Review of Kuo et al., Evaluating critical rainfall conditios for large scale landslides

Kuo et al., present a landslide catalogue in Taiwan, obtained by rmote sensing, from which they extract 62 large landslides that can be accurately timed thanks to seismic detection, and compared to local rainfall gaging data. Then they assess which type of rainfall threshold could be derived for this dataset, including a threshold guided by physical considerations, and compare it to a dataset of smaller landslides in Taiwan. The paper ends with a rather unconvincing or unclear discussion on potential variabiliy of the thresholds and on issues sith seismic detection.

Overall, the authors present an interesting, novel dataset (although relatively modest) and do a series of classic (rainfall threshold) and less classic (physically based threshold) analysis that can be worth publishing, but the discussion and some of the analysis need to be improved before that.

Major comment:
1/ Timing is an issue but rainfall estimation as well. Notably because rain gage may be far from the landslides and not experiencing similar rainfall especially due to orographic effects. The author explain they only associate landslide with rainfall measured within 100km2. I think this is a good start but in the analysis it would be good to indicate (by a color coding ?) the horizontal distance from the landslide, as well as to discuss difference in elevation between station and landslide median elevation for example. This would allow the authors to discuss uncertainty and the degree of reliability of rainfall estimates for the landslides.

2/ I think the attempt of the authors to define a threshold based on physical considerations is worth, but insufficient in th present form : the assumption and limit of the model lack validation/discussion, and the practical utility/validity of the model copared to purey empirical ones is poorly demonstrated. I give detailed proposition to test and refine the model, but in any case a more quantitative comparison of the validity of the different threshold seems important if the author want to underline the physical model has a path forward. I think also this part may benefit from being put in perspective compared to other work on physically based threshold. For example :

-- Salciarini and Tamagni 2013, Physically based rainfall thresholds for shallow landslide initiation at regional scales.
-- Papa et al., 2013, Derivation of critical rainfall thresholds for shallow landslides as a tool for debris flow early warning systems
-- Alvioli et al., 2014, scaling properties of rainfall induced landslides predicted by a physically based model.

3/ I think the discussion needs to be revised sigificantly.
The authors seek to discuss effects oncritical threshold that cannot really be assessed with the data they have, while several points are not really discussed : For example 1/ uncertainty on rainfall parameters, 2/ the added value of seismic dating of landslide and its limit (size of landslide distance from stations (currently section 5.3 needs significant clarification) , 3/ The value of the critical rainfall volume : how better compare with other, how to determine or constrain I0 etc

4/ Last, I strongly suggest the authors to define variable names for antecedent rainfall (e.g. Ra), cumulated rainfall (e.g. Rc) to later compare with Rt (Rt = Rc + Ra) and to be consistent in text and figure when they talk about rainfall amount.

Line by Line comments :

P2 L 5 : LSL / SSL : this is heavy and makes the draft harder to read. Why not simply use small and large landslide and indicating the boundary is at 0.1km2 ?

P2 L21 : State in the text how was estimated the occurrence time. Based on peak rainfall correct ?
In Fig 1 Caption you say that in general peak rainfall intensity is used.This may go int the main text, with one or two references. Indeed, simple groundwater modelling (e.g. Wilson and Wieczorek 1995) could estimate soil moisture based on the rainfall data and find a maximal pore pressure after the peak rainfall. Other simple modelling approach or assumption may give different estimation times.

P2 L34 : Fractural geological conditions >> Fratcured rock mass
P2 L35 : slope disasters >> I would suggest slope failures , more general (here and at other place in the text)

P3 L21 : By a rainstorm (which one ?) or by the Morakot typhoon ? Please clarify.

P3 L25 : end of the setnece unclear. Main factor to separate SSL from LSL or to relate to rainfall triggerring ? If so how ?

P3 L 30 : Ok the triangular signature is typical, but could you cite and discuss what are other typical properties ? I know there are quite some papers discussing how to detect and classify landslides based on various properties of the spectrogram or of the waveform.

P4 L 3 : Only now we learn that the landslide mapping was done between 2009 and 2014. Please indicate it at the start of the mapping section.

P4 L 35 : Could you give an estimate of how often the location point and landslide maps matched ? And what was the maximal acceptable offset from a mapped landslide ?

P5 L 4 : Need some reference for that : the track does not necessarily say so much given the size of the diameter of typhoons are some times similar to Taiwan island size... And the windward slope is not obvious. If you refer to orographic effects say it clearly, but this also occur ar large scale not a fine scale.

P5 L5-10 : Very true indeed. Another important point may be the altittude of the gauging station and of the upper part of the landslide. If the gage is near the river at the outlet of the 100km² catchment possibly 500m or more below slopes where landslide happen the rainfall may be quite different.

P5 L 14 : Say if this is your definition (we define the beginining of a rain event) or a general one (then cite other studies.

P5 L18-20 : I understand it is hard to choose objectively which time should be considered for antecedent rainfall, but an arbitrary threshold without temporal weighting seems disingenuous... It is fair to use the official definition but what about testing a coupd other antecedent rainfall conditions : for example the cumulated rain over 3 or 5 days. Or a weighted sum over the 10 preceding days (with weight decreaseing with time before the event).

P6 L 4-7 : How was the occurrence time obtained for SSL ? Not by seismic means ? SO how accurate are these times ? Are we back to the same uncertainties as shown in Fig 1 ? Authors should clarify that.

P6 : Subsection 2.4 : missing "l", >> water model ?

P6 EQ 1 and 2 : ok but the assumption C' = 0 maybe quite a big one , especially for large bedrock landslides... Need to be discussed at some point, because Qc would be larger with non zero C.

P6 EQ 4 : Qc is actually the height of saturated regolith above the failure plane, in mm. Maybe clearer than calling it a critical volume. Note that in EQ 3 it is a critical height. But in EQ 4 it is simply a height assuming I0 is correctly estimated.
Another key issue is that this equation does not account for the antecedent rainfall. As I and D are for the triggering storm only, correct ?
Finally, I do not see why the authors assume a linear drainage. Most hydrological simple model of soil drainage (backed up by theory and observations) show a non linear drainage rate, where drainage increase with the amount of water in the soil (e.g., Wilson and wieczorek, 1995). I think the uthors should discuss this choice here or in discussion. Ths model is very easy to implement and use to obtain soil water level, only requiring the hourly estimate of rainfall and an assumed drainage parameters. I think it may be a interesting addtion to the paper to really make the authors model physical.
I note that a number of recent attempt to model physically landslide threshold (cf major comments) should be mentionned and discussed here and/or in discussion these models and how they compare to the author proposition.

P7 L5 : "their slope angles"
Do you mean the mean slope within the landslide body ?

P7 L7 : " This increase was most likely due to the fact that during the extremely heavy rainfall of Typhoon Morakot in 2009, more than 2,000 mm precipitated in four days, causing numerous landslides on lower slopes and reducing the stability of the steeper slopes in the following years."

>> I do not think this claim is supported by the data of Fig 4 : First in 2009 Morakot did not seem to be so different from 2005-2008 in terms of slope distribution. 2nd it only affected the southern half of the distribution of 2010-2014. If the hypothesis of the author is true, comparing only pre 2009 amd post 2009 in the Morakot area only (i.e. southern half of the dataset) would yield an even more pronounced shift, while the northenr half should have no shift. I invite the authors to check and show this to support their claim.
Alternatively they should try to check that statistical uncertainties may not be responsible for shift, and it would be interesting to compute a confidence interval on each histogram.
Last point, either if morakot did perturb the slope distribution the author need to clarify their argument, as it is not obvious how failing gentle slope would weaken steeper slopes (as a start the author could try to demonstrate that failing slopes in 2010-2014 are spatially related to 2009 failures)

P7 L18-20: Yes probably.

P7 L 24-26 : Not clear. To clarify.

P7 L 26 : Could you explain with some more details how these 62 LSL were obtained ? Is it the combination of neary gages and seismic signal quality ? Anything else ? One sentence for recalling the reader of thr criteria used would be helpful.

P8L25 : Interesting, but size is not the only difference with these other thresholds.
The fact you focussed on large landslides, requiring higher total rainfall, and thus higher I-D lines is likely contributing. However, how much of the difference could be due to seismic dating ? To the regional characteristics of the landslide (as some threshold are global, other taiwanese or japanese). I think these should be menionned here or in discussion, because your threshold for SSL is also much larger than most other threshold, and these SSL are more siilar in size to past study.

P9 L 1-2 : it was determined that Rt–D analysis could be used effectively to distinguish SSLs from LSLs.
>> I think it is very interesting to see in Fig 5B that the landslide size groups shift from small for relatively short duration and low rainfal amount to large landslides for long and very large cumulative rainfall.

P9 L8 : "conditions for SSLs included high average rainfall intensity but relatively low cumulated rainfall"
>> You plot Rt that is the total effective rainfall in Fig 5. So do SSL have low cumulated rainfall or low Rt or both (if Ra is low...)

In any case this plot is also quite interesting, as it matches well the theoretical expectations (Van asch 1999, Iverson 2000) stating that very large landslides will require high cumulated rainfall (unlikely to accumulate over short timescales) while small landslides may be caused by transient pulse of water accumulation in the shallow regolith relating to very high intensity, but that do not need to cumulate large amount of water.

P9 L14-15 : Not only Wieczorek and Glade could cited here. Van asch 1999, Iverson 2000 discussed that earlier.

P9 L20 : This seems like a very crude approach. I would strongly encourage the author to have a Compute Qc based on an actual estimation of the landslide slope and the landslide depth : Using Larsen 2010 or a local Area Depth relation from Taiwanese dataset (Chen 2013) the authors could use A to derive Z and thus obtain a more realistic extimate of Qc as a function of Z and the mean slope. The effect of small variations in porosity or friction angle could also be computed and shown.

I understand you want a single average threshold to compare to a population. Nevertheless you can make an almost individual prediction of each large landslide (with Depth and Slope) and compare it to uniquely constrain rainfall information, thanks to your seismic dating. I think it would be worth checking the validity of the model this way, and potentially refining the drainage model that seems critical to really obtain a physically based threshold.

P9 L 23 -25: Is this curve allowing to better predict the LSL compared to the other plots in Fig 5 (Especialy I – Rt or I-D?) Same question for the separation from SSL/LSL . The authors should provide some statistices confirming that this model is better than a Rt -I for example. Log Logplot is absolutely necessarry for all plot.
Further, the very low drainage found by the authors, mean their threshold is almost ID ~452 or R~452. And indeed a vertical line in the I -Rt graph at about 500 may be as good...

P10L14 : If so you should observe a larger fraction of the LSL in 2010-2014 neighborih a 2005-2009 landslide, compare to LSL in 2005-2009 bein the reactivation of older landslides. Given the small dataset (62?), I encourage the authors to check each LSL and report the proportion of reactivation

before and after 2009. Then they can support and discuss this hypothesis.

P10 Section 5.1 and 5.2 Strange writing: the authors oscillate between presenting new result about shift between threshold for different subset and then concluding that they are insignificant. Based onFig 7 and 8 I do think the dataset of the author is insufficient to discuss these two topics and I would strongly suggest the author to remove these two sections (or just mention rapidly that sub dividing the the dataset does not give clear difference and send Fig 7 and 8 in Supplement.) ans give more space to discussing other points, like their critical rainfall model, or the uncertainties on rainfall.

P11 section 5.3 : maye interesting but Fig 9 is too confusing. So I suspect text and Fig 9 should be clarified a bit.
P11 Eq 5 : to discuss validity and limits of EQ 5 it should be made clearer how (empirically?) and with which dataset/environment this relationship was obtained.

Fig 3 : closest station is MASB (in the caption) or SGSB (in the map) ? It means 90% of the landslide and seismic signal

Fig 5 : The last panel is not very clear : Cumulated rainfall is the total rainfall in the triggering storm. Antecedent rainfall has no reason to be compared directly with landslide occurrence, but only when summed with the cumulated rainfall. So why not show Rt the total effective rainfall together with Rc the cumulative rainfall (Given that Rt>= Rc it should be easy to visualize).

Fig 6 : Log Log scale is needed on all panel. Right now we do not see clearly the position of the different datapoints.

Fig 7 and 8 : I do not believe any of the subset can be significantly distinguished. What is driving the (small) difference in threshold curve is only 1 or 2 points out of each subset (that seems to be15-25 points). These low points shift the threshold while the bulk of each popuation do not seem different in any way. I am convinced this can only be due to chance and not to a shift of the whole population. I am even surprised that the curve are so low because if they are the 5% exceedance probability ~ 1point should be left out in subset of ~20...

Fig 9 : I really tried, but did not understand it... I got that the line, is an empirical estimation of the distance at which station should be able to detect a landslide of a given size.
What are the points ? The 62 LSL ? If yes why are they all above the line ? Does that mean only distant station detect the slides? I can believe for some but not the whole dataset, and this seems contradictory with Fig 3

References not used in the manuscript

-- Wilson and Wieczorek 1995, **Rainfall** thresholds for the initiation of debris flows at La Honda, California
-- Iverson, 2000, Landslide triggering by rain infiltration
-- Van asch et al., 1999, A view on some hydrological triggering systems in **landslides**
Larsen et al., 2010, **Landslide** erosion controlled by hillslope material