Review of “Global estimates of 100-year return values of daily precipitation from ensemble weather prediction data” by F. Ruff and S. Pfahl

General comments

The authors develop a method to reduce the sampling error in estimates of very extreme daily precipitation, based on a new dataset consisting of a vast number of weather forecasts.

The manuscript is of a very high standard in terms of clarity, and contains a thorough description of data, methods and results. However, there is one science aspect that needs further consideration. While statistical uncertainty in estimates of extremes is reduced in this new dataset, the evidence on their bias is limited to precipitation amounts for normal weather types (1-in-2 day, and 1-in-10 day) rather than the extreme weather conditions producing the most intense rainfall. A fuller evaluation is required to prove ensemble forecast data from a relatively short period of real weather situations can provide more accurate estimates of very extreme precipitation.

If a revised version included more appropriate validation, and considered the extra points below, then publication as a NHESS Highlight article may be suitable since it has potential to make a substantial contribution to a subject of growing importance to society.

Major comments

1. The assessment of the new dataset versus observed values is described in lines 208-231, and Figures 2 and 3. It is limited to an evaluation of the 1-in-2 day, and 1-in-10 day precipitation, which concern the most common events producing precip at locations. However, the new dataset provides information on those very extreme events caused by rare atmosphere conditions, and no evidence is given on forecast bias for highly unusual weather events. As a consequence, little confidence can be attached to estimates of the 1-in-100 year return values from this new dataset.

The evaluation of the extreme right tail of forecast data is required. The observed data in this study are sufficiently long to contain relatively small statistical uncertainty for precip amounts exceeded every one year, up to perhaps once in 10 or 20 years. If the new dataset is consistent with such rare events (or can be calibrated to be more consistent), then more confidence could be placed in the new estimates of 1-in-100 year return values.

The last sentence in the Abstract may need revision after the model is validated.

2. There are many land areas in Figures 2 and 3 (b, c and d) which contain almost no information on bias due to the use of absolute values. Could the authors plot the relative differences (in %) between EPS and the three obs datasets in Figures 2 and 3 (or any new exhibits made as a result of comment 1 above)?

3. Could the authors include return value plots for a selection of locations representing their global results? For example, return period on the x-axis (a log scale, from 1 to 100 years, or more) and

return value on the y-axis, for EPS and the three observational datasets, and locations in interesting regions such as northern and southern Europe, west and east US, Arabian Peninsula, India, Brazil etc? This would be a very useful addition in Section 4 (after any bias correction is applied to EPS in Section 3, see major comment 1 above). Inclusion of Confidence Intervals for these representative locations, perhaps for EPS and REGEN, would be very useful for the reader too.

4. This comment is intended as a suggestion. Many of the maps are dominated by values over the ocean, which distracts the eye away from the changes over land. Given how this study is focused on 100 year return values, and these are most relevant to flooding on land, do the authors think it worthwhile to focus on land-only changes in all figures in the main manuscript? Maps with values over oceans could be included in a Supplementary section, if the authors wish to include them?

5. The added value from the eight maps in Figures 6 and 7 is quite limited, since EPS naturally has smaller sampling error than observed datasets due to its much longer record (1224 years versus a few decades). Perhaps 2 of these 8 maps would suffice, e.g. Figures 6a and b, leaving space for more locations to be examined in detail, discussed in point 3 above?

Minor comments/corrections

1. line 37: should the author be referred to as World Meteorological Organization, rather than 'Organization'? (With consistency in References section too.)

2. lines 115-116: could the text include the year of these ECMWF Cycles, to make them more meaningful for the reader?

3. line 144: change '135.000' to '135,000'

4. colours chosen in most figures: the authors choose different shades of red to represent wetter conditions, and blue to represent drier. In my experience, most researchers choose the reverse, I've become familiar with 'blue = wetter'. Would the authors consider reversing their colour scale, or are they more familiar with 'red = wetter'?

5. caption in Figure 1: "Mind the logarithmic colour scale" is acceptable spoken English, but not standard written English. This applies to captions for figures 5 and 6 too. A small tweak such as "Note the logarithmic colour scale" would be sufficient.

6. line 360: I recommend deleting 'hence, reduce flood risks' (not necessary and breaks the flow of sentence).

7. line 370 ff: it may be worth adding how these quoted precip values refer to averages over 1 deg grid cell areas? (For example, some northern Europe locations have recorded daily total far in excess of 100mm, due to events such as storm Bernd.)