

## review 1:

The manuscript presents a study on key debris flow characteristics (velocity and volume) from three study sites. The flow characteristics are estimated using a predominantly seismic approach. The goal of the study is timely, relevant and matches the scope of the journal well. From my perspective, the study design is well suited and the gathered data is substantial, providing proper substance to pursue most of the discussed topics. That said, the manuscript also shows a few drawbacks that need to be addressed before it has reached a state that renders it eligible for publication. I believe all these points can easily be addressed with the data that is inherent to the study.

In the scope section (lines 40-42) the reader gets the impression that the approach would not require any site-specific calibration efforts. This is not true. Evidently, the authors record independent data (flow stage at all four sites, and velocity by a Doppler Radar sensor at one site) and use that (table 2, fig. 8) to convert the integrals of squared seismic velocity to volume (integral of flow stage throughout an event), the corresponding fit function is presented as Eq 1. So, there is a need for calibration.

Yes, we used a site-specific calibration based on complementary data to derive equation (1). However, the aim of this study is to propose a simple method based on few metrics of the seismic signal to obtain a rough estimation of debris-flow volumes. We tested equation (1) against independent debris-flow volumes collected at Illgraben, Switzerland, from 2015 to 2017. The results show that the method offers a good estimation for the volume on an independent test site without additional calibration. We will add the results of this test in the revised manuscript.

Now, it is a matter of discussion whether the authors have demonstrated that this fit function is a universal law to reach suitable results or not. My visual impression of the content of fig. 8 is: No, this is not the case. If one would fit the data of the three sites (coloured symbols) individually, the resulting regression coefficients would be quite different, and that difference could and should actually be tested quantitatively. Hence, when propagating the impact of these supposedly three different regression coefficients to the results shown in fig. 9, I would assume that we would see for each subset of dots quite different results of predicted volumes. In any way, to make things short here: if the authors wish to claim there is a universal law to relate seismic energy integrals to total flow volumes then they have to prove this hypothesis in a proper way.

The aim of our study is to find a simple method to provide an approximate volume estimation that can be implemented in a detection system. By no means we believe that, with such a limited dataset, we have found a universal law relating seismic energy to debris-flow volumes in any conditions. In particular, the higher range of the volume distribution of our dataset is poorly represented, and by using three different regression equations, one per sites, the results would not be statistically significant. In addition, differences in rheology and flow dynamics can have an impact on results. We will better clarify the aims of our study, as well as we will expand the discussion of our results using the validation dataset from the Illgraben mentioned above.

It remained unclear to me, which stations the authors used to calculate the integral of squared ground velocity, one sensor (which one) or all available? Was an average value used (and can the scatter be quantified), or have the amplitudes been scaled by distance to source?

The geophones used for the volume estimation are G4 at Gadoria, G2 at Lattenbach, and G2 at Cancia. This information is given in the text describing the test sites, we will clarify this point in the methods section 2.2 Magnitude estimation. At each test sites, we selected the geophone that recorded the larger dataset and we performed the volume estimation using data from that geophone. The geophones used for the volume estimation are marked with a yellow circle in Fig. 2-4 - we will use a more contrasting color for indicating these geophones in the revised version. Since the distance geophone - channel is little and rather similar at the three sites, no scaling based on the distance has been applied.

The authors raise the claim that they can deliver a metric like average flow velocity based on seismic sensors. This is only true for that stretch of the channel that lies between two sensors, some tens of metres, and thus close to an “extended point measurement” considering the total channel lengths under investigation. That information is kind of implicit but should be brought up explicitly in the abstract and other appropriate places, because seismic sensors can also be used to study the average velocity of a debris flow as it propagates down the channel (Walter et al., 2017), which is a quite different type of information.

We agree and will better clarify this point in the text.

Following the authors approach to study flow volume and velocity using seismic sensors, raises the question why this approach would really be superior to other, more classic measurements. Sure, the other measurements require infrastructure built to host the required sensors (Doppler systems and flow stage meters) while seismometers can be installed relatively easy adjacent to the channel – a point that the authors correctly elaborate on. But, at least for the calibration work to be able to relate seismic signals to volume, some independent measurements need to be gathered, as well, no? Thus, I suggest to authors spend a few words to frame this topic a bit: advantage of seismic approach in the light of efforts for calibration work.

Our ultimate goal is to test whether debris-flow volumes can be estimated reasonably well by means of few simple metrics derived from the seismic signal. The advantages of such an approach will be discussed more in detail in the revised version. We agree on the need of calibration work, and we will add a paragraph to explain this point.

Line 7, “methods was”, change to “methods were”

OK

Line 12 (and further cases), the terms magnitude and volume are used alternatingly. And it is not clear to me how especially the term magnitude is defined. As I read, volume is defined as the time integral of stage height. But what about magnitude?

We agree, we will use volume instead of magnitude throughout the manuscript.

Line 27-28, the study of Manconi on rockslides (and also the work of Perez-Guillen of snow avalanches) does not really match with the context of the introduction (and actually the scope of the manuscript as a whole). Either provide a more elaborated overview on more of the existing approaches to relate seismic signal properties to material volume or leave this part out. This issue links to another point, see my comment regarding lines 189-198, which describe such additional (but by far not all relevant) approaches to turn seismic metrics into volume and other kinetic process attributes.

We agree, we will expand a bit the introduction presenting an overview on the few other studies relating seismic signal properties to volume of mass movements (e.g., Lai et al., Le Roy et al.). This would be useful for discussing the pros and the limitations of our methods.

Line 33-34, this reads like magnitude is volume times velocity. Is that so? If yes, this should be mentioned explicitly and also it should be discussed to give more substance, I see there is a reference to Coviello et al. (2019), but a few more words would be really helpful, here. At a first glance the product has the unit  $m^4/s$ , does this make sense?

We will reword this sentence and clarify that the mentioned reference only provides a first insight on a possible method for debris-flow volume estimation that is developed in the present paper.

Line 37-38, this is a fair point and I fully agree. However, actually the results of this study show exactly this point: there is no universal “approach” to scale seismic signals to flow properties, without any site-specific calibration. I suggest this part should be revised to not raise the implication that this study solves this issue. I do not see the point that no other studies have yet presented a universal simple method. There are numerous other studies that have used seismic sensors to investigate debris flows and have come up with methods to reveal key flow properties, including those studies cited by the authors, and especially previous studies by the author team (see line 34).

As already answered above, a universal approach is not the goal of our work. We try to find a simple, practical method for the volume estimation of debris flows detected by geophone installed along the active channel, which could be implemented in the future in monitoring and – possibly – warning systems. We will delete the word “universal” to avoid misunderstandings.

Line 38, again I see some ambiguity in the wording, here. The presented study does not at all use seismic amplitude data only, to provide an estimate of flow velocity and volume: figures 5-7 and 8-9 basically show independent data, and the data from table 2 is used to fit a regression model, which ultimately allows relating seismic data to total volumes. So in essence, this study also uses additional data as many other studies did and which is logical because as the authors mention, the seismic signal properties depend on a lot of site specific parameters. Please revise this section to be clear. This includes especially line 40-42, which raises the claim to overcome site specific calibration needs.

Correct, for calibration and verification other data are needed, but for further use no additional data should be necessary. This will be clarified.

Fig.1, this figure is of very limited use, providing no relevant information, not even a scale bar. Either remove it or expand its content, for example by combining it with figs 2-4. Regarding these latter figures, just as a hint, I would double check if using Google Earth screen shots is in agreement with the CC-BY license of the journal. It would any way be better to provide proper topographic maps instead of the perspective views, unless they reveal content that a proper map would not be able to deliver.

We agree, Figure 1 will be removed. We will add an inset showing the catchment location to Fig. 2-4, which will be also improved. We will check if Google Earth images can be used under CC-BY license and change to topographic maps if necessary.

Line 52, “G1 and G2 ... marked with yellow circle ... 75 m”, these information bits do not add up. Yellow circles are around G2, G3 and G4, 75 m are between G2 and G3. Please clarify. In addition, I think it is not clear at all why the velocity was only estimated between two geophones when a nice linear array of four sensors is present that can be exploited. Imagine the increased depth of information and robustness if you would use four sensors, i.e., six possible combinations of velocity estimates! Why did you limit your study so drastically? Here would be an excellent chance to estimate the robustness of your velocity estimation approach, and also at the Cancia site (station pairs 1-2, 2-3, 1-3) this would be possible.

A mistake happened in line 52 - we used geophone G2 and G3 for the analysis as reported in the caption of Fig. 6. The point raised by the reviewer will be clarified, and the graphical quality of the figure will be enhanced to convey the message in a clear way. We fully understand the reviewer’s comments on the possibility to increase the number of reaches analysed in terms of velocity. While not all of combinations are possible due to the sufficient quality of the geophone signal, we will add flow velocity estimations at one additional reach per site in the revised version where the data is fully available. Also, a table summarizing the different results of flow velocity at each site will be added.

Line 55, “which reliable detects”, first correct wording to “reliably” and second, I feel more detail is needed, here. What gives rise to that reliability? Can you quantify that based on the referenced work, e.g., ratio of correct versus incorrect detections, or a confusion matrix? What is this “specially designed detection algorithm” and how is it related to the STA-LTA algorithm mentioned above? I know there are other articles about this system, which I actually really see as an asset to the field of seismic hazard detection, but a few lines of explaining text would be great in this context here, as well.

All the information about the detection algorithms employed at Gatria and Lattenbach is available in the referenced papers (Coviello et al., 2019 and Schimmel et al., 2018). Yes, the automatic detection is not the focus of this paper but we will provide some additional information on both systems that can be useful in the perspective of future works (i.e., integration of real-time velocity and volume calculations in an automatic detection system).

Line 62, “two stations for testing the warning system”, this section does not make sense. Do these two stations belong to the system, or is MAMODIS an extra device/setup? In the former case, how can the system be tested by the system, in the latter case, where is the system in the map shown in fig. 3? Please clarify.

MAMODIS is an extra setup consisting of one geophone and one infrasound sensor. This will be clarified.

Line 81, I think I understand there are two approaches to get flow velocity from two seismic stations. It is not so clear from the wording, that you actually applied these two methods independently. Can you please revise the text to make this obvious to the reader? I only got this information when I looked at the legend of fig. 5 and then trying to move myself backwards through the manuscript to find the indication of these two methods. I see you mentioned a reference for the amplitude modelling approach but a few lines of explaining text would be very helpful to understand the context without needing to search for the referenced article.

The reviewer is correct, the text was not clear enough. This paragraph will be rewritten as suggested. The use of cross-correlation of two seismic signals for the velocity estimation is

the main method analyzed in that work, the amplitude maximum values are only used for validation.

Line 84, “mean surge velocity”, you need to be specific here, this is only valid for that stretch of the channel/flow between the the two sensors, not the entire flow as it propagates down the channel.

Right, we agree on this point. We will add that the velocity values are only valid for the specific channel section.

Line 85, “manually analysed”, my first impression was that the study pursues an automatic detection and characterisation of debris flows. How does this match up? Can you clarify, ideally at the first introduction of the idea of an automatic system. More importantly, based on which criteria did you identify comparable peaks? Was that just a subjective eye-spotting approach? What would the uncertainties be that arise here?

The manual analysis is used for validating the results of application of the cross-correlation method. The “manual” velocity values are found by an eye-spotting method. We will clarify this point in the revised manuscript, and we will include an estimation of the related uncertainties.

Line 89, the phrasing of the window size definition is somewhat unclear to me. How is the “number of samples equal to the distance” defined? Number of samples relates to the temporal domain and distance to the spatial domain. The linking factor would be velocity – which is not known beforehand. Please clarify because it seems like this selection of the window size appears to be a very sensitive parameter.

For the cross-correlation analysis, we have to set up a number of samples to define the starting window size. After testing several settings, we decided to use a starting window size related to the distance of the two geophones in meters. This choice offers the best result for the cross-correlation and provides an objective method, based on one parameters (distance) only, to adapt the cross-correlation analysis at new sites. Number of samples equal distance means that a resolution from 1 m/s is possible which is a good starting value for describing turbulent flows. A “methods figure” for the velocity calculation with cross-correlation will be added to clarify the method steps, the used window sizes and overlaps.

Line 95, if cross correlation is performed twice, what are the two pairs for time series that are used? Or do you mean within a fixed time window you do a cross-correlation of amplitudes (actually you should rather call this envelopes, because you only have positive values) in a sliding sub-window? If so, why only twice and based on which sub-window size and overlap? This information is not really clear, and I suggest to simply add more detail, here.

More details will be added here and also the figure will clarify this. We tested several settings for window size and overlap. An overlap of the half sample numbers used for the window size offers the best results.

Line 101, for which stations does this energy estimate hold? Again, the source energy requires application of a bit more calculus and information about ground parameters, e.g as done by Le Roy et al. (2019, JGR). And thus, the information about which station or ideally which stations is essential. In addition to this, you mention to unit Joule, but to get from  $m^2/s^2$  to kg

m/s<sup>2</sup> there is a bit more necessary. In other words, just squaring the signal envelope will not give you seismic energy on a short track.

This is correct, we neglected site effects due to characteristics of the media and attenuation. The relatively little and uniform source-sensor distance at the different sites supports this choice. In the revised manuscript, we will clarify that we use the squared amplitudes as a proxy of the seismic energy to estimate the debris-flow volume and we will expand the discussion comparing our results with others such as Le Roy et al., 2019.

Line 103, this links to my above comment. “estimation of seismic energy” is misleading here, at best you get a rough proxy of seismic energy following  $v^2 \sim E$ , but you need to estimate the scaling factor that turns this relation to a real function to estimate the energy from seismic amplitude values. Again, my suggestion is to either reword the text (relax the tough claim on energy estimate) or follow a similar approach as Le Roy et al. did.

We agree, we will reword this clarifying that we use a proxy of the seismic energy. We will also add that neglecting some site-dependent parameters is acceptable when source-sensor distances are little and similar but this simplification can represent a possible source of uncertainty in the volume estimations.

Line 107, can you really justify that especially in the near field, where it is far from easy to understand the wave field, it is legitimate to ignore any attenuation of the signal? If not, consider reworking the text to be less “confident” and glossing over this topic.

We will reword this. Actually, given the similar geomorphology (glacial-fluvial deposits) and source-sensor distances (from 10 to 20 m) at all sites, we assume that the attenuation level is similar. This simplification of course produces some uncertainty in the volume estimations but a simple approach like that have some practical advantages. In the revised manuscript, we will expand the discussion on that point.

Line 111-114, you can significantly shorten/consolidate this part. The first sentence repeats things we know from previous sections, The second sentence would make more sense in the second paragraph.

We agree, we will shorten the first sentence and move the second one as suggested.

Line 117, Defining peak discharge as all periods above a threshold of 3.5 m does not make sense. Peak discharge is the one value of maximum discharge, not several values. I think you mean local maxima in the amplitude time series, right?

The reviewer is right, this was an error. We will amend the sentence as indicated.

Line 121, this paragraph should be connected to the one above. The same holds for the description of the third study site.

OK

Line 123, be specific and replace “several” by the actual number, ideally illustrated also in the figure, for example by small numbers denoting the selected surges.

Ok, the number of surges will be indicated.



Line 124-127, can you please be more specific and elaborate regarding these results (range of values, number of surges, signal-to-noise ratio, range of cross-correlation values, and so on)? In the methods you mention a lot of things that you did but here we see only a very shortend presentation of the results of these methods. Specifically, the results of the two velocity estimate approaches (cross-correlation and amplitude modelling) should be presented in a more elaborate form. Also, avoid interpretations of your results already in this chapter but move them to the discussion section.

Additional available results will be added. We will add a new table summarizing the different results of flow velocity and the additional information at each site.

More to the above point, I am missing any presentation of the Doppler velocimeter results as well as of the other non-seismic instruments you use to calibrate the seismic signals. All we get is the condensed version in table 2.

We will clarify that Doppler velocimeter data are not available for debris flows analyzed in this study.

Table 2, where do these numbers come from? From the gauge measurement devices? Please specify and if from independent measurements, these results deserve a bit more content than just their appearance in the table.

Different methods are used at the different test sites to measure the debris-flow volumes (e.g., topographic surveys, stage sensor measurements integrated in time, etc.). We will add some information in the text and a column in Table 2 to clarify how the independent volumes were calculated.

Line 138, a) please also give uncertainties on the fit coefficients and b) – related to my general comments – these fits should be performed also for each site individually so that the reader can judge how justified a global fit approach is.

We understand the reviewer's request, and uncertainties on the fitting coefficients will be provided. We are a bit skeptical about the possibility to provide a linear regression for each site individually. The dataset is limited and three equations would probably not be statistically significant as the higher part of the volume distribution of our dataset is poorly described (only two debris flows with volumes >40,000 cubic meters recorded at Lattenbach). In the revised paper, we will enlarge the dataset using published data from Illgraben (Switzerland) for results validation.

Figure 9, what are the implications of the 20 % error range with respect to the point that only the blue dots fall roughly into it while the majority of the other coloured dots do not? How much % scatter would you need to catch all dots? Or 95 % of the dots? What do the +/- 2 sigma lines depict exactly and how does this metric relate to the data you show?

We agree, the 20% range lines are confusing and will be removed, defining another % scatter will also make not much sense. The idea was to show that the error estimate does not increase with volume, but we will mention this more clearly in the text. The +/- 2 sigma lines represent the confidence interval of the distribution.

Line 143, add at the end of the sentence something like “between two closely spaced seismic sensors”.

OK

Line 145-148, this is a very broad and arm-waiving statement with little crisp information. Either include more specific (and thus justified) content or leave it out.

The reviewer is right, the statement will be removed.

Line 154-155, I think you can and should be more specific here. You can for example quantify the ratio of channel distance to station distance (20/90) as a metric to better define the term “significant difference”.

Good suggestion we will adopt the ratio indicated by the reviewer

Line 156-160, Ideally, you would test explicitly by signal aggregation and inspection of the impact of different sampling frequencies on the results. This could be easily done and I encourage the authors to do so, in order to be able to replace some of the “may”s by justifiable hard results.

We agree, but this is not doable for all sites since we have to work with the available data. For instance, at Cancia only 10s maximum amplitude values are recorded. We can perform a synthetic test on one single event recorded in another site at a higher frequency rate, we will consider to include the results of such a test in the revised manuscript.

Line 165-174, this section is not really helpful in the discussion. The information given there is material I would rather expect in the introduction part, giving an overview of possibilities to measure debris flow height and/or velocity and therefore in the end justifying the usefulness of seismic sensors. Here, you have little results to raise a discussion about other potential approaches. The discussion should be based on your study’s findings.

The reviewer is right, and we will remove this section from the discussion to move it to the introduction.

Line 176-188, I would welcome also a bit more discussion on the actual downsides of the seismic approach. It is good to underline its strengths, but there are also obvious weaknesses that deserve a discussion.

We agree, a discussion on the disadvantages of the seismic approach will be added.

Line 178, that “variance” is indeed due to the multitude of site specific parameters, parameters that must and can be accounted for by a calibration of the seismic data.

We agree, we have to clarify this. On one hand, the site-specific parameters affect the seismic amplitudes recorded, on the other hand the process type and flow regime also has a large influence on the seismic signal. We will discuss this point properly in the revised manuscript, better presenting the uncertainty of the methods.

Line 189-190, can you explain how/why the velocity would affect the frequency spectrum? This does not seem intuitive for me.

Thanks, this is a mistake. As described in Lai et al. and other authors, the seismic amplitudes and the PSD are influenced by bolder snout, grain size and flow velocity. We will reword as



follows: “Studies of different events also showed a large dependency of the seismic amplitudes and their PSD on solid particle characteristics and flow velocity”.

The conclusion is a weak one. It merely repeats what has been discussed before, rather than putting the findings into a wider context. Can you reach out a bit more and touch this wider impact? What is this study relevant for? What are the great assets? Which fundamental research gap/questions has been tackled? What is possible to engage with, now that the technique is there to seismically estimate important debris flow parameters?

We agree, we will expand and revise the conclusion and try to handle some of the relevant questions raised by the reviewer.

## review 2:

In their submission, Schimmel et al. test relationships (some of them empirical), which compare metrics of seismic measurements near a torrent to debris flow velocities and volumes. Similar calculations have been made in the past, but this study uses data from different torrents, which provides valuable insights for potential alarm and monitoring systems.

The scope of the study falls into seismic monitoring of Alpine mass movements, which is currently an active research field. So I expect this study to be met with interest within the journal's readership. The study is more on the technical side, which is OK given the journal scope, although I suggest some more discussion in terms of physical mechanisms as I outline below. Moreover, the details of the documented calculations are unclear and should be rewritten, especially since they constitute the paper core. The manuscript is concise, structured and easy to follow. However, the English contains numerous grammar mistakes and has to be revised before the paper can be published. Overall I recommend major revisions.

Fabian Walter.

## MAIN COMMENTS

The velocity calculations are not well described. Which three different sliding windows do the authors refer to? What is the relation between minimum and maximum amplitudes? A ratio? How can the number of samples be equal to some distance? Which distance? Distance in which unit? What is a "significant signal shape"? In its current state, it is not possible for a reader to use the explanations to reproduce the calculations.

The reviewer is right, the explanations provided are too succinct. This section will be expanded to present a clear flow chart on the different steps adopted, and the motivations for the window size used will be argued with numeric examples as well as with a figure to illustrate graphically the rationale.

I may have missed this, but how are the ground-truth debris flow volumes calculated, which the seismically-derived values are compared to? I was surprised that the authors do not discuss Schimmel et al. (2018), who use seismic and infrasound data to calculate discharge and estimate debris flow volumes. Is the current technique an improvement compared to this previously suggested one?

The volumes are calculated by means of different techniques at the three test sites (e.g., topographic surveys, stage sensor measurements integrated in time), flow height measurements or 2-D laser scans and velocity measurements. We will add some information on that point in the text and we will add one column with the reference (i.e., the paper/report used as source of volume data). In the discussion, we will add a comparison of volume estimations presented here with results from Schimmel et al. (2018) only based on infrasound data.

Except for a small part of the discussion, the authors give no explanations on the physics behind debris flow seismicity. The cited papers by Lai et al. (2018) and Farin et al. (2019)

make specific predictions between seismic signature and debris flow velocities, grain size distributions and other parameters. Even if the authors do not want to dive into details, they should use these theoretical assertions to offer explanations for their observed volume scaling.

We will add some additional discussion of our experimental results in comparison with the theoretical predictions proposed by the mentioned papers.

In several parts of the manuscript, the authors refer to the turbulent flow front. They need to provide evidence that their flow fronts were indeed turbulent and that this explains their observed signals. Some video or still footage could serve this purpose. Alternatively, I would expect that boulders in the flow front cause a distinct seismic signature compared to the flow tail. In a recent paper (Zhen et al., 2020, in GRL), we showed how the flow front's seismic signature is dominated by ground impacts of the largest boulders.

We agree, we will use the video footage of a few selected events to support the interpretation of the effect of flow turbulence. Of course, the effect of coarse front vs liquid tail on the seismic signal is an important point, we will expand the discussion on that also considering recent the paper by Zhen et al.

Finally, Figures 8 and 9 should include error bars or at least some short discussion on uncertainties should be offered.

The reviewer's suggestion is highly valuable, and we will add the error bars in the revised version of these figures.

#### OTHER COMMENTS

Line 18: "feasible" should be deleted, as it is implied by "effective".

ok

Lines 24ff: What are the physical concepts these velocity estimates?

In several instances, the authors use the word "magnitude". If this is synonymous with "volume", then use the latter, only.

Volume will be used throughout the ms.

Lines 35ff: Here it seems that the authors argue that mass could be estimated with the Coviello et al. (2019) approach, but volume is poorly constrained. The difference between the two is the factor density. Why is this so poorly constrained?

Yes, the mass could be estimated with the Coviello et al. (2019) approach but the volume is poorly constrained because of the lack of the variability of sediment concentration, which is not measured. We will reword these lines providing more information on the uncertainties of the methods.

Lines 40ff and elsewhere: Avoid 1-sentence paragraphs.

Ok

2 Methods: It would be interesting to see rough numbers of debris flows per year for the different sites.

Ok, numbers can be added.

Line 113: this peak discharge seems rather high.

That's the data I got from BOKU.

Line 126: Velocity measurements around 2500 s in Figure 6 do not seem “consistent” as asserted in the text.

I think they are in an acceptable range.

Lines 142-143: “can be an useful tool to analyze the flow behavior” This statement is trivial.

Ok

Line 148: “permitting to avoid wrong correlation results” unclear

This can avoid wrong correlation results since the possibility to have significant signal patterns for the cross-correlation increases with the window length. We will reword this.

Line 152: It is not clear how longer distances offer better resolution (resolution should be lower ...?).

Ok, wrong wording, will be rewritten.

Line 155: “so that the cross-correlation offers useful results” You should be more specific here.

Like a correlation coefficient above 0.8 - will be specified.

Line 159: “determine problems for the cross-correlation analysis” is unclear.

Correlation of wrong signal patterns. Will be clarified.

Line 160: “exaggerated averaging” is unclear.

averaging of surges

Line 164: I do not understand how the authors arrive at granularity here. This should be explained.

Flow regimes have been characterized using the video of the events. This will be explained in the text.

Line 169 and elsewhere: Is “process” synonymous with “debris flow”? If so, only use the latter.

Yes it is, and we will use the latter only as suggested.

Lines 166-168: This sentence needs a reference.

Line 174: “velocity measured by the radar is often lower ...” needs a reference.

We understand the reviewer’s request, and we will add a reference as requested

Lines 184-185: I am not sure that the volume estimate would always come in too late. It should be OK if the measurements were made high up in the catchment.

Ok, depending on the site and application. We will discuss, that an installation in the upper catchment might can offer a volume estimation in time, but there are high uncertainties, while installations in the lower part offers more accurate results, but come late.

Line 193 and elsewhere: “sediment concentration” Do you mean “grain size distribution”?

Yes, actually we did, it was a mistake. Thanks for highlighting this.

Line 193: “calculation of the magnitude” of what?

We will use “calculation of debris flow volume”

Lines 200-201: “among the different methods deployed and the different catchments” be more specific.

This sentence will be rewritten.

Lines 203-204: “but still further research on different ...” this is an unnecessarily generic statement. Why exactly is more research needed?

The reviewer is right, and this statement will be removed.

Lines 206-209: I suggest discussing Zhen et al. (2020) in the context of this sentence earlier in the manuscript.

Good suggestion, we will anticipate the reference to this paper in the introduction of the revised ms.

## FIGURES

Captions of Figures 5, 6 and 7: Rewrite so that site name appears in the first sentence of each caption and so that it is clear which “two geophones” are meant.

Ok

### **review 3:**

This is a well-written and clear presentation of seismic signal analysis for debris flows. In my opinion the paper is very close to publication as is. I especially liked the discussion, especially the paragraphs concerning volume estimation. The conclusions are straightforward and not overly “hyped”.

I have only one request. I found myself studying the debris flow hydrographs, because I am especially interested in the surge behavior as a function of time. I wish I could see the elevation profiles of the three sites, with the location of the gauging stations. The figures containing the torrent overview are based on google maps and contain a 100m scale, but I am interested in the elevation (slope) and distance travelled. I would very much appreciate three distance/elevation plots of the torrents. I would like to know if there are strong slope changes that would generate strong signals, etc. and, of course, where.

**This is not in the scope of this study. We only analyze the data in the section where the geophones are installed. We might add the information of the elevation difference in that part.**