Dear Authors

I read your paper with great interest and I think that the topic is perfectly within the scope of the journal. Your approach tackles a very important problem regarding the design of countermeasure work for flow processes characterized by higher uncertainties. You propose a very honest and straightforward procedure which is well described in the paper that can be applied in other contexts and also for landscape evolution analysis. Having said that, I have only some open questions regarding your work that should be addressed before publication.

Broad comments:

Debris flow or debris flood? Please address the terminology issue also in the light of this new paper Church, Michael; Jakob, Matthias (2020). What is a debris flood?. Water Resources Research, doi:10.1029/2020WR027144

L75: the Crucecita Alta deposit is described in bulk with other fans in paper (Cabrè 2020a). Figure 6 of that paper shows the inferred sediment concentrations during the event but, since we are focusing on this test site now, I would advise you to provide a paragraph discussing how you estimated the sediment concentrations for each surge. Moreover, in Cabrè 2020a it is stated that the maximum thickness of the deposits is 100 cm while your simulations of the viscous debris flow show larger thickness of deposits for Surge 1. Is the remoulding of the subsequent floods that reduces the deposit thickness in the upper part of the fan? Is it possible to add a map showing the deposit thickness inferred from the geomorphological survey? The model calibration of debris flows just on the basis of the impacted area instead of deposits can lead to errors. In case it is not possible, please insert two sentences in the text highlighting this potential problem.

L259: I completely agree with you that for modelling calibration is sometimes difficult or not robust to rely only on a automatic algorithms to select the best fit parameters; it is in fact usually best to incorporate the expert knowledge euristic to select the best fit parameters and call the algorithm a Decision Support System rather than a automatic calibration algorithm. To better explain this to the readers you can provide a figure with the cloud points of your optimization indexes to show how clustered or nonclustered your 100 simulations were, so as to also understand how arbitrary is your “five runs” threshold.
Specific comments:

L10: if a channel is wide enough

L50: citation missing (?)

Figure 1: it is not clear what do the black lines represent in part b. Can you show in a less stylized way where exactly the element at risk (roads and buildings) are located in fig b?

L386: to design mitigation works

L389: the broad flow typology

L398: are deviated and forced to deposit

L399: With surge 4 avulsion is present, inundating the southern portion of the fan

Figure 8: please explain why the deposition pattern in surge 4 is so rectified, it seems a little bit unrealistic - did you experience problems with DTM interpolation? Can you show the contour lines of your model DTM topography?

L435: buffer?

L441-446: please revise this paragraph as the concepts expressed are clear but their formulation is not so

L484: incide