Dear Editor G. Sofia,

we revised the paper following the detailed and constructive suggestions of the reviewers. We thanks Dr. M.D. Hurst, an anonymous reviewer and Dr. L. Marchi for appreciating the paper and for the valuable suggestions that permitted to improve the paper. In the following lines, we report the replies to the reviewers' comments. As requested, the corresponding changes in the manuscript are highlighted yellow or marked by means of track changes in the document in word format.

Sincerely,

Sebastiano Trevisani

Reply to reviewer Dr. M.D. Hurst

We thanks very much Dr. M.D.Hurst for the positive and precise review of our paper; his suggestions and comments have been very useful for improving the paper and for highlighting relevant aspects. In the following lines, we address the highlighted critical points.

General comment 1.

Reviewer:
"I have some minor suggestions to improve the manuscript and data presentation. In particular, I suggest the use of tables to present data that would be much better plotted graphically be addressed. The figures plotting spatially distributed results are consistently too small to identify details discussed in the text. Perhaps the authors might need to review the figure guidelines in terms of figure size, resolution and font size, the editor can also advise."

Authors:
As reported in the replies to specific comments, we added graphical plots of the data to improve data interpretation as also requested by Reviewer #1. Concerning the figures, we increased the resolution and, where requested, we used a larger font size.

General comment 2

Reviewer:
"I attempted to download the MAD software from the 2015 paper in Computers and Geosciences but found that the zip file was invalid. I would suggest the authors correct this and provide the link somewhere, or alternatively providing a link to the github page where the code is maintained so that other scientists can apply these methods. A "code and data availability" statement with the acknowledgements will suffice for this."

Authors:

We thank the reviewer for reporting this issue. We added the link to the GitHub page of the tool in the paper (in the new text at page 5 - line 7) (https://github.com/cageo/Trevisani-2015) and in the acknowledgements.

Specific comments

Reviewer:P1401 L5: I suggest changing "Ad hoc" to "bespoke"; Similar P1404 L 11.

Authors: We rephrased accordingly.

Reviewer: P1402 L20: Could also cite DiBiase et al. 2010 and Milodowski et al. 2016 (currently an ESURF discussion paper) have used isotropic surface roughness map bedrock outcrops.

Authors: We prefer to not add the suggested references, because of the references that we inserted in that point of the paper provide quite comprehensive reviews of roughness indices used in scientific literature more than specific applications of a single index (as in the suggested additional references). Moreover, referencing to Milodonwski et al. 2016 paper could be redundant since a similar approach using surface roughness to detect rocky outcrops (and other morphologies) is reported in the Trevisani et al. 2012 paper, that we already mentioned in our work.

Trevisani, S., Cavalli, M. & Marchi, L. 2012, "Surface texture analysis of a high-resolution DTM: Interpreting an alpine basin", Geomorphology, vol. 161-162, pp. 26-39.

Reviewer: P1404 L16: A physical description of what h represents is required. By "lag" do you mean the search distance in the local neighbourhood? Or is it the search direction? I find the way these methods are presented difficult to penetrate but conceptually relatively simple. I think the authors could spend some time refining the description of the methodology.

Authors: We rephrased the sentence trying to be clearer in the explanation. The lag **h** is just another term for the separation vector **h**; the lag **h** is at the base of the selection of the sample pairs used for Variogram or Mad calculation; it is not the radius of the search neighborhood (i.e., moving window) used to search the pairs of values separated by the lag **h**. See page 5 lines 9-14 in the new text.

Reviewer: P1404: Eq 1: nalpha is not defined anywhere Figures

Authors: alpha is an index, is defined at line 20 page 1404, alpha= 1,…,N(**h**).

Reviewer: P1406 L1-2: This smoothing method will result in systematic bias on ridges and in valleys towards positive and negative residuals respectively, since smoothing will lower ridgelines and raise valleys.

Authors: We are aware of this behavior. It is a characteristic inherent to smoothing methods. From the perspective of DTM residual derivation, negative anomalies in incisions and positive anomalies on ridges is what we aim to derive. Given the high sensitivity of the mean estimator to extreme values, in some circumstances the mean surface tends to follow too much ridges and valleys, leading to a low-variance residual DTM. For this reason some authors (see Hillier et al. 2008, reported in the paper references) use the local median instead of the local mean (from your perspective still giving more bias). An alternative to the median approach is based on multi-pass moving averages (see for example Trevisani et al., 2010), used in study site 1, which reduces the impact of the high sensitivity of mean estimator.

Reviewer: P1406 L7: "any direction" should be all directions.

Authors: we corrected accordingly

Reviewer: P1406 L9: So are Riso and Rflow both just values of MAD but changing the neighborhood search to only look in the D8 flow direction for the latter? This is not clear.

Authors: The rephrasing of P1404 L16 should clarify the difference between the separation vector **h** (or lag **h**) and the search neighborhood. Consequently, the interpretation in the present form should be more straightforward. For calculating R_{iso} we use all the pairs in the neighborhood separated by h in all directions (i.e., we do not consider along which direction the sample pairs are aligned, but only their separation distance); for calculating R_{flow} we select only the pairs in the neighborhood separated by the vector **h**, with the **h** aligned with the local flow direction. We also modified the caption of figure 2 because it could be misunderstanding; in particular we clarified that the dashed circle around the highlighted pixel is reported to show the pairs separated by the modulus of h (i.e., the diameter is equal to the modulus of **h**), not the search neighborhood.

Reviewer P1406 L11: D8 limits the flow directional analysis to 8 directional, 45 degree bins. This is a significant limitation that should be discussed further. There are plenty of alternatives (e.g. polynomial-derived aspect, or d-inf).

Authors: We highlighted that the proposed approach is a simplified one. Moreover, in the following lines where we argue, "the methodology can be extended using directions determined with other approaches" we included also d-inf as a potential example. In the new text page 6, lines 21-24.

Reviewer: P1411 L2: Was this done? I presume this is a qualitative confirmation process. Please be more specific, or delete this.

Authors: We modified the sentence to clarify the meaning: “site-specific field surveys have been carried out for confirming the results”. In the new text page 10, line 25.

Reviewer P1412 L27: “have to be evaluated critically” so where the differences are on the order of the DTM accuracy then roughness differences could just be artefacts. Earlier you reported the vertical accuracy at 0.15m-0.3m depending on the dataset, so should you consider roughness differences at smaller values than this?

Authors: The answer is yes, from manifold perspectives. We understand the point raised by the reviewer. For this reason in various parts of the paper we referred to DTM accuracy, its spatial characteristics (e.g., spatially correlated error or not) and possible impacts on roughness calculation. Moreover, it have to be considered that for study site 1 the upscaling to 2 m pixels DTM, in the low lying areas, brings to an hypothetical accuracy (assuming an uncorrelated Gaussian error) of 0.075 m. We do not think that we can add more information regard this point in the paper; however, given the opportunity of the open discussion, we provide in the following lines some considerations to address this point.

- Even if the reported accuracy of the DTM is only an indicative value (e.g., can be locally higher or lower than reported, and in no way it takes into consideration if the error is spatially correlated or not, etc.) we can formulate some hypothesis based on declared accuracies. A first hypothesis, the worst scenario for roughness calculation, is to have a spatially uncorrelated error. As a simple case, let‘s assume to have a spatially uncorrelated Gaussian random error with standard deviation of 0.15 m (the hypothesis for low lying areas in case study 1 could be 0.075 m), corresponding to a variance of 0.0225 m^2. This means that the differences between residual elevations, being related to a difference of two random variables, should have a related error variance of 0.045 m^2 (a standard deviation of 0.212132 m). Let’s now assume to be on a flat and smooth surface, so as the differences in residual elevation between pairs of values are only related to the uncorrelated error (and independent to the modulus of separation vector, being a white noise). The differences can be viewed as generated by a Gaussian random variable with 0 mean and standard deviation of 0.212132 m. Given this, it can be shown numerically (we used simulated values, we can furnish the R script used for these experiments if needed) as the standard deviation of median estimates of the absolute differences, using 29 samples (i.e., as for the MAD index calculated in the flow direction with a search radius of 3 pixels) is approximately of 0.031 m (0.0153 m for the areas with an accuracy of 0.075m): a much lower value (i.e., a higher accuracy) than the reported accuracy of the DTM.
- Another relevant point of the hypothesis of uncorrelated Gaussian random error with 0 mean, is that the median of absolute differences is related to the standard deviation of differences: i.e., larger the standard deviation of differences larger is the median of absolute differences. In case of a Gaussian random error, it is possible to show numerically that the median of absolute differences is approximately 0.674 time the standard deviation of differences. So, in case of a standard deviation of differences of 0.212132 m, we should expect a median of absolute differences (i.e., a MAD index) of 0.143 m. Accordingly, if the hypothesis of an uncorrelated random error is true, we should find minimum MAD values not much smaller than 0.143 m. Differently, we observe that MAD, especially in correspondence of gentler slopes, can have much smaller values; this suggests that the error can be

spatially correlated or that the accuracy in the representation of local morphology, in terms of spatial variability, is higher than expected. The last point is reasonable given that generally the accuracy is evaluated at the pixel scale (sometimes directly on LiDAR points cloud) comparing the pixel elevation with the "true" elevation: a small horizontal shift of the pixels induces large elevation errors. However, here we are evaluating the spatial variability on a neighborhood and in particular, the relative differences between neighboring pixels: in this context, a horizontal shift slightly affects measures of spatial variability.

We think, beyond statistical reasoning, that the best we can do to critically evaluate the results of this analysis, is to explore the spatial patterns of differences and relative differences in roughness and compare these patterns with fine-scale morphologies. Moreover, as reported in figure 13, in some circumstances, small differences in the two roughness indices, near the accuracy estimated for MAD indices, can still be useful from the interpretative viewpoint.

Reviewer: P1413 L1-2: Difficult to see this without a nice plot to look at. Please plot this data. What does a positive skew mean in terms of the landscape? Flow direction roughness tends to be lower than isotropic? That's neat if I've interpreted correctly so you could make more of this result.

Authors. We added on fig.10 the boxplots of differences and of relative differences. The slight, almost imperceptible, prevalence of positive differences means that flow-directional roughness, on the whole area, tends to be higher than isotropic (differently from study site 2). Clearly, looking at the table and the boxplots, there is an almost complete balance between positive and negative differences; this is expected since the area studied has a very large extent and covers a high variety of morphologies, some with higher R_{flow} than R_{iso} and some with R_{iso} higher than R_{flow}.

Reviewer: P1416 L22-24: I disagree with this statement. Looking at Fig 14b there are areas of both high and low relative differences adjacent to the channel and towards the divides headwaters.

Authors: The reviewer is right: the sentence is wrong and we rephrased it. The correct sentence (In the new text page 15, line 19), is "the analysis of relative differences versus DCiso indicates that there is a prevalence of positive differences in areas of lower connectivity".

Reviewer: P1417 L1: Does your method predict a more or less connected landscape overall? It's obvious from fig 17d that it predicts greater connectivity but you should state this explicitly if so.

Authors: We rephrased the sentence accordingly (In the new text page 15, lines 28-31). We added the sentence:
"The use of flow-directional roughness permits to describe better the connectivity in areas with a prevalence of erosional processes. In fact, in correspondence of gullies R_{flow} is lower than R_{iso}, because of it is not affected by the high variability of the slopes and channel banks."

Reviewer: P1418 L12: e.g. Wavelets (Lashermes et al 2007) or FFT (Perron et al 2008).

Authors: We prefer to not add the suggested references both because of we have too many references as well as because of we are thinking that the methodology to be followed to derive the residual DTM should be related to the target of the study and to the characteristics of the processes to be modeled. The decomposition according to wavelets or in frequency components via fast Fourier transform is interesting but at the end produces values that are not easily interpretable in term of processes.

Reviewer: P1418 L5-25: Discussion of future research avenues should come at the end of the discussion. This is not a conclusion of your work. The conclusion should highlight your main findings.

Authors. We prefer to maintain this part on “future research” at the end of the paper in order to give more relevance to these considerations

Reviewer: Fig 4: Figure text to small.

Authors: we improved the figure

Reviewer Fig 8: There seems to be significantly more negative residual than positive (I can’t see many white pixels but there are plenty near-black). This may be my eyes! A color image rather than grey-scale would be helpful and a CDF plot (i.e. showing the data in table 1) would also be helpful.

Authors: we improved the figure and we inserted a boxplot. We opted (here and for the other tables) for a boxplot because of a histogram or a CDF is not easily interpretable given the high kurtosis (and for some also skewness) of the presented distributions (the alternative would be to derive a histogram of trimmed data). Unfortunately, even if we inserted a color figure, it is difficult to appreciate a residual DTM without zooming it since it reproduces the high-frequency component of spatial variability.

Reviewer: Fig 10-13: These are difficult to interpret when printed as too small, and not high enough resolution when zoomed on a comp. I would want these will be bigger/higher resolution in the final paper.

Authors: We improved the resolution of these images and we inserted boxplots of the related distributions.

Reviewwer: Fig 13: Caption “hortophotos” typo.

Authors: We corrected the typo.

Reviewer:Fig 14 and 15: What is the blue line? Channels defined how?

Authors: we inserted the stream in the legend. In regard to the procedure followed for defining the streams we already inserted the references to the previous work in which the target channels have been defined (page 14 lines 23-25, in the new text)

Reviewer: Fig 17: This really demonstrates the application of your approach nicely!

Authors: Thank you!

Reviewer: Table 1: This could be better represented with a box and whisker or a cumulative probability plot, perhaps as an inset to Figure 8.

Authors: we inserted a boxplot

Reviewer: Table 2: Again I'd like to see a plot of CDF with different coloured lines for each method.

Authors: we added boxplots in the corresponding maps

Reviewer Table 3: Plot the data!

Authors: we added boxplot in the corresponding map

Reviewer: Table 4: CDF plots or similar

Authors: we integrated the table with boxplots in the corresponding maps

Replies to specific comments of reviewer 1 (reply to general comments already published in open discussion)

Reviewer P1400L5: Please specify which process and factors

Authors: We are referring generically to the different geomorphic processes and factor that can be potentially involved in the shaping of landscape.

Reviewer P1402L18. I would not present figures in the introduction. If you need do describe this, do it in material and methods.

Authors: We prefer to maintain here the figure, because of it is a conceptual one which describes the key ideas of the paper. In the method section, we prefer to focus on quantitative and mathematical aspects related to the implementation of these ideas.

Reviewer. P1402L27-28 P1403 L1-6. You can delete this. be objective in the introduction. No need to describe so many previous works.

Authors. We removed the sentence related to previous works (In the new marked text page 3, lines 24-29). However, one of the main motivations that inspired us to develop flow-directional roughness is the anisotropy in surface roughness that often is a characteristic and distinctive component of fine-scale morphologies (as highlighted in the cited paper).

Trevisani, S., Cavalli, M. & Marchi, L. 2012, "Surface texture analysis of a high-resolution DTM: Interpreting an alpine basin", Geomorphology, vol. 161-162, pp. 26-39.

Reviewer P1403 L15-24. Divide this in aim and specific objectives.

Authors: We rephrased to include specific objectives (In the new text page 4, lines 11-15).

Reviewer P1403 L25-29. Move this to materials and methods.

Authors: We moved in materials and methods and rephrased accordingly (in the new text page 4, lines 26-30).

Reviewer P1404 L9-10. Delete this. You mentioned if before.

Authors: Deleted.

Reviewer P1409L13. Dataset description should come before MAD and computation of flow-directional roughness

Authors: We prefer to maintain the present structure of the paper because of the datasets presentation is functional to the use of the developed methodologies, which are the kernel of material and methods section.

Reviewer P1411 L20 Is this the Root Mean Square Error? If yes write it

Authors: Done

Reviewer P1412L11 Is it possible to provide an histogram of the errors, instead of a table. it would be easier to see the distribution.

Authors: As also requested by reviewer 1, we inserted for all the tables a related boxplot. However, we are not dealing with errors but with a residual DTM

Reviewer P1412L15 and L21-22 This needs to be measured and ploted. Would be possible to provide the correlation between both indexes?

Authors: We reported the correlation coefficient (0.966, at page 11 line 29 in the new text). We avoid to insert a scatterplot given the strong correlation, that is expected since both indices are measuring spatial variability of residual elevations.

Reviewer P1412L27-29 You could compare if these differences are statistically significant. This is other reason why a scatterplot showing the correlation between these indexes would be helpful

Authors. We think that this kind of test performed on a global level would be not particularly explanatory. We think that the crucial point is to analyze these differences with morphologies locally. Moreover, see reply to M.D. Hurst regarding DTM accuracy.

Reviewer P1413 L3 A Morans I analysis would help to see if these errors have some specific spatial pattern (dispersed, random or clustered)

Authors: We are not dealing with errors (e.g., coming from an interpolation). Moreover, a Morans I analysis will be not too much useful to describe the patterns of differences at global level. First of all, because of the important point is to analyze the patterns of differences in comparison to local morphologies. Secondly, the patterns of differences are too complex to be described by means of a Moran I analysis (for example, the patterns are characterized by high anisotropy). We should use a variogram map (or MAD), but, again, not computed globally but on a moving search.

Reviewer P1415L5-6 Can you provide the RMSE results of both indexes?

Authors: We cannot provide the RMSE of the indexes because of we did not performed an interpolation from which calculate a RMSE (e.g., by means of cross-validation).