Comment on hess-2021-336
Anonymous Referee #1

Referee comment on "Modeling seasonal variations of extreme rainfall on different time scales in Germany" by Jana Ulrich et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-336-RC1, 2021

General comments:

The manuscript provides a novel analysis on seasonal variation of IDF curves for Germany. For estimation of the IDF curves well-chosen state of the art approaches are used. The results are interesting; they reveal the seasonality of extremes for different durations and show smaller uncertainties for IDF estimations when monthly models are used. The manuscript is well written. I have only one general point and a four minor comments for improvement (see specific comments).

The general point concerns the small block size of one month for the extreme value analyses. Even if the GEV is fitting well as shown by the authors the basic assumption of sufficient large n for the validity of Fisher–Tippett–Gnedenko theorem might be questionable also considering the fact that in certain months of some years no extremes might be observed. This problem becomes especially relevant given the results from which months the maxima from different durations originate with respect to the effective block size. Especially the latter is much larger than one month for long durations. I would suggest to discuss this a little further, may be also considering future research regarding specific analyses periods/ seasonal sub-divisions depending on the duration.

Specific comments:

- Please indicate the specific GEV parameters on the figures of relate a) – e) to the specific parameters in the caption.
- Lines 233ff: Please add some more explanation how the bootstrap is carried out.
- Equation 16: The equation is not completely clear to me. The variable u is not explicitly defined. However, if I assume u=(\(o_n-q_p\)) then it should read e.g. \(p_u=pu\) and not \(p_u=u\)?
- Lines 425ff: The uncertainties are estimated with different methods, the Fisher information matrix and the bootstrap method. Are these results comparable? Why not using bootstrap for all uncertainty assessments?