Review of Earth Surface Dynamics submission: “Comparative analysis of SRTM, TanDEM-X and UAV-SfM DEMs to estimate lavaka (gully) volumes and mobilization rates in the Lake Alaotra region (Madagascar)” by Brosens et al.

The study of Brosens et al. examines the applicability of three DEMs at three different resolutions (and different sources) to measure the erosion of lavaka (gully) features in Madagascar. The gullies are carefully identified from satellite and aerial imagery at three different dates (1949, 1969, 2011-2018) and their extents are digitized. Following this, pre-erosion surfaces are created for each gully and the volume of excavated sediment is measured. This allows the authors to build area-volume relationships to apply to the three time steps, and measure volumes from area alone. Of great note, they find two order of magnitude higher erosion rates in comparison to cosmogenic radionuclides.

First off, I want to thank the authors for exposing me to these dynamic and fascinating geomorphic features. While I am familiar with significant gully erosion in other places, I had never heard of the lavakas of Madagascar, and they are impressive. I read the study with great interest since bridging gaps in satellite and aerial (i.e., drone) measurements to quantify geomorphic processes presents exciting opportunities, albeit with significant challenges. These challenges are both in terms of spatial and temporal resolution differences (which the authors cover) and dataset accuracy (which the authors mention but do not consider). Overall, I found the paper interesting and think it should be published in ESurf, but there needs to be some major revisions.

I begin by listing my primary concerns, followed by some more specific comments. The references that are not already included in the submitted manuscript are provided at the end of this review.

Primary Concerns

The paper could do with a dedicated discussion section. There are many points I make below which would be valid items for a discussion, and other points that may expand the methods and results. Furthermore, the discussion section would allow the authors to more fully place the study and results in the context of other work on gully erosion cited (e.g. Cox et al., 2009, 2010; Perroy et al., 2010; Vanmaercke et al., 2021). Placing the chosen methods, observed volumes, and the causes of (increased?) gully erosion in the context of these other studies will present a fuller, and more citable, study.

I do not agree with using the SRTM DEM as the 30 m resolution dataset for a number of reasons. Firstly, as stated by the authors, SRTM uncertainties are often > 5 m, precluding accurate volume calculations in many cases. Secondly, previous work (Smith and Sandwell, 2003; Farr et al., 2007) has shown that the actual resolution of SRTM is likely on the order of 45-60 m. Finally, better open access DEMs exist. The authors could instead use the ALOS World 3D 30 m (available here: https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm) or, the cutting edge Copernicus DEM (available here: https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.032021.4326.1). The ALOS data has low vertical uncertainties and is based on a resampled 5 m DEM (cf. Purinton and Bookhagen, 2017) and the Copernicus DEM is essentially a 30 m version of the TanDEM-X (references: https://spacedata.copernicus.eu/documents/20126/0/GEO1988-CopernicusDEM-RP-001_ValidationReport_V1.0.pdf and https://spacedata.copernicus.eu/documents/20126/0/GEO1988-

CopernicusDEM-SPE-002_ProductHandbook_I1.00.pdf) I strongly recommend the authors utilize the newer Copernicus DEM in place of SRTM. It may be the case that the newer dataset can be included in the breakpoint analysis, especially considering that the largest lavaka make up the majority of exported sediment (as shown in Figure B2). If a 30 m open-access, near-global DEM (Copernicus or ALOS) provides some decent results, then the impact of this study would be greatly increased.

Vertical uncertainties of the different DEMs are never considered in the analysis. But this is a vital step when using spaceborne DEMs for volume estimations, for instance in the cryospheric (e.g., Brun et al., 2017) and geomorphic (e.g., Purinton and Bookhagen, 2018; Bessette-Kirton et al., 2018) communities. While I think the authors can safely argue for negligible vertical uncertainties (with the proper citations) in the UAV-DEM, these cannot be ignored in the case of TanDEM-X and SRTM (cf. Purinton and Bookhagen, 2017, 2018). In the case of the Copernicus DEM I reference above, TanDEM-X uncertainties can be used given this DEM was generated from the same source. Preliminary reports on Copernicus DEM accuracy are available here: https://spacedata.copernicus.eu/documents/20126/0/GEO1988-CopernicusDEM-RP-001_ValidationReport_V1.0.pdf. Furthermore, depending on the presence of vegetation (i.e., forests) in the DEM pixels, these different datasets (radar X-band for TanDEM; C-band for SRTM; optical images for ALOS) may have additional uncertainties. These uncertainties with regards to the study area characteristics (bare-earth, forests, bushes) should be mentioned.

I think the handling of vertical uncertainties can be done in a discussion section, but it may be better inserted in the methods and results. I suggest: an uncertainty value (e.g., RMSE or NMAD) is selected from the literature regarding each DEM and this uncertainty is propagated to the volume estimates. This would then put error bars on the regressions in e.g., Figures 5 and 6, which could be considered during the power-law fitting. This uncertainty will also propagate in the negative vs. positive volume calculations, presenting a range of percentages rather than an exact value, which implies perfect DEM accuracies.

In Purinton and Bookhagen (2018) we also attempted volume estimation (in this case using the more common DoD approach) between the SRTM and TanDEM-X DEMs. In this case, the uncertainties associated with the SRTM precluded widespread geophysical results, except in the areas of very rapid / large magnitude change. This study highlights the influence of spaceborne DEM accuracy and the care that must be taken when combining older and newer datasets from different sources, with different errors, in a given analysis. While detailed correction steps are not necessary in this case, this work should be noted, particularly since it is in the same journal.

In the interpolation step, I think some more work and justification is needed. I see the temptation to generate curved surfaces using splines, but I think nearest neighbor void filling is somewhat safer (not based on any parametrized curve fitting) and simpler. This may result in "staircase" artifacts, but those may not have a huge impact on the final volume estimates, or the vertical uncertainties of the spaceborne DEMs may have a larger impact. The authors should experiment with the GDAL gridding methods accessible in QGIS (Processing Toolbox > GDAL > Raster Analysis: GRID (nearest neighbor, linear, etc.)), since these are robust and widely used. Previous authors working on gullies have had success generating pre-erosion surfaces with simpler (i.e. parameter-free) interpolation (e.g., Perroy et al., 2010; Eustace et al., 2009, Evans and Lindsay, 2010). Furthermore, I'm not convinced the random-points approach is entirely necessary. Couldn't the authors just use each elevation value that the "horseshoe" overlaps with one time? Otherwise some elevations (single pixels) are used multiple times, which I don't understand the reason for and seems inappropriate.

I suggest the authors at least attempt the processing with the nearest neighbor approach, mentioning that this is parameter free and uses only the original elevation values, and if the results create significantly more negative volume, then mention this as justification for higher order techniques. This can be an item in the discussion section.

Specific Comments

Line 3: “high resolution DEMs”, I prefer that the exact resolution in meters is mentioned or the terms high, medium, and low resolution are clearly defined and a range of values for each is given. Twenty years ago 30 m DEMs were very high resolution. For posterity, best to be exact. Please note this change in other places in the manuscript (e.g. Line 31)

Line 10: “SRTM DEM is too coarse”, or too inaccurate? This pertains to one of my primary concerns. Also bear in mind the SRTM realistically has a ground resolution closer to 45-60 m (Smith and Sandwell, 2003; Farr et al., 2007)

Line 19-20: What do these different rates correspond to? Are they the range and average of the six different study areas?

Line 26: Rephrase “more and more”

Line 30: “remote sensing product”. Well, UAVs are also remote sensing if we define remote sensing as measurements that don’t disturb the surface. I would change this to “spaceborne product”

Line 32: They don’t necessarily need to all be here, but somewhere (perhaps in a new dedicated section) references to the accuracy of these various datasets should be cited (e.g., Purinton and Bookhagen, 2017; Rizzoli et al., 2017; Wessel et al., 2018)

Line 36: “Gully erosion…”, awkward sentence, rephrase.

Line 39: From “where”, awkward clause. Consider new sentence and/or rephrasing.

Line 41: What is being referred to by “high resolution surface imagery”? If the authors are referring to GoogleEarth then be explicit and I suggest also referencing Fisher et al. (2012).

Line 63-64: For this step the lavaka area was taken from the most recent (2010s) polygons? Maybe state this.

Line 80: “are available” maybe change to “were generated”, since these are data the authors created for this study. And what a nice dataset it is!

Line 81 and Table A1: I’m not familiar with Maxar-Vivid-WVO2, but I suppose this refers to the WorldView-2 satellite? Please call them WorldView-2 if so. Also, are these images proprietary or received through grant? They should be mentioned in the “Code and data availability” section. One more point: the ground resolution of the aerial photos and satellite images should be mentioned here (and/or in the table).

Figure 1: The color-scale could be improved here and elsewhere using e.g. the instructions here: https://gis.stackexchange.com/questions/94978/elevation-color-ramps-for-dems-in-qgis. The authors can decide for themselves, but the current blue to red scale is odd for topography. The red-blue could then be saved and used for topographic difference as is often done (e.g., Wheaton et al., 2010). Furthermore, in (a) the topography around the Alaotra catchment should also be shown, otherwise it looks like this implies ocean around it, which makes the inset map of Madagascar also confusing. Note the Krieger (2007) citation appears twice, drop one of them.

Line 88-89: The method of resampling to UTM coordinates should be mentioned (bilinear?). Also, importantly, the SRTM is likely referenced to the EGM96 geoid, whereas the TanDEM-X is referenced to the WGS84 ellipsoid. Was a vertical datum conversion done to bring the datasets into the same vertical reference? And what is the vertical datum for the UAV DEM?

Line 95-99: A few points here. While the UAV DEM was likely of high quality, the authors should confirm that there is no notable doming effect from using a fish-eye lens and no ground control points. Doming (cf. James and Robson, 2014) is a known issue with SfM-MVS from drones, and fish-eye lenses will exacerbate this. I suggest the authors examine the raw UAV point cloud in either Pix4D or perhaps CloudCompare over flat surfaces in the study area (I realize this may be difficult to find) and take visual note of any large-scale warping. Ideally a reference dataset or independent GNSS points could be used for validation, but that is missing here, correct? One more thing: there are many other citations regarding UAV-DEMs for geomorphic analysis and I recommend including them alongside or in place of Grohmann, 2018 (e.g. Cook, 2017 and others therein and referencing this study). The authors could even give a range of expected vertical accuracy from their UAV-DEM taken from the literature in other cases where GCPs were not used. This would provide justification for passing these off as “negligible”.

Line 116: The auxiliary files delivered with TanDEM-X (the COV.tif file) would allow the authors to specifically report the number of coverages used to generate the final DEM in this study area. I suggest reporting this value (mean +/- standard deviation, or range) since it has a large impact on DEM quality (vertical uncertainty).

Line 124-125: As noted, these height errors are really high for the SRTM and this is likely an inappropriate dataset for this application, particularly when I consider the bottom row of Figure 2 where the SRTM differences is maybe less than 5 m? Although again, as I note below, an improved (classified) color-scale would be helpful here.

Line 141: What is meant by “precise identification”? In this case, is the horseshoe drawn to not include the other lavakas? So a sort of broken horseshoe shape? Please provide a visualization of the points selected for interpolation on Figure B1. From this figure I don’t see in many cases how enough points could be selected to provide a reasonable interpolation in some of the locations where the lavakas appear to have no, or very little, pre-erosion topography preserved where they are touching.

Line 146: Delete hyphen in “DEM-pixel“

Line 160-162: I’m concerned about this 1-pixel lower limit. A single pixel can easily represent an inaccurate measurement. It would be better to have a multi-pixel lower limit. This could be 5 pixels (i.e., lavakas smaller than 5 pixels in each DEM are not considered), but that would remove a significant number of lavakas from the e.g. 30 m DEM. On the other hand, per the scaling relationships, perhaps these <5 pixel lavakas do not contribute significantly to the sediment budget?

Figure 2: The elevation colorbar is reversed with respect to Figure 1, please check it here and elsewhere (and as suggested consider different color scales). In this figure the elevation difference colorbar should really be classified, not continuous. For instance, if I look at the SRTM it’s hard to tell but those green values are maybe < 10 m (maybe even < 5 m)? That’s getting awfully close to the vertical uncertainty of SRTM. Classified color scales broken into ~5 m ranges with perceptually distinct colors would help a lot.

Line 216: See primary concerns above. Here the uncertainties on the a and b coefficients may be too small, since the regression does not consider uncertainties on the volume calculation which may be significant for the TanDEM-X and SRTM (if the authors continue to use the SRTM and not ALOS or Copernicus).

Line 226: “1.3% vs. 0.3%” and others, values should be switched (0.3 vs. 1.3) to match the preceding sentence.

Line 233: “increasing” should be “decreasing”

Line 241: Remove hyphen in “data-areas”

Line 260: Reference Smith et al. (2019) (in this journal) when discussing “optimal grid resolution”

Line 275: “at” should be “of”

Line 295-298: No mention of TanDEM-X vertical uncertainties. Here and elsewhere the uncertainties are only considered with regards to resolution.

Line 308-313: The scaling relationships are considered in the context of landslide studies. Do these results really hold for gully scaling? Can the authors include references or justification for the landslide comparison? I could accept an argument that the scaling relationship is similar, but the processes are different.

Figure 6: Am I correct in my visual interpretation that the break-point is found at the point where the relationship becomes 1:1? In that case, is the broken-stick analysis entirely necessary, or could a simpler approach be to just consider where the RMSE (or some measure of spread) from a 1:1 line passes below some limit (e.g. 5%). Just food for thought, I think the broke-stick is valid, but it may be worth mentioning this 1:1 change-point.

Line 322-323: And this is why uncertainties on the volume estimates are important for the coefficient estimation (small differences in coefficients lead to large differences in volumetric growth and mobilization).

Lines 337-345 and Figure 7: This would be good fodder for a discussion section and would allow expansion and inclusion of other references to gully studies (e.g. Vanmaercke et al., 2021 and references therein). I recommend removing the correlation coefficients from plots b-d in Figure 7 and where they are referenced in the text. These are only six data points and it may be best to just discuss the graphical trends observed, since these are not robust statistics in this case.

Line 345: This Brosens et al. (in review) paper is seemingly important for the discussion of results (reasons for gully changes). Hopefully review progresses there quickly and a final citable result is available for this paper / discussion. This paper leaves me asking “why are the lavaka erosion rates increasing?” In Perroy et al. (2010) (worth citing here) gully erosion increased in response to grazing.

Line 361: “Table A1 and 1”, not sure what the 1 is supposed to refer to.

Line 386: “area” should be “are”.

I had a look at the three example lavaka files and code on GitHub, thanks for publishing that, it is really useful. However, when I tried to run the LavakaVolumesPyQGIS.py script, I received error messages about importing modules and functions. There are no imports at the top of the script, is this something that is missing? Or could the authors add to the GitHub README the steps to actually run the script? Maybe this is done through a GRASS shell, but I’m not familiar with those steps. I tried adding “import os” and “from qgis.core import *” but then got an error about an undefined “processing” variable. These do not need to be detailed instructions, but the common way of running a script at the command line (python <script name>) does not work in this case, so maybe just note the steps to open and run the script.

I hope these are helpful and constructive criticisms, and I welcome further discussion in the open online forum.

Sincerely,

Ben Purinton

References

Bessette-Kirton, E. K., Coe, J. A., and Zhou, W.: Using Stereo Satellite Imagery to Account for Ablation, Entrainment, and Compaction in Volume Calculations for Rock Avalanches on Glaciers: Application to the 2016 Lamplugh Rock Avalanche in Glacier Bay National Park, Alaska, J. Geophys. Res.-Earth, 123, 622–641, https://doi.org/10.1002/2017JF004512, 2018.

Brun, F., Berthier, E., Wagnon, P. et al. A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016. Nature Geosci 10, 668–673 (2017). https://doi.org/10.1038/ngeo2999

Cook, K. L.: An evaluation of the effectiveness of low-cost UAVs and structure from motion for geomorphic change detection, Geomorphology, 278, 195–208, 2017.

Fisher GB, Amos CB, Bookhagen B, Burbank DW, Godard V, Whitmeyer SJ. Channel widths, landslides, faults, and beyond: The new world order of high-spatial resolution Google Earth imagery in the study of earth surface processes. Geological Society of America Special Papers. 2012 Oct 1;492(01):1-22: http://burbank.faculty.geol.ucsb.edu/Site/Publications_files/Fisher_etal_GSASpecPap492_2012.pdf

James MR, Robson S. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. Earth Surface Processes and Landforms. 2014 Aug;39(10):1413-20. https://doi.org/10.1002/esp.3609

Purinton, B., & Bookhagen, B. (2017). Validation of digital elevation models (dems) and comparison of geomorphic metrics on the southern central Andean plateau. Earth Surface Dynamics, 5(2), 211–237. https://doi.org/10.5194/esurf-5-211-2017

Purinton, B., & Bookhagen, B. (2018). Measuring decadal vertical land-level changes from srtm-c (2000) and tandem-x in the south-central Andes. Earth Surface Dynamics, 6(4), 971–987. https://doi.org/10.5194/esurf-6-971-2018

Smith, B., and D. Sandwell (2003), Accuracy and resolution of shuttle radar topography mission data, Geophys. Res. Lett., 30(9), 1467, doi:10.1029/2002GL016643.

Smith, T., Rheinwalt, A., and Bookhagen, B.: Determining the optimal grid resolution for topographic analysis on an airborne lidar dataset, Earth Surf. Dynam., 7, 475–489, https://doi.org/10.5194/esurf-7-475-2019, 2019.

Wessel, B., Huber, M., Wohlfart, C., Marschalk, U., Kosmann, D., and Roth, A.: Accuracy Assessment of the Global TanDEM-X Digital Elevation Model with GPS Data, ISPRS J. Photogram. Remote Sens., 239, 171–182, 2018.