Response to review by Luca Malatesta:

First, we would like to thank the reviewer for his detailed and constructive review of our submission. We believe that addressing his comments will increase the quality of the manuscript. His two main comments were related to (1) the usage of terminology and (2) the structure of the manuscript. In addition, he provided several line comments related to science and bibliography. While the original review comments are shown in *italics*, our responses are given in regular blue font.

1. Fill vs. Cut-in-fill terraces:

The authors introduce the object "fill-terrace" on page 2 and thereafter it is inferred that all terraces recorded in their flume experiments are such. I would object to this use of the term. A fill terrace, as described on page 2, is a morphologic datum recording the culmination of sediment aggradation immediately preceding a phase of incision and thus abandonment (Howard, 1959; Bull, 1991; Pazzaglia, 2013). In several experimental runs, it seems that the entire active floodplain is being eroded before it narrows its width and starts entrenching, thus abandoning terraces. In that situation, these terraces are not "fill-terraces" but cut-in-fill as they record a moment during the incisional phase and not the culmination of alluvial aggradation. The title of the article needs to be accordingly modified. Then, the difficulty resides in reliably identifying if a given "top" terrace (top as in being the highest from the last incision episode) is indeed a fill terrace. To me it is very interesting that the authors identify cases where barely any fill terraces are abandoned. And that instead two large cut-in-fill terraces replace the fill terraces one would commonly expect. It appears to capture the moment when vertical incision is promoted over lateral erosion leading to fast autogenic entrenchment of the channel (Malatesta et al., 2017; Bufe et al., 2018) but the two experiments with a drop in Qs suggest that this inflexion point does not always occur at a similar moment. Finally on that point, the rationale behind picking the terraces TA and TB should be fleshed out because at least in the case of the DQsin run, they capture cut-in-fill terraces. More about that with the comment on p. 12 l. 13.

Indeed, the terrace terminology in the literature is rather inconsistent. Often, terraces are subdivided into two main categories: strath and fill (e.g. Howard 1959, Pazzaglia 2013). Fill terraces have been further subdivided into the 'highest' terrace that preserves the original deposited surface and 'lower' terraces with surfaces below the original deposited surface that have been eroded laterally into the fill. While the first type is referred to as 'filltop' (Howard 1959) or just 'fill terrace' (e.g., Bull 1990, Merritts et al., 1994), the second type has been described as 'fill-strath' (Howard 1959), 'cut-terrace' (e.g. Merritts et al., 1994), 'fill-cut terrace' (Mizutani 1998, Bull 1990, Pazzaglia 2013, Malatesta et al. 2017) or 'cut-fill terrace' (e.g., Norton et al., 2015). As such, for simplification, we only referred to fill terraces in general with the aim to include both subtypes. Especially since a distinction between the two subtypes in the field is often not possible without detailed stratigraphic or geochronological analysis.

However, we agree that a distinction of the two fill terraces sub-categories would be helpful to clarify the description and especially several points within the discussion. A distinction of the two subtypes within the experiments can easily be made, as we covered the surface with a thin layer of red sand prior to each perturbation. The preservation of the red sand is a clear indication for no further overwash after the perturbation and as such identifies the first subtype of fill terraces (filltop). Any later formed terraces will consequently be cut-terraces. In a way, we already made the distinction as the filltop terraces are those with a lag-time of 0 min (Fig. 5), while the cut terraces have lag-times of > 0 min. But a formal definition of the two subtypes will clarify the difference. For a better visualization, we will include photos of each terrace section and label the terraces accordingly. For later analysis (e.g. terrace surface slope) we always chose the most extensive terrace surfaces on each side of the river. With this approach we aim to mimic common field approaches.

We think that such a distinction between the two terrace subtypes will also clarify the discussion about the degree of reworking of terrace material. Different techniques can be applied to date terrace surfaces. Most of them, however, include sample collection at the terrace surface or within the upper couple of meters. As this part is often equivalent the active layer of the river bed (the depth range over which gains are actively remobilized and deposited), the lag-time between the onset of perturbation and the abandonment of a surface determines what we referred to as 'reworking of terrace material'. As this point was not clear in the discussion (see reviewer comment p.12 l. 1 below), we will clarify this point.

We also agree that the definition of what constitutes a paired or unpaired terrace is not clear (see reviewer comments p.9 l. 21-22 and p. 12 l. 5). Often, paired or unpaired terraces are distinguished based on height similarities or differences. However, as far as we are aware, there is no common rule where to draw the threshold. Instead, we will follow the reviewer's suggestion and instead refer to the ages/ lag-times of the terrace surfaces and describe successive abandonment instead of referring to 'unpaired' terraces.

Also, we agree that the cut-terraces capture the moment when vertical incision outcompetes lateral erosion. However, we disagree that this process should necessarily be referred to as 'autogenic entrenchment'. In the literature, the term 'autogenic' has been used inconsistently and no one definition of the term seems to exist. For the purpose of this manuscript, we have decided to define the term to include terraces that are formed without any external perturbation (i.e. under constant external boundary conditions). Terraces formed due to meander-bend cutoffs are one example of such autogenic terraces. In contrast, the cut-terraces observed in our experiments, which were formed after a lag-time with respect to the time of perturbation, are clearly linked to the external perturbation and are formed during a transient phase in which the channel adjusts to the new conditions. The transient is observed in a number of measureable quantities such as the channel slope and the discharge of sediment out of the basin. We will suggest a differentiation of the terms "autogenic" and "allogenic" for future use (see comment to p. 5 l. 10-13 below).

2. Structure of the manuscript

I think that a weakness of the current manuscript structure is that it is difficult to understand what the novel advances are and what the narrative of the work is. That is especially true for readers who are familiar with the existing, extensive, body work on alluvial geometry dating starting with Gilbert and Murphy (1914). The results are presented as if they almost provided a first-time observation of such alluvial dynamics. However, most of the observations from the flume experiments have already been observed, predicted, or discussed in previous bodies of work. What is novel is the documentation of the transient response itself. The manuscript could be somewhat modified to make this clearer and better highlight the contribution of the authors to this larger body of work. In that spirit, I would suggest to move elements of the discussion to the review section "2 Formation of fluvial fill terraces" so as to clearly establish what is acquired knowledge and to underline the gap that the authors want to fill here. In particular, section 2.1 could be augmented with large parts of sections "5.1 Channel response to perturbations and conditions of terrace formation" and "5.3 Differences in terrace surface slope". By explicitly introducing the theoretical framework used to describe the relationships between alluvial slope and fluxes of sediment and water (Qs and Qw), the authors would build a better launchpad for their study in my opinion. The Meyer-Peter Müller (MPM) equation revised byWong and Parker (2006) or more recent derivations of slope as a function of Qs and Qw (e.g. by Malatesta and Lamb, 2017 GSAB, or Wickert and Schildgen, 2019) can help establish clearly what is known so far, and what is not. The latter being a good understanding of the transient behaviour from one equilibrium configuration to the next. I believe that this modification to the structure of the manuscript would help the reader better navigate the coexistence of the review and experimental aspects of the paper.

We apologize for giving the impression that all our observations were novel. This was not our intention. The reason to not include the theoretical framework on channel geometry (relationship between Qs, Qw, S and W) in section 2 and only bring it up during the discussion was to keep the focus of the paper on fluvial terraces. But we agree that it might be better to expand section 2 to better distinguish our experimental results between novel observations and those validating existing theories.

As such, we will follow the reviewer's suggestions of and (1) rearrange Section 2 by including background on channel geometry (moving parts of the sections 5.1 and 5.3 into section 2 and expand it) and (2) rephrase the sentences that implied our observations are novel despite being a confirmation of earlier observations or ideas (e.g. p. 13 l. 11-12, p. 16 l. 18-19). We hope that both of these adjustments will help to better focus our work on the 'transient response of an alluvial channel to external perturbation'.

3. Science and bibliography comments

p. 1 l. 9-10: This is a pretty strong statement. I would argue that published work provide a pretty good understanding of the impacts of such forcing on terrace formation and sediment dynamics. What is lacking and provided by the authors here is rigorous observations of the transient response.

We agree that the original statement was rather vague and as such could be understood in several ways. Therefore we will adjust the sentence to: "However, we currently lack a systematic understanding of the timescales of terrace formation, the transient channel evolution, and associated sediment storage and release in response to changes in base-level, water discharge, and sediment discharge"

p. 2 l. 27-30: Malatesta, Prancevic and Avouac (2017, JGR) explicitly target lateral feedbacks with a numerical model.

We will include this reference.

p. 2 l. 31: Limaye and Lamb (2016, JGR) could also be mentioned here as an example of an excellent bedrock model.

We agree that the work of Limaye and Lamb (2016) is an important paper. But as our main focus is on alluvial rivers formed in response to external perturbation, we prefer to not include another bedrock model in the introduction. Please note though, that we cite this paper in the section on autogenic terrace formation, as it particularly focuses on the formation of autogenic terraces (p. 2 l. 13 & p. 5 l. 28).

p. 3 l. 8-10: I strongly encourage the authors to have a look at the 2003 Geology paper by Bonnet and Crave. Therein the authors investigate the impact of climatic (Qw) vs. tectonic forcing (base level) on an experimental landscape. While not targeting terraces in particular, it is one of the most insightful papers I've read on the subject. I strongly encourage the authors to read through it and incorporate some thoughts in their work.

Although not investigating terrace formation, their measurements of denudation rates go well along with our Qs,out measurements and we will incorporate their findings within the discussion section on 'Signal propagation and implications for stratigraphy'.

p. 3 l. 20: "upstream" [and along stream] (to take into account extra Qs from local incision)

Will be clarified.

p. 4 l. 3: If incision supplies sediment to Qsin along stream, then Qsin is not the input sediment flux. It might be useful to separate Qsin, Qsc (sediment transport capacity at any point along stream), and Qsout.

As also suggested by the other two reviewers, our measurements on $Q_{s,in}$ and $Q_{s,out}$ actually allow a quantification of the contribution of upstream supplied sediment ($Q_{s,in}$) and remobilized sediment from within the channel to the total sediment discharge ($Q_{s,out}$). We will include those absolute values to the lowest panels of Fig. 5 and adjust the text accordingly, including a clear differentiation between sediment supplied upstream ($Q_{s,in}$) and sediment remobilized within the bed.

p. 5 l. 10-13: I understand and appreciate the distinction here, and it is quite useful to separate the two. But is it a new refined definition? It seemed to me that fill-cut terraces are commonly considered both "complex response" and "autogenic" at the same time (Schumm's work and Pazzaglia's review paper). If you indeed propose this new,useful, distinction here, I would encourage you to take ownership of it.

Please see comment to 1.

p. 5 l. 28: There is a new paper by Johnson and Finnegan that is in revision at Geology on "Tributary Channel Transience Triggered by Bedrock River Meander Cutoffs." I don't know when it will come out. But regardless, it might interest you for the future.

Thanks for the suggestion.

p. 6 l. 5: As the reference codes of the experiments are going to be used thereafter, I would suggest to make a reference to Table 1 here.

Will be included.

p. 6 l. 16: what is the vertical resolution?

Will be included.

p. 6 l. 29: It could be helpful to mention that water is tainted blue in the photos.

Will be included.

p. 6 l. 32: why can it be considered unaffected?

We agree that this statement was too strong as we cannot 'prove' the upstream part to be unaffected. Instead, we will correct the statement to "we consider this part as least affected by the fixed position at the outlet". The second and more important reason to analyze the upstream part is because the terraces were preferentially formed in this part.

p. 7 l. 30: I would argue that change in channel width is not required to form fill terraces. What needs to be reduced is the breadth of the active floodplain (in which the channel, of potentially fixed width, migrates left and right).

We agree and will adjust the text accordingly.

p. 9 l. 4: The nature of terraces TA and TB could be mentioned here to simplify the reading of the paragraph.

This section will be adjusted in accordance with the subdivision into the two sub-types of fill terraces. See comment to 1.

p.9 l. 21-22: Is there a threshold for what constitutes a pair? Is there a way to define that objectively, or at least in a consistently arbitrary way?

Please see comment to 1.

p. 10 l. 6-7: Not sure I understand the rationale behind the ratio of vertical and horizontal erosion. A terrace of width W is preserved for a time T with a river lateral erosion Eh such that T=W/Eh. Preservation is independent from the vertical incision rate. However, deep incision will result in higher walls that are costlier to erode.

What we meant is that vertical incision needs to outcompete lateral erosion to even form terraces. We agree that the term 'preservation' used in the text was misleading. It will be corrected. However, the preservation is in that sense dependent on the vertical incision rate, as faster vertical incision will reduce lateral erosion due to a limited transport capacity of the available water.

p. 10 l. 15-19: Field studies such as Tofelde et al. (2018), Malatesta et al. (2017, Basin Research), or, and especially, Dzurisin (1975). More on the latter below.

Yes, we can also refer to the field studies here. Will be included.

p. 11 l. 4-5: a comment only valid if the theoretical framework for alluvial rivers is not beefed up above: I suggest to state that +Qs leads to +S in order to preserve eq. 1 under constant Qw, just as to explain the rationale between Qs and S which is not directly derived from Eq. 1 and 2.

The theoretical framework will be included in section 2. See response to 1.

p. 11 l. 8: This dynamic is described and discussed by Malatesta et al. (2017, JGR). It is also worth noting two earlier flume experiments by Schumm et al. [1987, chapter 6] and Meyer et al. [1995] describe the evolution of a channel profile after it reaches a new equilibrium post-incision (see description of that work in section 5.1 in Malatesta et al. 2017, JGR).

We agree that it is a good idea to compare our observations to other studies that have observed a channel widening after the slope has adjusted to its new equilibrium profile. In the three mentioned studies, widening is related to the input of sediment from the walls or the channel bed. We will expand our discussion to include those studies.

p. 12 l. 1: What exactly is the degree of reworking of terrace material? The amount of vertical incision?

Please see comment to 1.

p. 12 l. 5: I am a little hung up on paired/unpaired and the threshold it implies. Wouldn't it be more informative to simply write that the terraces are abandoned successively?

Please see comment to 1.

p. 12 l. 13: Runs DQsin and IQsin_DQsin both lead to entrenchment when sediment flux drops. So, why does the same forcing cause very different terrace creation, or at least be considered as two different systems? To me, it seems that the different terrace record of the two runs could be explained as reflecting the inherent variability in the abandonment of cut-in-fill terraces. See point about fill terraces written at the beginning of the review. It should be however noted that, in the experiment DQsin, there are two slivers of what was probably the original floodplain datum. As such, these slivers should be TA and TB for comparison with IQsin_DQsin.

The discussion about this point will be adjusted as we will distinguish between the two different types of fill terraces (see response to 1). We agree that the different responses to the same forcing probably indicate the inherent variability.

p. 12 l. 17: this feedback has also been extensively discussed and explored by Malatesta et al. (2017, JGR).

We will add this reference.

p. 12 l. 19-21: yes, but the two effects mitigate each other. If the incision rate is slow, the later terrace will also not have been lowered that much such that the geometrical difference remains about the same.

The sentence refers to the time when the switch from dominantly lateral erosion to dominantly vertical incision happens. The earlier the switch, the better the preservation of the initial profile. When the channel continues to planate laterally, it lowers the entire bed surface and when rapid incision initiates, the cut-fill terrace has a lower slope than the channel at the onset of the perturbation. Given the observations we make (Fig. 6), lateral erosion and incision do not seem to completely trade-off so as to keep the geometry constant as suggested by the comment. Instead, we see a good preservation of profiles in cases of instant incision (very low lag-times), compared to lower channel profiles in cases with longer lag-times.

p. 13 l. 11-12: The formulation used here suggests that the authors have observed and established ("we found that") this relationship for the first time, along the 2018 Wickert & Schildgen paper. Yet, the fact that terraces have a steeper gradient than the stream's for Qs or Qw forcing is not a new observation or theoretical construct, it is built-in in theory since early fluvial geomorphology work (Mackin, 1948; Meyer-Peter & Müller, 1948; Léopold & Maddock, 1957; Hooke, 1968; Schumm, 1973; Leopold and Bull, 1979; Wells and Harvey, 1987; Harvey et al., 1999; DeLong et al., 2008; Rohais et al., 2012). Recently Malatesta & Lamb (2018) used a derivation of MPM to constrain alluvial slope as an explicit function of Qs and Qw. This passage is one that inspires my earlier suggestion to provide a more complete overview of current knowledge, in particular in terms of theories of transport and geometry.

As already stated in our response to comment 2, we will expand section 2 in include information on sediment transport and channel geometry. Also, we will carefully rephrase the sentences that were pointed out as misleading.

p. 13 l. 14: I would also point to the absolutely remarkable site of the Gower Gulch alluvial fan in Death Valley. There, a man-made diversion instantaneously changed the hydrology of the catchment leading to sudden incision of the alluvial channel. Details are found in the work of - Troxel, B.W. (1974, ManÂ made diversion of Furnace Creek Wash, Zabriskie Point, Death Valley, California: California Geology, v. 27, p. 219– 223), - Dzurisin (1975,Channel responses to artificial stream capture, Death Valley, California: Geology, v. 3, p. 309–312, doi:10.1130/0091Â 7613(1975)3<309 :CRTASC> 2.0.CO;2.), - Snyder & Kammer (2009), - Malatesta & Lamb (2017). [you will find the two 70's papers on Gower Gulch attached hereby]

We will carefully study the suggested manuscripts and implement the findings of that work.

p. 13 l.30 - p. 14 l. 9: I am not sure that I follow the argument here. When terrace treads are used to quantify tectonic deformation, the gradient of the terrace does not matter as it is always detrended to retrieve local deformation (e.g. from an anticline, Lavé Avouac, 2000). As long as the tread is straight, tectonic deformation can be well-constrained.

We think the confusion is between using deformation of the tread of a single terrace and using slope differences between different terraces to reconstruct tectonic deformation rates. We agree that we have not clearly differentiated between the two in the manuscript. We will clarify the differences and adjust the text as well as the references accordingly.

p. 14 l. 12-15: this context could be introduced much earlier in the manuscript to better motivate the study.

This part will be moved to section 2.

p. 15 l. 7: It can be noted that this illustrates predictions of laws like MPM whereby no geometric change at the downstream end of the reach demands that the sediment flux transport capacity does not change either.

Unfortunately, we do not follow the comment of the reviewer. The sentence refers to changes in upstream sediment supply and the according adjustment of the channel reach. Although the base level at the downstream end is fixed, changed in upstream sediment supply do result in changes of channel geometry, i.e. slope and width of the channel reach.

p. 16 l. 6-7: Wouldn't chemical signals be best transferred during phases of bypass? Or is recycling more important in such phase than during aggradation?

We would expect that recycling due to lateral movement plays a greater role during bypass that during an aggradation event. Bypass, in the sense of no net deposition or erosion because the channel is in equilibrium, does not exclude the mixing of older and younger material during lateral movement.

p. 16 l. 18-19: I understand that these are observations from the runs, but I think it would be advisable to add that these "findings" validate existing theories. Though grammatically correct, the word suggests an unwarranted degree of novelty to my ears (non-native english hearing ears, mind you) . That is well known and demonstrated already. The same comment is also valid for point 5 of the conclusion.

We will rephrase the sentence accordingly (see reply to 2.).