Interactive comment on "Effects of sample size on estimation of rainfall extremes at high temperatures" *by* Berry Boessenkool et al.

Anonymous Referee #2

Received and published: 2 August 2016

Although interesting I find this paper quite confusing. It combines different distributions (GPD, Wakeby, Weibull, etc.) and different parameter estimation procedure (L-moment, maximum likelihood, etc.), different threshold values. It is difficult to come out with a clear picture of the whole think as many elements can have an impact on their conclusion that decreasing extreme rainfall below C-C scaling above 20°C can be explained by sampling issue. Many authors (e.g. Lenderink and van Meijgaard 2010; Panthou et al. 2014) used dew point temperatures to check for possible changes in relative humidity with temperature and concluded that, for some regions, decreasing intensities at higher above 20°C disappeared suggesting that humidity is a limiting factor for these regions. In these cases, sample sizes remain unchanged as the number of rainfall recorded for these temperature bins remains unchanged and therefore the change in the C-C scaling cannot be attributed to sampling issue. Therefore I would suggest

that the authors, if these records are available, use dew point temperatures to look at possible change in the shape of the extreme rainfall temperature scaling. I also have some concerns regarding some of the results. For instance in Figure 5, it is disturbing that the parametric estimates doesn't converge to the empirical value. Regarding that point, the authors mentioned on lines 26: 'Due to the inherent structural differences between distribution functions, parametric quantile estimates range from 20 to 40 mm/h, thus the weighted average is slightly higher than the empirical value of the full dataset.' This is not a convincing explanation (and not an explanation at all). This Figure is important, and for me, this discrepancy between empirical and parametric estimates cast a shadow on all the other results. Figure 6 raises also important questions. Usually L-moment estimates are less bias for small sample size than MLE. However, in this figure it seems to be the opposite. How can this be? The authors should look at the (huge) literature on the subject. The authors also refer to various R packages without any further details about the methods behind. A more complete and rigorous scientific background need to be provided on these methods. I recommend a major revisions for this paper. The authors need to a comprehensive revision of their paper. The authors need to clarify the whole methodology (please get to the point) and convince the reader that their development is free of any problems or bugs. They also need to look at C-C scaling using dew point temperature series to see if the sampling problem is still apparent there. I think that this should be minimally done to consider a possible publication of this paper.