We thank Kendra Kaiser for reviewing our manuscript and for providing helpful comments. Below we respond (in bold type) to the reviewer's comments (in normal type).

This study synthesizes information collected from a dense soil moisture monitoring network in a watershed with deep loess soils. They have highlighted the limitations of using the topographic wetness index for understanding the spatial variability of soil moisture in places that do not have a shallow confining layer, and that the spatial variability in these hillslopes is largely driven by evapotranspiration. The introduction is well written, while the discussion could use some editing to highlight the main findings and the more clearly share the associated story. Below in my specific comments I give suggestions on how to edit and trim the discussion to help focus on the most interesting findings.

Thanks very much for these comments.

L23 – confusing since only part of the year is captured in these distinctions

Based on the seasonal patterns in soil moisture (Figs. 6a and 6b), it is evident that May to July and October to March were under dry and wet conditions, respectively. April was defined as the wet-to-dry transition month, and August and September were defined as the dry-to-wet transition months (L314-315), because their deviation from annual average are around 0 (Fig. 6b). These three months bridge the wet and dry conditions and it is challenging to precisely assign them to a specific wet or dry condition.

L114 – Gully land consolidation --- I'm not sure if this is necessary here, but I am curious why this is being done and if it is occurring in other watersheds in the area.

The Gully land consolidation is a project undertaken by the Chinese government in the Chinese Loess Plateau with an investment of approximately 4.6 billion dollars, aimed to address the problem of farmland losses caused by the local reforestation activities. Starting in 2000, the Chinese government undertook a large-scale reforestation effort to address the severe erosion problem in the loess plateau. In this process, certain agricultural lands in the hillslope were abandoned and dedicated to grass/tree planting or natural vegetation recovery. To prevent the adverse impact of farmland abandonment on local farmers' income, the gully land consolidation project was implemented alongside reforestation: filling and levelling the gullies with soils from adjacent hillslopes, converting them into new farmland suitable for crop cultivation. As a consequence, the topography of the catchments has been altered compared to natural undisturbed ones, especially in the gully areas, making them into a newly typical and widespread catchment type in the loess plateau. Although our study was not primarily focused on the hydrological changes resulting from this project, we think it is necessary to acknowledge this point to distinguish such a disturbed catchment from natural ones.

L187 – why the 20% trimmed standard deviation?

To avoid such similar confusion for readers, we will use non-trimmed standard deviation to calculate the temporal and spatial variabilities. Re-calculation using non-trimmed standard deviation gives similar results to our robust estimates of variability (20% trimmed standard deviations used now), thus it does not alter our conclusions. We will update the new results in Fig. 4 and Fig. 8.

L238 & 305 – suggest to edit sentences as to not start with "Fig. 2/6"

We will edit the sentence in L238: "*The seasonal changes in the average soil moisture for each depth for the hillslope ($\theta_{hillslope,jk}$) and the gully ($\theta_{gully,jk}$), respectively, are shown in Fig. 2.*"

We will edit the sentence in L305: "*The seasonal patterns in average soil moisture over the top 100 cm for the hillslope and gully sites ($\theta_{SEf,j0-100}$, $\theta_{NWf,j0-100}$, and $\theta_{gully,j0-100}$), and the deviations from their annual averages ($\delta\theta_{SEf,j0-100}$, $\delta\theta_{NWf,j0-100}$, and $\delta\theta_{gully,j0-100}$), are illustrated in Fig. 5.*"

L290 – figure 4 caption is a little unclear, it seems as though the figure labels should be before the description, e.g. a) depth of max trimmed stdv, b) depth of collapse. How are 90% of sites "collapsing" between 20-100cm and 82% of sites collapsing between 160 and 260?

We will relocate the figure labels to appear before the corresponding descriptions for Fig. 4, also for the other figures.

Fig. 4 displays the tallied histogram representing the frequency distribution in depth of maximum trimmed standard deviation (TSD) and in depth at which TSD collapses for 72 monitoring sites. The x-axis indicates the count of the monitoring sites. In panel (a), we observed that the maximum TSD occurred in the depth of 0-20 cm for 15 monitoring sites, in the depth of 20-40 cm for 20 sites, 40-60 cm for 8 sites, 60-80 cm for 17 sites, and 80-100 cm for 5 sites. Therefore, a total of 65 sites (=15+20+8+17+5) showed the maximum SD within the top 100 cm soil depth. As a percentage of the total number of sites (excluding those were measurements are truncated before 5m depth), this is 65/72 ≈ 90%. Therefore, we concluded that the depth of the maximum SD was between 0 and 100 cm for 90% of the sites. By the same way, we concluded that the depth at which the SD collapses was between 160 and 260 cm for 82% of the sites. We will change the x-axis title from "*Frequency*" to "*Count of sites*" to enhance the clarity. The percentage number will be a bit different after the result updated (re-calculation by non-trimmed SD), but this will not affect our conclusions.

L282-286 is a longer description of the content in L 316-321 which is easier to read and is associated with a mechanism, consider removing the former.

We will remove the content in L282-286. Additionally, we will exchange the order of Fig. 5 and Fig. 6, then include a simple sentence describing the tallied histograms to complement the main finding in L316-321: "*…, with wet-to-dry transitions around April and dry-to-wet transitions between August and September <u>(Figs. 5a-b). Together with the tallied histograms of the months in which the annual maximum and minimum soil moisture occurred for each soil layer (Fig. 6),</u> these results suggest that soil moisture in the top 100 cm of soil was at a minimum in the late spring and early summer, …*"

L298 – What is the color scheme in Fig 5 for?

We will remove the color scheme and fill it with single color for Fig. 5a and 5b, respectively.

Fig 6 could be easier to read as a vertically stacked figure so that months are aligned

We will re-stack the panels in a vertical way.

L341 – figure color descriptions should be in a legend or in caption, not in the text

We will remove figure color descriptions from L341 and add them in the caption of Fig. 7*: "<u>Deviation in volumetric soil moisture (0-100 cm, $\delta'\theta_{ij0-100}$) from monthly average across hillslope sites. The lighter colors (deviation close to 0) indicate smaller deviations from the average soil moisture, while darker colors indicate larger deviations. Blue and red colors indicate soil moisture above and below the average, respectively.</u>"*

L 349 – this is slightly misleading as written given that the max soil moisture is below 20% in the hillslopes. You get there in the paragraph, but the set up suggests that there is something different happening here than in the literature.

We will revise this paragraph considering the comments from you and reviewer #1: *"<u>The most widely reported model for describing the relation between spatial heterogeneity and mean soil moisture is a convex-upward parabola, with spatial variability peaking at intermediate values of soil moisture content (approximately 20%) (Brocca et al., 2010; Famiglietti et al., 2008; Jarecke et al., 2021; Peterson et al., 2019; Tague et al., 2010; Western et al., 2003). This convex parabola has been observed in loess catchments as well (Gao et al., 2011; Gao et al., 2015; Shi et al., 2014), where spatial variability peaked at soil moisture within 15%-20%. However, in a similar loess system, Hu et al. (2011) found that the spatial variability slightly increased with increasing soil moisture, even in wet conditions (20%-25%), indicating that a natural logarithmic curve might better describe the relationship between spatial variability and average soil moisture. In the Gutun catchment, the average soil moisture mainly concentrated between 5%-15%, which means we may only be observing the short rising segment of a convex parabola below the variability peak, or the middle</u>*

section of a logarithmic curve."

Figure 7 –an aspect layer as a basemap to help support the findings re: aspect and moisture patterns would be helpful; label the legend and edit the caption to start with "Deviation in VWC from ...

Thanks for this suggestion. We will add an aspect layer as a basemap in Fig. 7.
The legend will be labelled with "*Deviation in soil moisture*".
We will edit the caption: *"Deviations in volumetric soil moisture (0-100 cm, $\delta'\theta_{ij0-100}$) from monthly averages across hillslope sites."*

L374 – is this statement supported given the high amount of overlap between boxplots in Fig9? Is it statistically significant? There can be a pattern without being statistically significant, but it should be clarified.
We have tested the significance of difference with one-way ANOVA, the result is as follows:

Month	p-value	p-range	note
1	0.000715	p<0.001	***
2	0.000773	p<0.001	***
3	0.000912	p<0.001	***
4	0.00000863	p<0.001	***
5	0.0213	p<0.05	*
6	0.0191	p<0.05	*
7	0.0297	p<0.05	*
8	0.626	ns	ns
9	0.0000535	p<0.001	***
10	0.0000915	p<0.001	***
11	0.703	ns	ns
12	0.00180	p<0.01	**

For most of the months (not August and November), the differences in soil moisture between the two slopes were statistically significant (under dry conditions, May to July, the difference is less distinct). We have noted the result for significance test in the Fig. 9(a) and its caption. We will further note it in L374 in the revised manuscript: "*Fig. 7 also reveals a spatial pattern in soil moisture, with the NW-facing slope being much wetter than the SE-facing slope, which was also seen in Fig. 5a and is quantified in more detail, including significance test, in Fig. 9.*"

L410-412 – this is the thing that I find most interesting and broadly relevant, but clarify this is referring to spatial variability in the hillslopes (not the hillslope vs gully)

We will clarify this as suggested: *"We found no statistically significant correlation between TWI and the soil moisture patterns on the slopes for any soil depth, or averaged over the top 100 cm of the soils in each month. It is important to emphasize that we focus on the relationship between TWI or*

***aspect with soil moisture patterns at the hillslope scale, excluding the gully.* Soil moisture *at the catchment scale* is, however, markedly higher in the gully (Figs. 2-3), consistent with the high TWI values there."**

Section 4.3 –consider editing these first two paragraphs to be more integrated, e.g. L422-425 reads like results while L435-437 gets into the context that makes them interesting.

We will edit these two paragraphs to be more integrated.

L450 – confusing as written – most of the precipitation fell between driest and wettest months?

We will remove the entire paragraph in which the L450 is located as suggested by the next comment.

L449 – this paragraph isn't particularly compelling as written, it walks through calculations and results presented in the table. Move the table to results and retain the last sentence in the discussion.

We will remove this paragraph, and move the last sentence into the first paragraph of Section 4.3: "*Spatial patterns of annual soil moisture storage change (ΔS) at depths of 0-100 cm, 100-200 cm, and 200-300 cm are illustrated in Fig. 10. The average ΔS that we measured in the 0-300 cm soils for the entire catchment was 110 mm. This is broadly similar to the water balance estimate (Table 1), suggesting that the 0-300 cm soils account for most or all of the seasonal water storage in the Gutun catchment.*"

We will also fix a few minor errors in the annual water balance estimation resulting in slight discrepancies in the revised manuscript. This correction does not affect our conclusions.