historical maximum index indicates the renewal of the most dangerous state of the landslide and may server as a new clue for landslide early warning. But this new clue should not be exaggerated to such an extent that other clues can all be replaced. Once historical maximum is renewed frequently, other clues such as macro cracks should also be taken into account to fully determine landslide early warning level. [Has been added in "Discussion and conclusion"]