NOTE : In the following documents referees comments are in normal fonts and the answer to them are in bold fonts.

Review of Marc et al. “Towards a global database of rainfall-induced landslide inventories: first insights from past and new events” by David Milledge.

Major Comments

This is a well executed study with novel and interesting findings. I have three general comments and a large number of minor comments but neither the major nor minor comments reflect a fundamental problem in the research in my view.

I am not convinced that it is essential (or helpful) to present your inventories as the only inventories that are suitable for this type of analysis (as you seem to do on P2-3). Instead you could simply say they are one set of inventories and they demonstrate the power of this type of approach. I am not convinced of the need for landslides beneath an entire storm footprint to be mapped and am sceptical that entire storm footprints can be convincingly defined so I’m not convinced by your critique of studies that analyse far smaller study areas (other than on sample size grounds).

>> The mostly agree with the referee. We will develop the description along these lines:

1/ number is always important, in absolute term because below 50-100 landslide the reliability of any statistical treatment is uncertain but also in relative term, because a statistical study based on 500 landslides out of a storm that caused ~5000 take the risk to have biased interpretation if the (potential) specificity of the subset population are not notcied and understood (e.g., mostly large landslide ?, mostly landslide near river ? Mostly landslide in a given lithology ? Etc etc).

2/ If enough landslides are mapped within an AOI (e.g., >50-100 in total, > 75% of landslide above the resolution limit, leading to a reasonable frequency-size distribution) the inventory above the AOI is likely to be statistically usable and representative of the various processes and conditions affecting the process in the AOI. Then a partial inventory will indeed allow to study any local parameter and their variations within the AOI : e.g. landslide density, landslide size distribution, relations to slope etc.

3/ However, a comprehensive inventory may have the additionaly advantage to gather enough landslide across different areas (in terms of lithology, relief etc) potentially allowing to establish a hierarchy in controlling parameters and also allowing to study an averaged landslide response less likely to be dominated by specific site effects. comprehensive inventories are the only ones allowing to study the variations of total landsliding.

We will include a synthetic description of these point in the early part of the manuscript. These should answer the various minor comments about the importance of "comprehensive" landslide catalog, and better acknowledge the potential use of "partial" inventories.

The methodology description could be more consistent between inventories. Similar information is reported for each case but the style of the reporting differs and some key information reported in some cases is not present in others (e.g. image source, image resolution, acquisition date).

>> We will work on a higher degree of consistency.
However, we disagree on key image source/resolution/date not beeing present, as all this information is in Suppl Table 1. Is the referee suggesting that we integrate this table within the main text ?
Or simply specifying some terms in the main text ?

I am not convinced that your focus on ‘comprehensive’ inventories is necessary nor that examination of total landslide numbers, volumes or areas are particularly meaningful in relation to rainfall triggered landslide inventories (though I think the findings on landslide density and slope are extremely interesting and thought provoking). This focus might reflect a desire for comparability to co-seismic landslides but I think the two triggers are importantly different. For example, it is extremely difficult to define the spatial and temporal limits on a single storm. In addition I find the results relating to total numbers, volumes and areas less convincing because they are predicted from a small number of point rainfall records. A clearer explanation of why ‘comprehensive’ inventories and total statistics are important would be a valuable addition to the paper.

Minor Comments

P2
L30: comprehensive: this term needs defining.
>> We mean that all landslides detectable above the resolution limit were mapped, and that we could observe the landslide density fading away in all direction, indicating the limits of the footprint of the high intensity part of the storm.
We will specify this two criterium in the main text.

P3
L7: comprehensive mapping: where do you start and finish. Your definition of a storm is very important here and I don’t see it at the moment. For example shouldn’t the Morakot mapping extend to the Phillipenes and China on this basis?
>> The pragmatic answer is that the landslide response in Philippines and China was negligible compared to the one in Taiwan (We doubled check that quickly by looking at Landsat images where no hyperpicnal flows or alluviations in stream exiting hilly areas are visible, contrarily to Taiwan where these processes are very clear). Same is true if we look at landsliding in the rest of Japan hit by Typhoon Talas progressing Northward after hitting the Kii peninsula. This may be in part due to topographic difference, but I think this is also due to the fact enormous amount of rainfall was poured on this topography, probably largely because of orographic effects (cf Chien and Kuoa 2009, Taniguchi et al., 2009). Thus preceding rainfall on less high topography (e.g. in the Philippines) probably received much less rainfall, and the following rainfall over China (or Japan for Talas) was also likely less simply because little or no recharge ove the ocean was possible and a significant fraction of the typhoon moisutre was used up.

A theoretical answer is more difficult to find and would require a proper definition of a storm event. A tentative meteorological definition could consider a mass of moving moisture with a single source of moisture. A typhoon or (afternoon) convective cell could be such an object, that can then travel and pour its accumulated moisture as rainfall (and or snowfall) over an area limited in space an time. However, for the landslide community, only the relatively high intensity part of this rainfall matters, (as the part being below the landslide threshold can be neglected) and the spatial and temporal limits of a storm event could be further limited (as in the case of Morakot and Talas, where the orographic rainfall effect limit greatly the part of the typhoon relevant for landsliding).

Chien, F.-C. and Kuo, H.-C.: On the extreme rainfall of Typhoon Morakot (2009), J. Geophys. Res., 116(D5), D05104, doi:10.1029/2010JD015092, 2011.

Taniguchi, A., Shige, S., Yamamoto, M. K., Mega, T., Kida, S., Kubota, T., Kachi, M., Ushio, T. and Aonashi, K.: Improvement of High-Resolution Satellite Rainfall Product for Typhoon Morakot (2009) over Taiwan, J. Hydrometeor., 14(6), 1859–1871, doi:10.1175/JHM-D-13-047.1, 2013.

L30: this gets at a difficult issue, what do you include as a landslide? I think you need a clear definition that can be applied across all inventories and I don't see one at present. Divergence from the definition in different inventories will introduce bias to your results.
>> We tried to avoid mapping (or remove in the already mapped inventories (J11, TW9) deposition and erosion in the fluvial system, broadly defined as the areas with permanent flows, visible in the high resolution image. This meant that debris flow on hillslope would be mapped but not its prolongation within the fluvial system. We considered bank collapses as a disturbance that would be localized, usually not symmetric and not necessary linked to a landslide/debris flow on the hillsllopes. Clearly we may miss some bank collapse, and where to put the limit between a debris flow on a hillslope and its continuation on the fluvial channel is difficult and somewhat subjective. However, if amalgamation is avoided, the width estimate (and thus scar area and volume) will be relatively insensitive to these issues, that mainly affect the total runout and aspect ratio.

P7
L4: fluvial system: How do you define the fluvial system and how did you identify it for the study area?
>>The fluvial system was broadly defined as the areas with permanent flows, visible in the high resolution image. As we do not perform any anaylsis on the relation between the hydrographic network and landslide this appraoch is just aiming at making sure landslides are limited to hillslopes (cf comment P3 L30).

L20-21: maximal forcing: this doesn't seem to be consistent with your argument for the importance of complete landslide footprints. You are comparing the forcing at a single location within the footprint to the properties of the entire footprint.
>> Well the maximum forcing is taken as a "storm magnitude", and it is compared to total landsliding and peak landslide density close of this maxima l forcing, so we think the approach is reasonable.
We think that the issue of the referee is that in the introduction we push for "comprehensive landslide inventory" and later do not make fully use of it.

This is for 2 reasons: 1/ We have access to extensive rainfall data constraining the spatial pattern of rainfall for only a few cases, and the analysis of the spatial pattern is beyond the scope of this study and left for a future study (Marc et al., in Prep).
2/ Generally a comparison of storm magnitude with total landsliding requires an accurate order of magnitude of the total landsliding. We think that the fact that storm magnitude correlates well with total landslidng suggest some internal correlation between the peak total rainfall, and the mean rainfall and its variability within the storm footprint. Still we acknowledge that such correlation may not hold for all type of rainfall events: In our database, small events are likely

brief convective thunderstorm (C99,C15), while large ones are typhoons (M02,TW8,TW9, J11), which are very large singular system that are loaded during their displacement over ocean, and unloaded on landfall, even more importantly when hitting high relief. They fit well in a tentative definition of a storm event based on a single process / source of moisture accumulation and subsequent downpour on a given spatial/temporal zone (cf comments P3 L7).

The cases of B08 and B11 are more complex as they may results from interactions between multiple oceanically sourced moisture and specific meteorologic conditions on land. As mentionned in the main text, other rainfall period such as a monsoon could also rather be characterized as the sum of repeated convection events then transported across India, not allowing to differentiate individual meteorological event.

We will try to add some elements of this discussion into section 2.3

P9
L5: how, and where, did you measure landslide width?
--> The width was initially measured by GIS on a limited number (~50) of randomly selected slides in Colombia and Japan. To make this point more robust we proceeded as follow:
Text added in the revised manuscript:
"We measured systematically the width of 419 randomly selected landslides across all range of polygon area and aspect ratio, in the following inventories : J11, TW8, B11, and C15. The width was measured on the upper part of the landslide only, the likely scar, and ~4 width measurements made in arcGIS were averaged. When compared to the equivalent width obtained through our runout correction, 72% of the polygons are within 30% of the measured width and 96% within a factor of 2 (Fig Suppl X). We do not observed a trend in bias with area nor aspect ratio, except perhaps for the automatically mapped landslide in B11, where high aspect ratio correlates with underestimated width"

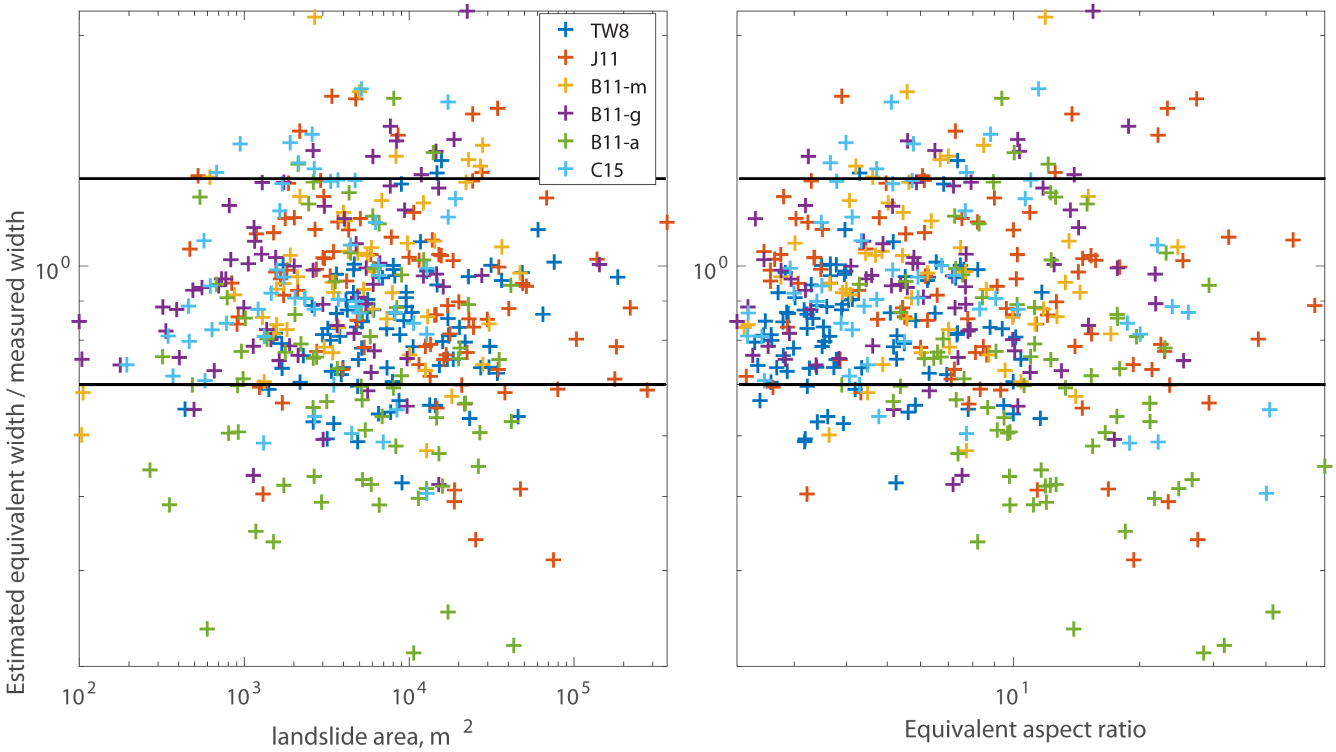

Figure Suppl X : Ratio of Estimated to measured width against landslide area (left) and landslide equivalent aspect ratio (right) for 418 landslides randomly selected during 4 rainfall events (TW8, B11, J11, C15). Most of estimated width are within 30% of the measured width, as shown by the horizontal black lines.

L6: I think you could state this more simply by saying that you assume that scars have equal length and width. This is the same assumption used by Pelletier et al., 1997.

--> This would have been an option. However, we became aware of a study presented at EGU general assembly of 2018, where the aspect ratio of a number of landslide scar has been analyzed. Domej et al., 2017, reported that in average the length-width ratio remained close of 1.5 for all landslide size. So we follow them and assume that 1.5 represent a good average of the length width ratio of landslide scar.

References:

Domej, G., Bourdeau, C. and Lenti, L.: Mean Landslide Geometries Inferred from a Global Database of Earthquake- and Non-Earthquake-Triggered Landslides, Italian Journal of Engineering Geology and Environment, (2), 87–107, doi:10.4408/IJEGE.2017-02.O-05, 2017.

P10

L24: to what extent is the landslide distribution area constrained by your study area (i.e. the extent of available images). Taking this to an extreme did Typhoon Morakot trigger landslides in China or the Phillipenes and should these also be included? This again reflects something that I think you need to discuss somewhere, the differences between rain storms and earthquakes as triggers: where are they similar enough to borrow frameworks from one another and where do they differ?

>> Well it is clear that, contrarily to earthquake that have a well delimited source (across the fault), the rainfall forcing is moving together with the storm and can travel other significant areas.

P11

L24: peculiar distributions: are these distributions peculiar if you are seeking power laws but not if other alternatives are considered? Have you tried a log-normal distribution? Negative curvature of the tail in log-log space sometimes indicates better fits for log-normal distributions?

>> The question is difficult to solve and not so important for our studies: Fitting log-normal distribution by MLE we obtain better agreement for some distributions and worst for others (comparing the Kolmogorov-Smirnov Test statistic and Anderson Darling test statistic obtained for log-normal or IGD fit obtained by MLE).

In other work the Inverse Gamma distributions has been found to provide the best fit to 3 large landslide catalogue (Malamud 2004). Further some work on the theoretical emergence of landslide size distribution also predict power-law decaying tail (Stark and Guzzetti 2009), with a tail related to the mechanical properties of the medium, implying the debate may not only be a question of goodness of fit, especially given that some datasets may be affected by artifacts.

Although we can mention these facts, solving such a debate is clearly out of the scope of our work. We will of course double check wether or not LogNormal fit parameters (Mu and Sigma) are correlated to rainfall parameters.

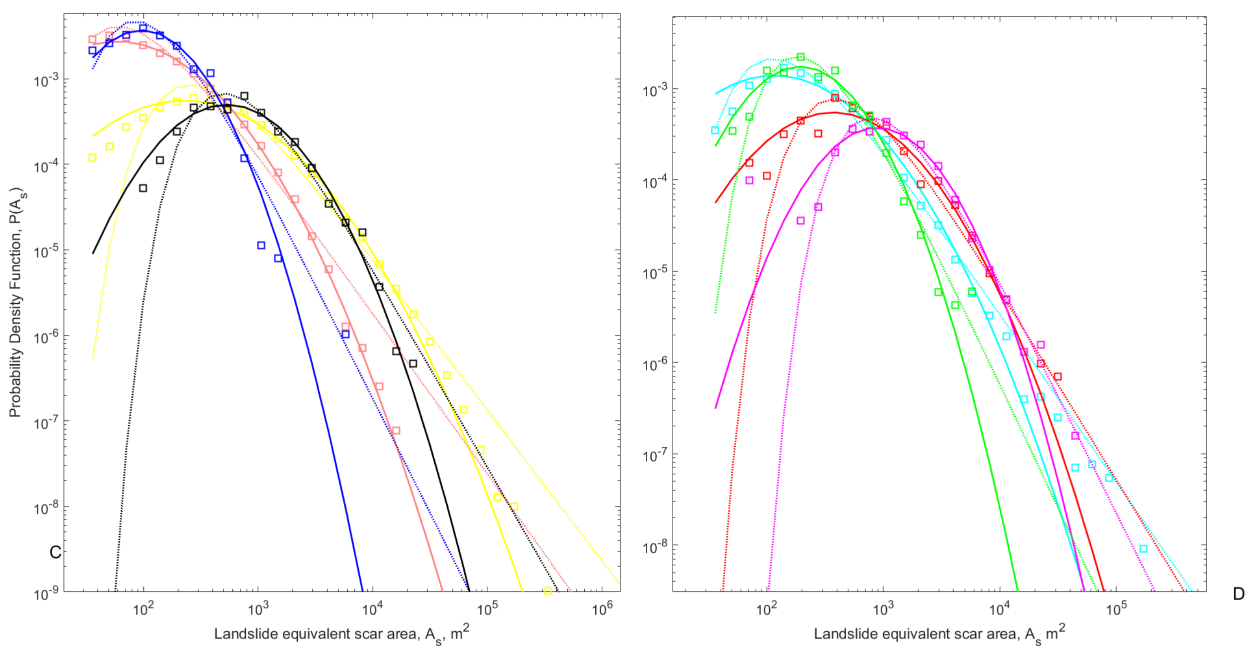

Figure for the discussion: Comparison of lognormal distribution (Solid) and Inverse gamma distribution (dashed) for the best fit of landslide scar size distribution. Fit are obtained by MLE.

L29: aspect ratio below 2: why below 2? What are the specifics of the equation? I had understood it to be A=w^2, which would give an aspect ratio of 1.

--> This is now updated based on Domej 2017 (Cf comments above). This is a reduction to an

aspect ratio of 1.5.

L11: focussing on scar areas seems sensible but this particular approach seems strange and the choice of modal topographic slope somewhat arbitrary, could you provide a more robust explanation for this choice? Alternatively couldn't you have used your previously defined scar area (w^2) to identify scars as the highest w^2 area of each polygon? This would be consistent with your previous definition and would avoid introducing an arbitrary slope threshold which could bias the results.
--> We note that for C99 there is not much difference between initiation point and the steeper part of the landslides (i.e, after the mode). Additionally we simply do not interpret what happens below the topogaphic mode but can certainly show it. Reducing all landslides polygon to their scar would require some work but is possible and we will try to examine whether or not it creates any difference in the results.

P13
L14: Could you use line thickness to indicate the slope beyond which small numbers of cells in the value range preclude interpretation of the line? It would be useful for the reader to know where that point is for each dataset. Also could you colour the lines in Fig 5 by storm duration? This might make it easier to pick out the behaviour you are identifying in the text and to make a connection between 5A and 5B.
>> Actually we plan to use the notion of prediction interval of the landscape slope distribution to assess if the landslide-affected slope distribution is staistically different from the landscape one.
With this method we will indicate which part of the distribution can be robustly interpreted and which one are less robust.
We will change the colour-code to reflect the duration, this is a good suggestion.

P17L9: Total storm rainfall: These results are extremely interesting. They suggest that absolute rainfall properties are good predictors for landslide properties. In the rainfall threshold literature there has been debate over whether absolute rainfall properties are driving failure or whether it is the degree of deviation from normal conditions (e.g. expressed as percentiles). It might be useful if you could reflect on this in relation to your findings. Would a plot of rainfall percentiles for these storms look very similar to the plot of absolutes that we see here?
>> By deviation of from normal conditions, do the referee means the comparison between the sotm rainfall and for example mean annual rainfall or mean seasonnal rainfall ? Or more something like the estimated return period of such a storm ? It may be difficult to estimate one or the other for a number of events but we can try (Mean annual or seasonnal rainfall may be tractable).
It is clear that although total rainfall may be a good predictor of the relative amount of landsliding between different storm (as shown in Fig 6 and 7) the control on landsliding must be more compliacted as in the surrounding area similar rainfall occur without triggering landslides (Taiwan, Japan), or in the same season similar total rainfall did not trigger landslide (in Colorado), so either antecedent rainfall or some constraints on intensity will be needed to generalize /strengthen the results we found.
In any case we will add such caveat somewhere in the discussion.

P18L17: we have no clear physical explanation: isn't this something that either extreme rainfall community or the hurricane community have thought about? It would be useful to point readers to

key reference from that literature here even if you don't strongly back one particular explanation.
>> We will look at this literature to try to suggest interesting reads. Some work indeed study the correlation between the total rainfall on land of hurricane and tropical storm with their diameter and travel velocity (Jiang, et al., 2008). We will try to relate to such work.

Jiang, H., Halverson, J. B., Simpson, J. and Zipser, E. J.: Hurricane "Rainfall Potential" Derived from Satellite Observations Aids Overland Rainfall Prediction, J. Appl. Meteor. Climatol., 47(4), 944–959, doi:10.1175/2007JAMC1619.1, 2008.

References
Clauset, A., Shalizi, C.R. and Newman, M.E., 2009. Power-law distributions in empirical data. SIAM review, 51(4), pp.661-703.
Cox, N.J., 2007. Kernel estimation as a basic tool for geomorphological data analysis. Earth surface processes and landforms, 32(12), pp.1902-1912.
White, E.P., Enquist, B.J. and Green, J.L., 2008. On estimating the exponent of power-law frequency distributions. Ecology, 89(4), pp.905-912.

Anonymous Referee #2

Received and published: 16 May 2018

General Comments **(Summary paragraphs NOT reproduced here)**

In terms of the general aspects of the manuscript presentation and layout, I find that the paper needs some fine-tuning and clarification, but overall it is close to being a finished product. The abstract provides a clear summary of the work, and the overall structure and segmentation of the manuscript is easy to follow. The title largely makes sense, although I find the last section stating "first insights from past and new events" confusing since all datasets are within the past twenty years and the youngest event occurred in 2015, and I'm not sure what that phrase adds to the description of the research. There are a number of places where language needs to be altered slightly, and I've tried to provide examples below in the technical comments section of the review. Mathematical formulae appear to be largely correct, but abbreviations for the landslide inventories (although intuitive) are not defined before they are used. Additionally, I found that Figure 2 should be modified, as it is very difficult to see the pink landslide polygons draped over the red and green topography. I imagine this would be especially difficult for people who are red-green colorblind. The supplementary material complements the manuscript well, and I have a few comments regarding supplementary figures below. Although I have a few additional concerns related to content and clarification, overall I think this paper will make an interesting contribution to Esurf..

>> We thank the reviewer for its interest in our study and findings.

From these general comments we retain :
1/ A clarification of Figure 1 (not 2) showing the landslide inventories.

2/ Some edits of the title : We could indeed drop the second statement.
An interesting alternative, slightly more descriptive could be "Towards a global database of rainfall-induced landslide inventories: first insights on landscape scale landsliding caused by rainfall event"
This option includes somewhat the notion of global magnitude of landsliding (e.g., Fig 6,7) and spatial distribution witin the landscape (e.g., Fig 5).
I would welcome comments of Referees and AE on such a title, and if they oppose it, we could simply stop with "Towards a global database of rainfall-induced landslide inventories", although it does not leave a hint than the paper do not only report on collating data but also analyze and interpret them.

3/ Check that all landslide variables are defined in the text, and improve and correct texts, following both Referees technical comments.

1st Paragraph: I'm not sure I agree with the statement that the goal of constraining quantitative relationships between landslide occurrence and rainfall is out of reach. The authors cite examples of this in the same paragraph. I do agree that there is certainly room for improvement in his area, which I think is the implied sentiment here.
>>Ok, we will rephrase in this direction.

Equation 1. I appreciate that the authors' goal here is to try to bridge the gap between purely deterministic models and purely statistical models, but I think that there needs to be a little more clarity. At the end of Paragraph 1, for example, you state that certain parameters such as permeability and cohesion that are required for deterministic approaches make a landscape-scale approach in data-poor regions inapplicable, yet you specifically include those parameters in your idealized semi-deterministic Equation 1. Why then is a deterministic approach not appropriate? I think that a bit more discussion might clarify these discrepancies.
>> Deterministic approach will require a fine scale representation of porosity and its variability at a fine scale: more or less the one of the landslides, so at 10-100m. In a semi-deterministic approach we may need only a constrain on the mean porosity (and perhaps some other aspect of its distribution like its variance or skewness) within a whole catchment or 10x10km catchment. Obtaining such information remains a challenge but may be more tractable, and may be correlated to other large scale observable (From hydrological behavior to soil maps ?).

Page 4
Lines 31-34: I'm confused by this sentence. When specifically are data from 2010-2012 used? When May-June 2009 data are not available for a specific location in the landslide-affected area?
>> Exactly. We will clarify this, but in most places in this case imagery just after the event is available in Google Earth.

Section 2.2 overall.
I'm also confused with the general methodology here. You map landslides on 30 m Landsat imagery, as well as on higher resolution imagery within Google Earth, but only in areas where a negative change in NDVI was observed at the 30 m scale. You then say that field mapping in the area reports twice as many landslides than was observed via remote sensing, but that the missing landslides must be smaller than #1 m resolution. Could the missing landslides not just be in areas that didn't result in a

negative NDVI shift in the landsat imagery? For example, a small translation or slump in a forested area may not affect a 30 m pixel.
>> This is true, it would however be relatively small landslide: indeed landslide much smaller than a landsat pixel (e-g, 10x10m rather than 30x30m) that looked fresh were almost systematically causing reduction of NDVI in one or 2 pixel, although the NDVI reduction was smaller than for large landslides.
So the Landsat NDVI is very sensitive to sub-pixel size landsliding, but the Google Earth imagery is essential to only map landslides and not many other anthropogenic/biological processes changing the NDVI.

Page 11
Lines 23-27: In your discussion of peculiar landslide frequency distributions, you focus on deviations from the (perhaps) expected Inverse-Gamma distributions at the large end of the distributions. What about deviations on the smaller end? For example, in the Total Area distributions, TW9, B11, C99, and J11 deviation from the maximum likelihood estimations pretty substantially for small landslide areas. Converting total landslides area to landslide scar area (As), the TW9 distribution especially deviates quite far from the expected P values. Is there a known reason for these deviations? I am far from an expert on landslide frequency distributions, but it seems worth discussing since it is quite apparent on Figure 3!
>> In agreement with Referee 1 we will also add a few line about the quality of IGD vs LogNormal (or essentially exponential tail). However, this will remain superficial given deciding on the functional form of size distribution is clearly out of the scope of our paper.

For the deviation in small landslide size, they are difficult to interpret because after the roll-over, censoring issues and difficulty to distinguish multiple adjacent landslides are likely important even with high resolution imagery. Given that these deviations are important almost only in TW9 where amalgamation and mapping artifacts were frequent but could not be mitigated down to the smaller landslide sizes, I would not try to interpret them.

Figure 3: Similarly, I don't believe it is mentioned why the authors choose to break up their landslide populations into two groups. Is this just to more easily visualize? Or is it based on the quality of datasets?
>> Indeed the split is just for visualization. We will specify it in the caption.

Page 15
Lines 11-12: Is there a plot that shows the relationship described on these lines? I couldn't find one. Maybe it would be worth including these in the supplemental material.
>> Ok we will add such a figure in the supplement.

Page 16
Lines 8-9: Very cool.
>> Thank you, this is indeed exciting.

Page 18
Lines 12-15: If there is a continuous forcing of heavy rainfall over an extended period of time, it is not clear to me why a monsoon would not fit in with the scaling relationships

derived in this paper. Would that not be an end-member condition for considering the role of water infiltration in setting the spatial distribution of landslides on lower slopes? If not, then why not? I imagine other people not as familiar with monsoon dynamics like myself might ask the same question.

>> The reviewer is certainly right about the slope sitribution, but not for the scaling between total rainfall and landsliding. To clarify both point we added:

Page 18 L20: "Indeed, in a long period with fluctuating rainfall such as the monsoon, drainage and storage of water will certainly not be negligible and the derivation of a soil water content proxy will be necessary (cf., Gabet et al., 2004) ."

P20 L23 we will add "A testable hypothesis would be that large populations of landslides caused exclusively during the monsoon would exhibit a strong undersampling of the steepest slopes"

Page 19

Line 14: Does the proportion of flat ground affect the slide aspect ratios as well, since the flat ground may provide more accommodation space for runout?

>> Maybe, but these sentence really relates to the availability of topography within the footprint of the storm event, so the landscape scale rather than the landslide scale, where local slope variations can indeed influence runout.

Line 22: This sentence cannot be true, as Figs. 6 and 7b all show a relationship between storm metrics on landslide scar areas. Do you mean other storm metrics outside of storm total rainfall?

>> Here, the reviewer confuses the total area (counting only scar or whole landslide) shown in Fig 6A, 6B and 7, and the individual scar area distribution.shown in Figure 3.

To avoid such confusion for other readers we rephrase to :

"We do not find a clear influence of storm metrics on the probability distribution of individual landslide scar areas or landslide runout (Fig 3)."

Page 20

Line 20: This is almost certainly true, especially for the smaller-area landslides that depend on local slope smaller than what a 30 m pixel can resolve.

>> We also think that. Thus an analysis with a high resolution DEM maybe needed.