I read the reply letter and the revised manuscript. I shall express that I am quite a bit disappointed about both documents. Evidently, the manuscript improved in quality and clarity in places. However, there still remain several weaknesses that were not changed, despite the notion in the reply letter that the authors would have changed them. This is quite bit from fair and transparent science, I think. I have the following general points to raise, followed by a list of more detailed comments/suggestions.

It was most frustrating to see that the authors promised in their reply letter to change a series of flaws of the previous MS version. However, when I read the new version I saw many of these changes not addressed at all. This concerns my previous comments (counted as paragraphs in the review letter from the start): 4, 5, 8, 13, 15, 18, 32, 36, 39. In detail, the not really useful fig. 1 is still in the MS, the Google Maps images are still in the MS without any notion if they are in agreement with a CC-BY license, fit parameter uncertainties are still not presented, the term magnitude is still used without definition and instead of volume, and so on. I encourage the authors to have a look at these unaddressed points of my previous review.

In the previous version, Figure 1 was kept because we added data from the Illgraben basin, and thus we thought that a general map showing the location of all study basins could be useful. In this revised version, we have removed Figure 1 as suggested by the reviewer. We have included inset maps showing the basin locations in the new figure presenting the three main study sites. We used “open street map” to draw this new figure, so we think we are now fully in agreement with a CC-BY license.

Overall, and partly as a result of the unaccounted change requests, the MS is still not well organised and leaves several statements unsupported (see below and the detailed comments). If the goal is to have a rough estimator of flow volume from integrated seismic energy, then this needs to be made clear in the text, including the abstract. Further, if this estimator consists of a linear regression model, then the uncertainties of the prediction must be mentioned and added to the test data set (Illgraben), beyond the plot in fig. 10, for example as RMS or relative error for each of the test events. If the regression model shows such strong deviations from the data structure of the three individual sites (fig. 9) then this point must be addressed, ideally by adding regression models for all sites individually, before the claim can be made that one universal model is a valid decision.

The previous abstract already included two sentences presenting the goal of the research, which is to provide a first order estimate of the total volume of debris flows based on seismic amplitude (squared), i.e., a proxy of the seismic energy in the near field. However, we have reorganized the abstract to make this message clearer, also adding a final sentence about the uncertainties of the methods used.

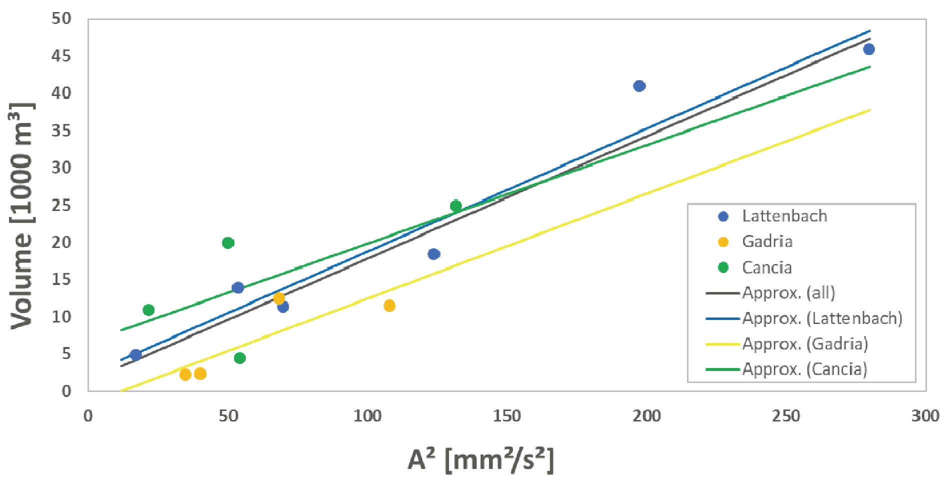
As suggested, in order to present more information about the performance of debris flow volumes prediction, in the discussion we have added a new table (Table 3) providing the relative errors of the prediction for Illgraben test dataset. Table 3 also shows how - in the near field - the impact of the different sensor-source distances on the volume estimation is quite small. More details on this specific point are given below in response to another comment by the reviewer.

Event date	Observed	Predicted eq. 1	Error eq. 1	Predicted scaled A^2	Error scaled A^2	notes
22.07.15	8700	7866	10%	8167	6%	
10.08.15	6100	24208	-297%	24667	-304%	high flow velocity

14.08.15	25000	11907	52%	12247	51%	
15.08.15	2000	7503	-275%	7800	-290%	smallest event
12.07.16	10000	13148	-31%	13500	-35%	
12.07.16	60000	10933	82%	11263	81%	liquid front, viscous tail
22.07.16	11000	16414	-49%	16798	-53%	
09.08.16	9000	7739	14%	8038	11%	
29.05.17	70000	57544	18%	58324	17%	
04.06.17	24000	20344	15%	20765	13%	
14.06.17	33000	27498	17%	27988	15%	

Following the suggestion by the reviewer, we have carried out regression analysis for each individual dataset, i.e., for each study basin (see table and figure below). Considering the relatively small size of each dataset, we do not think that including such results in the paper would add solid and valuable information, also because for the Gadria the regression turns out not to be statistically significant, as a consequence of the limited range in debris flow magnitude for which seismic data are available. In contrast, in the other basins the relationships are statistically significant and quite similar to each other. Therefore, we believe that presenting the regression analysis conducted on the multiple-basins dataset is most valuable and solid.

	R^2	p-Value
Lumped dataset	0.84	0.0006
Lattenbach	0.95	0.0213
Gadria	0.70	0.0839
Cancia	0.46	0.0456



The ambiguity of magnitude versus volume is important. Yet it remains unresolved in the MS. Worse, the term magnitude is not defined (l. 23) but then used in a statement that its estimation is still an open problem (l. 26-27). How can we judge this without even knowing what debris flow magnitude is supposed to mean? This issue needs to be fixed. In the conclusions, the term magnitude appears again.

In the literature about debris flows, the terms “magnitude” and “volume” are often used interchangeably. However, to avoid any possible misunderstanding the term “magnitude” has been modified into “volume” throughout the manuscript.

The discussion section is still weak and misses obvious points that need to be addressed. The validation approach for the cross correlation based velocity estimates, i.e. manually determining matching amplitude peaks, is certainly not ideal, but in the absence of other independent data it may be the only solution. But, this needs to be mentioned (in the methods) and discussed (with all its weaknesses and uncertainties). Also, the discussion should make use of findings of other authors regarding typical debris flow velocities. Are the values presented in this study within the range of other findings?

In order to strengthen the discussion about the results obtained by the cross-correlation method, we have added the following sentences (lines 175-183): “The validation approach for the velocity estimates, i.e. manually determining matching amplitude peaks at the two stations, is also affected by uncertainty. In the Gadria this is particularly evident for the tail of the debris flow of 08 June 2015 (Figure 4c, from t = 2400 s) during which roll waves propagate and produce multiple peaks, one following the other. The uncertainty in the manual velocity calculation was also observed in previous analysis based on data gathered with a pair of flow stage sensors and led to the approximation of lumping multiple waves into one single surge for the subsequent volume estimation (Coviello et al., 2021). The velocity estimates of surges lacking multiple peaks (i.e., from t = 200 s to 2400 s in Figure 4c) are consistent with those performed with the flow stage sensors located downstream from G3 (Figure 1b). Indeed, they are slightly higher (i.e., differences from 0.3 to 1.7 m/s) than those calculated with the flow stage data on a milder sloping channel reach (Coviello et al., 2021).”

Also, in the discussion we have added a comparison with results stemming from the use of debris flow radar at Lattenbach: "When velocity data measured by the debris flow radar in the Lattenbach (unpublished data) are compared against values calculated from the geophones there installed, very similar results can be observed. In fact, the maximum velocity measured by the debris flow radar for the event on 30 July 2017 is 10.0 m/s, while the maximum value calculated from the geophone data is 9.0 m/s. The mean velocity of the whole event is 1.8 m/s based on the debris flow radar, and 1.9 m/s for the presented method based on the geophone data."

The authors should comment on the fact that seismic sensors do not necessarily record just the debris flow signal but many others, as well. Particularly, rain and wind can cause quite strong signals in a frequency range that is definitely overlapping with the debris flow signal. And since debris flows are mainly triggered by rain storms, it is likely that these additional sources contaminate the record. Has this been accounted for? If so how? If not, at least the relative effect of these mechanisms must be discussed.

The reviewer is correct, other seismic sources can produce ground vibration but their intensity is low. To clarify this pint, we added the following sentences at lines 131-134: “Also intense rainfall and wind can produce ground vibration that geophones detect. However, seismic signals recorded by sensors installed at small distance from the channel (from 15 to 25 meters, in our case) are dominated by in-channel processes. This is particularly true in our study sites, which are located in lower reaches of the main channels where the debris-flow surges are well formed and characterized by velocities of several meters per seconds and flow depth in the order of meters”. The seismic effects of rainfall and wind in the upper Gadria basin have been recently investigated through a master thesis (Ioratti, 2022), and we are thus confident in neglecting their contribution to the seismic signal recorded at the monitoring stations considered in this paper.

Event onset and duration is another unaddressed topic. The severity arises because the seismic energy is derived from the signal integral, so overly long (or short) event

definitions will directly affect that energy estimate and thus affect the linear regression. How first of all have the onset and end been defined, second how have these definitions been implemented in the software that runs on site (or at least in the presented analysis), third what are the uncertainty estimates arising from this issue of missing the correct onsets and duration?

The reviewer is correct in highlighting how the duration of the event is a crucial information for the prediction of debris flow volume. However, in this study we have used a manually determined duration to establish relationships between seismic amplitudes and volumes for the available monitored events. Indeed, a sentence (line 238) was already present in the discussion to stress the larger uncertainties arising in the case the volume prediction has to be undertaken based on an automatic determination of debris flow durations.

The authors claim not to present a universal scaling model for debris flows. Yet, this is what the results boil down to: figure 9. And, further, this is what the application of the regression model to the Illgraben data actually is: using one generic model to predict event properties at another site. I understand that the goal is to have a simple tool to convert integrated seismic energy to debris flow volume. However, this is not what the text in its current style of writing implies.

Our study builds on solid physical principles, but it remains anchored to empirical relationships obtained in few – but of extremely high value – study sites. Therefore, we are humble enough not to claim that we can propose a universal scaling model, but “just” a statistical, data-driven relationship which appears to have the capability to provide reliable and thus useful estimation of debris flow volumes for practical applications.

In any way, another severe issue is that the authors still adhere to the regression of a lumped data set from three individual sites. I agree that the individual data sets are sparse. But also when I look at fig. 9, the yellow dots to not at all match up with the regression line, which is dominated by the blue lines. Also, to a certain extent, the green dots would result in a quite different regression line. In summary, the results shown in fig. 9 do not let me agree with the statement that all three sites can be lumped together. I suggest to at least add regression models for all individual sites and discuss how they differ from the lumped data set result, concluding with a statement why the authors think the lumped regression model remains a valid model.

Further, the data does not seem to be normally distributed (which would be reasonable for such stochastic processes). Without properly transforming the data to become close-to-normally distributed, the regression analysis is not in agreement with the basic assumptions of this approach.

See the above reply on the limited value of adding relationships for each single site, especially because the purpose of this paper is to obtain a statistical model with the potential to be utilized in other sites (as the case of the Illgraben shows).

Regarding comment about the non-normality of data, regression analysis – performed by the least squares approach - do not require data to be normally distributed, as described in all statistics manuals.

Finally, it is a pity that the authors wish to avoid accounting for the distance to source and ground property effects. This restricts their method severely to always the same distance to channel, glacial till deposits and similar channel properties (gradient, roughness, width, etc.) and stands in contrast with their stated goal to provide a simple but flexible estimator

of debris flow volume. Simple models to scale a seismic source amplitude to a given distance exist and could be readily added, for example: A(d) = A_0 exp(-pi f d / q v), all parameters that can be easily measured or simply assumed equal, but leaving the flexibility to apply the approach more adequately to other sites. I do not insist on implementing this model though. I just wanted to raise the possibility and express how easy this could be added.

We declare from the beginning that our methods is intended to provide a volume estimation based on seismic data gathered in the near field, i.e. with geophone stations located along the channel. For such a volume estimation, small differences in the distance sensor-channel are negligible compared to uncertainties descending from the variability of flow properties. Indeed, the distances sensor-channel for the different sites are 15 m at Lattenbach, 23 m at Gadria, 25 m at Cancia and 15 m at Illgraben. In any case, we followed the reviewer's suggestion and we explored the effect of amplitude attenuation on our dataset. We applied an empirical relation such as the power law suggested by the reviewer to model the decay of the seismic amplitude with distance. We used a value of quality factor of Q = 20 suggested as a reasonable approximation for the relatively high frequencies and shallow depths of interest (Tsai et al., 2012) and a reference value for group velocity of 1300 (Coviello et al., 2019). An additional approximation was needed to apply the formula using aggregated values of amplitude recorded with different sampling rate and recording frequency (Table 1), given that we do not dispose of the complete power spectra. We assumed that the coarse fronts of the different surges are the source dominating the seismic signal during the respective time window. This assumption is consistent with our approach of calculating the mean velocity of each surge by means of the cross-correlation technique, which needs to split the entire debris-flow signal. We tested values ranging from 10 to 20 Hz, typical of the main frequency of debris flows. Higher differences in the scaled amplitude are obtained with the lower frequency value (f = 20 Hz) so we used this latter value in the calculation to maximize the uncertainties. Finally, we recalculated the debris-flow volumes using a linear regression equation similar to eq. (1) but based on the square of the scaled amplitudes (Figure 8). Results clearly show how the differences in the calculated volumes with the non-scaled and the scaled amplitude equations are negligible (Table 3). In the discussion (lines 250-266), we have added text presenting the test we performed on the impact of the scaled amplitude on the volume estimation including new Figure 8.

Specific comments

l. 3, find a useful link between sentences to motivate the switch to the seismic approach after introducing debris flows as a hazard.

We have added the following sentence at line 22: "Seismic-based monitoring and warning systems have become increasingly applied worldwide to mitigate risks associated to debris flow processes"

l. 4, briefly add how the velocity can me constrained seismically. This is important, here.

We have modified the abstract to make this point clearer.

l.6, add here explicitly that you can only aim at rough estimates. Currently, the sentence raises the expectations that you can fully constrain volume and velocity, which is arguably not the case.

We have modified the abstract, we think this point is now very explicit.

l. 8, the Illgraben site and its purpose for testing the model should be mentioned here, as well.

We have modified the abstract, now we mention the test on the Illgraben data.

l. 32, "local and duration magnitude", is there a word missing? Please clarify.

Thanks, one "magnitude" was missing, now the sentence reads "…based on the ratio between local magnitude and duration magnitude".

l. 40, this does not make sense to me. How would a flow volume estimate be possible through velocity measurements? Please clarify.

We clarified this point revising the sentence as follows: "Using such scaling relationships, the estimation of the flowing mass is possible based on the seismic energy detected by a geophone and on the information about the flow velocity". Before this latter sentence, we added the following lines to to clarify that we are talking about scaling relationship between the seismic energy and flow characteristics: "Recently, Andrade et al. (2022) observed a linear correlation between the seismic amplitudes and the discharge rates of lahars at Cotopaxi and Tungurahua volcanoes".

l. 47-48, suggest to remove this sentence. It reads like a conclusion, not a description of scope. Certainly it is not a useful statement in the introduction.

We have modified as follows the entire paragraph: "This paper explores the possibility to develop a simple method to predict debris flow velocity and volume based on seismic sensors installed along the channel, with a limited calibration dataset. The aim is not to seek a universal law relating seismic energy to debris flow characteristics, but just to provide robust tools for debris flow risk management. Specifically, the proposed method is intended to be easily applicable in different catchments, at least for first order estimations of debris-flow volumes."

l. 52, "data of" change to "data from the"

Done, thanks.

Remove figure 1, see my previous statement about this.

Done, see above.

l. 64, "reliable" change to "reliably"

Ok, done.

l. 91, "(celerity)", add "between the two stations".

Ok, done.

l. 92, You must also mention the ambiguities that arise when no sharp amplitude peaks emerge, or when too many of them appear close to each other. How will you be able to identify the matching ones, the right ones, to derive your velocity estimate? How did you actually do this with your data? What were the criteria? When I look at fig. 6, for example, I find it tough to judge the validity of all the manually constrained velocity values.

We added one sentence (line 173) about the possible limitation to the application of the cross-correlation methods in case of signals characterized by many amplitude peaks close to each other produced, for instance, by the propagation of roll waves.

l. 93, revise the text. The word "only" implies that validation is just a minor thing. However, it actually should be one of the back bones of your study. You have to convince the readers (and me as referee) that the cross correlation approach can yield valid results. Manually matching amplitude peaks is one way to do this, albeit not really a bullet proof one.

We skipped the word “only” and modified the sentence as follows: “The manual analysis is time-consuming but it was needed for validating the results of application of the cross-correlation method and avoiding misinterpretation”.

l. 97, “Number of samples equal distance means”, clarify, this reads cumbersome.

The number of samples for the starting window size of the cross-correlation is set equal to the distance in m.

l. 107, “twice with a sliding”, clarify. It is not clear to me what this should mean. Did you repeat the correlation in the same window? Did you double the window size? Same for the next bit “an overlap of the half sample numbers offers most consistent results”. How can half the sample numbers overlap? Why does this give more consistent results?

The Cross-correlation is performed with an overlap of half of the sample numbers (n/2 in Figure 2).

l. 116, has the data not been filtered? How can you be sure you get a proper signal of the debris flow? Seismologists have presented ample examples of the typical frequency content of debris flows. More importantly, how did you account for the possibility that the recorded signals are not just caused by the flow but many other sources: wind, rain, other meteorological and anthropogenic events? Many of these have a significant frequency overlap and since debris flows are often caused by rainstorms, it is quite likely that these processes can happen at the same time, and they do indeed have a quite strong seismic fingerprint in the data.

As responded above, our geophone dataset is controlled by channel processes given the small distances sensor-channel and the location of monitoring sites. We analyzed seismic data produced by well-documented debris flows collected in experimental basins that also benefit of complementary data (e.g., flow stage data, see Figures 6-7-8) that can be used for validation.

l. 158-159, you should discuss your own results, not start with statements of another study.

We reorganized the entire paragraph as follows: “Our results suggest that the cross-correlation method we used - based on a window length adaptable according to the signal waveform - provides solid estimates of debris-flow velocity, as the temporal resolution of the calculation is high during the fast, initial stages of the flow, while longer window length are applied for smoother flows, thus permitting to avoid wrong correlation results. Arattano et al. (2012) already showed how the cross-correlation technique can provide a reliable estimation of the flow velocity even when the signals recorded at the two monitored cross-sections do not present a clear, common feature - typically the passage of the debris-flow front. Nonetheless, some significant signal features are required such as a progressive rise and subsequent decrease of the signal amplitude. Signals characterized by many amplitude peaks close to each other produced, for instance, by the propagation of roll waves can represent a limitation to the application of cross-correlation methods (Figure 5c)”.

l. 159, what is a “main front”?

We meant the main front of the debris flow. We reworded the entire paragraph, see response above.

l. 159-161, is this based on the data you present? Not sure I can follow these arguments. Please be more concise.

We reworded the entire paragraph, see response above.

l. 164, “the influence of different”, influence on what? Signal amplitude, flow velocity? Please be specific.

We modified as follows: “The influence of different longitudinal distances between the geophones is evident on the velocity estimation.”

l. 167, “a longer distance offers the possibility to use higher resolution for the velocity calculation”, this is confusing. Above you say that longer distances make things complicated. Here you say it can improve the resolution. Please clarify.

We agree that this is confusing and removed this sentence (this only takes in account the higher mathematic resolution).

l. 169-170, “significant difference”, please be precise here. How many sampling intervals is that? Try to quantify statements when it is possible.

The sentence above (line 188) already describes a minimum distance between the geophones.

l. 198, “linear model performs definitely better”, this has simply not been tested. Please test other models or leave this unsupported statement.

We tested other regression models, and the results are summarized in the table below.

regression model	R^2
linear model	0.84
exponential model	0.57
power model	0.55
logarithmic model	0.66

l. 218-219, The method will always have to be far from real time volume estimation of debris flows. Simply because, as the authors express, the volume can only be estimated after the debris flow has passed the geophone array and the energy integral can be completed. This is a structural issue of the approach and there is no future development that can change that. This is OK, but it needs to be expressed in the text.

We contend that the achievement of an algorithm for the automatic determination of debris flow duration will make feasible a rapid estimation of debris flow volumes by the application of our approach. Indeed, in these lines we exactly talk about rapid response as a possible practical outcome of the method, we clarified this point also in the previous sentence: “Nonetheless, adopting such a physically-sound empirical model, a rapid estimate of the order of magnitude of the debris-flow volume is possible”.

l. 240, I think the 20 % will be a quite optimistic estimate when I look at fig. 10. I suggest you calculate the RMS of the Illgraben volume predictions and present/discuss this, instead.

We modified as follows: “the order of magnitude of debris flow volumes can be correctly estimated in most cases from seismic data only”.

The data availability statement is also a fair bit from in agreement with the FAIR principles. Please consider providing it in an appropriate way, through a data repository.

The data availability statement has been modified has follows: “A temporary private link to the geophone dataset gathered at Gadria and Lattenbach is provided as supplementary material to support the peer review process. In the final paper, this link will be substituted with a persistent link to an Open Access data repository (http://ds.iris.edu/ds/products/esec/).”

This is the temporary private link to the data of Gadria and Lattenbach analyzed in the paper: www.almosys.at/seismic-data/

Data from Cancia have been provided by a third part (i.e., Regional Department for Land Safety, Hydrogeological Services Center, ARPA Veneto) and we need a specific permission to share them.

The material in the appendix might be better moved to the supplementary information.

We agree, done.

With our best regards,

Velio Coviello (corresponding author)

References

Andrade, S.D., Almeida, S., Saltos, E. (2022). A simple and general methodology to calibrate seismic instruments for debris flow quantification: application to Cotopaxi and Tungurahua volcanoes (Ecuador). Landslides 19, 747-759. https://doi.org/10.1007/s10346-021-01784-5

Coviello, V., Arattano, M., Comiti, F., Macconi, P., Marchi, L. (2019). Seismic characterization of debris Flows: Insights into energy radiation and implications for warning. Journal of Geophysical Research: Earth Surface, 124, 1440-1463. https://doi.org/10.1029/2018JF004683

Coviello, V., Theule, J. I., Crema, S., Arattano, M., Comiti, F., Cavalli, M., Lucia, A., Macconi, P, and Marchi, L. (2021). Combining instrumental monitoring and high-resolution topography for estimating sediment yield in a debris-flow catchment, Environmental and Engineering Geoscience, 27(1), 95-111. https://doi.org/10.2113/EEG-D-20-00025

Ioriatti, E. (2022). Study of debris-flow initiation through the analysis of seismic signals. Master thesis unibz.

Tsai, V. C., Minchew, B., Lamb, M. P., & Ampuero, J.-P. (2012). A physical model for seismic noise generation from sediment transport in rivers. Geophysical Research Letters, 39, L02404. https://doi.org/10.1029/2011GL050255