Manuscript: Annual flood sensitivities to El Niño Southern Oscillation at the global scale
Authors: P. J. Ward, S. Eisner, M. Flörke, M. D. Dettinger, and M. Kummu
Response to Reviewer #1

General response points

We thank the reviewer for the very positive comments and are very pleased that he/she values the scientific relevance of our research. The reviewer provides several very useful comments/suggestions for minor revisions. We will address these in the revised manuscript, as per our responses to each comment below.

Specific points

1. Meaning of the word "flood" is not immediately obvious. A common definition of a flood is a hydrological event that results in the inundation of land, as well as the definition used here, the annual peak flow of a river. However, it is not until the very end of the Introduction that this definition is provided. This is potentially confusing, especially in the abstract, where the first sentence seems to refer to inundation flood. I suggest either using a different term – e.g. annual peak flow – or a definition of the term flood both in the abstract and close to the start of the Introduction.

   Indeed, there are many definitions of the word "flood", both between communities and within the same community. So, the reviewer makes a very valuable observation to define how we are using this as early as possible. We propose to keep the terminology used in the manuscript, but to make the following clarification in the abstract, namely replace the text "...Here, we present the first global assessment of the influence of El Niño Southern Oscillation (ENSO) on river floods..." with ".... Here, we present the first global assessment of the influence of El Niño Southern Oscillation (ENSO) on annual river floods, defined here as the peak daily discharge in a given year...". We will also add this clarification to the third paragraph of the introduction, which is where we first mention what we will do in this study (and remove the definition that was in the final paragraph of the original introduction, to remove redundancy). We will also specifically note that not all annual floods lead to inundation.

2. P10233. Long lists of past studies on climate change impacts on floods are given, but no indication is provided as to the kind of results these studies have found. It would be useful to provide a more select list of previous studies, in association with some detail on how these studies provide context for the present study. As they are, I find the lists too long and with insufficient detail to be useful.

   Many thanks for the suggestion. We attempted to be as exhaustive as possible in formulating the lists. However, we agree that the lists have become very long, and that the context of why all the papers are given is therefore lost.

   In fact, the purpose of listing the vast body of studies on hydrology and climate change, was to demonstrate that this has received a lot of attention from the scientific community, whilst climate variability (which is also important) has received little. We therefore suggest amending this part by only citing large scale studies (and not the local studies). However, for this section, we do not want to spend too much time summarising all of the findings of this section, since this has been done very well already by IPCC. Moreover, we stress again that the point is simply to *note that there are many such studies, whilst climate variability is understudied. We propose to amend this part of literature review as follows:*

> *"In recent decades, a large number of studies have examined instrumental discharge records to identify possible changes in flood frequency and/or magnitude due to climate change at national to continental scales (e.g. Allamano et al., 2009; Conway et al., 2009; Cunderlik and Ouarda, 2009; Di Baldassarre et al., 2010; Douglas et al., 2000; Hannaford and Marsh, 2008; Hirsch and Ryberg, 2012; Mudelsee et al., 2003; Shiklomanov et al., 2007; Villarini et al., 2009; Villarini and Smith, 2010), with many more studies than can be listed here focusing at basin scales. There is also a growing literature on possible changes in flood frequency and/or magnitude based on future hydrological projections. Studies at the continental scale (Dankers and Feyen, 2008, 2009; Feyen et al., 2012; Kitoh et al., 2011; Lehner et al., 2006) to global scale (Hirabayashi et al., 2008, 2013; Milly et al., 2002) show differing signals of potential change across regions and between models and/or scenarios. Specific studies at the local to national scale are too numerous to be listed here, but are summarised in past reports of the Intergovernmental Panel on Climate Change (IPCC), including the Fourth Assessment Report (Kundzewicz et al., 2007) and the Special Report on Extremes (IPCC, 2012).".*

In the next part of the literature review, which describes past studies on ENSO and river discharges, we will add a few sentences stating that many of the past studies (at local/regional scale) found significant relationships between ENSO and discharge. We propose to amend the text to read:

> *"Many past studies have assessed ENSO's impacts on average river flows at the local to basin scale (see, for example, Dettinger et al., 2000, and references therein). Since many of these studies were carried out in regions known to be sensitive to ENSO, many have found significant correlations between average river flows and various indices of ENSO. A few studies have examined global scale relationships between ENSO and average river flows (Chiew and McMahon, 2002; Dettinger and Diaz, 2000; Dettinger et al., 2000; Labat, 2010), based on discharge measurements from gauging stations, and have found significant relationships in many regions. In contrast, only a few studies have examined relationships between ENSO and peak flows. Most of these studies have focused on the United States (e.g. Bell and Janowiak, 1995; Cayan and Webb, 1992; Cayan et al., 1999), although studies have also been carried out in northern Peru (Waylen and Caviedes, 1986), South Asia (Mirza, 2011), and the Mekong Basin (Räsänen and Kummu, 2013). To a large extent, the lack of observed daily discharge data in many regions has hampered the kinds of consistent global scale assessments that are needed. Ward et al. (2010) examined the relationship between ENSO and observed annual peak discharge for 622 gauging stations, but the geographical coverage of those stations was highly biased towards a few regions (particularly North America and Central Europe), and for many regions data were limited or lacking".*

3. P10235. The Watch Forcing Dataset (WFD) extends to 2001, not 2000 as stated here.

That is correct; we will amend this error to read 2001. Note that we used the data to 2000.

4. P10236, line 7: why specifically was the number 34 chosen (for the 34 largest basins only to be divided into sub-basins). This seems like an arbitrary decision, so some explanation is needed.

The 34 largest 'first-order' basins are those with areas greater than 750,000 km^2, and were further sub-divided to improve the assessment of the water resources. We will clarify this in the revised manuscript.

5. P10236, line 11. Why did the analysis stop in 1999 when it is stated on line 18 that the WaterGap model was run until 2000, and the WFD extends to 2001?

It is correct that we simulated daily discharge using WaterGAP until 31 December 2000. However, since we were using hydrological years (and not calendar years), this meant that the final hydrological year that we could use was 1999 (i.e. Oct 1999-Sept 2000). However, this was not so clear in the original manuscript. In our revision, we will clarify this by amending the section around P10236 line 11 to read "For each grid cell and hydrological year, we calculated the maximum annual discharge, or annual flood discharge (Qmax), and the mean annual discharge (Qann) from the simulated daily discharge time series for ***hydrological years*** *1958-1999". Also, we will add a statement in section 2.2 and the caption of Table 1, clarifying that we refer to hydrological years using the year number in which they begin (rather than the year in which they end).*

6. P10236, section 2.2. Definition of the water year. Whilst I appreciate the desire to categorise the world into a small number of categories, I am a little concerned that the results shown in Figure A1 are not realistic in some locations. For example, in southern hemisphere temperate locations one might expect the hydrological year to start and end in the autumn months. Can the authors comment on the basis for choosing only July-June as an alternative to October-September, and on whether there is a possibility that the definition of the hydrological year could have affected the results.

For this paper, when defining the "hydrological year", we were specifically concerned with choosing a 12-month period in which the simulated Q_{max} of each year would occur. We chose to use the standard hydrological year as default due to its prevalence in the scientific literature. However, the reasoning for choosing an "alternative hydrological year" in some regions is that standard hydrological year could be problematic for those regions in which the peak annual occurs in the months of September-November. In those basins, the peak annual flow occurs at around the change of the "standard hydrological year". Hence, an alternative was chosen that would ensure that the peak annual flow in those aforementioned regions could be correctly categorised into a hydrological year. For that purpose we used July-June. Indeed, as the reviewer mentions, we could have also chosen a year change around the (arboreal) autumn, such as April-June (i.e. mirroring the standard hydrological year). However, this should yield more or less the same Q_{max} time-series, since the peak discharges in those regions are occurring in September-November (and therefore occur in the same hydrological year whether we use July-June or April-March as the hydrological year).

However, to test this, we have re-run the correlation analyses but instead using an alternative hydrological year of April-March. The impacts on the results are minimal. For example, for more than half of the basins classified as having the "alternative hydrological year", the change in Pearson's r is less than 0.02, and less than 0.10 for more than 95% of the basins. There are only a few basins where the selection of the hydrological years does have a clear impact on the strength of the correlations/sensitivities. This is demonstrated in the figure below. Here, we show: (a) the *sensitivity results based on the alternative hydrological year July-June (as presented in the paper); and (b) the same results based on the alternative hydrological year April-March. The figure also shows the differences to be minimal, although there are a few basins, particularly in hyper-arid regions, where the results do change a little (for example in the western parts of the Sahara). This is because the month of peak discharge is highly variable in these regions, and therefore difficult to categorise. However, the figure demonstrates that overall our results are very insensitive to the assumed hydrological years.*

In the revised manuscript, we will add the following text to section 2.2 to show that we have investigated this: "We also tested the sensitivity of the results to the choice of the months for the alternative hydrological year. To do this, we also used April to March (instead of July-June) as the alternative hydrological year. The differences in the correlation and sensitivity results were found to be minimal in most regions, with some small difference in hyper-arid regions (such as western parts of the Western Sahara), where the month of peak discharge is variable."

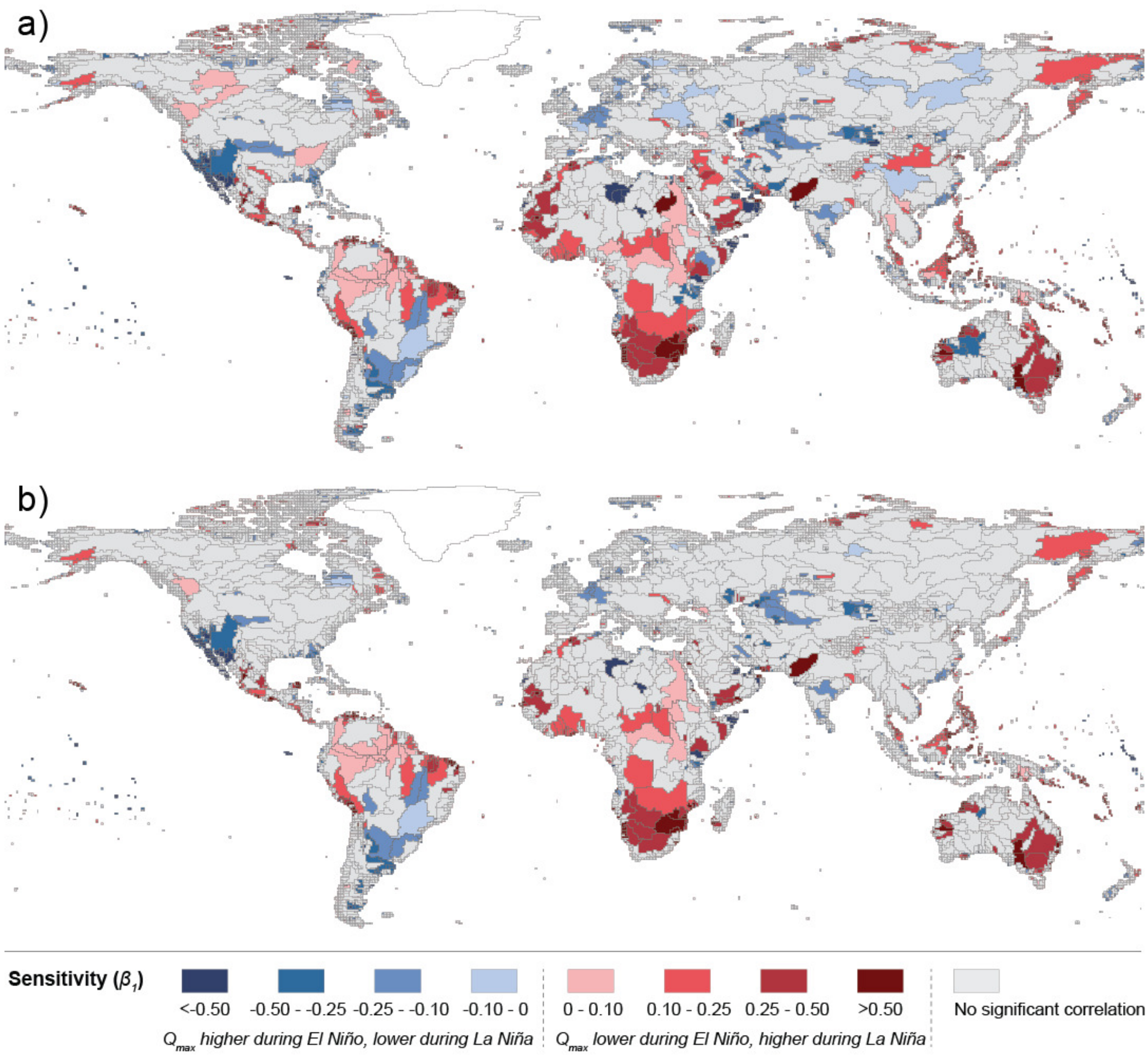

Figure: Sensitivity (β1) of ln Q_{max} to variations in SOI for basins with significant correlation (Pearson's r, α = 0.10) (basins where the correlation is not significant are shown in grey). Part a shows the results using the original "alternative hydrological year" of July-June, and part b shows the results using the "alternative hydrological year" of April-March.

7. P10241, lines 19-20. Could the authors comment on why annual maximum flow might be more sensitive to the SOI than annual mean flow – perhaps in relation to the previous studies cited?

There is a section stating that other studies have also noted this difference using global gauged data, and using gauged data for the USA (p. 10241, line 24+). In the revise manuscript we will also add that another study found similar results for a set of different atmospheric circulations in Europe. Whilst we agree that a research priority should now be to attempt to identify why this is the case, we feel that answering this is beyond the scope of the current paper. Hence, we propose to add the following amended passage: "In our earlier work (Ward et al., 2010) based on observed discharges at 622 gauging stations, Ward et al. (2010) we also found that, on average, ENSO has a greater impact on annual flood discharges than on mean discharges. Similarly, for observed discharges in the western USA, Cayan et al. (1999) found ENSO to have a greater impact on the number of days exceeding the 90 percentile values of streamflow as compared to the number of days exceeding the 50 percentile (i.e. median) values. In Europe, Bouwer et al. (2008) also found annual peak discharge to be more sensitive than annual mean discharge to variability in various large-scale atmospheric circulation patterns. Research is now required to examine the mechanisms behind these apparent differences in the sensitivity of peak and mean discharges to large scale atmospheric circulation".

8. Figure 4. I'm not sure of the relevance of part a of this figure. What is the point of presenting the sensitivity to the SOI when that sensitivity does not result in statistically significant correlation? Deletion of the current part a and enlargement of part b would also help the readability of what is at present quite a complex figure to decipher.

Indeed, we had many discussions on whether to only present the results in Fig 4b (i.e. with significant correlations only), or whether to also present all results (as is now the case). A problem with only showing the results for the significant correlations is that the choice of the confidence limit to show will always be arbitrary, and this is why we originally chose to show both sets of results. However, in light of the reviewers comments, and the suggestions that the current figures is difficult to read, we have decided to do the following: (1) present only Fig 4b in the main manuscript of the revised manuscript, whereby we will rotate the figure so that it can be displayed larger for better readability; (2) move current fig 4a to the appendices so that the information are still available to those readers who wish to see this; and (3) in the appendices also present the results using a confidence level of 95%. Generally, we find the same overall regional patterns, though of course the number of basins with significant correlations does reduce, particularly for many of the very small single cell "coastal basins". This information will also be added to the revised manuscript. The results for the 95% confidence level area also shown below.

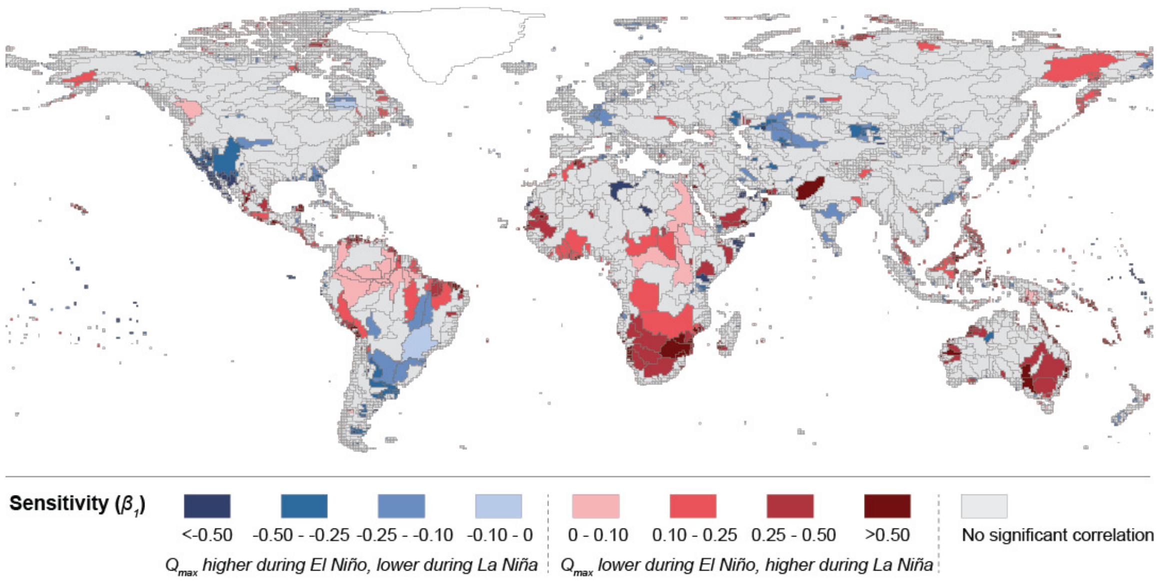

Figure: Sensitivity (β1) of ln Q_{max} to variations in SOI. Sensitivity is only shown for basins with significant correlation at a 5% confidence interval (Pearson's r, t-statistic, $\alpha = 0.05$) (basins where the correlation is not significant are shown in grey). Blue indicates negative correlation (higher annual floods in El Niño years/lower annual floods in La Niña years); and red indicates positive correlation (lower annual floods in Niño years /higher annual floods in La Niña years).

9. P10242, line 19. Please can the authors comment on whether the apparent high sensitivity of mean annual peak flow in arid regions could be linked to the use of percentage, rather than absolute, values? For example, a relatively small change in absolute flow in a region with low peak annual flows would appear as a large percentage change – whereas a much larger absolute change in a river with high peak annual flows would appear as a much smaller percentage change.

It may be the case that the anomalies in peak discharge between the ENSO phases (such as those shown in Figure 8) are higher in arid regions as a result of this. However, for the sensitivities for these analyses we did not use discharge (percentage) anomalies, but rather absolute values of (log) Q_{max} per year.

With regards the anomalies, there could be a whole range of explanations, including that one suggested. In some regions, another possible explanation could be that the large anomalies could be related to the fact that these biggest (annual) floods reflect a regime where most of the precipitation (and flows) come in just a very few isolated storms; thus all ENSO has to do is to modify the number of storms or the magnitude of a few storms, and it can have a large influence on any given year's Q_{max}. In other settings where the number of storms is larger (more wet days overall), ENSO may need to have a much larger more persistent influence to be reflected in Q_{max}. Another possibility is related to the characteristics of arid areas themselves: arid regions are particular challenging for hydrological simulations, with particular characteristics such as high rainfall variability, extensive surface runoff, reduction of infiltration capacity (crusted soils), lack of vegetation, and so on.

We therefore feel that further research would be required to give a sound answer to this question, and therefore add the amended passage: "However, less research has assessed the *influence of ENSO on the hydroclimatology in arid regions. Whilst the paucity of observed discharge data in many of these regions limits the validation of our model results there, the strength of the signal provides motivations for enhancing research activities in those regions, in order to examine whether this is related to physical processes (and if so which), and/or whether this is related to the high coefficient of variability in peak flows".*

10. P10244, line 6. Why was a 21 year moving window used? This is another apparently arbitrary choice that requires some justification.

A very relevant question. In fact, we made a trade-off between the number of years per window (21) and the number of windows (21), in order to try to preserve power in both the correlation test (Pearson's r) and the test of trend (Mann-Kendall). We do believe that the 21 values therefore available for the Mann-Kendall test are sufficient to detect a trend, and therefore non-stationarity, as can also be seen by the fact that statistically significant trends are indeed detected for some basins. However, we have now also tested how stable these trends are if we use a shorter window. We repeated the analyses using a moving window of 15 years, which therefore gave us 27 "windows" from which to detect the trend. Based on this, we found only a small number of changes, namely the Yellow River, Murray, and Ohio displayed no significant trend (instead of strengthening), and the Tocantins displayed no significant trend (instead of weakening). These changes do not affect the overall storyline and conclusions of the paper, and it should be noted that the 15-year time period used to calculate the individuals values of Pearson's r is short. Hence, we choose to keep the 21-year moving window. We will add the following statement to explain why this was chosen: "A 21-year moving window was used as a trade-off to maximise both the number of years per window (21) and the number of windows (21)". Moreover, we will also add the following text "We also repeated the analyses using a 15-year moving window (which yields 27 windows for the trend detection). The results of the latter analyses were similar to those using a 21-year moving window, with the following differences: the Yellow, Murray, and Ohio rivers displayed no significant trend (instead of strengthening), and the Tocantins displayed no significant trend (instead of weakening)." We also considered testing the results using a 30-year moving window. However, this yields just 12 windows, which we consider to be too short for a meaningful assessment of the trend.

11. P10244, line 10. Could the authors confirm the p-value used as the threshold for statistical significance.

Correct, we used $\alpha = 0.10$ due to the low number of observations (n=21). See also response to comment #10, above.

12. P10246, line 16. The PDO is the north Pacific representation of the Pacifc-wide Interdecadal Pacific Oscillation (IPO). The IPO (and so PDO) underwent a phase change in 1977-1978, which has been linked to a change in the variation in ENSO. In this context, it would be extremely interesting if the authors could comment on whether they see any difference in SOI relationship to annual peak flows across the IPO phase change in 1977-1978.

We agree that it would be very interesting to examine possible shifts in ENSO correlations associated with different large-scale circulation patterns. However, based on the analyses and data available in this paper, we do not feel it is possible to give any scientifically-sound remarks *on this, and that anything we mentioned would be too speculative at this point. A simple visual inspection of the results in Figure 7 could suggest a possible change in the correlation strength around 1977-78 in some basins, but certainly not all, and certainly not conclusive. The results in Figure 7 do however show that the ENSO relationships with Q_{max} show non-stationarity over time. We believe that what may be causing these "changes in teleconnection strength" is a matter for further research, as discussed in Section 4.4.*

13. P10247, line 10. Asymmetries between El Nino and La Nina influence on annual peak flows may also be indicative of a nonlinear effect on climate of El Nino vs La Nina – e.g. as shown previously in New Zealand: Mullan, 1995, International Journal of Climatology, 15, 1365-1386.

Indeed, we agree with the reviewer. This is also reflected in the manuscript with the reference to Hoerling et al (1997). We thank the reviewer for pointing us to the paper of Mullan, which we have added to this section.