Earth Surf. Dynam. Discuss.,
doi:10.5194/esurf-2017-21-AC1, 2017

Earth Surface Dynamics Discussions

Interactive comment on "Geomorphometric delineation of floodplains and terraces from objectively defined topographic thresholds" *by* Fiona J. Clubb et al.

Fiona J. Clubb et al.

f.clubb@ed.ac.uk

Received and published: 5 May 2017

We would like to thank the reviewer for their comments, and their positive response to our manuscript. We have edited our manuscript accordingly including expanding the discussion section, and including an overall quality measure to compare the results of our methods to the published datasets. Details of our responses to the individual comments are outlined below.

1. First of all, I do appreciate the effort of creating an entirely automated procedure: this is the ultimate goal of many research, providing tools to avoid time consuming field surveys over large areas, in addition to allow understanding earth

surface processes at the landscape scale. The paper states that prior approaches required manual editing by the users, and they suggest their work is a step forward from these issues. They underline this fact many times in the manuscript, describing how their method is 'fully automated'. However, I think the authors should note that indeed, the procedure is still not fully automated. At page 7 - line 203: there is a suggested threshold, but such threshold can be changed by the user after visually inspecting the landscape - line 207: the user must provide the latitude and longitude do focus on a specific channel of interest (of course, in the case the user wants to focus on a specific channel on the whole landscape, which is understandable) - line 212: the user must specify the width of the swath, and this value can be estimated by a visual inspection of the DEMs. So it appears there is still some user-related parameters. I think, actually, what the authors propose is a procedure based on a fully automatic threshold (based on statistic) for the extraction (as the paper title correctly indicates). And statistic itself has been proven very useful in this task in many other research papers also in other fields, in addition to those mentioned by the authors in the introduction e.g. (Molly and Stepinski, 2007; Thommeret et al., 2010; Pelletier, 2013).

We would like to thank the reviewer for their appreciation of our goal in the paper of creating a fully objective method of feature extraction. We agree that there are user-defined parameters which are set in the method, and we have stated this clearly when outlining our methodology. In order to make this clearer we have changed references to this throughout the text to highlight that the threshold selection is fully automated, but that the method does require some visual inspection of the DEM prior to running the analysis. We have also added in references to the studies suggested here by the reviewer:

'Many methods of channel extraction employ statistical selection of topographic thresholds (e.g. Lashermes et al., 2007; Thommeret et al., 2010; Passalacqua et al., 2010a; Pelletier, 2013; Clubb et al., 2014), but this has yet to be developed for the identification of floodplains or terraces. We identify thresholds for R_c and S

using quantile-quantile plots, which have previously been used in the detection of hillslope-valley transitions (e.g. Lashermes et al., 2007; Passlacqua et al., 2010a).'

2. Table 3 reports the accuracy of the floodplain extraction. I tried to do the math myself but I do not get the value of 8m for the Mid Bailey Run. Maybe I am missing something? Also, the mean distance is not a reliable information, the authors errors arrive to values of ≈90 m. This measurement might not be that influent for landscape scale processes, but for flood inundation maps, especially near human settlements, it might make a difference, so I think it is worth discussing it, unless the authors believe that this error is an outlier due to specific reasons (but it still might be worth mentioning it). Maybe they could evaluate reliability and sensitivity for the FIP (and not just for the overall floodplain extraction) as (Orlandini et al., 2011) did to assess the goodness of its point identification. This would also make the floodplain initiation point analysis consistent with the floodplain identification and terraces extraction analysis.

In response to this comment we have calculated the reliability and sensitivity of the method compared to the mapped FIPs instead of reporting the mean distance, in order to make the comparison more robust and to keep it consistent with the analysis for the rest of the data. We have also added a discussion of the reliability and sensitivity compared to the mapped FIPs.

3. Lines from 285 to 335 should be in the method section. This is not a result, but rather the metrics the authors choose to evaluate the quality of their results. Concerning this approach (also for the previous point), I think the use of an overall Quality measure would be appropriate, rather than just using reliability and sensitivity. Overall quality can be evaluated according to (Heipke et al., 1997), which is the first one proposing the sensitivity and reliability formulation. This would allow the authors also to compare their quality with other works about feature extraction in literature. I would also argue that reliability and sensitivity in their broad sense do not report an

overall 'spatial correlation' between the datasets, as stated by the authors (line 365), but only a specific relation between either false negatives or true positives. Hence why I would suggest to use an overall measure as well.

We have moved this section from the results to the methodology. In terms of the quality analysis, we would argue that using the reliability and sensitivity values does allow comparison of the quality with other works in the literature: numerous studies presenting methods of feature extraction have reported the reliability and sensitivity, for example in channel extraction (e.g. Orlandini et al., 2011; Clubb et al., 2014) and in floodplain identification (e.g. Manfreda et al., 2014). In order to determine the performance of the method spatially, we also report the flow distance between the mapped and predicted floodplain initiation points for the field mapped data (Table 3). However, for the terraces and the published flood maps, metrics of length do not provide a good predictor of the performance of the method, therefore we decided to report reliability and sensitivity values which take into account the true or false positives or negatives based on the entire DEM. However, as suggested, we have also added in the overall quality analysis based on Heipke et al. (1997) for each of the comparisons (Table 4).

4. Line 383: Floodplain inundation and alluviation changes through time. However I am not sure these changes would affect the geomorphological floodplain in the timeframe expressed by the authors (2-5 years' differences) unless significant events happened in that timeframe.

We have removed this sentence from the discussion - although it's interesting to note the timescales of the formation of geomorphic floodplains are not well understood. A potential application of our method could be to compare the different floodplains predicted geomorphically with those of a specific magnitude event predicted through hydrological modelling.

5. Results discussion. Can the author explain why their method performs better for floodplain delineations rather than for terraces? Is there a reason related to the *method itself, or to the topography under analysis?is it related to the method they use to extract the channels? I think this is worth discussing more. Also, can the authors provide information about what influences the rate of TP or FN (so reliability and sensitivity, and eventually overall quality if they decide to evaluate it)? I think this is an important information to give, so users willingly to apply the proposed method in other areas can understand where to expect better or worse results.*

The sites used for terrace identification were generally lower relief than those for the floodplain extraction, which is a potential reason for the worse performance of the method in these sites. In the terrace site with higher relief (South Fork Eel River), the method performed as well as for the floodplain identification. The method of channel extraction will not influence the results of the algorithm, as we only extract floodplains or terrace on higher order channels which are not affected by the locations of the first order channels. We have expanded the discussion to include a section on comparison of the performance of the method between floodplain and terrace extraction:

'The results of the quality analysis for the eight field sites (Table 4) showed that the method performed better in the floodplain identification compared to the terrace identification. This may be due to the fact that, with the exception of the South Fork Eel River, the sites used for terrace extraction are lower relief than those used to test the floodplain extraction (e.g. Figures 6 - 8).'

Our manuscript includes information in the discussion about potential influences on the reliability, sensitivity, and overall quality for both the floodplain extraction and terrace extraction. We have also included discussion of the types of landscape in which the method may work best:

'As our method relies on the distribution of relief relative to the channel in order to select the threshold for terrace identification, it will work best in areas where there is a greater contrast between the slope and relief of the terrace surfaces compared to the surrounding topography. This is similar to other semi-automated terrace extraction methods (e.g. Stout and Belmont, 2014; Hopkins and Snyder, 2016).'

6. Line 418- on. The authors state their method is relatively insensitive to grid resolution. However, their optimum value of reliability is obtained with a 5m DEM rather than for a 1 m DEM, and there are variations in reliability and sensitivity when changing the resolution: in some cases, the r and s are higher for the 10m DEM. I wonder if the authors have an idea on why this happens (maybe less noise on the 10m DEM that can influence their evaluations? Maybe too much noise on the 1m?). I think this part is also worth discussing a bit, since the procedure is available to the public, and users might have different datasets (not necessarily Lidar at 1m). I understand the shifts in the two indices are low in magnitude, but I think discussing them makes sense.

We have expanded the discussion to suggest potential reasons why grid resolution may cause some small changes in the values of reliability, sensitivity, and overall quality:

'High-resolution topographic data may contain both small-wavelength topographic noise caused by tree throw and biotic activity (Roering et al., 2010; Marshall and Roering, 2014), as well as synthetic noise from point cloud processing (Liu, 2008; Meng et al., 2010). This noise may affect the calculation of topographic metrics (Grieve et al., 2016c), potentially leading to differences in the location of extracted floodplains or terraces compared to the lower resolution data.'

7. Figures The figures are clear and well described. Just a curiosity: figure 8c and d: the predicted terrace is quite different from the digitised one in the central part of the river. From a visual inspection, this appears as a quite well define terrace, what is this difference's cause? Also, is it possible to have a map of an area showing both the identified terrace and floodplain?

We think that the very subtle change in elevation between the different terraces in the central part of the valley in this landscape compared to the others (e.g. the Mattole River, Figs 8a and b) makes it difficult to identify these accurately compared to the digitised terraces. If the terraces are very close in elevation to the modern floodplain

then it can be difficult to distinguish between these from the DEM alone: we think that some of the portions of the landscape identified as digitised terraces may be selected as modern floodplain in our method. We tested the ability of the method to distinguish between floodplains and terraces, and found that the best way of separating between floodplains and terraces is to use a threshold height of the terraces above the modern river channel, which is user-defined, as in lower resolution DEMs patches of predicted floodplain/terrace may be connected. We have added in a section to the discussion of the problems of distinguishing between modern floodplain and terraces:
'Another potential cause of error between the predicted and digitised terrace locations may be problems in distinguishing whether features represent the modern floodplain or terraces. In our method a minimum height above the modern channel is set, where pixels above this height are classified as terrace, and below this height are classified as floodplain. In some cases, particularly where the terraces are at a similar elevation to that of the modern channel, our method may mistakenly identify terraces as being part of the modern floodplain, or vice versa. An example of this may be the Clearwater River site, where our method had lower values of the quality metrics (Figures 8c and d, Table 4). In this site, the digitised terraces are close in elevation to the modern channel, with a maximum terrace height of 13 m.'

We have not included the combined floodplain and terrace maps into the paper, as we want to keep the sections on floodplain and terrace identification separate to fit in with the overall structure of the paper. However, we include the combined map for the Clearwater River here for reference.

Interactive comment on Earth Surf. Dynam. Discuss., doi:10.5194/esurf-2017-21, 2017.

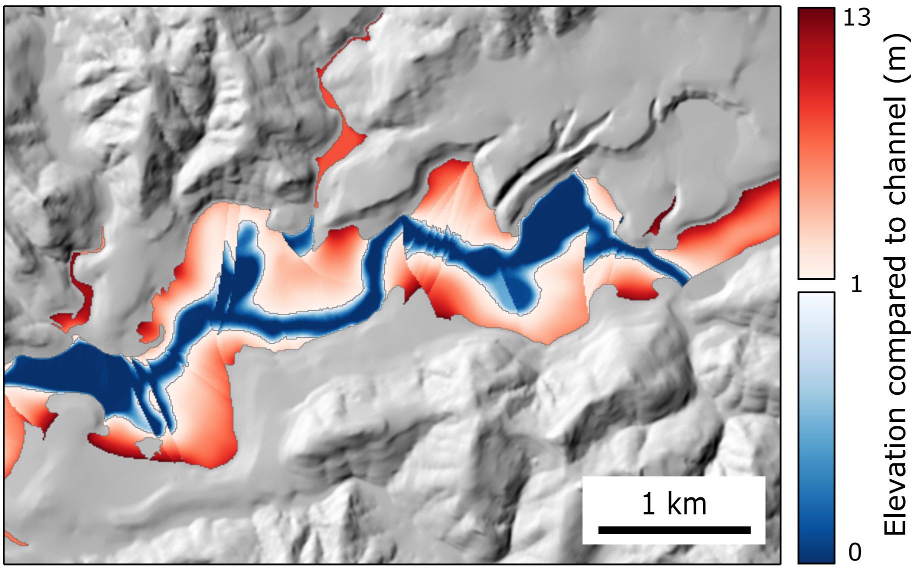

Fig. 1. Shaded relief map of the Clearwater River, WA, showing combined floodplains (blue) and terraces (red)