Response to Reviewer #1:

This study applies multiple methods to identify critical soil moisture (CSM) that separates water- and energy-limited regimes using several satellite-based data and in-situ observations with a specific spatial scope. Then, it explores the factors that dominate the variations using a feature regularization technique. I support this study as I think their analyses on CSM and the determinant factor advance the science on the water and energy cycle over land. However, I think the readability of the paper and the description of the analyses should be further improved. Moreover, I have concerns about the confidence in the CSM estimates. I suggest a major revision. Please see my comments below:
I do not wish to remain anonymous – Hsin Hsu

Response: We thank Dr. Hsu for the positive evaluation and constructive comments. We have carefully revised the manuscript to address the concerns and to improve the paper. Specifically, we are very grateful to Dr. Hsu for asking professional questions and providing the latest research literature. Based on these comments, we have filled the statistical knowledge gaps of critical soil moisture (CSM), such as the application of the BIC and Chi-square test, and so on. For Figure 3, we first added the BIC method to evaluate the performance of three different linear models and further constrained the occurrence of multiple critical values. For Figure 4, the chi-square test was then introduced to test the differences between the models. In addition, the previous figures were well aligned among different products and soil layers, so we improved them to ensure that each figure showed clear differences and could provide a clearer conclusion. For the results of CSM, we added the following constraint: "in the entire period, negative metrics are displayed when SM is less than CSM, and positive metrics are shown when SM is greater than CSM. If there is more than one value where SM shifts between positive and negative metrics, CSM is treated as unidentified." Further, the performances of the regression model were assessed by five-fold cross-validation using the mean absolute percentage error. Temperature and VPD were considered in the attribution analysis. The comments are shown in grey shadow; our reply is shown in regular text and all the changes are highlighted. The main responses are as follows:

Major Comments

1. The article is very hard to read because the amount of **abbreviation is overwhelming**. I had to look up what an abbreviation stands for **multiple times** in just one sentence as they are from different categories (variables, locations, names of in-situ sites, land cover types, algorithms, statistical parameters, data products, etc.). I suggest **retaining the full names** for land cover types and some algorithms in the writing, as many do not occur frequently in the paper. The authors could also consider separating the abbreviations by different systems. For example, use Greek alphabets for statistical parameters and italics for variables.

The main methodologies used in this study may need abbreviations due to their lengthy names (e.g., Corr(ET,VPD) - Corr(ET,SM)). Sometimes they occur many times in one paragraph (or even one sentence), but the key information separating different correlation-difference methods is the second variable used in the first correlation calculation. The authors could consider **modifying the notation** of Denissen et al. 2020 to define:

ΔCorrVar = Corr(ET,Var) - Corr(ET,SM).

Response: Thanks for your suggestion. In the revision, we have substantially reduced the usage of abbreviations. We chose to retain the abbreviations of the eight satellite-based products, six main variables (ET, GPP, VPD, SM, EF, and CorrVPD), and site names. Acronyms for temperature and radiation in the relevant Correlation-difference method were deleted. Thus, we only kept CorrVPD = Corr(ET,VPD) - Corr(ET,SM). In addition, the full names of land cover types and sub-basins were used and the occurrence of land cover types and sub-basins was also reduced.

2. Most of the cited work on critical soil moisture and regime examination is published before 2022. There are many new aspects of regimes and CSM since 2023. Not required to reference but the authors could consider integrating these recent studies:

New method for calculating CSM based on satellite data:

Fu et al. (2024). Global critical soil moisture thresholds of plant water stress. https://doi.org/10.1038/s41467-024-49244-7

Global estimation of CSM based on soil moisture dry-down framework and an index to quantify vulnerability:

Dong et al. (2023). Land Surfaces at the Tipping-Point for Water and Energy Balance Coupling. https://doi.org/10.1029/2022WR032472

In the abstract (line 11), author mentions that regimes can shift under climate change. This is not discussed in the introduction:

Hsu, H., Dirmeyer, P.A. (2023). Soil moisture-evaporation coupling shifts into new gears under increasing CO2. https://doi.org/10.1038/s41467-023-36794-5

Hsu, H., Dirmeyer, P.A. (2023). Uncertainty in Projected Critical Soil Moisture Values in CMIP6 Affects the Interpretation of a More Moisture-Limited World. https://doi.org/10.1029/2023EF003511

Duan et al. (2023). Coherent Mechanistic Patterns of Tropical Land Hydroclimate Changes. https://doi.org/10.1029/2022GL102285.

Response: Thanks for your recommendation. We have added these studies in the Introduction and Methods section as follows:

Lines 39–44: "The ratio increases with increasing SM during water-limited regimes. This approach has limitations due to sparse eddy covariance observations (Feldman et al., 2019; Fu et al., 2022a) and poses challenges in adequately capturing comprehensive regional or continental-scale CSM and its variations (Dong et al., 2023; Hsu and Dirmeyer, 2023b). Globally, some model-based analyses used the ratio of LE to net radiation (Seneviratne et al., 2010, Schwingshackl et al., 2017), surface temperature diurnal amplitude (Feldman et al., 2019; Fu et al., 2024), and LE (Hsu and Dirmeyer, 2023a; Duan et al., 2023)."

Lines 209–210: "The alignment of CSM obtained by different methods was determined using the Chi-square test (McHugh, 2013; Hsu and Dirmeyer, 2023b)."

3. The robustness of each technique for estimating CSM is not described:

From **Figure 4**, it seems that CSM can be identified in almost all grid cells by the covariance method. **Is this also true for all other methods?** If not, what is the **rate of agreement** on the existence of CSM among all the methods?

Response: No, CSM cannot be identified in all grid cells. Thanks for reminding me. In the previous version, we did an average for each tipping point in each year and treated it as a critical value, which will reduce the difference between different products. In the revised version, we considered that only one tipping point in each year is the critical value.

We added the constraints in Method:

Lines 163–165: "For each grid cell and the entire period, negative metrics were displayed when SM was less than CSM, and positive metrics were shown when SM was greater than CSM. If there was more than one value where SM shifts between positive and negative metrics, CSM was treated as unidentified."

We added the disparity of soil moisture regime and percentage of area with $p < 0.05$ in Result:

Lines 279–281: "The number of wet binary bit was used to quantify the agreement among eight ET and GPP-based models at 10 cm soil depth. If CSM was identified, the SM wetter-than-CSM was assigned as 1, and others as 0. If CSM was not identified within a year, the digits of the mode were treated as 0. If CSM was not detected for all 18 years, it was displayed as empty."

For Figure 4, the agreement and chi-square test were introduced to test the differences between the models:

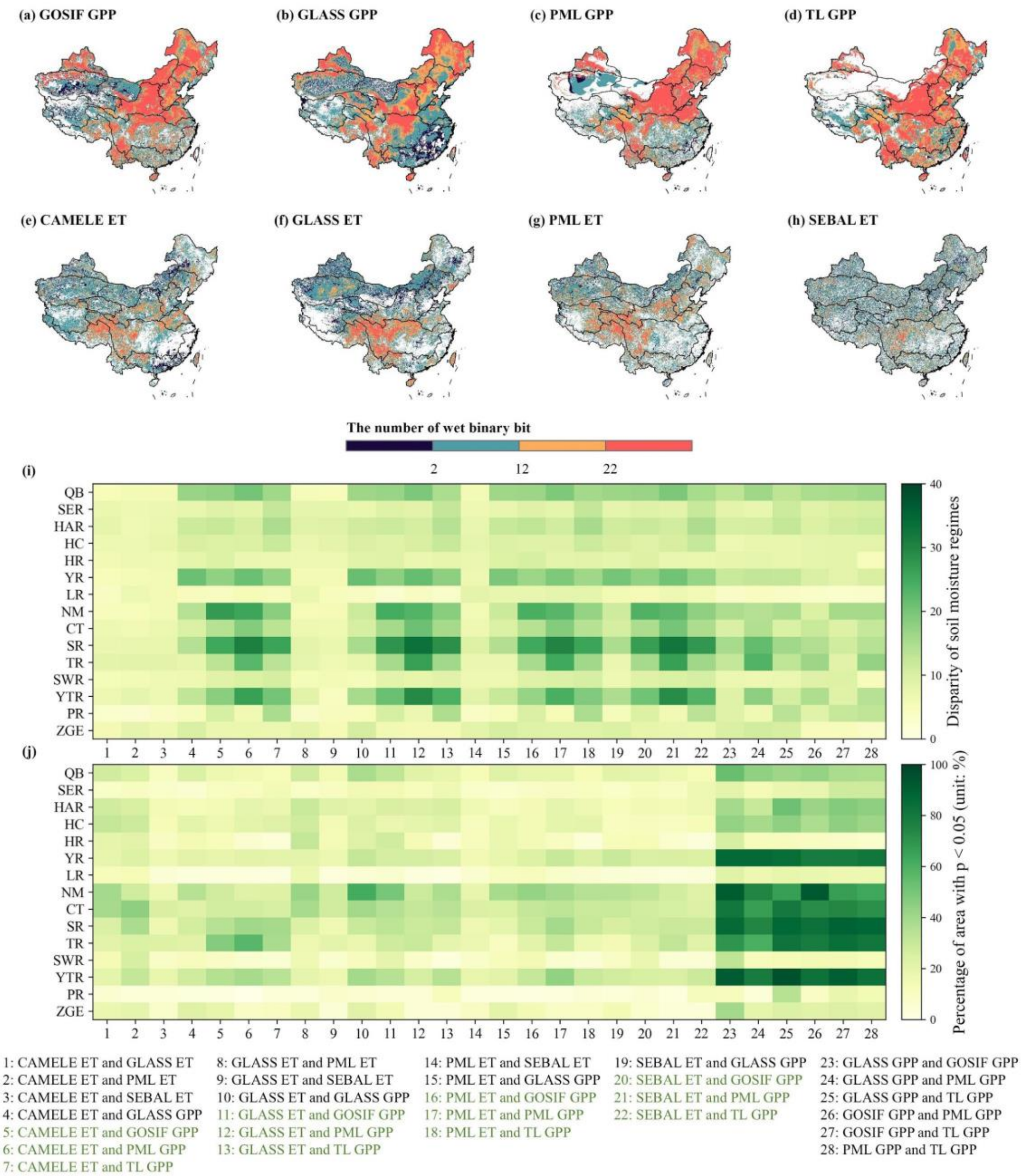

Figure 4: Spatial pattern of number of wet binary bit at 10 cm depth using covariance between vapor pressure deficit (VPD) and gross primary production (GPP) from (a) GOSIF, (b) GLASS, (c) PML, (d) TL. CSM using correlation-difference metric with Kendall's rank correlation (Corr) between detrended anomaly soil moisture (SM) and evapotranspiration (ET) from (e) CAMELE, (f) GLASS, (g) PML, and (h) SEBAL and Corr between detrended anomaly VPD and those ET products. (i) Disparity of soil moisture regimes among all the methods and (i) their percentage of area with $p < 0.05$. ZGE: Zhungaer Basin, PR: Pearl River Basin, YTR: Yangtze River Basin, SWR: Southwestern River Basin, TR: Tarim Basin, SR: Songhua River Basin, CT: Changthang Region, NM: Inner Mongolian Plateau Region, LR: Liaohe River Basin, YR: Yellow River Basin, HR: Huaihe River Basin, HC: Hexi Corridor Region, HAR: Haihe River Basin, SER: Southeastern River Basin, QB: Qaidam Basin.

Do all methods for estimating CSM have significant tests? For example, when using the SM-EF method, it is common to use three different linear models (flat line, a positive slope line, and a linear-plus-plateau) and select the best one based on the **Bayesian information criterion (BIC)**. If flat-line regression or positive-slope regression outperforms linear-plus-plateau regression, CSM should be considered as not identified. This procedure is also used for detecting CSM by the soil moisture-drydown framework.

Response: Thanks for the question. No, we didn't have significant tests in the previous version. Thus, for Figure 3, we have added the BIC method to evaluate the performance of three different linear models and further constrained the occurrence of multiple critical values.

We added the description in Method:

Lines 183–186: "In addition, Bayesian Information Criterion (BIC) (Schwarz 1978) was used to select the best fit among the three-segmented regression candidates (the flat line, the positive slope line, and the linear-plus-plateau). If SM was greater than or less than the CSM in some regions, the relationship between SM and EF appeared a flat line or a positive slope line. If the flat-line regression or the positive-slope regression outperformed the linear-plus-plateau regression, CSM was considered as not identified."

We added the description in Results:

Lines 252–255: "Overall, the linear-plus-plateau regression with the lowest BIC outperformed the flat line and the positive slope line in the study period of all eight sites, indicating that CSM can be identified. Specifically, CN-Du2 and Qianyanzhou sites showed a great slope at low SM values with BIC of -80.29 and -98.64, respectively."

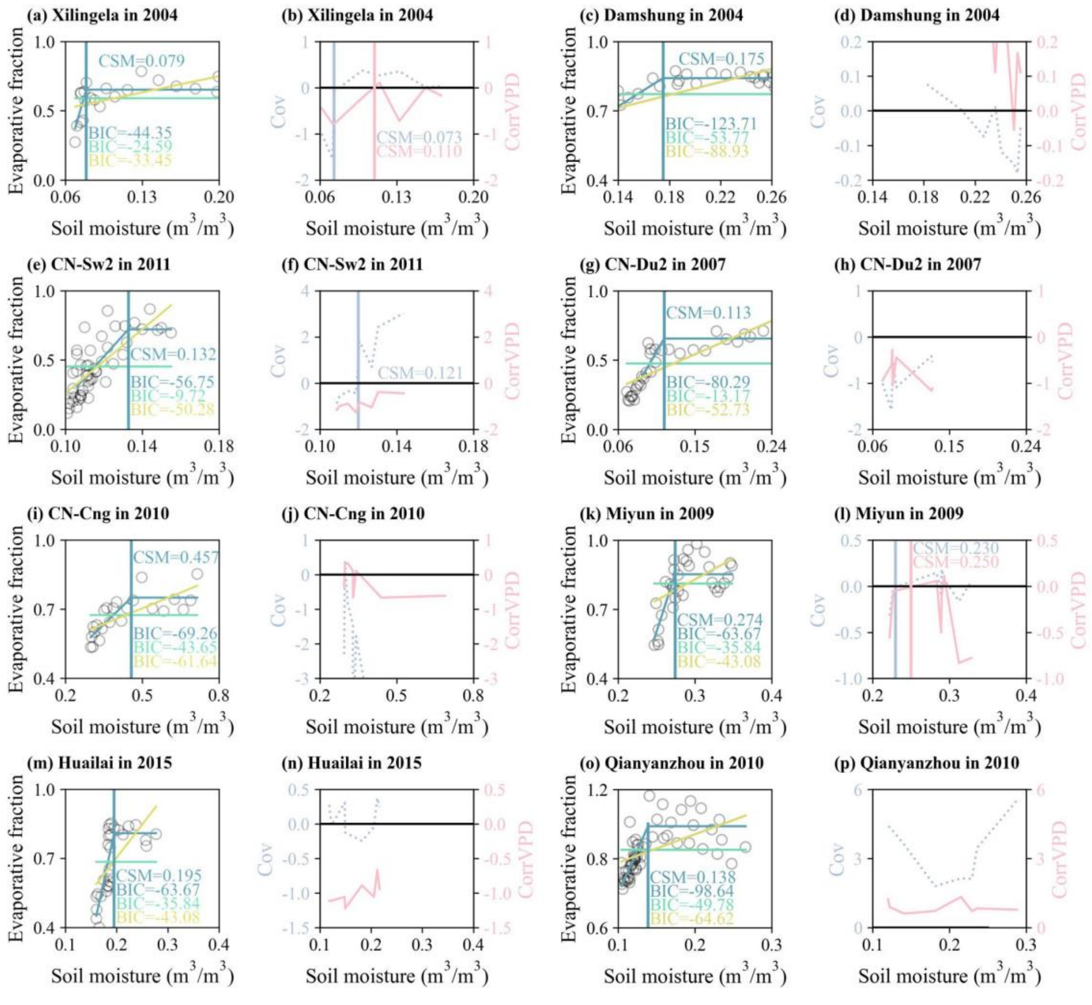

Figure 3: Variations of soil moisture and evaporative fraction at (a) Xilingela in 2004, (c) Damshung in 2004, (e) CN-Sw2 in 2011, (g) CN-Du2 in 2007, (i) CN-Cng in 2010, (k) Miyun in 2009, (m) Huailai in 2015, (o) Qianyanzhou in 2010. Variations of covariance (referred to as Cov) between vapor pressure deficit and gross primary production, and correlation-difference metric (referred to as CorrVPD) at (b) Xilingela in 2004, (d) Damshung in 2004, (f) CN-Sw2 in 2011, (h) CN-Du2 in 2007, (j) CN-Cng in 2010, (l) Miyun in 2009, (n) Huailai in 2015, (p) Qianyanzhou in 2010.

When using the correlation-difference method, if there is more than one SM value where the correlation-difference is zero, which SM value is identified as CSM? In **Figure 3b**, the red line locates at the wetter SM value when the correlation-difference is zero, but blue-greenline locates at the drier SM value when the correlation-difference is zero. This should be clarified.

Response: In the revised version, we consider that only one tipping point in each year is the critical value. We added the constraints in Method:

Lines 163–165: "For each grid cell and the entire period, negative metrics were displayed when SM was less than CSM, and positive metrics were shown when SM was greater than CSM. If there was more than one value where SM shifts between positive and negative metrics, CSM was treated as unidentified."

We added the description in Result:

Lines 264–268: "For CN-Du2 and Qianyanzhou sites, only positive or negative Cov and CorrVPD were found. For Damshung, CN-Cng, and Huailai sites, we found more than one SM value where the Cov or CorrVPD is zero. Along with surface soil wetting, there was a change of Cov and CorrVPD from positive to negative at these sites, inconsistent with the transition from water to energy limitation, indicating that CSM was not identifiable."

Additionally, I assume that if either correlation is **not statistically significant** when calculating the correlation before taking the difference, CSM should also be treated as not identified. Is this the case in this study?

Response: No, since each set of correlations has only 9 values, satisfying a significant correlation at all times is hard to achieve. Thus, although no significant correlation is verified, we added the following constraint:

Lines 163–165: "For each grid cell and the entire period, negative metrics were displayed when SM was less than CSM, and positive metrics were shown when SM was greater than CSM. If there was more than one value where SM shifts between positive and negative metrics, CSM was treated as unidentified."

Lines 260–261: "To explore the performance of both methods on sites and whether they can be used on a large scale, the data applied to both methods was averaged for 8 days, consistent with gridded data with the 8-day time scale."

4. When examining the alignment of CSM between different methods, **statistical significance** is needed. I recommend a **Chi-square test** as it can address scenarios involving categorical data: comparing rates or proportions between two groups when the outcome is a binary variable, such as negative and positive outcomes. In this specific case, categorical data represent soil moisture (SM) values tagged as "drier-than-CSM" and "wetter-than-CSM". So, a Chi-square test can be used to compare the proportions of SM values below and above CSM between two sets of variables or groups (obtained by different methods). If there are significant differences, it means the CSM is different.

Response: Thanks for your suggestion. We have added statistical significance in Figure 4 shown in the previous question and added the chi-square test to the Method and description of the Results.

We added the description in Method:

Lines 209–213: "The alignment of CSM obtained by different methods was determined using the Chi-square test (McHugh, 2013; Hsu and Dirmeyer, 2023b). SM values were divided into two groups below and above CSM. In this case, categorical data represented SM values tagged as a binary variable of 0 "drier-than-CSM" and 1 "wetter-than-CSM". If there were significant differences with a 95% confidence level, it meant that the CSM was different. CROSSTAB in MATLAB was used to perform the chi-square test."

We added the description in Result:

Lines 283–288: "Figure 4 shows the strong disparity in North and Central China, especially in Inner Mongolian Plateau Region, Songhua River Basin, Yangtze River Basin, and Yellow River Basin. In these regions, the chi-square test showed significant differences among GPP-based models due to their large number of wet binary bit. In addition, GLASS GPP displayed no CSM value in Northwest China. Note that the SM wetter-than-CSM showed agreement in eastern and southern basins, such as Huaihe River Basin, Liaohe River Basin, Southeastern River Basin, and Pearl River Basin, indicating that the ET and GPP-based models were consistent in these basins."

5. Figure 5 seems problematic. The CSM is extremely well aligned among different ET products. However, I assume the spread in temporal variation among ET products over many locations based on Figure 2a, where the correlation of each product's ET to in-situ data can be very different. Does that not lead to a different CSM estimate? The authors could provide some supporting information to justify the **consensus**, which looks too good. The tick labels on the y-axis in each bar chart are incorrect. The bars are spatial means, so **error bars** should also be provided.

Response: Thanks for the question. Yes, products lead to a different CSM. This alignment comes from two reasons:

(1) the first is that we have done an average for each tipping point in each year and treated it as a critical value, which will reduce the difference between different products. In the revised version, we consider that only one tipping point in each year is the critical value. We added the constraints in Method "2.2 Determination of CSM":

Lines 163–165: "For each grid cell and the entire period, negative metrics were displayed when SM was less than CSM, and positive metrics were shown when SM was greater than CSM. If there was more than one value where SM shifts between positive and negative metrics, CSM was treated as unidentified."

In the response to question 3, we can see that in the process of SM from low to high, if there are multiple times that metrics (Cov and CorrVPD) change from negative to positive, it means that this CSM does not conform to the change process of water stress to energy stress caused by SM from low to high.

We can see that the critical values obtained for the eight products in Figure 5 also show varying undetected areas.

(2) then, the spatial variation in the study area is too great, and the CSM and SM in spatial and depth are represented by a gradual color, resulting in the depth difference is not obvious compared to the spatial difference. The color of the previous drawing can only reflect the difference in space, not the difference in depth. Therefore, we averaged the CSM and SM by land cover and soil texture, highlighting differences in SM and CSM at different depths for each type.

After the constraint was added in question 3, we modified the code, redrawn Figure 5 and added the error bars, and redrawn Figure 6.

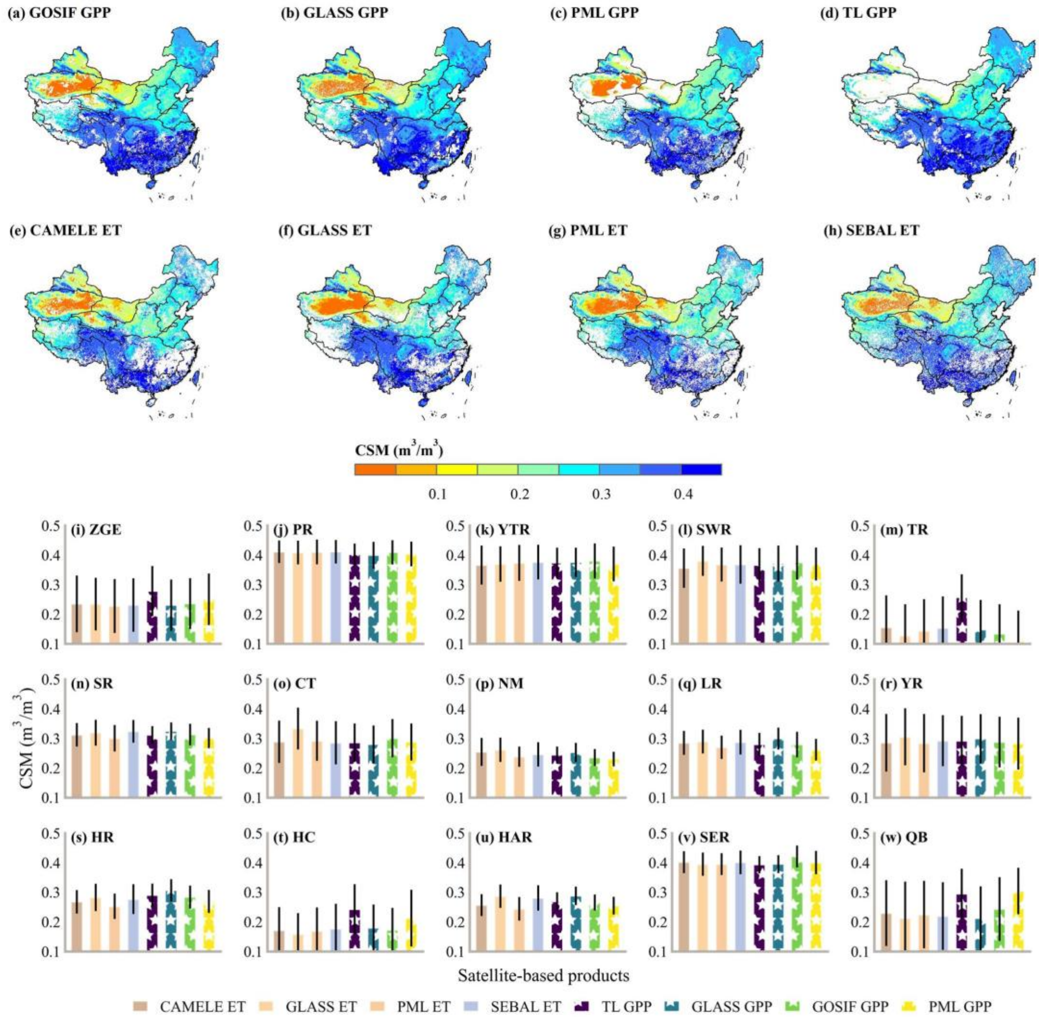

Figure 5: The spatial pattern of critical soil moisture (CSM) at 10 cm depth using covariance between vapor pressure deficit (VPD) and gross primary production (GPP) from (a) GOSIF, (b) GLASS, (c) PML, (d) TL and CSM using correlation-difference metric with Kendall's rank correlation (Corr) between detrended anomaly soil moisture (SM) and evapotranspiration (ET) from (e) CAMELE, (f) GLASS, (g) PML, and (h) SEBAL and Corr between detrended anomaly VPD and ET. (i–w) the basin-average values of ZGE:

Zhungaer Basin, PR: Pearl River Basin, YTR: Yangtze River Basin, SWR: Southwestern River Basin, TR: Tarim Basin, SR: Songhua River Basin, CT: Changthang Region, NM: Inner Mongolian Plateau Region, LR: Liaohe River Basin, YR: Yellow River Basin, HR: Huaihe River Basin, HC: Hexi Corridor Region, HAR: Haihe River Basin, SER: Southeastern River Basin, QB: Qaidam Basin.

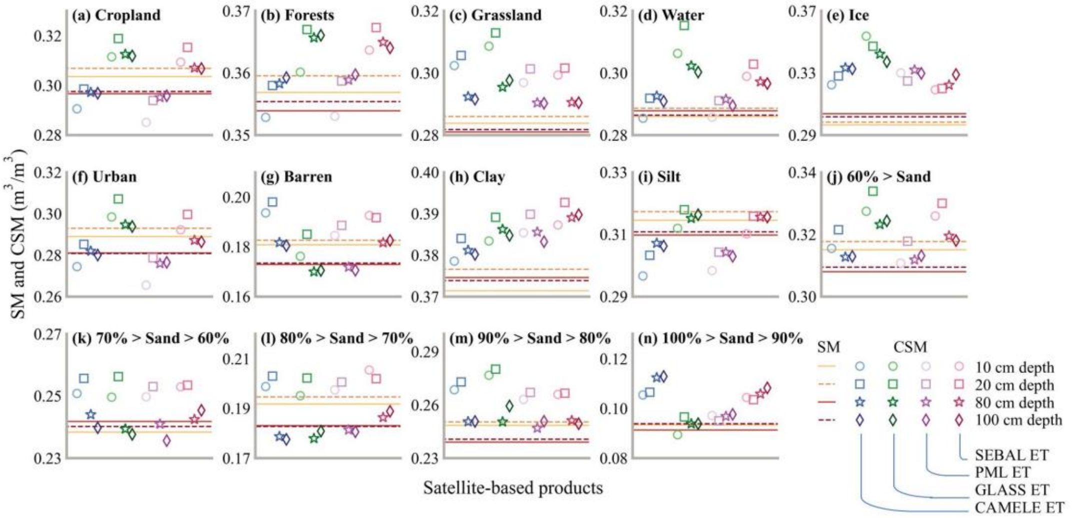

Figure 6: Soil moisture (SM) at (a) 10 cm, (b) 20 cm, (c) 80 cm, and (d) 100 cm soil depths across China, and critical soil moisture (CSM) derived from CAMELE, GLASS, PML, and SEBAL ET and SM at corresponding soil depths for (a) cropland, (b) forests, (c) grassland, (d) water, (e) ice, (f) urban, (g) barren, soils with a majority of (h) clay, (i) silt, and sand with content (j) less than 60, (k) between 60% and 70%, (l) between 70% and 80%, (m) between 80% and 90%, and (n) higher than 90%.

6. In Figure 6, the CSM among different SM layers is also **extremely well aligned** (if I interpret it correctly). This makes me doubt the reliability of SM in deeper layers. I assume SM-EF could be decoupled in deeper soils if roots do not reach that deep in some places, so there should be some inconsistency in CSM values.

Response: Thanks for the question. We improved the representation of Figure 6 to make the differences between the different depths more noticeable and changed the corresponding description as follows.

Lines 300–310: "Furthermore, large-scale CSM varied across vegetation types and soil textures at four soil layers (Figure 6). With shorter root systems and less vegetation (i.e., barren), areas with low CSM were water-limited. Forested regions displayed a relatively high CSM (e.g., 0.18 m3/m3 using PML ET and 10 cm depth SM). As for the soil types, sand covering the large area was further part into content of less than 60%, 60–70%, 70–80%, 80–90%, and higher than 90%. Soils with a majority of clay had a wetter CSM than others (e.g., 0.38 m3/m3 using PML ET and 10 cm depth SM) and was to be expected given that clay had a larger negative matric potential compared to coarse soil textures dominated by sand and silt. In summary, fine soils and luxuriant vegetation had wetter CSM. Additionally, a layer-wise CSM analysis was conducted to highlight variations in SM properties for different soil layers. It was evident that there were variations in the CSM behavior across layers with higher SM and CSM at 20 cm soil depth than at other depths. We also found that there was higher CSM than SM at all four layers for grassland and clay, which identified a large range of SM within the water-limited regimes. However, for cropland and forests, differences existed in the CSM among four ET-based methods, with higher CSM from GLASS and SEBAL than others."

In either case, I suggest the authors provide **additional analysis** to examine soil moisture at deeper layers (maybe taking some grid cells for example and put as supporting information) **and discuss** the uncertainty of using these data products in the discussion.

Response: According to the previous version, the results were too consistent, and additional analysis was required to demonstrate the differences between the different depths. However, according to our latest version, the differences in different depths have been shown by land cover types and soil textures as shown in Figure 6 in question 5. Therefore, instead of providing additional analysis, we have modified the original representation as shown in Figure 6.

Because this SM dataset is based on site observation data, the uncertainty of SM comes from the spatial interpolation, and the interpolation factors affecting different depths are different. Here we add a description to the Section Discussion:

We discussed the uncertainty of four soil layers in Discussion:

Lines 383–385: For gridded SM, surface climate shows a significant effect on the upper soil layer SM modeling, while the background aridity could lead to low variability of the deeper layer SM (Li, Q. et al., 2022)."

For example, some ET products are estimated; are the sources of input to derive ET independent of each other for all of them?

Response: No, CAMELE ET combined PML ET. We discussed the sources of input to derive ET in Discussion:

Lines 378–382: "ET and GPP exhibited great uncertainties (Liu et al., 2021) in areas with barren land as indicated in Section 3.1. Thus, eight satellite-based SM regimes were in good agreement in the eastern and southern regions (Figure 4) where satellite-based methods were more reliable. Secondly, since the CAMELE ET combined PML ET, they showed consistency in cropland and forests with a lower CSM than GLASS and SEBAL (Section 3.3)."

Does method to derive SM at different layer inherently lead to consistency in CSM?

Response: No, the CSM obtained from different layer SM is definitely different. However, the spatial variation in the study area is too great, and the CSM and SM in spatial and depth are represented by a gradual color, resulting in the depth difference is not obvious compared to the spatial difference. The color of the previous drawing can only reflect the difference in space, not the difference in depth. Therefore, we averaged the CSM and SM by land cover and soil texture, highlighting differences in SM and CSM at different depths for each type as shown in Figure 6 in question 5.

7. How does the author determine the set of input features for the ridge regression and why air temperature and VPD are not considered in the analysis?

Response: Thanks for the question. Given the similar information among radiation, temperature, and VPD, only radiation represented the atmospheric energy drive. In the revision, we have changed ridge regression to partial least square regression, added cross-validation as shown in Figure 7, and added temperature and VPD in partial least square regression as shown in Figure 8.

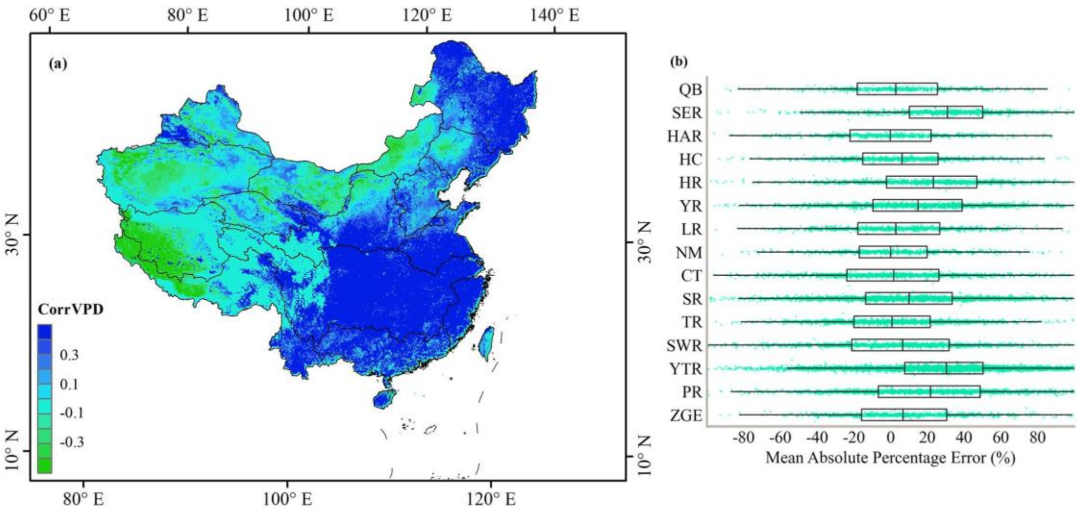

Figure 7: Spatial pattern of (a) CorrVPD from PML ET and 10 cm soil depth soil moisture and (b) the mean absolute percentage error based on partial least square regression for the predictions of CorrVPD. ZGE: Zhungaer Basin, PR: Pearl River Basin, YTR: Yangtze River Basin, SWR: Southwestern River Basin, TR: Tarim Basin, SR: Songhua River Basin, CT: Changthang Region, NM: Inner Mongolian Plateau Region, LR: Liaohe River Basin, YR: Yellow River Basin, HR: Huaihe River Basin, HC: Hexi Corridor Region, HAR: Haihe River Basin, SER: Southeastern River Basin, QB: Qaidam Basin.

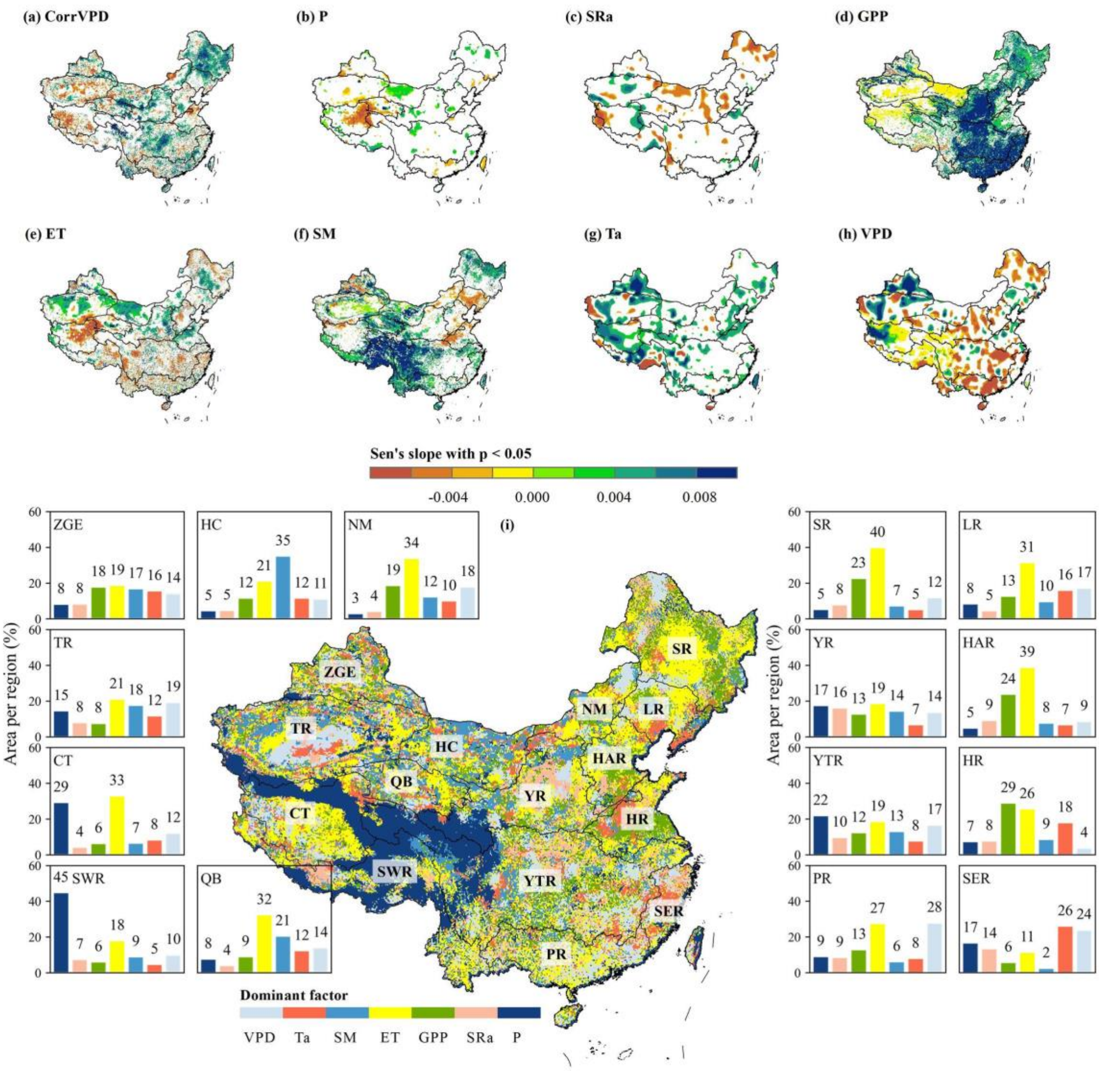

Figure 8: The sen's slope with p<0.05 of normalized (a) CorrVPD, (b) precipitation (P), (c) incoming shortwave radiation (SRa), (d) GOSIF gross primary production (GPP), (e) PML evapotranspiration (ET, (f) soil moisture (SM), (g) temperature (Ta), and (h) vapor pressure deficit (VPD) during the CorrVPD detection using the Mann–Kendall test (Mann, 1945; Kendall, 1948). Attribution of CorrVPD variations from PML to land-atmosphere variables. Colors indicate the variables that best predict the CorrVPD dynamics. ZGE: Zhungaer Basin, PR: Pearl River Basin, YTR: Yangtze River Basin, SWR: Southwestern River Basin, TR: Tarim Basin, SR: Songhua River Basin, CT: Changthang Region, NM: Inner Mongolian Plateau Region, LR: Liaohe River Basin, YR: Yellow River Basin, HR: Huaihe River Basin, HC: Hexi Corridor Region, HAR: Haihe River Basin, SER: Southeastern River Basin, QB: Qaidam Basin.

Minor Comments

1. Line 213: should be "ea" not "ea."

Response: We have modified it in the text.

2. Line 122: The term REddyProc is not explained nor mentioned elsewhere.

Response: We have added it in the text as follows:

Lines 90–94: "Given the fact that Huazhaizi, Dashalong, Luodi, Arou, Guantao, Huailai, Miyun, and Daxing did not have GPP data, the REddyProc website (https://www.bgc-jena.mpg.de/5622399/REddyProc/) was used to calculate GPP. The REddyProc imported half-hourly net ecosystem exchange, LE, H, and meteorological measurements to partition net ecosystem exchange into GPP and ecosystem respiration."

3. I suggest putting the unit of variables in every chart.

Response: We have added the unit in all figures.

4. Is ΔCorr calculated monthly between June and September and then an annual mean is obtained?

Response: As mentioned in our response to question 3 of Major Comments above, we got at most one critical value per year, and then an annual mean was obtained.

5. The word "Slope" in figures 7 and 8 is confusing. Is this a temporal trend? What is the statistical significance?

Response: Yes, it is a temporal trend. We changed the "slope" to "Sen's slope" using "the Mann–Kendall test". The significance was added in Figure 8.

6. Does the author perform cross-validation or bootstrapping for ridge regression?

Response: In the revision, we have changed ridge regression to partial correlation and added cross-validation for partial least square regression as shown in Figure 7.