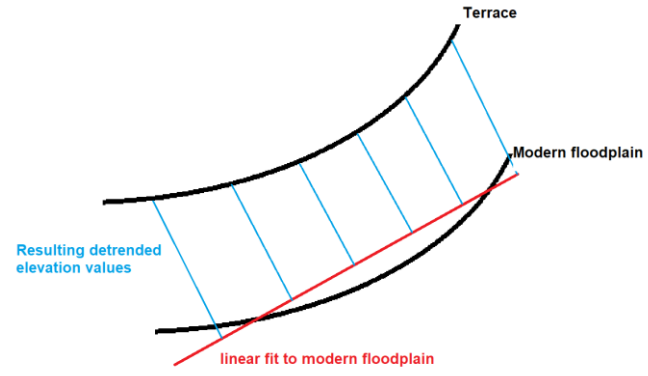
Review of *A multi-proxy assessment of terrace formation in the lower Trinity River valley, Texas*

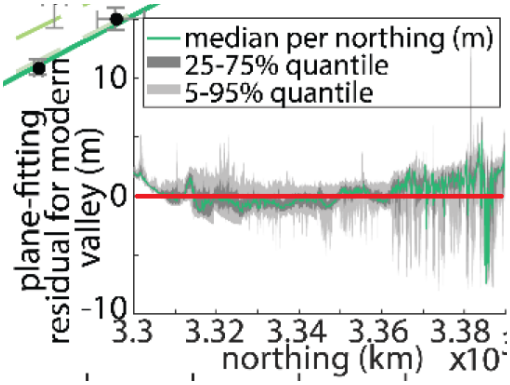
I have reread the manuscript *A multi-proxy assessment of terrace formation in the lower Trinity River valley, Texas* by Hassenruck-Gudipati and colleagues and their responses to the two previous reviews. The authors addressed most of the comments raised in the two reviews and implemented them in their 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.

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. However, there is a spatial trend in the residuals that will also affect the detrending of the terrace elevation data. 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?

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?

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.

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.

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.

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.

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’.

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

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

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.

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.

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.

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.

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.

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?

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

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

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.

Line 506: Remove ‘introduce’?