RESPONSE TO REFEREES

Referee 1: Review V2 for "Classification of Large-Scale Environments that drive the formation of Mesoscale Convective Systems over Southern West Africa" by Nkrumah et al.
Overview: Compared to the initial version, the authors were able to make some improvements on the paper. However, after having received the responses, I do not feel that every question I had was fully answered. Furthermore, the authors should overall consider transferring some more of the explanations they have given in response to my questions to the main text. I therefore recommend minor revision.

Thank you for the in-depth comments and specific analysis advice the referee provided for our manuscript. This helped significantly improve our results' robustness and sharpen our discussion. Below, we give a point-by-point response to the comments with line references to the new manuscript version. We also provide a tracked changes document highlighting the differences from the previous version.

• General comments/questions on the responses

- The authors argued in their response that the "scope of the work looked at investigating environmental conditions favorable for MCSs over the region". One aspect of it includes under which conditions MCSs might be triggered in the first place and further develop in southern West Africa. Therefore, it was a bit disappointing to see that there was no attempt to address some of the general comments I had (e.g CIN). For instance, it is known that MCSs can develop in high-CAPE/high-CIN situations where high CIN inhibits a premature initiation of smaller-scale convection that allows CAPE to further build up. Once CIN breaks down or is overcome, e.g. through moisture convergence or convergent motions at elevated terrain, vertical wind shear becomes relevant for the consequent evolution. While this has been observed for the midlatitudes and also partly for the Sahel, MCSs southern West Africa may be initiated differently in a moister environment. I do believe that this aspect is missing in the paper and is not beyond the scope. My suggestion: Have a look into anomalies of CIN and moisture (flux) convergence the same way as CAPE.

Thank you for this comment, it is, of course, correct that anomalies of certain variables will be more strongly linked to what allows MCSs to be maintained and organized (as the reviewer said, wind shear) and what may allow or inhibit their initiation (as CIN). We cannot disentangle these different life stages here as we're not looking at MCS tracks, but we now include anomalies of CIN in the supplementary as suggested by the reviewer, finding that all nodes show decreased CIN over SWA on MCS days. This is included in the main text in section 4.2:

" *Overall, all nodes show positive CAPE and negative convective inhibition (c.f. Supplementary Fig. S4) anomalies for MCS days in parts of SWA, creating an environment sufficiently unstable to support the development of convection."*

Specific comments/questions on the responses

- On the question why 925 hPa specific humidity was replaced by TCWV: "As pointed out by the referee, we considered the TCWV due to its ability to represent the total gaseous water in the vertical column of the atmosphere which is influenced by the evolution of the humidity field. TCWV represents the precipitable water the atmosphere holds better than the humidity. We, therefore, had to show both since in the first instance (i.e. 925 hPa humidity) we were looking at an environment that is suitable for instabilities in the atmosphere, of which humidity forms a part." I think this needs to be added to the text then since there was no motivation given of why TCWV was suddenly used in Fig. 10 and not elsewhere.

Thank you for pointing this out. We now specify the link between low-level moisture anomalies and instability changes in Section 4.2:

"Overall, all nodes show positive CAPE and negative convective inhibition (c.f. Supplementary Fig. S4) anomalies for MCS days in parts of SWA, creating an environment sufficiently unstable to support the development of convection. The close alignment with regions of increased low-level humidity (Fig. 9) suggests increased low-level moisture advection as the main driver for these instability changes."

and added an introductory sentence as justification for the use of TCWV in Section 4.3 as follows:

"TCWV instead of single-level specific humidity is used here to capture the changes in total moisture available to MCSs under the different regimes."

- On the question how the authors determined the rainfall amount: "The rainfall amount was determined from rainfall snapshots of the ''high-quality precipitation'' (HQ) a field within the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG; Huffman et al. 2019) dataset. This has been included in the manuscript as follows: This can include the same MCS at several timesteps in a day. Corresponding rainfall snapshots were sampled from the ''high-quality precipitation'' (HQ) field within the Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG; Huffman et al. 2019) dataset.". Why did the authors use the HQ fields of IMERG only? If really variable "HQprecipitation" was used then the authors should have experienced large data gaps since PMW satellites alone cannot fully cover the region at a given timestep. Sure that the variable "precipitationCal" was not used instead?

Thank you for catching this. In previous studies, we indeed used the PMW product to match up cloud top temperature with rainfall, as for applications where it is not crucial to capture all MCS cases (the reviewer correctly points out that HQprecipitation only covers swaths), the PMW rainfall better conserves the spatial structure of the rainfall field and thus provides a cleaner cloud top temperature/precipitation field relationship. However, given that in this work we only use the maximum rainfall rather than its spatial structure within the MCS cloud envelope, we used the spatially continuous "precipitationCal" product to capture as many raining MCSs as possible. The full paragraph will be included in the manuscript as follows:
"Twelve years of MCS snapshots (2004–15) detected from Meteosat Second Generation 10.8 µm-band brightness temperatures (Schmetz et al., 2002, EUMETSAT 2021) are used to define MCS days in this study. Following (Klein et al., 2021), an MCS is defined here as a -50°C contiguous cloud area larger than 5000 km^2. We consider the MCS images every half hour, for which they are matched up with the half-hourly Integrated Multi-satellite Retrievals for Global Precipitation Measurement (IMERG; Huffman et al. 2019) dataset, using the merged microwave / infra-red ("precipitationCal") rainfall product. An "MCS day" is then defined as a day with at least one hour containing 5 simultaneously existing MCSs between 16 and 1900 UTC with maximum rainfall >5mm within the SWA domain."

- On why no MCS are seen in DJF in Fig. 5: "The focus of MCSs over the study area in this study is during the rainfall season of the SWA domain which mainly starts in March and ends in November. February recorded zero because it wasn't considered in the frame of this work.". Then Fig. 5 is misleading, and the x-axis should be truncated to the relevant months. Is this also the case for the numbers in Fig. 6? In any case, unless I missed it, this should be mentioned in the data section as well.

We apologize for the confusion as to which months were included in the analysis. For completeness, we now processed and included the MCSs for the full annual cycle (Jan-Dec), which is reflected in Figs. 1, 7, and Fig. 8, and all node anomalies with MCS days in DJF.

Other specific comments/questions

- L219: 925 hPa is not exactly surface level.

The statement has been rewritten to reflect the 925 hPa level.

- L235: “The patterns demonstrate northward transport...”. Of what?

The statement has been rewritten to read “The patterns in zonal wind shear from first to third column nodes illustrate a strong link of high-shear areas to the propagation of the WAM cycle, and these areas widen as the zonal shear band moves further inland.”

- L241: “Generally, the presence of zonal wind shear can be seen as a necessary condition...” For What?

This statement has been removed.

- Fig.4: How is it possible that the wind shear is negative? Is the directional information included as well, i.e. the direction of the shear vector? Then the authors need to provide more information on the sign of the wind shear.

As outlined in section 2.1., the shear indicator used here corresponds to the zonal wind difference between 925hPa and 600 hPa, which is the main direction in which MCSs propagate (east to west). We thus present enhanced easterly shear as positive anomalies and follow the convention that easterly shear and easterly anomalies are presented with a positive sign across all plots, as these are positively related to storm development. This is now explained in the methodological data section as follows:

“Due to the main direction in which MCSs propagate across the SWA region (east to west), enhanced easterly zonal wind shear anomalies are presented as positive anomalies as these are positively related to storm development.”

- Fig. 8: Following on the comment on Fig. 4 above, a negative anomaly in wind shear can have a different meaning when wind shear itself can be negative. Node 3 for instance, where the western Sahel exhibit a negative anomaly on climatologically negative wind shear values. What does that mean?

This means a strengthening of westerly shear, which is not favoring MCS development or maintenance.

Referee 2:

- The revised manuscript shows the full SOM analysis in supplementary material but does not sufficiently address the previously raised concerns. Specifically, major questions are still pending regarding the SOM classification, which lies at the heart of the analysis. The authors refrain from providing any metrics to estimate the SOM robustness, nor do I better understand why the original 9-node SOM is not used.

Thank you for your advice. Regarding the SOM classification, we evaluate the quantization error. An optimal SOM is obtained when the average Euclidean distance is the minimum (the quantization error is the smallest). Once the average quantization error has been minimized, the relationships between the predictor and node data are investigated. Based on evaluating the quantization errors for node sizes 2x3, 3x3, and 3x4, it was observed that node size 3x3 had the smallest quantization error. On testing various sizes, a 9-node SOM was selected that adequately picks out the seasonal variation of rainfall over the region of study. The 2x3 resulted in a more generalized circulation archetype while the 3x3 represented a wider range of circulations with fewer redundancies.

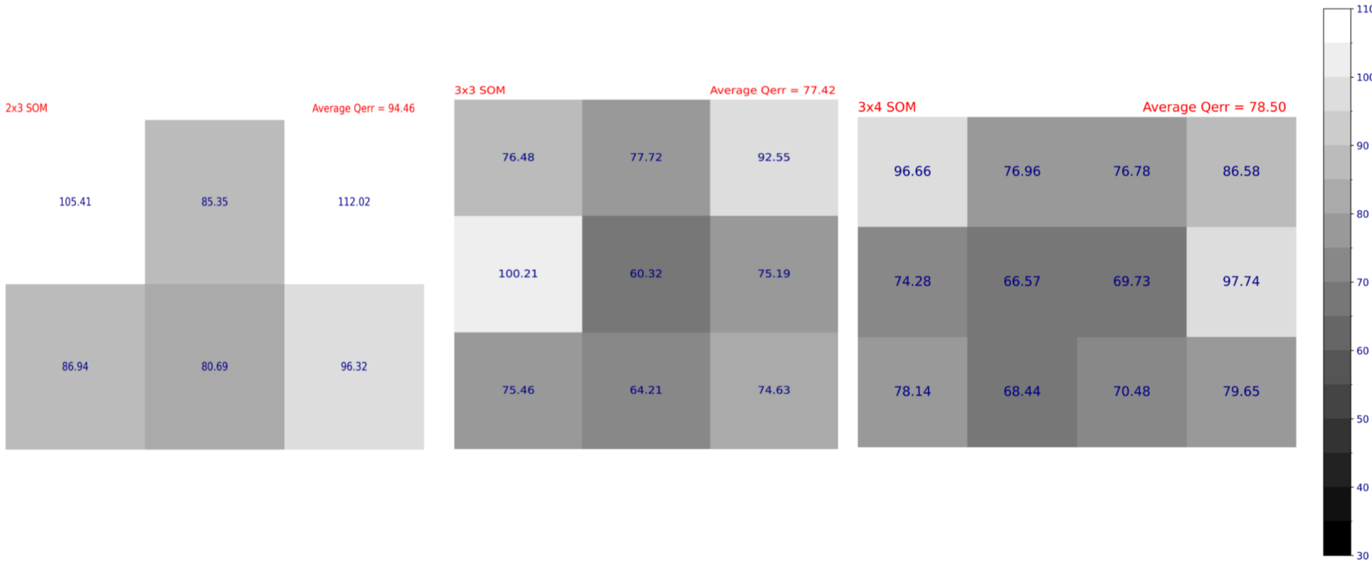

Figure S1. Total quantization error (gpm) of all grid points in the geopotential height patterns mapped to each SOM node.

We have attached here figures of calculated quantization errors for node sizes 2x3, 3x3, and 3x4. This figure has been added to the supplementary material as Figure S1.

Following this comment, we now provide the original 9-node SOM information in the manuscript.

- The authors do not share my concern regarding the evidently low correspondence between low-level geopotential heights and winds in the region of interest.

We assume the reviewer refers to Fig. 3 (old manuscript, Fig. 5 in new manuscript), which however shows geopotential height at 925 hPa (the field the SOM analysis is based on) and flow vectors at 650 hPa (the MCS steering wind level in the region), where consequently the fields do not evidently correspond. For illustration, please refer to the plots below (Figs. S2 and S3) where 925 hPa and 650 hPa geopotential height and wind field are plotted together for the same pressure levels. These are now also available in the supplementary material. This is now better clarified in the text:

"The SOM classification of different synoptic states was based on 925 hPa geopotential heights, with resulting patterns shown in Fig. 4. [...] The overlaid 650 hPa wind field reveals mean easterly wind conditions at MCS steering levels across all nodes, suggesting that the dominant propagation direction for MCSs remains east to west for all synoptic states."

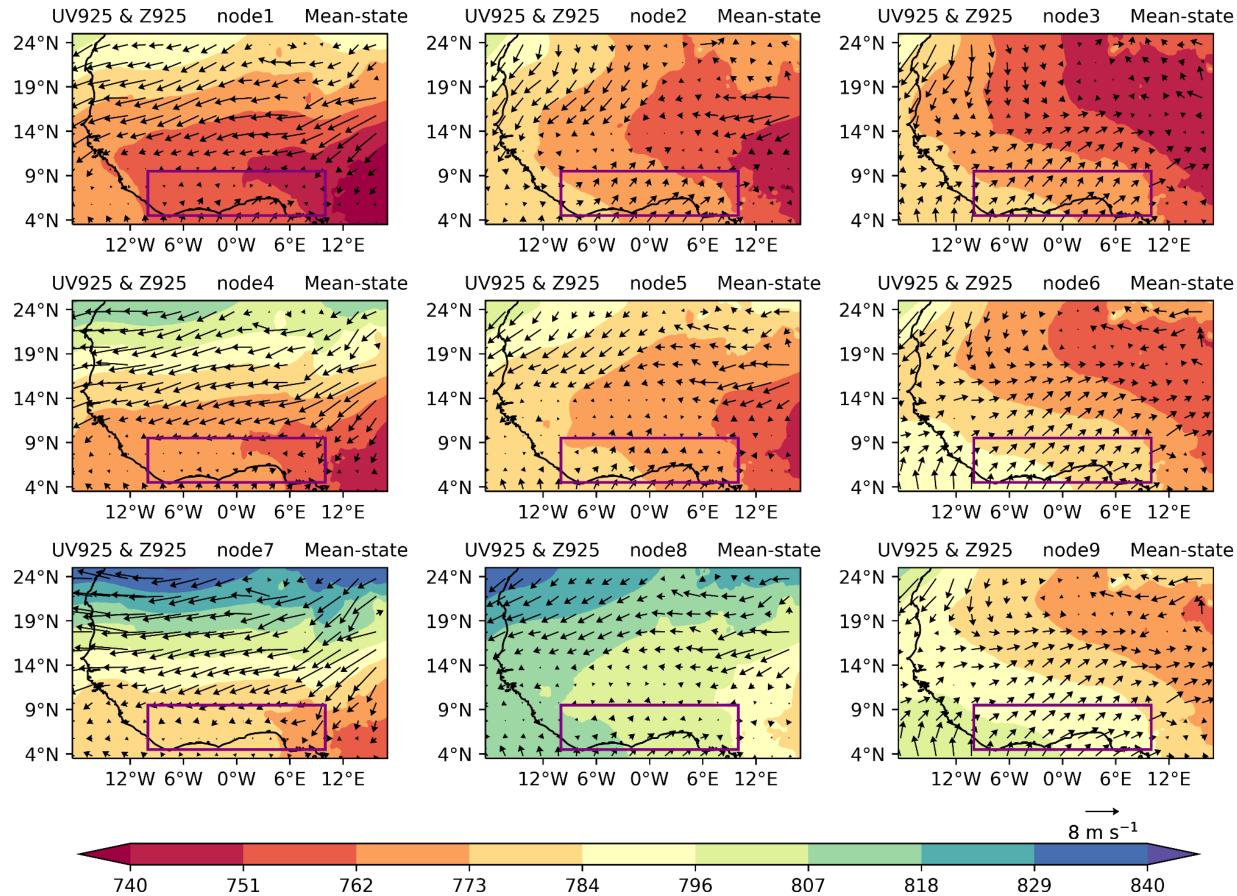

Figure S2. 12 UTC composites of 925-hPa geopotential height (shading; gpm) and 925-hPa winds (vectors; m s^{-1}) in 9 nodes based on SOM analysis. The purple box depicts the SWA region (5º–9ºN, 10ºW–10ºE)

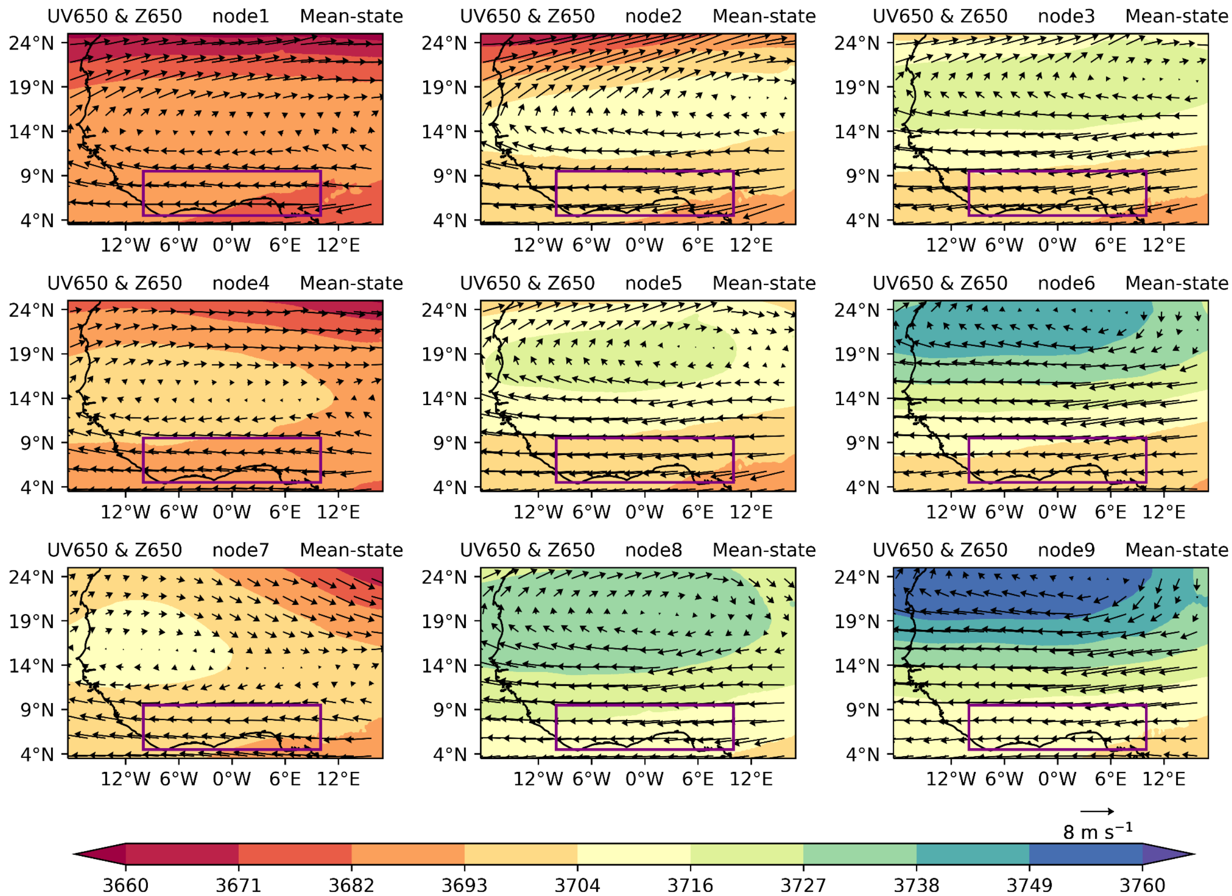

Figure S3. 12 UTC composites of 650-hPa geopotential height (shading; gpm) and 650-hPa winds (vectors; m s^{-1}) in 9 nodes based on SOM analysis. The purple box depicts the SWA region (5^o–9^{o}N, 10^{o}W–10^{o}E)

- While the subject of the study is important and relevant for the scope of the journal, substantial issues arise from the present methodology. I conclude that the manuscript should still undergo major revisions before being accepted.

Major revisions:

- The subjective formation of a 6-node SOM from an objectively defined 9-node SOM raises issues. The classified geopotential patterns' robustness (or confidence level) is unclear. It is important to account for each node's internal variability and to mark which regions of the flow are significant.

Thank you very much for this comment, we now show the original 9-node SOM for the manuscript which, while creating some redundancies, also allowed us to discuss the node matrix structure in more detail, as illustrated by new Figures 2 and 3 looking at node persistence and node-to-node transitions, respectively, as later suggested by the reviewer.

Regarding gph pattern robustness, we evaluated the quality and/or robustness of the SOM in this study based on the learning quality indicator, which is determined through the measurement of the quantization error (QE). Based on evaluating the quantization errors for node sizes 2x3, 3x3, and 3x4, we found that node size 3x3 showed a quantization error minimum compared to the higher/lower node number cases. On testing these various sizes, we hence selected the 9-node SOM as it also adequately picks out the seasonal variation of circulation over the region of study. The 2x3 resulted in a more generalized circulation archetype while the 3x3 represented a wider range of circulations with fewer redundancies. This is now detailed in the SOM Methods section: "*Based on SOM_PAK, we tested node sizes 2x3, 3x3, and 3x4, using the quantization error (QE) as an indicator for quality and robustness of the respective node size. We find a minimized QE for 3x3 (c.f. Supplementary Figure S1), which, from visual inspection, also shows a larger number of distinct circulation features than 2x3 while producing fewer redundancies than 3x4. Thus, all following analyses are based on the 3x3 node matrix.*"

- From previous experience, the statistical significance of the SOM nodes may not include the entire domain, and it is important to clarify the regions where the pattern is indeed robust.

Thank you, it is true that the node presentation on the large West African domain does not necessarily imply statistically significant differences between nodes for each part of the domain or on a pixel basis. As outlined above, the SOM nodes were identified based on synoptic-scale patterns and we now provide the QE errors.
Unfortunately, the reviewer has not provided a test strategy or particular test they would like to see performed for further 'node significance' testing. We hope the information we now provide on node persistence and node-to-node transition demonstrates node differences more satisfactorily.
To explore within-node variability, we also performed a node-mean bootstrapping with 5000 resamples per node and per domain pixel, with Fig. S4 below representing the width of the 95% confidence interval for average geopotential height maps (testing representativeness of geopotential means as presented in the revised manuscript in Fig. 4). We find that for all tested ERA5 pixels, the calculated means as shown in Fig. 4 lie within the respective 95% confidence interval. As visible from Fig. S4, per-pixel 95% confidence intervals are generally small in comparison to large-scale geopotential height gradients and decrease further towards the SWA domain (red shading).

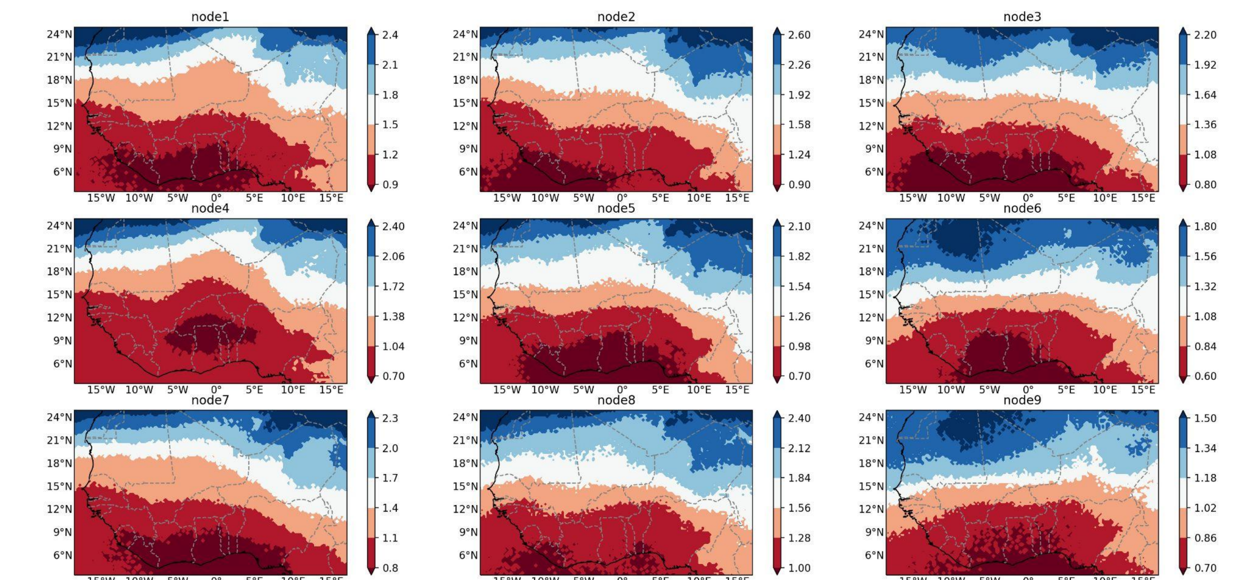

Figure S4. 12 UTC 925-hPa geopotential height 95% confidence interval widths (shading; gpm) based on bootstrapping with 5000 resamples for the 9 node day groups based on the SOM analysis.

- This is helpful to determine the relevance of the mean node states displayed in figure 10, i.e., what is the standard deviation of the mean environmental conditions? If they largely overlap with that of MCS days, then the separation may not be as meaningful as suggested by the authors.

In Figure 10 (Fig.13 in updated manuscript), instead of adding standard deviation whiskers to node mean states and MCS-day states (which creates visualization problems), we now provide the standard deviations of the node mean states and perform Welch's t-tests between node mean states and MCS-day states to indicate for which nodes the mean MCS-day states are significantly different in TCWV and shear. As expected, this highlights that certain nodes are favorable for MCSs in SWA as per their mean state (e.g. node 3, which does not show a significant difference in either TCWV or shear and thus suggests the most favorable mean synoptic conditions for MCSs), while other nodes show significant within-node differences for MCS days only in shear (e.g. peak monsoon nodes 6,9) or both TCWV and shear.

- Furthermore, no topological or quantization errors are displayed, nor is a Sammon map, making assessing the SOM's performance virtually impossible.

Thank you for raising this point.
Here we display the quantization errors for the different SOM sizes we tested. Based on evaluating the quantization errors for node sizes 2x3, 3x3, and 3x4, it was observed that node size 3x3 had the smallest quantization error. On testing various sizes, a 9-node SOM was selected that adequately picks out the seasonal variation of rainfall over the region of study. The 2x3 resulted in a more generalized circulation archetype whiles the 3x3 represented a wider range of circulations with fewer redundancies.

We have included a figure (**Fig. S1**) of calculated quantization errors for node sizes 2x3, 3x3, and 3x4 in the supplementary material.

- Also, the relative frequency of each node is not given, though it is clear that the distribution is far from equal. This leads to peak Monsoon node 6 showing a reduced MCS frequency despite containing the largest amount of MCSs. This is problematic since the main frequency of each node is unclear. I.e., if node 6 is twice (or more) as frequent as node 5, it is reasonable to expect a lower MCS frequency that may be an attribute of inner cluster variability (or noise) rather than a dynamic feature.

Thank you for your comment, the absolute and relative frequency of each node is now given in the title of Fig. 1, alongside the node cases per month. We wouldn't consider the fact that (previous) monsoon node 6 showed reduced MCS frequency per day despite containing the largest number of MCSs (solely due to it containing the largest number of node days) as problematic. The metric MCS frequency/day implicitly contains the number of days per node and answers the question of whether a particular node is more or less favorable for MCSs in SWA on a particular day, as intended. Please note that in SWA, the "peak monsoon" period includes a dip in MCS frequencies in southern West Africa, including the little dry season, when MCS frequency is in fact lower, as shown in Fig. 7. We now revised our node classification into "dry season, transition season and monsoon season" and discuss the relationship of MCS frequencies/day with the presented MCS annual cycle in more detail in the respective first paragraph of section 4.2:
"*The location of MCSs during node days is represented in Fig. 8. Comparing daily MCS frequencies, we find that MCSs are most likely to develop under transition node (2,5,8) conditions (2.8 MCSs per day) featuring a northward-displaced moisture anomaly (Fig. 9). Given the transition nodes occur predominantly during pre-monsoon (late March to June) and post-monsoon (from September to November) - the major and the minor rainy season respectively in SWA (cf. Fig.~1), these patterns may in some cases be representative of early monsoon onset and a delayed monsoon retreat respectively. MCSs seldom develop under dry node (1,4,7) conditions, with as low as 0.6 MCSs per day*"

- There is a problem with the logic of the present analysis: Nodes are associated with monsoon phases mostly based on their seasonality. MCSs make up 30-80% of the monsoon precipitation, so <u>they are clearly more common during peak monsoon stages.</u>

We apologize that the given information in the manuscript seems to give rise to this misunderstanding, but for southern West Africa, MCSs are not the most common during peak monsoon stages, as Fig.5 (Fig. 7 in the revised manuscript) illustrates. Overall MCS numbers in southern West Africa are lower during peak monsoon months July and August, due to the bulk monsoon band moving further inland during the monsoon peak, creating drier conditions in southern and coastal regions. Large parts of the evaluated region have a bimodal rainfall regime during pre/early and post/late monsoon stages, created by the northward passing and subsequently southward passing monsoonal rain band. Depending on exact locations, MCS frequency peaks thus lie in pre-/and post-monsoon phases (now termed "transition season"). This is reflected in the MCS/day metric. The "peak monsoon" expression is not to be used synonymously with peak rainfall or peak MCS occurrence but reflects the monsoon movement stage. This has been added to the introduction section to read as follows:

"*The SWA region is different from its Sahelian counterpart in its closer proximity to the ocean and a distinct bimodal rainfall seasonality. The WAM stages can broadly be classified into a dry season when north-easterly Harmattan winds prevail over most of West Africa during December-February when rainfall mostly occurs off the southern coast of the continent (Thorncroft et al 2011), and the monsoon season from July-September, initiated by a striking jump of the monsoonal rainfall band from coastal regions to the Sahel (Hagos and Cook, 2007). The monsoon months thus represent the unimodal Sahelian rainfall season. In SWA however, the majority of rainfall occurs between the dry months and monsoon months, when the monsoon rainband first passes northward over southern regions from March to June and subsequently moves southward again when the monsoon retreats in October (e.g. Maranan et al 2018, Klein et al 2021). Here, we summarize these months when SWA receives most of its rainfall as transition season.*"

- This suggests that pre-monsoon nodes will have significantly fewer MCSs and hence their response to MCS appears larger. As a result, seen in Figure 10, only node 1 seem to show a significant response to MCS, while for the other nodes, the mean state is well within the STD of the MCS conditions. This undermines the main point of the paper.

Thank you for this comment. We first want to point out that in our manuscript, we are transparent about the differences in MCS numbers per node as per the given MCS/day frequencies (old Fig. 4, now Fig. 8), which exactly addresses the question as to what within-node conditions would allow MCS development. We additionally now added absolute MCS numbers per node to Fig. 1 in the updated manuscript.

We would argue that illustrating that certain nodes are MCS-favorable in their mean state while others are not (i.e. show large anomalies in Section 4.2) is exactly one of the important main points of the paper, rather than undermining any point made, i.e. no or little difference between mean node state and MCS-day state identifies nodes (large- scale synoptic states) that are on average MCS favorable in southern West Africa. However, as our updated Fig. 13 illustrates, only node 3 shows insignificant changes in both TCWV and shear for the MCS day mean compared to the node mean. Higher or lower MCS frequencies within a node are not a problem of the analysis but a consequence of how frequently respective SOMS-defined synoptic conditions allow the development of convection in SWA (producing the MCS-day anomalies shown in Fig 13 in the updated manuscript). It is thus perhaps not an entirely unexpected result that nodes that represent on average dry season conditions (e.g. node 7) "need" larger increases in TCWV and shear to allow MCS development.
In addition to updating Figure 13, we now clarify that nodes that display the largest changes needed for MCS development are nodes that exhibit on average atmospheric conditions most hostile to convection and thus feature the smallest total within-node MCS numbers as part of our revised discussion & conclusion section.

- The MCS data should be better treated and displayed. As innovative as combining satellite data to such a large-scale perspective is, it eventually ends up only as a cloud indicating all MCSs within each node. It is worth visualizing the data in a manner that will ease the interpretation of the results. For instance, it is unknown what the MCS spatial distribution actually looks like for each node. Overlapping dots are invisible and may hide preferable locations for MCSs. The mean number of MCSs per MCS day per node is not given, thus it is possible for example that while MCS "daily" frequency peaks on node 5, the overall MCS frequency may be larger for node 6, simply by having more MCSs per MCS day. Such discrepancies in the MCS data should be addressed and studied within the present manuscript, including spatial variability.

The MCS/day metric was calculated as the absolute number of detected MCSs at 1800 divided by the number of node days. If more than one MCS exists at that time on any single day, this frequency contribution is thus taken into account (I.e. the number does not merely reflect the number of MCS days per node). As per the explanation of the seasonal cycle of MCSs in southern West Africa given above, our result of a higher MCS frequency for node 5 in the previous version of the manuscript was reasonable given that node 5 predominantly covered September and October days, the peak of the second rainfall season in southern West Africa (these nodes are now termed "transition nodes"). We added a new figure showing a density plot of MCSs (rather than individual points) with a zoomed-in domain to facilitate pattern evaluation and provide the needed detail on how the MCS frequency/day is calculated in the caption of that figure (Fig. 8 in the revised manuscript). We also include a description of the typical annual cycle and the relations of the rainy season to the West African monsoon progression to the introduction as mentioned above.

- Information on the SOM dynamics also seems relevant, e.g., how long does each node persist? are the nodes typically changing on a time scale of days or weeks? Which nodal transitions are frequent, which are rare, and how does this relate to the dynamical interpretation of the SOM as indicating the monsoon phases? Such information is valuable to evaluate the SOM's consistency in mapping consecutive days and support the derived conclusions.

Thank you for this question which helped us better explore the node differences - and made the presentation of the full 3x3 SOM matrix very useful. We now added a figure showing typical (and clearly different!) node persistence as well as node transitions in the manuscript as Fig. 3 and Fig. 4, which includes an added new section on node persistence and node-to-node transitions, including in the entirely revised discussion part.

Minor comments:
L140: the heat low is not well captured relative to similar SOM analysis. I've commented more about that above.
See provided plot that illustrates consistency above.

L144: ts = typo?
Yes it is a typo. This has been corrected

L156: Q and T errors should be presented per cluster
The errors are now shown in Fig. S1 in the supplementary material.

L160: what led to the choice of 9 members? Was it purely a qualitative choice?

On testing various sizes, a 9-node SOM was chosen that adequately picks out the seasonal variation of rainfall over the region of study. The 2x3 resulted in a more generalized circulation archetype whiles the 3x3 represented a wider range of circulations with fewer redundancies. This was based on evaluating the quantization errors for node sizes 2x3, 3x3, and 3x4, which saw that node size 3x3 had the smallest quantization error. The choice was therefore based

L210: Why is a heat low evident as a high-pressure area?

As is typical for thermal lows, the SHL exhibits low pressure for near-surface levels (the 925hPa level at which the geopotential height is shown), which turns into a high pressure circulation above at 650hPa for which the wind pattern is shown. To avoid any more confusion, we now show both, the 925hpa geopotential with circulation field, and 650hPa with circulation field here and in the supplementary materials – illustrating correspondence between geopotential height and flow, and only refer to the low pressure system in the low-levels that is known as the Saharan Heat Low.

L212: “and linked southward retreat” – rephrase
This statement has been removed.

L238: “weaker geopotential heights representative of high-pressure areas” – unclear. Rephrase.
This statement has been removed.

L335: the use of the term “storm” is confusing. Stick to MCS.
The term “storm” has been changed to “MCS”.

L356 & 357: use “show/shown” instead of the passive “observe/observed”.
These changes have been done

L359: remove double dots
Double dots have been removed

L364: “with frequent convective activities during peak monsoon” - Why is the same not true for node 5 with the most frequent MCS days?
This has been revised, now considering the full 9 node structure.

L376: “..states and then examined..” - .. “states and examined...”
Changed.