Response to Reviewer Stuart Grieve for manuscript submission to Earth Surface Dynamics,

Validation of digital elevation models (DEMs) and comparison of geomorphic metrics on the southern Central Andean Plateau (esurf-2017-4)

We appreciate this review and we are thankful for the insights and close reading of the manuscript. Highlighted below in **bold are the reviewer comments**. Below is our response, which addresses all changes in the final manuscript submission following completion of the interactive review period.

General comments

I found the abstract very dense, in particular because of the use of a large number of acronyms. While I recognize that the results for individual data products need to be specified, it may help the reader get to grips with the aims of the paper, if in addition to grouping the datasets in the 4th line of the abstract by resolution, they were also grouped into radar and optical sources, as occurs later in the manuscript.

The density of the abstract reflects the density of the study, but we appreciate the comments with understanding the scope given all the information (and plethora of acronyms). The second reviewer suggested simplification of the abstract, with less reference to specifics of the results. We've chosen to meet the requests somewhere in between and have removed several abbreviations from the abstract. Below we show what we hope is a simplified version of the abstract, which perhaps adds the necessary clarity:

“

In this study, we validate and compare elevation accuracy and geomorphic metrics of satellite-derived digital elevation models (DEMs) on the southern Central Andean Plateau. The plateau has an average elevation of 3.7 km, and is characterized by diverse topography and relief, lack of vegetation, and clear skies that create ideal conditions for remote sensing. At 30 m resolution, the

SRTM-C, ASTER GDEM2, stacked ASTER L1A stereopair DEM, ALOS World 3D, and TanDEM-X have been analyzed. The higher resolution datasets include 12 m TanDEM-X, 10 m single-CoSSC TerraSAR-X / TanDEM-X DEMs, and 5 m ALOS World 3D. These DEMs represent the state-of-the-art for optical (ASTER and ALOS) and radar (SRTM-C and TanDEM-X) spaceborne sensors.

We assessed vertical accuracy by comparing standard deviations of the DEM elevation versus 307,509 differential GPS measurements across 4,000 m of elevation. For the 30 m DEMs, the ASTER datasets had the highest vertical standard deviation at > 6.5 m, whereas the SRTM-C, ALOS World 3D, and TanDEM-X were all < 3.5 m. Higher resolution DEMs had generally lower uncertainty, with both the 12 m TanDEM-X and 5 m ALOS World 3D having < 2 m vertical standard deviation. Analysis of vertical uncertainty with respect to terrain elevation, slope, and aspect revealed the low uncertainty across these attributes for the SRTM-C (30 m), TanDEM-X (12-30 m), and ALOS World 3D (5-30 m). Single-CoSSC TerraSAR-X / TanDEM-X 10 m DEMs and the 30 m ASTER GDEM2 displayed slight aspect biases, which were removed in their stacked counterparts (TanDEM-X and ASTER Stack).

Based on low vertical standard deviations and visual inspection alongside optical satellite data, we selected the 30 m SRTM-C, 12-30 m TanDEM-X, 10 m single-CoSSC TerraSAR-X / TanDEM-X, and 5 m ALOS World 3D for geomorphic metric comparison in a 66 km^2 catchment with a distinct river knickpoint. Consistent *m/n* values were found using chi plot channel profile analysis, regardless of DEM type and spatial resolution. Slope, curvature, and drainage area were calculated and plotting schemes were used to assess basin-wide differences in the hillslope-to-valley transition related to the knickpoint. While slope and hillslope length measurements vary little between datasets, curvature displays higher magnitude measurements with fining resolution. This is especially true for the optical 5 m ALOS World 3D DEM, which demonstrated high-frequency noise in 2-8 pixel steps through a Fourier frequency analysis. The improvements in accurate space-radar DEMs (e.g., TanDEM-X) for geomorphometry are promising, but airborne or terrestrial data is still necessary for meter-scale analysis.

“

In the discussion surrounding the measurement of hillslope length, the challenges of interpreting hillslope length from slope area plots are highlighted clearly and although I prefer the use of a flow path method to discern hillslope length I appreciate the utility of the slope area method in this study and believe it provides some very interesting results. However, one additional issue I would like to see highlighted is the assumption that grid resolution is equivalent to the unit contour width and can be used to convert a drainage area into a characteristic hillslope length. In some cases this may be appropriate, but these two parameters are distinct and to my knowledge no work has been done to attempt to correlate these parameters. I do not expect your analysis to change, as varying a constant in the calculation of hillslope length will not change the trends of your results, however, adding a sentence to highlight this assumption within the methodology on Page 12 would enhance the clarity of this section.

We appreciate the flow path method for calculating hillslope length, and experimented with these calculations via the tools available on LSDTopotools (https://github.com/LSDtopotools; Grieve et al., 2016a,b). After much trial we were unable to get consistent results with the flow path method. We suspect this has to do with the relative coarseness of our data, compared with the high-resolution lidar data that these methods were developed on. Particularly in the processing step where soil-mantled, continuous hilltops are identified, we found widely varying results. Importantly, our study area does not have the extent of soil-mantled hilltops as those on which the algorithms were developed (e.g., Gabilan Mesa), and the algorithms are therefore perhaps inappropriate in our rocky, thinly soil-mantled, hyper-arid desert study area. Furthermore, we were not comfortable applying the flow path method that we had not personally developed and written the code for and are familiar with all boundary conditions. We are currently experimenting with different approaches to derive flowpaths, but none has been satisfying yet. Due to these reasons, we instead chose the simplified area-slope plotting technique. Although problematic for a number of reasons (including the use of unit contour width), we feel that this method nonetheless provided an interesting comparison of the various DEMs. Below we copy a passage to be modified on Page 12, Line 14-19 to highlight the contour width assumption:

“

Following the method of Roering et al. (2007), we divide the rollover drainage area by DEM resolution to approximate the characteristic horizontal hillslope length (L_H), providing an additional check on DEM applicability to geomorphology. This method relies on the assumption that DEM resolution is equivalent to unit contour width, which may be an oversimplification. Despite this caveat, the resolution, or unit contour width, serves as a constant for division and differences between values will not alter the trend of the results. We use the horizontal definition of L_H since the difference between horizontal and downslope L_H (as measured by Grieve et al. (2016a,b,c)) should be minimal, except for very high slope angles.

“

Throughout the analysis, curvature and slope is calculated using a 9 cell window, which suggests that as the grid resolution is varied between data products, these derivatives of elevation will be calculated across differing length scales, potentially capturing the signals of processes operating at distinct spatial scales. I would be interested to see a small discussion of this difference between this paper’s approach and other approaches to measuring curvature and slope from kernels of a variable radius.

This is indeed an important distinction, as differences in slope, curvature, and other derivatives may not only be related to the different sensors, but also to the different resolutions utilized. By using an equally sized 9 cell window (3 by 3) across the datasets (with resolutions of 5-30 m), we are measuring different length scales. A few references (Albani et al., 2004; Sofia et al., 2013) pointing this effect out were suggested by the second reviewer, which we will include in the revised manuscript with regards to this discussion. As a matter of fact, we already did a small analysis of this effect by bilinear resampling the 5 m ALOS World 3D DEM to 10 m and 30 m, and running the analyses again (area-slope plotting, curvature and slope distribution comparison). By resampling the data to the coarser resolutions, we are essentially changing the length scale over which these factors are calculated (still in a 9 cell window). Our results show that with coarsening resolution the 5 m ALOS data still shows high-frequency noise, but overall become increasingly similar to TanDEM-X and SRTM-C. The resampled data have a similar median and range of slope and curvature values and similar slope rollover to the coarser data, though more outliers are still measured in the resampled 5 m data. Please see the figure included below for clarification (Fig. S11 in the revised Supplement). This result indicates that the different resolutions capture different geomorphic signatures independent of sensor, which is a relevant result for the limitations of coarser data (particularly greater than ~15 m resolution). We include a new paragraph about these results in the revised manuscript discussion section (Sect. 5.2.2.) Page 35, Line 21 - Page 36, Line 6:

“

One important distinction to make with slope and curvature measurements is the window size used for calculation as differences may not only be related to the different sensors, but also to the different resolutions. By using an equally sized nine cell window (3 × 3) across the datasets, we are measuring different length scales. Numerous authors (e.g., Albani et al., 2004; Sofia et al., 2013) point this effect out with regard to elevation error propagation. To test this, we bilinear resampled the 5 m AW3D5 to 10 m and 30 m, and examined the slope and curvature distributions compared to the 12 m TanDEM-X and 30 m SRTM-C. By resampling the 5 m data to the coarser resolutions, we are essentially changing the length scale over which the derivatives are calculated (still in a nine cell window). Our results show that with coarsening resolution the 5 m AW3D5 still shows high-frequency noise, particularly with respect to curvature, but overall become increasingly similar to TanDEM-X and SRTM-C (Fig. S11). This result indicates that higher resolution data captures more information even when measuring over the same length scale as coarser data. However, we demonstrate with the Fourier analysis that, in this case, the additional information is just sensor-related noise.

“

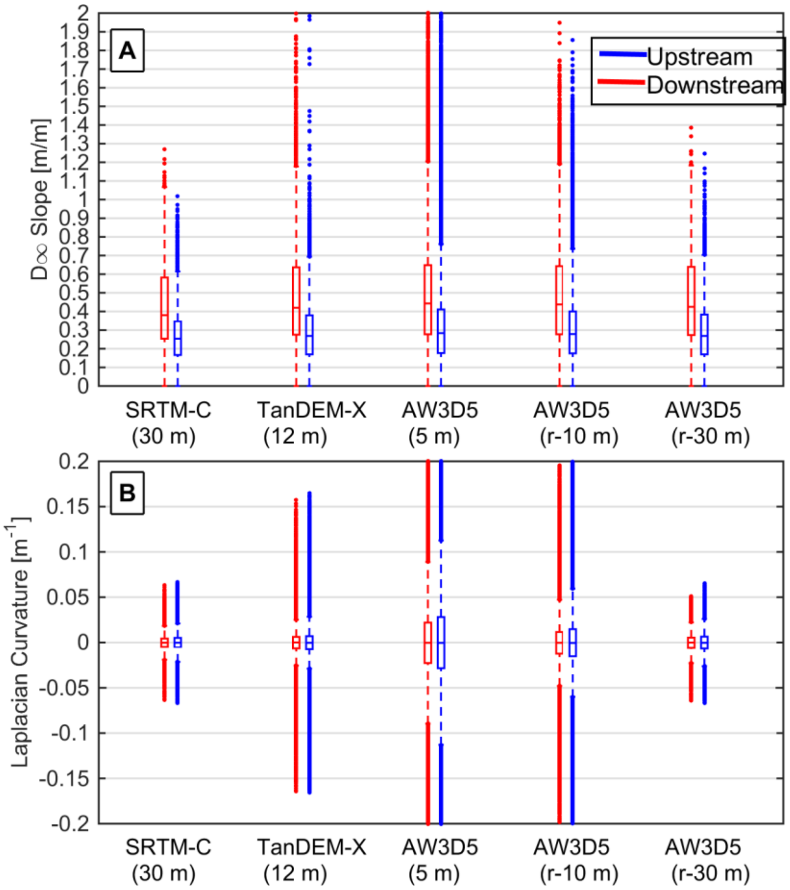

Figure S11. Slope (A) and curvature (B) box plots separated upstream (blue) and downstream (red) of the knickpoint. The AW3D5 DEM (native 5 m resolution) has been bilinear resampled (r) to 10 m and 30 m for comparison with TanDEM-X and SRTM-C. Slope and curvature are calculated in a 9 cell window. In (A) greater magnitude slopes in the interquartile range are still measured in the AW3D5, even after resampling, when compared to the other datasets at their respective original resolutions. Slope outliers, however, remain similar. In (B) particularly, we note that there are a greater number of outlier measurements and wider interquartile range for the AW3D5 resampled. We note that in both plots the differences between datasets becomes reduced in the 30 m resampled AW3D5 and 30 m SRTM-C, whereas the difference is more apparent in the 10 m resampled AW3D5 compared to the 12 m TanDEM-X.

In the line by line comments below I have identified some confusion between the terminology of grid resolution increasing or grid size decreasing. Please check the manuscript for any other instances of this.

Thanks for the note. We will be sure to alter accordingly.

A lot of reference is made to supplemental figures S4 to S9, I appreciate that these are large figures, but it may be more valuable to present this information in the main manuscript to ensure readers who don't always read supplements will still see the interesting results from these datasets. However, I will leave this up to the authors and the AE to decide whether this will result in too many figures in the main manuscript.

The reference to these figures is unfortunately unavoidable, however, given the length of the manuscript, we will not include S4-S9 in the main text, which would lead to 21 figures. The most important results are referenced in the text, and if the reader would like to confirm these statements, we leave it to them to go check the supplement. We would like to avoid adding too much information about the elevation accuracy into the manuscript, which may obscure the main points we wish to make about the geomorphometric potentials of the data.

<u>Line by line comments</u>

Page 2, Line 23 - There have been developments in the production of adaptive resolution DEMs (e.g. Liu et al., 2014), it would make this section more complete to direct an interested reader to some of these papers.

Reference added:

“

These datasets are commonly received in gridded format – rather than point cloud, triangulated irregular network (TIN), or other recently developed adaptive formats (e.g., Liu et al., 2014) – resulting in a defined measurement interval (grid resolution) that may oversimplify fine landscape variability.

“

Page 2, Line 31 - Grid resolution is increasing, grid size is decreasing.

Confusion over these terms has been noted throughout.

Page 3, Line 26 - With this list of lidar applications it would be better if the references were placed alongside their examples, rather than a long list of uses followed by a long list of references.

Changed accordingly.

Page 32, Line 7 - The global compilation of m/n values and other properties by Harel et al. (2016) would be a good reference to add in here to place these results in their full context.

Reference added.

Page 35, Line 27 - Is there a reason for the selection of a 1 km radius for the estimation of relief?

This discussion section is mostly speculation for our hillslope length results, but it definitely makes more sense for us to use the L_H value we measured (109 ± 26 m) as the radius for relief measurement, as is done in the "How does grid-resolution modulate the topographic expression of geomorphic processes?" in Earth Surface Dynamics (Grieve et al., 2016c). Below is the modified passage using a more sensible relief radius:

"

Given the presence of a knickpoint, we expect longer hillslopes in the gently sloped upstream catchment, whereas our data demonstrate equally long hillslopes in the steeper downstream section. This result may be caused by the graphical selection of a rollover value in slope-area plots, which have known deficiencies in assuming uniform basin shape at these scales (Grieve et al., 2016a). However, it could be that upstream and downstream L_H are in fact similar. In that case, steeper slopes downstream given the same curvature and drainage area (Figs. 9 and 11) would be compensated by greater relief rather than longer hillslopes in order to maintain steady state of hillslopes with respect to changing baselevels caused by the migrating knickpoint. Relief measured across DEMs with a 100 m radius (consistent with the measured L_H) is ~35 m greater downstream, indicating that this hypothesis is a possible explanation for the similar L_H.

"

Page 39, Line 4 - This is the only paper title in the reference list which is in block capitals.

Changed accordingly.

Page 39, Line 8 - Check the formatting of the author's name for this paper.

This is a technical report for the original ASTER release, we changed the reference to better reflect this:

"

METI/NASA/USGS: ASTER Global DEM Validation Summary Report, Tech. rep., METI/ERSDAC, NASA/LPDAAC, USGS/EROS, 2009.

"

Figures and Tables

Figure 10 - As the points obscure each other, would it be possible to introduce some transparency to the diamonds to more clearly show both datasets?

This figure will be adjusted in the revised manuscript.

Supplement

The supplement is an excellent addition to the main text and provides detailed information on the data and the methodologies employed. Everything is very clear, aside from the description of the use of "standard GIS tools", it would be helpful to indicate which program you used to help future authors reproduce your work.

Sorry for being vague there. The supplement is changed to say "using the Point to Raster tool in ArcGIS". We also tried this gridding by writing a python script using the GDAL module (no arcpy), but this was slow given the large number of points (>300,000) that needed to be averaged into a grid. Perhaps this is a task that can be revisited, but in the meantime a simple arcpy script calling on the "Point to Raster" tool works nicely and only takes a few minutes to run.

I have also gone through the provided Matlab code and although I have not run it as I do not have access to the right licenses, from a close reading of the code it appears to implement the analysis which is described in the paper. I would also like to thank the authors for sharing their code.

Thanks for taking a look at the code, and sorry to hear about the license issues. In the future we will move entirely in the open-source (Python) direction as we have done for other projects. Currently, the Fourier functions (and the other tools we rely on) are written quite nicely and easily accessible in Matlab.

Sincerely,
For both authors,

Ben Purinton
Universitaet Potsdam, Germany
purinton@uni-potsdam.de

References

Roering, J., Perron, J., and Kirchner, J.: Hillslope morphology and functional relationships between topographic relief and denudation, Earth and Planetary Science Letters, 264, 245-258, 2007.

Sofia, G., Pirotti, F., and Tarolli, P.: Variations in multiscale curvature distribution and signatures of LiDAR DTM errors, Earth Surface Processes and Landforms, 38, 1116-1134, 2013.

Grieve, S. W. D., Mudd, S. M., and Hurst, M. D.: How long is a hillslope?, Earth Surface Processes and Landforms, 41, 1039-1054, 2016a.

Grieve, S. W. D., Mudd, S. M., Hurst, M. D., and Milodowski, D. T.: A nondimensional framework for exploring the relief structure of landscapes, Earth Surface Dynamics, 4, 309-325, 2016b.

Grieve, S. W. D., Mudd, S. M., Milodowski, D. T., Clubb, F. J., and Furbish, D. J.: How does grid-resolution modulate the topographic expression of geomorphic processes?, Earth Surface Dynamics, 4, 627-653, 2016c.

Harel, M.-A., Mudd, S. M., and Attal, M.: Global analysis of the stream power law parameters based on worldwide ^{10}Be denudation rates, Geomorphology, 268, 184-196, 2016.

Liu, Z., Peng, M. and Di, K.: A continuative variable resolution digital elevation model for ground-based photogrammetry, Computers & Geosciences, Elsevier, 79, 71-79, 2014.