Dear Prof. Blöschl,

Thank you very much for your help in the review process of this manuscript. Find enclosed the revised copy of the text and figures. We have made all the specific changes suggested by Carles Balasch (Reviewer 2). Reviewer 1 was very critical with the paper but never suggest any specific change, just some general comments. Several critical questions addressed by Reviewer 1 are now included in the discussion section of the manuscript (attached below).

The replies to comments and questions from the Reviewers are attached in this report.

We appreciate very much the time and comments made by the reviewers and the editor which were very valuable to improve the quality and understanding of the material presented in our study.

Very best wishes,

Maria J. Machado

The following text (in blue) corresponds to the paragraphs added in the discussion:

Flood frequency analysis with systematic and additional non-systematic information have been widely demonstrated to reduce the quantile estimate uncertainty (e.g. Stedinger and Cohn, 1986; Francés et al., 1994; O'Connell, 2005). In the analysis of frequency using non-systematic data it's more important to know the number of floods exceeding a certain stage (threshold) over a period of time than the precise value of the historical flood peaks (Viglione et al., 2013). The location of the royal palace and gardens on the River Tagus floodplain (Aranjuez) made it highly likely that any flood causing damages to these royal constructions would have been reported and recorded on administrative documents, providing an accurate account of all floods exceeding the bankfull discharge. The statistical analysis with non-systematic records is suitable to deal with uncertainties related with the precise estimation of flood discharges (Stedinger and Cohn, 1986) particularly when the historical information is used as binomial censored (i.e., without defining their exact value, just the lower limit of censoring) without losing statistical significance (e.g. Francés et al, 1994). In our FFA the 32 historical floods, exceeding 350 m^3s^{-1} over the period AD 1779-1912, have been treated as lower bound information, reducing in this way the potential negative impact of the information error.

In recent decades, land use change and anthropogenic climate change impacts on hydrology have largely questioned the stationarity hypothesis (Milly et al., 2008), and several non-stationary models have been applied in the study of annual maximum flood series (e.g. Cunderlik and Burn, 2003; Ouarda and El-Aldouni, 2011; López and Francés, 2013). However, there is still a lack of research assessing the reliability of non-stationary flood frequency models to represent hydrological disturbances due to low-frequency climate variability over centennial historical periods. This study presents a unique record of censured flood discharges over a 300 years period that made possible to test the weight of temporal variability on statistical parameters under a non-stationary assumption.

Anonymous Referee 1

Dear reviewer,

We appreciate very much this critical point of view to our non-stationary analysis of the historical flood period. We understand that from a conventional-type engineering approach, this paper using non-systematic data brings up a number of questions on how to deal with uncertainty inherent to documentary flood data. However, the reviewer should consider this work as a scientific-exploratory approach in the application of non-stationary analysis to centennial flood data series in a site with a complete and continuous record of peak over threshold discharges on the basis of documentary evidence.

1.- In the opinion of the reviewer "Discharge time series are the worst one to make any kind of conclusions in term of climatological forcing. Rainfall and Temperature are surely better since they are direct measurements and the human effects are limited".

Response: This paper does not try to make conclusions of the climatological forcings in the region, and neither is a palaeoclimatological study. Our paper presents a long and complete record of censured discharges during the last 300 years, and tries to understand how statistical parameters may change or not under a non-stationary assumption. In the Iberian Peninsula previous papers have already demonstrated that runoff and peak discharge are influenced by North Atlantic mode of the atmospheric circulation, which we can express in term of NAO index (e.g. Silva et al., 2012). In our study site, we have got the same conclusion by matching flood magnitudes with NAO index, and therefore, it is evident that probability of flooding increase under NAO negative conditions.

2.- Another statement of the review is: "As the authors well known, homogeneity and uncertainty of measured data are not only a statistical hypothesis to be followed but also a philosophical constrain to include in the analysis. While the historical reconstruction is really accurate and fascinating, I do not think that can have a statistical relevance to support any general conclusion. Maybe it can be useful for local hydrological analysis to have an idea of the historical behavior of the river and so can suggest some operational rule for the dam."

Response: We think that such a critic is a general one applicable to any type of flood frequency analysis. Even modern floods are affected by such kind of uncertainty due to potential land-use changes, with a larger impact on small floods that on large floods. Regarding accurate measure of flood discharges, the statistical analysis with non-systematic records has been able to deal with such kind of uncertainties and there is large number of papers indicating the gain of using historical floods in flood frequency analysis (Leese, 1973; Stedinger and Cohn, 1986). As we deal with a methodological development, this study has a general interest to other sites worldwide, and particularly to European rivers.

3.- Another comment is "The uncertainty that affects historical values is not quantifiable and surely is variable event by event. Moreover, it is totally different to the systematic data uncertainty, that is still very high since they are indirect observations."

Response: We agree the inclusion of information error into flood frequency analysis is a very interesting topic, but unfortunately there are few papers addressing it and also it is not covered in this paper. However, we don't agree errors of historical data are larger than errors of systematic record in a gauge station. Historical data focus in medium and large floods and, in our experience, the errors for these floods in the systematic record can be similar or even larger. The main reason is that for the systematic record the data rely in the precision of the stage-discharge curve, which is usually calibrated only for low flows. On the other hand, historical floods usually are reconstructed using hydraulic models. Of course, it doesn't mean they are free of errors, but probably their magnitude is smaller or at least not necessarily larger. Additionally, in case this reconstruction can have serious doubts, it has been demonstrated the historical information can be used as binomial censored (i.e., without defining their exact value, just the lower limit of censoring) without losing too much statistical gain (e.g. Francés et al, 1994). In this paper 46 historical floods have been treated as lower bound information, reducing in this way the potential negative impact of the information error.

4.- On the following comment: "As usual, the discharge time series are flagellated by human activities. Authors underlines that there is an abrupt changes in 1957 for the dam constructions, however I think their influence (statistically speaking) should be related to several years (i.e. 1952, 1953, 1954, 1955, show very low values). If I am not wrong we are talking about reservoirs of several cubic kilometers that affects 80% of the watershed. From the beginning of the construction to the end there are surely some effects on the discharges. Furthermore, the dam operations, at the beginning of the dam life, surely influences data. Maybe contemporaneously rainfall data would have been helpful for understanding what happened in these years as well as the reconstruction of natural discharge could be also useful."

Response: In our record at Aranjuez, low annual peak floods in that decade start on 1953 (62 m^3s^{-1}), 1954 (59 m^3s^{-1}), 1955 (64 m^3s^{-1}), and 1955 (97 m^3s^{-1}). However, it would be difficult to link these low discharge values to the dam construction itself because there are historical precedents of a sequence of years with low discharges due to droughts. In fact, during the previous decade to the starting dam operation, the rainfall in the whole Tagus catchment (see below figure) as well as the runoff produced upstream of the reservoir, were lower than average due to drought conditions. In addition, we don't think that adding these years to the dam regulation will produce any major change in our results, and we rather prefer to stay with the year of starting dam operations (1957) in the application of the Reservoir Index.

See attached Figure 1. Total rainfall in the Tagus River Basin in hm3. Source: Water Authority of the Tagus River Basin.

See attached Figure 2. Annual runoff produced upstream of the Entrepeñas and Buendia reservoirs. Note the decrease in runoff during the decade preceding the starting of dam operation in 1957.

5.- Regarding the last comment "Finally, I kindly disagree with the authors when they say (page 540 - lines 7-11) that the land use changes are out of scope of the paper. In my opinion, this is one of the most important reasons of non-stationarity in a discharge time series that usually is much more relevant than a potential climate change (hydrological change, in my opinion, is more impacting than the climate change). Maybe this watershed is not affected, however, it is important to support it showing land use maps of the past comparing it to the present ones."

Response: Our text in the paper it was not intended to diminish the importance of land use changes on catchment hydrology, but to indicate that we cannot cover in our paper all the aspects influencing flood peaks. Moreover, including several maps with land use and cover changes in this paper will make it more difficult to read and it is not going to provide further insight into our methodological analysis. As indicated in the paper, the area upstream of the reservoirs corresponds to forest and shrubs and the population density has been historically very low. In a general map (next figure) we have marked in a red rectangle the area upstream of the reservoir sites, and it shows in green colours the cover by forest and shrubs, typical of a mountain region. In any case, we will improve this sentence in the final manuscript.

See attached Figure 3. Land use of the Tagus River Basin. The red rectangle shows the area upstream of the study reach, which is dominated by green colour corresponding to forest and shrubs. Source Tagus River Basin Authority.

Dear Dr. Balasch,

Firslty, thank you very much for your time and thoughtful review and comments. All the comments were very constructive to improve the manuscript towards the final publishable format.

Going to the answers to the comments:
1.- p. 530 lines 6-10: “Moreover: : :: : :”). The sentence seems to contradict the data deduced from figure 1c

This sentence referred to the paper by Salgueiro et al 2013, a study carried out in the lower Tagus river basin, where reservoir regulation is not as effective as in the upper catchment. In order to avoid confusion we have change this sentence as follow:

“Recent studies on the influence of the North Atlantic Oscillation (NAO) on the Tagus River flooding show evidence that the largest floods (average recurrence intervals >25 years) are associated with negative mode of NAO during the 20-25 days (of a total 40 days period length) before the flood peak (Salgueiro et al., 2013). Analysis of flood response under natural and dam-regulated regimes (before and after the construction of reservoirs ca 1957-60) revealed changes in the behaviour of flood peaks. In particular moderate floods (return intervals of 10-25 years) were blurred during the post-dam period due to flood peak discharge attenuation by reservoirs”.

2.- Suggestion: In section 4 (Historical flood occurrence and discharge estimates), historical floods reconstruction methodology is described (p. 536 line 8 to p. 538 line 8). This could be presented alongside the rest of the methods in section 3.1 (Floods records database). This section 3.1 could then be retitled as Flood hydraulic reconstruction and database and section 4 as Historical flood occurrence.

Thanks for this interesting suggestion. However, we think that the text presented in section 4 Historical flood occurrence are part of results, which if there are moved to section 3.1, it may result in a mix between Methods and Results which we would like to avoid.

3.- Section 4: It is convenient to add a table with the reconstructed floods (year and peak flow).

There is a total of 59 historical floods, and it would be a long list, plus some comments regarding the hydrological meaning of each discharge (minimum, maximum, double bound). It may result in the long list. All this is already in the figures and if someone wants the data, always can contact the authors.

4. Figure 1c is profusely cited in the text due to its importance. However, it is too small to be read with ease. I suggest to detach this figure from figures 1a and 1b so as to enable its enlargement.

Thanks for this comment. We agree that this part of the figure should be enlarged. In the revised manuscript will appear as a separate figure.

TECHNICAL CORRECTIONS

Typing errors in the manuscript
- p. 527 line 25: accent in Francés
Done

- p. 529 line 14: in a basin of more than 9000 km2, instead of flash floods, it would be preferable to speak of floods or riverine floods.

We fully agree, now is read only “floods”

-p.533 line 16: accent in Francés.
Done

- This reference (as it appears in the text) is not included in the references section p. 536 line 1: Jiang et al. (2014) differs from Jian et al. in the references section (p. 553 line 13)

Corrected to Jiang et al.. The reference now is updated

- p. 538 line 11: it is Figure 1c instead of Figure 1
Done

- p.538 line 24: it is 1916 instead of 1616
Done

- p. 539 lines 9 and 10: specify the area where this happens: the Iberian Peninsula?

Done

- p. 540 line 14: it is Figure 1c instead of Figure 1

Done

- p. 545 lines 4 and 10: accent in Francés

p. 545 line 13-14: reference Cunderlik and Burns, 2003 differs from the one in the references section, where the second author appears as Burn

“s” in Burn is deleted.

p. 546 line 6: it is 1878 instead of 1787

Thanks!. Done

p. 546 line 17: according to table 1 and figure 4, the return period of a peak flow of 400 m3s-1 is 5 years instead of 10, as the text says

Thanks. Changed to 5 years.

p. 547 line 13: Hall et al., 2013 differs from Hall et al., 2014 in the references section
Changed to Hall et al., 2014