We firstly thank both reviewers for agreeing to review the manuscript and appreciating its originality. Their thoughtful comments helped improving the revised-version.

As the four main concerns of both reviewers (implying a major revision) are broadly consistent (text in black), we first propose a concise answer to each of these (text in blue). A detailed answer to each reviewers' comment is provided afterwards. Four new supplementary materials (A-B-C-D) are attached to our revisions as well.

1) SV-error is not put to good use because it is averaged over all of our sub-reach.

We agree with this remark. We changed our methodology accordingly, which now uses a "node-specific" error. Supplement A and B illustrates and describes our new methodology, respectively.

2) The methodology is not clear enough, especially the way the channels are translated at each run.

We agree that the first draft did not allow the readers to fully understand the methodology. We thus produced a new figure (Supplement A) and rewrote the methodology accordingly (Supplement B). We think it is now possible to thoroughly understand how channels are translated (directions and values) at each MC run. We also propose to integrate Supplement A into a new figure of the revised-manuscript.

3) The use of the proposed SoD is not correct. Reviewers ask for a threshold of change detection.

We agree with this remark. The term SoD has been abandoned. We think it was potentially misleading for readers as it was technically not a detection threshold. We now propose focusing on a well-known and reliable uncertainty indicator: the relative percentage of uncertainty. It allows us to describe variability (and so uncertainty) of our MC results.

Firstly, the relative percentage of (total) uncertainty is calculated as following: 0.5x(max-min)x100/mean

Secondly, we propose the following significance threshold: percentage of uncertainty > 50%. Indeed, if the percentage of uncertainty exceeds 50%, the mean value is then necessarily lower than the total range of measured value (max-min), which leads to the assumption that the uncertainty is too high to consider the change (mean value) as significant.

4) Consequence of the last three points: The hypotheses, as stated (except hypothesis 1), cannot be verified.

We agree with this remark. We thereby reformulate the second and third hypothesis, according to the new methodology, the new results and reviewers' concerns.

They are:

1. Orthophotos are affected by a local significant SV-error;
2. SV-error highly affects variability of MC simulated measurements of eroded and/or deposited surfaces;
3. Uncertainty of surficial changes depends on their magnitude

Referee 1
General comments: This paper addresses an important topic in fluvial geomorphology, analyzing channel change from time series of remotely sensed data, and offers a new perspective on evaluating the uncertainty inherent to this approach. Recent studies have shown that image co-registration errors are spatially variable and this paper takes an additional step by performing Monte Carlo simulations to assess the significance of observed changes in channel planform. While the idea has merit, I have some serious reservations about the way the approach is implemented in this study.

We thank the reviewer for appreciating the originality of our study. We think that his/her reservations about the methodology are relevant and we address them in the replies below.

1) The authors produce a continuous spatially variable error surface but then aggregate the error over a reach-scale by spatial averaging, which is an unnecessary loss of information.

We agree with this remark. Whilst we initially thought that error aggregation was necessary to avoid channel distortion, we managed to find a technically possible and geomorphologically valid way to assign a specific error to each node. Details of our new methodological approach is illustrated in supplement A, the latter being incorporated as a supplementary figure in the revised manuscript (either in the main text or as supplementary material). The part 3.4.1 in the initial manuscript has been rewritten accordingly. It is attached as Supplement B.

We think this allows a precise understanding of our methodology, as we are fully aware that it is imperative in this kind of study.

2) The manner in which nodes of digitized bank lines (are moved?) is not explained well and might be conceptually flawed.

We agree with this remark : see Supplement A providing the details of our new methodological approach (and see our reply above). Our "old" methodology consisted in geometrically translating nodes with a unique value for each nodes. This has been changed by using node-specific errors.

3) The surface of detection introduced by the authors should be used as a threshold, not subtracted from the observed changes.

We agree with reviewer 1. We believe that comparing our SoD to Lea and Legleiter's (2016) LoD indeed was a mistake, as the SoD is technically not comparable to the LoD. We therefore propose to leave the term SoD as it might be confusing for readers.

Instead, we now use in the revised manuscript the relative percentage of uncertainty associated with the variability of the measured changes through the MC runs. Because it is derived from the confidence interval, which is a simple and well-known indicator of variability, we think it will simplify the interpretation and comparison of our results.
The percentage of (total) uncertainty is calculated as following: 0.5x(max-min)x100/mean
Then, the (conservative) significance threshold becomes: percentage of uncertainty > 50%. Indeed, if the percentage of uncertainty exceeds 50%, the mean value is then necessarily lower than the total range of measured value (max-min), which leads to the

assumption that the uncertainty is too high to consider the change (mean value) as significant.
Part 3.4.3 of the current manuscript has been rewritten accordingly.

Those are the key issues, but please refer to the attached PDF for more detailed comments and text edits. While I think this manuscript has potential, the authors must address the concerns listed above, as well as the various minor edits, before the paper can be published.

We thank the reviewer 1 for highlighting the potential of our manuscript. We believe his concerns have greatly improved the revised manuscript.

Specific comments:

Title: I recommend modifying the title to emphasize that you are accounting for spatially variable error

The title now is : *Short communication: Measuring river planform changes from remotely-sensed data: A Monte-Carlo approach to assess the impact of spatially-variable geometric error*.

Title: I think surficial is a more commonly used term than surfacic, which I have never seen before, so please replace this with surficial throughout.

We modified this term as suggested.

Page 1 Line 1: The more common phrase is remotely sensed data, so please replace throughout.

We replaced it throughout.

Page 1 Line 8: You need to be clear what you mean by this: is it an area where erosion occurred, followed by deposition?

We mean an area where erosion first occurred followed by deposition, as illustrated on Figure 1. We specified it in the revised manuscript as following: "(i.e. *quantification of eroded, deposited, or eroded then deposited surfaces)*" (Page 1 Line 8).
We also specified it in Figure 1 by modifying the legend as following: "*Erosion then deposition*".

Couldn't that lead to no net change and thus not be detectable even if the images were perfect?

It could actually lead to "no net change" only if the channel left its original position and then recovered its original position. This would indeed be undetectable with photos collected before and after the migration. However, according to the evolution of the active channel between both orthophotos, we believe this situation did not happen.

Page 1 Line 18: What is the distinction between these two?

Contrary to exclusively co-registered aerial photos, orthophotos have undergone an orthorectification process. We propose to modify the sentence as following: "*Taking the SV-error into account is strongly recommended even on orthorectified aerial photos, especially in the case of mid-sized rivers* ..." (Page 1 Line 18).

Page 2 Line 4: Missing page numbers in Cadol reference

Page numbers have been added.

Page 2 Line 5: Only use author's last name: Lauer, not Wesley Lauer - please change throughout

This has been changed throughout.

Page 2 Line 18: Replace this with co-registered throughout

This has been replaced throughout.

Page 2 Line 27: Include a URL for this reference.

URL has been added.

Page 3 Line 28: What were the discharges for these two time periods? If the flow was much different, that could lead to a false impression of erosion or deposition.

Unfortunately, the river wasn't gauged before 1965. Nevertheless, as we strictly delineated the channels by referring to the active channel concept, we believe that confusion was not possible. Moreover, according to the observation of both orthophotos, the Bruche river wasn't at bankfull stage at these dates.

Page 4 Line 7: This figure should use one of the photos, probably the 2015 image used as a base, not the lidar hillshade. The lidar is not even used in this study and seeing the photo used to define the control points would be much more informative.

We agree. We replaced the LiDAR by the 2015 orthophoto.

Page 4 Line 18: What is "it"? Please clarify by writing out what it is.

We clarified this sentence. It has been changed to: "*Because of the difficulties to select a high number of independent control points spatially uniform over time in old spatial data and in accordance with findings from Tan and Xu (2014) (previously described), we argue that IDW is a reliable method to interpolate the registration error over old aerial photographs*." (Page 4 Line 18)

Page 4 Line 25: This seems like a big step backward to me, as you're taking a spatially variable error but then averaging it over a large area (the entire reach length) so you lose all of that spatial information. We agree.
Why not just use the actual SV error values at each location rather than aggregating over the reach?

We followed this advice to produce the new results in the revised manuscript (see our first remarks above).

Page 4 Line 26: Exactly, but your whole point is to NOT assume or use a uniform error metric. The way you've approached this, your just making the area assumed to be uniform a bit smaller (i.e., the reach rather than the whole image).

We agree. Our new methodology now takes into account the real SV-error (node specific). See supplement A and B.

Page 4 Line 27: Perhaps to some degree, but the error surfaces you produce by IDW will have a different value at every pixel, not just one value for each of the four reaches. If you're going to emphasize SV error you need to actually use SV error.

We agree. See comment above.

Page 5 Line 9: You could use this pooled-over-the-reach standard deviation to parameterize a unique normal distribution for each pixel, with the mean being the local SV error for that pixel, not just the reach-averaged error.

We thank the reviewer 1 for this suggestion. To produce the new results, we calibrated a Gaussian distribution from the neighborhood (5m buffer) of each node (see Supplement A and B).

Page 5 Line 13: But couldn't that digitization be based on images that were not co-registered accurately? I'm not sure I agree that the digitization should not be altered, so please elaborate on your reasoning for this approach.

We agree with your first point. Indeed, the digitization could have been produced from images that were poorly registered. Then, nodes of the digitized objects could suffer from a strong error in their position but their general shape would not be altered dramatically. Let's imagine that 10 different producers co-registrates the same aerial photo and digitize the same channel. Probably, it will appear a variability in the position of the nodes placed by each user. However, if a similar rule base has been correctly respected by every user, the global shape cannot be changed drastically.

Page 5 Line 13: Please include a figure that illustrates how the digitized bank nodes are shifted, as this is a key part of your method that is not well-described verbally but would be easier to understand with a graphic.

We thank the reviewer 1 for this suggestion. A "methodological figure" has been added to the revised manuscript (Supplement A).

Page 5 Line 15: This is an important point that needs to be described more explicitly and thoroughly.

Because we reshaped our methodology according to both reviews, the part 3.4.1 "Channel boundaries simulation method" has been rewritten in the revised manuscript. It is attached as the Supplement B.

Page 5 Line 17: Wouldn't this only work if the river moves straight north, south, east, or west, but not if it's movement is not in a cardinal direction? In other words, what about cases where x is positive and y is negative or vice versa. Overall, your method of shifting the digitized bank nodes seems oversimplified in some ways and could be refined.

Thanks for your observation. We first imagined that we could move each node of the polygon in a positive or negative X direction as well as a positive or negative Y direction as you said. In the present study, since the distance between two nodes is always higher than the error (its standard deviation) assigned to each node, the operation is feasible and does not alter the shape of the polygons. However, when the distance between nodes is lower than the error assigned to each node (in historical maps for instance, see Herrault et al., 2013), the operation can potentially lead to strong geometrical errors (cf "Butterfly polygon issue" in Supplement A). These errors could be corrected (moving average algorithms, Douglas Peucker, etc..) but the shape of polygons could thus be wrongly modified.

Therefore, we proposed an alternative solution to move nodes in space : (1) nodes from one sub-reach can move in any Y directions (positive or negative) at each run; (2) nodes from one sub-reach can only move in one X direction at each run. That latter rule allows to avoid topological errors while simulating the most probable displacements of polygon channels. See Supplement A for illustration. We also believe this choice is preferable to allow transferability of our method to other fluvial contexts.

Page 6 Line 9: Subtracting the SOD is not correct. You need to use the SOD as a threshold for assessing whether the measured value is significant. If the measured value exceeds the SOD at a given location, then the change is significant, but if the measured value is less than the SOD, the change is not significant. In either case, the SOD is used to establish the threshold, not subtract from the measured value.

We fully agree with your explanation and we erroneously interpreted this in the first version. We thus propose to calculate the relative percentage of uncertainty as regard to the reported mean value of change, .i.e erosion, deposition or erosion/deposition.
The percentage of (total) uncertainty is calculated as following: 0.5x(max-min)x100/mean
If the reported mean value exceeds the uncertainty threshold equal to 50%, then the change is considered as insignificant. Indeed, if the percentage of uncertainty exceeds 50%, the mean value is then necessarily lower than the total range of measured value (max-min), which leads to the assumption that the uncertainty is too high to consider the change (mean value) as significant.

Page 6 Line 12: OK, but you're losing a lot of SV information by aggregating over the entire sub-reach like this.

We agree on this and changed our methodology. See Supplement A and B.

Page 6 Line 12: Please add panels showing the error in the x and y directions separately, as well as the total error you have now.

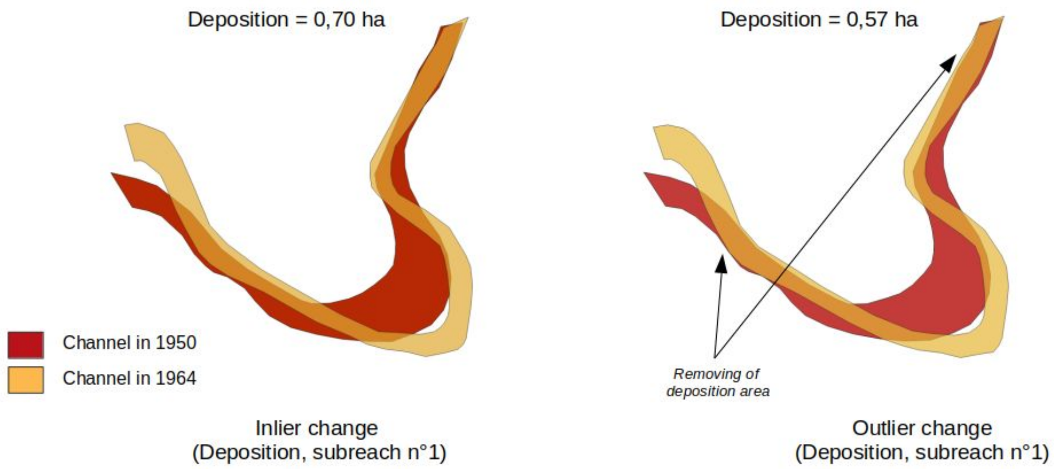
This is indeed a pertinent proposition. The figure has been reworked to display panels showing the error in x and y separately, as well as the total error. However, we only display the panels for 1950 in the main text; panels for 1964 have been added in supplementary material (Supplement C).

Page 7 Line 1: This is kind of getting at the threshold criteria you should be applying.

We agree. We now propose in the revised manuscript the following threshold: relative percentage of uncertainty > 50%. It corresponds to the following equation : 0.5x(max-min)x100/mean
Indeed, if the percentage of uncertainty exceeds 50%, the mean value is then necessarily lower than the total range of measured value (max-min), which leads to the assumption that the uncertainty is too high to consider the change (mean value) as significant.

Page 7 Line 13: Do you have any ideas for filtering out or eliminating these outliers?

In our opinion, these outliers should be visually checked to see if they represent geomorphologically plausible situations. This is mentioned in the discussion of the revised manuscript. We also considered the 95% confidence interval in order to propose a less conservative uncertainty percentage. This is mentioned in the discussion of the revised manuscript. Finally, we propose to add a figure which shows how an outlier can look to feed the discussion (see below).



Page 7 Line 20: Please clarify what you mean by assess

We were suggesting "to assess changes significance". But we agree the sentence was confusing and has been changed to "*This emphasises the need to take the SV-error into* *account and, importantly, to assess its impact on significance of the measured changes*. (Page 7 Line 20)"

Page 8 Line 16: Your approach would be greatly improved by leaving out this reach-averaging step and using the actual SV error at each location along the bank.

This is done in the revised manuscript. See comments above.

Technical corrections:

Please note that every technical and grammatical corrections have been taken into account in the revised manuscript.

Altogether, we thank the reviewer#1 for his/her in-depth reviewing, detailed corrections and the integration of his/her thoughtful remarks into the revised version of the manuscript will enhance its quality.

Referee 2

This manuscript outlines a new method to quantify errors in measurements of channel change calculated from repeat aerial image overlays. The method is a valuable contribution in that uncertainty in measurements of channel change are estimated from polygons of erosion and deposition; this makes the method generalizable to multiple river types (e.g., braided). However, the methodology fails to retain the spatial variability of geometric error, which previous studies have demonstrated to be an important source of uncertainty. The proposed methodology uses a spatially variable error to calculate geometric error statistics (e.g., mean and standard deviation) and generate a distribution of geometric errors that are randomly sampled and applied uniformly over each sub-reach. Thus, the proposed methodology assesses how the variability of geometric error influences measurements of channel change, and this differs from the stated aim of the manuscript: to create a generalizable spatially varying error assessment method. While I appreciate that the authors developed a method that can be applied to polygons of erosion and deposition and I believe the use of a Monte Carlo approach as merit, I have significant concerns with the proposed methodology.

We thank the reviewer for appreciating the originality of our study. We think that his/her reservations about our methodology are very relevant. We address them in the comments below.

Technical comments:

Page 2 Line 9: Image co-registration does not affect measurements of channel width because the images do not have to be overlaid to calculate the width.

We agree. This sentence was meant to summarise different kinds of metrics generally extracted from planimetric studies. However, as it might have been confusing for readers, we thus replaced the sentence by: "*Requiring data coregistration and river bank digitisation, these planimetric studies often result in the quantification of lateral migration rates (Hooke and Yorke, 2010; Janes et al., 2017; Mandarino et al., 2019; O'Connor et al., 2003).*" (Page 2 Line 9)

Page 2 Line 30: Why are medium-sized rivers more prone to digitization and coregistration error? I would think that small-sized rivers might be more prone to these issues because the digitization and co-registration error potentially accounts for a larger portion of the active channel.

We agree on the fact that the smaller rivers are, the more prone they are to be affected by spatial errors in planimetric analysis. According to the European Water Framework Directive, the Bruche river however falls into the medium-sized category (catchment >100 to 1000km²). Smaller rivers (streams) are generally too small to be studied with planimetric analysis with the channel polygon method. We added the reference to the Water Framework Directive classification. (Page 2 Line 31)

Page 3 Line 2: You need a sentence defining the channel polygon method.

We added a sentence defining the channel polygon method: "*The latter consists in the extraction of eroded and/or deposited surfaces from overlaid diachronic channels.*" (Page 3 Line 2)

Page 3 Line 7-10: Using the methodology proposed in this manuscript, I believe that you can only test hypothesis 1. This is because the spatial errors are aggregated to estimate a population of uniform errors which are sampled in the Monte Carlo framework. What you are actually testing is how the variability of error affects polygons of erosion and deposition (i.e., the effect of changing the mean and standard deviation of the populations of errors in a reach).

We agree and we reworked our methodology in the revised manuscript accordingly. Details of our new methodological approach is illustrated in supplement A, the latter being incorporated as a supplementary figure in the revised manuscript (either in the main text or as supplementary material). The part 3.4.1 in the current manuscript has been rewritten accordingly. It is attached as Supplement B.

Hypotheses 2 and 3 have also been redesigned:

1. Orthophotos are affected by a local significant SV-error;
2. SV-error highly affects variability of MC simulated measurements of eroded and/or deposited surfaces;
3. Uncertainty of surficial changes depends on their magnitude

Page 3 Line 8: “the higher the SV-error is, the less significant the measured changes are.” More description is needed for the word “higher”. Do you mean the larger the mean of the SV-error, the larger the standard deviation of the SV-error, or a combination of both?

We actually meant “the higher the LSE (Figure 5 in the current manuscript), the less significant the measured changes are”.
The LSE was however removed as it consisted in uniformizing the error over the sub-reaches. We now use a node-specific error, extracted from a normal distribution in the local node neighborhood (5m buffer). See our new formulated hypotheses above as well as supplement A and B for the new methodological approach.

Page 3 Line 29: What was the discharge on the day each image was collected?

Unfortunately the river wasn’t gauged before 1965. Nevertheless, as we strictly delineated the channels by referring to the active channel concept, we believe that differences in discharge do not have any impact on our methodology. Moreover, according to the observation of both orthophotos, the Bruche river wasn't at bankfull stage at these dates.

Page 4 Line 10: Note that the RMSE of a single GCP is the Euclidean distance between the two points. See equation 1 verses 2 in Lea and Legleiter (2016).

Thank you for this precision. We propose to supplement the sentence by some precisions as following: “*Local Root Square Error (RSE) is then measured for each of the 18 GCPs, on both orthophotos. Error in x or y correspond respectively to the euclidean distance between the two points for x and y coordinates. SV-error is calculated by interpolating local RSE on our whole study area with an Inverse Distance Weighting (IDW) technique at the original spatial resolution (Fig. 4)*.” (Page 4 Line 10)

Page 4 Line 14: The sentence starting with “First, Lea and Legleiter (2016) showed” is incorrect. Lea and Legleiter (2016) simply stated that linear and nearest neighbor reduced the spatial extent of large co-registration errors. The authors did not evaluate which interpolation method should be used.

We thank the reviewer for this comment. We modified our sentence by : “*First, from a comparison of five interpolation methods, Lea and Leigleter (2016) stated that linear and nearest neighbour methods reduces the areal extent of large co-registration errors. Thus, these methods could strongly limit the influence of large co-registrations errors on our estimations of surfacic changes, so we decided to eliminate them. Then, in a comparative study of spatial interpolation methods to produce Digital Elevation Model from a small set of points not spatially uniform, Tan and Xu (2014) showed that IDW provided better results than Spline or Kriging. Because of the difficulties to select a high number of independent control points spatially uniform over time in old spatial data and in accordance with findings from Tan and Xu (2014) (previously described), we argue that IDW is a reliable method to interpolate the registration error over old aerial photographs*.” (Page 4 Line 14)

Page 4 Line 21: What is the length of each sub-reach?

Length of each sub-reach are added in the revised-manuscript, as following: “*Their mean talweg lengths are respectively 530; 380; 700 and 890 meters long.*” (Page 4 Line 21)

Page 4 Line 25: The method to determine the LSE needs to be more clearly stated. Is the LSE calculated using the SV-errors extracted from each channel boundary vertex or all SV-errors within the sub-reach?

LSE is not used anymore in our revised methodology, which now takes into account node specific error. Please see Supplement A and B for more details.

Page 5 Lines 4-6: These sentences seem to contradict one another. In one sentence the authors state that MC simulations are useful because the method assumes “spatial continuity and a relatively spatial homogeneity of the error”, while in the second sentence the authors note that the method can improve the “generalization of methods for calculating planform changes and spatially variable uncertainty. . .”. This is a major problem with the proposed method. Lea and Legleiter (2016) and Donovan et al. (2019) demonstrated the importance of using a spatially varying co-registration error to estimate uncertainty at individual points; however, the authors use the SV error to estimate the mean and standard deviation of the co-registration error population in each sub-reach.

Thanks for your comment, we agree it was contradictory in the first version since uniform mean and standard deviation were assigned to all nodes of one sub-reach. In the revised manuscript, we have calibrated one normal distribution for each node from the local neighbourhood. Thus, we consider that the mentioned sentences can be held in the old form. A relative spatial homogeneity of error is assumed in the local neighbourhood of each node and we argue that our proposal can improve the “*generalization of methods for calculating planform changes and spatially variable uncertainty*. “

Page 5 Line 8: Have you tested whether the distribution of raw LSE values is normal? Would another distribution better model these values?

LSE is not used anymore in our revised methodology. We thus tested the normal distribution of SV-error in a 5m buffer around 10 channel nodes randomly selected along the 1950 and the 1964 channels, with the Shapiro test. Please check Supplement D for details (histograms and test results).

Page 5 Line 30: Note that the metric "erosion/deposition", as shown in Figure 1, does not always required erosion and deposition (e.g., channel avulsion or meander cutoff).

Thanks for this thoughtful comment; this short complementary text in the revised manuscript now specifies it : "*Note that the metric "erosion then deposition" measured in the area located between the former channel (T1) and the new one (T2) does not always imply continuous lateral channel migration followed by deposition. Sudden lateral shifts of meanders (i.e. through meander cutoff) or meander belts (i.e. through channel avulsion) may be involved as well and require specific geomorphological attention.*" (Page 5 Line 30)

Page 6 Section 3.4.3: Virtual Surface of Detection (SoD) is not an appropriate description and this is NOT equivalent to the LoD in Lea and Legleiter (2016). In my opinion, the SoD cannot be used to distinguish significant from non-significant changes. The SoD is simply a statistical description of the MC results. Because the authors adjust the channel delineations by the registration and digitization error for each MC iteration (equations 1 and 2), the individual iterations already take into account uncertainty and therefore should be significant. The SoD simply shows the variability of channel changes based on the distributions of error in the x and y directions for each image.

We believe that comparing our SoD to Lea and Legleiter's (2016) LoD was a mistake. As the SoD is technically not comparable to the LoD, we agree with reviewer 2. We therefore left the term SoD, as it might be confusing for readers. For these reasons, we chose instead to focus on the relative percentage of uncertainty associated with the variability of the measured changes through the MC runs in the revised manuscript. Because it is a simple and well-known indicator of variability, we believe it might simplify the interpretation and comparison of our results. The percentage of (total) uncertainty is calculated as following: 0.5x(max-min)x100/mean
Then, the (conservative) significance threshold becomes: percentage of uncertainty > 50%. Indeed, if the percentage of uncertainty exceeds 50%, the mean value is then necessarily lower than the total range of measured value (max-min), which leads to the assumption that the uncertainty is too high to consider the change (mean value) as significant.

Page 6 Line 24 to page 7 Line 2: It is not appropriate to directly compare results from each sub-reach without a normalization, such as by sub-reach length. The difference between sub-reaches could be caused by reaches being smaller or larger.

We agree. New results have been normalized by sub-reach length in Figure 7 which displays the mean surficial changes.

Page 7 section 4.3: The method cannot show the percentage of individual measurements of erosion or deposition retained because each MC iterations is treated as a single value, so sentences like: ": : :significant change globally increases from 17% using the raw-SoD to 37% using the 95-SoD" are incorrect.

We agree. Our methodology unfortunately does not allow dealing with a percentage of significance on surficial measurements. We therefore describe our new results using the relative percentage of uncertainty, which we believe to be a more appropriate indicator of the variability (and so uncertainty) of the results.
As our methodology and our results have changed, 4.2 and 4.3 have been re-written accordingly.

Page 7 Line 20: The authors state, "This emphasizes the need to take the SV-error into account: : :", yet their method does not include a SV-error.

We agree. Our new methodology now takes the real SV-error (node specific) into account. Please see supplements A and B for details.

Page 7 Line 25-31: The authors were not able to test their second hypothesis because the error ultimately did not vary spatially.

We agree, see our newly formulated hypotheses above.

In addition, the authors cannot identify the number of channel change measurements statistically retained and the results are not comparable to Lea and Legleiter (2016) or Donovan (2019).

We agree. Unfortunately, because (1) our surficial metric differs from the linear one used by Lea and Legleiter (2016) and Donovan et al. (2019) and (2) the way we deal with significance necessarily differs from the one proposed by Lea and Legleiter (2016), it is not possible to directly compare our results nor their significance to Lea and Legleiter's (2016) and Donovan et al.'s (2019) results.
As our results have changed since the first version of the manuscript, part 5 has been rewritten accordingly.

Page 8 Line 1-10: I do not believe that the authors successfully tested the third hypothesis because they did not directly include the SV-error nor was the significance of channel change measurements accurately determined.

We agree. Node specific SV-error is now actually included in our new methodology. Concerning the significance of channel change measurements, we now propose the relative percentage of uncertainty as a threshold. See comments above. See supplements A and B.

Page 9 Line 1: What is the appropriate sub-reach size and how sensitive are the results to the sub-reach size? How do you recommend users delineate sub-reaches for different channel types?

In our opinion, the suitable sub-reach size is dependent on the way the user intends to quantify planform changes. Our first thought was to focus on independent morphological units, such as meanders. We think it would be appropriate to decrease the sub-reach size when planform changes are more complex or channel pattern complexifies. On the contrary, a straight channel does not necessarily need to be divided into many sub-reaches. It would also be interesting to increase the size of the documented sub-reach, to check how sensitive the results would be. We propose to add the following sentence: "*Concerning the size of the sub-reaches, we recommend to adapt it according to the complexity of the planform changes and/or the channel pattern*." (Page 9 Line 2)

General editorial comment: The manuscript has numerous sentences that are awkwardly worded and could benefit from line-by-line edits to improve the readability. In addition, citations need to be checked (e.g., Wesley Lauer et al. (2017) should be Lauer et al. (2017)).

Proofreading of the revised manuscript was carried out and improved its formal quality. Citations have been checked too.