Since I'm a late addition to the review here, in fairness to the authors I've tried to primarily focus on how the authors addressed concerns from the previous stage. I'm currently stuck in a cycle of receiving a new set of concerns from a new set of reviewers repeatedly, and I don't want to do the same thing to someone else! In general, I think the authors have done a pretty good job of addressing prior reviewers' concerns in the previous round of revision. I'll raise some issues in the line-by-line comments that are mostly related to clarity and presentation and could be addressed without any new analysis. The biggest issue I see scientifically is the centrality of the argument about the levee effect being driven by check dam construction – this is an important thing to get right for work like this that could potentially influence decision making and I don't think that a dataset of development 3 catchments, two with check dams and one without is compelling support for the levee effect here. However, there's still some interesting stuff here that should be more or less ready to publish with some reframing and clarification.

28-30 – incomplete sentence – needs copy edit

107-108 – Please clarify this sentence – it sounds like landslides are being included as a subcategory of debris flows as written. Also – why would debris flows be constrained to the south? I can find lots of documentation about debris flows in northern provinces in China, maybe eliminate.

179 – Clarify "properties"

190 – commas make this sentence awkward

231 – Is this assessed cell by cell?

236 – Why is fragility just assessed as a binary? Would be nice to have a little clarification there.

Figure 3 – What do the numbered tributaries correspond to?

362-363 – This is a little confusing per earlier – isn't exposure just being assessed by whether a structure is in the path of a debris flow or not?

Figure 5 – In Cutou Gully, the shading of the tributary channel with the pale blue color only implies that it wasn't there 2005-2010? Also, some of the roads (dashed purple, solid red) lack symbols in the legend. Not a huge deal though. Also, were any of the areas shaded as built environment in 2005-2010 subject to debris flows? It's a little hard to parse whether the

changes in built environment from 2011-2019 represent just additions to pre-existing infrastructure or replacement to areas that were damaged. More importantly, the spatial scale is not correct. The small scale bar seems to represent 5 km (it's pretty small, I may be misreading) but checking out the field sites on Google Earth they're an order of magnitude smaller than that.

383-390 – Given the questions I had about figure 5, maybe it would be helpful to include some example satellite images showing damage in these events and your methodology of assessment. This would be fine just as a supplement too.

397-398 – Is there an available volume estimate for any of the debris flows in these catchments? Forgive me if you already listed it somewhere and I just couldn't find it.

Figure 6 – By "low and high runouts" do you mean the 10^4 and 10^5 m^3 debris flow simulations in LAHARZ? Please clarify. Maybe it would be helpful to show the result of some of the smaller simulations at a smaller scale, though, to visualize the effect of the check dams.

412-415 – Please clarify what we're looking at in 6iii that shows the effectiveness of the check dams – I'm struggling to see what's illustrative in the figure. Also, the confusion I had a little earlier with visualizing expansion in the built environment in Figure 5 is compounded here – the text states that expansion was restrained in Xiaojia Gully but in Figure 6iii it appears that all the development was post-2005 at the very least.

415-417 – I agree that the levee effect might be an important consideration here and the check dams might have played a role in encouraging development, but three catchments is an extremely small sample size to rely on. I hesitate to ask the authors to expand on their analysis at this point, but it would substantially strengthen the paper to look at a few other catchments with and without check dams in place and see whether this trend holds up to a larger sample size.

420-423 – I don't quite understand what this means. The landscape seems like a really important factor in all these simulations, but I wouldn't invoke it if you're not directly analyzing it here – it just gets the reader thinking that it might be more important than the check dams.

Figure 7 – What drives the extreme jump up in destruction in Cutou under the largest debris flow volume simulation? Also, I think it would be more effective to have at least a little analysis of debris flow inundation decoupled from building damage since that's dependent on a lot of

factors not controlled for in your analysis. How do the check dams control the distribution of erosion and deposition?

469 – I don't think this even needs to be stated

481-482 – How did the dams fail? Collapse? Overtopping?

519-520 – Declined from what? Earlier industrial presence? Or is it just less industrialized than the other catchments?

525-526 – Doesn't it go without saying that a larger debris flow is going to subject a larger area to potential damage?

529-530 – is it that check dams are not effective against the largest flows or that the check dams just aren't big enough to be effective?

531 – What results if you simulate large debris flows without the check dams present?

532 – "raised development level" to me implies an increase in elevation – maybe reword.

551 – Needs more than 3 data points for substantive support of this claim, I'd say. A lot of other factors could be encouraging or discouraging development.

566 – Could it also be that construction of check dams has been prioritized in areas with a lot of development?