Dear editor:

On behalf of all the contributing authors, I would like to express our sincere appreciations of your letter and the constructive comments from Referee #1 concerning our article entitled "Flash drought characteristics based on three identification methods in the North China Plain, China". All the comments are very helpful for revising our paper. We have studied and discussed all the comments point-by-point carefully, and accordingly made substantial revisions to our paper. All the changes we have made were in the red-colored text. If the response to the latter comments has already been mentioned in the previous response, it is provided in the pink text without the detailed explanation. Our point-by-point responses to all the comments are provided below in the blue-colored texts.

****

############ Major Comments

1. Uncertainty in Data Sources:

The study implemented reanalysis ET, PET, and SM data. Due to the uncertainty inherent in these datasets, it is challenging to confirm whether the proposed MESR methodology accurately captures flash drought events. While the authors evaluated the methodology using three historical drought events (1981, 1983, and 1989), the nature of these events (flash or conventional drought) remains unclear, and their characteristics are not provided. Comparing MESR performance with recent, well-documented flash drought events would strengthen the reliability of the findings.

Response: Thank you for the comments. The two typical FD events occurring in 2017 and 2019 are utilized to evaluate the applicability of three FD identification methods in Section 3.4 (see Section 3.4 in lines 481-502). The FD events in 2017 and 2019 are identified by soil moistures, and both are well recorded with the detailed development and evolution records in Xue (2023) and Yao et al. (2022), further confirmed in Chen et al. (2024) as well. Grids suffering FD_{RZSM}, FD_{SESR}, and FD_{MESR} are identified in the revised manuscript and compared with the development and evolution records of the two real FD events, demonstrating good consistency.

"3.4 Typical historical events

To demonstrate the applicability of three FD identification methods, an evaluation is conducted based on two typical drought events occurring in 2017 and 2019 (Chen et al., 2024; Xue, 2023; Yao et al., 2022). Xue (2023) identified the FD events in the NCP between 1978 and 2020 using the soil moisture. It was found that the FD_{RZSM} *event began in late July 2017, terminated in the mid-August, and became serious in early August. Figure 9 (a) shows the spatial evolution of* FD_{RZSM}, FD_{SESR}, *and* FD_{MESR} *from July to August 2017.* FD_{SESR} *and* FD_{MESR} *started in the southwestern NCP on July 5th, and alleviated on August 4th. After then, there were only sporadic FDs. A* FD_{RZSM} *event occurred on July 15th and eased until August 9. After that, the affected area rapidly shrank and ended on August 29th. In late July, the affected area of* FD_{RZSM}, FD_{SESR}, *and* FD_{MESR} *were all large, indicating a serious FD. Therefore, it can be considered that the* FD_{RZSM}, FD_{SESR}, *and* FD_{MESR} *in 2017 in this study agree with the findings of Xue (2023).*

Furthermore, the FD_{SESR} and FD_{MESR} started and developed before FD_{RZSM}, indicating that they somewhat spread towards FD_{RZSM}.

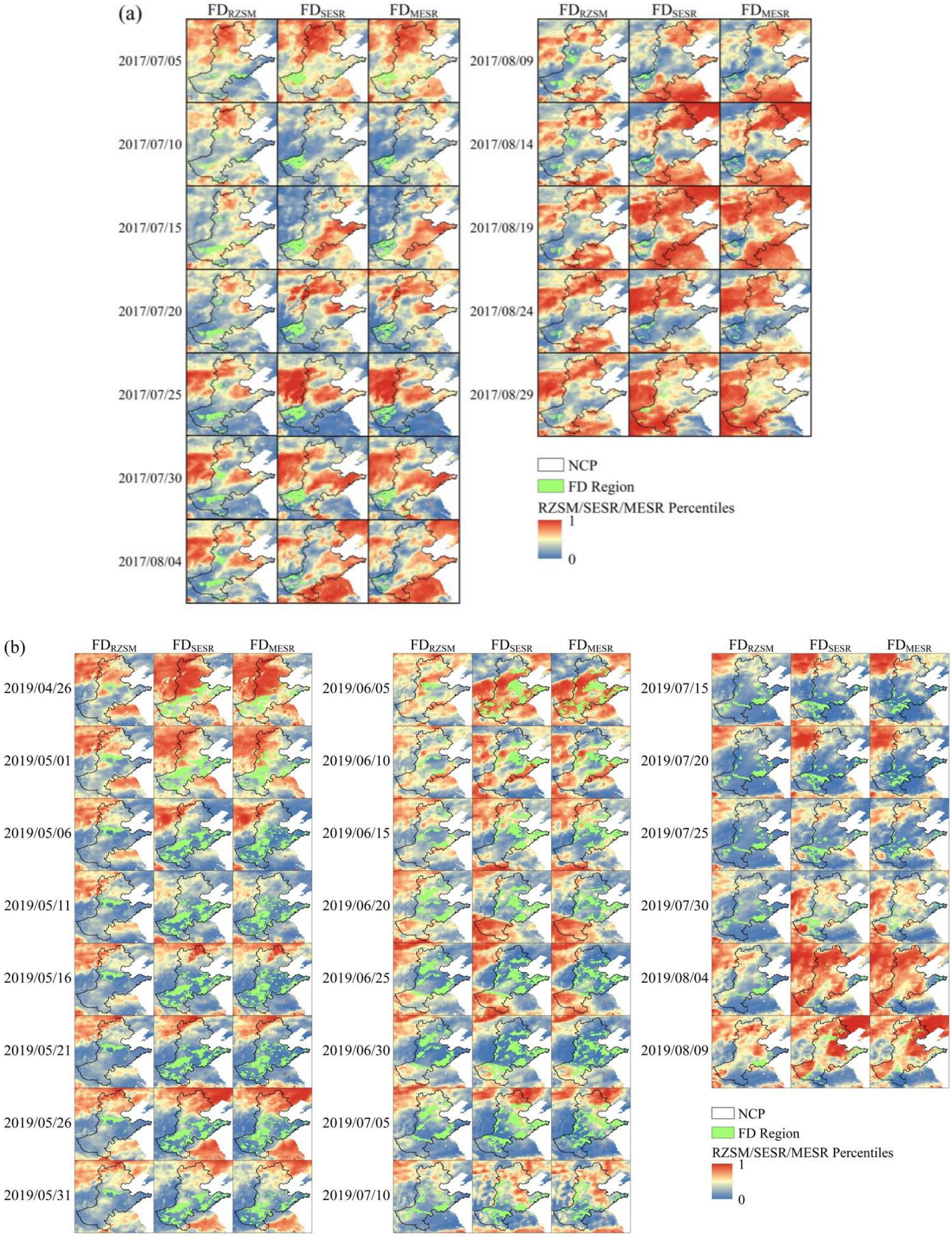

Figure 9 The spatiotemporal evolution process of FD events in (a) 2017 and (b) 2019.

Yao et al. (2022) discovered that FD_{RZSM} in 2019 rapidly developed from April 30th to June 9th, during which the RZSM percentiles decreased sharply from 86% to 25%. Afterwards, the RZSM percentile decreased once again, and the FD_{RZSM} severity peaked in July and recovered in August. However, Fig.9 (b) shows that the FD_{RZSM} started on April 26th and recovered for a short time on June 5th, but it worsened again since June

15th with the largest affected area from late June to early July, then gradually recovered and terminated in August. FD_{SESR} and FD_{MESR} exhibited a similar evolution as FD_{RZSM}. It began on April 26th, recovered from May 31st, then continued to develop on June 15th, eased on July 10th, and ended on July 30th. The evolution is comparable to that from Yao et al. (2022). Therefore, the FD identification by RZSM, SESR, and MESR in this study might be in good agreement with the actual FD events, and the findings are trustworthy."

2. Spatial Heterogeneity and Climate Regimes:
The study area is semi-humid, and identifying flash droughts requires consideration of background aridity and land cover impacts. One concern is the spatial heterogeneity in FD frequency detected by MESR, with significant differences between adjacent pixels scattered across the area, and such patterns are not evident in the other two methods. Evaluating MESR's performance in different climate regimes, such as semi-arid or sub-humid regions, using the Aridity Index (AI), would improve the robustness and generalizability of this research.

Response: Thank you for the comments. The spatial heterogeneity of the FD characteristics in the original manuscript has been analyzed, combined with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET in the revised manuscript (see lines 321-334, lines 361-364, lines 381-386, lines 398-405).

Furthermore, the performance of MESR has been evaluated in Section 4.3 in the revised manuscript (see lines 616-647). The coefficients of determination (R^2) are used to measure the capacity to explain the variance in the dependent variable by the linear regression between the independent and dependent variables. Considering the certain relationship between FD_{RZSM}, FD_{SESR}, and FD_{MESR}, the relationships between FD_{RZSM}, FD_{SESR}, and FD_{MESR} have been explored by the R^2. In Section 4.3, "RZSM ~ SESR", which represents the relationship between FD_{RZSM} and FD_{SESR}, is determined by the R^2 via the linear regression between the RZSM percentile (dependent variable) and SESR percentile (independent variable) of FD_{RZSM} pentads. Meanwhile, the "RZSM ~ MESR", "SESR ~ RZSM", "MESR ~ RZSM", "SESR ~ MESR", and "MESR ~ SESR" are determined as well. Besides, the impact of AI on the spatial distributions of R^2 has also been explored.

"The frequency of FD_{RZSM} is high in the central and northeastern NCP and low in the southern NCP with two high-frequency regions in the central and northeastern NCP. The AI values of central and northeastern NCP are 0.2 ~ 0.3 and less than 0.2, respectively (Fig.S1 (b)). Therefore, they are comparatively dry. The lack of precipitation and an increase in evapotranspiration would fasten the decline in soil moisture and increase the probability of FD in the central NCP than the southern NCP (Gou et al., 2022; Yuan et al., 2023). Besides, the northeastern NCP is woodland with high ET (Guo et al., 2007). Adding the high-water demand, it occurs frequent FD_{RZSM}. The frequency of FD_{SESR} and FD_{MESR} are both high in the north-central NCP and low in the northeastern and southern NCP and are opposite to the ratio of annual ET and PET (Fig.S1 (c)), indicating that a region with greater evaporative stress would encounter more FD_{SESR} and FD_{MESR}. North-central NCP is cultivated land. The

evaporative stress of north-central NCP fluctuates significantly due to the influence of irrigation, which causes frequent FD_{SESR} and FD_{MESR}. In contrast, the woodland in the northeastern NCP is usually not irrigated artificially, and the evaporative stress is mostly influenced by climate conditions. Therefore, northeastern NCP has less evaporative stress fluctuations and low FD_{SESR} and FD_{MESR} frequency (Guo et al., 2007). Southern NCP has higher temperature, greater evapotranspiration, and more abundant precipitation as latitude decreases. The balanced hydrothermal conditions lead to a greater AI and lower FD frequency in the southern NCP."

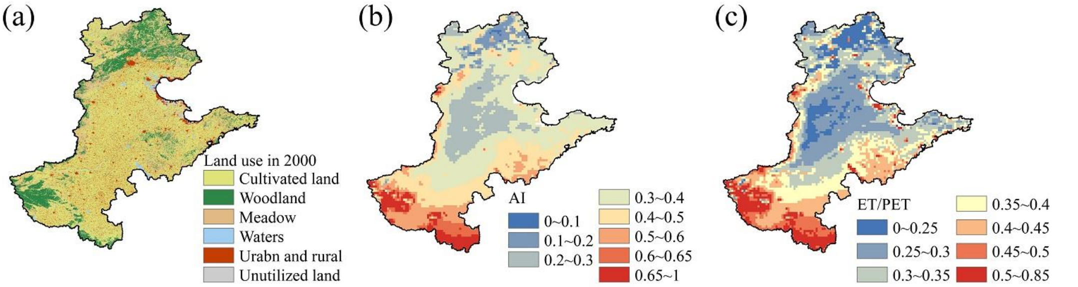

Figure S1 Spatial distribution of (a) the land use in 2010, (b) AI, and (c) the ratio of annual ET and PET in NCP.

"Warmer temperature may result in longer FD duration (Zhang et al., 2022c). Woodlands take longer to recover from drought than the cultivated lands (Wu et al., 2024). Additionally, human activities might also have an impact on the FD $duration_{Total}$. Irrigation might significantly alleviate FD in cultivated land, but the woodland in the northeastern NCP is less impacted by human activity and might have a longer FD $duration_{Total}$."

"Warming not only lengthens the drought durations of the southern NCP, but also exacerbates them by increasing surface evapotranspiration losses and decreasing the soil moistures (Yuan et al., 2019; Zhang et al., 2021). The long duration might also result in the great severity. Whereas FD_{MESR} has high severity in the northern and central NCP and low severity in the southern NCP. The spatial distribution of the FD_{MESR} severity is opposite to the ratio of annual ET and PET, which is lower in the northern NCP but higher in the southern (Fig.S1 (c)). It illustrated that the $duration_{Total}$ and severity of FD_{MESR} have a stronger correlation with the evaporative stress than that of FD_{SESR}."

"For FD_{RZSM}, RZSM percentiles decrease slower in west-central NCP than in other regions. Even though the west-central NCP is in the arid state with a low AI of 0.2 ~ 0.4, it might be because the west-central NCP is cultivated land and irrigation has a significant impact on the soil moisture, which might effectively alleviate the decline in RZSM. Although the AI in northern NCP is less than 0.4 as well, the woodland in this region is less impacted by human activities like irrigation, which causes RZSM to rapidly decrease. Additionally, the high temperature and great ET in the southern NCP hasten the RZSM decline rate. For FD_{SESR} and FD_{MESR}, even though southern NCP has great ET, the PET is great as well, which might not lead to a low ESR and high evaporative stress (Fig.S1). Abundant precipitation and low evaporative stress ease the declining rate of SESR and MESR."

*"**4.3 Explanatory ability between different FD types***

Given the impact of climate control on the FD occurrence (Mukherjee and Mishra, 2022), there might be a certain relationship between different FD types. Therefore, the relationships among FD_{RZSM}, FD_{SESR}, and FD_{MESR} are explored by the coefficient of determination (R^2) which stands for the capacity to explain the variance in the dependent variable by the linear regression between the independent and dependent variables (Mukherjee and Mishra, 2022). In particular, the relationship between FD_{RZSM} and FD_{SESR}, which is referred to as "RZSM ~ SESR", is represented by the R^2 determined via the linear regression between the RZSM percentile (dependent variable) and SESR percentile (independent variable) of FD_{RZSM} pentads. Meanwhile, the "RZSM ~ MESR", "SESR ~ RZSM", "MESR ~ RZSM", "SESR ~ MESR", and "MESR ~ SESR" are determined, as shown in the first two columns in Fig.S15.

In Fig.S15 (a) and (b), both "RZSM ~ SESR" and "RZSM ~ MESR" explain more than 40% of the variance in RZSM percentile in the central NCP, but less than 30% in the other regions. In Fig.S15 (e) and (f), the southern NCP has more "SESR ~ RZSM" explaining the variance in SESR and more "MESR ~ RZSM" explaining the variance in MESR percentile (mostly about 15% ~ 25%) than the northern (less than 15%). Whereas "SESR ~ MESR" explains over 90% of the variance in SESR, as well as "MESR ~ SESR" explains the variance in MESR, as shown in Fig.S15 (i) and (j). Overall, the explanatory ability from high to low is: "SESR ~ MESR" and "MESR ~ SESR" > "RZSM ~ SESR" and "RZSM ~ MESR" > "SESR ~ RZSM" and "MESR ~ RZSM". SESR and MESR are all based on the linear transformation of ESR, which makes them good in explaining each other with a high R^2 of over 90%. The relationship between MESR and RZSM ("RZSM ~ MESR" and "MESR ~ RZSM") is quite comparable to that between SESR and RZSM ("RZSM ~ SESR" and "SESR ~ RZSM"), highlighting the reliability of FD identification based on MESR. The differences in Fig.S15 (c), (g), and (k) further demonstrate the similarities between SESR and MESR as well.

The spatial distributions of R^2 also point to the sensitivity to the AI, as shown in Fig.S15 (d) (h) and (l). For "RZSM ~ SESR" and "RZSM ~ MESR", the variance explanatory ability increases with the increasing AI in the region with AI below 0.3, but decreases in the region with AI above 0.3. In the region where AI is between 0.2 and 0.3, they explain the greatest variance in the RZSM percentile (about 60%). Overall, the RZSM percentiles could be explained more by the SESR and MESR percentiles in the dryer region with the exception of the region with AI less than 0.2, which might be related to that RZSM is greater initially in the wet region with extended memory (Mukherjee and Mishra, 2022). For "SESR ~ RZSM" and "MESR ~ RZSM", the variance explanatory ability is less than 20%, and it is obviously greater in the region with AI more than 0.2 than in the region with AI less than 0.2. SESR and MESR percentiles could be better explained by the RZSM percentile in the wetter region with less evaporative stress and higher evaporation. The FD_{SESR} and FD_{MESR} belonging to meteorological drought might lead to the FD_{RZSM} belonging to agricultural drought, making that R^2 of "SESR ~ RZSM" and "MESR ~ RZSM" is lower than that of "RZSM ~ SESR" and "RZSM ~ MESR". For "SESR ~ MESR" and "MESR ~ SESR", the explanatory ability exceeding 90% increases with the increasing AI overall."

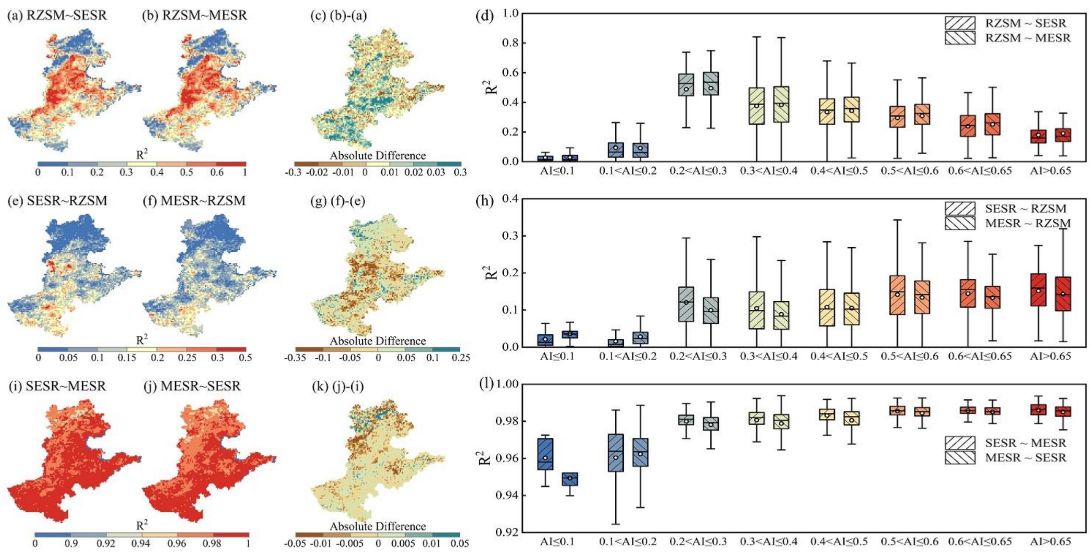

Figure S15 Spatial distribution of the R^2 determined by (a) "RZSM ~ SESR", (b) "RZSM ~ MESR", (e) "SESR ~ RZSM", (f) "MESR ~ RZSM", (i) "SESR ~ MESR", and (j) "MESR ~ SESR", as well as the differences of R^2 (c) between "RZSM ~ SESR" and "RZSM ~ MESR", (g) between "SESR ~ RZSM" and "MESR ~ RZSM", and (k) between "SESR ~ MESR" and "MESR ~ SESR". The boxplots in the (d), (h), and (l) illustrate the R^2 in the "RZSM ~ SESR" and "RZSM ~ MESR", "SESR ~ RZSM" and "MESR ~ RZSM", and "SESR ~ MESR" and "MESR ~ SESR" over different AI values.

3. Justification of Methodology:

The paper lacks a clear explanation for multiplying ESR values by their mean (climatological or long-term) to create MESR. This difference appears to be a primary factor distinguishing MESR from SESR in terms of frequency. The authors should further clarify and justify this decision in the main text for better understanding and transparency.

Response: Thank you for the comments. FD_{MESR} is proposed based on the limitations of FD_{SESR} in lines 84-92. The first one is that whether ESR follows a normal distribution requires further determination, which makes the rationality of normalizing ESR into SESR is still up for debate. The second one is that the $SESR_{40th}$ and $\Delta SESR_{40th}$ might be greater than 0. Then we have added the detailed reasons for using MESR instead of SESR to identify FD in lines 198-209. There are three main differences between FD_{MESR} and FD_{SESR} identification. The first one is that the ESR value is divided by its mean to construct the MESR series instead of normalizing the ESR to create SESR. The second one is that the percentiles of MESR and ΔMESR for each pentad are fitted using the optional PDF instead of utilizing EDF to convert into percentiles. And the third one is the variable thresholds of MESR and ΔMESR are employed in the process of FD identification.

"However, the SESR application has some problems. When identifying FD using SESR, both SESR and the change of SESR (ΔSESR) theoretically follow normal distributions, therefore the 50th percentile ($SESR_{50th}$ and $\Delta SESR_{50th}$) equal 0, indicating that

$\Delta SESR_{50th}$ denotes no change in SESR ($\Delta SESR = 0$), whereas ΔSESR below the 40th percentile of ΔSESR values ($\Delta SESR_{40th}$) denotes a declining in SESR. However, whether SESR and ΔSESR follow a normal distribution remains to be determined, and if not, the $\Delta SESR_{40th}$ may not be less than 0. In the study by Gou et al. (2022), it was found that the 36th percentile of ΔSESR ($\Delta SESR_{36th}$) corresponds to an increase in SESR, where $\Delta SESR_{36th}$ is greater than 0. This phenomenon may be because that SESR gradually decreases during periods without precipitation, and increases during the precipitation process. It is conceivable for ΔSESR to be less than 0 during periods with no precipitation and larger than 0 during precipitation periods. Consequently, the FD events may be underestimated or overestimated."

"There are three main differences between MESR and SESR identification. Firstly, the ESR value is divided by its mean to construct the MESR series instead of normalizing the ESR to create SESR. Regardless of whether ESR is standardized as SESR or MESR, their percentile values are unaffected by the linear transformations based on ESR. However, whether ESR follows a normal distribution requires further determination, which makes the rationality of normalizing ESR into SESR is still up for debate. Whereas, when dividing ESR by its mean and transforming it into MESR, it is not necessary to take into account the PDF that ESR follows. Secondly, the percentiles of MESR and ΔMESR for each pentad are fitted using the optional PDF instead of utilizing EDF to convert into percentiles. Since the distribution function that MESR and ΔMESR follow is yet unknown, several PDFs are fitted in order to select the best PDF, which can produce more precise percentiles. Lastly, the variable thresholds of MESR and ΔMESR are employed in the process of FD identification. Due to the uncertainty surrounding whether SESR and ΔSESR follow a normal distribution, as well as the phenomenon that the $SESR_{40th}$ and $\Delta SESR_{40th}$ might be greater than 0 (as found in Gou et al. (2022)), the variable percentiles of MESR and ΔMESR are used as the threshold for FD identification to make sure the threshold is less than 0."

############Minor Comments

4. Lines 26-27: Replace "becomes" with "become" to align with the plural subject "droughts."

Response: Thank you for the comments. We have revised "becomes" into "become" (see lines 26-27).

"*The terrestrial water cycle accelerates and droughts become more frequent under global warming.*"

5. Line 40: Remove "respectively" for clarity.

Response: Thank you for the comments. The "respectively" has been deleted in the revised manuscript (see lines 38-40).

"Currently, the FD identification methods can be mainly divided into three categories based on conventional drought indicators, soil moisture, and atmospheric evaporation demand."

6. Lines 84-85: SESR is a method for identifying flash droughts, so it is better to use

the word 'using' in this sentence. Here is the revised version:
"However, the SESR application has some problems. When identifying FD using SESR, both SESR and the change in SESR"

Response: Thank you for the comments. We have revised the presentation (see lines 84-87).

"However, the SESR application has some problems. When identifying FD using SESR, both SESR and the change of SESR (ΔSESR) theoretically follow normal distributions, therefore the 50th percentile ($SESR_{50th}$ and $\Delta SESR_{50th}$) equal 0, indicating that $\Delta SESR_{50th}$ denotes no change in SESR (ΔSESR = 0), whereas ΔSESR below the 40th percentile of ΔSESR values ($\Delta SESR_{40th}$) denotes a declining in SESR."

7. Line 96: It would be better to rephrase this sentence for better clarification, and start with 'In this study, ... ' Here is the revised version:
In this study, a new method based on the ......

Response: Thank you for the comments. We have revised the presentation (see lines 96-97).

"In this study, a new method based on the multiples of the mean evaporative stress ratio (MESR) for FD identification has been developed to address the aforementioned problems in SESR identification method."

8. Line 111: One of the most important factors in characterizing droughts is considering background aridity. One of my concerns regarding this paper is that the study area is mainly a semi-humid region. The baselines of SM percentile or SESR vary across different climate regimes.

Response: Thank you for the comments. The thresholds in the FD identification on one grid are the specific percentiles (like 40^{th} or 20^{th} percentiles) of RZSM, SESR, or MESR on the specific grid for the specific pentad throughout the entire study period. Although the percentiles of the FD thresholds are fixed for all grids ($RZSM_{40th}$, $RZSM_{20th}$, $SESR_{40th}$, $SESR_{20th}$, $\Delta SESR_{40th}$, $MESR_{Pr1-10th}$, $MESR_{Pr1-30th}$, and $\Delta MESR_{Pr2-10th}$), the determination of the percentiles is based on the RZSM, SESR, and MESR series on the specific grid for the specific pentad during the whole study period. It can be considered that the FD thresholds are temporally and spatially influenced. Besides, SESR and MESR are both obtained through the linear transformation of ESR, which does not change their percentiles. The 40^{th} and 20^{th} percentiles of RZSM and SESR, as well as the $(Pr1\text{-}10)^{th}$ and $(Pr1\text{-}30)^{th}$ percentiles of MESR, indicate a low level of soil moisture and evaporative stress value. Meanwhile, the $\Delta SESR_{40th}$ and the $\Delta MESR_{(Pr2-30)th}$ indicate a significant decreasing in the SESR or MESR. Therefore, the thresholds for FD identification in different climate regions do not affect the FD identification results, and these thresholds could be widely applied in various climate regions.

9. Lines 120-127: Soil moisture, ET, and PET datasets used in this study are reanalysis and there is uncertainty in these data sets. Additionally, different reanalysis datasets have great differences in their values, so this study can benefit from using different SM, ET, and PET datasets. GLEAM, or MERRA2?

Response: Thank you for the comments. The soil moisture, ET, and PET obtained from the GLEAM and GLDAS 2 datasets are detailed in lines 133-145. Based on the GLEAM and GLDAS 2 datasets, the uncertainties of the different reanalysis datasets have been evaluated in Section 4.1 of the revised manuscript (see lines 552-574). We have analyzed the bias of the pentad RZSM, SESR, and MESR percentiles from 1981 to 2022 between ERA5-Land and GLEAM and GLDAS 2 datasets by Taylor diagrams in Figure S11. The spatial distribution of the pentad RZSM, SESR, and MESR percentile correlations between ERA5-Land and GLEAM and GLDAS 2 datasets in Figure S12 have also shown great similarities. The FD characteristics identified based on ERA5-Land, GLEAM, and GLDAS 2 datasets have also been compared in Figures S5, S13 and S14. The similarity of pentad RZSM, SESR, and MESR percentiles from various datasets, as well as the FD characteristics based on various datasets, effectively demonstrates the reliability of findings.

"Two additional reanalysis datasets, fourth version of the Global Land and Evaporation Amsterdam Model (GLEAM v4.1a; Miralles et al., 2011) and Global Land Data Assimilation System version 2 (GLDAS 2; Beaudoing &Rui, 2019; Beaudoing &Rui, 2020) were introduced to evaluate the uncertainty of FD identified by the ERA5-Land dataset. The surface (0 ~ 10 cm) soil moisture (unit: m^3 m^{-3}), root-zone (0 ~ 100 cm) soil moisture (RZSM, unit: m^3 m^{-3}), actual evaporation (E, unit: mm day^{-1}), and potential evaporation (PET, unit: mm day^{-1}) provided by GLEAM v4.1a dataset are on a 0.1°×0.1° latitude-longitude grid and with a daily temporal resolution (Xue and Wu, 2024). Evapotranspiration (ET, unit: kg m^{-2} s^{-1}), potential evaporation rate (unit: W m^{-2}), and soil moistures from 0 ~ 10 cm, 10 ~ 40 cm, and 40 ~ 100 cm (unit: kg m^{-2}) derived from the GLDAS 2 dataset are obtained at 0.25°×0.25° spatial resolution and 3-hourly temporal resolution. The potential evaporation rate was transformed into the potential evaporation (PET) by calculating the accumulation over time. Similar to the ERA5-Land dataset, the soil moistures of layers 0 ~ 40 cm and 0 ~ 100 cm in GLDAS 2 were determined by the weighted average. Because the GLDAS 2.0 dataset spans between 1948 and 2014, while the GLDAS 2.1 dataset is from 2000, the GLDAS 2.0 dataset from 1981 to 1999 and the GLDAS 2.1 dataset from 2000 to 2022 were utilized in accordance with the method of Wang and Yuan (2021). The findings in this study are based on the ERA5-Land dataset unless otherwise noted."

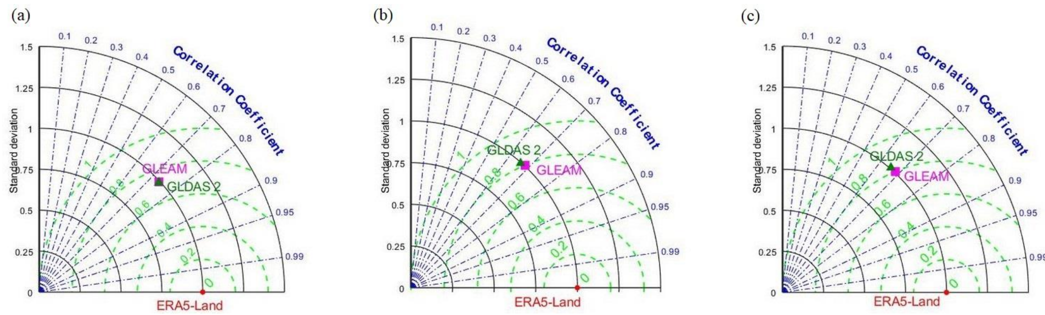

Figure S11 Taylor diagram for the pentad (a) RZSM, (b) SESR, and (c) MESR percentiles based on ERA5-Land, GLEAM, and GLDAS 2 datasets.

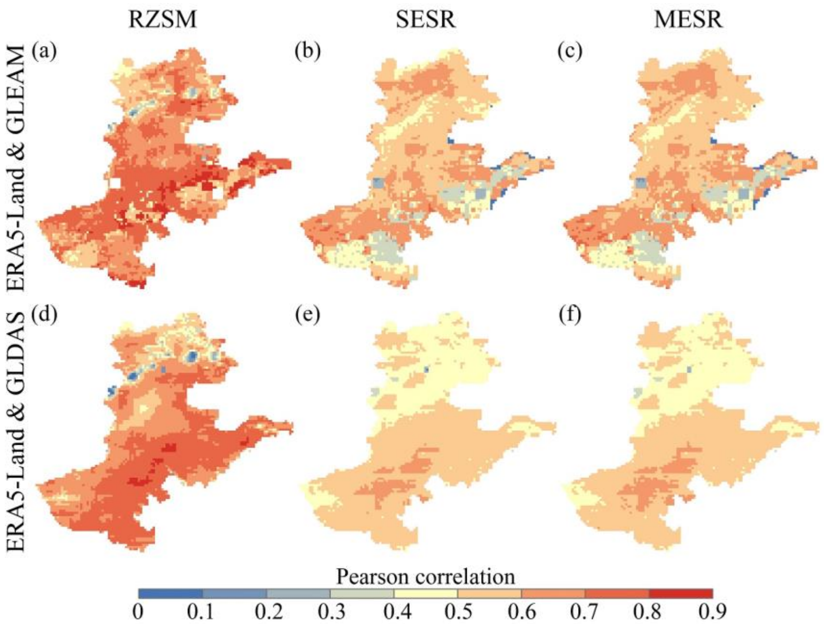

Figure S12 Spatial distribution of the Pearson correlation of the pentad RZSM, SESR, and MESR percentiles between ERA5-Land and (a) ~ (c) GLEAM and (d) ~ (f) GLDAS 2 datasets.

"4.1 Uncertainties from the reanalysis datasets

To explore the data-related uncertainties, the soil moisture, ET, and PET obtained from two additional reanalysis datasets, GLEAM and GLDAS 2 datasets, are utilized to identify the FD_{RZSM}, FD_{SESR}, and FD_{MESR}. Due to that the RZSM, SESR, and MESR are the basis for FD identification, the pentad percentiles of RZSM, SESR, and MESR are determined. Figure S11 illustrates the Taylor diagrams of the pentad RZSM, SESR, and MESR percentile series from 1981 to 2022 for ERA5-Land, GLEAM, and GLDAS 2 datasets, with the ERA5-Land dataset as the observation. The points representing the pentad SESR and MESR percentiles from GLEAM and GLDAS 2 datasets are all very close. The correlation coefficients between pentad RZSM percentile and SESR or MESR percentiles are around 0.7, the centered root mean square differences are around 0.8, and the standard deviations are approximately 1. Therefore, the pentad percentiles of RZSM, SESR, and MESR of the ERA5-Land dataset are similar to those of GLEAM and GLDAS 2. Figure S12 displays the spatial distribution of the pentad RZSM, SESR, and MESR percentile correlations between ERA5-Land and GLEAM and GLDAS 2 datasets. As shown in Fig.S12 (a) and (d), the RZSM percentile correlation between ERA5-Land and GLEAM is comparable to that between ERA5-Land and GLDAS 2, both of which are primarily greater than 0.6. While the SESR percentile correlation between ERA5-Land and GLEAM is similar to the MESR percentile correlation, which both exceeds 0.5 mostly. The correlation of SESR and MESR percentiles between ERA5-Land and GLDAS 2 is mostly between 0.4 and 0.7, with a comparable spatial distribution pattern. The higher correlation between ERA5-Land and GLEAM than between ERA5-Land and GLDAS 2 might be due to the coarse spatial resolution of GLDAS 2.

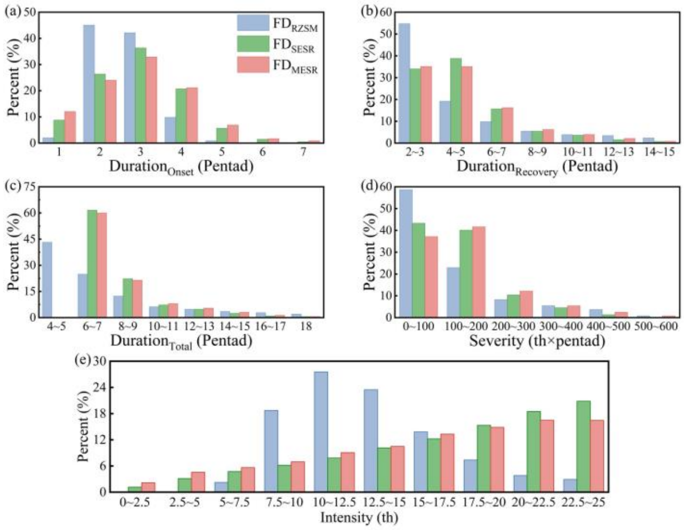

Figure S5 Histogram of FD characteristics identified by RZSM, SESR, and MESR based on ERA5-Land.

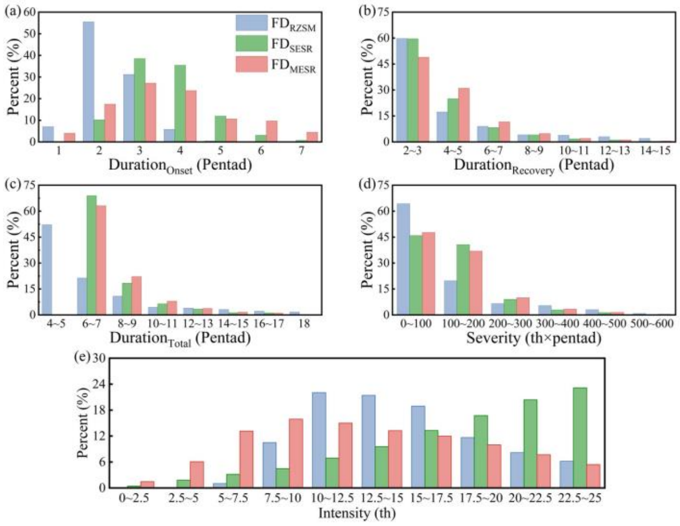

Figure S13 Same as Figure S5, but based on GLEAM.

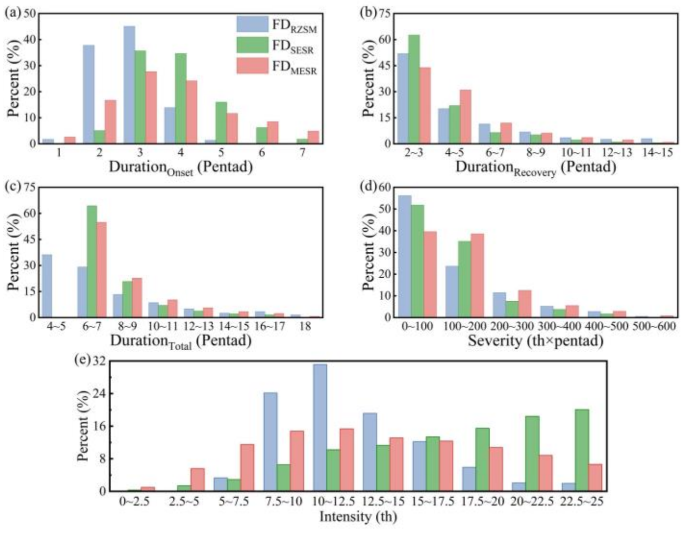

Figure S14 Same as Figure S5, but based on GLDAS 2.

Besides, the FD characteristics identified based on ERA5-Land, GLEAM, and GLDAS

2 datasets are displayed in Figs.S5, S13 and S14. The distributions of the FD_{RZSM}, FD_{SESR}, and FD_{MESR} characteristics based on various datasets are comparable with the exception of the FD_{MESR} intensity based on GLDAS 2. Meanwhile, the proportions of various FD grades determined by intensity from diverse data sources also demonstrate an indisputable resemblance, as seen in Fig.S6. The similarity of pentad RZSM, SESR, and MESR percentiles from various datasets, as well as the FD characteristics based on various datasets, effectively demonstrates the reliability of our findings.

10. Lines 182-183: what is the main reason for dividing ESR value by its mean? Although there are some benefits to doing this, I am concerned that normalizing ESR by the mean could reduce the sensitivity of the method in detecting flash droughts, especially in periods or regions with naturally higher evapotranspiration stress. The underlying issue with normalizing ESR by its mean is that it reduces the relative magnitude of $MESR_0$ values when the baseline ESR is high, effectively making it harder to detect rapid changes. If the authors have any explanation for this, it would be helpful to include it in the main text.

Response: Thank you for the comments. In the process of identifying FD based on SESR and MESR, the first step is the percentile transformation, followed by the FD identification by the FD identification criteria that are related to the percentiles. The thresholds of FD identification based on SESR and MESR are both related to the percentiles of SESR and MESR. In the process of standardizing ESR into SESR or MESR, a linear transformation is conducted on ESR, which would not influence the corresponding percentiles. The percentiles are determined by the PDF that ESR fits. ESR might not follow a normal distribution, making that the rationality of normalizing ESR into SESR needs further discussion. Whereas, standardizing ESR by its mean and converting it into MESR does not involve the problem of whether ESR follows a normal distribution. Therefore, ESR is divided by its mean and standardized as MESR. The reason for using MESR instead of SESR to identify FD has been added in lines 198-209, which has been mentioned in Comment 3. Therefore, it can be considered that the linear transformation methods conducted on ESR would not affect the sensitivity of FD identification, and using the mean ESR for standardization would not affect the FD identification based on the percentiles. The detailed reason that why normalizing ESR by the mean would not reduce the sensitivity of FD identification has also been supplemented in lines 248-252.

"In the process of FD_{SESR} and FD_{MESR} identification, the first step is converting the SESR, $\Delta SESR$, $MESR_0$, and $\Delta MESR_1$ into percentile values, and then identify FD by the criteria related to the percentiles. Since the linear transformation applied to the ESR would not change the percentiles, the linear transformation methods would not have an impact on the percentiles of SESR, $\Delta SESR$, $MESR_0$, and $\Delta MESR_1$ and the thresholds of the FD criteria, as well as the FD identification results. It is said that the FD identification based on the percentiles would be unaffected by using the mean ESR for standardization."

11. Line 184: It is climatological mean or long-term mean?

Response: Thank you for the comments. $\overline{ESR}$ is the climatological mean value of ESR at a specific grid for a specific pentad for all years during the study period (see lines 214-215).

"where $\overline{ESR}$ is the climatological mean value of ESR at a specific grid for a specific pentad for all years during the study period, and $MESR_0$ is the multiple of $\overline{ESR}$ at the specific grid for a specific pentad."

12. Line 184-191: Specify whether the mean is climatological or long-term.

Response: Thank you for the comments. The mean is climatological mean (see lines 222-224).

"where $\Delta MESR_0$ is the change in MESR between adjacent pentads, $\overline{\Delta MESR_0}$ is the climatological mean value of $\Delta MESR_0$ at the specific grid for a specific pentad over the whole study period, and $\Delta MESR_1$ is the multiple of $\overline{\Delta MESR_0}$ for the pentad o all study period."

13. Line 191: Again, climatological mean or long-term mean?

Response: Thank you for the comments. The mean is climatological mean (see lines 222-224 in Comment 12).

14. Line 205: Likely a typo; change FD (SESR) to FD (MESR).

Response: Thank you for the comments. It is a typo. We have revised "FD_{SESR}" into "FD_{MESR}" (see line 237).

"FD_{MESR} can be identified by the following four criteria:"

15. Line 213: This sentence seems to be uncompleted!

Response: Thank you for the comments. We have corrected the original manuscript and removed this sentence (see line 245).

16. Line 242: Clarify that "mean value" refers to a single point, and specify whether it is climatological or long-term.

Response: Thank you for the comments. The $frequency_0$ is the FD frequency from 1981 to 2022 in the NCP on the grid; the $severity_0$ and $intensity_0$ are the average severity and intensity from 1981 to 2022 in the NCP on the grid. The "mean value" is long-term mean, and the detailed explanations for the "mean value" have been supplemented in lines 273-286.

"$$frequency' = (frequency_0 - frequency_{min}) / (frequency_{max} - frequency_{min}) \quad (9)$$

$$severity' = (severity_0 - severity_{min})/(severity_{max} - severity_{min}) \quad (10)$$

$$intensity' = (intensity_0 - intensity_{min})/(intensity_{max} - intensity_{min}) \quad (11)$$

$$\overline{frequency'} = 0.5frequency'_{RZSM} + 0.25frequency'_{SESR} + 0.25frequency'_{MESR} \quad (12)$$

$$\overline{severity'} = 0.5severity'_{RZSM} + 0.25severity'_{SESR} + 0.25severity'_{MESR} \quad (13)$$

$$\overline{intensity'} = 0.5intensity'_{RZSM} + 0.25intensity'_{SESR} + 0.25intensity'_{MESR} \quad (14)$$

$$Hotspots = \left(\overline{frequency'} + \overline{severity'} + 1/\overline{intensity'}\right) / 3 \quad (15)$$

where the $frequency_0$ is the FD frequency from 1981 to 2022 in the NCP on the grid; the $severity_0$ and $intensity_0$ are the average severity and intensity from 1981 to 2022 in the NCP on the grid; the $frequency_{max}$, $frequency_{min}$, $severity_{max}$, $severity_{min}$, $intensity_{max}$, *and $intensity_{min}$ are the maximum and minimum values of the $frequency_0$, $severity_0$, and $intensity_0$ in the NCP; frequency', severity', and intensity' are the max-min normalization of the $frequency_0$, $severity_0$, and $intensity_0$; and the* $\overline{\text{frequency'}}$, $\overline{\text{severity'}}$, *and* $\overline{\text{intensity'}}$ *are the weighted average frequency', severity', and intensity' of FD_{RZSM}, FD_{SESR}, and FD_{MESR}. Due to that the FD_{MESR} method is considered an extension of the FD_{SESR} method, the weights for the characteristics of FD_{RZSM}, FD_{SESR}, and FD_{MESR} are 0.5, 0.25, and 0.25, respectively."*

17. Lines 259-264: Although the discussion in this paragraph is statistically correct, I believe an important point has been overlooked: the most critical aspect of flash droughts is their rapid onset and intensification. As long as the SESR method can detect this characteristic of flash droughts, the exact value of the 50th percentile is less significant, particularly in humid or hyper-humid climate regimes where the ESR baseline might be higher than in arid or semi-arid regions. In regions with lower background aridity, the 50th percentile of SESR might be higher than zero, primarily due to dense vegetation. In such cases, it is crucial to capture the rapid reduction in SM/ESR that leads to flash drought.

Mukherjee and Mishra (2022) demonstrated that using different indicators, such as ESR and soil moisture, to identify flash droughts results in varying frequencies across different climate regimes. This is an important factor to consider, especially in this study.

Mukherjee, S., & Mishra, A. K. (2022). Global Flash Drought Analysis: Uncertainties From Indicators and Datasets. Earth's Future, 10(6), e2022EF002660. https://doi.org/10.1029/2022EF002660

Response: Thank you for the comments. The rapid reduction in RZSM, SESR, or MESR that leads to flash droughts are judged by the percentiles of ΔRZSM, ΔSESR, or ΔMESR. If SESR follows normal distribution, the 50^{th} percentile being equal to 0 represents the median of the SESR series, and the 40^{th} percentile being less than 0 represents a low value in the SESR series. If ΔSESR follows normal distribution, the 50^{th} percentile being equal to 0 represents no change in SESR, and the 40^{th} percentile being less than 0 represents the decreasing SESR. Therefore, the 40^{th} percentile of SESR represents a low ESR, and 40^{th} percentile of ΔSESR represents a decreasing ESR. Figure 3 (b) illustrates that amount of $\Delta SESR_{50th}$ is larger than 0, which might lead to the $\Delta SESR_{40th}$ greater than 0, reflecting an increasing SESR. It represents a rapid decline in SESR theoretically, but in reality, represents the increasing in SESR, which might lead to the misjudgments of FD events. So does the $\Delta SESR_{50th}$ less than 0. Moreover, the SESR and ΔSESR percentiles are both calculated on the grid by pentads, meaning that the $SESR_{50th}$ and $\Delta SESR_{50th}$ on different grids and pentads have different $SESR_{50th}$ and $\Delta SESR_{50th}$ values, and the $SESR_{50th}$ and $\Delta SESR_{50th}$ are temporally and spatially affected (see lines 289-300).

"To graphically illustrate the limitations of the FD_{SESR} method, the distributions of $SESR_{50th}$ and $\Delta SESR_{50th}$ are shown in Fig.3 (a) (b). The SESR and ΔSESR percentiles are both calculated on the grid by pentads, meaning that the $SESR_{50th}$ and $\Delta SESR_{50th}$ on various grids and pentads have various $SESR_{50th}$ and $\Delta SESR_{50th}$ values, and $SESR_{50th}$ and $\Delta SESR_{50th}$ are temporally and spatially affected. Theoretically, SESR and ΔSESR follow the standard normal distribution, where $SESR_{50th}$ = 0 and $\Delta SESR_{50th}$ = 0,

indicating that $SESR_{40th}$ is below the average level and $\Delta SESR_{40th}$ is less than 0 with the decreasing SESR. However, the skewed distributions of $SESR_{50th}$ and $\Delta SESR_{50th}$ in NCP demonstrate that $SESR_{50th}$ and $\Delta SESR_{50th}$ are generally not 0. When $SESR_{50th} > 0$, it is possible that $SESR_{40th} > 0$, indicating that ESR_{40th} exceeds $\overline{ESR}$ and $SESR_{40th}$ cannot indicate the low evaporative stress value. When $\Delta SESR_{50th} > 0$, maybe the corresponding $\Delta SESR_{40th} > 0$, reflecting an increasing SESR, which could result in the underestimation of evaporative stress value and inaccurate capture of FD events that would not occur. When $SESR_{50th} < 0$, ESR_{40th} may be significantly lower than $\overline{ESR}$, which would indicate a lower evaporative stress value. When $\Delta SESR_{50th} < 0$, $\Delta SESR_{40th}$ is also significantly less than 0, indicating a severe decreasing SESR, leading to the overestimation of evaporative stress value and disregard for the FD events that actually occurred."

18. Figure 4, first row: As shown in this figure, the frequency of events detected by MESR is significantly lower than that detected by SESR. It would be helpful if the authors could show how many events detected by MESR are similar to those detected by SESR. Since both methods use ESR, it would be beneficial to compare these methods. One possible approach would be to compare binary time series of drought events detected by these methods and calculate their overlap.

Response: Thank you for the comments. The rationality of the spatial distributions of FD_{RZSM}, FD_{SESR}, and FD_{MESR} frequency has been analyzed in conjunction with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET (see Comment 2 and lines 321-334). Besides, the overlapping events between FD_{SESR} and FD_{MESR} have been detected in Figure S4. The great overlapping event proportion and the significant correlation relationship between the pentad SESR and MESR percentiles demonstrate the reliability of the FD_{MESR} identification (see lines 347-354).

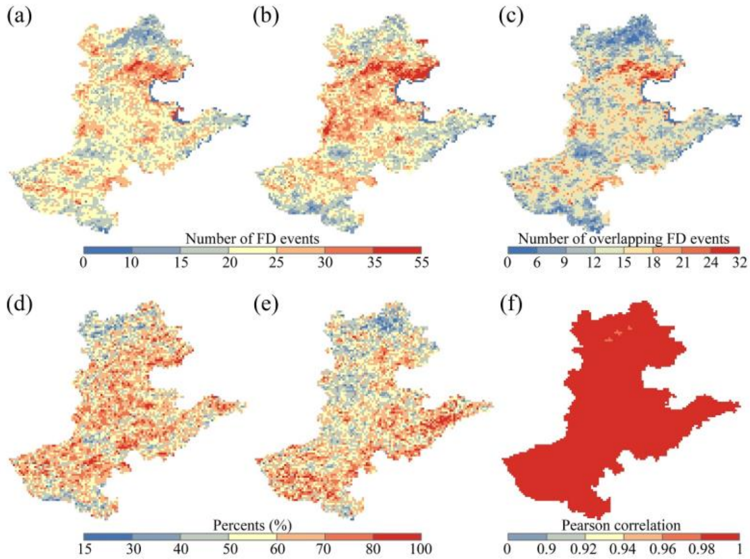

Figure S4 The number of (a) FD_{SESR}, (b) FD_{MESR}, (c) overlapping FD_{SESR} and FD_{MESR} events, the proportion of the overlapping events in (d) FD_{SESR} and (e) FD_{MESR} events, as well as (e) the correlation between pentad SESR and MESR percentiles.

"FD events that exhibit temporal overlap between FD_{SESR} and FD_{MESR} are regarded as

overlapping events, and Fig.S4 shows the number of overlapping events between the two. According to Fig.S4 (a) and (b), the spatial distributions of the number of FD_{SESR} and FD_{MESR} events are similar to their frequencies. Furthermore, overlapping events have a similar spatial distribution to FD_{SESR} events in Fig.S4 (c), which might be because the FD_{SESR} events are less than FD_{MESR} events. As can be seen in Fig.S4 (d), the overlapping FD events make up more than 50% and even more than 60% of FD_{SESR} events with the exception of the northwestern NCP. Figure S4 (f) also shows that the correlation between the pentad SESR and MESR percentiles is around 1, demonstrating a strong linear relationship between the two. The large overlapping proportion and linear relationship display the reliability of the FD_{MESR} identification."

19. Line 305: Is 5 pentads for the flash drought onset stage not too long? Flash droughts are characterized by their rapid onset, and 5 pentads is not particularly rapid. I would suggest setting a limitation on the duration of the flash drought onset stage in this research, similar to the RZSM method, which has an onset duration limitation. Additionally, are Figures 4 and S1 showing the average duration? If so, this implies that in some events, the onset stage of the flash drought is longer than 5 pentads!

Response: Thank you for the comments. Unlike the FD_{RZSM} method, which manages the onset speed by limiting the onset duration in criteria a), criteria d) of FD_{SESR} ensures that the drought has a rapid development rate and would not be affected by the temporary moderation of SESR (see lines 195-196).

The spatial distribution maps of FD characteristics in Figures 4 and S1 are the mean characteristics of all FD events that occur grid by grid, therefore the characteristics for the FD events might be larger or smaller than the mean. We have determined the characteristics of all FD events during 1981 ~ 2022, and displayed the histograms in Figure S5. Mostly $duration_{Onset}$ of the FD_{RZSM}, FD_{SESR}, and FD_{MESR} range from 2 to 4 pentads (see lines 372-378).

"Unlike the FD_{RZSM} method, which manages the onset speed by limiting the onset duration in criteria a), criteria d) of FD_{SESR} ensures that the drought has a rapid development rate and would not be affected by the temporary moderation of SESR."

"Since the spatial distribution maps of FD characteristics in Fig.4 are the mean characteristics of all FD events that occur grid by grid, the characteristics for the FD events might be larger or smaller than the mean. The histograms for the characteristics of all FD events during 1981 ~ 2022 are displayed in Fig.S5. It can be seen in Fig.S5 (a) ~ (c) that FD_{RZSM} mostly onsets in 2 ~ 3 pentads and recovers in 2 ~ 5 pentads, with the $duration_{Total}$ of 4 ~ 7 pentads. All FD_{SESR} and FD_{MESR} events have similar $duration_{Onset}$, $duration_{Recovery}$, and $duration_{Total}$, with $duration_{Onset}$ mainly ranging between 2 and 4 pentads, $duration_{Recovery}$ between 2 and 7 pentads, and $duration_{Total}$ between 6 and 9 pentads."

20. Line 321: The maps of severity, duration, and intensity of flash droughts detected by MESR are almost evenly distributed, especially compared to the other two methods. Although there are some concentrated areas, overall, these maps appear evenly distributed, suggesting that this method does not respond significantly to regional

characteristics, unlike the RZSM and SESR methods. Are there any reasons for this? If so, it would be helpful to discuss this further in the main text.

Response: Thank you for the comments. The spatial heterogeneity of the FD characteristics has been analyzed, combined with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET in the revised manuscript, which has been mentioned in Comment 2 (see lines 321-334, lines 361-364, lines 381-386, lines 398-405).

21. Lines 325-326: Are there any specific characteristics in the northern NCP that cause this change (e.g., land cover or background aridity)? It would be helpful if the authors could justify the spatial differences in the characteristics of droughts using this method.

Response: Thank you for the comments. The spatial heterogeneity of the FD characteristics in the original manuscript has been analyzed, combined with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET in the revised manuscript (see Comment 2 and lines 321-334, lines 361-364, lines 381-386, lines 398-405).

22. Line 361: A considerable number of flash droughts detected by MESR are categorized as grade 4, but they are not confirmed or detected by the other two methods. Moreover, this study relies on reanalysis ET, PET, and SM data, which contain uncertainties. Therefore, it is unclear whether the FD4 events detected by this method are actual flash drought events.

Response: Thank you for the comments. We have reanalyzed the characteristics of FD events based on two other reanalysis datasets, GLEAM and GLDAS 2 datasets, and obtained similar findings (see lines 437-442 and lines 569-574).

"Figure S6 (a) displays the proportion of FD1 ~ 4. FD_{RZSM} is primarily focused on $FD_{RZSM}3$, which is followed by $FD_{RZSM}2$ and $FD_{RZSM}4$, and finally, $FD_{RZSM}1$. On the other hand, FD_{SESR} is concentrated on $FD_{SESR}1$ ~ 2, $FD_{SESR}3$, and $FD_{SESR}4$. $FD_{MESR}4$ accounts for the largest amount of FD_{MESR}, while the proportions of $FD_{MESR}3$, $FD_{MESR}2$, and $FD_{MESR}1$ steadily decrease. The FD1 ~ 4 proportions based on GLEAM and GLDAS 2 datasets in Fig.S6 (b) and (c) also exhibit a great similarity to those based on the ERA5-Land dataset in Fig.S6 (a)."

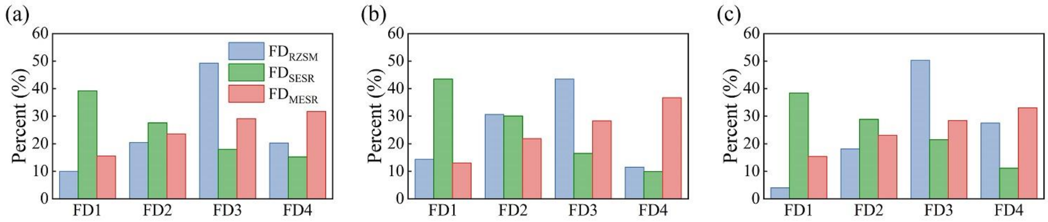

Figure S6 Proportion of FD1 ~ 4 for FD_{RZSM}, FD_{SESR}, and FD_{MESR} based on (a) ERA5-Land, (b) GLEAM, and (c) GLDAS 2 datasets.

"Besides, the FD characteristics identified based on ERA5-Land, GLEAM, and GLDAS 2 datasets are displayed in Figs.S5, S13 and S14. The distributions of the FD_{RZSM}, FD_{SESR}, and FD_{MESR} characteristics based on various datasets are comparable with the exception of the FD_{MESR} intensity based on GLDAS 2. Meanwhile, the proportions of

various FD grades determined by intensity from diverse data sources also demonstrate an indisputable resemblance, as seen in Fig.S6. The similarity of pentad RZSM, SESR, and MESR percentiles from various datasets, as well as the FD characteristics based on various datasets, effectively demonstrates the reliability of our findings."

23. Line 366: "Decrease trend" should be changed to "decreasing trend" for grammatical correctness.

Response: Thank you for the comments. We have revised the presentation (see lines 453-455).

"The affected area of FD_{RZSM} fluctuates between 0 and 60% with a clear decreasing trend, while that of FD_{SESR} ranging from 0 to 80% gradually rises and FD_{MESR} decreases ranging from 0 to 80%."

24. Line 387: "Increases" should be changed to "increase" to match the plural subject ("duration and severity").

Response: Thank you for the comments. We have revised "increases" into "increase" (see lines 475-476).

"Similarly, the $duration_{Total}$ and severity of FD_{SESR} and FD_{MESR} increase in the southern and west-central NCP, but decrease in the east-central NCP."

25. Line 389: "Increases" should be changed to "increase" to match the plural subject ("duration and severity").

Response: Thank you for the comments. We have revised the presentation and corrected the "increases" in the revised manuscript.

26. Line 399: Are these drought events categorized as flash drought or they are conventional drought events?

Response: Thank you for the comments. We have researched again for well-documented FD events and used two typical FD events occurring in 2017 and 2019 to evaluate the applicability of three FD identification methods in Section 3.4 (see Comment 1; Section 3.4 in lines 481-502). The FD events in 2017 and 2019 are identified by soil moistures, and both are well recorded with the detailed development and evolution records in Xue (2023) and Yao et al. (2022), further confirmed in Chen et al. (2024) as well. Grids suffering FD_{RZSM}, FD_{SESR}, and FD_{MESR} are identified in the revised manuscript and compared with the development and evolution records of the two real FD events, demonstrating good consistency.

27. Lines 395-495: I am not sure if this section can provide any reliable results, mainly because it is unclear whether those events are flash droughts or not, and if they are, how they were detected. Moreover, this study utilizes reanalysis data, which have inherent uncertainties. Therefore, this study could benefit from a comparison between these three methods and some real flash drought events in the study area that occurred recently and have reliable information on their characteristics.

Response: Thank you for the comments. We have researched again for well-documented FD events, and used two typical FD events occurring in 2017 and 2019 to evaluate the applicability of three FD identification methods in Section 3.4, which has been mentioned in Comment 1 (see Section 3.4 in lines 481-502).

28. Figure 9: What is the meaning of 'The color bands represent the pentads with FD events from June to August of that year.' For example, in 1989 event, what does it mean to have 44 (max) FD events from June to August in each pixel?
Response: Thank you for the comments. We have modified the presentation of the typical FD events to the pentad evolution of FD events, where the color bands have been removed (see Figure 9 in lines 500-502).

29. Lines 418-419: Do you have any reference for this? What are the main characteristics of these regions, mainly in term of land cover?
Response: Thank you for the comments. We have added the reasons why these regions are the hotspots combined with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET (see lines 507-513).
"These hotspots are situated in the northeastern and southwestern NCP, respectively. High hotspot indicator regions are likely to encounter frequent and severe FD events. Northern NCP has low AI and significant evaporative pressure. There is a significant water demand in the woodland. The eastern NCP is similar to the northern NCP with low AI and high evaporative pressure. Although it is cultivated land where irrigation might effectively alleviate the FD development, its coastal position promotes evaporation. In the southern NCP, the high temperature prolongs the FD $duration_{Total}$ *and increases the evaporation loss despite of the high AI and low evaporative pressure, accelerating the decline of soil moisture and rendering it prone to FD events."*

30. Line 448: It would be better if the authors could start this section with a sentence stating that they developed MESR method in this study.
Response: Thank you for the comments. We have revised the presentation (see line 576).
"This study developed a new FD identification method called MESR, which is a modified version of SESR."

31. Line 476: change 'decrease' to 'decreasing trend'
Response: Thank you for the comments. We have corrected the "decrease" into "decreasing" (see lines 605-606).
"The annual affected area of FD_{RZSM} *and* FD_{MESR} *in this study exhibit a significant decreasing trend, but that of* FD_{SESR} *exhibits a slow increase."*

32. Line 477: change 'slowly' to 'slow'
Response: Thank you for the comments. We have corrected the "slowly" into "slow" (see lines 605-606).
"The annual affected areas of FD_{RZSM} *and* FD_{MESR} *in this study exhibit a significant decreasing trend, but that of* FD_{SESR} *exhibits a slow increase."*

33. Lines 489-490: The 50th percentile of SESR is not greater than zero in all regions. As shown in Figure S7, in the vast majority of regions, it is around zero or even lower. Moreover, in regions with a higher evaporation baseline, the 50th percentile is slightly higher than zero, but this is not a disadvantage of the SESR method. Perhaps one reason your method shows spatially heterogeneous frequencies is this, which could lead to missing some rapidly developing events, particularly in wetter climate regimes.

Response: Thank you for the comments. Section 4.4 illustrates the rationality of the spatially heterogeneous frequencies between FD_{SESR} and FD_{MESR}. The spatial distribution of mean $SESR_{50th}$ and $\Delta SESR_{50th}$ in Fig.S16 on the FD_{SESR} pentads determined that there might be an overestimation in the FD_{SESR} frequency. The distributions of average $SESR_{40th}$ and $\Delta SESR_{40th}$ further show that the unobserved FD_{SESR} events would be captured in the northern NCP. Figure S17 displays the frequency of $SESR_{50th}$, $\Delta SESR_{50th}$, $SESR_{40th}$, and $\Delta SESR_{40th}$ when FD_{SESR} occurs, supporting the likelihood of FD_{SESR} frequency overestimation in the NCP as well. The difference between FD_{SESR} and FD_{MESR} identification results can be traced back to two aspects: PDFs fitting and variable thresholds. The PDFs fitting decreases the FD frequency while the variable threshold increases that. The contribution of PDFs fitting is less than that of variable thresholds in the northeastern and west-central NCP but greater in the other regions. Therefore, the frequency difference between FD_{SESR} and FD_{MESR} is mainly due to the variable thresholds in the northeastern and west-central NCP, and due to the PDFs fitting in the other regions. The relative contributions of PDFs fitting and variable thresholds demonstrate the rationality of the FD frequency difference between FD_{SESR} and FD_{MESR} (see lines 648-691).

"4.4 Attribution analysis of the frequency difference between FD_{SESR} and FD_{MESR}

Compared to FD_{SESR}, the FD_{MESR} frequency shows spatial heterogeneity, which is connected to the frequency distribution of $SESR_{50th}$ and $\Delta SESR_{50th}$. Figure S16 illustrates the spatial distribution of mean $SESR_{50th}$ and $\Delta SESR_{50th}$ on the FD_{SESR} pentads. Except for the southern NCP where the $SESR_{50th}$ is less than 0, the $SESR_{50th}$ in the northern and central NCP is mainly more than 0. Besides, the $\Delta SESR_{50th}$ in NCP is all larger than 0, which makes it possible that $SESR_{40th}$ cannot indicate the real low evaporative stress value, as well as the underestimation of evaporative stress value and inaccurate capture of FD_{SESR} events that would not occur. Therefore, it may result in the overestimation in the NCP (see Sect.3.2). Theoretically, the $SESR_{40th}$ and $\Delta SESR_{40th}$ in the NCP should be smaller than 0, representing the low evaporative stress value and decreasing ESR. The distributions of average $SESR_{40th}$ and $\Delta SESR_{40th}$ in Fig.S16 (c) (d) indicate that both are less than 0 except for the northern NCP, where the $SESR_{40th}$ is greater than 0. It further shows that the unobserved FD_{SESR} events would be captured in the northern NCP. The frequency of $SESR_{50th}$, $\Delta SESR_{50th}$, $SESR_{40th}$, and $\Delta SESR_{40th}$ when FD_{SESR} occurs are determined in Fig.S17. The $SESR_{50th}$ and $\Delta SESR_{50th}$ values are mostly greater than 0, particularly the $\Delta SESR_{50th}$, corresponding to Fig.S16 (a) (b). However, most $SESR_{40th}$ and $\Delta SESR_{40th}$ are below 0. It could not be ignored that both $SESR_{40th}$ and $\Delta SESR_{40th}$ greater than 0 account for around 35%, supporting the likelihood of FD_{SESR} overestimation in the NCP.

The FD identification based on SESR and MESR differs in that the ESR is standardized by its mean to construct MESR rather than being normalized into SESR, MESR is fitted with various PDFs rather than EDF, and the variable thresholds are utilized in the FD_{MESR} identification. MESR and SESR are both linearly converted from ESR in order to facilitate comparison of the FD identification results between different regions. Converting ESR to MESR rather than SESR only makes the ESR standardization process more reasonable, but the linear translation of ESR into SESR or MESR has no effect on its corresponding percentile. Therefore, the difference between FD_{SESR} and FD_{MESR} identification results can be traced back to two aspects: PDFs fitting and variable thresholds.

To demonstrate their contribution to the difference between FD_{SESR} and FD_{MESR}, the thresholds in FD_{SESR} method are referred to. The fixed thresholds, MESRonset1 = 40, MESRonset2 = 20, ΔMESRonset = 40, and MaxMESRchange = 25, are applied in the FD_{MESR} method, which is called $FD_{MESR\text{-}invariable}$. Figure S18 (a) ~ (c) displays the frequency of $FD_{MESR\text{-}invariable}$ and the differences between $FD_{MESR\text{-}invariable}$ frequency and FD_{SESR} and FD_{MESR} frequency. Figure S18 (d) (e) also show the contribution of PDFs fitting and variable thresholds to the differences between FD_{SESR} and FD_{MESR} frequency, respectively, as well as the relative contribution in Fig.S18 (f). $FD_{MESR\text{-}invariable}$ and FD_{MESR} have a comparable frequency spatial distribution, with higher frequency in the north-central and lower in the south of NCP. Besides, $FD_{MESR\text{-}invariable}$ has a lower frequency than FD_{SESR} and FD_{MESR} in NCP, indicating that the PDFs fitting decreases the FD frequency while the variable thresholds increase that. According to Fig.S18 (d) (e), the PDFs fitting has a negative impact on the difference between FD_{SESR} and FD_{MESR} frequency in the northeastern and west-central NCP, whilst the variable thresholds have a positive impact. However, it is opposite in the other NCP regions. As shown in Fig.S18 (f), the relative contribution of the variable thresholds to the PDFs fitting mostly ranges between -2 and -1 in the northeastern and west-central NCP, but between -1 and 0 in the other regions. Therefore, the absolute values of the relative contributions of the PDFs fitting are larger than 1 in the northeastern and west-central NCP, but less than 1 in the other regions. Taking into account the negative contribution of PDFs fitting in the NCP, it can be assumed that the contribution of PDFs fitting is less than that of variable thresholds in the northeastern and west-central NCP but greater in the other regions. Therefore, the frequency difference between FD_{SESR} and FD_{MESR} frequency is mainly due to the variable thresholds in the northeastern and west-central NCP, and due to the PDFs fitting in the other regions.

The FD_{SESR} frequency is lower than the FD_{MESR} frequency in the northeastern and west-central NCP while larger in the other regions. Although the FD_{SESR} frequency is overestimated in the NCP, particularly the northern NCP, PDFs fitting decreases the FD frequency in the FD_{MESR}. But the variable thresholds increasing the FD frequency contribute more in the northeastern and west-central NCP, resulting in a greater FD_{MESR} frequency than FD_{SESR} frequency. In the other regions where PDFs fitting plays a larger role, the decreasing from PDFs fitting surpasses the increasing from the variable thresholds, making that the FD_{MESR} frequency is lower than the FD_{SESR} frequency."

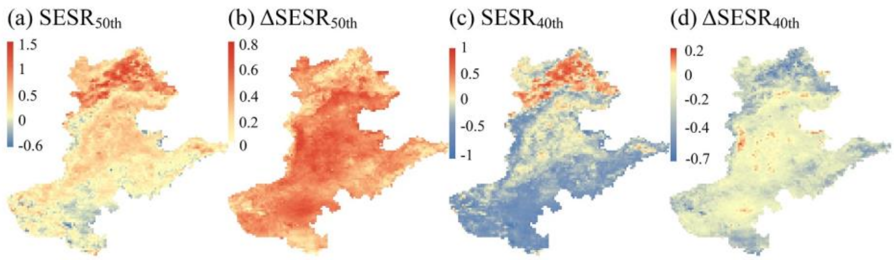

Figure S16 Distribution of the mean (a) $SESR_{50th}$, (b) $\Delta SESR_{50th}$, (c) $SESR_{40th}$, and (d) $\Delta SESR_{40th}$ on FD_{SESR} pentads.

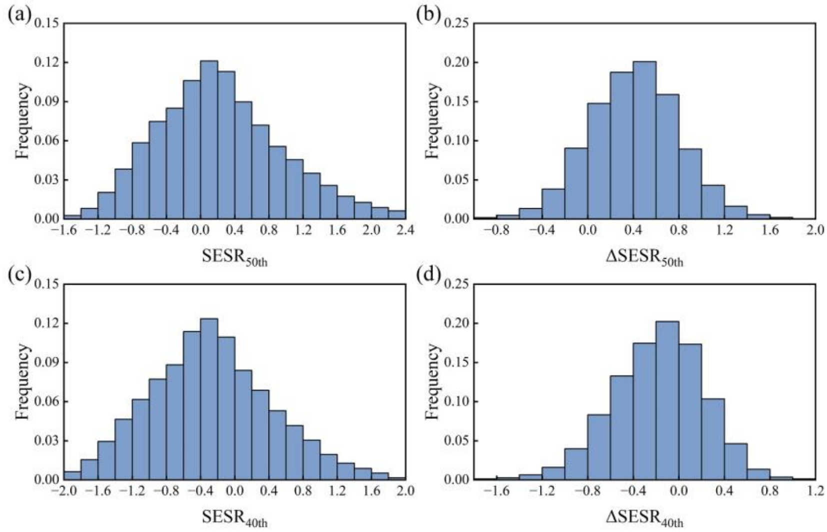

Figure S17 Histogram of (a) $SESR_{50th}$, (b) $\Delta SESR_{50th}$, (c) $SESR_{40th}$, and (d) $\Delta SESR_{40th}$ on FD_{SESR} pentads.

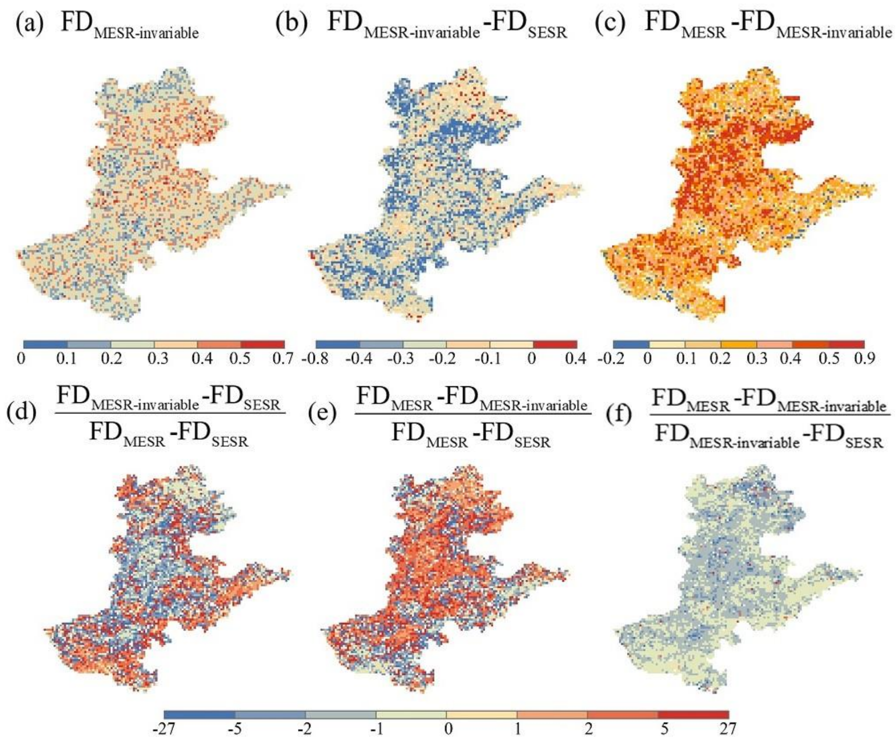

Figure S18 Frequency of $FD_{MESR\text{-}invariable}$ and its difference with FD_{MESR} and FD_{SESR}.

34. Lines 498-499: Actually, there are three differences. In addition to what you have mentioned (PDF fitting and threshold), your method multiplies ESR by its mean, whereas in the original SESR, ESR anomalies are standardized. Please include this in the main text.

Response: Thank you for the comments. The three differences between FD identification using MESR and SESR have been supplemented in lines 662-668. Due to that the linear transformation conducted on ESR does not influence the percentile results, which only increases the rationality of the ESR standardization process, the difference between FD_{SESR} and FD_{MESR} identification results can be traced back to two aspects: PDFs fitting and variable thresholds.

"The FD identification based on SESR and MESR differs in that the ESR is standardized by its mean to construct MESR rather than being normalized into SESR, MESR is fitted with various PDFs rather than EDF, and the variable thresholds are utilized in the FD_{MESR} identification. MESR and SESR are both linearly converted from ESR in order to facilitate comparison of the FD identification results between different regions. Converting ESR to MESR rather than SESR only makes the ESR standardization process more reasonable, but the linear translation of ESR into SESR or MESR has no effect on its corresponding percentile. Therefore, the difference between FD_{SESR} and FD_{MESR} identification results can be traced back to two aspects: PDFs fitting and variable thresholds."

35. Line 528: Again, these historical events are not reliable, mainly because as the authors mentioned that there is no detailed information on these events, and it is unclear whether they were flash droughts or not. The combination of uncertainty in the reanalysis datasets used in this study and the lack of adequate information on these events cannot lead to a reliable conclusion.

Response: Thank you for the comments. We have researched again for well-documented FD events, and used two typical FD events occurring in 2017 and 2019 to evaluate the applicability of three FD identification methods in Section 3.4 (see Comment 1; Section 3.4 in lines 481-502 and lines 707-708). Besides, the uncertainties from the different reanalysis datasets have been evaluated in the Section 4.1 of the revised manuscript (see Comment 9 and lines 552-574).

"Meanwhile, this study demonstrated the three FD methods could accurately identify the true FD events through the two typical historical events in Sect.3.4."

36. Lines 447-448: Please rephrase this sentence as follows: "There are notable differences in the spatial distribution of severity among the three FD methods."

Response: Thank you for the comments. We have rephrased this sentence (see lines 729-730).

"There are notable differences in the spatial distribution of severity among the three FD methods."

37. Lines 545-549: This study could benefit from a deeper discussion on the main

reasons for changes in the frequencies and characteristics of FD events. Such a discussion could incorporate land cover or the background aridity of the study area.

Response: Thank you for the comments. The spatial heterogeneity of the FD characteristics in the original manuscript has been analyzed, combined with the land use types, Aridity Index (AI), and the ratio of mean annual ET and PET in the revised manuscript, which has been mentioned in Comment 2 (see lines 321-334, lines 361-364, lines 381-386, lines 398-405).

38. Lines 561-565: Isn't it obvious? As the depth of soil moisture increases, the impact of flash droughts on SM can become less pronounced. Moreover, this study used reanalysis SM data, so finding these trends in reanalysis datasets is not surprising!

Response: Thank you for the comments. We have identified the FD events based on GLEAM and GLDAS 2 datasets as well, and obtained similar findings (see lines 522-523). Moreover, the mean impacts of the thresholds for each unit on the FD frequency based on different datasets are shown in Table S3, which further measures the impacts of thresholds on the FD frequency. Overall, the minimum FD duration, as well as the continuous pentads that FD indexes such as RZSM, ΔSESR, and ΔMESR should exceed the FD termination thresholds, is sensitive to the FD identification, while the thresholds for the decreasing indexes in the FD onset stage are insensitive (see lines 536-546). The similar effects of various thresholds on the frequency based on various datasets increases the credibility of the findings (see lines 715-718).

"The mean FD_{RZSM} frequency based on various soil layers from GLEAM and GLDAS 2 datasets are listed in Table S2, where the FD frequency follows the same pattern."

"To further demonstrate the impacts of thresholds on the FD frequency, the mean impacts of the thresholds for each unit on the FD frequency are shown in Table S3, corresponding to the FD frequency in Figs.11, S9, and S10. For FD_{RZSM}, the impacts of thresholds on the mean frequency are ranked as follows: "RZSMtermination" < "RZSMpentad1" < "RZSMonset1" < "RZSMonset2" < "MinDuration" < "RZSMpentad2". For FD_{SESR}, the impacts of thresholds on the FD frequency ranging from small to large are: "ΔSESRonset" < "SESRonset1" < "SESRonset2" < "MaxSESRchange" < "SESRpentad" < "MinDuration". For FD_{MESR}, the impacts of thresholds are: "ΔMESRonset" < "MESRonset1" < "MESRonset2" < "MaxMESRchange" < "MESRpentad" < "MinDuration". Therefore, the minimum FD duration ("MinDuration"), as well as the continuous pentads that FD indexes such as RZSM, ΔSESR, and ΔMESR should exceed the FD termination thresholds ("RZSMpentad2", "SESRpentad", and "MESRpentad"), is sensitive to the FD identification, while the thresholds for the decreasing indexes in the FD onset stage ("RZSMtermination", "RZSMpentad1", "RZSMonset1", "ΔSESRonset", "SESRonset1", "SESRonset2", "ΔMESRonset", "MESRonset1", and "MESRonset2") are insensitive."

"The sensitivity of thresholds to the FD_{RZSM}, FD_{SESR} and FD_{MESR} frequency are measured, as well as that based on various datasets. Table S3 demonstrates that the rankings of the effects of various thresholds on the mean frequency based on various datasets exhibit notable similarities, increasing the credibility of the findings."

Table S3 The mean frequency change of various FD events caused by one unit

change in the FD identification thresholds

FD identification method	*Parameter*	*ERA5-Land*	*GLEAM*	*GLDAS 2*
FD_{RZSM}	*RZSMpentad1*	*0.0031*	*0.0024*	*0.0025*
	RZSMonset1	*-0.0121*	*-0.0167*	*-0.0075*
	RZSMonset2	*0.0227*	*0.0297*	*0.0137*
	RZSMpentad2	*-0.0788*	*-0.1493*	*-0.0411*
	RZSMtermination	*-0.0022*	*-0.0005*	*-0.0014*
	MinDuration	*-0.0387*	*-0.0857*	*-0.0179*
FD_{SESR}	*Minduration*	*-0.6137*	*-0.1468*	*-0.1043*
	SESRonset1	*0.0022*	*-0.0018*	*-0.0002*
	SESRonset2	*0.0277*	*0.0094*	*0.0108*
	ΔSESRonset	*0.0007*	*0.0016*	*0.0010*
	SESRpentad	*-0.0978*	*-0.0299*	*-0.0196*
	MaxSESRchange	*0.0457*	*0.0161*	*0.0136*
FD_{MESR}	*Minduration*	*-0.5410*	*-0.2507*	*-0.1126*
	MESRonset1	*0.0021*	*0.0018*	*-0.0007*
	MESRonset2	*0.0249*	*0.0164*	*0.0131*
	ΔMESRonset	*-0.0009*	*0.0090*	*0.0040*
	MESRpentad	*-0.1260*	*-0.0054*	*-0.0122*
	MaxMESRchange	*0.0476*	*0.0096*	*0.0075*