THE UNIVERSITY *of* EDINBURGH
School of Geosciences

Fiona J. Clubb
School of Geosciences
University of Edinburgh
Drummond Street
Edinburgh, EH8 9XP
Phone: +44 (0)131 650 9170
Email: f.clubb@ed.ac.uk

Greg Hancock
Associate Editor, Earth Surface Dynamics

May 26, 2017

Dear Dr. Hancock,

Thank you for considering our manuscript 'Geomorphometric delineation of floodplains and terraces from objectively defined topographic thresholds'. We are grateful to the reviewers for providing constructive feedback and allowing us to improve the manuscript.

We have made significant changes to our manuscript following the comments we received. Throughout the manuscript we have clarified the sections of our method that are fully objective (the selection of the thresholds for channel relief and gradient), and those for which user-defined parameters are required. We have included an additional measure of quality (Q) into our analysis along with the reliability (r) and sensitivity (s) metrics, which allows the overall performance of the method to be compared more easily to other studies of automatic feature extraction in the literature. We have also expanded the paper to include more discussion of the results, including the impact of grid resolution and comparison between the performance of the method for floodplains and terraces. Alongside this, we have added in a new section to the discussion (Section 5.3) on future research needs and the development of fully-automated methods of geomorphic feature extraction.

Please find below detailed responses to the individual points raised by each of the reviewers, along with a version of our manuscript highlighting the changes we have made to answer the reviewer comments. Throughout our responses we refer to line numbers in our manuscript: these are the correct line numbers in the manuscript with the changes incorporated. We have endeavoured to address all concerns and return the manuscript in a publication-ready state.

Sincerely,

Fiona Clubb

Fiona J. Clubb

Reviewer 1

The paper by Clubb et al. is an interesting and valid contribution to the journal. The authors propose a digital approach to mapping floodplains and terraces in different landscapes and compare their results with field measurements or flood maps derived from other sources. The paper is very well written and I enjoy reading it. Overall I think the authors provide a clear and detailed example of the validity of their procedure. However, I have few comments that I think might help to improve the paper.

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 (e.g. Lines 200 - 217). 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:
Line 221-226: '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. The reliability and sensitivity values for Coweeta and Mid Bailey Run are reported on Lines 316-320. We have also added a discussion of the reliability and sensitivity compared to the mapped FIPs on Lines 392-400.

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 (Section 2.3, Comparison with published data). 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).

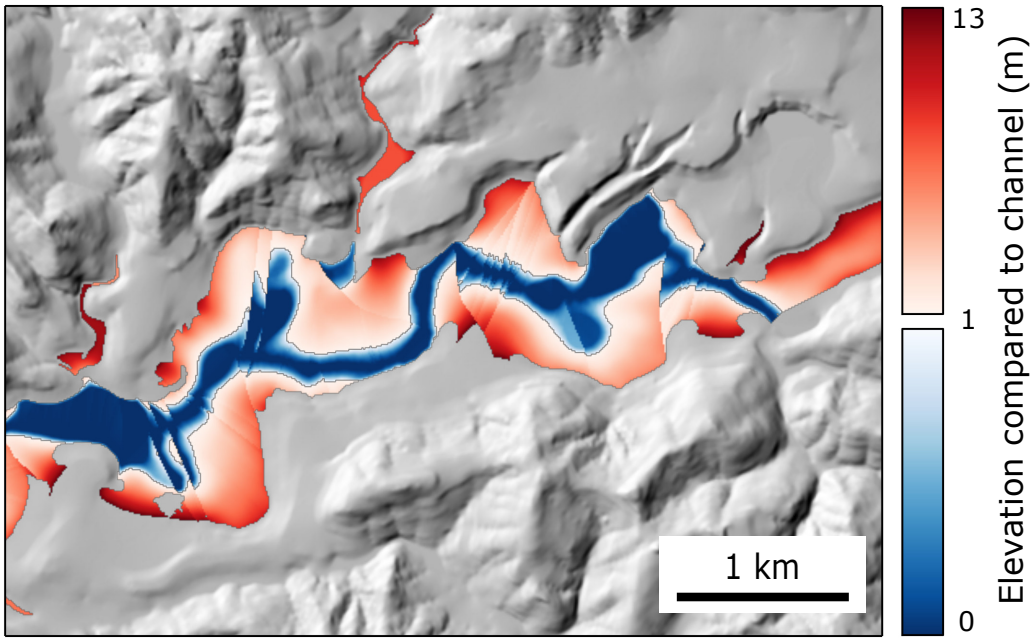
4. Line 383: Floodplain inundation and alluviation changes through time. However I am not sure these changes would affect the geomorphological floodplain in the timeframe expressed by the authors (2-5 years differences) unless significant events happened in that timeframe.
We have removed this sentence 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:
Lines 465-469: '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 (Lines 406-422) and terrace extraction

(throughout Section 5.2). We have also included discussion of the types of landscape in which the method may work best:
Lines 456-460: '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:
Lines 441-445: '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 differences 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:
Lines 470-478: '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.



Shaded relief map of the Clearwater River, WA, showing combined floodplain (blue) and terrace (red) map. Where terraces are close to the elevation of the modern channel it can be difficult to distinguish between terraces and the active floodplain.

Reviewer 2: Andrew Wickert

This paper is a strong contribution to ESURF and a clear step in the right direction towards a mapping floodplains and terraces. I am particularly pleased about the idea of using the quantilequantile plot approach, which provides a null hypothesis (in this case, normally-distributed topography) against which the landscape may be tested. The majority of my comments are in the paper itself, an annotated version of which is attached.

General comments

1. First, I will echo the first reviewer in writing that there needs to be a more clearly-defined line between "fully automated" and "semi-automated". In other words: define the realms within which your model is automated or is not. Currently, the lack of a well-defined separation undercuts the advances that you really have made by making it seem as if you overstate the work and making the focus on the "it isnt that far" rather than "it is a big step beyond prior work". I have read that Clubb et al. have responded to the first reviewer already in response to this general concern, so I will go on to a couple more specific points:

In response to this comment, and the comments from Reviewer 1, we have edited our manuscript to highlight the distinction between the parts of our method that are fully automated (the statistical selection of the thresholds from the quantile-quantile plots) and the need for some user-defined parameters. Our method does still have some parameters that are user-defined (threshold stream order for running on a landscape scale, width of the swath, and minimum height above the channel for floodplain/terrace distinction). However, in general these parameters can be estimated easily by the user from visual inspection of the DEM, and don't require the input of any independent datasets, unlike previous methods. However we agree that future research is needed in order to create a fully autonomous method, which is beyond the scope of our paper at the moment. We have added a section

to the discussion on future research directions, highlighting the points raised in the review comments: Section 5.3: 'A key goal for the Earth surface research community is to develop fully-automated methods of feature extraction from DEMs in order to avoid expensive and time consuming field-mapping, and to investigate the controls on geomorphic processes at a landscape scale. Our new method of floodplain and terrace delineation attempts to meet some of these research needs, by allowing the statistical determination of the thresholds for feature extraction. However, our method still requires the input of some user-defined parameters. If the method is run across the whole landscape, the user must set a threshold stream order for the calculation of elevation compared to the nearest channel. This is necessary so that each pixel is mapped to the main channel along which floodplains or terraces have formed, rather than narrow tributary valleys. This threshold can be determined by the user based on a visual inspection of the DEM compared to the channel network. If the user runs the method based on the swath mode, the width of the swath profile must be set. This can also be done based on a visual inspection of the DEM to provide a sufficiently wide swath compared to the valleys in the landscape. Furthermore, if the method is run in the swath mode, then a minimum terrace height must be set in order to delineate between floodplains and fluvial terraces.

However, future development of new algorithms, such as extraction of valley widths, would allow these parameters to be set based on the topographic data alone. Our method represents a first step towards this goal of fully-automated geomorphic feature identification, which can be improved upon with future research. The combination of different algorithms for terrain analysis, such as hillslope flow routing, channel network extraction, floodplains, and fluvial terraces, would allow an objective landscape-scale investigation of the controls on geomorphic processes.'

(a) One arbitrary piece is the decision about how wide of a swath should be used to search for terraces. To me, this highlights something that has long been on my "to do" list: a tool to automatically compute the widths of river valleys (see Shaw et al., 2008, for an analogous problem in coastlines). So I think that your use of an user-defined parameter is due to the lack of a tool that is outside your current scope, making this a placeholder for a better method!
This is definitely an area that needs further research, and would improve our method along with other algorithms for digital terrain analysis. We have added this to our new section in the discussion (see reply to general comment above).

(b) Your wording hints that there are problems in terrace identification when a river exists below a high plateau surface, and that these require some special parameter choices. This could also be aided by a tool to identify valley widths, but a more satisfying explanation about possible failure modes and ways around them especially considering the range of upland topographies from steep lands with ridges to flat upland plateaus would be more satisfying.
Yes, the method does not work in areas as well where there is less distinction between the relief structure of the surrounding topography (for example, the plateau surface in the Le Sueur River site) compared to the floodplains or terraces. We have added in some more discussion of the results for the Le Sueur River to clarify the difficulties of automatic feature extraction in these landscapes:
Lines 460-465: 'The Le Sueur River is currently incising through Pleistocene tills, forming a low-gradient surface or plateau (Fisher, 2003; Gran et al., 2009; Belmont et al., 2011). High-altitude, low-gradient surfaces, such as relict plateaus, may result in error in the method due to the difficulty in distinguishing the distribution of terrace elevations from these low-relief surfaces. The Le Sueur River basin is also heavily influenced by human land use, which makes feature extraction challenging (Passalacqua et al., 2012).'

2. Second, and related: I wonder why you chose a Gaussian distribution as the 'landscape null

hypothesis' from which you search for variations. I see the power in its simplicity, but do wonder whether you could replace the Gaussian distribution with the distribution expected from a stream-power-erosion plus hillslope-diffusion (I'll write it in a linear way here) simple model: $\delta z/\delta t = -k_{SP}A^m S^n + k_{HS} \bigtriangledown^2 z$. *By integrating through time (e.g., numerically with a landscape evolution model), one can generate a non-Gaussian 'landscape null hypothesis'. This to me would seem a more powerful approach insofar as it represents what is expected on the landscape in absence of floodplains and terraces, but does have sensitivity to the k values chosen (or calibrated to the given landscape with another automated procedure). Nevertheless, I think that some of the by-hand "tweaking" with the quantile-quantile plots could be reduced by comparing the measured landscape against a more physically-based elevation distribution. To be clear: I am happy to see this paper published without changing its entire basis, but would feel remiss to not leave a record of this idea as a potential future avenue for improvement.*
We chose the Gaussian distribution to use as the reference distribution for the elevation distribution in the landscape in order to keep the approach general, so that it could be applied across multiple landscapes with varying relief, and to limit the amount of user-defined parameters in the method as much as possible. Furthermore, the Gaussian distribution has also been used in feature extraction algorithms previously (e.g. Lashermes et al., 2007; Passalacqua et al., 2010). The idea of using a simple stream power and hillslope diffusion model to generate a distribution of elevation and slopes as a 'null hypothesis' is a very interesting one, which we could potentially apply to improve our method in the future. However, it may actually generate more user-defined parameters than the Gaussian distribution does at the moment (as you say, this may be sensitive to both erodibility and hillslope diffusivity). Although we feel it is beyond the scope of the paper to add this in at the moment, we have expanded our discussion to include some more of the potential limitations of using the Gaussian distribution to model relief.
Lines 480-489: 'This may be the case if the threshold for elevation compared to the channel selected by the quantile-quantile plot is lower than that of the highest terrace elevations. This can be examined for the landscape in question by a visual inspection of the quantile-quantile plots and the location of the threshold compared to the distribution of channel relief (e.g. Figure 2). Our method fits a Gaussian distribution to the quantile-quantile plots, and selects the thresholds as the deviation of the real data from this distribution, as a simple general model of elevation distributions that can be applied across multiple landscapes. However, in some landscapes, the distribution of elevations may not be accurately represented by a Gaussian distribution. A future avenue for development of this method may be to include multiple models for elevation distributions from which to select the thresholds of elevation and gradient.'

In both of these cases, I think that your approach is the right set of steps towards a process that is fully automated, and think that the places in which it is not fully automated serve to highlight areas in which advances are needed; such advances can lie outside of the scope of this paper.
This is a good point - again we have tried to address this by adding in our new section to the discussion (Section 5.3).

Specific comments

Line 130: I'm interested to see how well the method of selecting breaks in the quantile/quantile plots performs, especially in landscapes with differing hypsometries
See response to general comment 2.

Line 164: I understand what you mean, but not how you wrote it
We have reworded this sentence to make it clearer:
Line 162: 'Following the methodology of Passalacqua et al. (2010a), we set the number of iterations (t)

to 50 and the calculation of λ as the 90% quantile.'

Line 205: Does this take away from the full automation? Maybe "much more fully automated" but not completely? (I can also imagine that you see this as negligible compared to other factors that led to more emphasis on the "semi" part of "semi-automation".)
See response to general comment 1.

Line 214: This also leads to more 'semi' automation, but I don't think this is a bad thing. The problem of automating a method to find the width of a valley is, in itself, an unsolved problem!
Again, we have added in a section to the discussion about the user-defined parameters, and suggested that a future research need is a method to automatically determine valley width, although in order to get a representative hypsometry distribution for the quantile-quantile plots, the width of the swath needs to include the hillslopes on either side of the valley. It should represent the full length scale of the ridge-to-valley topography. Again, this is a difficult problem to determine automatically from the topography alone.

Line 231: Is there any landscape characteristic that you could use to help guide which values of these to pick? Otherwise I worry that you are undercutting your point that you have a fully-automated method.
At the moment, the user can set the values of these percentiles for the reference Gaussian distribution. The best way to do this is to visually assess the quantile-quantile plots for the landscape (e.g. Figure 2), and determine whether the Gaussian distribution is fit through the real data appropriately. We have added this in to the discussion (see answer to general comment 2).

Line 250: You jump into 'two of the four', etc. below; could you tell the reader which are included in each cluster of four? (Addendum: you define the second cluster of four for terraces, below, but not the first cluster of four for floodplain extent.)
Done

Line 428: This is probably just a sign of the times changing, but when I hear "USGS NED", I think of the digitized contours and the associated stair-step errors. Could you clarify if these are from SRTM, perhaps above at the first introudction of the NED?
The DEMs used in the study were from the USGS NED as stated our manuscript: the NED comes from a variety of sources, including SRTM, and LiDAR, which is merged together into a seamless DEM for the US. It is now updated on a two-monthly cycle to include new elevation data as it is produced. In some places we think it may still use digitised contours, which does have its problems, but it is difficult to determine where this is the case. We have added in a sentence clarifying the source datasets for the USGS NED:
Line 355: 'The USGS NED is a seamless dataset created for the conterminous US, using a variety of elevation products which is updated on a two-month cycle.'

Line 436: Is the issue that the Des Moines Lobe till surface (flat!) that lies above the incised valley confounds attempts to look just at the valley bottom? In other words – will the algorithm in general perform poorly on incised-plateau landscapes, and well on landscapes with ridges and no high flat surfaces that are not terraces?
We have added in some extra discussion of the limitations of the terrace method in the Le Sueur site, and generally in areas where there may be relict plateaus or low gradient surfaces that are not terraces (see answer to general comment 1(b)).

Line 445: As I was thinking – fully-automated within a specific scope.
See answer to general comment 1.

Line 479: I think you still need to define better the scopes within which it is objective, or how it is better than the previous methods. I like the quantile-quantile approach, but do have concerns that the user's ability to pick better intercepts 'by eye' begs a clear definition between the objective and subjective portions of your metrics. It also opens the door to questions regarding the subjectivity, which may result from other landscape metrics not yet being developed.
We have added in extra clarification of the automation of the method and the setting of the user defined parameters (answer to general comment 1, and Section 5.3 of the revised manuscript). We have also added in extra discussion of the quantile-quantile plots and the potential limitations of using the Gaussian distribution to fit as a reference, in answer to general comment 2.

Line 484: rm preceding comma if this clause is related only to "swath profile"
Done

Table 2: dolostones [this is picky, but "dolomite" is the mineral even though it is colloquially used for the rock]
Done

Table 2: Was this a placeholder?
Yes...we have changed this in the revised manuscript!

Reviewer 3

This paper presents a new technique for mapping floodplains and terraces from digital elevation models. The paper is generally well written and the approach is both novel and useful. My biggest concern is the authors claim that the tool is fully automated, when it does not really produce reliable maps in fully automated mode and would require users to manually edit maps to make them reliable, just as is the case with any of the other semi-automated techniques out there. I would suggest the authors tone down the somewhat disparaging comments regarding existing semi-automated techniques and at the same time tone down the sales pitch on their method being fully automated (just add a caveat that user interaction is needed to produce reliable maps). Aside from that concern and a few other minor question and suggestions below I believe the paper will make a nice contribution to ESD.
Thank you very much for your comments on our manuscript. In response to your concerns, along with those from the other two reviewers, we have made a clear distinction in the paper about which part of the method are automated, and which parts still require user-defined parameters. We did not intend to be dismissive of other techniques of identifying floodplains and terraces - we agree that these methods are very useful, and have stated this in our manuscript. We have tried to build on these methods by developing statistical techniques for the selection of the thresholds in our method of elevation compared to the channel and local gradient. We have made clearer in our discussion that we believe that the different methods are valuable depending on the scale of the analysis, as well as the field site from which the floodplains/terraces are being extracted.

Lines 92-99: This explanation is not articulated well. I suggest revising, and perhaps condensing this section on Dodov and Foufoula-Georgiou. It seems to be a disproportionate amount of information compared to other studies discussed and the extent to which this information is utilized in the rest of the paper.
We have condensed this section as suggested.

Line 113: Overprediction is a feature, not a bug. These are decidedly semi-automated approaches and it is a benefit if the automated portion of the tool slightly overpredicts because it is easy for the user to manually clip polygons.
We have added in a sentence here to state that the user can manually clip the over-predicted surfaces and remove areas selected incorrectly:
Line 108-111: 'These semi-automated methods allow the user to manually clip over-predicted terrace surfaces based on field data and DEM observations, and remove selected surfaces that do not represent terraces, such as roads, alluvial fans, or water bodies (Stout and Belmont, 2014)'

Line 179: So in the end you use Optimal Weiner filter, correct? If so, why go into detail about Perona-Malik? I suggest either making a better connection between the two filters and explaining how the Perona-Malik equations relate to the Open Weiner filter, or reduce discussion on P-M and instead provide more detail on the OW filter.
We use the Perona-Malik filter for the method of floodplain/terrace extraction. The Perona-Malik filter is a non-linear filter which enhances the transition between features, such as hillslopes/valleys, while preferentially smoothing low gradient surfaces, such as floodplains or terraces. The Optimal Wiener filter is only used here for the extraction of the channel networks using the method outlined by Grieve et al. (2016, ESURF). We have added a sentence to clarify this in the manuscript:
Line 173: 'The Optimal Wiener filter is only used to extract the channel network: we use the Perona-Malik filtering to extract the floodplains and terraces.'

Line 202: terrace should be terraces
Done

Line 203: The authors dont provide any evidence that third order is a reasonable threshold. I have frequently seen terrace features on first and second order streams in places in the northeastern, Midwestern and western US. I suggest removing this arbitrary suggestion and simply explaining how the user should determine what the threshold should be for their particular landscape.
In each of our field sites we found that a third order threshold was appropriate for where the terraces initiated in the landscape (see Figures 7 and 8). We have changed this section to state this, and we have clearly stated that a visual inspection of the DEM compared to the channel network should allow the user to select the appropriate threshold stream order:
Lines 203-207: 'We found that a threshold of third order channels was appropriate for each of our field sites, based on a visual inspection of the DEM. One of the outputs of our software package is a raster of the channel network labelled by the Strahler stream order. The user can identify an appropriate threshold stream order based on visual inspection of floodplain and terrace surfaces compared to this network.'

Lines 220-234: The authors spend a lot of time explaining quantile-quantile plots. Such explanations may be best left for textbooks as q-q plots are fairly routine, but I leave it to the authors to decide whether or not it is necessary to include. More importantly, I think it is important that the authors explain why it is reasonable to assume that local gradients would follow a Gaussian distribution and why deviations from Gaussian are likely to be transitions between process domains.
We believe that it is important to include the description of the quantile-quantile plots as this is a key part of our methodology for selecting the thresholds of gradient and elevation compared to the channel from the DEMs. We chose a Gaussian distribution as a simple model, which can be applied generally over a range of landscapes, and has been used in previous methods of feature extraction (Lashermes et al., 2007; Passalacqua et al., 2010). We have added in some more discussion about the Gaussian distributions in response to this comment plus comments from Reviewer 2.

Lines 478-489: 'Furthermore, in some cases our method did not select all of the terraces identified by the field mapping, particularly at the highest elevations compared to the modern channel (e.g. Figure 7c and d). This may be the case if the threshold for elevation compared to the channel selected by the quantile-quantile plot is lower than that of the highest terrace elevations. This can be examined for the landscape in question by a visual inspection of the quantile-quantile plots and the location of the threshold compared to the distribution of channel relief (e.g. Figure 2). Our method fits a Gaussian distribution to the quantile-quantile plots, and selects the thresholds as the deviation of the real data from this distribution, as a simple general model of elevation distributions that can be applied across multiple landscapes. However, in some landscapes, the distribution of elevations may not be accurately represented by a Gaussian distribution. A future avenue for development of this method may be to include multiple models for elevation distributions from which to select the thresholds of elevation and gradient.'

Line 240: In what way to do you mean connected to the modern channel? Certainly terraces can abut the modern channel.
The method identifies patches of floodplain as those which are at a similar elevation to the modern channel (based on the extracted channel network), whereas terraces should be at a higher elevation compared to the channel. This was not clear in our original wording: we have rephrased this and added more discussion about the separation between floodplains and terraces to the manuscript based on comments from Reviewer 1.

Line 296: How and why did you separate flood zones into 100 year and greater than 100 year flood risk? Just based on comparison with the FEMA maps? If so, are the FEMA maps necessarily reliable? Many would consider floodplains above the 100 year flood flood zone to be terraces. At what point do you make this distinction?
The separation of flood zones into 100-year and greater than 100 year was on the FEMA maps which are classified based on the annual percentage chance of flooding. There may be some errors with the FEMA flood maps based on this: this may a cause of some of the discrepancies between the floodplains extracted from our method and with these published maps. We have a section in our manuscript discussing some of the potential problems with the FEMA flood maps:
Lines 423-431: 'Published flood maps are useful in providing an independent estimate of likely floodplains in each field site. However, there are potential limitations to these maps which must be carefully considered, and may result in some of the differences compared to geomorphic floodplain prediction techniques. Hydrodynamic models have a large number of parameters, which require careful calibration with field and hydraulic data, such as channel roughness and discharge data from gauging stations. Furthermore, due to the time-consuming and expensive nature of these studies, flood maps are often not produced for small catchment sizes, and may therefore be incomplete on a landscape-scale (e.g. Figure 4). There may also be differences in the methodology used in producing these maps for each site, depending on the input topographic data and modelling software used.'
The distinction between floodplains and terraces is something that may also cause some problems in our method, especially when the terraces are close in elevation to the modern channel. We have also added in more discussion about this to our manuscript based on comments from Reviewer 1.

Table 4: The authors were somewhat disparaging about semi-automated approaches that have been developed earlier. Seeing these reliability and sensitivity values, I would suggest that the tool they have developed is no different. In comparisons with mapped terraces the tool is mapping a lot of false positives and false negatives. To map terraces reliably a user would need to manually edit these extensively...thats fine...its to be expected, really...and thats why previous algorithms have claimed to be semi-automated. But I would urge the authors not to make claims about it being a fully automated process when the

automated process fails to produce a reliable map.
We did not intend at all to be disparaging about semi-automated approaches that have been previously developed: we think these methods are very useful, particularly in areas where there is some field data available to calibrate the selection of thresholds and user-defined parameters. We have tried to build on these methods by developing statistical techniques for the selection of the thresholds in our method of elevation compared to the channel and local gradient. As previously stated, we have now made a clear distinction in our manuscript between the user-defined parameters and these thresholds which are calculated statistically. We have made clearer in our discussion that we believe that the different methods are valuable depending on the scale of the analysis in question and location from which the floodplains/terraces are being extracted:
Lines 492-497: 'Semi-automated methods of terrace identification, where the terrace polygons are manually edited by the user, are particularly useful in areas where independent datasets of terrace locations are available for calibration, and may be more appropriate than our method on site-specific scales (e.g. Stout and Belmont, 2014). However, the selection of thresholds based on a objective statistical approach means that our method can be applied in areas where these data do not exist, on a broader landscape scale, or as a rapid first-order predictor of terrace locations.'

Lines 445-450: I dont think the authors have made a strong case that their method produces reliable maps as a fully automated system. I agree that their method is a useful first cut, but this is no different from Stout and Belmont or any of the other semi-automated approaches mentioned in the paper.
See reply to comment above.

Line 469: There are several other key papers that could be cited as examples of using terraces to quantify sediment budgets: Trimble, S. W. (1999). Decreased rates of alluvial sediment storage in the Coon Creek Basin, Wisconsin, 1975-93. Science, 285(5431), 1244-1246. Belmont, P., Gran, K. B., Schottler, S. P., Wilcock, P. R., Day, S. S., Jennings, C., ... & Parker, G. (2011). Large shift in source of fine sediment in the Upper Mississippi River. Environmental science & technology, 45(20), 8804- 8810. Brown, A. G., Carey, C., Erkens, G., Fuchs, M., Hoffmann, T., Macaire, J. J., ... & Walling, D. E. (2009). From sedimentary records to sediment budgets: multiple approaches to catchment sediment flux. Geomorphology, 108(1), 35-47.
We have added in the suggested references.

Line 474: Several key papers needed to substantiate this statement as well. Lots of examples, such as: Pazzaglia, F. J., & Brandon, M. T. (2001). A fluvial record of long-term steady-state uplift and erosion across the Cascadia forearc high, western Washington State. American Journal of Science, 301(4-5), 385-431. Avouac, J. P., & Peltzer, G. (1993). Active tectonics in southern Xinjiang, China: Analysis of terrace riser and normal fault scarp degradation along the HotanA RQira fault system. Journal of Geophysical Research: Solid Earth, 98(B12), 21773-21807. Viveen, W., Schoorl, J. M., Veldkamp, A., & Van Balen, R. T. (2014). Modelling the impact of regional uplift and local tectonics on fluvial terrace preservation. Geomorphology, 210, 119-135.
We have added in the suggested references.

Manuscript prepared for Earth Surf. Dynam.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 25 May 2017

Geomorphometric delineation of floodplains and terraces from objectively defined topographic thresholds

Fiona J. Clubb[1], Simon M. Mudd[1], David T. Milodowski[2], Declan A. Valters[3], Louise J. Slater[4], Martin D. Hurst[5], and Ajay B. Limaye[6]

[1]School of GeoSciences, University of Edinburgh, Drummond Street, Edinburgh, United Kingdom, EH8 9XP
[2]School of GeoSciences, University of Edinburgh, Crew Building, King's Buildings, Edinburgh, United Kingdom, EH9 3JN
[3]School of Earth, Atmospheric, and Environmental Science, University of Manchester, Oxford Road, Manchester, United Kingdom, M13 9PL
[4]Department of Geography, Loughborough University, Loughborough, United Kingdom, LE11 3TU
[5]School of Geographical and Earth Sciences, East Quadrangle, University of Glasgow, Glasgow, United Kingdom, G12 8QQ
[6]Department of Earth Sciences and St. Anthony Falls Laboratory, University of Minnesota, Minneapolis, Minnesota, USA

Correspondence to: Fiona J. Clubb (f.clubb@ed.ac.uk)

Abstract. Floodplain and terrace features can provide information about current and past fluvial processes, including channel response to varying discharge and sediment flux; sediment storage; and the climatic or tectonic history of a catchment. Previous methods of identifying floodplain and terraces from digital elevation models (DEMs) tend to be semi-automated, requiring the input of independent datasets or manual editing by the user. In this study we present a new ~~, fully automated~~ method of identifying floodplain and terrace features based on two thresholds: local gradient, and elevation compared to the nearest channel. These thresholds are calculated statistically from the DEM using quantile-quantile plots and do not need to be set manually for each landscape in question. We test our method against field-mapped floodplain initiation points, published flood hazard maps, and digitised terrace surfaces from seven field sites from the US and one field site from the UK. For each site, we use high-resolution DEMs derived from light detection and ranging (LiDAR) where available, as well as coarser resolution national datasets to test the sensitivity of our method to grid resolution. We find that our method is successful in extracting floodplain and terrace features compared to the field-mapped data from the range of landscapes and grid resolutions tested. The method is most accurate in areas where there is a contrast in slope and elevation between the feature of interest and the surrounding landscape, such as confined valley settings. Our method provides a new tool for rapidly and objectively identifying floodplain and terrace features on a landscape

scale, with applications including flood risk mapping, reconstruction of landscape evolution, and quantification of sediment storage and routing.

1 Introduction

Identifying the location of floodplains and fluvial terrace features can provide important insights into geomorphic and hydrological processes. Understanding the controls on floodplain inundation carries increasing societal importance, as the frequency of flood events is predicted to increase with the rise in global temperatures and varying patterns of precipitation caused by climate change (Schreider et al., 2000; Booij, 2005; Hartmann et al., 2013). Although there are still large uncertainties regarding the impacts of climate change on flood frequency (Booij, 2005), identifying floodplains is crucial for forecasting and planning purposes. On longer timescales, the morphology and structure of fluvial terraces can provide important information on channel response to climatic, tectonic, and base-level variations (Bull, 1991; Merritts et al., 1994; Pazzaglia et al., 1998); the relative importance of lateral and vertical channel incision (Finnegan and Dietrich, 2011); and sediment storage and dynamics (Pazzaglia, 2013; Gran et al., 2013).

Attempts to identify floodplains can be classified into two broad families of methods: (i) flood risk mapping and hydrological modelling~~;~~, and (ii) geometric terrain classification. Traditionally, identification of floodplains has relied upon the creation of flood hazard maps, produced through detailed hydraulic modelling studies (e.g. Noman et al., 2001; Grimaldi et al., 2013). These studies tend to incorporate historical flood event information, hydrological analyses, and hydraulic flow propagation models (Degiorgis et al., 2012). These mature techniques can lead to accurate flood inundation predictions down to the level of a single building (e.g. Horritt and Bates, 2002; Cobby et al., 2003; Guzzetti et al., 2005; Hunter et al., 2007; Kim et al., 2012). However, these models can be computationally expensive and time-consuming to run, even in one dimension, requiring the calibration of large numbers of parameters, all with their own uncertainties (e.g. Beven, 1993; Horritt and Bates, 2002; Liu and Gupta, 2007). This means that hydraulic simulations are usually performed at cross sections across the channel and interpolated to cover the rest of the stream network (Noman et al., 2001; Dodov and Foufoula-Georgiou, 2006). For example, floodplain mapping tools have been developed that incorporate either field-based or modelled stage-duration information at multiple cross sections along the channel, and interpolate a three-dimensional water surface between these sections (e.g Belmont, 2011; Yang et al., 2006).

The introduction of high-resolution digital elevation models (DEMs) has provided the opportunity ~~of mapping~~ to map floodplain features much more rapidly and over larger spatial scales than previously possible (Noman et al., 2001). This had led to the development of many different methods that rely on extracting a variety of topographic indices from DEMs, such as local slope, contributing area, and curvature (Manfreda et al., 2014). One common metric used to predict floodplains is the

topographic index ($\phi = ln(A/(tan\beta))$), where A is the contributing area to each cell (m^2) and β is the local slope in degrees (e.g. Kirkby, 1975; Beven and Kirkby, 1979; Beven et al., 1995; Quinn et al., 1995; Beven, 1997). The contributing area term reflects the tendency of water to accumulate at certain regions of the basin, whereas the slope term represents the tendency for gravity to transport water downhill. Therefore, high values of the topographic index represent areas which are likely to saturate first, as they have a large contributing area compared to local slope (Beven, 1997). Manfreda et al. (2011) suggested a modified version of the topographic index, changing the weighting on the area term by raising it to an exponent n. This modification allows the relative importance of slope or contributing area to be changed by varying the n parameter. They proposed that floodplains can be identified as cells with a modified topographic index (ϕ_m) greater than a threshold value, τ. However, this method requires calibration of the parameters τ and n through comparing the output floodplain map with a pre-existing hazard map, and noting the occurrence of true and false positives and negatives (Manfreda et al., 2011).

Another geometric method that has been developed to identify floodplains uses a series of linear binary classifiers for a number of topographic metrics (Degiorgis et al., 2012). Five different parameters are sampled from the DEM (slope, contributing area, elevation from nearest channel, distance from nearest channel, and curvature), and each cell is classified as either 1 (floodplain) or 0 (non-floodplain) depending on whether these parameters are above or below threshold values. Each of these five metrics can be considered in isolation or in pairs. The thresholds are calibrated using flood hazard maps, where the number of true and false positives and negatives are noted, similar to the approach of (Manfreda et al., 2011). For each parameter and threshold value the Receiver Operating Characteristics (ROC) curve (e.g. Fawcett, 2006) is calculated, which is defined by the number of true and false positives. The maximum area under the curve is determined to allow the threshold value for each parameter to be calibrated, as well as comparisons between each parameter to be found. The pair of best-performing features was identified as the distance (D) and elevation (H) from the nearest channel (m). This method is also semi-automated, as it requires the existence of flood hazard maps for at least some part of the catchment in order to select the correct binary classifiers for floodplain identification.

Dodov and Foufoula-Georgiou (2006) ~~distinguish between the 'geomorphic floodplain', or GF, which represents the morphology of the floodplain compared to its natural boundaries, and the 'submerged floodplain', SF, which represents the part of the floodplain inundated by a specific magnitude flood event. The GF will remain fixed over the scale of multiple flood events, and should be clearly distinguished based on geometric features extracted from the DEM. The SF, however, will vary through time with each flood event, and may be more appropriate to determine based on hydraulic modelling studies. Dodov and Foufoula-Georgiou (2006)~~ present an algorithm for identifying ~~the GF~~ floodplains over large scales based on information on bankfull channel depths. They suggest that the morphology of the ~~GF~~ floodplain is defined by the lateral channel migration rate

through time, and is controlled by the transport of water and sediment by the channel. Therefore, they assume that the geometry of the ~~GF~~ floodplain is related to that of the channel, and demonstrate a relationship between bankfull channel depths and floodplain inundation depths which is linear over a range of scales (Dodov and Foufoula-Georgiou, 2006). Floodplain delineation is carried out by locally filling the DEM up to the depth of inundation, which is determined based on bankfull channel depths, calibrated using data from United States Geological Survey (USGS) gauging stations across Oklahoma and Kansas, along with field measurements. The depth of inundation at points along the channel network is then used to find the lateral extent of the floodplain by using the planform curvature of the channel. This method also requires significant user input, as the channel bankfull depths are required in order to estimate the ~~GF~~ inundation depth.

The extraction of fluvial terraces (the remnants of previous floodplains) represents a closely related problem to the delineation of presently active floodplain surfaces. Previous studies have also used a geometric approach to identify terrace features from DEMs. For example, Demoulin et al. (2007) identified terrace surfaces based on local slope and height of each pixel compared to the channel. They used these attributes in order to reconstruct palaeo-channel profiles from terrace surfaces, but their methodology was not designed to produce a map of terrace extents on a wider landscape scale. Therefore, following on from their approach, Stout and Belmont (2014) presented the TerEx toolbox, a semi-automated tool to identify potential terrace surfaces based on thresholds of local relief, minimum area, and maximum distance from the channel. After potential terrace surfaces are identified, their area and height above the local channel are measured. The tool then allows the user to edit the terrace surfaces based on comparison with field data. Hopkins and Snyder (2016) evaluated the TerEx toolbox, along with two other semi-automated methods for identifying terrace surfaces (Wood, 1996; Walter et al., 2007) at the Sheepscot River, Maine. They found that all of the methods ~~overpredicted~~ over-predicted terrace areas compared to the field-mapped terraces, and the accuracy of the methods decreased in lower relief landscapes. These semi-automated methods allow the user to manually clip over-predicted terrace surfaces based on field data and DEM observations, and remove surfaces selected that do not represented terraces, such as roads, alluvial fans, or water bodies Stout and Belmont (2014).

The geomorphic methods of mapping both terraces and floodplains outlined above are all semi-automated, requiring independent datasets and significant user input. For example, the method proposed by Manfreda et al. (2011) requires the parameters to be optimised using flood inundation maps from hydraulic simulations. The linear binary classifiers outlined by Degiorgis et al. (2012) and tested by Manfreda et al. (2014) use flood hazard maps to select the correct threshold for floodplain prediction from the geomorphic indices. The TerEx toolbox, developed by Stout and Belmont (2014), requires significant user input in order to manually edit the predicted terrace surfaces. No existing approach to mapping either floodplains or terraces from topographic data includes objective criteria for setting the thresholds that identify floodplains and terraces. As a result, the different

thresholds that a user might select can result in varying floodplain and terrace maps for the same input DEM, complicating efforts to consistently map geomorphic features between different landscapes.

Here we introduce a new ~~, fully automated~~ method of identifying floodplain and terrace surfaces from topographic data. This method uses two geometric thresholds ~~which~~ that can be readily extracted from DEMs: the gradient of each pixel, and the elevation of each pixel relative to the nearest channel. Importantly, this method does not require calibration using any independent datasets, as the thresholds are statistically calculated from the DEM using quantile-quantile plots. We test our method against field-mapped floodplain initiation points, published flood hazard maps, and digitised terrace surfaces from seven field sites throughout the US and one site in the UK (Figure 1). For each site, where available, we use high-resolution LiDAR-derived DEMs, as well as the corresponding national elevation datasets (10 m resolution for the US and 5 m for the UK) in order to test the sensitivity of our method to grid resolution.

2 Methodology

Floodplain and terrace surfaces can be defined as low relief, quasi-planar areas capped by alluvium and found proximal to the modern river channel. Therefore, field mapping campaigns typically identify these surfaces as spatially continuous areas with low gradients that occur next to the channel. We present a new geometric method which replicates this field approach as closely as possible by using two metrics which can be readily extracted from the DEM: elevation compared to the nearest channel, and local gradient. Our method is efficient to run and is ~~fully automated~~based on the fully automated selection of topographic thresholds, requiring no input of independent datasets or field mapping. We outline below the DEM pre-processing steps followed by the methodology for identifying floodplain and terrace features.

2.1 DEM pre-processing

The first step of the algorithm is to smooth the DEM in order to remove micro-topographic noise. Gaussian filters are often used to smooth DEMs, where the smoothing can be described by linear diffusion. A Gaussian filter results in the DEM being smoothed uniformly at all locations and in all directions (e.g. Lashermes et al., 2007). However, one consequence of the Gaussian filtering is the loss of information where there are sharp boundaries between features due to the uniform smoothing. Therefore, we filter the input DEM using a non-linear filter proposed by Perona and Malik (1990), and applied to channel extraction from high-resolution topography by Passalacqua et al. (2010a). The Perona-Malik filter is an adaptive filter in which the degree of smoothing decreases as topographic

gradient increases (Perona and Malik, 1990; Passalacqua et al., 2010a). This non-linear diffusion equation can be described as:

$$\partial_t h(x,y,t) = \bigtriangledown.[p(|\bigtriangledown h|)\bigtriangledown h] \quad (1)$$

where h is the elevation at location (x,y) and time t, $\bigtriangledown$ is the gradient operator, and $p(|\bigtriangledown h|)$ is an edge-stopping function that specifies where to stop diffusion across feature boundaries, where:

$$p(|\bigtriangledown h|) = \frac{1}{1+(|\bigtriangledown h|/\lambda)^2} \quad (2)$$

where λ is a constant. Importantly for the identification of low-gradient surfaces, the Perona-Malik filtering enhances the transitions between features, such as the low-gradient valley floor and the surrounding hillslopes, while preferentially smoothing low gradient reaches of the DEM. Following the methodology of Passalacqua et al. (2010a), we set the ~~time of forward to diffusion~~ number of iterations (t) to 50 ~~iterations~~ and the calculation of λ as the 90% quantile. We keep these parameters constant across each site tested in the study. A full explanation of these parameters and derivation of the Perona-Malik filter is described by Passalacqua et al. (2010a).

After the DEM is smoothed, we then extract the channel network. Many studies have proposed different methods for identifying channel networks from high-resolution topography (e.g. Lashermes et al., 2007; Tarolli and Dalla Fontana, 2009; Passalacqua et al., 2010b, 2012; Pelletier, 2013; Clubb et al., 2014). Grieve et al. (2016c) tested the validity of channel extraction algorithms at coarsening DEM resolution, and found that a geometric method of channel extraction was consistent up to DEM resolutions of 30 m. This method, described by Grieve et al. (2016b), uses an Optimal Wiener filter to remove micro-topographic noise from the DEM (Wiener, 1949; Pelletier, 2013). The Optimal Wiener filter is only used to extract the channel network: we use the Perona-Malik filtering to extract the floodplains and terraces. Channelised portions of the landscape are selected using a tangential curvature threshold (Pelletier, 2013), which is defined using quantile-quantile plots as described by ~~Lashermes et al. (2007); Passalacqua et al. (2010a)~~Lashermes et al. (2007) and Passalacqua et al. (2010a). These channelised portions of the landscape are combined into a channel network using a connected components algorithm outlined by He et al. (2008), and thinned using the algorithm of Zhang and Suen (1984). We chose this algorithm for channel extraction to allow consistency when running our method on DEMs of varying grid resolutions.

2.2 Floodplain and terrace identification

After smoothing the DEM, the user can choose to run the terrace and floodplain mapping algorithm across the whole DEM, or to extract the floodplains and terraces relative to a specific channel of interest. If the algorithm is run on the whole DEM, the local gradient, S, and relief relative to the

nearest channel, R_c, are calculated for each pixel. These two parameters were chosen on the basis that floodplains and terraces tend to form low-gradient regions ~~which~~ that are close to the elevation of the modern channel. Local gradient has been used in previous geometric methods of floodplain and terrace identification, both in the calculation of the topographic index (Kirkby, 1975; Manfreda et al., 2011), and in combination with other topographic metrics (e.g. Degiorgis et al., 2012; Stout and Belmont, 2014; Limaye and Lamb, 2016). Local gradient was calculated by fitting a polynomial surface to the DEM with a circular window (e.g. Lashermes et al., 2007; Roering et al., 2010; Hurst et al., 2012; Grieve et al., 2016a). The radius of the window is calculated by identifying breaks in the standard deviation and interquartile range of curvature with increasing window size, following Grieve et al. (2016a). This allows the window size to be calculated for each DEM to ensure that the slope values are representative at the hillslope scale, rather than being influenced by smaller-scale variations from vegetation (e.g. Roering et al., 2010; Hurst et al., 2012). R_c has also been used in previous geometric methods (e.g. Degiorgis et al., 2012; Manfreda et al., 2014; Limaye and Lamb, 2016), and is calculated as the difference in elevation between the starting pixel and the nearest channel pixel, identified using a steepest descent flow routing algorithm (O'Callaghan and Mark, 1984; Braun and Willett, 2013). A threshold Strahler stream order is set by the user such that the nearest channel must have a stream order greater than the threshold. This is necessary so that each pixel is mapped to the main channel along which floodplains or ~~terrace~~ terraces have formed, rather than narrow tributary valleys. We ~~suggest~~ found that a threshold of third order channels ~~is appropriate for most landscapes, but this can be determined easily by the user from~~ was appropriate for each of our field sites, based on a visual inspection of the ~~channel network~~DEM. One of the outputs of our software package is a raster of the channel network labelled by the Strahler stream order. The user can identify an appropriate threshold stream order based on visual inspection of floodplain and terrace surfaces compared to this network.

As well as running the algorithm on the whole landscape, the user can also choose to extract floodplains or terraces relative to a specific channel of interest. The user must provide the latitude and longitude of two points defining the upstream and downstream end of the channel. The algorithm then defines a channel network between these points using a steepest descent flow routing algorithm (O'Callaghan and Mark, 1984; Braun and Willett, 2013). After the identification of the channel, a swath profile is created along it following the method outlined in Hergarten et al. (2014) and applied by Dingle et al. (2016). The user must specify the width of the swath, which can be estimated by a visual inspection of the DEM, to provide a sufficiently wide swath compared to the valleys in the landscape. The same two parameters (S and R_c) are used for feature classification for each pixel in the swath profile, except that R_c is calculated compared to the nearest point on the reference channel.

After the calculation of slope and R_c, we identify thresholds for each metric in order to provide a binary classification of each pixel as either floodplain/terrace (1) or hillslope (0). A key feature of our new method is that the thresholds for R_c and local gradient do not need to be set by the user

based on independent validation, but are calculated statistically from the DEM. ~~These thresholds are identified~~ 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 ~~geomorphic process domains~~ hillslope-valley transitions (e.g. Lashermes et al., 2007; Passalacqua et al., 2010a). Quantile-quantile plots are used to determine if a probability density function of real data can be described by a Gaussian distribution. The transition between process domains can be determined by the value at which the probability density function of the real data deviates from the Gaussian function (Lashermes et al., 2007). The real data ~~is~~ are plotted against the corresponding standard normal variate, which indicates how many standard deviations an element is from the mean. For example, if a value has a standard normal variate (or z-score) of 1, then it is one standard deviation above the mean, which has a z-score of 0. A Gaussian distribution plots as a straight line on a quantile-quantile plot, and is modelled for each DEM based on a lower and upper percentile of the real data. The percentiles chosen to represent the reference Gaussian distribution can be set by the user based on the landscape in question, but are generally set as the 25th and 75th percentile (Passalacqua et al., 2010a). For each value of the real data, we calculate the difference between the real data and the Gaussian distribution as a fraction of the range of the real data (Figure 2). The threshold values for R_c and slope are then identified as the lowest value at which there is less than 1% difference between the two distributions. Figure 3 shows an example of the channel relief and slope maps for the Russian River field site, with the calculated thresholds for each field site presented in Table 1.

~~After the selection of pixels which are below the threshold for both S and R_c, the next step of the algorithm is to assign each pixel as either floodplain or terrace. In order to identify discrete patches of floodplain or terrace, we run the connected components algorithm of He et al. (2008), which assigns a unique identifier to each patch. If a patch is connected to the modern channel network it is defined as part of the modern~~ If the user wishes to extract only the terraces, then a threshold height above the modern river channel must be set: any pixels below this height will be identified as floodplain, and ~~if not it is defined as a fluvial terrace. This tool~~ any pixels above this height will be identified as terraces. This threshold height can also be determined based on a visual inspection of the DEM. Our method allows the analysis of spatial extent of floodplain and terrace features (if run across the whole DEM) as well as the distribution along a specific channel of interest (if run with the swath mode). For example, in swath mode, the elevation and slope of the terraces can be mapped as a function of distance upstream along the channel network. This provides numerous potential applications of the method for understanding controls on terrace formation and morphology.

2.3 Comparison with published data

In order to test the results of our method we compare the predicted floodplain and terrace locations to field-mapped floodplain initiation points, published flood hazard maps, and digitised terrace surfaces. In order to quantify the performance of our methods compared to these datasets, we assess the rates of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (e.g. Heipke et al., 1997; Molloy and Stepinski, 2007; Tarolli et al., 2010; Orlandini et al., 2011; Manfreda et al., 2014; Clubb et al., Each pixel is assigned to one of the four categories:

1. True positive TP: The pixel is identified as floodplain/terrace by both the geomorphic method and the independent dataset.
2. False positive FP: The pixel is identified as floodplain/terrace by the geomorphic method, but not by the independent dataset.
3. True negative TN: The pixel is not identified as floodplain/terrace by either dataset.
4. False negative FN: The pixel is identified as floodplain/terrace by the independent dataset but not by the geomorphic method.

We report the reliability (r), sensitivity (s), and overall quality (Q) for each field site:

$$r = \frac{\sum TP}{\sum TP + \sum FP} \tag{3a}$$

$$s = \frac{\sum TP}{\sum TP + \sum FN} \tag{3b}$$

$$Q = \frac{\sum TP}{\sum TP + \sum FP + \sum FN} \tag{3c}$$

The reliability, r, is a measure of the ability of the method to not generate false positives. The r value can vary between 0 and 1: if the r value is low, then the method is predicting a large amount of pixels as floodplain or terrace which are not identified by the independent dataset, whereas as high r value indicates that the majority of pixels mapped as floodplain or terrace are also identified by the independent map. The sensitivity, s, is a measure of the ability of the method to not generate false negatives; a low s value indicates that the method is not identifying many of the floodplain or terrace pixels selected by the published maps. The overall quality, Q, combines both the number of false positives and false negatives to give an overall 'goodness' of the feature classification. It also varies between 0 and 1, where 0 represents no correlation between the predicted and observed features, and 1 represents a perfect match (Heipke et al., 1997).

3 Study areas

We ran our new method on a total of eight field sites, located in Figure 1. Four of these field sites (the Russian River, CA; Mid Bailey Run, OH; Coweeta NC; and the River Swale, UK) were selected

to test the ability of the algorithm to identify floodplains, using published flood maps for the regions. The remaining four sites were selected to validate the algorithm against digitised terrace maps (South Fork Eel River, CA; Le Sueur River, MN; Clearwater River, WA, and Mattole River, CA). Table 2 summarises the mean annual precipitation and mean annual temperature of each site, based on data from the PRISM Climate Group (http://prism.oregonstate.edu) for the US sites and the Met Office (http://www.metoffice.gov.uk/public/weather/climate/) for the UK site. It also summarises the underlying lithology, the source of the data used for validation, and the grid resolution. The algorithm was run based on topographic data derived from 1 m LiDAR data for the sites where these were available (the Russian River, CA; Mid Bailey Run, OH; Coweeta, NC; the South Fork Eel River, CA; and the Le Sueur River, MN). For the remaining field sites the topographic data were generated from the United States Geological Survey National Elevation Dataset 1/3 arc sec DEM, sampled at 10 m resolution for the US sites, and from the Ordnance Survey Terrain 5 dataset for the UK site, sampled at 5 m resolution. All DEMs were converted to the Universal Transverse Mercator (UTM) coordinate system using the WGS84 datum.

4 Results

4.1 Comparison with mapped floodplains

We compare the floodplain extent predicted by the our method to field mapped floodplain initiation points (FIPs) from two of the four study areas: Mid Bailey Run, OH, and Coweeta, NC. A FIP was defined as the upstream limit of low gradient surfaces at the same elevation as the channel banks. As the valley opens out from its more confined upper reaches, these surfaces transition from discontinuous depositional pockets to more continuous floodplain surfaces (Jain et al., 2008). In this study we consider the FIP to start at the onset of alluviation outside the channel banks: therefore, we mapped the start of the discontinuous floodplain pockets at the FIPs in each channel. The onset of alluviation often occurred at multiple locations along the same channel: in these cases we took the location of each FIP downstream along the channel.

A total of 19 FIPs were mapped in Mid Bailey Run, OH, during May–June 2011, and eight FIPs were mapped in the Coweeta catchment, NC, in May 2014. FIPs in the Mid Bailey Run catchment were mapped using a Trimble GeoXM GeoExplorer 2008 series GPS with a mean horizontal accuracy of 6 m. Point locations in the Coweeta catchment were mapped using a Trimble GeoXR GeoExplorer 6000 series GPS with a mean horizontal accuracy of 1.01 m and a mean precision of 1.3 m. Figure 4 shows the relationship between the field mapped initiation points and predicted floodplain extent. In order to compare these field mapped FIPs to our predicted floodplain extents, we measured the flow distance between the field mapped point and the furthest upstream point of the nearest predicted floodplain patch. The distances for each FIP are reported in Table 3, where negative values indicate that the predicted floodplain initiation was upstream of the mapped, and

vice versa for positive values. ~~There was a mean flow distance of 8 ± 10 m between the mapped and predicted for the~~ We also report the r, s, and Q values for the predicted floodplain initiation points. Following the methodology of Orlandini et al. (2011), we classify a point as a TP if the predicted FIP is within a 30 m radius of the mapped FIP. The comparison with the mapped FIPs resulted in $r = 0.83$, $s = 0.67$, and $Q = 0.59$ for Mid Bailey Run~~field site, and a mean flow distance of -6 ± 7 m for the Coweetafield site~~, and $r = 0.78$, $s = 1$, and $Q = 0.78$ for Coweeta.

Along with these field mapped floodplain initiation points, we also compare our predicted floodplain extent to published flood risk maps for three out of the four study areas. For the sites in the US, flood risk maps were obtained from the Federal Emergency Management Agency (FEMA)'s National Flood Hazard Layer (https://msc.fema.gov/portal/). The National Flood Hazard Layer is a compilation of GIS data consisting of a US-wide Flood Insurance Rate map. It contains information on the flood zone, base flood elevation, and floodway status for a location. Floodplain extents are calculated using a hydraulic model, such as HEC-RAS (Hydrologic Engineering Center-River Analysis System), incorporating discharge data, cross sectional survey data, and stream characteristics. These studies can be expensive, with a detailed survey on a mile-long reach typically costing between $10,000 and $25,000 (Committee on FEMA Flood Maps, 2009). The original data were in the geographic projection NAD1983, and were converted to the projected UTM WGS84 coordinate system (Ohio and NC Zone 17N, Russian River Zone 10N). We separate the flood zones into two categories: areas within the 100 year flood (blue), with a 1% annual chance of flooding, and areas with a greater than 100 year flood risk (less than 1% annual risk of flooding). In order to compare these maps to our method, we gridded the FEMA flood risk maps with a resolution of 1 m. The Coweeta field site in North Carolina did not have a complete flood risk map for the catchment and therefore could not be included in this analysis.

For the River Swale field site in the UK, flood risk maps were obtained from the Environment Agency's (EA) Risk of Flooding from Rivers and Sea dataset, which divides the landscape into 50 by 50 m cells (https://data.gov.uk/dataset/risk-of-flooding-from-rivers-and-sea1). Each cell is categorized into one of four flood risk likelihood categories: high (3.3% annual chance of flooding); medium (between 3.3% and 1%); low (between 1% and 0.1%); or very low (<0.1%). The dataset is created by hydraulic modelling, including information about the state of flood defenses and local stage heights as inputs to the model. The data were re-projected from the British National Grid coordinate system to the UTM WGS84 datum, Zone 30N. In order to keep the comparison consistent with the sites from the US, each pixel was classified into the same two categories as for the FEMA maps, with areas of flood risk identified as having greater than 1% annual chance of flooding. The dataset is provided as vector data: to compare with the floodplain identified by the our method, we gridded the vector dataset at 5 m resolution (the same as the input DEM). Figure 5 shows examples of the FEMA and EA flood maps for each study area.

The ~~performance of our geomorphic method of predicting floodplains was compared to flood hazard maps by assessing the rates of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (e.g. Orlandini et al., 2011; Manfreda et al., 2014; Clubb et al., 2014). Each pixel is assigned to one of the four categories:~~

1. ~~True positive TP: The pixel is identified as floodplain by both the geomorphic method and the flood hazard map.~~
2. ~~False positive FP: The pixel is identified as floodplain by the geomorphic method, but not by the flood hazard map.~~
3. ~~True negative TN: The pixel is not identified as floodplain by either dataset.~~
4. ~~False negative FN: The pixel is identified as floodplain by the flood hazard map but not by the geomorphic method.~~

~~Following the methodology of Orlandini et al. (2011), we report the reliability (r) and sensitivity (s) for each field site:~~

$$r = \frac{\sum TP}{\sum TP + \sum FP}$$

$$s = \frac{\sum TP}{\sum TP + \sum FN}$$

~~The reliability, r, is a measure of the ability of the method to not generate false positives. The r value can vary between 0 and 1: if the r value is low, then the method is predicting a large amount of pixels as floodplain which are not identified by the flood hazard maps, whereas as high r value indicates that the majority of pixels mapped as floodplain are also identified by the flood hazard maps. The sensitivity~~, s, ~~is a measure of the ability of the method to not generate false negatives: a low s value indicates that the method is not identifying many of the floodplain pixels selected by the published maps. The r and s~~ and Q values for each site are reported in Table 4, with a visual comparison between the method and the published flood maps shown in Figure 6. We also report the ~~r and s~~ quality values for floodplains extracted from the United States Geological Survey's 1/3 arc second National Elevation Dataset (NED), gridded at 10 m, in order to test the sensitivity of our method to grid resolution.

The USGS NED is a seamless dataset created for the conterminous US, using a variety of elevation products which is updated on a two-month cycle. The method was most similar to the flood risk maps for the Russian River, CA ~~, with high values of both reliability (r) and sensitivity (s~~with the highest overall quality value ($Q = 0.67$ for the 1 m DEM and 0.68 for the 10 m DEM). The method has a higher sensitivity than reliability for both DEM datasets, with $s = 0.97$ and $r = 0.74$

for the 1 m DEM; compared to $s = 0.96$ and $r = 0.70$ for the 10 m DEM. For both the Mid Bailey Run and Russian River field sites, the sensitivity is higher than the reliability for all of the DEM resolutions tested (Table 4). However for the River Swale site, the reliability is higher than the sensitivity ($r = 0.84, s = 0.65$).

4.2 Comparison with mapped terraces

We also compare the features extracted by our method to field-mapped terraces from four field sites throughout the US: the South Fork Eel River, CA (Seidl and Dietrich, 1992); the Le Sueur River, MN (Gran et al., 2009); the Mattole River, CA (Dibblee and Minch, 2008); and the Clearwater River, WA (Wegmann and Pazzaglia, 2002). Two of these sites had 1 m LiDAR-derived DEMs (the South Fork Eel and Le Sueur Rivers). For the remaining two sites, we used 10 m DEMs ~~were created~~ derived from the USGS 1/3 arc second NED, and used by Limaye and Lamb (2016). Terraces in the South Fork Eel River and the Le Sueur River were digitised from field mapping carried out in previous studies (Seidl and Dietrich, 1992; Gran et al., 2009), constrained by the hillshaded DEMs. Terraces from the Mattole River and the Clearwater River were digitised by Limaye and Lamb (2016) from geological maps, with the terraces mapped by Dibblee and Minch (2008) for the Mattole River and Wegmann and Pazzaglia (2002) for the Clearwater River. We ran our method in the swath setting for each of these sites, so that the terraces were mapped compared to the main stem channel of interest in each site. The thresholds for terrace identification (R_c and S) were set statistically for each site using the quantile-quantile plots. In order to quantify the difference between our method and the digitised terraces, we calculated the r and s values following the same methodology as for the floodplain comparison (Table 4).

Figure 7 shows a visual comparison of the predicted and digitised terraces from the two sites with 1 m LiDAR-derived DEMs. In general there was good spatial correlation between the two terrace datasets for each field site, although in some cases the automated method did not identify all terraces at high elevations compared to the modern channel. The South Fork Eel River had the highest values of both r (0.65) and s (0.72). The comparison between the two terrace datasets for the field sites with 10 m DEMs is shown in Figure 8. These sites had lower r and s values than that of the South Fork Eel River, but were comparable to the values for the Le Sueur River (e.g. Table 4).

5 Discussion

5.1 Floodplains

The results outlined above compare our method of automatic feature extraction to various datasets of both floodplains and terraces. In order to test the ability of our method in identifying floodplains, we compared the delineated geomorphic floodplain to both field-mapped floodplain initiation points and hydrological modelling predictions. We found that our method predicts the location of the field-

mapped FIPs to within tens of metres for both field sites (Mid Bailey Run, OH; and Coweeta, NC). The ~~best agreement between the mapped and predicted floodplain points occurs at~~ reliability and sensitivity values were highest for the Coweeta field ~~site~~sites, with a ~~mean horizontal error of $\pm$ 6 m. These results suggest that our method is reliable in predicting the geomorphic floodplain as identified in the field , as~~ value of $r = 0.78$ and $s = 1$, which indicates that there were no false negatives in this field site. Table 3 shows that in many cases the error between the mapped and predicted FIPs is within the same order of magnitude as the error on the field-mapped coordinates ($\approx$ 1 m for Coweeta and $\approx$ 6 m for Mid Bailey Run). ~~Some discrepancies may also be expected due to the difference in dates between the field mapping (carried out in 2011 for~~ In isolated cases in the Mid Bailey Run ~~, and 2014 for Coweeta) and the LiDAR collection (2008/2009 for Mid Bailey Run, and 2009 for Coweeta), as the extent of floodplain inundation and alluviation may vary through time. In the Mid Bailey Run field site~~site, the error was higher between the mapped and predicted FIPs (around 90 m for two of the points), where the mapped FIP was located in narrow headwater valleys (Figure 4). Furthermore, the predicted floodplain in the majority of cases was located downstream of the mapped FIPs in Mid Bailey Run (Table 3)~~, which were contained in narrow headwater valleys (Figure 4)~~. This is not surprising, as our method is based on identifying areas of low gradient, which is calculated based on polynomial surface fitting with a specified window radius (Sect. 2.2). Small pockets of alluviation in narrow valleys may therefore be missed by the method if the width of the floodplain is less than that of the window radius or the DEM resolution.

We also validated our method against published flood maps for three of our field sites (Mid Bailey Run, OH; Russian River, CA; and River Swale, UK). The quality analysis for this comparison (Table 4 and Figure 6) suggests that there is in general a good correlation between our method and the published flood maps, with high values for ~~both~~ reliability ($r \geq 0.7$)~~and~~ , sensitivity ($s \geq 0.65$), and overall quality ($Q \geq 0.58$) for each field site. The results for both the Russian River and Mid Bailey Run showed higher sensitivity values than reliability, suggesting that the our method predicted more false positives than false negatives. In each field site, the published flood maps were classified to define the 1% annual chance of flooding, or the 100 year return period flood event. It may therefore be expected that our geomorphic-based method would delineate a larger floodplain than is flooded in a 100 year return period event. The results for the River Swale, however, show a higher reliability than sensitivity, suggesting that more false negatives were predicted than false positives. This may be due to methodological differences in the production of this flood map by the Environment Agency (UK) compared to the US sites. Figure 6f shows the published flood map for the River Swale site which, in comparison to the FEMA flood maps (Figures 6b and 6d) extends into the headwaters of the channel network. As these areas do not have low gradient surfaces next to the channel, they may not be selected by our method. This may account for the higher number of false negatives predicted at this site.

Published flood maps are useful in providing an independent estimate of likely floodplains in each field site. However, there are potential limitations to these maps which must be carefully considered, and may result in some of the differences compared to geomorphic floodplain prediction techniques. Hydrodynamic models have a large number of parameters, which require careful calibration with field and hydraulic data, such as channel roughness and discharge data from gauging stations. Furthermore, due to the time-consuming and expensive nature of these studies, flood maps are often not produced for small catchment sizes, and may therefore be incomplete on a landscape-scale (e.g. Figure 5). There may also be differences in the methodology used in producing these maps for each site, depending on the input topographic data and modelling software used. However, despite these discrepancies between the flood maps we find a good spatial correlation between these and the predictions from our method (Figure 6).

In order to test the sensitivity of our method to grid resolution, we also ran the floodplain extraction using 10 m DEMs derived from the USGS NED for two of the field sites (Russian River, CA, and Mid Bailey Run, OH), as well as testing it on the River Swale in the UK (5 m resolution DEM). We found ~~no observable~~ there was little difference in the reliability and sensitivity results when compared to the 1 m DEMs (Table 4). This suggests that our method is relatively insensitive to grid resolution, allowing the identification of floodplain features on coarser-resolution DEMs. Furthermore, in the Mid Bailey Run field site, the method performed better on the 10 m data compared to the 1 m DEM. 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.

5.2 Terraces

We also tested the ability of our method to identify fluvial terraces in four field sites (South Fork Eel River, CA; Le Sueur River, MN; Mattole River, CA; and Clearwater River, WA) by comparing to digitised terrace maps. Two of these field sites had 1 m LiDAR-derived DEMs (Figure 7) whereas two had 10 m DEMs from the USGS NED (Figure 8). The quality analysis for the 1 m DEMs showed the higher reliability and sensitivity values for the South Fork Eel River site ($r = 0.65$ and $s = 0.72$), with comparable values for the remaining three field sites. This may be due to the influence of topographic structure on terrace identification. The portion of the Eel River DEM analysed here has higher relief, with a maximum elevation of 290 m above the nearest channel, compared to the lower-relief landscape covered by the DEM for the Le Sueur River, with a maximum elevation of 40 m above the nearest channel. 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, such as steep mountainous areas. This is similar to other semi-automated terrace extraction methods (e.g. Stout and Belmont, 2014; Hopkins and Snyder, 2016). The Le Sueur River is currently incising through Pleistocene tills, forming a low-gradient surface or plateau (Fisher, 2003; Gran et al., 2009; Belmont et al., 2011a). High-altitude, low-gradient surfaces, such as relict plateaus, may result in error in the method due to the difficulty in distinguishing the distribution of terrace elevations from these low-relief surfaces. The Le Sueur River basin is also heavily influenced by human land use, which makes feature extraction challenging (Passalacqua et al., 2012). 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).

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 indices of r and Q (Figures 8c and d and Table 4). In this site, the digitised terraces are close in elevation to the modern channel, with a maximum terrace height of 13 m. Furthermore, in some cases our method did not select all of the terraces identified by the field mapping, particularly at ~~higher~~ the highest elevations compared to the modern channel (e.g. Figure 7c and d). This may be the case if the threshold for elevation compared to the channel selected by the quantile-quantile plot is lower than that of the highest terrace elevations. This can be examined for the landscape in question by a visual inspection of the quantile-quantile plots and the location of the threshold compared to the distribution of channel relief (e.g. Figure 2). ~~However, despite this limitation~~Our method fits a Gaussian distribution to the quantile-quantile plots, and selects the thresholds as the deviation of the real data from this distribution. Our method uses a Gaussian distribution as a simple general model of elevation distributions that can be applied across multiple landscapes. However, in some landscapes, the distribution of elevations may not be accurately represented by a Gaussian distribution. A future avenue for development of this method may be to include multiple models for elevation distributions from which to select the thresholds of elevation and gradient.

However, despite these limitations, the selection of the threshold from quantile-quantile plots ~~allows our method to be fully automated. It~~ means that our method does not require the input of any independent datasets or field-mapping~~, unlike previous~~. Semi-automated methods of terrace identification~~which are semi-automated (e.g. Stout and Belmont, 2014). These semi-automated~~

methods , where the terrace polygons are manually edited by the user, are particularly useful in areas where independent datasets of terrace locations are available for calibration, and may be more appropriate than our method on site-specific scales (e.g. Stout and Belmont, 2014). However, our fully automated approach the selection of thresholds based on a statistical approach means that our method can be applied in areas where these data do not exist, on a broader landscape scale, or as a rapid first-order predictor of terrace locations.

In addition to the field sites with LiDAR-derived DEMs, we also tested our method against digitised terraces from two sites with 10 m DEMs gridded from the USGS NED, to examine the performance of the method at lower grid resolution. Figure 8 shows the results of the terrace identification on the 10 m resolution data. The reliability and sensitivity of the method for these two sites (Table 4) was lower than that of the South Fork Eel River, but comparable to that of the Le Sueur River. This suggests that the method is able to successfully select terraces at lower grid resolutions. Although there are some differences between the terraces predicted by the method and those digitised in the field, the majority of the terrace features evident from a visual inspection of the hillshaded DEMs are correctly identified by the algorithm (Figure 8). In some cases, some terrace-like features that can be seen on the hillshaded DEMs are not identified in the digitised terrace maps (e.g. Figure 8b). This may be due to error in the mapping of terrace surfaces in the field, or discrepancies resulting from the digitisation process.

Fully automated identification of An objective, landscape-scale method of identifying floodplain and terrace features has numerous applications in the geomorphological and hydrological communities. For example, terrace surfaces have been used to examine the response of fluvial systems to tectonic and climatic perturbations (e.g. Merritts et al., 1994), and to investigate the relative importance of lateral and vertical channel incision (e.g. Finnegan and Dietrich, 2011). Analysis of terrace areas can be used to quantify sediment budgets and estimate storage volumes over millenial timescales (e.g. Blöthe and Korup, 2013)(e.g. Trimble, 1999; Brown et al., 2009; Belmont et al., 2011b; Blöthe and Korup, 2013). Our new method facilitates the rapid extraction of terrace surfaces either across the whole landscape or compared to a representative channel of interest. It allows the user to investigate how various metrics, such as elevation compared to the channel, slope, and curvature, vary both within and between individual terrace surfaces (e.g. Figure 7). These metrics could be used in order to examine how terrace heights vary with distance along channel profiles, for example, or to identify signatures of deformation corresponding to tectonic processes (Avouac and Peltzer, 1993; Lavé and Avouac, 2000; Pazzaglia and Brandon, 2001; Vive

5.3 Research needs: fully-automated feature extraction

A key goal for the Earth surface research community is to develop fully-automated methods of feature extraction from DEMs in order to avoid expensive and time consuming field-mapping, and to investigate the controls on geomorphic processes at a landscape scale. Our new method of floodplain

and terrace delineation attempts to meet some of these research needs, by allowing the statistical determination of the thresholds for feature extraction. However, our method still requires the input of some user-defined parameters. If the method is run across the whole landscape, the user must set a threshold stream order for the calculation of elevation compared to the nearest channel. This is necessary so that each pixel is mapped to the main channel along which floodplains or terraces have formed, rather than narrow tributary valleys. This threshold can be determined by the user based on a visual inspection of the DEM compared to the channel network. If the user runs the method based on the swath mode, the width of the swath profile must be set. This can also be done based on a visual inspection of the DEM to provide a sufficiently wide swath compared to the valleys in the landscape. Furthermore, if the method is run in the swath mode, then a minimum terrace height must be set in order to delineate between floodplains and fluvial terraces.

However, future development of new algorithms, such as extraction of valley widths, would allow these parameters to be set based on the topographic data alone. Our method represents a first step towards this goal of fully-automated geomorphic feature identification, which can be improved upon with future research. The combination of different algorithms for terrain analysis, such as hillslope flow routing, channel network extraction, floodplains, and fluvial terraces, would allow an objective landscape-scale investigation of the controls on geomorphic processes.

6 Conclusions

We have presented a novel method for the ~~automated identification~~ geomorphometric delineation of floodplain and fluvial terrace features from topographic data. Unlike previous methods, which tend to require calibration with additional datasets, our method ~~is fully objective. Our method~~ selects floodplain and terrace features using thresholds of local gradient and elevation compared to the nearest channel, which are calculated statistically from the DEM. Furthermore, the floodplain or terrace surfaces do not need to be manually edited by the user at any point during the process. Our method can ~~either be run~~ be run either across the whole landscape, or from a topographic swath profile ~~,~~ where features can be compared to a specific channel of interest.

In order to test the performance of our method we have compared it to field-mapped floodplains and terraces from eight field sites with a range of topographies and grid resolutions. We find that our method performs well when compared to field-mapped floodplain initiation points, published flood risk maps, and digitised terrace surfaces. Our method works particularly well in higher relief areas, such as the Russian and South Fork Eel Rivers (CA), where the floodplain and terrace features are constrained within valleys. It is relatively insensitive to grid resolution, allowing the successful extraction of floodplain and terrace features at resolutions of ~~1–10~~ 1 - 10 m.

Our new method has numerous applications in both the hydrological and geomorphological communities. It can allow the rapid extraction of floodplain features in areas where the data required

for detailed hydrological modelling studies are unavailable, facilitating investigation of flood response, sediment transport, and alluviation. Furthermore, the automated extraction of terrace locations, heights, and other metrics could be used to examine the response of fluvial systems to climatic and tectonic perturbations, as well as the relative importance of lateral and vertical channel incision.

7 Software availability

Our software is freely available for download on GitHub as part of the Edinburgh Land Surface Dynamics Topographic Tools package at https://github.com/LSDtopotools. Full documentation on download, installation, and using the software can be found at http://lsdtopotools.github.io/LSDTT_book/

Author contributions. FJC, SMM, DTM, and DAV wrote the software for the feature extraction. MDH, LJS, and FJC collected the field data for floodplain validation; ABL collected the field data for terrace validation. FJC performed the analyses, created the figures, and wrote the manuscript with contributions from the other authors.

Acknowledgements. FJC is funded by the Carnegie Trust for the Universities of Scotland and NERC grant NE/P012922/1. SMM is funded by NERC grant NE/P015905/1 and U.S. Army Research Office contract number W911NF-13-1-0478. DTM is funded by NERC grant NE/K01627X/1 and DAV is funded by NERC grant NE/L501591/1. LJS was supported by a NERC PhD studentship. ABL acknowledges support from the National Center for Earth-Surface Dynamics 2 Synthesis Postdoctoral Program. We are also grateful for additional financial support from the British Society for Geomorphology and the Royal Geographical Society with IBG. We would like to thank Stuart Grieve and Elizabeth Dingle for their help with fieldwork.

References

Avouac, J. P. and Peltzer, G.: Active tectonics in southern Xinjiang, China: Analysis of terrace riser and normal fault scarp degradation along the Hotan-Qira Fault System, Journal of Geophysical Research: Solid Earth, 98, 21 773–21 807, doi:10.1029/93JB02172, http://onlinelibrary.wiley.com/doi/10.1029/93JB02172/abstract, 1993.

Belmont, P.: Floodplain width adjustments in response to rapid base level fall and knickpoint migration, Geomorphology, 128, 92–102, doi:10.1016/j.geomorph.2010.12.026, http://www.sciencedirect.com/science/article/pii/S0169555X10005702, 2011.

Belmont, P., Gran, K., Jennings, C. E., Wittkop, C., and Day, S. S.: Holocene landscape evolution and erosional processes in the Le Sueur River, central Minnesota, Field Guides, 24, 439–455, doi:10.1130/2011.0024(21), http://fieldguides.gsapubs.org/content/24/439, 2011a.

Belmont, P., Gran, K. B., Schottler, S. P., Wilcock, P. R., Day, S. S., Jennings, C., Lauer, J. W., Viparelli, E., Willenbring, J. K., Engstrom, D. R., and Parker, G.: Large Shift in Source of Fine Sediment in the Upper Mississippi River, Environmental Science & Technology, 45, 8804–8810, doi:10.1021/es2019109, http://dx.doi.org/10.1021/es2019109, 2011b.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Advances in Water Resources, 16, 41–51, doi:10.1016/0309-1708(93)90028-E, http://www.sciencedirect.com/science/article/pii/030917089390028E, 1993.

Beven, K.: TOPMODEL: A critique, Hydrological Processes, 11, 1069–1085, doi:10.1002/(SICI)1099-1085(199707)11:9<1069::AID-HYP545>3.0.CO;2-O, http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-1085(199707)11:9<1069::AID-HYP545>3.0.CO;2-O/abstract, 1997.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43–69, doi:10.1080/02626667909491834, http://dx.doi.org/10.1080/02626667909491834, 1979.

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J.: TOPMODEL, in: Computer Models of Watershed Hydrology, edited by Singh, V. P., pp. 627–668, Water Resource Publications, Colorado, 1995.

Blöthe, J. H. and Korup, O.: Millennial lag times in the Himalayan sediment routing system, Earth and Planetary Science Letters, 382, 38–46, doi:10.1016/j.epsl.2013.08.044, http://www.sciencedirect.com/science/article/pii/S0012821X13004822, 2013.

Booij, M. J.: Impact of climate change on river flooding assessed with different spatial model resolutions, Journal of Hydrology, 303, 176–198, doi:10.1016/j.jhydrol.2004.07.013, http://www.sciencedirect.com/science/article/pii/S002216940400397X, 2005.

Braun, J. and Willett, S. D.: A very efficient O(n), implicit and parallel method to solve the stream power equation governing fluvial incision and landscape evolution, Geomorphology, 180–181, 170–179, doi:10.1016/j.geomorph.2012.10.008, http://www.sciencedirect.com/science/article/pii/S0169555X12004618, 2013.

Brown, A. G., Carey, C., Erkens, G., Fuchs, M., Hoffmann, T., Macaire, J.-J., Moldenhauer, K.-M., and Walling, D. E.: From sedimentary records to sediment budgets: Multiple approaches to catchment sediment flux,

Geomorphology, 108, 35–47, doi:10.1016/j.geomorph.2008.01.021, http://www.sciencedirect.com/science/article/pii/S0169555X09000336, 2009.

Bull, W. B.: Geomorphic responses to climatic change, Oxford University Press, New York, https://www.osti.gov/scitech/biblio/5603696, 1991.

Clubb, F. J., Mudd, S. M., Milodowski, D. T., Hurst, M. D., and Slater, L. J.: Objective extraction of channel heads from high-resolution topographic data, Water Resources Research, 50, 4283–4304, doi:10.1002/2013WR015167, http://onlinelibrary.wiley.com/doi/10.1002/2013WR015167/abstract, 2014.

Cobby, D. M., Mason, D. C., Horritt, M. S., and Bates, P. D.: Two-dimensional hydraulic flood modelling using a finite-element mesh decomposed according to vegetation and topographic features derived from airborne scanning laser altimetry, Hydrological Processes, 17, 1979–2000, doi:10.1002/hyp.1201, http://onlinelibrary.wiley.com/doi/10.1002/hyp.1201/abstract, 2003.

Committee on FEMA Flood Maps: Mapping the Zone: Improving Flood Map Accuracy, The National Academies Press, 500 Fifth Street, N.W. Washington, DC 20001, http://www.nap.edu/catalog/12573/mapping-the-zone-improving-flood-map-accuracy, 2009.

Degiorgis, M., Gnecco, G., Gorni, S., Roth, G., Sanguineti, M., and Taramasso, A. C.: Classifiers for the detection of flood-prone areas using remote sensed elevation data, Journal of Hydrology, 470–471, 302–315, doi:10.1016/j.jhydrol.2012.09.006, http://www.sciencedirect.com/science/article/pii/S002216941200755X, 2012.

Demoulin, A., Bovy, B., Rixhon, G., and Cornet, Y.: An automated method to extract fluvial terraces from digital elevation models: The Vesdre valley, a case study in eastern Belgium, Geomorphology, 91, 51–64, doi:10.1016/j.geomorph.2007.01.020, http://www.sciencedirect.com/science/article/pii/S0169555X07000487, 2007.

Dibblee, T. W. and Minch, J. A.: Geologic Map of the Point Delgada and Garberville 15 minute Quadrangles, Humboldt and Mendocino Counties, California, 2008.

Dingle, E. H., Sinclair, H. D., Attal, M., Milodowski, D. T., and Singh, V.: Subsidence control on river morphology and grain size in the Ganga Plain, American Journal of Science, 316, 778–812, doi:10.2475/08.2016.03, http://www.ajsonline.org/content/316/8/778, 2016.

Dodov, B. and Foufoula-Georgiou, E.: Floodplain morphometry extraction from a high-resolution digital elevation model: a simple algorithm for regional analysis studies, IEEE Geoscience and Remote Sensing Letters, 3, 410–413, doi:10.1109/LGRS.2006.874161, 2006.

Fawcett, T.: An introduction to ROC analysis, Pattern Recognition Letters, 27, 861–874, doi:10.1016/j.patrec.2005.10.010, http://www.sciencedirect.com/science/article/pii/S016786550500303X, 2006.

Finnegan, N. J. and Dietrich, W. E.: Episodic bedrock strath terrace formation due to meander migration and cutoff, Geology, 39, 143–146, doi:10.1130/G31716.1, http://geology.gsapubs.org/content/39/2/143, 2011.

Fisher, T. G.: Chronology of glacial Lake Agassiz meltwater routed to the Gulf of Mexico, Quaternary Research, 59, 271–276, doi:10.1016/S0033-5894(03)00011-5, http://www.sciencedirect.com/science/article/pii/S0033589403000115, 2003.

Gran, K. B., Belmont, P., Day, S. S., Jennings, C., Johnson, A., Perg, L., and Wilcock, P. R.: Geomorphic evolution of the Le Sueur River, Minnesota, USA, and implications for current sediment loading, Geological

Society of America Special Papers, 451, 119–130, doi:10.1130/2009.2451(08), http://specialpapers.gsapubs.org/content/451/119, 2009.

Gran, K. B., Finnegan, N., Johnson, A. L., Belmont, P., Wittkop, C., and Rittenour, T.: Landscape evolution, valley excavation, and terrace development following abrupt postglacial base-level fall, Geological Society of America Bulletin, 125, 1851–1864, doi:10.1130/B30772.1, http://gsabulletin.gsapubs.org/content/125/11-12/1851, 2013.

Grieve, S. W., Mudd, S. M., and Hurst, M. D.: How long is a hillslope?, Earth Surface Processes and Landforms, 41, 1039–1054, doi:10.1002/esp.3884, http://onlinelibrary.wiley.com/doi/10.1002/esp.3884/abstract, 2016a.

Grieve, S. W. D., Mudd, S. M., Hurst, M. D., and Milodowski, D. T.: A nondimensional framework for exploring the relief structure of landscapes, Earth Surface Dynamics, 4, 309–325, doi:10.5194/esurf-4-309-2016, http://www.earth-surf-dynam.net/4/309/2016/, 2016b.

Grieve, S. W. D., Mudd, S. M., Milodowski, D. T., Clubb, F. J., and Furbish, D. J.: How does grid-resolution modulate the topographic expression of geomorphic processes?, Earth Surface Dynamics, 4, 627–653, doi:10.5194/esurf-4-627-2016, http://www.earth-surf-dynam.net/4/627/2016/, 2016c.

Grimaldi, S., Petroselli, A., Arcangeletti, E., and Nardi, F.: Flood mapping in ungauged basins using fully continuous hydrologic–hydraulic modeling, Journal of Hydrology, 487, 39–47, doi:10.1016/j.jhydrol.2013.02.023, http://www.sciencedirect.com/science/article/pii/S0022169413001352, 2013.

Guzzetti, F., Stark, C. P., and Salvati, P.: Evaluation of Flood and Landslide Risk to the Population of Italy, Environmental Management, 36, 15–36, doi:10.1007/s00267-003-0257-1, http://link.springer.com/article/10.1007/s00267-003-0257-1, 2005.

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L. V., Bronnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P. M.: Observations: Atmosphere and Surface, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Stocker, T. F., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

He, L., Chao, Y., and Suzuki, K.: A Run-Based Two-Scan Labeling Algorithm, IEEE Transactions on Image Processing, 17, 749–756, doi:10.1109/TIP.2008.919369, 2008.

Heipke, C., Mayer, H., Wiedemann, C., and Jamet, O.: Automated reconstruction of topographic objects from aerial images using vectorized map information, International Archives of Photogrammetry and Remote Sensing, 23, 47–56, 1997.

Hergarten, S., Robl, J., and Stüwe, K.: Extracting topographic swath profiles across curved geomorphic features, Earth Surface Dynamics, 2, 97–104, doi:10.5194/esurf-2-97-2014, http://www.earth-surf-dynam.net/2/97/2014/, 2014.

Hopkins, A. J. and Snyder, N. P.: Performance evaluation of three DEM-based fluvial terrace mapping methods, Earth Surface Processes and Landforms, 41, 1144–1152, doi:10.1002/esp.3922, http://onlinelibrary.wiley.com/doi/10.1002/esp.3922/abstract, 2016.

Horritt, M. S. and Bates, P. D.: Evaluation of 1D and 2D numerical models for predicting river flood inundation, Journal of Hydrology, 268, 87–99, doi:10.1016/S0022-1694(02)00121-X, http://www.sciencedirect.com/science/article/pii/S002216940200121X, 2002.

Hunter, N. M., Bates, P. D., Horritt, M. S., and Wilson, M. D.: Simple spatially-distributed models for predicting flood inundation: A review, Geomorphology, 90, 208–225, doi:10.1016/j.geomorph.2006.10.021, http://www.sciencedirect.com/science/article/pii/S0169555X07001304, 2007.

Hurst, M. D., Mudd, S. M., Walcott, R., Attal, M., and Yoo, K.: Using hilltop curvature to derive the spatial distribution of erosion rates, Journal of Geophysical Research: Earth Surface, 117, doi:10.1029/2011JF002057, http://onlinelibrary.wiley.com/doi/10.1029/2011JF002057/abstract, 2012.

Jain, V., Fryirs, K., and Brierley, G.: Where do floodplains begin? The role of total stream power and longitudinal profile form on floodplain initiation processes, Geological Society of America Bulletin, 120, 127–141, doi:10.1130/B26092.1, http://gsabulletin.gsapubs.org/content/120/1-2/127, 2008.

Kim, J., Warnock, A., Ivanov, V. Y., and Katopodes, N. D.: Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow, Advances in Water Resources, 37, 104–126, doi:10.1016/j.advwatres.2011.11.009, http://www.sciencedirect.com/science/article/pii/S0309170811002211, 2012.

Kirkby, M. J.: Hydrograph modelling strategies, in: Process in Physical and Human Geography, edited by Peel, R., Chisholm, M., and Haggett, P., pp. 69–90, Heinemann, London, 1975.

Lashermes, B., Foufoula-Georgiou, E., and Dietrich, W. E.: Channel network extraction from high resolution topography using wavelets, Geophysical Research Letters, 34, L23S04, doi:10.1029/2007GL031140, http://onlinelibrary.wiley.com/doi/10.1029/2007GL031140/abstract, 2007.

Lavé, J. and Avouac, J. P.: Active folding of fluvial terraces across the Siwaliks Hills, Himalayas of central Nepal, Journal of Geophysical Research: Solid Earth, 105, 5735–5770, doi:10.1029/1999JB900292, http://onlinelibrary.wiley.com/doi/10.1029/1999JB900292/abstract, 2000.

Limaye, A. B. S. and Lamb, M. P.: Numerical model predictions of autogenic fluvial terraces and comparison to climate change expectations, Journal of Geophysical Research: Earth Surface, 121, 2014JF003 392, doi:10.1002/2014JF003392, http://onlinelibrary.wiley.com/doi/10.1002/2014JF003392/abstract, 2016.

Liu, X.: Airborne LiDAR for DEM generation: some critical issues, Progress in Physical Geography, 32, 31–49, doi:10.1177/0309133308089496, http://journals.sagepub.com/doi/abs/10.1177/0309133308089496, 2008.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resources Research, 43, W07 401, doi:10.1029/2006WR005756, http://onlinelibrary.wiley.com/doi/10.1029/2006WR005756/abstract, 2007.

Manfreda, S., Di Leo, M., and Sole, A.: Detection of Flood-Prone Areas Using Digital Elevation Models, Journal of Hydrologic Engineering, 16, 781–790, doi:10.1061/(ASCE)HE.1943-5584.0000367, http://ascelibrary.org/doi/10.1061/(ASCE)HE.1943-5584.0000367, 2011.

Manfreda, S., Nardi, F., Samela, C., Grimaldi, S., Taramasso, A. C., Roth, G., and Sole, A.: Investigation on the use of geomorphic approaches for the delineation of flood prone areas, Journal of Hydrology, 517, 863–876, doi:10.1016/j.jhydrol.2014.06.009, http://www.sciencedirect.com/science/article/pii/S0022169414004697, 2014.

Marshall, J. A. and Roering, J. J.: Diagenetic variation in the Oregon Coast Range: Implications for rock strength, soil production, hillslope form, and landscape evolution, Journal of Geophysical Research: Earth Surface, 119, 2013JF003 004, doi:10.1002/2013JF003004, http://onlinelibrary.wiley.com/doi/10.1002/2013JF003004/abstract, 2014.

Meng, X., Currit, N., and Zhao, K.: Ground Filtering Algorithms for Airborne LiDAR Data: A Review of Critical Issues, Remote Sensing, 2, 833–860, doi:10.3390/rs2030833, http://www.mdpi.com/2072-4292/2/3/833, 2010.

Merritts, D. J., Vincent, K. R., and Wohl, E. E.: Long river profiles, tectonism, and eustasy: A guide to interpreting fluvial terraces, Journal of Geophysical Research: Solid Earth, 99, 14 031–14 050, doi:10.1029/94JB00857, http://onlinelibrary.wiley.com/doi/10.1029/94JB00857/abstract, 1994.

Molloy, I. and Stepinski, T. F.: Automatic mapping of valley networks on Mars, Computers & Geosciences, 33, 728–738, doi:10.1016/j.cageo.2006.09.009, http://www.sciencedirect.com/science/article/pii/S0098300406002159, 2007.

Noman, N. S., Nelson, E. J., and Zundel, A. K.: Review of Automated Floodplain Delineation from Digital Terrain Models, Journal of Water Resources Planning and Management, 127, 394–402, doi:10.1061/(ASCE)0733-9496(2001)127:6(394), http://dx.doi.org/10.1061/(ASCE)0733-9496(2001)127:6(394), 2001.

O'Callaghan, J. F. and Mark, D. M.: The extraction of drainage networks from digital elevation data, Computer Vision, Graphics, and Image Processing, 28, 323–344, doi:10.1016/S0734-189X(84)80011-0, http://www.sciencedirect.com/science/article/pii/S0734189X84800110, 1984.

Orlandini, S., Tarolli, P., Moretti, G., and Dalla Fontana, G.: On the prediction of channel heads in a complex alpine terrain using gridded elevation data, Water Resources Research, 47, doi:10.1029/2010WR009648, http://onlinelibrary.wiley.com/doi/10.1029/2010WR009648/abstract, 2011.

Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., and Dietrich, W. E.: A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths, Journal of Geophysical Research: Earth Surface, 115, doi:10.1029/2009JF001254, http://onlinelibrary.wiley.com/doi/10.1029/2009JF001254/abstract, 2010a.

Passalacqua, P., Tarolli, P., and Foufoula-Georgiou, E.: Testing space-scale methodologies for automatic geomorphic feature extraction from lidar in a complex mountainous landscape, Water Resources Research, 46, W11 535, doi:10.1029/2009WR008812, http://onlinelibrary.wiley.com/doi/10.1029/2009WR008812/abstract, 2010b.

Passalacqua, P., Belmont, P., and Foufoula-Georgiou, E.: Automatic geomorphic feature extraction from lidar in flat and engineered landscapes, Water Resources Research, 48, W03 528, doi:10.1029/2011WR010958, http://onlinelibrary.wiley.com/doi/10.1029/2011WR010958/abstract, 2012.

Pazzaglia, F. J.: 9.22 Fluvial Terraces, in: Treatise on Geomorphology, edited by Shroder, J. F., pp. 379–412, Academic Press, San Diego, http://www.sciencedirect.com/science/article/pii/B9780123747396002487, dOI: 10.1016/B978-0-12-374739-6.00248-7, 2013.

Pazzaglia, F. J. and Brandon, M. T.: A Fluvial Record of Long-term Steady-state Uplift and Erosion Across the Cascadia Forearc High, Western Washington State, American Journal of Science, 301, 385–431, doi:10.2475/ajs.301.4-5.385, http://www.ajsonline.org/content/301/4-5/385, 2001.

Pazzaglia, F. J., Gardner, T. W., and Merritts, D. J.: Bedrock Fluvial Incision and Longitudinal Profile Development Over Geologic Time Scales Determined by Fluvial Terraces, in: Rivers Over Rock: Fluvial Processes in Bedrock Channels, edited by Tinkler, K. J. and Wohl, E. E., pp. 207–235, American Geophysical Union, http://onlinelibrary.wiley.com/doi/10.1029/GM107p0207/summary, dOI: 10.1029/GM107p0207, 1998.

Pelletier, J. D.: A robust, two-parameter method for the extraction of drainage networks from high-resolution digital elevation models (DEMs): Evaluation using synthetic and real-world DEMs, Water Resources Research, 49, 75–89, doi:10.1029/2012WR012452, http://onlinelibrary.wiley.com/doi/10.1029/2012WR012452/abstract, 2013.

Perona, P. and Malik, J.: Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, 629–639, doi:10.1109/34.56205, 1990.

Quinn, P. F., Beven, K. J., and Lamb, R.: The ln(a/tan/β) index: How to calculate it and how to use it within the topmodel framework, Hydrological Processes, 9, 161–182, doi:10.1002/hyp.3360090204, http://onlinelibrary.wiley.com/doi/10.1002/hyp.3360090204/abstract, 1995.

Roering, J. J., Marshall, J., Booth, A. M., Mort, M., and Jin, Q.: Evidence for biotic controls on topography and soil production, Earth and Planetary Science Letters, 298, 183–190, doi:10.1016/j.epsl.2010.07.040, http://www.sciencedirect.com/science/article/pii/S0012821X10004784, 2010.

Schreider, S. Y., Smith, D. I., and Jakeman, A. J.: Climate Change Impacts on Urban Flooding, Climatic Change, 47, 91–115, doi:10.1023/A:1005621523177, http://link.springer.com/article/10.1023/A%3A1005621523177, 2000.

Seidl, M. and Dietrich, W. E.: The problem of channel erosion into bedrock, Catena Supplement, 23, 101–124, 1992.

Stout, J. C. and Belmont, P.: TerEx Toolbox for semi-automated selection of fluvial terrace and floodplain features from lidar, Earth Surface Processes and Landforms, 39, 569–580, doi:10.1002/esp.3464, http://onlinelibrary.wiley.com/doi/10.1002/esp.3464/abstract, 2014.

Tarolli, P. and Dalla Fontana, G.: Hillslope-to-valley transition morphology: New opportunities from high resolution DTMs, Geomorphology, 113, 47–56, doi:10.1016/j.geomorph.2009.02.006, http://www.sciencedirect.com/science/article/pii/S0169555X09000646, 2009.

Tarolli, P., Sofia, G., and Fontana, G. D.: Geomorphic features extraction from high-resolution topography: landslide crowns and bank erosion, Natural Hazards, 61, 65–83, doi:10.1007/s11069-010-9695-2, http://link.springer.com/article/10.1007/s11069-010-9695-2, 2010.

Thommeret, N., Bailly, J. S., and Puech, C.: Extraction of thalweg networks from DTMs: application to badlands, Hydrol. Earth Syst. Sci., 14, 1527–1536, doi:10.5194/hess-14-1527-2010, http://www.hydrol-earth-syst-sci.net/14/1527/2010/, 2010.

Trimble, S. W.: Decreased Rates of Alluvial Sediment Storage in the Coon Creek Basin, Wisconsin, 1975-93, Science, 285, 1244–1246, doi:10.1126/science.285.5431.1244, http://science.sciencemag.org/content/285/5431/1244, 1999.

Viveen, W., Schoorl, J. M., Veldkamp, A., and van Balen, R. T.: Modelling the impact of regional uplift and local tectonics on fluvial terrace preservation, Geomorphology, 210, 119–135, doi:10.1016/j.geomorph.2013.12.026, http://www.sciencedirect.com/science/article/pii/S0169555X1300634X, 2014.

Walter, R., Merritts, D. J., and Rahnis, M.: Estimating volume, nutrient conent, and rates of stream bank erosion of legacy sediment in the Piedmont and Valley and Ridge physiographic provinces, Southeastern and Central PA, Tech. rep., Pennsylvania Department of Environmental Protection, Harrisburg, PA, 2007.

Wegmann, K. W. and Pazzaglia, F. J.: Holocene strath terraces, climate change, and active tectonics: The Clearwater River basin, Olympic Peninsula, Washington State, Geological Society of America Bulletin, 114, 731–744, doi:10.1130/0016-7606(2002)114<0731:HSTCCA>2.0.CO;2, http://gsabulletin.gsapubs.org/content/114/6/731, 2002.

Wiener, N.: Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications, Technology Press of the Massachusetts Institute of Technology, 1949.

Wood, J.: The geomorphological characterisation of Digital Elevation Models., PhD thesis, University of Leicester, https://lra.le.ac.uk/handle/2381/34503, 1996.

Yang, J., Townsend, R. D., and Daneshfar, B.: Applying the HEC-RAS model and GIS techniques in river network floodplain delineation, Canadian Journal of Civil Engineering, 33, 19–28, doi:10.1139/l05-102, http://www.nrcresearchpress.com/doi/abs/10.1139/l05-102, 2006.

Zhang, T. Y. and Suen, C. Y.: A Fast Parallel Algorithm for Thinning Digital Patterns, Commun. ACM, 27, 236–239, doi:10.1145/357994.358023, http://doi.acm.org/10.1145/357994.358023, 1984.

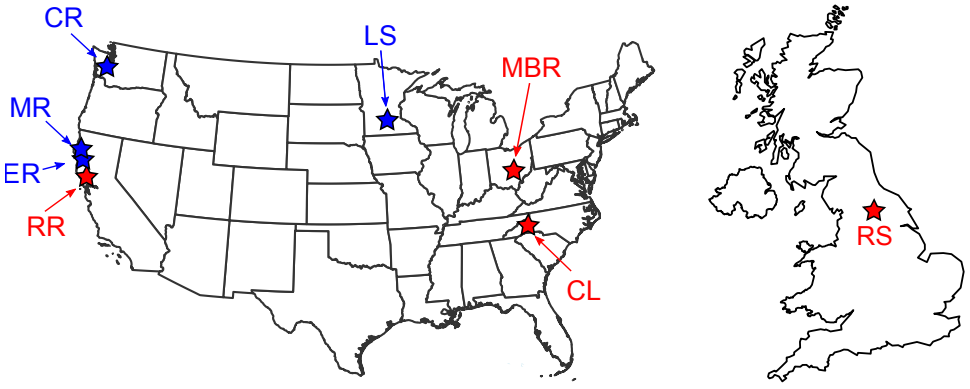


Figure 1. Maps of the US and UK showing the location of the eight field sites in the study. Red stars represent floodplain sites; blue stars represent terrace sites. RR = Russian River, CA; ER = South Fork Eel River, CA; MR = Mattole River, CA; CR = Clearwater River, WA; LS = Le Sueur River, MN; MBR = Mid Bailey Run, OH; CL = Coweeta Hydrologic Laboratory, NC; RS = River Swale, Yorkshire, UK.

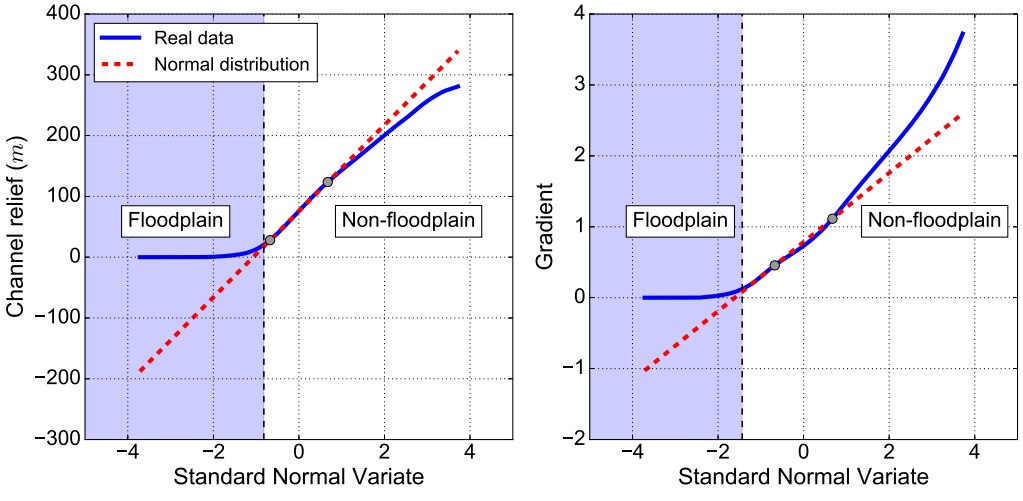


Figure 2. Example quantile-quantile plots for Mid Bailey Run, Ohio, showing probability density function of relief relative to the channel and slope. The probability density function of each is shown in blue, with the reference normal distribution shown by the red dashed line. The threshold (black dashed line) is selected where there is less than 1% difference between the real and reference distributions. The blue box highlights the portion of the distribution identified as floodplain.

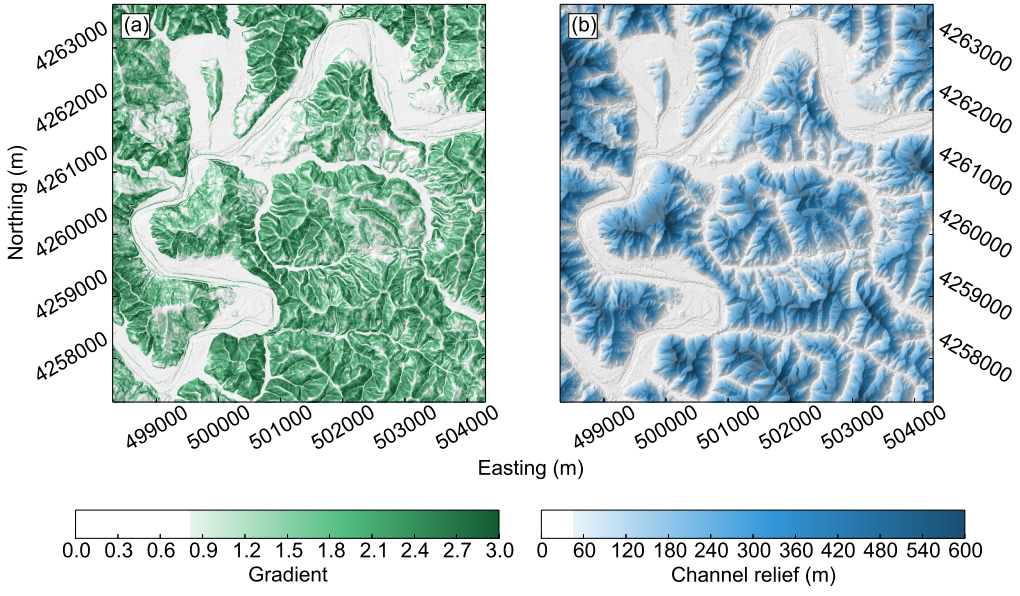


Figure 3. Maps showing a) gradient and b) relief relative to the nearest channel, R_c, for the Russian River field site. The areas of the landscape identified as below the threshold are shown in white, with values above the threshold then grading to darker colours. In order to be selected as floodplain, each pixel must be below the threshold for both gradient and R_c. The coordinate system is UTM Zone 10N.

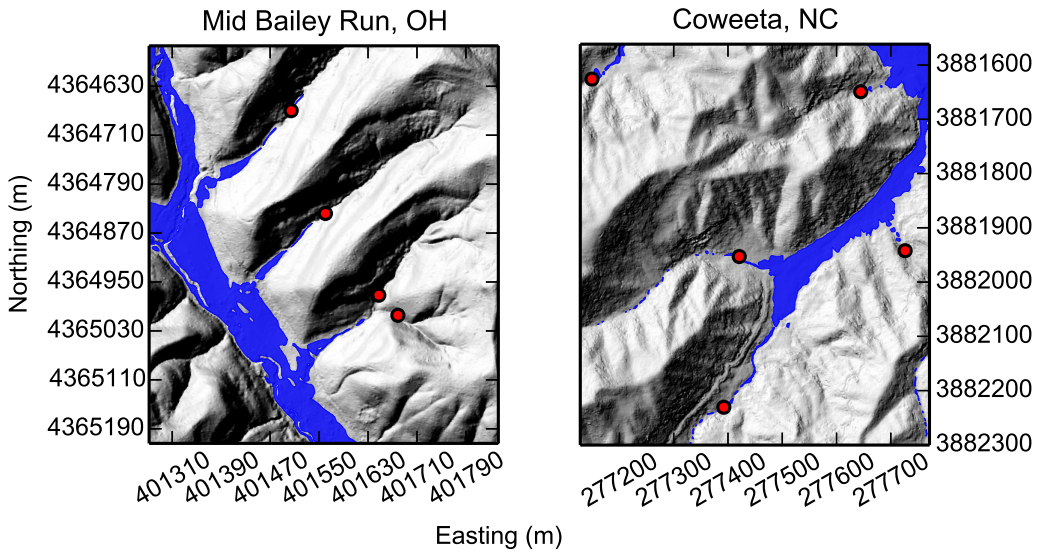


Figure 4. Shaded relief maps of Mid Bailey Run and Coweeta field sites showing the relationship between the predicted floodplain (blue) and the mapped floodplain initiation points (red). The UTM zone is 17N.

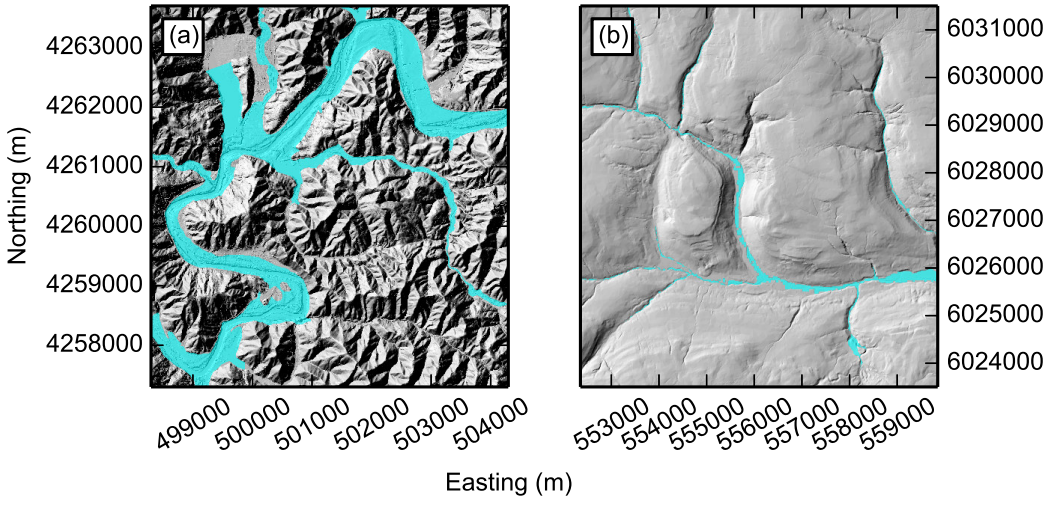


Figure 5. Shaded relief maps showing a) FEMA flood risk map for the Russian River, CA, UTM Zone 10N and b) EA flood risk map for the River Swale, UK, UTM Zone 30N. In some parts of the landscape the published flood maps do not extend all the way up the catchments.

~~Shaded relief maps of Mid Bailey Run and Coweeta field sites showing the relationship between the predicted floodplain (blue) and the mapped floodplain initiation points (red). The UTM zone is 17N.~~

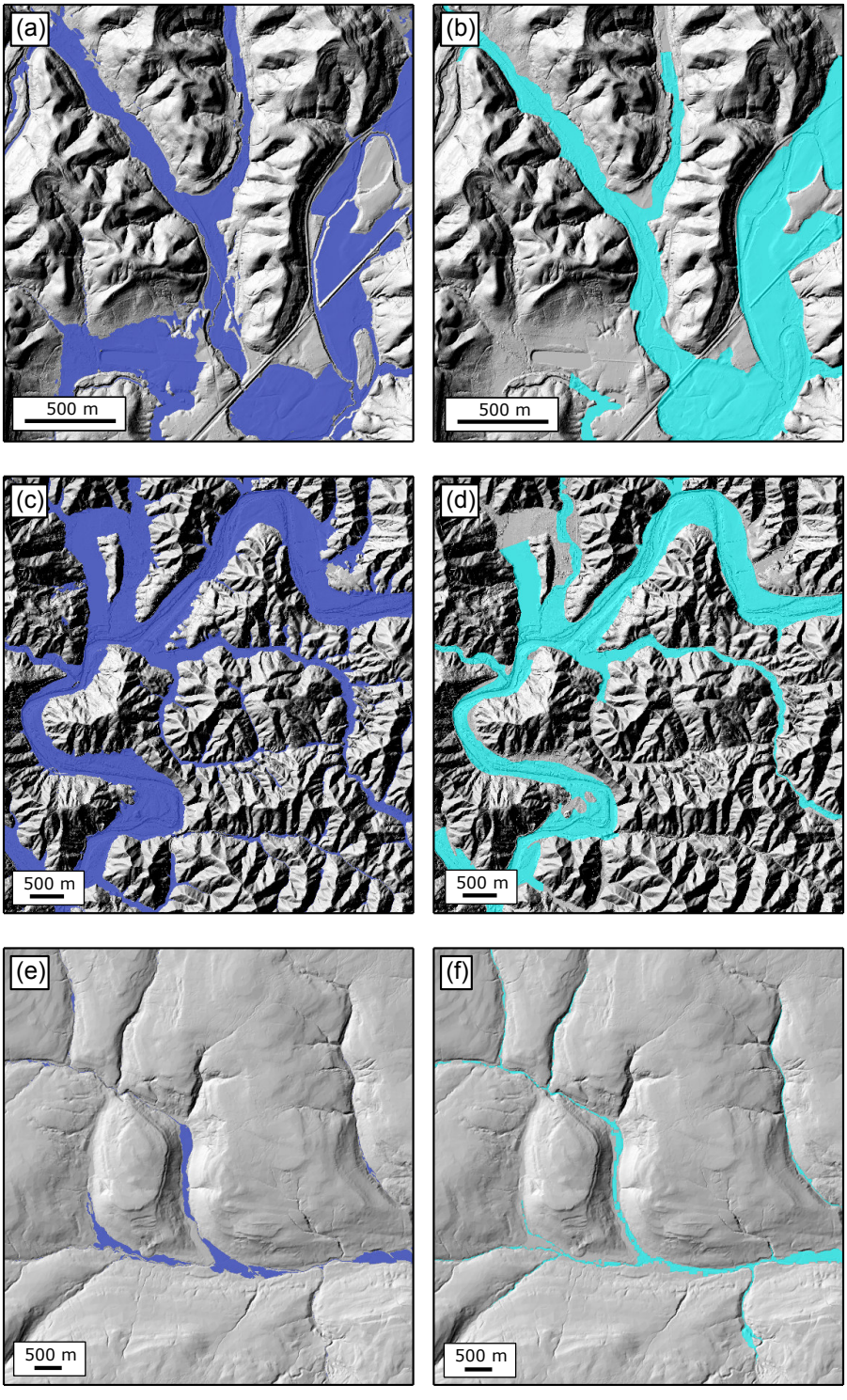


Figure 6. Shaded relief maps for each field site showing a comparison between the predicted floodplains (first column) and the published FEMA/EA maps (second column). (a) - (b) Mid Bailey Run, OH. (c) - (d) Russian River, CA. (e) - (f) River Swale, UK.

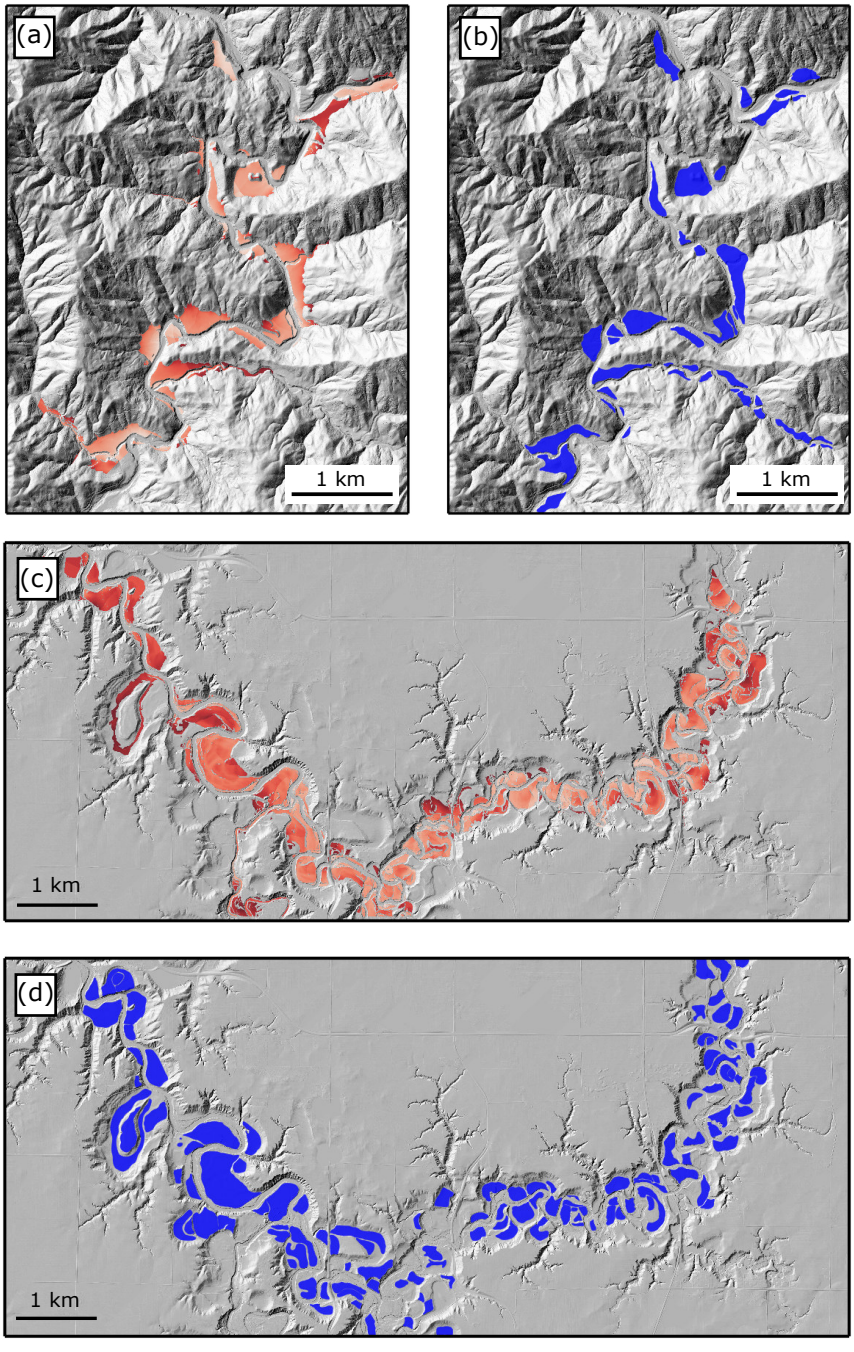


Figure 7. Shaded relief maps for the two field sites with LiDAR-derived DEMs showing a comparison between the predicted terraces (red) and the digitised terraces (blue). The predicted terraces are coloured by elevation compared to the channel, where darker red indicates higher elevation. (a) - (b) South Fork Eel River, CA. Maximum terrace height is 43 m. (c) - (d) Le Sueur River, MN. Maximum terrace height is 9.5 m.

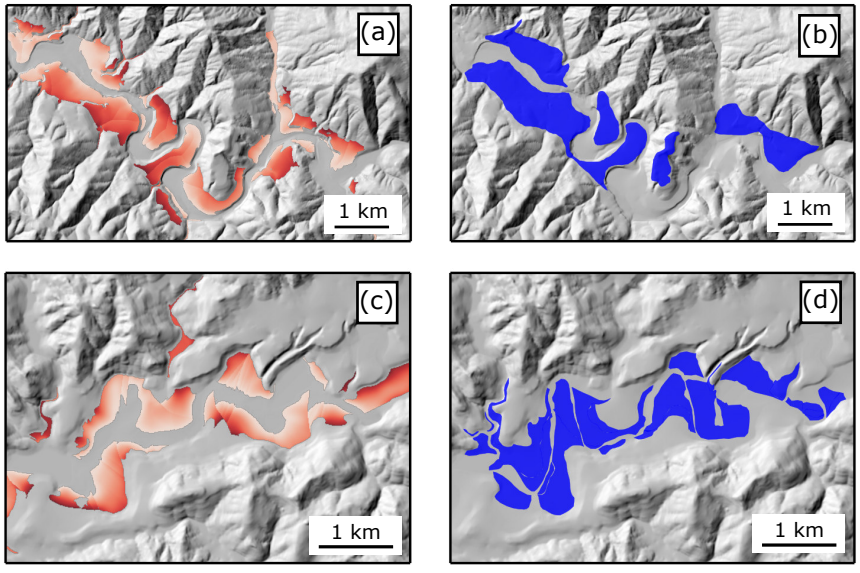


Figure 8. Shaded relief maps for the two field sites with 10 m resolution DEMs from the USGS NED showing a comparison between the predicted terraces (red) and the digitised terraces (blue). The predicted terraces are coloured by elevation compared to the channel, where darker red indicates higher elevation. (a) - (b) Mattole River, CA. Maximum terrace height is 50 m. (c) - (d) Clearwater River, WA. Maximum terrace height is 13 m.

Table 1. Channel relief and slope threshold for each field site

Field site	Channel relief threshold	Slope threshold
Mid Bailey Run, OH	23.69	0.15
Coweeta, NC	32.80	0.11
Russian River, CA	43.51	0.81
River Swale, UK	39.40	0.05
South Fork Eel River, CA	42.96	0.05
Le Sueur River, MN	9.42	0.05
Mattole River, CA	50.25	0.17
Clearwater River, WA	12.67	0.06

Table 2. Details of climate and lithology for each field site

Field site	UTM Zone	MAP (mm)	MAT(°C)	Lithology	Comparison datasets	Grid res. (m)
Russian River, CA	10°N	1396	14.1	Sandstones and shales, Quaternary alluvial deposits	FEMA flood hazard maps	1
Mid Bailey Run, OH	17°N	1005	10.9	Sandstones, siltstones, shales	FEMA flood hazard maps Field-mapped FIPs	1
Coweeta, NC	17°N	1792	12.3	Meta-sedimentary units	FEMA flood hazard maps Field-mapped FIPs	1
River Swale, UK	30 °N	898	8.4	Limestones and sandstones	EA flood hazard maps	5
South Fork Eel River, CA	10°N	2009	12.7	Greywackes and shales	Digitised terraces (Seidl and Dietrich, 1992)	1
Le Sueur River, MN	15°N	793	7.5	Pleistocene tills and Ordovician ~~dolomites~~ dolostones	Digitised terraces (Gran et al., 2009)	1
Mattole River, CA	10°N	2593	12.8	Sandstones and shales, Quaternary alluvial deposits	Digitised terraces (Dibblee and Minch, 2008; Limaye and Lamb, 2016)	10
Clearwater River, WA	10°N	3126	9.9	~~some rocks~~ Sandstones with interbedded shales	Digitised terraces (Wegmann and Pazzaglia, 2002; Limaye and Lamb, 2016)	10

Table 3. Flow distances between the field-mapped FIPs and predicted floodplain extents

Field site	Mapped FIP	Easting (m)	Northing (m)	Flow distance[1]
Mid Bailey Run, OH	T2FPI1	401513	4364940	59
	T3FPI1	401622	4364773	85
	T3FPI2	401661	4364732	-49
	WBT1FPI	400090	4363977	-23
	WBT2FPI1	399865	4364215	-1
	T4FPI	401342	4365472	28
	T5FPI2	401072	4365675	0
	T7FPI2	400670	4366152	2
	T5FPI1	401208	4365807	0
	T1FPI1	401443	4365150	0
	TX3D3-FPI0	400718	4366277	-42
	TX3FPI1	400644	4366126	-5
	MBFPI	400449	4366130	-34
	T7FPI1	400600	4366074	-19
	T4FPI2	401391	4365514	92
	T6FPI1	400900	4365921	-20
Coweeta, NC	SF5	277212.380	3882554.000	-51
	BC1	276326.800	3880661.200	-3
	HCW	277641.5	3881694.2	2
	BC3	277584.633	3881138.653	-3
	HW1	278252.652	3881715.719	13
	CB1	278089.041	3882301.638	12
	HB1	277444.900	3882919.685	-16
	CC2	277098.745	3882348.108	-2

[1] The distance between the mapped FIP and the upstream extent of the nearest floodplain patch predicted by our geomorphic method

Table 4. Results of the reliability (r)~~and~~, sensitivity (s), and overall quality (Q) analysis for each site

Field site	Grid resolution (m)	r	s	Q
Mid Bailey Run, OH	1	0.73	0.76	0.59
	10	0.77	0.80	0.65
Russian River, CA	1	0.74	0.97	0.67
	10	0.70	0.96	0.68
River Swale, UK	5	0.84	0.65	0.58
South Fork Eel River, CA	1	0.65	0.72	0.52
Le Sueur River, MN	1	0.58	0.54	0.39
Mattole River, CA	10	0.58	0.65	0.44
Clearwater River, WA	10	0.56	0.55	0.39