Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2021-541-RC1, 2021

Comment on hess-2021-541

Anonymous Referee #1

Referee comment on "Analysis of flash droughts in China using machine learning" by Linqi Zhang et al., Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2021-541-RC1, 2021

This paper studies the predictability of flash drought over China using machine learning methods. The starting point is ERA5 soil moisture over China for the period 1979-2021. They use a definition of flash drought based on changes in soil moisture percentiles (SMP) which they term the rate of intensification (RI) during periods when SMP is decreasing. They define flash droughts as occurring when SMP crosses the 40th percentile and is decreasing at a rate of at least 6.5 percent per week (time step is weekly). There is some confusion in Figure 1 and text surrounding it as to whether crossing of the 20th percentile of SMP is also required (the figure implies this, but text does not). There also is a criterion for a termination time Tn "when the rapid decline of soil moisture ceases", but this is not shown in Figure 1 nor are specifics in the text.

My main problem with this paper is philosophical. Why are you using machine learning at all? It reflects no physical process understanding water – you just throw a bunch of variables that you think could possibly have something to do with RI and turn the crank. Rather obviously, flash droughts are going to occur during dry periods (during precipitating periods, presumably soil moisture increases rather than decreases). So given that it's dry, it must have to do with evaporative demand, and the soil moisture you start with. We do understand those processes (albeit imperfectly), so surely you could use a physically based model to predict the RI. Now, if you did that first, and then applied ML and could somehow (not clear at all to me how) use the ML predictions to diagnose the physically based ones so as to improve them, I would be interested. But I don't really see where the hydrologic content is in this paper.

My other complaint is that key information needed to understand the results is either buried in text or missing altogether. For instance, were flash drought periods extracted from the entire period of record, without regard for season? Ordinarily, one would expect such events to occur primarily in summer, when evaporative demand is the highest. But RI is determined in terms of soil moisture percentage changes, which complicates the picture considerably. In winter, for instance, evaporative demand will be reduced, but the

range of soil moisture percentages likely is also reduced, so it could be that the statistics of RIs are being dominated by events that in a practical sense aren't really droughts at all. I don't know if this is true but constraining the analysis to a window in the summer (if this hasn't already been done – I searched the document and didn't find any indication that it was) would make the most sense.

My suggestion is, the paper needs to go around the track again, and the authors need to include a physically based alternative. If ML provides better predictions, they need a very good explanation for why, and some diagnosis of why the physically based predictions are failing. Ate this point, I don't see that this paper is really about hydrology.