Interactive comment on "Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models" by I. Giuntoli et al.

Referee: Antonio Speranza

The addressed topic is interesting and quite a bit of performed work is reported in the paper; in particular, I appreciated the effort in assessing the relative contribution of GCMs and GHMs to "uncertainty". I think, however, the reader could be helped in forming his own judgement if some more information was provided here and there; for example, by showing specific examples (giving an idea of the nature of analysed fields) of runoff fields for different models.

At some points I had problems in following the proposed reasoning:

- The proposed analysis suggests major impact of GHMs on polar and arid areas, where the runoff component is minor. Doesn't this go without saying?

- In Fig. 1 to the observed increase in the percentage of HFI days in high latitudes in the northern hemisphere (Fig. 1a) values of signal-to-noise ratio less than 1 correspond; only in some regions - Alaska, north-west America, north China - S2N>1 (model agreement): what motivates, then, the conclusion "…exhibit a number of robust large-scale features. Increases in high flow days were found at northern latitudes, with a strong signal over…"? Similar considerations hold for seasonal maps.

I share with Referee 1 doubts concerning the "novelty" of proposed work, but I leave to him evaluation of answers proposed by the Authors.

All in all, I believe the paper is in line with similar literature and it can be published with the suggested amendments.

But, just for the sake of scientific discussion, let me raise a general question concerning the specific use of observations over threshold proposed in this paper.
The authors mention at pag. 12, lines 15–29 that "The identification of high and low flows over long time series, and particularly over climate projections, is non trivial. As an illustration, van Huijgevoort et al. (2014) in their multi-model ensemble study on droughts, report that applying the threshold level method to the future period using a threshold derived from the control period can lead to spurious pooling of drought events. They suggest that future changes could be counted for by linking the drought threshold to adaptation scenarios like Vidal et al. (2012) did over France. Wanders et al. (2014) used a transient threshold level method for a moving reference period, in order to reflect the changes in hydrological regime over time, finding that the non-transient threshold method projected larger shares of areas in drought (except in snow dominated regions).".
Now, use of very low (because of trend) thresholds for the selection of "extremes" generates problems typical of "fat distributions": the sets of events over threshold are rich in number, but include non extreme events! Use of different models further complicates the situation. As a consequence, even under conditions allowing application of Gnedenko ("three type") theorem[•], the distribution of the selected events is non-parametric and the reliability of the statistical inference is very poor (in particular for what concerns the "tail events"). But, at the same time, "central limit" conditions are not fulfilled and the distributions cannot be considered normal[•]. Geographical non-uniformity further complicates the picture.
As a consequence, the inferred statistical estimators are presumably characterized by (very) weak reliability (in particular for what concerns "real extremes") and have, at most, a qualitative meaning: I would not base any relevant decision on them!

• Note that in the specific case in question even a simply linear trend would introduce an additional dimension to the parametric space of the distribution.

• By the way, isn't normality a necessary requisite in ANOVA?