Review of *A multi-proxy assessment of terrace formation in the lower Trinity River valley, Texas*

I have reread the manuscript *A multi-proxy assessment of terrace formation in the lower Trinity River valley, Texas* by Hassenruck-Gudipati and colleagues and their responses to the two previous reviews. The authors addressed most of the comments raised in the two reviews and implemented them in their revised manuscript. The revised manuscript reads well, is much clearer and well structured, and most concerns have been addressed. I think that the manuscript will be an important and timely contribution to the community. But there is one important point that I raised during the first review, which is still not entirely clear to me. Therefore, I need to address it again and I suggest to clarify this prior to publication.

We thank the reviewer for their feedback on the updated manuscript and highlighting outstanding concerns. We address these below and in the updated the manuscript.

The authors propose to use variability in terrace heights as a test to assess the plausibility of an allogenic terrace formation mechanism. To do this, they (1) subtract terrace heights from a plane fitted to modern floodplain heights (Figs. 3 and 4), and (2) compare the RMSE of terrace heights relative to a plane fitted to all data points on the same terrace with the RMSE of a plane fitted to randomly selected terrace segments, which is an indicator of autogenic terrace formation (Figs. 5 and 6). If I understand correctly, this is to investigate whether all terrace segments of a terrace (low, medium, high) are similar in height and belong to one large, externally-driven incision event or whether the heights are scattered and the terraces were formed by individual, localized incisions (autogenous terrace formation). However, using a plane as a reference surface introduces some uncertainty in the data, which I have tried to outline in the figure below. Although the modern river profile is fairly straight, elevation values are above the trend line near the outlet (probably due to recent sea-level rise and corresponding sediment deposition), below the trend line in the middle part, and above the trend line again in the upper part (Fig. b, taken from the manuscript). This variability in modern floodplain elevations results in an overall RMSE of 1.36. Is it possible, then, that most of the scatter in the terrace data is caused by the method, while only a fraction is truly due to variability in terrace heights?

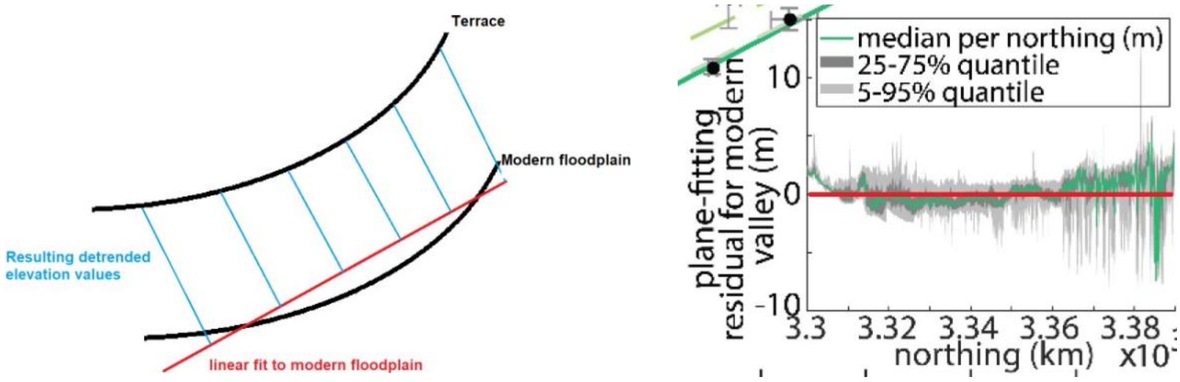
We thank the reviewer for this question on what causes the variation in RMSE in the floodplain. It is true that this method does not allow to distinguish where the RMSE error is coming from. While the above is the case we do recognize that any potential systematics in the residual are what would be expected from a plane fit to a concave up longitudinal profile. However, we suspect that the slight deviations in residuals in the Figure 3 (now Figure 4) a) are due to 1) aggradation at the downstream end and 2) the river deviates from the N-S valley axis in the upstream. We also note that the actual magnitude of variability is very small, ~1-2 m over a reach of >90 km, and less than the terrace heights above the modern valley floor (see Fig. 3B).

In the schematic figure on the left (a), the offset between the terrace surface and the modern floodplain is constant along the channel, as assumed for an allogenic forcing such as a base-level drop. However, the chosen approach systematically results in lower detrended values for the terraces in the middle of the reach compared to the upstream and downstream ends (blue lines). To me, this means that any distribution of residuals in the terrace data that results in an RMSE on the order of 1.36 is entirely due to the method itself.
The RMSEs for the three terrace data sets are only slightly higher (1.43 m, 1.54 m, and 1.41 m). Is it possible, then, that most of the scatter in the terrace data is caused by the method, while only a fraction is truly due to variability in terrace heights?

We thank the reviewer for pointing out the link between assumptions of a plane fitted to modern floodplain and the resulting implications assessing the variability in terrace heights. However, we think it important to note that the plane fitting to the terrace groupings was done on the raw elevation data, not the detrended data, so there would be no propagation of error from any systematic residuals in our planar fit to the modern valley floor to terrace analysis. The same basic concern could be raised for a planar vs. polynomial fit, but again, we argue that any systematics are within the error/natural variability of the system, and so fitting with a 2D polynomial would be an over-reach for the data.

Furthermore, terraces like those described on the Trinity River only have variations in their detrended elevation of maximum 15 m (Fig. 4). This is on the order of a channel depth for the modern Trinity River. Larger structural variation in along channel variation as describe in Figure a., below, likely only have a small effect on the RMSE results for a low-sloping river like the Trinity River. Assessing the variability in long-profile elevation versus terrace elevation variation might for rivers with larger concave up structures might be important.

a b



The authors then compare the RMSE, which describes the offset between each terrace height and a best-fit plane, to the RMSEs of only randomly selected terrace segments (Fig. 6) to test the null hypothesis. Figure 6 shows that the overall RMSE for randomly selected terrace segments increases with the number of segments selected, which I would expect since a larger number of randomly selected terrace segments causes a wider distribution of residuals. Currently, the authors do not reject the null hypothesis for the lower terraces because the RMSE of the lower terrace overlaps with the RMSE distribution of the Monte Carlo fits (Fig. 6). However, the non-rejection is not due to the fact that the elevation data of the lower terrace has a larger dispersion compared to the other terraces (given the RMSE, it is quite similar to the middle and high terrace levels), but because fewer segments are preserved. Does this mean that even if the distribution of the residuals and the RMSE are very similar, the question of whether a terrace group can be considered allogenically formed or not depends solely on the number of preserved terrace segments?

We thank the reviewer for this great summary of the impacts of the number of terraces preserved on assessing if these terraces were allogenically formed. We think that it is harder to justify an allogenically formed label for the case where only a small number of terraces are formed, something our Monte Carlo approach quantifies, and which also makes sense for the reasons the reviewer points out. However, we would argue it does not "solely" depend on this, as there is nothing precluding a set of allogenic terraces from having a very low RMSE over such a short reach as studied here, the studied sets simply do not.

In any case, an RMSE is only a single parameter describing a distribution of residuals. Wouldn't it therefore make more sense to compare the full distributions of residuals to assess the scatter in the terrace survey data, perhaps using a Kolmogorov-Smirnov test? After all, as long as the distributions of the residuals are quite similar to the modern flow, an allogenic driver seems quite reasonable.

We have considered a K-S test for the residuals but given the macro-scale roughness of the floodplain and terraces, we are not convinced the residual structure would be sufficiently more informative than a simple RMSE metric. We prefer to use the RMSE metric as a simple means of evaluating the plane fit rather than the evaluating of the more floodplain structure-dependent distribution of residuals.

On the other hand, I understand that the authors prefer to test the null hypothesis of an autogenic driving mechanism. The current overlap of the RMSE values of the lowest terrace and the randomly selected terrace segments cannot falsify this hypothesis. However, this means that there are not enough segments left to identify an allogenic drive. This does not mean that these terraces were autogenically generated. It just means that not enough segments are preserved to determine this. In this case, I cannot support the conclusion that the lower terrace was formed by an autogenic mechanism, as stated in the abstract (line 18) and in several places in the manuscript.

We agree with the reviewer that the limited number of terrace segments limits the ability to reject our null hypothesis that terraces were formed autogenically. That said, we would also expect an allogenic forcing mechanism to abandon a large number of terraces considered over the same valley reach length with more than one channel bend preserved on the terraces. Aggradation history at the downstream end prevents us from assessing the first expectation but number of channel bends are less than or equal to 1 for low Deweyville terraces (Figure 12). Regardless, we wholeheartedly agree with the reviewer that our failure to reject the null hypothesis of autogenic formation for the Lower Deweyville set does not disprove allogenic formation, it simply means we should not reject autogenic formation. We have tried to make this clear in the abstract on Line 23-26. Our interpretation that favors autogenic formation of this terrace set is our interpretation based on both these results, and the rest of the analysis presented in the manuscript.

Overall, I think the manuscript is an important contribution to the community and that we lack methods to serve as a "quality control" before using terraces for paleoenvironmental reconstructions. But I am not yet fully convinced that the proposed approach or the conclusion drawn are correct. I also realize that the elevation data are only one of several proxies analyzed. However, given the exceptional preservation of the paleochannels at the study site that were used for the other proxies, the analysis of the elevation data is the one that can be most easily applied to other study sites. Therefore, I would be grateful if the point raised above could be clarified before publication. I provide some further line-by-line comments below.

We thank the reviewer again for their comments, and hope our above response have proven satisfactory in addressing their remaining concerns.

Line-by-line comments

Lines 15-16: A cluster in elevations is not necessarily expected for terrace formed by a change in hydroclimate, as is also explained well later in the manuscript.

We thank the reviewer for catching this. The lines now read (L15-16): "For 52 distinct terraces, we quantify whether terrace elevations fit distinct planes…"

Lines 45-46: To make the sentence easier to read, it might helpful to add a "(1)" before 'punctuated decreases' and a "(2)" before 'punctuated base-level fall'.

We thank the reviewer for the helpful suggestion. The text now reads (L45-46): "Commonly invoked allogenic triggers connected with terrace formation are (1) punctuated decreases in sediment-to-water flux that are assumed to embed a signal of regional climate change and (2) punctuated base-level fall controlled…"

Line 62: 'This reduction from the measured paleo-slopes of terrace sets…' This sounds a little strange, I suggest rewording.

We thank the reviewer for the feedback and have rewritten the sentence to read (L62-63): "This reduction in slopes from older terrace sets to the modern floodplain has been observed in both natural (Poisson and Avouas, 2004) and experimental (Tofelde et al., 2019) systems."

Line 77: Channel bed slope instead of bedrock slope? (last word in line)

We have changed the wording to read (L77-78): "include local variations in channel dynamics, channel bed slope, and sediment contribution from tributaries."

Figure 1: Maybe add an arrow indicating the flow direction in the figure. Also, is it correct that river discharge decreases in downstream direction? The downstream gauging station (Liberty) has a lower discharge value compared to the upstream one.

We have added an arrow to identify the downstream direction. The downstream gauging station does have a lower mean discharge. This could be because not all discharges are found for USGS stage measurements for Liberty since tides do influence discharge in this region.

Figure 2: Please add coordinates to the map (A) to allow the reader to find the site in other datasets and Google Earth. Is the legend in (B) displayed correctly? To me the colors for post-Deweyville, Beaumont and Lissie all look white.

For Figure 2, we have added the coordinates to the map (A) and added colors to the legend in (B) for stratigraphic unites outside of the study period.

Lines 131-132: The information about floodplain aggradation during the Holocene is an important point. It means that we cannot directly compare the slope of the valley floor with the slope of the terraces, because the valley floor slope at the end of incision phase is not preserved anymore. Hence, this argument cannot be used to rule out hydroclimatic changes as terrace formation drivers, because it is possible that the channel slope at the end of the incision phase was different than the terraces. Instead, the similarity in slopes of the three terrace themselves could be used as an indicator that any potential changes in water discharge were complemented by changed in sediment discharge.

We agree with the reviewer that both water and sediment discharge likely increased and point this out in the manuscript in Lines 453-455: “We suspect that the switch in discharge is not directly recorded in the terrace elevation because the change in water discharge appears to have been approximately matched by a sediment-flux increase, as recorded in the constant long-profile slope for the paleo-river. With no slope reduction, no incision would have occurred.” Therefore, even if hydroclimate changed it might not have been the trigger for terrace formation. As the reviewer points out, this assessment is harder for the low Deweyville terraces that have be partially buried with Holocene deposition (L131-132).

Line 157: The summary of the null hypothesis and approach in section 3 is really helpful. Just a suggestion, but the authors could even consider to summarize their approach in a simplified, schematic sketch, especially since they want to ‘sell’ this approach for future studies.

We thank the reviewer for this suggestion and have summarized the approach to section 3 in Figure 3 and updated our RMSE fitting in Figure 4.

Line 187: It is unclear if the median values were calculated for each of the 52 terrace segments or only for the 3 terraces. Please clarify.

We have updated the text to specify that each of the 52 terrace segments are quantified in this way. The text now reads (L193): “From these elevations, the median value and interquartile range were found for each of the 52 mapped terraces.”

Line 191: As stated above, the higher elevation values close to the outlet that plot above the plane are probably related to sediment deposition since sea-level rise?

We reemphasized this likely cause for this trend is sediment deposition. The text now reads (L198-200): “Plotting the residuals to the best-fit plane along UTM northing reveals some structure in the most downstream southern long profile extent (Fig. 4A insert), likely due to Holocene sedimentation (Blum et al., 1995; Blum and Aslan, 2006).”

Line 192-194: I suggest to move this sentence up to line 189 to state from the beginning, why this analysis is done.

We have moved the suggested text up in the paragraph (L195-196). The text now reads “The best-fit plane for the modern valley was used to generate detrended elevations for each terrace DEM measurement by subtracting it from the spatially corresponding modern valley best-fit plane value.”

Figure 3: It would be helpful to color the datapoints in (A) and (B) according to the terrace they belong to.

We have added the grouping to points in Figure 3.

Figure 6: The actual RMSE values for the three terraces are not give here, they only come up later in section 4. Please briefly give the values already when describing the fits in the results.

We have updated the figure caption to read (L241): “The low, intermediate, and high Deweyville terrace sets have RMSEs of 1.43m, 1.54m, and 1.41m, respectively.”

Line 506: Remove ‘introduce’?

We thank the reviewer for catching this mistake. The text now reads (L514-516): “We suggest that paleo-channel characteristics are a more faithful record of discharge changes in fluvial systems and that additional bend metrics can differentiate autogenic terrace formation processes, specifically bend cut-off from unsteady lateral migration rates.”