Response to Reviewer Comments: *Reviewer #1*

Journal: ESurf
Title: Quantifying uncertainty of remotely sensed topographic surveys for ephemeral gully channel monitoring
Author(s): Robert R. Wells et al.
MS No.: esurf-2017-3
MS Type: Research article
Special Issue: 4-D reconstruction of earth surface processes: multi-temporal and multi-spatial high resolution topography

***Interactive comment on* "Quantifying uncertainty of remotely sensed topographic surveys for ephemeral gully channel monitoring" *by* Robert R. Wells et al.**
Anonymous Referee #1
Received and published: 20 February 2017

The paper is interesting but in its present form, and based on the comments highlighted below, should be modified to improve the readability. The ideas could be interesting for the scientific communities, the methods and the assumptions resulted valid, but I suggest to improve some aspects of the analysis, the description of experiments and calculation. The reproduction of the experiment should be very difficult with the paper in this form. The conclusions resulted too general and no information repeatable and useful were provided. The authors should give proper credit to related work and the discussion (and conclusion) should clearly indicate their own new/original contribution. Therefore I suggest to improve the structure of the paper for a better reading, the discussion to compare the results obtained in the paper to others recent paper, the conclusions have to be enlarged because in present form resulted very insignificant. I have highlighted many of these instances below in the specific comments. For the language, I suggest to verify the comma in the text. All the formulae, figures and tables have to be verified. I highlighted some corrections, but probably my suggestions are not complete. I suggest to improve the references considering some recent paper that compared different survey technology for erosion studies.

Specific Comments (obtained from RC1-supplement):
P1L25: I suggest to improve the introduction and the discussion with more recent studies. For example:
Di Stefano, C., Ferro, V., Palmeri, V., Pampalone, V., Agnello, F., (2017) Testing the use of an image-based technique to measure gully erosion at Sparacia experimental area. Hydrological Processes, 31 (3), pp. 573-585.
Hayas, A., Vanwalleghem, T., Laguna, A., Penã, A., Giráldez, J.V., (2017) Reconstructing long-term gully dynamics in Mediterranean agricultural areas. Hydrology and Earth System Sciences, 21 (1), pp. 235-249.
Smith M.W., Vericat D., (2015) From experimental plots to experimental landscapes: Topography, erosion and deposition in sub-humid badlands from Structure-from-Motion photogrammetry. Earth Surface Processes and Landforms, 40 (12): 1656-1671.
Vinci, A., Todisco, F., Brigante, R., Mannocchi, F., Radicioni, F., (2017) A Smartphone camera for the Structure from Motion reconstruction for measuring soil surface variations and soil loss due to erosion. Hydrology Research, in press
Vinci, A., Brigante, R., Todisco, F., Mannocchi, F., Radicioni, F., (2015) Measuring rill erosion by laser scanning. Catena, 124, pp. 97-108.
Vinci, A., Todisco, F., Mannocchi, F., (2016) Calibration of manual measurements of rills using Terrestrial Laser Scanning. Catena, 140, pp. 164-168.
Wheaton, J.M., Brasington, J., Darby, S.E. and Sear, D.A., (2010) Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets. Earth Surf. Process. Landforms, 35: 136–156. doi:10.1002/esp.1886
Zheng, F., Xu, X., Qin, C., (2016) A Review of Gully Erosion Process Research. Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery, 47.

We agree and have updated the manuscript, especially the introduction and discussion, to include some of the suggested references listed by the reviewer, as well as other pertinent sources.

Casalí, J., Loizu, J., Campo, M.A., De Santisteban, L.M., and Álvarez-Mozos, J.: Accuracy of methods for field assessment of rill and ephemeral gully erosion, Catena, 67(2): 128-138, 2006.

Casalí, J., Giménez, R., and Campo-Bescos, M.A.: Gully geometry: What are we measuring?, The Soil, 1: 509-513, 2015.

Di Stefano, C., Ferro, V., Palmeri, V., and Pampalone, V.: Testing the use of an image-based technique to measure gully erosion at Sparacia experimental area, Hydrol. Proc., 31: 573-585, doi: 10.1002/hyp.11048, 2017.

Eitel, J.U.H., Williams, C.J., Vierling, L.A., Al-Hamdan, O.Z., and Pierson, F.B.: Suitability of terrestrial laser scanning for studying surface roughness effects on concentrated flow erosion processes in rangelands, Catena, 87: 398-407, 2011.

Gómez-Gutiérrez, A., Schnabel, S. Berenguer-Sempere, F., Lavado-Contador, F., Rubio-Delgado, J. Using 3D photo-reconstruction methods to estimate gully headcut erosion, Catena, 120, 91-101, 2014.

Micheletti, N., Chandler, J.H., Lane, S.N.: Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smart phone, Earth Surf. Process. Landforms, 40(4): 473–486, 2015.

Smith, M.W. and Vericat, D.: From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from structure-from-motion photogrammetry, 40: 1656-1671, 2015.

Vinci, A., Brgante, R., Todisco, F., Mannocchi, F., and Radicioni, F.: Measuring rill erosion by laser scanning, Catena, 124: 97-108. doi: 10.1016/j.catena.2014.09.003, 2015.

Wheaton, J.M., Brasington, J., Darby, S.E., and Sear, D.A.: Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets, Earth Surf. Process. Landforms, 35: 136-156, 2010.

P2L11: Why? What's the problem?

With traditional airborne LiDAR, point collection is 12-30 points per square meter (depending on flight height). The mechanics of rill and ephemeral gully erosion require much finer resolution information. The DEM will be too smooth and will miss the position of the channel as well as morphology of the channel.

P2L40: at the same time? together?

During the same period. Each platform was run independent of the others at different times on the same day. (P3L17-22)

"Furthermore, measurements of the same site using identical methods and equipment, but performed at different time periods can also lead to three-dimensional registration errors. Therefore, the scope of this work was to evaluate multiple survey techniques and provide a framework for temporal studies of ephemeral gully channels. Three surveying platforms, with varying parameters, were independently evaluated for locational accuracy and applicability in generating information for model development/validation. The objectives of this study are twofold: to quantify the overall accuracy of the different survey configurations and to develop practical guidelines for the design and implementation of future ephemeral gully monitoring studies."

P3L3: could be very interesting, but where are?

The recommendations are outlined in Sec 4.2.

"Long-term photogrammetric monitoring of ephemeral gullies should be performed with systems and procedures that: (1) Provide a minimum sampling density to capture the overall and local terrain characteristics based upon study objectives (i.e. temporal headcut migration process understanding may require sub-centimeter resolution data, while temporal channel meander process understanding may only require decimeter resolution) (James and Robson, 2012; Gómez-Gutiérrez et al., 2014). The planning phase of the project must consider the physical characteristics of the process to be investigated, study site physical and environmental variables, and the available hardware and software. (2) Utilize static ground control points visible in comparable photo pairs in all time-step surveys (i.e. fixed known points within the scene provide checks to assure proper three-dimensional registration of temporal data) (e.g. Smith and Vericat, 2015). An organized scheme for control points must be realized for detailed multi-temporal quantitative assessment. Small variations in alignment within temporal surveys will introduce error into length, width, cross-sectional area and volume estimates (e.g. Casalí et al., 2015). Repeated realizations of GCP coordinates will always reduce error in survey solutions. (3) Collect the same number of photo pairs using the same sensor and with the same orientation in all time-step surveys (i.e. data collection strategies should not vary temporally and new sensors must be carefully calibrated to preexisting datasets). Consistency in photo collection (i.e. scheduling and no. of photo pairs) will enhance the comparison of temporal solutions (Gómez-Gutiérrez et al., 2014). Also, consider site visits at a particular time of day. (4) Process and generate photogrammetric solutions using the same software package and similar input processing parameters. And, a calibrated camera will always yield better solutions."

P3L27: Uppercase, verify in the text.

We agree and have made the correction.

P4L3: This technique was widely used, it is necessary this explanation? In the paper were used techniques less used and probably more interesting. I suggest a briefly description of the Terrestrial LIDAR. Probably the authors could be some photos of the equipment used for the paper.
The TLS was used as the "reference" and, this distinction required a more detailed description.

P6L23-24: where are the results of this analysis?
The results of this analysis and a more detailed description have been added to the manuscript. (P7L9-24; P9L8-20; Figure 8) Figure 8 is new to the revision.
"Two metrics were used to quantify the spatial pattern distribution: distances between events and between events and random points not in the pattern (void space). The G-Function, $G(r)$, defined as the cumulative frequency distribution of nearest-neighbor distances (Lloyd, 2010), provides the conditional probability that the distance between points (event-event) is less than the point distance threshold (r). The empirical distribution is obtained for each distance r by counting the number of points at distances less than or equal to r from each point within the AOI. The theoretical distribution is obtained by assuming a completely random pattern with density (λ; estimated by the ratio of total number of points, divided by the area of the AOI), modeled as a Poisson process. Empirical values closer to the theoretical values indicate a random distribution, empirical values above the theoretical values indicate clustering, and empirical values below the theoretical values indicate a more regular distribution (Bivand et al., 2008). The F-Function, $F(r)$, defined as the cumulative frequency distribution of the distance to the nearest point in the AOI from random locations not represented within the AOI (Lloyd, 2010), provides the probability of observing at least one point (event) closer than r to an arbitrary point within the AOI ("empty space" or "void" distances). Estimated and theoretical distributions are obtained similar to that of the G-function. The interpretation of graphed observed versus theoretical values indicates a regular pattern when the observed is above the theoretical values and clustering when it is below (Bivant et al., 2008). These point pattern analyses were performed using the Spatstat package in the R software package (Baddeley and Turner, 2005; Bivand et al., 2008).
Furthermore, points with the same X, Y and Z to the fifth decimal place were removed from the LiDAR dataset, indicating collection of redundancy information."
"Point pattern analysis was examined using the G- and F-functions (Figure 8). Each tests through an assumption of completely spatially random (homogeneous Poisson process), although interpretations for clustering and regularity are in opposition for each test (i.e. regular point spacing outcome for the G-function is below gray confidence bounds and for the F-function it is above). At first, the confidence bounds (gray envelope bounding the theoretical values (red dash line)) show that the Fixed_122 has a sparse point count. Looking at the G-function results, the data are clustered at distances of $0.02\ m$ (Fixed_122), $0.002\ m$ (Quad_20) and $0.001\ m$ (Grd_8A), then regularly distributed. This indicates that small distances occur less often than expected under the assumption of spatial randomness. The F-function results say that the data are randomly distributed to $0.06\ m$ (Fixed_122) and $0.025\ m$ (Quad_20), and clustered for Grd_8A (i.e. on short distances, fewer points are encountered than for a random pattern); although, the scale of r should be acknowledged ($< 3\ mm$). All twelve datasets yielded observed G-function values below the theoretical values, indicating a regular sampling pattern. The terrestrial photogrammetric surveys did show slight clustering at small distances ($< \ 3mm$). Therefore, based on these metrics, a regular sampling pattern was observed, indicating that all locations within the study area were sampled with a similar sampling intensity (no areas where over sampled nor under-sampled) for all twelve datasets in this study."

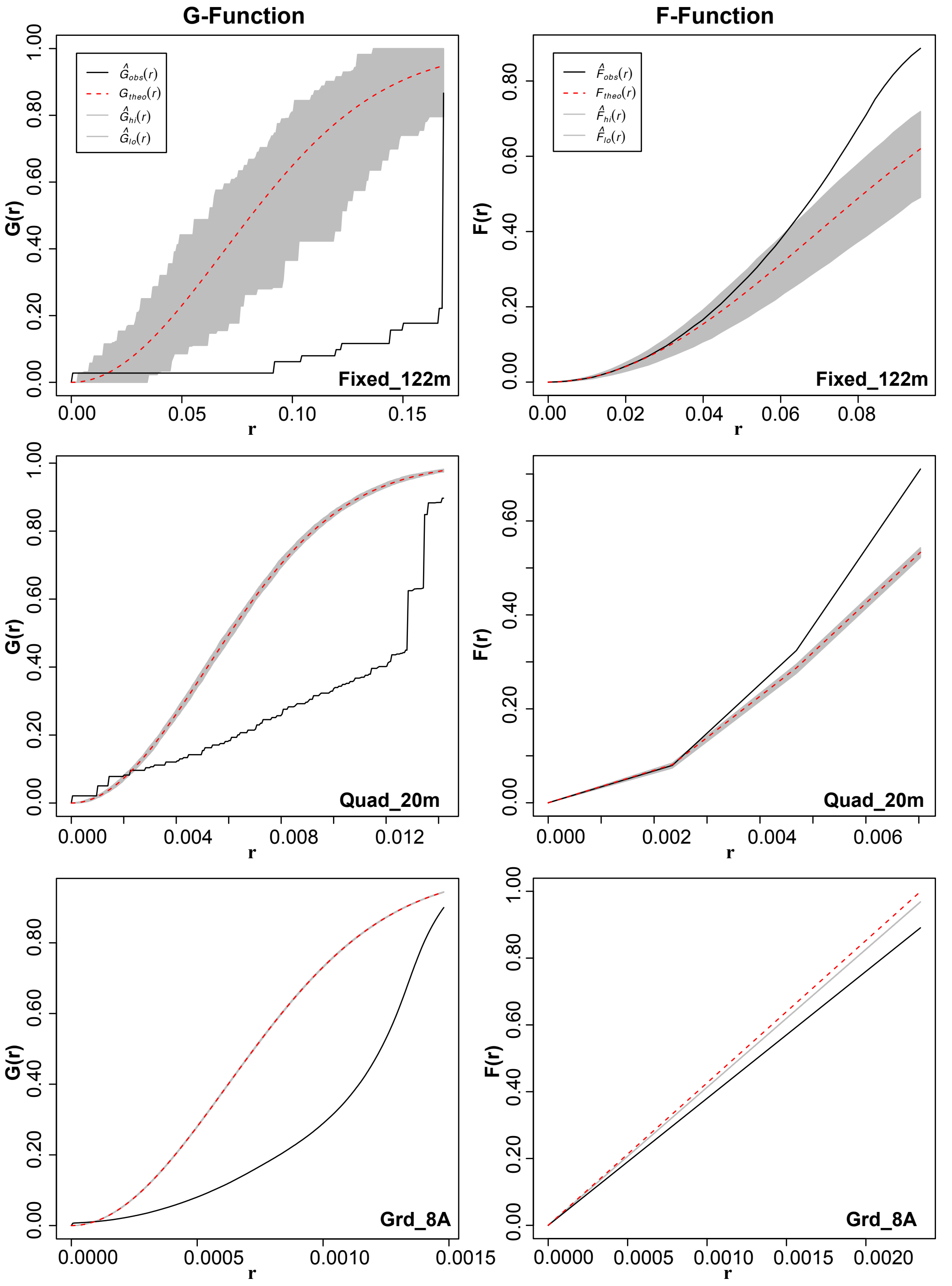

Figure 8: Results of G-Function and F-Function analysis for Fixed_122m, Quad_20m, and Grd_8A datasets.

P7L20: Verify the units of the equations in all the paper.

The reviewer is correct. In equation 4 and 5, the elevation difference is multiplied by the cell area. Here, the cell area is 2.5E-5 square meters. The equations have been corrected. (P8L11-14)

P7L22: what is cs? It is cz?

As a continuance from the previous comment, the variable in question is *ca*, the cell area. The change has been made. (P8L13)

P8L5: In this table were reported the evaluation category, not the score.
The reviewer is correct. It was not intended to provide the scores within the materials section. Table 7 has been moved to Table 2 and provides a description of the metric tests as stated within the text of this section. Scores are provided in Table 8 within the results section.

P8L10: I suggest to better explain the influence of the sampling intensity. I suggest to insert a classification of "low", "medium", "high" sampling intensity. Therefore, in this paragraph should be reported the results of the point pattern analysis.
We agree and apologize for the confusion concerning the point pattern analysis. As mentioned in a previous comment, we enhanced the section to further describe the F- and G-functions. (P7L9-24; P9L8-20; Figure 8) However, we do not share the reviewer's enthusiasm for ranking the intensity (i.e. low, medium and high).

P8L12: How? Probably a comment of a figure 11 or a better explanation could be useful.
Further discussion of the cross-section information is provided in Sec 3.4. (P10L33-P11L5) Figure 11 is now Figure 12.
"In the gridded elevation evaluations (Min, Max, Mean; Table 5; Figure 12), absolute difference from LiDAR was less than 0.02%, and these differences were only seen in the fixed-wing flights. The variance (i.e. roughness), however, does shoe that absolute differences from LiDAR were 131% (Fixed_122), 23% (Fixed_61), and 9% (Grround_4A). Comparing elevation information (Table 7; Figure 12) between photogrammetric cross-sections and LiDAR cross-sections through linear regression, indicates a coefficient of determination larger than 0.98 for all datasets, excluding the two fixed-wing flights. The standard error of this regression was less than 10 mm for Quad_20, Quad_35, Ground_2B, Ground_4B, Ground_4C, Ground_6A, and Ground_8A. Following that, Ground_2A and Ground_4A had a standard error of approximately 17 mm and the two fixed wings had a standard error of 25 mm. The average area percent difference for all cross-sections were within 1.5%, while the two fixed wings had 3% (Fixed_61) and 8% (Fixed_122). It is important to mention that the range of area percent difference is within ± 2%, while the fixed-wing systems had up to 15% percent difference. The error is huge for instance if this dataset was intended to be used for the purpose of development/calibration/validation of a soil erosion model."
The effect of sampling intensity can be explained as this: let us say you have one square meter and in one survey you collect 10 points and in another survey you collect a thousand points. If the surface of this square meter is a plane then both surveys will tell you the same thing; however, if the surface is rough, the survey with 10 points will not help you describe the terrain, as you may only collect peak or trough values, where the 1000 point survey will more effectively convey the characteristics of the terrain, the micro-topography (see papers by Haubrock et al., 2009 and Eitel et al., 2011 for discussions of surface roughness). So, if the surface is rough, low point counts will bias the volume estimates.

Eitel, J.U.H., Williams, C.J., Vierling, L.A., Al-Hamdan, O.Z., Pierson, F.B.: Suitability of terrestrial laser scanning for studying surface roughness effects on concentrated flow erosion processes in rangelands, Catena, 87: 398-407, 2011.

Haubrock, S.-N., Kuhnert, M., Chabrillat, S., Güntner, A., Kaufmann, H.: Spatiotemporal variations of soil surface roughness from in-situ laser scanning, Catena, 79: 128-139, 2009.

P8L27: What's mean? Clarify.
The two fixed-wing flights could not be adjusted by the methods described earlier within the text (see Sec. 2.3), since neither point cloud contained points on the channel GCPs. Therefore, the seemingly large deviation in elevation difference may be simply an alignment issue that we could not resolve/control or it is a bias we introduced in the raster process, since the results of the point analysis show that the fixed-wing data is clustered below 20 mm.

P10L3: I suggest to compare the results obtained with other recent papers.
We agree and have included appropriate references within the discussion for comparison.

Casalí, J., Giménez, R., and Campo-Bescos, M.A.: Gully geometry: What are we measuring?, The Soil, 1: 509-513, 2015.

Di Stefano, C., Ferro, V., Palmeri, V., and Pampalone, V.: Testing the use of an image-based technique to measure gully erosion at Sparacia experimental area, Hydrol. Proc., 31: 573-585, doi: 10.1002/hyp.11048, 2017.

Gómez-Gutiérrez, A., Schnabel, S. Berenguer-Sempere, F., Lavado-Contador, F., Rubio-Delgado, J. Using 3D photo-reconstruction methods to estimate gully headcut erosion, Catena, 120, 91-101, 2014.

Smith, M.W. and Vericat, D.: From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from structure-from-motion photogrammetry, 40: 1656-1671, 2015.

Wheaton, J.M., Brasington, J., Darby, S.E., and Sear, D.A.: Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets, Earth Surf. Process. Landforms, 35: 136-156, 2010.

P10L20: Where is the Sec. 4.2.1?? For the point pattern analysis see the comments reported above.
Dyslexia. The reference should have been 2.4.1. We made the correction and enhanced the explanation for the point pattern analysis (P7L9-24; P9L8-20; Figure 8).

P11L1: I suggest to improve the guidelines for give more useful information.
The guidelines were intended to be broad, as many will not be interested in sub-centimeter characterizations. However, we did add some more information from our experiences to enhance the section. Because each individual will study a different problem, it is not possible to develop a cookbook that works for everyone. Instead, we described a path to minimize problems in processing and future temporal comparisons. We have added more details to our recommendations in this section. Our main message concerns GCPs and consistency. (P12L36-P13L12)
"Long-term photogrammetric monitoring of ephemeral gullies should be performed with systems and procedures that: (1) Provide a minimum sampling density to capture the overall and local terrain characteristics based upon study objectives (i.e. temporal headcut migration process understanding may require sub-centimeter resolution data, while temporal channel meander process understanding may only require decimeter resolution) (James and Robson, 2012; Gómez-Gutiérrez et al., 2014). The planning phase of the project must consider the physical characteristics of the process to be investigated, study site physical and environmental variables, and the available hardware and software. (2) Utilize static ground control points visible in comparable photo pairs in all time-step surveys (i.e. fixed known points within the scene provide checks to assure proper three-dimensional registration of temporal data) (e.g. Smith and Vericat, 2015). An organized scheme for control points must be realized for detailed multi-temporal quantitative assessment. Small variations in alignment within temporal surveys will introduce error into length, width, cross-sectional area and volume estimates (e.g. Casalí et al., 2015). Repeated realizations of GCP coordinates will always reduce error in survey solutions. (3) Collect the same number of photo pairs using the same sensor and with the same orientation in all time-step surveys (i.e. data collection strategies should not vary temporally and new sensors must be carefully calibrated to preexisting datasets). Consistency in photo collection (i.e. scheduling and no. of photo pairs) will enhance the comparison of temporal solutions (Gómez-Gutiérrez et al., 2014). Also, consider site visits at a particular time of day. (4) Process and generate photogrammetric solutions using the same software package and similar input processing parameters. And, a calibrated camera will always yield better solutions."

P11L16: See the comments reported above.
See responses reported above.