Reviewer 1

Peng et al. generated a new SPEI drought index by refining the calculation method of potential evapotranspiration (PET), incorporating land surface characteristics driven mainly by leaf area index (LAI). They found that this new SPEI index has a higher correlation with surface moisture data and can explain 29% more variability within soil moisture. The improved index demonstrated good performance in humid regions and forest-dominated ecosystems, making the topic interesting. The manuscript is well-written; however, some concerns regarding methodology and evaluation remain evident. In general, I am favorable to the publication of the manuscript after a thorough revision.

Response: Thank you for your very positive evaluation.

1. **First, it appears that the evaluation throughout the paper relies on the correlation coefficient of the entire time series of SPEI and soil moisture. The increment in the correlation coefficient is almost less than 0.1, even if statistically significant. Since drought indices are typically used to identify and quantify drought events, I suggest the authors evaluate the skill of their improved SPEI index in detecting and quantifying extreme events rather than the dynamics of the entire time series.**

   Response: Thank you for the great suggestion. Our study primarily aims to assess the overall improvement in predicting of temporal variations in drought indices and soil moisture. While the absolute increments in correlation appear modest, the percentage change is quite significant, around 25-30%. Despite the small average increase, local improvements are notable (as shown in Figure 4). We acknowledge the above and the importance of capturing the extreme events in discussion section 6.4: "While the absolute improvements in correlation with soil moisture appear modest, they represent significant percentage changes of 25-30% and notable local improvements. We acknowledge the need for evaluation of the effectiveness in addition to the temporal correlations. Specifically, future studies should evaluate the capability of the land cover specific approaches to accurately capture extreme events."

2. **Secondly, I observed that many Ga and Gs parameters (in Table 1) have been used to incorporate features of aerodynamic and surface conductance. I wonder if substantial uncertainty arises from these prescribed parameters. In other words, does the subpar performance of the improved SPEI index in non-forest ecosystems relate to larger uncertainties in parameters for grassland, shrubland, or cropland compared to the forest?**

   Response: We agree that the uncertainties in the Gs parameter can potentially affect the results. For example, the better performance of the tall crop reference ET compared to the Land Cover approaches for the non-forest ecosystems suggests inaccuracies in the Gs parameters, given the similar big-leaf model. We compare Gs among the approaches and note the parameter uncertainty in the discussion section 6.2: "Given that the RC-tall method—a similar big leaf model—performs better than LC in these areas (Fig.4), it suggests that uncertainties in LC's Gst_{max} could result in these outcomes. Additionally, a comparison between Gst_{max} and Rst_{min} (used in SW) highlights uncertainties in this parameter. For instance, Rst_{min} in shrublands, grasslands, and savannas ranges from 100-180 s m^{-1} (equivalent to Gst_{max} of 5-10 mm s^{-1}), which is generally lower than 9-12 mm s^{-1} reported by Kelliher et al. (1995). These findings highlight the need for in-situ measurements of surface conductance in these areas."

3. **Thirdly, the improved SPEI exhibits better performance in humid regions, which aligns with expectations given the energy-limited water availability dynamics. However, in arid regions where water availability is more supply-dependent, the adjustment to PET has no significant effects and the uncertainty in precipitation data may be crucial. The authors may elaborate on this point in the manuscript.**

   Response: Thank you for pointing out the influence of aridity on our results. We acknowledge this in section 6.2: "In the meantime, these areas are located in the arid regions (Fig.7), the improvements of PET do not have significant effects on modeling the soil moisture, and precipitation dynamics may dominate the soil moisture variations".

4. **Specific comment:**
   Figure 2: It is unclear whether the correlation between soil moisture and SPEI reflects temporal or spatial variability or includes both signals. Additionally, please clarify what the white dots within each bar represent.

   Response: The correlation reflects spatially averaged temporal variability between SM and SPEI. The white dots indicate the average difference in correlations between the four methods and the reference method (current Table 2). We will clarify this in the legend of Figure 2: "Differences in spatially averaged correlation (ΔR) of pairs of PET methods that share the same surface characteristics except for one of the surface features: surface roughness, canopy conductance, albedo, and overall consistency among the above features. The white dots indicate the average ΔR between the four methods and the reference method."

Reviewer 2

Peng et al.'s manuscript provides a valuable estimate of global potential evapotranspiration (PET) and forms the basis for developing the SPEI index. The authors incorporate more realistic vegetation characteristics, such as Leaf Area Index (LAI) and conductance, to enhance PET estimation. However, some sections of the manuscript, particularly the structure and descriptions, could benefit from further clarity. The novel aspects of the PET calculation method should be more distinctly highlighted or enhanced.

Response: Thank you for your very positive evaluation.

5. **A more detailed description of the "two-source model" in Section 3.3 would be beneficial. The manuscript does not clearly articulate the relationship between this model, Equation (13), and the improved vegetation characteristics described in Section 3.2. The statement "We adopt the same parameterizations detailed in Zhou et al. (2006)" is too vague. It would be valuable to elaborate on how these parameter improvements are integrated into your PET method.**

   Response: Thank you for the suggestion. In order to clarify the parameter improvements for different models, 1) we add a detailed description of the two-source model in Appendix A; 2) we separate the original Section 3.2 into two parts, (a) surface characteristics and (b) parameterizations of surface characteristics, to elaborate how these parameters are integrated into different PET methods.

6. **The manuscript estimates PET over 1981–2017. This timeframe should be explicitly mentioned in Sections 2 and 3, such as "PET is estimated over 1981–2017 using [specific methodology]."**

   Response: We add the timeframe in Section 2.1 (P3): "To calculate the SPEI, PET is estimated on daily scale over the period of 1981-2017 using high-quality daily meteorology data from PRISM (Parameter-elevation Regressions on Independent Slopes Model) that employs weather stations and digital elevation model (Daly, Neilson, & Phillips, 1994; Daly et al., 2008).", and in Section 4.1 (P11): "We integrate the PET methods into the SPEI drought index across 1-, 3-, 6-, and 12-month time scales over the CONUS for the period of 1981-2017."

7. **Clarify whether PET calculations are based on monthly or daily scale meteorological inputs. The application of land surface ancillary data in your equations, such as the usage of "black- sky and white-sky albedo," is not clearly explained. For instance, how is albedo factored into the net radiation calculations in your equations?**

   Response: PET calculations are based on daily meteorological inputs, which is clarified in Section 2.1 (P3) as mentioned in #6. Regarding the processing of GLASS albedo, we resample the 8-day albedo product to a daily resolution, average the black- and white-sky albedos, and implement gap-filling for missing data using the average of adjacent years.

We add this detail in Section 2.3: "We resample the 8-day albedo to a daily resolution and obtain daily albedo by averaging the black- and white-sky albedos. Missing data are gap-filled using the average of adjacent years."

8. **On L121, you mention obtaining "canopy height data from a global tree height dataset at 1- km for 2005 using spaceborne lidar." It seems not clear how this dataset is used in your study? You also state that "As canopy height and frictional velocity are rarely measured continuously for each grid, we use a simple look-up table approach to provide roughness parameters." These statements seem contradictory and need clarification.**

   Response: We acknowledge the confusion of using both approaches in this study. We add a new section 3.2.3 Canopy height to clarify the combined usage of global tree height dataset and the literature values for roughness parameters.

   "Canopy height (h) is a key parameter in determining aerodynamic conductance. The OW and FAO methods generally assume it to be constant across vegetation types and temporal scales. To address this limitation, we introduce two methods for estimating canopy height. The first method, eventually used to obtain d_0 and z_{0m} for Eq.9, determines canopy height based on land cover type by calculating the median height within each land cover from the global tree height dataset. The second method, applied in the SW two source model (Appendix A, Eq. A9-10), takes into account both land cover type and dynamic LAI. Each land cover type has a range for canopy height defined by the minimum canopy height (h_{min}) and maximum canopy height (h_{max}). The actual canopy height is then determined by assuming a linear relationship with LAI following Zhou et al. (2006).

   $$h = h_{min} + \frac{(h_{max} - h_{min})LAI}{LAI_{max}} \tag{13}$$

   where LAI_{max} represents the annual maximum value at the grid cell level, obtained from the satellite data. Note that h is set to zero if LAI_{max} is zero."

9. **Section 3.1 lists different PET methods, most of which are derived from the Penman equation. Including the derivation process in the supplementary material and schematic figures illustrating the differences between these methods (e.g., big leaf models vs. two-source models) would enhance understanding. This suggestion is optional if it's difficult to implement.**

   Response: Thank you for the suggestion. We reorganize Section 3.1 by introducing all the existing PET methods at the beginning, including the SW model. We also explain the details of the SW model in Appendix A, as mentioned in #5.

6. **In Section 3.3.3, clarify the role of Gstmax in previous PET methods or equations mentioned earlier.**

Response: We clarify the role of Gstmax in previous PET methods in Section 3.2.2: "In previous PET methods, surface conductance is either not considered or assumed to be constant across vegetation types and over time. LAI plays a dominant role in determining the canopy-atmosphere coupling and ET partitioning (Peng et al, 2019; Wei et al., 2017; Forzieri et al., 2020). The OW and PT approach does not consider the role of LAI. The FAO approach uses a constant LAI throughout the growing season. Here we adopt a widely used method in estimating actual ET and assume a well-watered condition."

And "We introduce two options to incorporate an average LAI or the seasonal cycle of LAI into the surface conductance."

7. **While many surface vegetation characteristics are included to improve PET estimations, some easily accessible characteristics are not utilized. Global canopy vegetation height data (https://www.nature.com/articles/s41559-023-02206-6), which could be employed in Ga estimation, is now available. Other datasets like the 1k datasets (https://essd.copernicus.org/preprints/essd-2023-242/) may also be valuable for your study.**

   Response: Thanks for the recommendations. These datasets are useful for future improvements of our approach. We add this in Section 6.2: "In addition, future improvements to our approach could benefit from incorporating newly available datasets such as Lang et al. (2023) for canopy height."

8. **A recent study "Sun, S., Bi, Z., Xiao, J., et al." (2023) considers comprehensive parameters for improved PET estimation. If detailed consideration of vegetation characteristics is a novelty of your study, please explicitly explain its advantages compared to this study. Alternatively, if your focus is more on comparing different PET methods with limited vegetation considerations, clarify this in your introduction and discussion.**

   Response: Thanks for pointing out this new study focusing on the Shuttleworth-Wallace model. We clarify the novelty of our study compared to this study in the introduction: "A recent study by Sun et al. (2023) highlighted the importance of incorporating surface properties especially vegetation control in PET and used a two source model designed for sparse vegetation surfaces. However, the model's broader applicability beyond sparse vegetation is uncertain, and additionally it may increase data requirements and associated uncertainties." The advantage of our approach has been illustrated in the end of the discussion: "Our approach is a compromise between the above two types of models, which is more realistic and process-based than the commonly used drought index while being easy-to-implement and less data-intensive than a land surface model."

9. **Compare your PET estimations with reference datasets, such as Sun et al. (2023).**

   Response: Our estimates align with this study and we note this in the discussion section 6.3: "The LC method not only provides modest absolute PET values (Fig.5a) but also displays better performance across many areas (Fig.6). Specifically, LC estimates an annual PET of roughly 1200 mm, consistent with PET estimations for the same region as well as temperate zone reported in a recent study (Fig.8 in Sun et al., 2023)."

10. **Appendix A contains important information leading to the results in Section 5.1. Mentioning this in your method sections would prevent sudden introduction of these comparisons in the results. Some sentences around L280 could be moved to the method section.**

    Response: Thanks for the great suggestion. We move the results and table A1 in original Appendix A to a new section 5.1 Initial assessment of surface characteristics. The original paragraph around L280 has been moved to methods, a new section 4.2, to ensure a smoother transition to the results. Most of the original section 5.1 has been moved to a new section 5.2 for clarity.

11. **Move Figure A1 to the results section. The results section should feature PET estimations before transitioning to SPEI comparisons (starting in Figure 2).**

    Response: We move the original Appendix A to current results, as addressed in #10.

12. **Incorporate multiple soil moisture datasets in your comparison to account for the significant uncertainties among different soil moisture data.**

    Response: The ESA CCI SM v4 dataset is chosen for its widely accepted data quality, which is achieved by combining multiple single-sensor active and passive microwave soil moisture products to minimize uncertainty. Gruber et al. (2019) provides a more comprehensive understanding of the data accuracy. We add this in Section 2.2: "The dataset is chosen for its enhanced data reliability by integrating multiple single-sensor active and passive microwave soil moisture products to minimize uncertainty (Gruber et al., 2019)."

13. **On L329, introduce the full name 'LC-Kelliher' before its abbreviation. LC is "land cover" as detailed in the table of Figure 3. Please check the manuscript for any potential similar issues.**

    Response: We clarify the two land cover (LC) parameterizations for surface conductance in section 3.3: "To calculate surface conductance in Eq. 11-12, we provide two set of parameterizations based on land cover type. The first set is derived from the findings of Kelliher, Leuning, Raupach, & Schulze (1995)… The second set uses the minimum stomatal resistance Rst_min, following Zhou et al. (2006)." We also add a description in a new section 4.3 Comparision of PET parameterizations: "The LC method uses the same aerodynamic conductance method (Eq. 9) but differ in their surface conductance parameterizations: LC-Kelliher, which adopts Gst_{max} from Kelliher, Leuning, Raupach, & Schulze (1995), and LC-Zhou, which uses Rst_{min} from Zhou (2006)."

14. **On L61, provide examples of "conventional PET methods" versus non-conventional methods for clearer understanding. Regarding the statement "The vegetation**

control on transpiration is often neglected," comment on the impact of plant hydraulics on potential transpiration estimation, referencing relevant studies (e.g., https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018MS001500).

Response: In the line mentioned, we differentiate between "conventional" PET methods, which often assume no or simple universal vegetation control on transpiration, and "non-conventional" methods that account for vegetation control based on specific conditions. For instance, conventional methods that are often used in SPEI include the Thornthwaite and Hargreaves-Samani equations, as well as the Penman open water or Reference crop ET formulas. The "unconventional" methods in this study do not refer to the land surface models or dynamic vegetation models, which normally have representations of the transpiration process including plant hydraulics. This is clarified in L61.