Reply on Referee Comment 1: A modeling methodology to study the tributary-junction alluvial fan connectivity during a debris flow event

November 2021

We thank the reviewer for his time in commenting on our paper. In this document, we answer each individual comment. For clarity, the reviewer's comments are in black, while our answers are in blue.

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

The debris flood description provided by Church and Jakob (2020) is similar to our preprint description for inertial debris flows. Church and Jakob (2020) recognize that debris floods rely on the tractive forces of the water. This is the key difference with viscous debris flows described in our preprint, where the tractive forces of the mud dominate (water and fine sediment mixture). Church and Jakob (2020) mention that "debris floods represent water-driven flood flows with high bedload transport of gravel to boulder size material and significant damage potential." They also highlight that the term "hyperconcentrated flows" may lead to confusion when talking about debris floods. Even though a particular flood may be both a debris flood and a hyperconcentrated flow, the generalization of the term "hyperconcentrated flows" overlooks an essential distinction between suspension and bedload dominance in the sediment transport process. Therefore we will eliminate the term "hyperconcentrated flows".

Surges 3 and 4 are described as inertial debris flows in our work. However, we recognize that inertial debris flows have a broad significance, where many types of flows are lumped. The term debris flood for surges 3 and 4 is, therefore, more specific to depict what our model is able to simulate. Furthermore, our work shows that collisions and turbulent stresses dominate in surges 3 and 4, which result in erosion and sedimentation that modify the fan's topography. The sediment transport characteristics of debris floods described in Church and Jakob (2020), therefore, support the assumption in our paper that surges 3 and 4 can be simulated as a Newtonian flow together

with a sediment transport model. On the other hand, viscous debris flows (surges 1 and 2) are properly defined as debris flows.

In conclusion, we will add the debris flood classification to the description of surges 3 and 4 and a short description of what a debris flood is.

L75: the Crucecita Alta deposit is described in bulk with other fans in the paper (Cabré 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.

In our model, the sediment concentration is a function $C_V(t)$, where t is time, and it depends on the minimum and maximum plausible volumetric sediment concentrations. Zegers et al. (2020) present this function in detail. Specifically, in this article, we introduced a semi-automatic calibration algorithm (will be called Decision Support System in the future, see below). This algorithm evaluates the effect of different sediment concentrations and constrains the cases that reproduce the event flooded area and deposited volume. Therefore, sediment concentrations were not estimated directly but instead calibrated with our novel algorithm.

L75(continue): Moreover, in Cabré 2020a it is stated that the maximum thickness of the deposits is 100 cm while your simulations of the viscous debris flow show a larger thickness of deposits for Surge 1.

The resulting deposit thicknesses of the viscous debris flows (surges 1 and 2) are highly dependent on the Surface Detention parameter (SD). In Line 325, we analyze the effect of the SD parameter over the deposit thickness associated with Surge 2 but not for surge 1. We will add in this paragraph that Surge 1 resulted in similar behavior.

The problem with the SD parameter is that the final deposited volume is very sensitive to this parameter. In Fig. 4. of Zegers et al. (2020), it is shown that SD corresponds to the parameter with the highest sensitivity for alluvial fans that buffer the sediment discharge. Also, it is complicated to select an SD parameter value based on field data. We started with an SD parameter value of 1 m resulting in slightly thicker deposits. Since we have no clarity on the correct value for SD, our approach was to test lower values of SD (0.5m and 0.7m). Still, unfortunately, the flow spreads over the surface, losing any correlation with the affected areas. Finally, we decided to keep the SD parameter value equal to 1 m because this value could reproduce the event better than an SD value equal to 0.5 or 0.7 m. We tested a set of one hundred models for each SD parameter value with our novel algorithm to find the rheological parameters.

L75(continue): 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?

It is possible that remoulding of the subsequent floods could have reduced the deposit thickness, but, this is very difficult to affirm based on our field observations. However, as stated in the previous answer, the overestimation of the deposit thickness can be a spurious effect that is analyzed in line 325 of the preprint.

In Figure 1 we show the point measurements of the deposit thicknesses taken on the field for

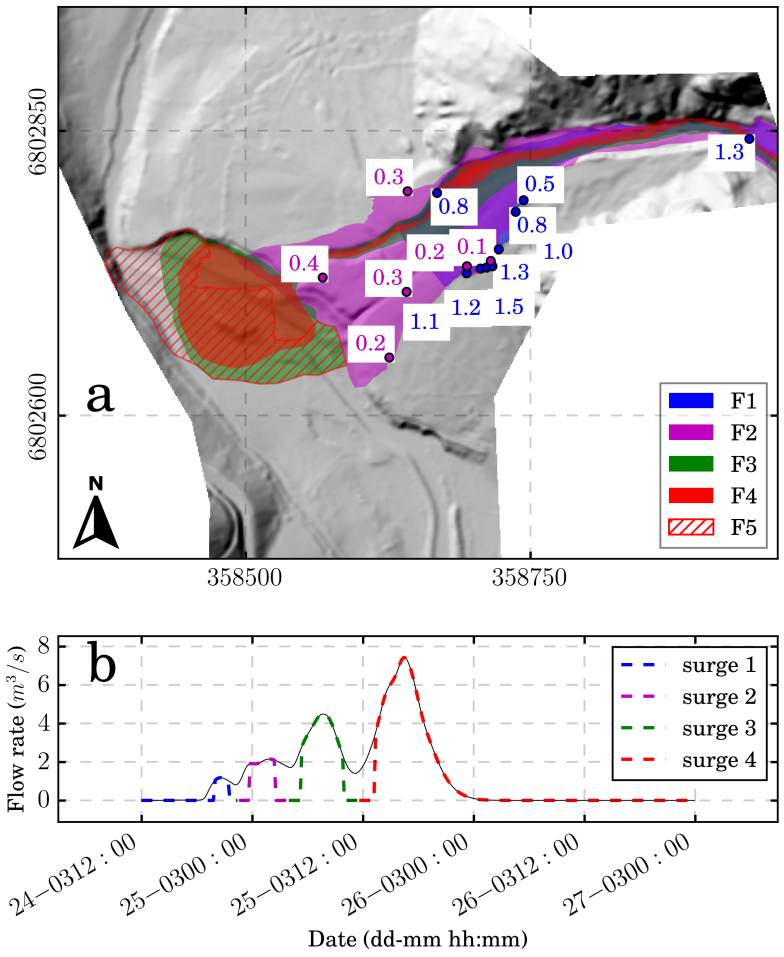

Figure 1: Point measurements of deposit thicknesses taken for F1 and F2 deposits.

F1 and F2 that we used to infer the mean deposit thickness.

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.

We based our methodology on the recent paper by Zegers et al. (2020). One of this study's main findings was that the calibration by flood area and deposited volume also constrained the flow velocity and flow height (Figure 5 in Zegers et al. (2020)). Therefore, in this study, we analyzed these two variables in a two-step procedure. First, we filtered by flood area indexes considering the correlation of the flooded pixels. Secondly, we analyzed the filtered cases by considering the resulting deposited volume and visual inspection, where we manually selected the best case out of the filtered runs.

Even though the flooded area filter returns a manageable amount of runs for visual inspection, we agree with the reviewer that including the volume filter directly to our Decision Support System (DSS) is a better approach. However, one must remember that the important goal of the DSS is to

find the best cases (for example, five out of one hundred) instead of finding specific thresholds for the volume and area filters. For instance, in order to obtain three runs for F1 (Fig. 3), a volume filter of ±60% was necessary. On the other hand, all cases in F2 filtered by area pass the volume filter of ±60% (Fig. 3). Therefore, it should be noted that this filter (±60%) is arbitrary, and it can change for every studied case.

We will add a short explanation of this to the document.

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.

We thank the reviewer for this helpful suggestion. We will change the algorithm's name to Decision Support System (DSS) since this name is closer to the actual capabilities of the algorithm. As the article and the reviewer highlight, this algorithm constrains the search for the rheological parameters that reproduce the event. However, the user has the final decision about which model is the best.

Thanks to this suggestion, we found an undesirable behavior in our Decision Support System (DSS). Even though the parameter sampling method distributed the rheological parameters within plausible value ranges, their combination could result in unrealistic values for τ and η. As a result, only 56 out of 100 cases have reasonable yield stress and dynamic viscosity values. We corrected this problem by adding a step where we check if τ and η have reasonable values. If they don't, new values of β_1 and β_2 and C_V^{max} are set until we obtain values of τ and η lower than 50000 $dyn\ cm^{-2}$ and 100000 Pa, respectively.

To be consistent with our population of 100 cases, we re-run our simulations with the corrected algorithm. In Figure 2, we present the distribution of the 100 runs for the area "True Positive" ($Area^{TP}$) vs. the area "False Positive" ($Area^{FP}$). These areas were defined in the preprint. In Fig 2.a, we present the old set of 100 runs where black stars correspond to values of τ and η higher than the maximum plausible values. Red stars correspond to the 56 cases with reasonable values for τ and η. Figure 2.b presents a new set of 100 runs where the algorithm was corrected as explained previously. From both distributions, we observe that the red cases in Fig 2.a have the same distribution as the cases in Fig 2.b. The only difference now is that instead of having 56 runs with plausible values, we have 100. Consequently, the runs that now pass the flooded area filters are ten instead of four for F1 and nine instead of four for F2.

In Figures 3.a and 3.b, we show maps of the areas considered as "Observed positive" ($Area_{OP}$) and "Observed negative" ($Area_{ON}$) for the facies F1 and F2, respectively. In addition, in Figures 3.c and 3.d, we present the distribution of $Area^{TP}$ vs. $Area^{FP}$ for surges 1 and 2 associated with F1 and F2, respectively. Our approach with the DSS is to find a manageable amount of runs for visual inspection that better reproduce the area and volume. Therefore, we looked for the runs that maximize $Area^{TP}$ and minimize $Area^{FP}$. For F1 (Fig. 3.c), for example, the best ten cases have an $Area^{TP} > 70\%\ Area_{OP}$ and an $Area^{FP} < 30\%\ Area_{OP}$. For F2 (Fig. 3.d), the best nine

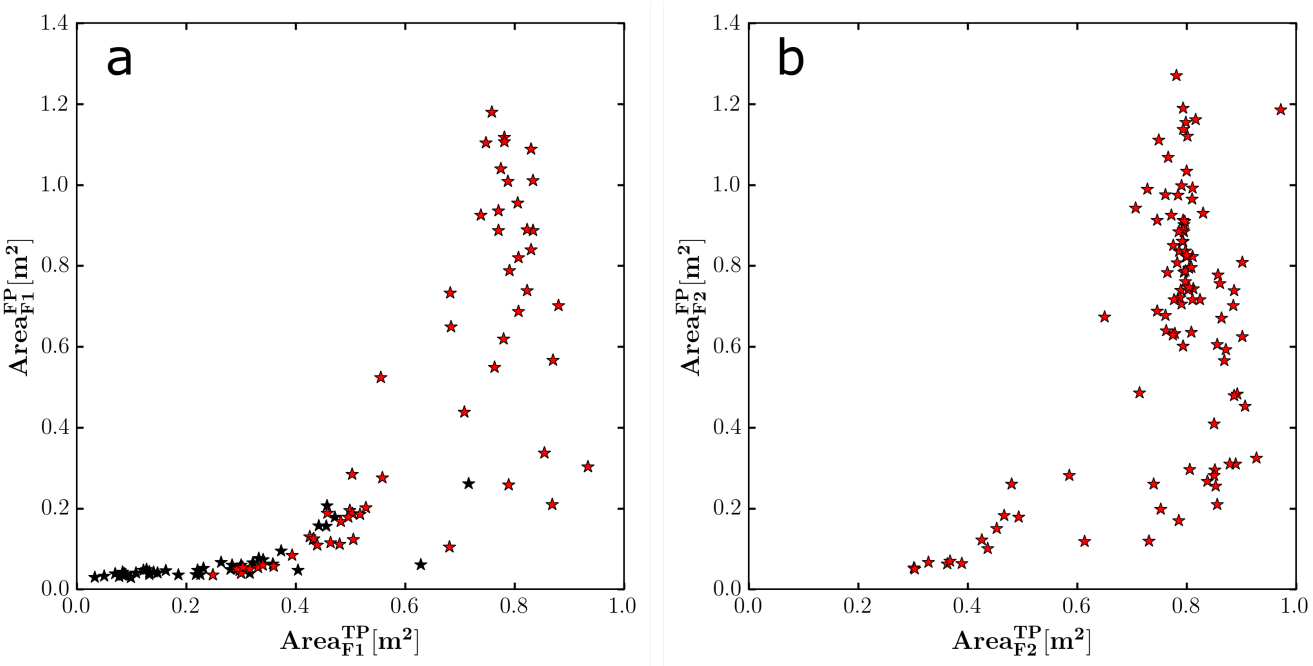

Figure 2: (a) the original set with τ and η values higher than the maximum plausible value. (b) the new set of runs with values of τ and η corrected to remain within a reasonable range of values.

cases have an $Area^{TP} > 90\%\ Area_{OP}$ and an $Area^{FP} < 10\%\ Area_{OP}$. In Fig. 3, the green hatched areas show the runs that pass the area filtering. It is important to note that the area thresholds for surge 1 and surge 2 are different due to their dissimilar distribution of $Area^{TP}$ vs. $Area^{FP}$. On the other hand, the cases that pass the volume filter are presented with blue stars (Fig. 3). Finally, **filtered runs by area and volume correspond to blue stars within the green hatched area.**

For F1, $Area^{TP}$ ranges from 0.3 to almost 1.0, whereas $Area^{FP}$ ranges from 0.05 to 1.4. For F2, $Area^{TP}$ ranges from 0.85 to almost 1.0, whereas $Area^{FP}$ ranges from 0.05 to 0.25.

About the data distribution, we found that, for F1, the points are spread, whereas for F2 are mainly clustered. The broader range of $Area^{TP}$ in F1 than in F2 is due to the higher C_V^{mean}. For F1, the inflow volume of water is less than for surge 2, but the volume of sediment is larger than for surge 2. As a result, F1 has higher C_V^{mean} giving the ability for this surge to stop before filling $Area_{OP}$. On the other hand, the more diluted nature of surge 2 makes it easier for this surge to flood $Area_{OP}$.

In the maps of Figure 3, we can see a crucial difference between F1 and F2 cases. In F1, the debris flows do not reach the river, so the area downstream of F1 is considered "observed negative". Contrary, since F2 is connected with the river, $Area_{ON}$ was selected only at both sides of the mapped deposit. This difference results in a broader range of $Area^{FP}$ for surge 1 than $Area^{FP}$ for surge 2.

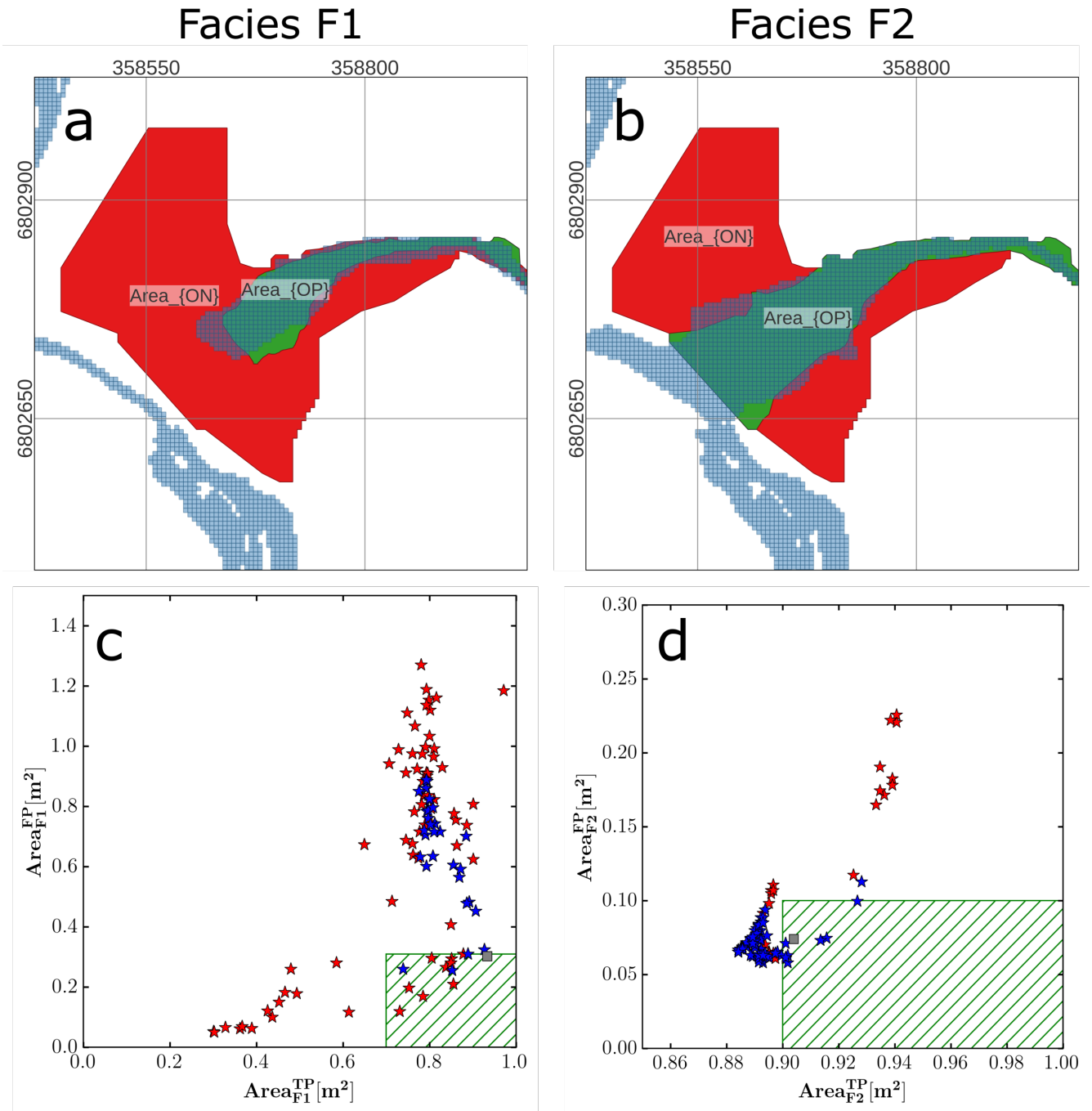

Figure 3: Correct vs. incorrect reproduced areas by one hundred runs tested by our Decision Support System.

Since the set of 100 runs is new, we added the original case used in our model to Figure 3.c and 3.d as a gray square. As seen here, the old selectedruns meet both the area and the volume filter. Therefore, it is not necessary to rerun the calibrated model. Moreover, the distribution of the results in Fig. 2 did not change due to the additional step in the DSS algorithm.

Specific Comments:

L10: if a channel is wide enough
We thank the reviewer's suggestions, we will replace it with: "if a channel is large enough".

L50: citation missing (?)
We apologize for the citation misspelling. It corresponds to Takahashi 2014 (book).

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?
We present a less stylized way in Figure 4.

L386: to design mitigation works
We thank the reviewer's suggestions, the word "test" will be replaced with "design".

L389: the broad flow typology
We thank the reviewer's suggestions, we will rephrase the line to "The same hydraulic works should be studied under the broad flow typology".

L398: are deviated and forced to deposit
We thank the reviewer's suggestions, we will rephrase the line to "are deviated and forced ~~to flow in the upstream direction of the river~~ to deposit due to the previous deposits".

L399: With surge 4 avulsion is present, inundating the southern portion of the fan
We thank the reviewer's suggestions, we will rephrase the line to "~~For surge 4, the channel overflows and inundates the fan.~~ 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?
In Figure 5, we present Figure 3 of the preprint with the contour lines every 2 m added. We did not experience any problems with DTM interpolation, but it should be noted that the model has a 5 $m/pixel$ resolution. FLO-2D has a grid of square cells, but these cells are also connected diagonally (Figure 6). Thus, the flow has eight possible flow direction. When the flow aligns with one of these directions, the solver tends to generate preferential flows following this alignment. As seen in Figure 8 in the preprint, these rectified patterns in surge 4 align in near 45°(south-west direction).

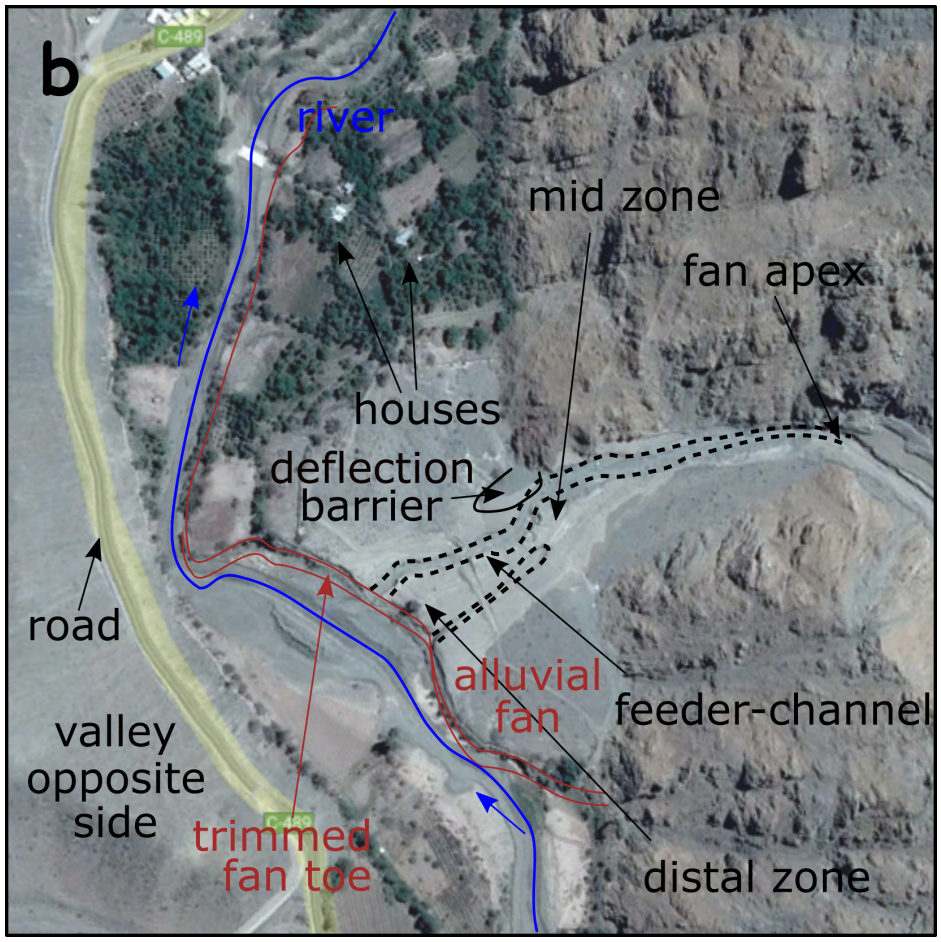

Figure 4: New figure with Google satellite imagery.

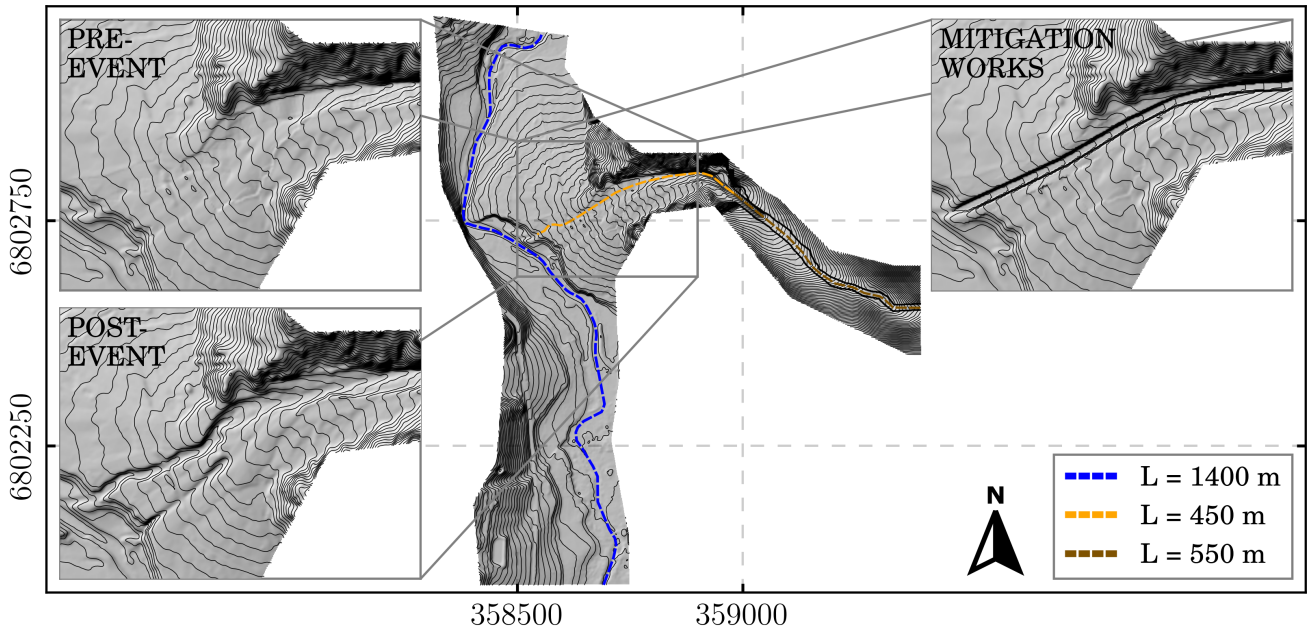

Figure 5: Contour lines every 2m of the topography.

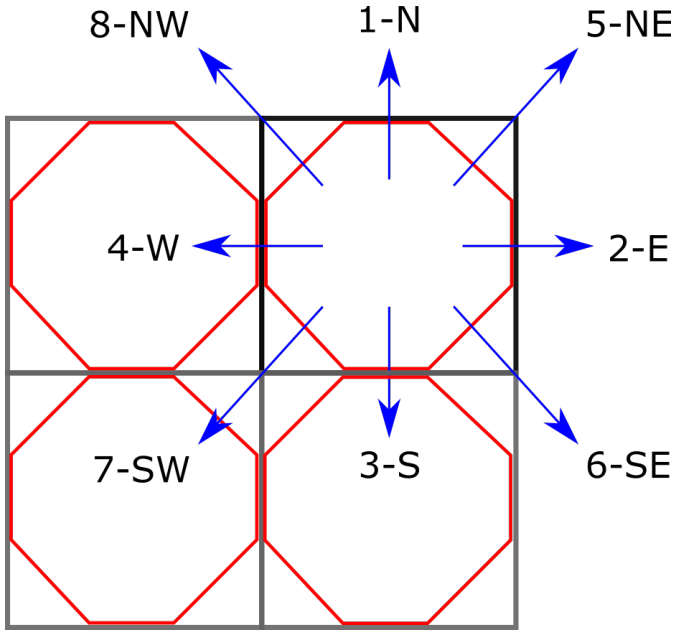

Figure 6: FLO-2D square cells and their eight flow connections between cells.

L435: buffer?
Fryirs et al. (2007) named the lateral impediment of sediment conveyance as a buffer, and the longitudinal impediment of sediment conveyance was named as a barrier. Since surge 1 could not reach the river, it is possible to say that the alluvial fan completely buffers this viscous debris flow.

Fryirs, K. A., Brierley, G. J., Preston, N. J., & Spencer, J. (2007). Catchment-scale (dis)connectivity in sediment flux in the upper Hunter catchment, New South Wales, Australia. Geomorphology, 84(3-4), 297-316.

L441-446: please revise this paragraph as the concepts expressed are clear but their formulation is not so
Original paragraph:
"We also studied how the Crucecita Alta alluvial fan would react to a channel. This approach has proven helpful because the channel (mitigation work) was expected to reduce affected areas on the alluvial fan and enhance the trunk river obstruction. However, the results in scenario 1 indicate that a channel reduces affected areas and avoids headward erosion. Thus, it reduces the amount of sediment that finally reaches the river reducing the river obstruction. In scenario 2, even though the affected areas are reduced, deposition occurs between the apex and the fan's mid zone and, consequently, a channel overflow. However, headward erosion is also avoided for scenario 2, and the new lobe at the fan toe reduces its size considerably."
Proposed new paragraph:
Our methodology is helpful to study hypothetical scenarios, such as a channel designed to mitigate affected areas but increasing the coupling status of the alluvial fan. The results of scenario 1 (Fig. 8.a to 8.d) show that viscous debris flows can reach the river since they remain confined to the channel. However, sediment deposition occurring during the flow reduces the channel transport capacity for following surges. Thus, when the inertial debris flow surges occur, we observe that the channel can convey surge 3 but not surge 4. Compared to the original 25M event, smaller erosion areas result in lower sediment loads, reducing the size of the new lobe at the fan's toe. The results of scenario 2 (Fig. 8.e and 8.f), which consider only the inertial debris flows, show that the channel avoids the headward erosion observed in the 25M event. Consequently, the amount of sediment deposited at the new lobe at the fan toe is reduced considerably compared to the original case. However, we expected that the presence of the channel would increase the structural connectivity.

L484: incide
We will change the word "incise" to "carve" in order to avoid confusion.