Response to Referee #3:

The authors would like to thank you for the constructive and thoughtful comments. We have addressed all your suggestions, leading to a much improved and complete manuscript. The following comments are addressed in the sequence as they were asked. We also respond to each point and clarify the corresponding changes adopted in the revised manuscript. The original comments are copied from the report with a **bolded font in black**, and our answers are in blue. Manuscript changes are in ***bold italic***.

Thank you again for your time and effort in reviewing our manuscript.

Yours sincerely

Baoqing Zhang (on behalf of all co-authors)

Reviewer #3

Tian et al. proposed a drought dataset for indicating drought across multiple categories and temporal scales. The proposed SZIsnow dataset includes different physical water-energy processes, especially snow processes. The evaluation for different spatiotemporal scales indicates the dataset can distinguish different types of drought. The SZIsnow shows superior performance over the cold regions. In addition, the dataset successfully described large-scale drought events over the world.

The purpose of the proposed work is clear and essential in order to establish a drought dataset including snow information for other studies related to drought. The manuscript is generally well-structured, the method description, evaluation, and data availability were well-written. However, several points should be addressed in the revised manuscript. I offer comments below in the hope this can be used to improve the paper further.

Reply: We appreciate the reviewer's comments on our work.

Major comments

First, in lines 102-103, the author stated that the GLDAS-2 drives the Noah land surface model (LSM), forced by the global Princeton meteorological forcing data, to approximate the observed land surface state. However, the Noah is not the only land surface model used by the GLDAS-2. I get the information from the URL: https://ldas.gsfc.nasa.gov/gldas. Could you explain why you chose the Noah model? Are there any differences among these land surface models?

Reply: We thank the reviewer for this excellent comment. At present, there are two versions of the GLDAS product: GLDAS-1 and GLDAS-2. The GLDAS-1 drives four land surface models (LSMs) at a spatial resolution of 1.0° from 1979 to 2017, including the Community Land Model (CLM), Mosaic model (MOS), NOAH model (NOAH1.0), and Variable Infiltration Capacity model (VIC). On the other hand, the GLDAS-2 only drives the upgraded version of Noah model (NOAH2.0) at a spatial resolution of 0.25°from 1948 to 2010. Here we chose 18 snow-influenced basins from the 32 global basins to discern the differences in the snow processes representation among the five LSMs mentioned above.

We compared and analyzed snow processes of the LSMs in terms of snowfall (Figure R1), snowmelt (Figure R2), and the snow water equivalent (SWE, Figure R3). As we can see, there are large differences for the snowmelt processes related to model snow physics, surface meteorological forcing (e.g., total precipitation, rainfall, snowfall, 2-meter air temperature), and sublimation in GLDAS-1 and GLDAS-2. As the same forcing is used for the four models in GLDAS-1, its snowfall is very similar, although CLM has much more snowfall than the other three models. The reason is that all other three models use 0 ºC 2-meter air temperature to separate total precipitation into rainfall and snowfall. However, CLM uses the Jordan algorithm (Jordan 1991) to separate total precipitation into rainfall and snowfall. The Jordan algorithm accounted for frozen rain, frozen rain and snow mix, and snowfall uses 2.5 ℃ 2-m air temperature to separate more total precipitation algorithm into snowfall. Additionally, small snowmelt values in GLDAS-1 Noah1.0 is due to larger sublimation. In general, compared GLDAS-2 with GLDAS-1, major SWE differences come from surface meteorological forcing data (precipitation and air temperature), although model snow physics also plays an important role.

We chose the GLDAS-2 as the forcing data to derive the SZI_{snow}, since our survey of the previous studies shows that the GLDAS-2 has a better performance than the GLDAS-1. Although the GLDAS-2 products have relatively large uncertainties in snow processes over regions with low quality meteorological forcing data and uncertain model physical processes, hydrological and meteorological studies over areas that lack complete set observations would benefit from GLDAS-2 products. This poses a dilemma. Nevertheless, using GLDAS-2 products to represent the land surface water-energy states and fluxes for large-scale hydrological or meteorological research is currently one of the commonly used methods, especially over regions lacking highquality *in-situ* observations. Despite uncertainties and errors in snow simulation of the GLDAS-2 product, the accuracy of the input water budgets does not influence the conceptual and technical improvement of the SZI_{snow} by considering the impact of snow dynamics on water supply and demand in drought characterization.

Jordan, R., 1991: A one-dimensional temperature model for a snow cover: Technical documentation for SNTERERM.89. Special Rep. 91–16, Cold Region Research and Engineers Laboratory, U.S. Army Corps of Engineers, Hanover, NH, 61 pp.

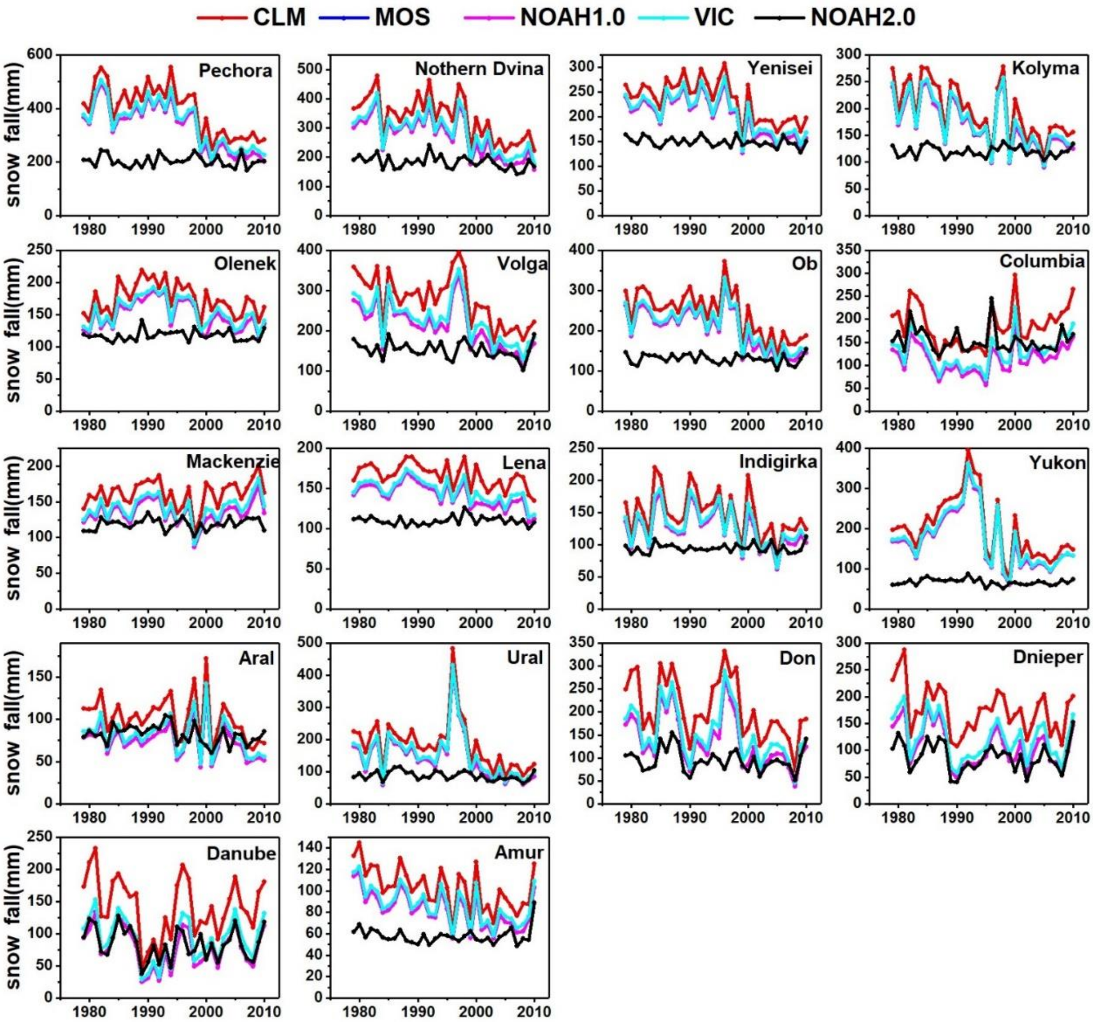


Figure R1. Monthly snowfall of different LSMs at 18 snow-influenced basins during 1979-2010.

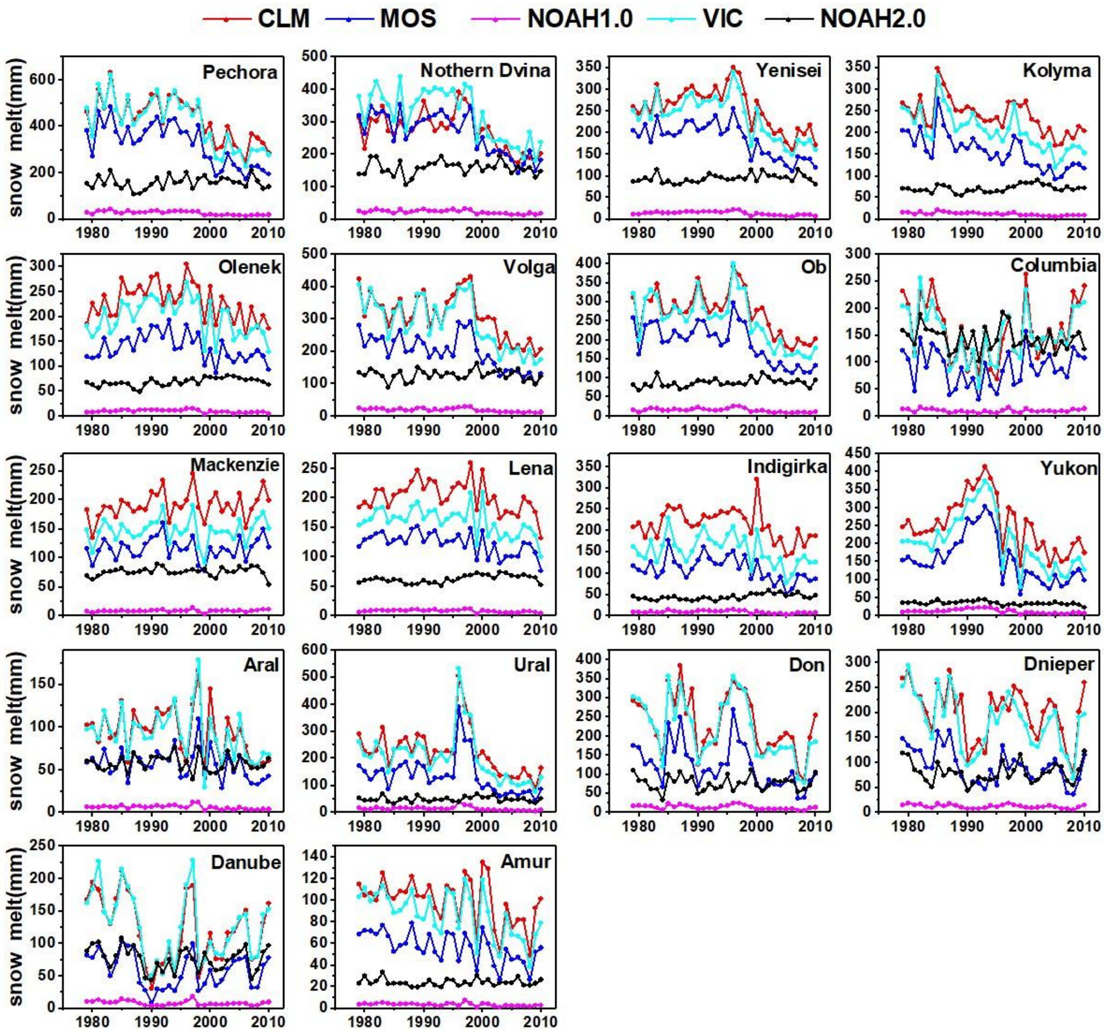


Figure R2. Monthly snowmelt of different LSMs at 18 snow-influenced basins during 1979-2010.

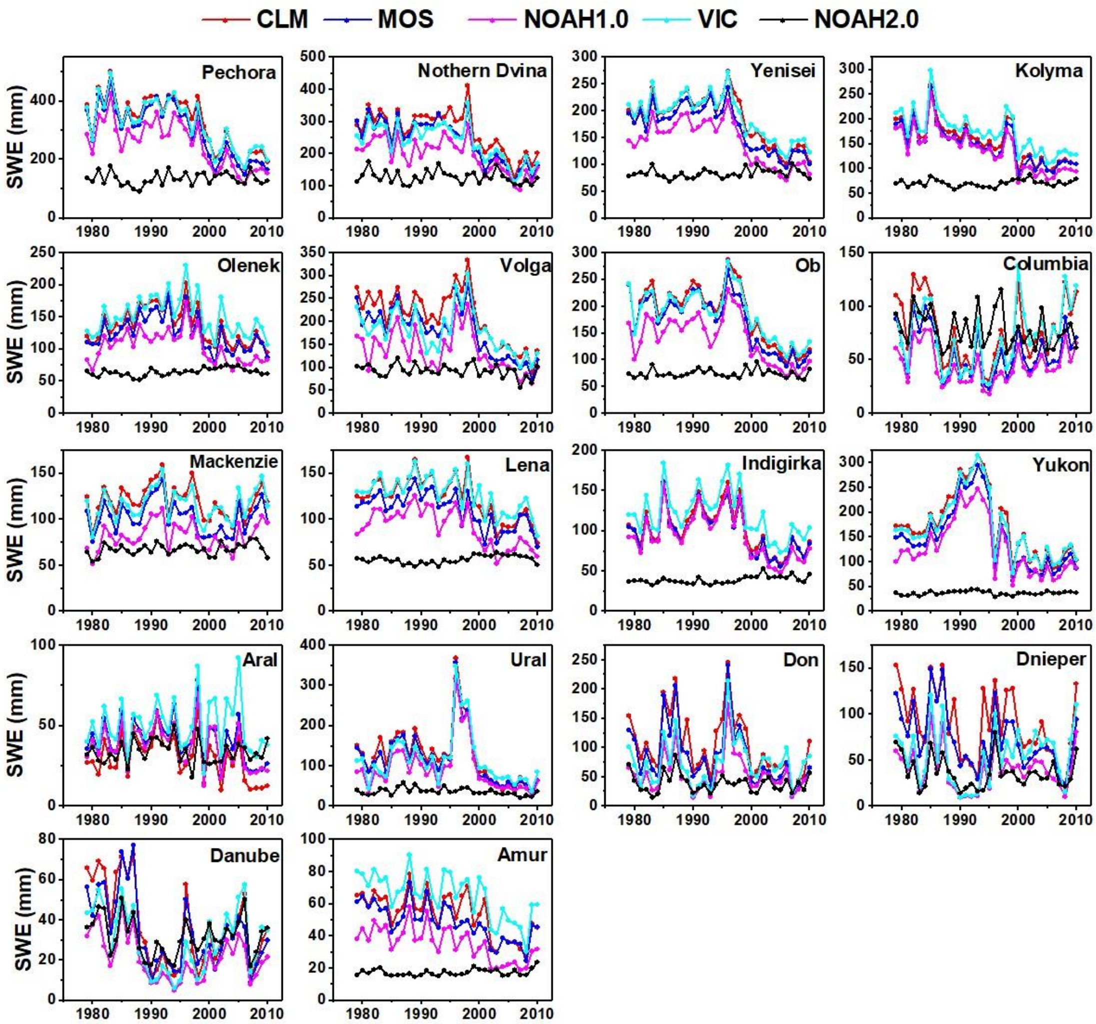


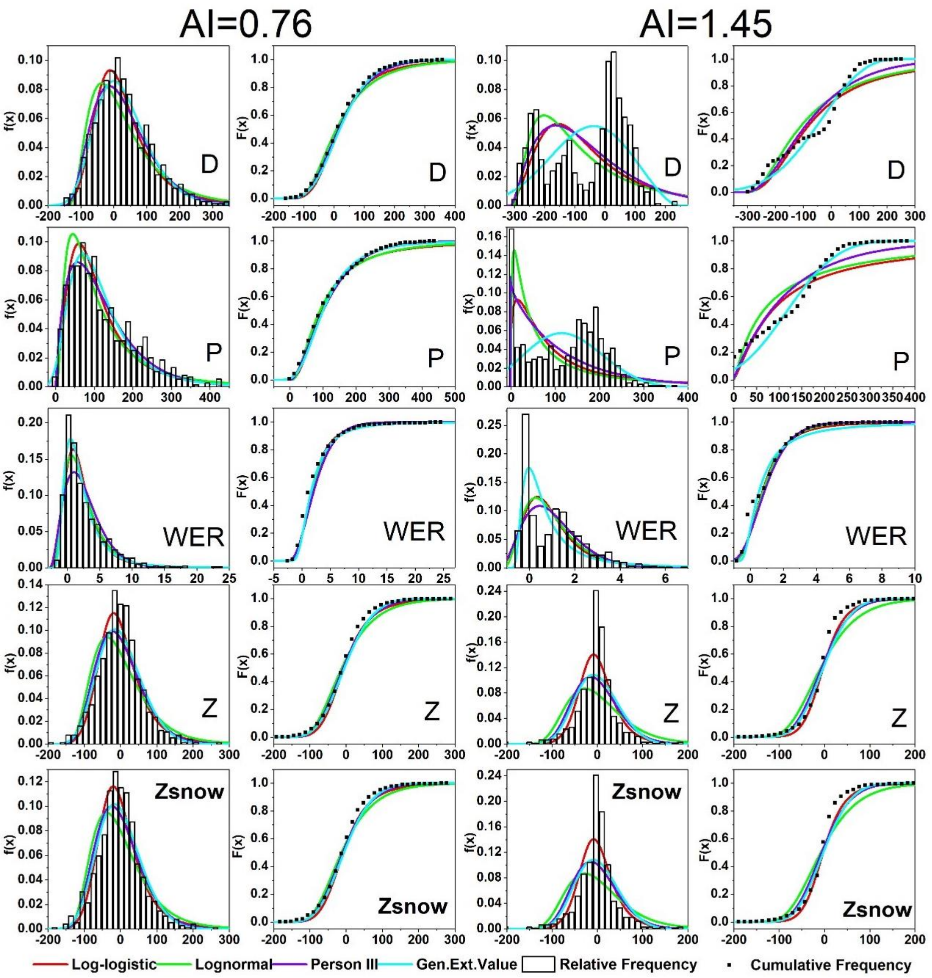
Figure R3. The monthly SWE of different LSMs in GLDAS at 18 snow-influenced basins during 1979-2010.

Second, the author adopted the log-logistic distribution to standardize precipitation, streamflow, and soil water storage to compute the Standardized Precipitation Index, Standardized Streamflow Index, and Standardized Water Storage Index. As I know, other probability distributions can be used to standardize. Are your evaluation results independent of different methods?

<u>Reply:</u> We have tested the validity of four applied distribution functions for precipitation (P), D ($D = P - PET$), WER, Z, and Z_{snow} across different climate zones. These climate zones were classified on the basis of the Aridity Index (AI), which is computed by the ratio of the long-term mean annual PET to P ($AI = PET/P$). A drier local climate condition corresponds to a higher AI value. Therefore, regions with AI

equal to 0.76, 1.45, 2.5, 4.4, and 79.9 represent humid, sub-humid, semi-arid, arid, and extreme arid regions.

As shown in Figure R4, we tested four possible three-parameter distributions to model the P, D, WER, Z, and Z_{snow} values at different climate zones, including Pearson III, log-logistic, lognormal, and general extreme value (GEV) distribution. For this purpose, the L-moment ratio diagram was used to examine the performance of different distributions because it allows comparing the empirical frequency distribution of P, D, WER, Z, and Z_{snow} time series at different climate regions with several theoretical distributions. The results show the log-logistic distribution generally fits the P, D, WER, Z, and Z_{snow} data the best across multiple climate zones. Therefore, the log-logistic distribution can reasonably be adopted for standardizing the P, D, WER, Z, and Z_{snow} data in calculating the SPI, SPEI, SWI, SZI, and SZI_{snow}.



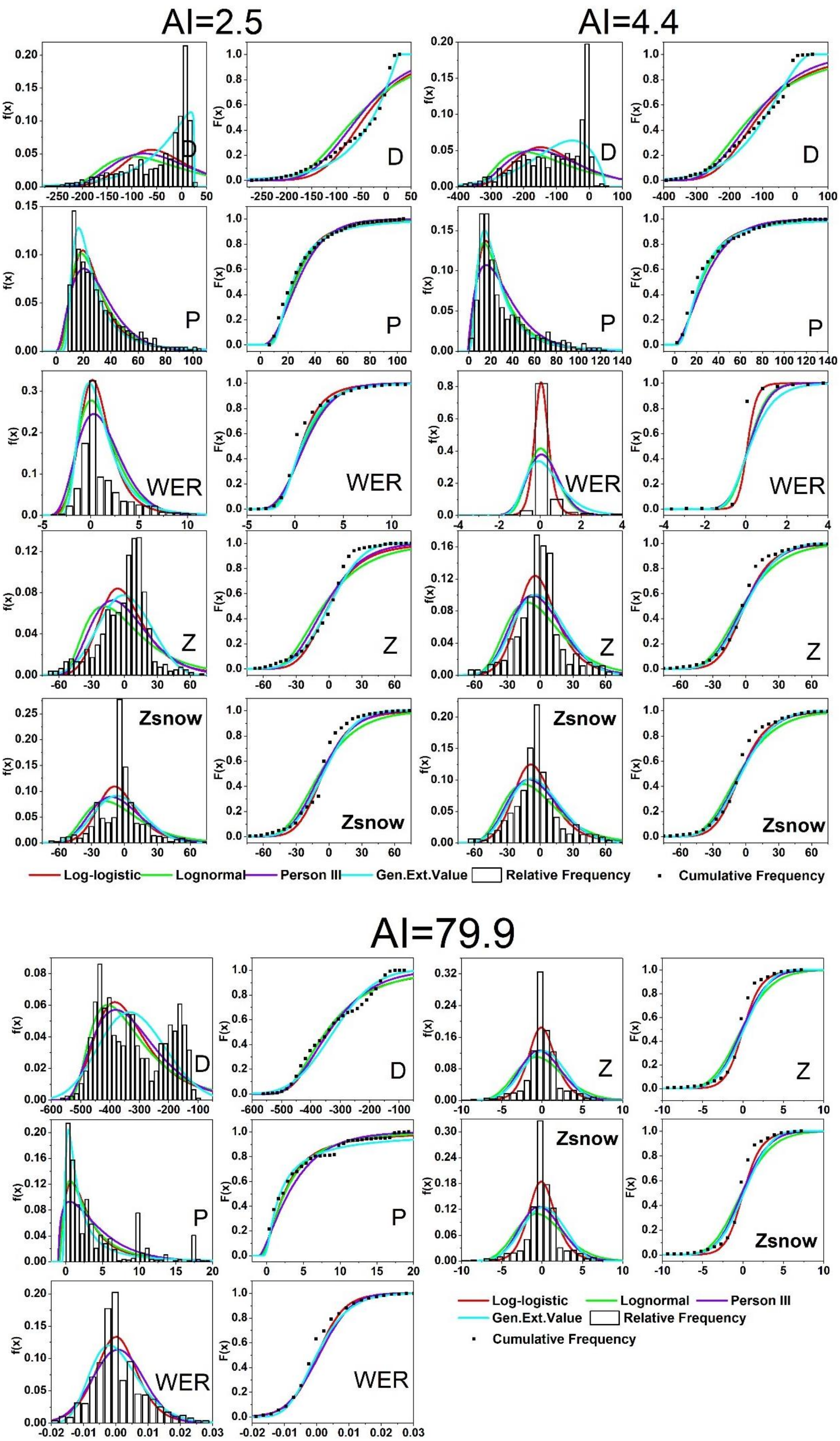


Figure R4. Empirical and modeled values (f(x) and F(x) or probability density function and cumulative distribution function, respectively) using the Pearson III, log-logistic, lognormal, and general extreme values (Gen. Ext. Value) distributions of the P, D, WER, Z, and Z_{snow} series at monthly scale for deriving the SPI, SPEI, SWI, SZI, and SZI_{snow}.

Minor comments

1) Lines 19-20, 22

Some numbers keep two decimals, and some do not.

Reply: The number in line 22 will be revised and consistent with other numbers. In the revised manuscript, this line will be changed to "***of which 68.38% had a duration longer than 6 months.***"

2) Lines 99-101

Show some reasons to explain why the better performance of GLDAS-2 compared to that of GLDAS-1?

Reply: Various studies have shown that the GLDAS-2 has a better performance than the GLDAS-1. These studies pointed out several causes for such better performance of GLDAS-2, and here we summarized the causes as follows. GLDAS-1 has serious discontinuity issues in its forcing data. Especially, GLDAS-1 precipitation data have larger errors in 1996, and the snowfall amount has approximately doubled after 2000. These discontinuity issues are primarily attributed to several switches of the forcing data of GLDAS-1. In contrast, GLDAS-2 precipitation has much better temporal continuity than GLDAS-1 precipitation, as GLDAS-2 uses the bias-corrected Princeton meteorological forcing dataset. Additionally, evaluations over the high-altitude regions indicate that GLDAS-2 models perform better than the GLDAS-1 models in simulating runoff. This is mainly because the GLDAS-2 models integrate glacier melt runoff in the simulation, whereas the GLDAS-1 models demonstrate no glacier melt runoff.

Following the comment, we will add these causes in the revised manuscript to explain the better performance of GLDAS-2 compared to GLDAS-1 as follows:

This is mainly attributed to that GLDAS-1 has serious discontinuity problems in its meteorological forcing dataset due to switches in its forcing data. In contrast,

GLDAS-2 has a better temporal continuity, using the bias-corrected Princeton meteorological forcing dataset. Additionally, evaluations for the high-altitude regions indicate that GLDAS-2 performs better in streamflow simulation because GLDAS-2 considers streamflow from glacier melt in its simulation, but the GLDAS-1 did not.

3) Line 140

I did not quite understand why the WER is regarded as a comprehensive drought indicator? What is a comprehensive drought indicator? Is there any definition for it? Please give more information.

Reply: The residual water–energy ratio (WER) was first suggested by Liu et al. (2017). As mentioned in the manuscript, the ratio of the residual available water to the residual energy (PET−ET) is relatively low (large) during drought (wet) events relative to normal conditions. Defining this ratio as WER = (P − ET)/(PET − ET), Liu et al. (2017) proposed a method for examining the response of the surface water-energy fluxes to drought based on WER.

Droughts are generally classified into three categories (meteorological, hydrological, and agricultural droughts) based on their physical characteristics. Nevertheless, these different categories make it difficult to objectively quantify drought features. Moreover, different types of droughts are considered to be interchangeable. Drought can convert from one type to another as it evolves in time and space. Zhang et al. (2019) thus recommend employing the WER, a comprehensive drought indicator, to comprehensively describe drought through a water-energy balance perspective. We can objectively and easily quantify drought across different spatial scales through such a perspective.

4) Line 155

Did you evaluate the capacity of your dataset across different climate zones? I did not see these mentioned climate zones in the following sections of the manuscript.

I guess you wanted to say that the evaluation was conducted over different geographical parts of the world. Please clarify this.

Reply: This is a typo. This line will be corrected in the revised manuscript as "***multiscalar drought across different geographical parts of the world***."

5) Lines 221-222

What is the name of dimensions in your 3D and 2D dataset? Please clarify this.

Reply: The names of dimensions in the 3D dataset are month, latitude, and longitude. The names of dimensions in the 2D dataset are latitude and longitude. This sentence will be modified in the revised manuscript as follows:

The SAD method firstly uses a monthly three-dimensional (3D, $\boldsymbol{month} \times \boldsymbol{latitude} \times \boldsymbol{longitude}$) gridded drought index dataset to identify two-dimensional (2D, $\boldsymbol{latitude} \times \boldsymbol{longitude}$) drought clusters in each time step.

6) Line 273

It seems this content has been repeated in the main text. Delete it to make the caption more concise.

Reply: This content will be removed from the revised manuscript following the comment.

7) Line 353

Is the geographical extent of Oceania equal to that of Australia? Give an exact definition.

Reply: The geographical extent of Oceania is not equal to Australia. This study defines Oceania as Australia, New Zealand, Papua New Guinea, and the Pacific Islands. This content will be added in the revised manuscript as follows:

Oceania is defined as Australia, New Zealand, Papua New Guinea, and the Pacific Islands.

8) Why wasn't Greenland included in Figure 6? It seems all the Greenland are missing values.

Reply: Greenland was excluded from our study because about 80% of its surface is ice-capped. In addition, Greenland has a small population of nearly 56,100 (in 2016) on an area of 2,166,086 km^2, making Greenland the least densely populated place on earth. Thus, the influence of droughts on this island is minimal, and the droughts in Greenland were not of interest for this study.

Following the comment, the reason for the exclusion of Greenland will be added in the revised manuscript as follows:

Our study excluded Greenland due to its sizeable ice-capped area (about 80% of the island) and low-density population.

9) When you did the SAD analysis, how to process the drought over the Sahara Desert?

Reply: When we did the SAD analysis, drought events were tracked through time by searching for overlapping grid cells between clusters at contiguous time steps. Clusters were allowed to propagate into the Sahara until their centroids fell within the specified domain that is outside the Sahara. Droughts whose centroids fell within the Sahara Desert (20°N–25°N, 17°W–34°E) were removed from our analysis since droughts in these regions were not of interest for this study.

Following the comment, how to process droughts over the Sahara in the SAD analysis will be added in the revised manuscript as follows:

Droughts in the Sahara Desert were not concerned in our study. During the SAD analysis, clusters were allowed to propagate into the Sahara (20°N–25°N, 17°W–

34°E), and these clusters would be retained if their centroids fell outside the Sahara Desert. In contrast, drought clusters were discarded if their centroids locate in the Sahara Desert.

Technical comments:

1) Figure 1: The font size of the description to compare the strength and weakness of each index is small.

Reply: In the revised manuscript, we will enlarge the font size of the description to compare the strength and weakness in Figure 1.

2) Figure 4: The font should be the same, use one type.

Reply: We will revise the font and use the same font in Figure 4 in the revised manuscript.

3) Figure 6: The stippling seems unclear.

Reply: We will increase the resolution of Figure 6 in the revised manuscript.

4) Figure 9: Please supply information on the geographic coordinate system used in this figure. It can help others to compare their results with yours.

Reply: Figure 9 uses the geographic coordinate system (latitude and longitude). We will add this information into the caption of Figure 9 in the revised manuscript.

5) Figure S2: Adjust the minimum value of the legend. There is no grid with a SZIsnow value less than -4.0.

Reply: The minimum level of the SZIsnow was set as -4.0 in Figure S2, since we wanted to keep the legend of Figure S2 to be consistent with that in Figure 9.

Comments for the dataset

1) Table 1 in the metadata file should be kept the same as Table 2 in the manuscript.

Reply: We will replace Table 1 in the metadata file with Table 2 in the manuscript.

2) Add information relative to the size of decompressed files in the metadata file.

Reply: Following the comment, we will add the below information relative to the size of decompressed files in the metadata file:

The size of each zipped file is 14.5 GB (gigabyte), and each file in a zipped file is 2.43 GB. The total size of all the decompressed files is 116 GB.

3) I recommend providing a thumbnail of your dataset in the metadata file.

Reply: Following the comment, we will add a thumbnail of our dataset in the metadata file as follows:

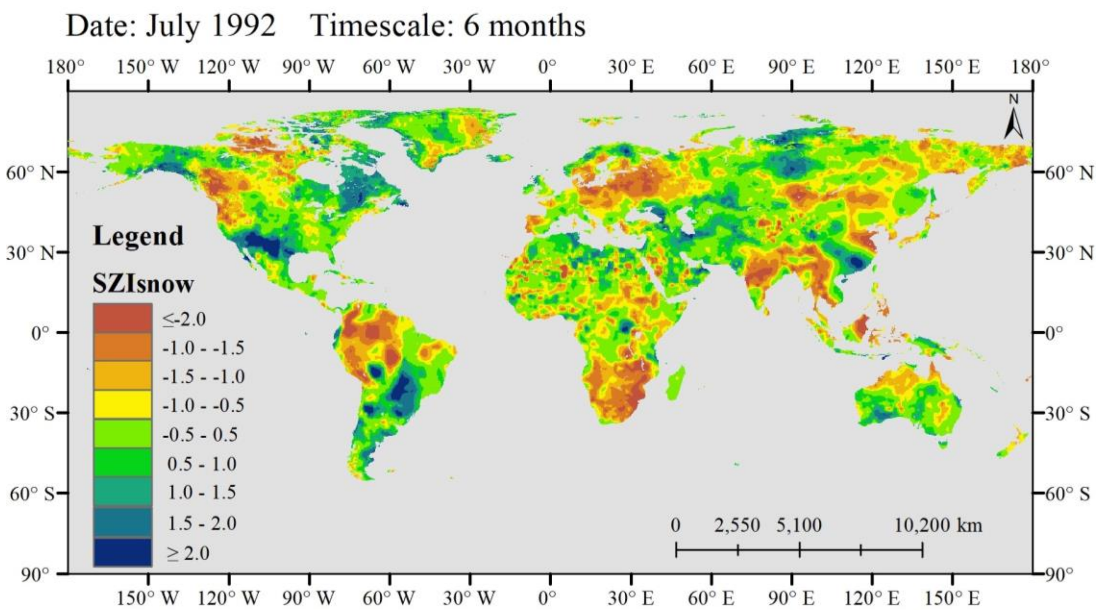


Figure R5. Spatial distribution of the SZI_{snow} on July 1992 at a 6-month timescale.

4) If possible, provide some scripts for potential readers to plot your data.

Reply: Following the comment, we will provide two scripts for potential readers to plot the SZI_{snow} data in the metadata file. The first script is coded based on Python programming language, and another is coded based on the NCAR Command Language (NCL).