The manuscript analyses large-scale synoptic conditions in West Africa, and aims to identify flow anomalies related to the occurrence of MCS in a smaller domain of South West Africa (SWA). The study relies on daily ECMWF ERA5 data for the period of 1981-2020 for the large-scale environment and cloud-top temperatures retrieved from satellite data to identify MCS events. The 925 hPa geopotential surface is classified using SOM, and 6 nodes out of 9 are discussed. The nodes correspond to the seasonal evolution of the West-African Monsoon (WAM). The flow features are discussed with an emphasis on winds, humidity, vertical zonal-wind shear, and CAPE. Nodal variations corresponding to MCS events are analyzed regionally and locally and solidify the link between increased TCWV and vertical zonal-wind shear to MCS-favoring conditions in the SWA. It concludes that local increases in wind shear and humidity are a common MCS feature for all MCS days, with strong wind shear correlated to large CAPE and dictating the onset of MCS. However, since this mechanism has been consolidated for the Sahel, there seems to be room for a deeper investigation here.

MCS studies dedicated to the SWA region are lacking, and the motivation for the present one is clear. The introduction is well structured and written for the most part, though some elaboration is required when it comes to the SOM classification. The methodological approach appears novel and highly practical. The results are clear and concise, perhaps to a fault – some aspects could be elaborated. Still, the results enable the conclusions drawn by the authors.

The subject of the manuscript is relevant and fits the journal's scope. However, there are major issues that should be addressed before consideration of this work for publication. I summarize the main issues in the following, along with more specific comments and suggestions for addressing the main concerns and improving clarity.

General comments

1) The SOM analysis should be presented as a whole, and the reason for rejecting certain nodes from the analysis should be better explained. No discussion is dedicated to the SOM configuration, primarily the choice of the number of nodes, but also other SOM parameters (neighborhood size, topology, initial coverage space, etc.). The robustness of the SOM clustering is not evaluated. A significance test should be added for the detected geopotential patterns, and the SOM errors (quantification and topological) should be discussed. Furthermore, the low number of nodes under consideration for this study appears to not fully justify a SOM analysis in the first place. Seeing as each node roughly corresponds to a certain season and is treated as a seasonal mean, it appears that the information presented here can be yielded by a simple seasonal decomposition. Ideas to enrich the SOM analysis and the gain from it can be found in the literature quoted by the authors. Otherwise, the authors may consider replacing the SOM analysis with a simple seasonal decomposition. Please also refer to more specific comments in this regard, below,

and the following highly relevant references with very similar motivations and methodologies:

- Liu, Y., Weisberg, R. H., and J. I. Mwasiagi (Eds.): A review of self-organizing map applications in meteorology and oceanography, Self-Organizing Maps: Applications and Novel Algorithm Design, InTech publications, Rijeka, Croatia, 2011.

- Gueye AK, Janicot S, Niang A, Sawadogo S, Sultan B, Diongue-Niang A, Thiria S 2010 Weather regimes over Senegal during the summer monsoon season using self-organizing maps and hierarchical ascendant classification. Part I: synoptic time scale. Climate dynamics. doi:10.1007/s00382-010-0782-6

- Espinoza, J. C., Lengaigne, M., Ronchail, J., and Janicot, S.: Largescale circulation patterns and related rainfall in the Amazon Basin: a neuronal networks approach, Clim. Dynam., 38, 121–140, https://doi.org/10.1007/s00382-011-1010-8, 2012

- Givon, Y., Keller Jr, D., Silverman, V., Pennel, R., Drobinski, P., & Raveh-Rubin, S. (2021). Large-scale drivers of the mistral wind: link to Rossby wave life cycles and seasonal variability. Weather and Climate Dynamics, 2(3), 609-630.

2) The choice of low-level geopotential heights as a clustering agent should be better motivated, given the relatively low correspondence between it and the low-level winds in the domain, which are described as the main process driver throughout the manuscript. Have the authors considered directly classifying the wind field?

3) Nodal trends – this section appears unrelated to the motivations of the paper and is very slim. I suggest a deeper analysis to explore, for instance, corresponding trends in MCS events. Otherwise, consider removing this section.

4) MCS data – I think this data should be further explored. For one, it can be better presented using a density plot. Secondly, spatial variability should be discussed and possibly explained, with an emphasis on variations between nodes and seasons within the nodes. Finally, it's worth checking for MCS behavior on off-season node days.

5) The link to predictability can be improved. For instance, can we learn anything from a lagged correlation between nodal transitions and MCS density?

Specific comments:

L24: Too vague. What is the input used for classification? i.e., how do you define a "synoptic circulation-type"?

L32: Unclear. Do you mean vertical/ horizontal wind shear? what is the field under discussion here?

L35: The use of the term "shear" or "wind shear" when alternatively referring to vertical and zonal shear is confusing. You should specify which shear is under consideration throughout the paper.

L49: Missing a link to WAM. The change of subject is too sudden and does not flow from the previous paragraph. Consider opening the section with lines 53-54

L94: "large-scale patterns" - Too vague. You should name the parameter used for the classification here.

L121: "SWA domain" - This domain should either be specified in latitude and longitude boundaries or displayed in a figure earlier on. Possibly both.

L124: This section requires more detail.
For instance, what is the SOM topology? It would be useful to add a neighbor distances map and to evaluate SOM errors. The number of members in each cluster should also be given, preferably in Fig 1.

L128-129: This statement is true for many optional classification inputs. In the present study, the focus is on the tropics where geostrophic balance is not obvious, as seen by your results. Therefore, the choice to classify patterns using geopotential heights should be justified.

L131: Each method has its advantages and disadvantages, and each can be more suited for a different study. Refrain from making conclusive statements.

L132: "data is not discretized and orthogonality is not forced" – Again, these are not clear advantages. The SOM's strengths and weaknesses should be discussed in the context of the present study.

L142: More information is required on what led to the choice of 9 clusters. Have you evaluated the network errors under the different configurations (SOM size and other parameters) to show that 9 is the most compatible?

L153: Why not compare to non-MCS days within the node? This may highlight the signal you are after.

L154: regarding the T-test – on Which confidence level was it conducted? have you used any method to detect false positives in the multi-gridded test? See Wilks 2016 for example.

Wilks, D.: "The stippling shows statistically significant grid points": How research results are routinely overstated and over-interpreted, and what to do about it, B. Am. Meteorol. Soc., 97, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1, 2016.

L165: Why is the complete SOM not shown? This is not clear.
If you choose to discard nodes altogether, you should show the full SOM map (9 nodes) first, then explain why not all nodes are relevant, and which ones were removed. The resulting 6-node SOM map should be shown in the context of the full SOM map, as the node locations on the SOM map are crucial for the SOM interpretation. This also raises the question: are the panels in Figures 2-9 arranged correctly? i.e., are neighboring nodes in these Figures also neighbors in the full SOM map?
I suggest repeating the analysis for 6 nodes if that's what you end up analyzing, while completely removing irrelevant dates from the SOM input.

L167-168: Even if some nodes are ignored, the numbering of the nodes should be as in the full SOM analysis, to be consistent with the complete SOM map.

Figure 1: Add the total number of members in each node. Consider normalizing per year and not per month.

Figure 2: Grey grid can be removed to improve visibility. Also, be consistent with X-label intervals. Clarify whether these are daily means or 12 UTC composite.

Figure 3: The low correspondence between winds and geopotential heights in the tropical region raises the question: what is the value of classifying by geopotential if it's not indicative of the flow field? Why not directly classify the velocity/ wind-speed fields?

L211-218: This subsection is too slim. Either remove it or expand it to get to a conclusion .
At the present state, this subsection does not contribute to the main motivation of this study and possibly draws the reader's attention from the main storyline.

Figure 6: This domain should be shown earlier when first presenting the SWA domain.

L258: unclear. Why does high humidity lead to cooling?

L279: This point was given as a well-known fact in the introduction, so I don't see what is the novelty here.

L290: "eastern patterns" – This is not evident in figure 6. This issue should be discussed

L140: "pure node analysis" – What do you mean by this?

Technical corrections

L25: which=that.

L38: variabilities=variability.

L42: "Change, 2014" is not a reference, or is missing from the reference list.

L88: environments= parameters?

L97: Is this a correct use of the word stratify? Seems confusing to me. How about grouped/ separated?

L108: product= data source.

L126: daily=daily mean.

L182: SOMs=SOM.

L200: "much more strengthened" – Rephrase. Consider "Intensified", "Increased" and so on.

L211: "A further" = Further.

L213: during=within.

Figure 4: The term "moving mean" seems more fitting.

L225: This second subtitle is redundant.

L244: Repetitive.

Figure 7: The colors appear saturated. Expand the color map beyond 2K to avoid this.

L274: observes= demonstrates/ exhibits.

L290: observe = show/ depict.

L308-311: Long sentence, consider splitting.

L323: "making node 2…" This sentence is unclear, rephrase.

L330: "This season" – Unclear which season is that.