Are historical stage records useful to decrease the uncertainty of flood frequency analysis? A 200-year long case study

Mathieu LUCAS, Benjamin RENARD, Jerome LE COZ, Michel LANG, Antoine BARD, and Gilles PIERREFEU

Peer Review for Journal of Hydrology - Manuscript ID HYDROL50412

Response of the authors

We are thankful to the Associate Editor and the reviewers for their appreciation and understanding of our work. Please find below in black the comments of the two referees and in blue our answers. Line numbers corresponds to the document “HYDROL50412_RevisedManuscript.pdf”. You can also find a document with tracked changes (in red) into the new submission (“HYDROL50412_TrackedChangesManuscript.pdf”)

Reviewer #1 :

John England, Risk Management Center (RMC)
john.f.england@usace.army.mil

Overall Summary Review Comments

This is an important and excellent paper that provides a case study illustrating the propagation of uncertainty from river stages to discharges, and to the flood frequency curve. The use of multi-period rating curves and illustrations of three sources of uncertainty for this unusually long stage record at Beaucaire (combined with Beaucaire Restitution) is appropriate. Demonstration of the total and streamflow uncertainty propagation in flood frequency estimation is novel. The work is appropriate for the Journal of Hydrology. I recommend the manuscript be published after minor revision. Some very minor (mostly editorial and comment) work is needed to improve the quality of the presentation (clarify some concepts/limitations, slight improvements to figures) as noted below.

J. Hydrol Review Notes

Attributes of Contribution
A. What is the major contribution of this work to the field of hydrology?
Clear framework for uncertainty propagation from stages to discharges to flood frequency. Demonstration via a case study with a very long stage record.
B. How novel is this contribution?
High – nice, novel contribution with rating uncertainty clearly shown for stable periods (Figure 7), with stage uncertainty and discharge uncertainty (Fig 9) clear.
C How significant is this contribution?
High – significant showing contribution to uncertainty – especially figure 10b through 10e.
D Is this contribution suitable for the journal?
High – yes excellent contribution suitable for Journal of Hydrology
E Is this contribution of broad international interest?
High – yes broad international interest, globally to address flood frequency and uncertainty.
Attributes of Research
F Are the objectives stated clearly?
High. Yes objectives clearly stated lines 107-114 (Intro) and lines 633-635 (Conclusions)
G Are the methods appropriate? Are the mathematics correct?
High for methods – appropriate for what is assumed and presented. Mathematics correct for what is presented. Authors neglect discussing/presenting gradually-varied flow rating curves besides their single-value rating curve. For large rivers with major floodplain storage (such as this Rhone application), they need to describe why a standard stage-discharge (SD) model is appropriate at this location, given the presentation and discussion of the stage-fall (SFD) rating curve for the Rhone at Valence (Mansanarez et al., 2016), located about 140 km upstream.
Unlike Valence, using a SFD rating curve is not justified at Beaucaire for several reasons. At Valence, a variable backwater from the CNR dam influences the stage. This is not the case at Beaucaire because there is no dam downstream from the gauge. In addition, the sea (and the tide) creates a backwater effect on the stage at Beaucaire. However, this effect only influences the very low flows. As we focus on the floods in this article, we assume that there is no reason to use a SFD model here. Therefore, just like the gauge operator (CNR), we use a SD model here. We added the following sentences at the beginning of section 3.2 (line 273):

"Many gauges of the Rhône River are subject to the effect of variable backwaters caused by the proximity of a dam, and therefore require the use of a stage-fall-discharge (SFD) rating curve model (for example, Valence gauge, 140 km upstream from Beaucaire described by Mansanarez, 2016). Beaucaire is located within a narrowing of the floodplain and there is no dam downstream from the gauge. However, a backwater effect from the sea has been observed at Beaucaire Restitution, but it only affects the very low flows. As this article focuses on floods, we assume that there is no reason to use a SFD model here. Consequently, like the gauge operator (CNR), a stage-discharge (SD) model is used for both Pont de Beaucaire and Beaucaire Restitution gauges."

H Are the data of good quality?

High – data are of good quality for this case study. However, the data URL provided in the Acknowledgements https://www.plan-rhone.fr/publications-131/actualisation-de-lhydrologie-des-crues-du-rhone-1865.html?cHash=5628938abe287dc9ca390dad7373ae0e does not provide data on stages and flows after 2016 in Annex B1-12 Beaucaire Courbes de tarage.xlsx or in Annexe B4-12 Beaucaire-Tarascon Quantiles extrapolés.xlsx. Files were posted in 2018. You can state in the Acknowledgements that the uncertainty propagation code (https://github.com/MatLcs/PropagMaxAn) contains the data, such as the annual max flows- maxpost, and uncertainty in 'Quantiles_Amax.txt'

Line 706 have been changed: "The data is available at https://www.plan-rhone.fr/ (up to 2016) and uncertainty propagation code with the whole data set ('Quantiles_Amax.txt') is available at https://github.com/MatLcs/PropagMaxAn"

I Are the assumptions and analyses valid?

High. Yes the assumptions and analyses are valid, noting the lack of discussion on SFD rating models. There is very minor improvement needed on the rating model fit, showing actual stage-discharge measurements for the largest floods, and uncertainty estimates. Fig 10 (notes below) can be improved with supplemental information (figure) clearly showing the actual measurements for stages at Beaucaire Restitution that exceed about 8.5 m and flows > 9,000 cms.

SFD rating models have been discussed in question G.

Rating model fits for the largest floods are presented in two new figures, see section O.2.

Stages measurements for the three greatest floods are presented in a new figure, see section O.1.

When one only has 3 floods >= (about) 9,500 cms (2003, 1994 and 2002) since the new gauge was established, I find that a 5% estimation error (line 472) is questionable. Stage and streamflow uncertainty for these largest floods may be greater than what is stated/estimated.

Even if this 5% estimation error seem optimistic, it can be explained by the large amount of gaugings during this period, and especially high flows gaugings. In addition, an international consensus conference on the 2003 flood concluded that:

"The most likely estimate of the maximum discharge of the Rhône River at Beaucaire during the December 2003 flood is 11500 m^3/s, corresponding to a return period slightly above 100 years. The maximum stage reached at Beaucaire is estimated at 11.35 m. This maximum discharge estimate is subject to an uncertainty of around 5%, resulting from the uncertainty of flow measurements, maximum stage, parameterization and extrapolation of the December 2003 gauging data." (translated from the French original document). This conference reunited the operators of hydroelectric schemes, the government services and international experts. The original document is no longer available on internet.

We added the following sentence in section 4.4 (line 495): "This 5% uncertainty on recent AMAX discharges is consistent with the results of the international consensus conference on the December 2003 flood : "The most likely estimate of the maximum discharge of the Rhône River at Beaucaire during the December 2003 flood is 11500 m^3/s, corresponding to a return period slightly above 100 years. […] This maximum discharge estimate is subject to an uncertainty of around 5%, resulting from the uncertainty of flow measurements, maximum stage, parameterization and extrapolation of the December 2003 gauging data." (MEDD, 2005)."

J To what extent are the interpretations/conclusions supported by the data/analyses?

High. Interpretations and conclusions are mostly supported by the data/analysis.

Statements about design flood uncertainty and varying record length need to be supported with an annual maximum maxpost Q time series figure before fig 10. Clearly show that the largest flood (given estimated maxpost flows) occurred in 1856, the second largest flood occurred in 1840, and the 2003 flood – which is the basis of the 2000s design? – was the third (!) largest flood.

We added a new figure: figure 3 in supplementary material (see section "Specific comments – Minor revisions needed")

K Overall quality of research

High.

Attributes of Presentation :

L Is this paper

1 well organized?

Yes

2 to the point/concise?

Yes

3 written clearly using correct grammar and syntax?

Yes

M Is the title informative and a reflection of the content?

Yes

N Are the objectives, methods, results and conclusions intelligible from the abstract alone?

Yes

O Are the figures/tables

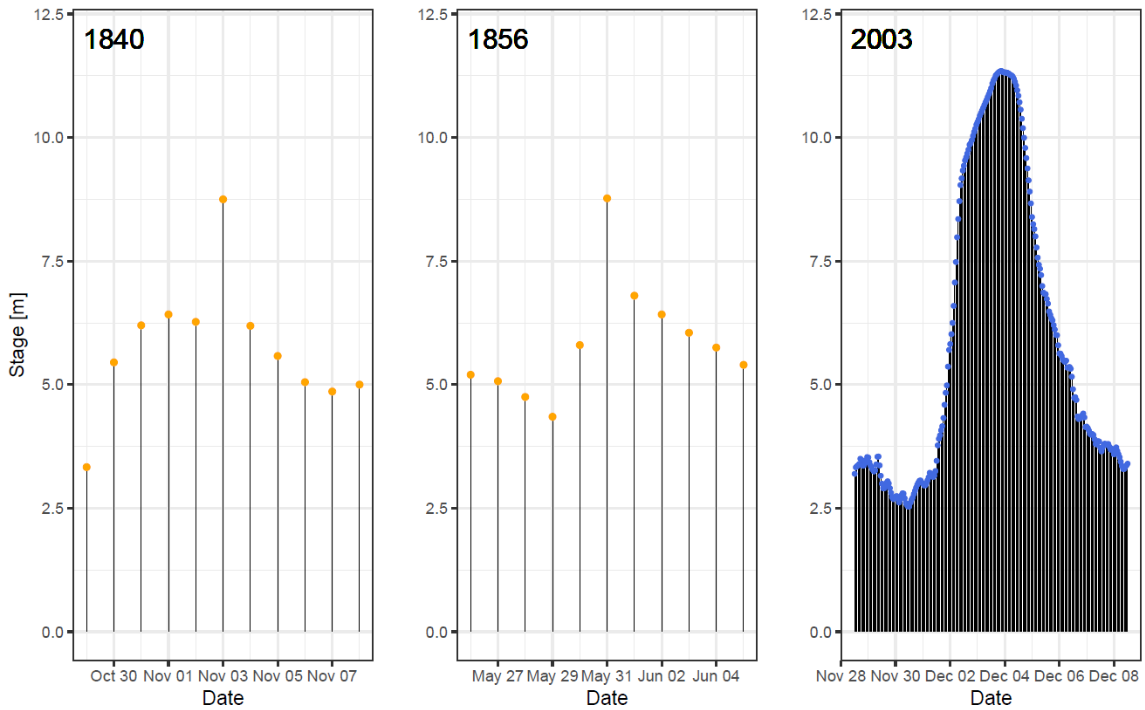
1 useful and all necessary?

Yes- require they add an annual maximum flow time series plot.

We added a new figure; see section "Specific comments – Minor revisions needed"

Could enhance information from fig 5 with supplementary information showing the stage observations/measurements for the 1840, 1856 and 2003 floods.

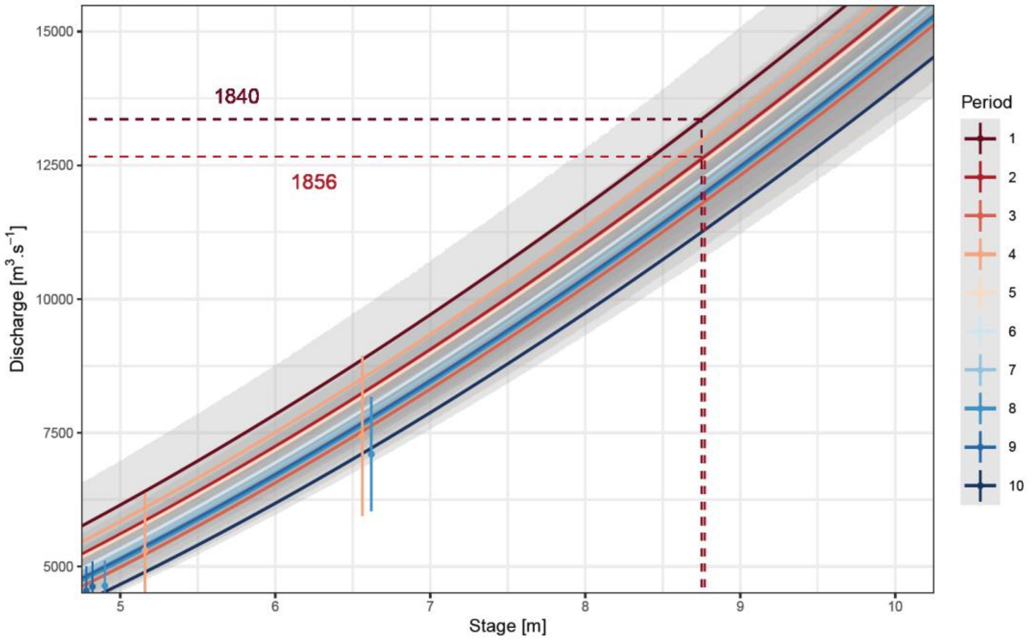
See the following figure added in supplementary material (figure 1).

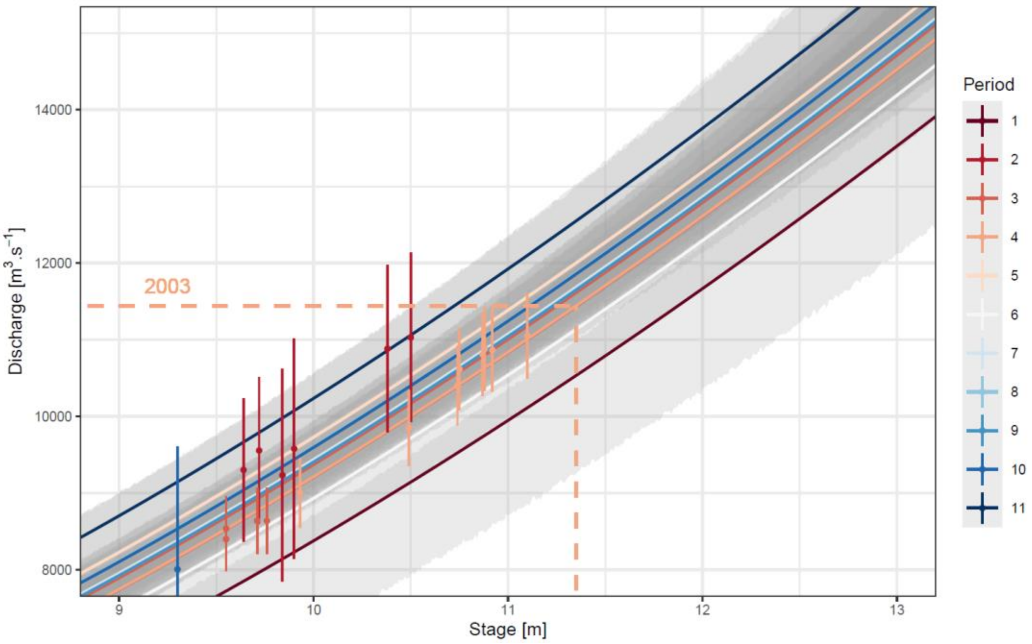


2 of good quality?

Yes. Figures 1 and 5-10 are excellent. Fig 7 b could be improved (via supplemental figure) to show a zoom in to the high-flow measurements, number of measurements, and estimated uncertainty for stage > 8.0 m (floodplain) and flow measurements > ~ 9,000 cms.

See the two following figures, for Pont de Beaucaire and Beaucaire Restitution gauges, added in supplementary material (figure 2a and 2b)





P Is the referencing

1 relevant and up-to-date?

Yes – mostly. Neglects paleoflood research and historical/paleoflood guidelines in US (Bulletin 17C) See suggestions below.

See the answer in the following section

2 excessive or unaccessible?

No.

Q Overall quality of presentation

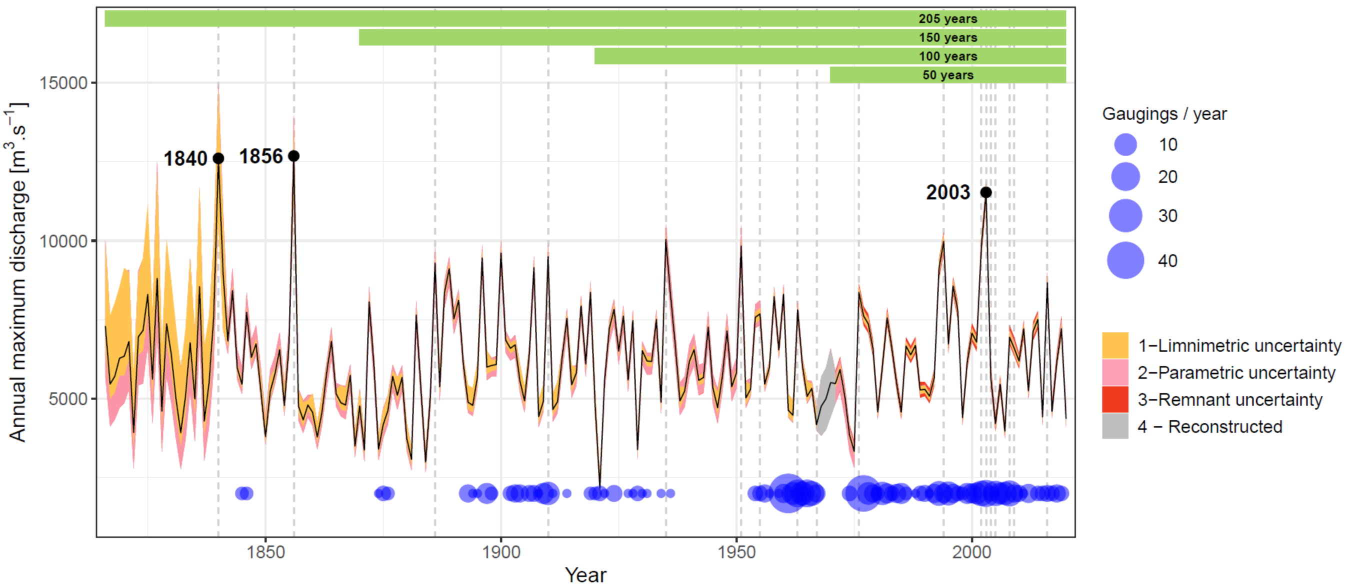
High

Some Specific Comments - Minor Revisions Needed

Explain differences in high-flow (floodway) ratings between the two stations – explain the 2.5 m shift between Fig 7a and 7b for flows > ~7,000 cms.
A few sentences have been added at the end of section 4.2 (after line 470).
"There are several reasons for the significant differences between the upper parts of the rating curves for Pont de Beaucaire (figure 7a) and Beaucaire Restitution (figure 7b). First, the gauge datum (the altitude of the stream gauge zero value) correspond to 3.37 m for Pont de Beaucaire and 0.06 m for Beaucaire Restitution. In addition, as the stations are 2 km apart, their cross-sections are very different. Pont de Beaucaire cross-section corresponds to a main channel splitted in two sub-channels (figure 3b) while Beaucaire Restitution cross-section corresponds to a unique channel (figure 3c). A magnified representation of the upper parts of the rating curves is available in supplementary material (figure 2)."

Show (supplementary material) ratings and extrapolations for the largest floods, especially 1840, 1856, 1935, 2003, 1994, etc.
We added two figures in Supplementary material (see answer to comment in section O.2.)

Show (could be supplementary material) a maxpost (with uncertainty) annual maximum flow time series plot for the entire period. Illustrate (vertical lines) the 50, 100, 150, and 205 year time slices from 2020. This way the reader can easily see in a figure that the < 205 year time series excludes the 2 largest floods. This is needed to support the statements in Section 5.1. A person affected by the 2003 flood should not have been surprised, given the prior flood history in the 1800s.
We added the following figure in Supplementary material and a sentence after line 479 (section 4.4): "An AMAX flow (with uncertainties) time series plot is available in supplementary material".
We also added a sentence in section 4.5.3 (line 507):
"A figure describing these sub-samples along the AMAX time series is available in supplementary material (figure 3)."



Section 5.1: Add a note about historical data and information given the settlement history of Tarascon (town hall in 1640s; Chateau de Tarascon in 1400s) and knowledge that large floods happened in 1840 and 1856, relevant for flood frequency. This additional information should suggest the start year could be easily extended before 1816.

We explained at the end of section 5.2 that additional information on historical floods is available since 1300 in the HISTRHONE database. However, the information before 1816 is not exhaustive and the discharge of those floods is not known. Therefore, the extension of the sample before 1816 requires the formalization of statistical assumptions to guarantee the exhaustiveness (such as perception threshold) and additional work to determine the discharge of those historical floods. This additional work is in progress and is part of Mathieu LUCAS PhD.

We added a sentence at the end of section 5.2 (line 633):

"In addition, as this information is not exhaustive (unlike the stage measurements used in this paper), the extension of the sample before 1816 requires a different statistical treatment."

Note in Sections 5.1 and 5.2 that this data set is a highly unusual situation to have a very long stage record. Instead, most rivers have sporadic flood marks that you note in lines 115-116, such as the classic Pegelhaus on the Rhine in Koblenz, Germany https://de.wikipedia.org/wiki/Hochwasserschutz_Koblenz

And the flood marks on the Pont Wilson, River Loire, Tours – figure 3 in Benito et al. (2015).

Section 5.2 – you completely neglect mentioning paleoflood data, that have been and are used by many countries (Spain, Australia), including in the United States flow frequency guidelines (England et al., 2019). Suggest you add some relevant references such as Benito et al. (2022), England et al. (2019), Harden et al. (2021) and USACE (2020).

We give additional information about the last two comments on line 620 (section 5.1):

« Finally, a promising development is to use sporadic flood evidences older than systematic stage measurements. This data can come from various origins such as testimonies (Pichard and Roucaute, 2014), flood marks (Renard, 2023), paleoflood evidences based on slack water deposits (Sheffer et al., 2003; Dezileau et al., 2014), lake sediments (Wilhelm et al., 2002) or sediments within the Rhone prodelta (Fanget et al., 2013). Various procedures have been developed in the literature to include such data in FFA, through the use of perception threshold and censored data (Brazdil et al. (2006); Kjeldsen et al. (2014); England et al. (2019); Harden et al. (2021)).

Define 'design flood' in lines 118-119 and what the 'official' design flood is (line 128) – it is a specified flood quantile (AEP=0.01 or 0.001)?

We added the following sentence in line 133 (introduction): "In France, the official design flood for flood risk mapping is based on the "largest known flood" or the Q100 flood (AEP=0.01) if the latter is greater. Uncertainty is not taken into account in the official rules"

Line 120 (section 1, after "the following questions… stage measurements"):

"This paper illustrates the chained application of methods to quantify and propagate uncertainty from stage records (and their limited time resolution) and stage-discharge rating curves to the estimation of ~~design flood~~ flood distribution with uncertainties."

Line 130 (section 1, after "the following questions… stage measurements"):

"This procedure is applied to the Beaucaire gauge on the Rhône River (section 3), ~~which official design flood estimates is based on~~ which previous official FFA only used a 80-year long discharge series (Rigaudière et al., 2000)."

Does the design flood estimate account for uncertainty, as in the posterior predictive (expected probability) flow (e.g., Kuczera, 1999)? If not - your FFA estimator can readily provide this mean hazard instead of the median (maxpost) shown in Fig 10a. This may be a recommended 'design flood' estimator if you do not translate the flow frequency curve to a stage-frequency curve (usual practice in United States by USACE).

New sentence in the conclusion, after line 688 "Finally, this article…estimation process":

"Discussions are currently underway with the competent authorities in flood risk management concerning the updating of design flood (Q100, Q1000, Q1500) with enlarged samples (from 80 to 204 year-discharge series), and on ways to account for estimation uncertainty. The use of estimates derived from the predictive distribution (also referred to as the expected probability approach, e.g. Kuczera 1999, Renard et al. 2013) could be considered since it is naturally suited to the Bayesian approach used here. Interesting improvements may come from the use of sporadic flood evidences [...]"

Section 5.3 – lines 605-607: show a posterior pdf of Q>=10,000 cms and Q >=12,500 cms for each case. Would the mean AEP be substantially different, depending on record length and flow of interest?
We explained in section 4.5.3 (lines 507-513 and figure 10) how the uncertainty on flood quantiles varies with the length of the flood sample (40, 100, 1500, 205 annual maximum values) and the return period of the flood quantile (Q10, Q100, Q1000). We add a reference to fig. 10 and section 4.5.3 on line 638 (section 5.3):
"For the specific case of Beaucaire, the use of historical stage records is clearly beneficial up to a 100-year sample size, but the added value is not as clear with longer samples (see section 4.5.3 and fig. 10)".
We do not understand your request concerning a posterior pdf of Q>=10,000 cms and Q >=12,500 cms

Line 544: clarify (I do not understand) "for usual and longer than unusual sample sizes"
We changed the sentence (line 572, section 5.1)

« The case study investigated in this paper is based on a 205-year long continuous sample, which enables an in-depth evaluation of the contribution of the different sources of uncertainty in design flood estimates ~~for usual and longer than unusual sample sizes~~ .from series of several decades to long series exceeding one century of systematic record »

Reviewer #2:

Prof. Neil Macdonald
University of Liverpool, UK

Thank you for the opportunity to review this paper, it is well written and very interesting, the case study presented is particularly valuable in demonstrating the approach applied and figures are well presented, though I have made some minor comments on size and presentation. I have provided an annotated manuscript which I hope will support the authors in making their amendments, it also includes minor suggestions.

It is unusual not to ask for some corrections, but in this instance I have no such requests other than you consider the comments provided in making any revisions, I believe it is an interesting research area and I believe that this manuscript advances current understanding.

Answers of the authors:

All the minor syntax and vocabulary suggestions were considered accurate and were included in the manuscript. The detail of those suggestions are not detailed here.

Abstract : In turn, this induces a 15% increase of the 1000-year flood estimates, a minor difference considering the associated uncertainty.
I would remove this [in red] - finish on the change, it provides more impact of the key messages.
Corrected

L 59 : I have mulled over this and looked at historic diaries to try and understand the uncertainty in the stage levels and diaries, if notable events they are often comparable - I suspect because of the interest in the 'peak' or highest water level is recorded, this can be seen in the relationship between flood marks and stage series in the same settlement even when formed independently.

We agree that additional investigation is needed when historical information relates to a place that have been flooded. We need to decide if it is a maximum level or not. On line 59, we are dealing with another uncertainty associated to the frequency of measurement (the real peak flow may have been missed between two successive measurement of stage). See our sentence in line 354 (section 3.3.1): “Additionally, after 1840, when the stage was rising above 5 m, the operators made more frequent visual records (supposedly hourly measurements). When these records are available, they are of course used to establish the daily maximum stages.”

L 61 : Gaugings -> We tend to not use the term 'gaugings' in English, either gauges or gauged series
As “Gaugings” is differently understood (UK, USA, Australia, New Zealand…), we explain in lines 39-40 and 61 our “definition”, and we changed the title of section 3.4.
Lines 39-40 (section 1, second paragraph):
“This procedure includes three types of errors on hydrological data: stage measurement; discharge measurement (gaugings); stage-discharge model (rating curve).”
Title of section 3.4:
“Gaugings (discharge measurements)”
Line 61 (section 1, page 2):
“Note that the term "gauging" corresponds in this paper (and in the French practice) to the sporadic measurement of the stage-discharge couple, which can correspond to the term "discharge measurement" in the British or American practice.”

L 79 : stability period -> stable period
Corrected all over the paper.

L82 : “Following the detection of rating shifts, rating curves should be estimated for each stability period. This task may not be straightforward, as the number of gaugings available within a stability period is not always sufficient to properly estimate the stage-discharge relationship for the whole discharge range. A common way to address this problem is to artificially repeat some gaugings from other stable periods (McMillan et al. (2012); Puechberty et al. (2017)).”
I am a little unclear what you mean here [in red].
Corrected. (line 83, Introduction):
“A common way to address the lack of gaugings (in particular flood gaugings) within some of the stable periods is to use gaugings from the other stable periods.”

L 115: ancient but regularly measured stage series -> old but continuously or regularly measured stage series
Corrected (Line 118, Introduction)

L 178 : cross channel surveys are often only undertaken post-flood event which means they capture the filling/erosion of the channel but not the state prior to the event, might be worth adding a statement recognising this uncertainty.
New sentence after line 182
“Specific investigation is needed to correctly account for past flood surveys of the cross section, as complex filling/erosion process may have been encountered during a flood with morphogenic changes”.

L 384 : Gaugings -> Gauge readings
We precised the term according to our definition
Section 3.4 Gaugings (discharge measurement)

L 401 (figure 6) : These are important figures, i think that they need to be larger, consider presented one below the other so that they are more easily read and then note as A and B. I appreciate you intention was to offer a continuation of time.
Corrected, figure 6 splitted into figures 6a and 6b

L 405: “This is reassuring about the method since there is indeed no information during long no-gauging periods to identify a shift time.” -> review sentence
Sentence corrected section 4.1., second paragraph, line 419:
“Without discharge measurement (gaugings) the method is not able to detect any rating shift. Additional information may be of interest for those first periods.”

L 599: “Moreover, the interest of long discharge series and flood evidences is not limited to standard FFA, but is also useful for studying the long-term historical variability of floods.” -> see Macdonald & Sangster 2017 HESS as an example for the Uk
Sentence added (line 634, end of section 5.2):
“Moreover, the interest of long discharge series and flood evidences is not limited to standard FFA, but is also useful for studying the long-term historical variability of floods (e.g. Macdonald and Sangster 2017).”