Manuscript reference number: hess-2017-308 - Response to Anonymous referee #2

We would like to thank the referee for his review and for the interesting comments. We provide here a response to his comments together with our proposed edits to the manuscript. The referee's comments are reported in black and denoted as RXCY where X is the reviewer number and Y is the corresponding comment number whereas our response is in blue.

In this manuscript the authors investigate the effects of spatial aggregation of precipitation on the power law Total Raingall-Duration thresholds for debris flow. The study is based on 11 storms inducing 99 debris flow events in a region in the North of Italy and uses 5-min radar data with a spatial resolution of 1 km2. The spatial aggregation ranges from 1 to 20 km cells, corresponding to resolutions typical of remote sensing data. Additionally the authors compare the results of the spatial aggregation with those obtained with a synthetic raingauge network of different densities (1/10 up to 1/100 km2). Overall the paper is well written, with a clear structure and objective. I believe it could benefit from some more elaborations on some of the aspects presented, mentioned here below. I recommend minor revisions before publication on the journal.

We would like to thank the referee for his review.

R2C1

P5 L27-31: The authors conclude from Figure 4 and the increase of the log-residuals' standard deviation with decreasing synthetic raingauges network density that "debris flow occurrence thresholds derived using very high density networks (1/10 km-2) are comparable to the ones obtained using aggregation scales of 20-km grid size, corresponding to averaging areas as large as 400 km2". This conclusions seems to be based on the relative error of the alpha parameter with p=5% (ca. 25% underestimation for both 20-km grid size and 1/10 km2 network density) and therefore be true only for this specific method chosen (following Brunetti et al., 2010). For instance the alpha for p=50% for 20km grid size corresponds roughly to 1/100km2 density.

Thank you for pointing out this aspect. We agree with the referee on this comment and we would like to update the text in the manuscript to highlight this aspect: "*As a result, debris flow occurrence thresholds obtained using aggregation scales of 20-km grid size (corresponding to averaging areas as large as 400 km2) are comparable to the ones derived from relatively high-density rain gauge networks, such as 1/10 km-2 for 5% exceedance probability thresholds or 1/100 km-2 for 50% exceedance probability thresholds.*" Moreover, a sentence in the conclusions would be updated accordingly (see response to R2C2).

R2C2

P6 Conclusions: on the same line of the comment above, the authors could elaborate a bit more on the effect of the specific choice of method made and its effects on the conclusions. For instance, how do the authors believe the results would changing when applying a different method for the definition of the thresholds? Already the choice of p seems to affect the conclusions.

As underlined in the response to R2C1, the choice of the exceedance probability level used for the thresholds does actually affect the quantitative values reported in the above mentioned comment and a sentence in the conclusions should be updated to decrease the impact of the reported example: "*Conversely, rainfall estimates from synthetic rain gauge networks present large estimation variance so that debris flow occurrence thresholds at 5% exceedance probability derived from densities as high as 1/10 km^{-2} are can be comparable to 20-km grid size spatial aggregation and thresholds from sparser networks are largely underestimated*"

However, this does not affect the conclusions of the study that are pointed towards the use thresholds as "lower envelop curves" [P4L18; P6L3; P6L23] and are not specific (except the particular example

mentioned above) of any particular choice of exceedance probability (in the case of frequentist) or of the triggering events-based method used, as soon as an exceedance probability lower than 50% is used ("the results we present hold for any probability level lower than 50%, i.e. for any threshold representing a lower envelop curve to the (E, D) pairs" [P4L17-18]).

R2C3
Furthermore, how would the results change when applying a method that accounts not only for triggering events, but also non-triggering events?
The use of non-triggering events for the definition of the thresholds has been thoroughly discussed in our response to referee #1. We would like here to refer, in particular, to our response to R1C1.

R2C4
P4 L1: should be Figure 2b
Thank for noticing. The figure reference will be updated.