Dear Dr. Benjamin Purinton,

First of all we would like to thank you for the very constructive and thorough feedback that you provided on our manuscript. We highly appreciate your suggestions, which will drastically improve the quality and rigor of our analysis and manuscript.

While a more in-depth response and the updated manuscript will be provided upon receiving the decision of the editor, we would already like to address your primary concerns briefly, as we have in the meantime already tested and implemented most of your methodological suggestions. We would be more than happy to further discuss these methods and results and are open to additional suggestions.

1) **Separate discussion section**
We do agree that the paper would benefit from a separate discussions section in which we can address both technical and contextual aspects in more detail. Frankl (reply 2021) and Cox (reply 2021) have also suggested a more elaborate discussion of possible technical caveats or limitations; this will be foreseen in the adapted manuscript.

2) Alternative 30 m DEM: Copernicus
We were happy to be introduced to this new 30 m Copernicus DEM, which will indeed provide many benefits compared to SRTM. We have therefore replaced the 30 m SRTM DEM by the 30 m Copernicus DEM for all our analysis. Using the Copernicus DEM furthermore facilitates better uncertainty assessment, which is addressed in more detail in point 4.

3) Vertical uncertainties
We agree that vertical uncertainties will be substantial for the TanDEM-X and Copernicus DEMs. After verification of the literature and consideration of our specific application, we have opted to address two specific uncertainties or errors, which we believe to be important in our case: i) the interpolation error, and ii) the relative height error.

<u>Interpolation error</u>
After considering your suggestions on the interpolation method (see point 4) we have decided to develop a more rigorous method to determine which interpolation method works best in our landscape setting. To calculate our volumes, we subtract the current DEM elevation from the interpolated pre-erosion surface. We previously assessed the performance of the interpolation methods tested by comparing the percentage of negative volume that resulted from these volume calculations. Therefore, we have now adopted the following method to assess i) which interpolation methods works best ii) what the error/uncertainty of this methods is.

We have taken five different lavaka polygons with different sizes that span the range of our lavaka polygons (100 m² (n = 10), 1000 m² (n=10), 5000 m² (n=10) and 20 000 m² (n = 10)). These 50 lavaka polygons were then placed on unaffected convex-shaped hillslopes on which lavaka typically occur, together with the corresponding horseshoe-shaped pre-erosion polygons. We then tested different interpolation methods (more details are provided in point 4) and calculated the difference between the interpolated surface and the DEM. This difference then gives the interpolation error, as a perfect interpolation would results in an identical interpolated and original surface.

Based on obtained height differences between interpolated and original surfaces several error metrics were calculated (mean, median, root mean squared error (RMSE), mean absolute error (MAE) and standard deviation (std)), which were then used to i) identify the best interpolation method and ii) estimate the interpolation error. Next, it was verified if the magnitude of the interpolation error depends on the size of the lavaka, in order to correctly account for these errors. Our main preliminary results are presented in Table 1, where our main conclusions are the following:

Table 1: Mean, median, mean absolute error (MAE), root mean square error (RMSE) and standard deviation (std) of the Interpolation error in meter for the different DEMs and interpolation methods

	UAV-SfM (0.20 m)					TanDEM-X (12 m)					Copernicus (30 m)				
	Linear	TIN	Spline bilinear	Spline bicubic	Spline reg.	Linear	TIN	Spline bilinear	Spline bicubic	Spline reg.	Linear	TIN	Spline bilinear	Spline bicubic	Spline reg.
Mean	-2.29	-1.83	-2.63	-2.66	-1.75	-1.93	-1.62	-2.63	-2.58	-1.76	-1.44	-1.58	-2.68	-2.54	-0.91
Median	-1.92	-1.32	-2.19	-2.21	-1.47	-1.40	-1.17	-2.16	-2.11	-1.38	-1.00	-1.19	-2.34	-2.13	-0.60
MAE	2.94	2.53	3.28	3.24	2.21	2.41	2.17	3.15	3.09	2.13	2.02	3.02	3.04	2.86	1.44
RMSE	4.21	3.90	4.66	4.41	3.05	3.48	3.18	4.22	4.14	2.97	2.86	3.95	4.11	3.91	2.05
Std	3.53	3.45	3.84	3.51	2.50	2.90	2.73	3.30	3.23	2.39	2.47	3.63	3.12	2.98	1.85

i) The interpolated surface is on average lower than the real surface (mean and median error < 0). Our calculated volumes will therefore, on average, be underestimated.

ii) Coarse DEMs have a lower error. As the coarser resolution DEM contains less topographic detail, it makes sense that the difference between a generalized interpolated surface and the real surface is smaller as when compared to high resolution DEMs that contain more detailed micro-topography that cannot be easily interpolated

iii) Overall, the Regularized Spline with tension interpolation results in the lowest errors.

iv) Based on both pearson and spearman correlation coefficients no significant relationship is observed between lavaka area and the mean or median error for the 30 m Copernicus DEM (Figure 2). However, for the finer resolution TanDEM-X and UAV-SfM DEMs we do find a significant correlation between lavaka area and the mean or median error of a lavaka (Figure 2).

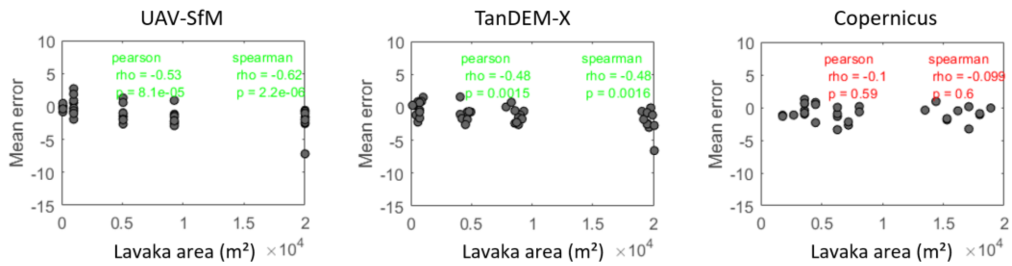

Figure 2:Ccalculated pearson and spearman correlation coefficients and their p-value for the relationship between lavaka area and the mean interpolation error of a lavaka.

Based on these results we now take into account the interpolation error for the different DEMs by running a Monte Carlo analysis:

i) Copernicus: draw random values from a normal distribution with the mean = mean lavaka error and std = std of the mean lavaka error (mean = -0.81 m, std = 1.21 m)

ii) TanDEM-X and UAV-SfM: Established a relationship between lavaka area and mean error. Use this relationship to estimate the interpolation error for each lavaka based on its size. Uncertainties are taken into account by considering the uncertainties on the fitted coefficients, which are again drawn from a normal distribution with known mean and std, where we use copula to account for the correlation between both coefficients.

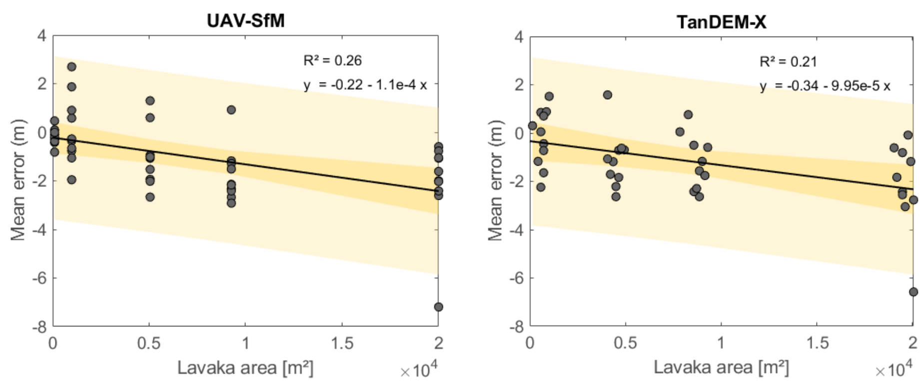

Figure 3: fitted linear relationships between lavaka area and mean error for the UAV-SfM and TanDEM-X DEM

<u>Relative height error</u>

Typically, the performance of a DEM is assessed by considering its absolute vertical accuracy. We, however, argue that this metric is not the most suitable in our case, as we are rather interested in relative pixel-to-pixel errors instead of absolute errors: the difference between the interpolated surface and DEM will determine the volume. An absolute error of X m will not result in a different volume estimate if this absolute error is the same for all DEM pixels. However, if the relative height of the pixels is not correct, this will result in different volume estimates. Therefore, we have opted to use the relative height error. We assume that this relative height error will be negligible for our high resolution UAV-SfM DEM and use the height error masks (HEM) that are

provided for the TanDEM-X and Copernicus DEM to assess this error. The height error mask gives the height error for each pixel in the form of the standard deviation. This error is considered to be a random error and does not include any contributions of systematic errors (Wessel, 2016).
For each lavaka we now calculate the mean HEM-value, which we then use in our Monte Carlo simulations: for the relative height error we draw random values from the normal distribution with mean = 0 (height of the pixel) and std = mean HEM of the lavaka.

Both for the interpolation and relative error we use a lavaka as the observational unit, where we work with the mean errors and standard deviations obtained for this observational unit. In this way we make abstraction of the fact that pixels within a lavaka will likely be autocorrelated and correct for the error on the lavaka as a whole.
By combining both errors in one Monte Carlo simulation, we have been able to calculate the mean volume and uncertainty on this volume (std) for each lavaka, which we can further propagate into the establishment of the Area-volume relationship and subsequent calculations.

4) Interpolation

Two concerns were raised by the reviewer: i) the use of parameterized spline interpolation and ii) random-points approach.

<u>Interpolation method</u>

First of all we have now improved our method to assess which interpolation method works best (see point 3 interpolation error). Furthermore, we have assessed the performance of different interpolation techniques, all available in Open Source QGIS:

- i) Linear (GDAL)
- ii) TIN (Qgis)
- iii) Spline: Bilinear (GRASS)
- iv) Spline: Bicubic (GRASS)
- v) Spline: regularized with tension (GRASS)

We did some manual try-outs with the nearest neighbor (GDAL) algorithm that was proposed by the reviewer. Visual interpretation of these results (Figure 4), together with the fact that this method is typically applied to categorical data made us decide to not consider this method in further analysis.

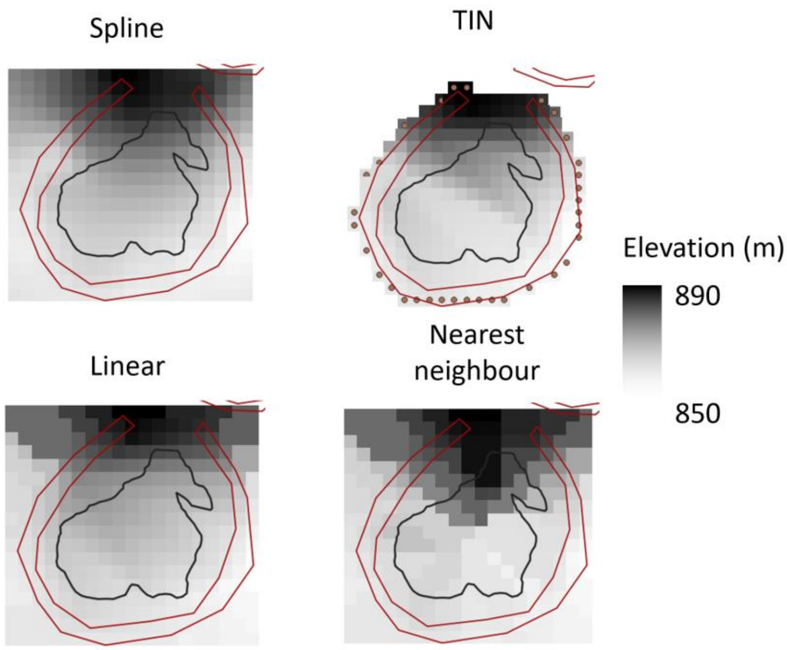

Figure 4: Examples of four different interpolation methods for the TanDEM-X DEM: Regularized spline with tension, TIN, Linear and Nearest neighbor interpolation.

From the different error metrics (Table 1) it was concluded that regularized spline with tension yields the lowest interpolation errors. Therefore, we decided to use this interpolation method for all further analysis. We will, however, add a discussion on the disadvantages and limitation of this method.

<u>Random-points approach</u>

We agree with the reviewer that the creation of the random points in the pre-erosion surface polygon is not necessary and might result in incorrect interpolation results as then indeed the same pixel value might be considered multiple times during interpolation. We have therefore changed this method by following the suggestion of the reviewer where we now create one point per pixel, which are then used for interpolation. The steps followed to derive the lavaka volumes are now the following:

i) Clip the DEM with the pre-erosion polygon. All bordering pixels are included in order to assure that the width of the pre-erosion raster has at least one pixel for the coarse resolution DEMs
ii) One point per pixel is created to which the height of the pixel is assigned
iii) The pre-erosion surface is constructed by interpolating between these points
iv) The volume is calculated by taking the difference between the interpolated pre-erosion surface and the current surface.

To summarize the changes in our methods:

i) we now use the 30 m Copernicus DEM instead of the SRTM DEM

ii) vertical uncertainties are calculated for each lavaka based on both the interpolation error and the relative height error. These errors will be propagated in all subsequent calculations.

iii) the performance of different interpolation methods is assessed based on the interpolation error. This is done by reconstructing intact hillslopes and by calculating the difference between the interpolated and real surface. Regularized spline with tension has the lowest errors. We therefore applied this interpolation method to our further analysis

iv) The volume calculation workflow is adapted so that we now use one point per DEM pixel for interpolation instead of the random points.

We would like to highlight that the figures presented here will be further modified/finetuned for the revised version of the manuscript. Minor concerns of the reviewer will be addressed in detail in the formal rebuttal letter and in the revised version of the manuscript.

We want to thank Dr. Benjamin Purinton for his thoughtful feedback that will improve the methodological robustness of this work and are open to any further suggestions to improve the quality of our work.

Sincerely,

Liesa Brosens on behalf of the co-authors

References

Cox (reply 2021). CC1: Comment on esurf-2021-64 by Rónadh Cox, 19 Oct 2021. https://doi.org/10.5194/esurf-2021-64-CC1

Frankl (reply 2021). RC2: Comment on esurf-2021-64 by Armaury Frankl Oct 2021. https://doi.org/10.5194/esurf-2021-64-RC2

Wessel B. (2016), “TanDEM-X Ground Segment – DEM Products Specification Document”, EOC, DLR, Oberpfaffenhofen, Germany, Public Document TD-GS-PS-0021, Issue 3.1, 2016. [Online]. Available: https://tandemx-science.dlr.de/