Review of Najibi and Devineni (2017) in ESD

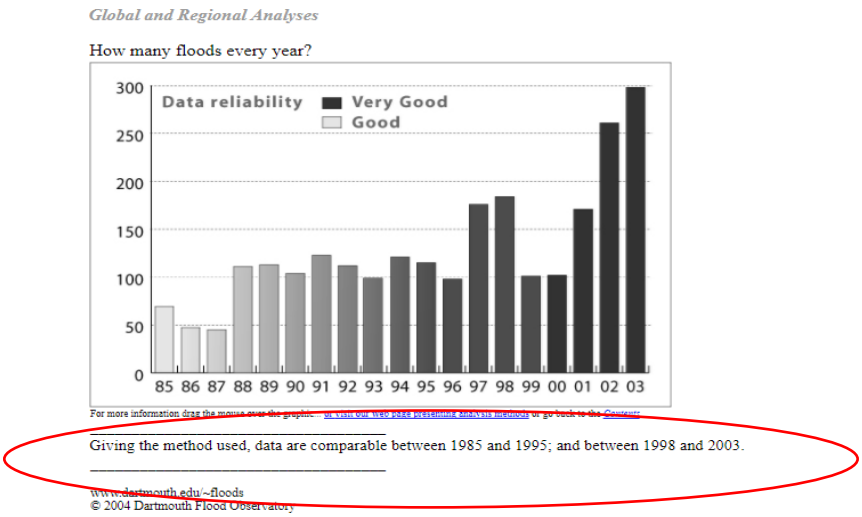
General Comment

Najibi and Devineni (2017) presents an analysis of trends in global flood frequency and duration between 1985-2015 based on flood data from the Dartmouth Flood Observatory's (DFO) global flood database. Using standard trend and change point analysis they find an increase in both the frequency and duration of floods in many regions (here using latitudinal belts), with a change point centred around the year(s) ~2000. A further step attempts to attribute these changes to large-scale drivers such as ENSO/PWC/GPH using a GLM framework. While the pursuit of understanding how floods are changing at a global scale is indeed a worthwhile topic, I have some serious concerns about the way the paper uses the DFO dataset as a single source of information on flood frequency and duration to go about this. I list a few of my major concerns about the use of this dataset, the methodology applied, and attempt to offer some suggestions to move towards a stronger manuscript. Until these major concerns are addressed I won't provide a full detailed review.

Specific Comments

1) I am not convinced it is appropriate to statistically test for trends/change points using a database such as the DFO, and then assume any changes are physically driven. I do not have an issue with the DFO archive in general, and in fact believe it is a very useful database of reported floods and impacts, however I don't think you can treat such a database like an observed flood event series derived from e.g. gauged river flow stations. While the authors acknowledge some uncertainties with the dataset (e.g. Section 4.1), they really do not address some key known errors and uncertainties of the use of such flood report databases. Here are a few:

a) How can you know that the increase in floods is not because of an increased level of the media reporting floods/access archives or changing quality of information entering the archive? Here is a graph taken directly from the DFO website (http://www.dartmouth.edu/~floods/archiveatlas/floodnumber85_03.htm) showing the changing reliability of the data in the archive over time. I also highlight in red the sentence "… data are comparable between 1985 and 1995; and between 1998 and 2003" [that analysis was done in 2004] – is there something within the DFO methodology that means these two periods are not comparable?



b) L88-93: A change in the information DFO uses is known to have changed majorly after 1999, by introduction of MODIS. Surely this would have a large impact on the temporal homogeneity of the frequency and duration of floods and thus could be a likely explanation for why so many change points were detected in the time around the 2000s, and hence the increasing trend?

c) DFO Duration data. The start date/end date within the DFO appears to be very uncertain and dependent on news reporting time-scales. This is very different from actual physical flood duration/inundation. There is a well-known issue also with separating multiple-events in close succession within these such databases that tend to aggregate many smaller events (Chapter 3 in Kundzewicz (2012) might be useful here). While I recognise some flood inundations can take weeks or even months to recede, many large floods recede in a much shorter time-frame (within a few days). How likely are the floods reported to last e.g. 21 or longer to be actual inundation of an affected populated area, or an artefact of news reporting etc.? Overall, more careful consideration of what is meant by flood duration and a discussion of uncertainties is needed.

d) A record of 31 years, no matter if using observations or DFO data, is very short for a trend analysis. I acknowledge that the spatial availability of observed gauged flow data is limited for many countries, but it would be important to know if trends even in the overlap period, for regions that have data, are similar to those shown for the DFO data. In those cases it would also be useful to know if the pattern seen in the 1985-2015 record is part of a longer-term trend or just short-term climate variability. Even though you use e.g. ENSO within the GLM framework to assess drivers, other drivers that operate at a longer time-scale could be important (e.g. AMO and floods, see Hodgkins et al., 2017, especially Fig. 10b). You will see from their paper that the period of ~1985-2015 really is on the short side. It is my strong opinion that there needs to be some corroboration with the DFO data and observed flood frequency datasets that could be extracted from e.g. Global Runoff Data Centre (GRDC) records (and for 'duration' metrics if enough data exists?) or at least for a select number of countries with good data in the overlap period. Do floods also increase in a similar manner, is the inter-annual variability/trends the same between different flood datasets over 1985-2015?

2) A note about the methods. The paper would benefit to have all the methods described within the methods section. E.g. The methods description (and results!) from the GLM analysis is given in the discussion section (L396-439) instead.

3) While the use of the Mann-Kendall and Pettitt tests are useful for working with environmental data as less assumptions are needed (non-parametric), they do still need to satisfy the assumption of independence in statistical hypothesis testing. So at a minimum the series should be tested for serial correlation and dealt with appropriately if detected (e.g. block bootstrapping perhaps).

4) Page 24 Fig. 2: The bottom right panel (i.e. Mid-latitudes (S)) have many zeros. This is an issue when using the Mann-Kendall test as it introduces many 'ties' that can impact the variance of the test. Does the Mann-Kendall version you use deal with ties? Or have you thought about using other methods often used for frequency or count data such as logistic regression? If not, perhaps there are too few data points within the Mid-latitude (S) region, in particular, to use the Mann-Kendall/Pettitt tests reliably.

5) The section from L326-333 does not belong in the results section. Further, the whole text in Sect. 4.3 is not formally being tested so does not deserve a separate section. It is useful to mention throughout the discussion in passing, but to aim of the paper is not to test "Have the exposures of residential/industrial sectors to flood events increased recently?"

6) Page 23 Fig. 1: Can the method differentiate floods spatially within a country? For example, if a flood was reported in 'Australia' how does the method determine if the flood occurs within the 'Subtropics (S)' or 'Mid-latitudes (S)' belts?

7) A key strength of the paper is the effort to explore the drivers of flood variability and change using large-scale predictors (e.g. ENSO). This does not feature in the original hypotheses tested in L58-65 + Sect. 2.4 + Table 1. In my opinion, if the paper was to re-focus and expand upon this element, and/or strengthen the trend analysis components with other methods and datasets, it would lead to a much stronger analysis.

I'm happy to give more specific and detailed comments if my above concerns are addressed in a future iteration of the manuscript. I do of course see merit in this work on such an important and challenging topic, but I feel additional effort is needed with regards to corroborating the DFO dataset and strengthening the methods in the first instance.

References

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., Fleig, A. K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C. and Wilson, D.: Climate-driven variability in the occurrence of major floods across North America and Europe, J. Hydrol., 552, 704–717, doi:10.1016/j.jhydrol.2017.07.027, 2017.

Kundzewicz, Z. W.: Changes in Flood Risk in Europe. [online] Available from: http://www.crcpress.com/product/isbn/9780415621892 (Accessed 20 August 2014), 2012.