Response to reviewer #1

1. 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.

Response: We thank the reviewer's comment. In the revised manuscript, we added some descriptions in Page 5 Lines 159-160 to clarify two detailed requirements for extracting drought events: "*Specially, the drought events are extracted by following two requirements below: (1) soil moisture falls below the 40th percentile, and (2) there exists the time period when soil moisture is less than the 20th percentile.*". The original symbol T_n which represents the termination time of the onset-development phase was replaced by T_{0+d} in order to keep its consistency between Fig. 1(a) and Fig. 1(b). We adjusted the formula of the intensification rate of drought events and revised some related descriptions as below:

In Page 6 Lines 174-178:

$$RI = \frac{1}{d+1}\sum_{i=0}^{d}\left[\frac{SM(T_{i+1})-SM(T_i)}{T_{i+1}-T_i}\right], \quad T_0 \le T_i \le T_{0+d}, \tag{1}$$

$$s.t = \{\min[SM(T_i)] \le 20th\}, \tag{2}$$

"*Where T_0 is the onset time, T_{0+d} denotes the termination time for the onset-development phase, d is the duration of onset-development phase, $SM(T_i)$ is the soil moisture percentile at time T_i in the rapid intensification process of drought.*"

In Page 6 Lines 168-171:

"*T_{0+d} denotes the termination time for the onset-development stage when the rapid decline of soil moisture ceases…*"

"*T_{0+d} can be determined through a polynomial function and located when the first derivative of the constructed polynomial equals zero in calculus.*"

Besides, we revised Figure 1 in the original manuscript by showing the termination time T_{0+d} clearly. For the specific identification method of the termination time, we added a sentence in Page 6 Lines 171: *"The detailed determination process of T_{0+d} is presented in our previous study (Liu et al. 2020a)."*

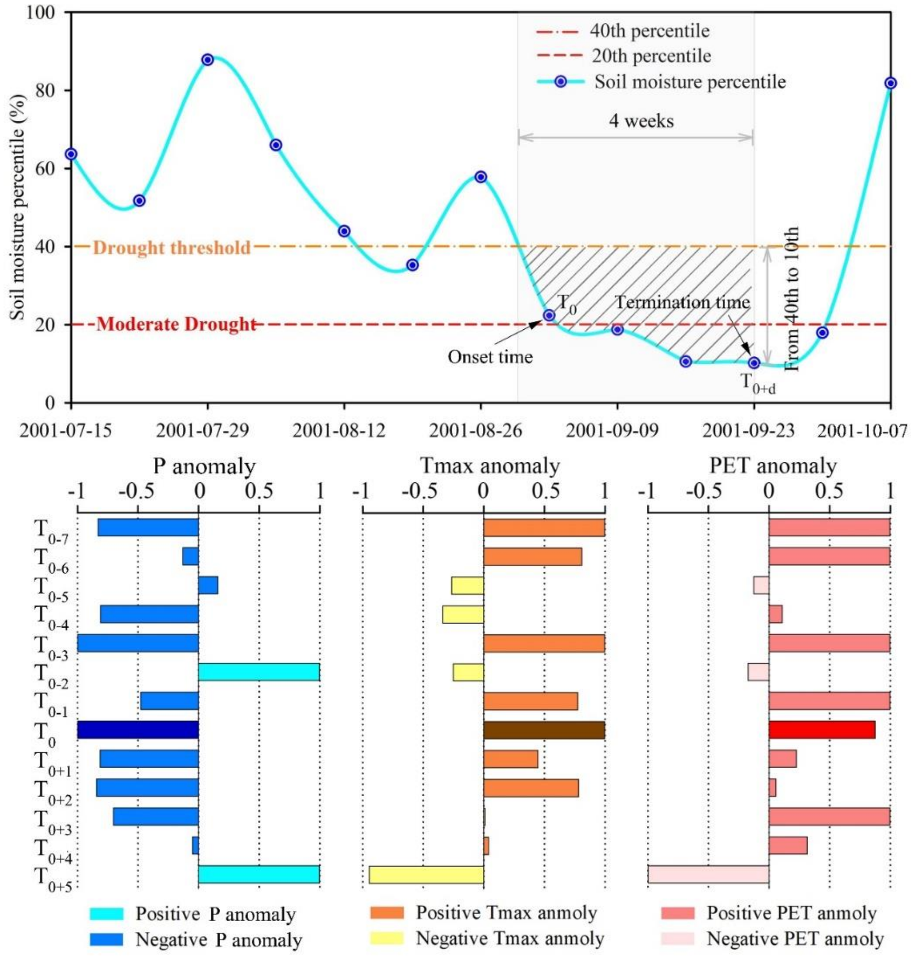

Fig.1 A concept map for identifying flash droughts.

2. 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.

Response:

Thanks for your comments. We agree that a physically-based model is helpful to understand the physical process of flash drought. Flash drought is an emerging and ongoing topic in the drought community during the past ten years. It has a rather complicated evolving process and is influenced by a variety of factors in its different development stages. A variety of studies have analyzed the characteristics of flash drought at global and regional scales, though there is no consistent notion on how we define flash droughts in the community (e.g., Ford et al. 2015; Otkin et al. 2018; Liu et al., 2020). Especially, the physically driven mechanism of flash drought is still uncertain. Given the current progresses of the flash drought field, it is difficult to predict RI using physical-based models. The three machine learning (ML) technologies (i.e., MLR, RF, and LSTM) were used to establish the relationship between RI and predictors, which can be served as references for recognizing and understanding the physical formation and development features of flash droughts.

We agree that it is difficult to use ML technologies to reflect the physical process of flash drought. However, these ML technologies have advantages in providing a fast and direct mapping pathway between the independent and dependent variables based on a combination of abundant data and advanced model architectures (Feng et al., 2021; Yang et al., 2020). Also, they are able to provide an accurate estimation of soil moisture, though the input samples are limited (Long et al., 2019). Given this, we considered using the ML models to evaluate the feasibility of flash droughts simulation over China. Our current study is to figure out the statistical interaction between RI and the anomalies of meteorological factors, which is beneficial for understanding the physical mechanism of flash droughts.

Besides, we appreciate the reviewer for providing a good idea that we can diagnose the physical-based models using machine learning technologies so as to improve the performance of the former models. This might be a study direction in the future flash drought field. However, considering the reason for the unclear physical mechanism of flash droughts, we applied the ML models to analyze the potential relationship between the RI and anomalies of climate factors in the current work. This is the first step before we effectively construct physical-based models to simulate and predict the RI in the future. The reason for using machine learning algorithms has been supplemented in the introduction section in Page 4 Lines 107-113:

"*These ML technologies have some advantages in providing a fast and direct mapping pathway between the independent and dependent variables based on a combination of abundant data and advanced model architectures (Feng et al., 2021; Yang et al.,* *2020). Also, they are able to provide an accurate estimation of soil moisture, though the input samples are limited (Long et al., 2019). Given the inconsistent definition of flash drought and its uncertain driving mechanism, we applied the ML models to simulate flash droughts, which is beneficial for understanding their physical mechanism in the future. Meanwhile, limited studies focused on flash droughts based on ML technologies.*"

References:

Feng, Z., Niu, W., Tang, Z., Xu, Y., Zhang, H.: Evolutionary artificial intelligence model via cooperation search algorithm and extreme learning machine for multiple scales nonstationary hydrological time series prediction, J. Hydrol., 595, 126062, https://doi.org/10.1016/j.jhydrol.2021.126062, 2021.

Ford, T. W., McRoberts, D. B., Quiring, S. M., Hall, R. E.: On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA, Geophys. Res. Lett., 42:9790–9798, https://doi.org/10.1002/2015GL066600, 2015.

Liu, Y., Zhu, Y., Ren, L., Otkin, J., Jiang S.: Two Different Methods for Flash Drought Identification: Comparison of Their Strengths and Limitations, J. Hydrometeorol., 21: 691-704, https://doi.org/10.1175/JHM-D-19-0088.1, 2020.

Long, D., Bai, L., Yan, L., Zhang, C., Shi, C., Yang, W., Lei, H., Quan, J., Meng, X., Shi, C.: Generation of spatially complete and daily continuous surface soil moisture of high spatial resolution, Remote Sens. Environ., 233, 10.1016/j.rse.2019.111364, 2019.

Otkin, J. A., Svoboda, M., Hunt, E. D., Ford, T. W., Anderson, M. C., Hain, C., Basara, J. B.: Flash Droughts: A review and assessment of the challenges imposed by rapid onset droughts in the United States, Bull. Am. Meteorol. Soc., 99(5), 911-919, https://doi.org/10.1175/BAMS-D-17-0149.1, 2018.

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Zhao, B.: A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data, J. Hydrol., 590(1): 125206, http://doi.org/10.1016/j.jhydrol.2020.125206, 2020.

3. 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.

Response:

Thanks for pointing out that. Yes, we extracted flash drought from the entire period of record. When we designed the manuscript, we also first focused on flash droughts in the summertime as the reviewer suggested. We analyzed the identification results carefully and found flash droughts are prone to occur during the cross seasons (e.g., spring-summer or the summer-autumn). Constraining the analysis to a window in the summer may miss the continuous development process of flash drought. Given the above considerations, we preferred to analyze flash droughts by using the entire period. The main reasons are listed below: Firstly, our method relies on continuous time series of soil moisture percentile. The intermittent data makes it hard to capture the onset, or termination of drought events accurately, and the continuity and integrity of the datasets are important for identifying the development process of drought. Secondly, some important information related to flash droughts might be ignored if we merely focus on them in the summer. Previous studies showed that flash droughts may coexist with the seasonal drought and cross-seasonal drought due to the diverse climatic conditions and underlying surface (i.e., the soil texture and vegetation cover) of China (Liu et al., 2020). Meanwhile, according to their study, cross-season drought events easily started from spring (April and May) and summer (June and July). We also analyze the frequency of flash drought occurrence (FOC, Mo et al., 2016) in different seasons, as shown in Fig. 2. According to our identification results, the frequency of flash drought occurrence in winter is the lowest, and for 84% of the study area, the FOC is no more than 5% (Fig. 2a). This low value may have tiny influences on the simulation results. Based on the above analysis, we were more inclined to use the entire period for RI simulation and prediction of flash droughts. We considered your suggestion carefully and supplemented some expressions in Page 6 Lines 179-186 in the revised manuscript to explain the reason why we focused on the entire period.

"We extracted flash drought from the entire period of record, the main reasons are listed: Firstly, our method relies on continuous time series of soil moisture percentile. The intermittent data makes it hard to capture the onset, or termination of drought

events accurately, and the continuity and integrity of the datasets are important for identifying the development process of drought. Secondly, enough important information related to flash droughts need be included in the ML models because flash droughts may coexist with the seasonal drought and cross-seasonal drought due to the diverse climatic conditions and underlying surface (i.e., the soil texture and vegetation cover) of China (Liu et al., 2020). Thirdly, the lower occurrence of flash drought in winter may have tiny influences on the simulation results.''

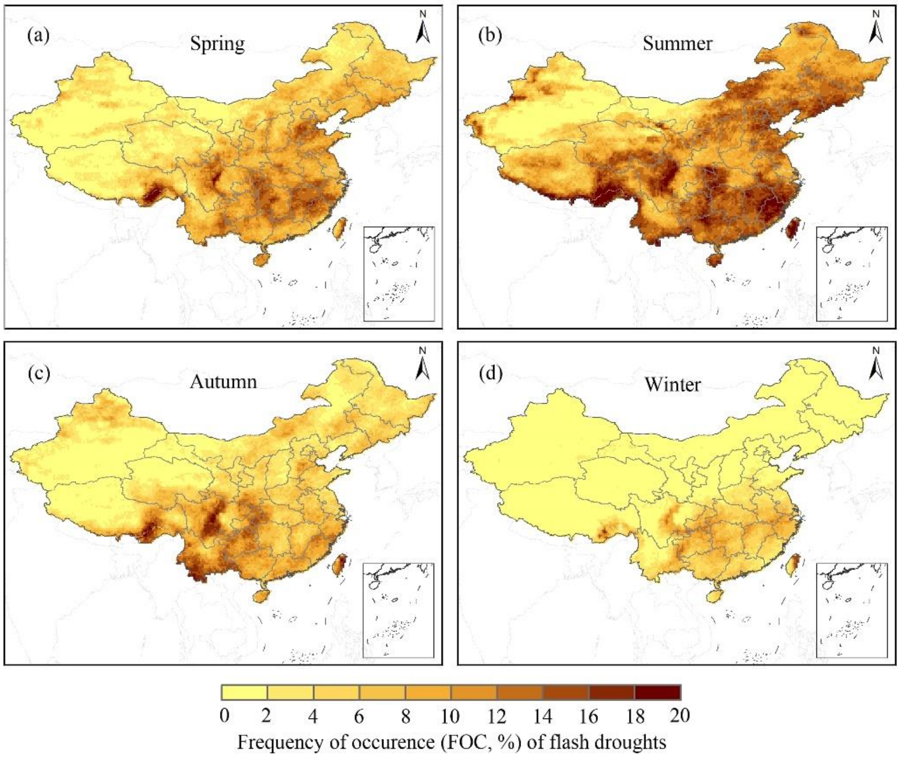

Fig.2 Spatial distribution of frequency of occurrence of flash droughts in different seasons over China

References:

Mo, K. C., Lettenmaier, D, P.: Precipitation deficit flash droughts over the United States, J. Hydrometeorol., 17(4): 1169-1184, https://doi.org/10.1175/JHM-D-15-0158.1, 2016.

Liu, Y., Zhu, Y., Zhang, L., Ren, L., Yuan, F., Yang, X., Jiang, S.: Flash droughts characterization over China: From a perspective of the rapid intensification rate, Sci. Total Environ., 704,135373, https://doi.org/10.1016/j.scitotenv.2019.135373, 2020b.