Review of the first revision of the manuscript entitled "Assessing the ability of the ECMWF seasonal prediction model to forecast extreme September-to-November rainfall events over Equatorial Africa", submitted by Nana et al. to Natural Hazards and Earth System Sciences (NHESS).

I appreciate the effort and the amount of work that the authors have put into the revision of the manuscript. The cohesiveness and comprehensiveness of the manuscript have increased considerably and the authors have successfully addressed most of the concerns from the referees. Nevertheless, I believe there are still some aspects that require revision. Please find below my remaining concerns (most of which are minor) and further suggestions which I hope are useful for the authors.

General comments

1. The Introduction is now very well structured and sets the context for the study appropriately. The authors motivate the study by stating that it is important to assess whether the seasonal model successfully reproduces the mechanisms driving rainfall variability over EA, and because the performance of SEAS5.1 has not been evaluated yet in comparison with the previous SEAS5. However, considering that the main difference between SEAS5 and SEAS5.1 is rooted in the interpolation and additional variables, rather than a change in the physics, dynamics or data assimilation of the model, in principle I would not expect any relevant differences in terms of performance between both. The authors could comment on this, and therefore focus the motivation mainly on the study of the mechanisms in the model. I therefore recommend that the authors emphasise this point and place less emphasis on the change in model version, since the scientific questions addressed are: does SEAS5.1 exhibit skill in precipitation over EA? And if so, does the skill come from the right sources? Or is the model producing precipitation for the wrong reasons? I think these are the main scientific questions, which are sufficiently relevant and make this study necessary.

Specific comments

1. I appreciate the authors' explanation of lead time definitions. I will try to clarify my concern regarding the lead times. When Lead-0 forecasts are used for SON (i.e. the September initialisation is used to predict SON), the three-month period under analysis includes one month which is Lead-0 (September). In this particular month, it is likely that short-range predictability or regional features are playing an important role, and probably SEAS5 is able to capture these processes, thus having higher skill. This is not the case for the forecasts initialised in August (Lead-1 forecasts), since for the three months in SON, the sources of predictability are mainly seasonal. Thus, it may be a bit tricky to compare skill between both initialisations. It is possible that the Lead-0 forecasts perform better in SON not because the teleconnections are reproduced more accurately, but mainly because SEAS5 successfully predicts shorter-range predictability features that increase forecast skill in September particularly. I agree that this influence is supposed to be limited when dealing with a 3-month period, but am unsure about to what extent it is negligible. This was the point I was trying to make and I hope it is now more clear.

2. Please specify in the Data and Methods section that SEAS5 initialisations for all the months (not only August and September) were used in order to study the rainfall annual cycle (shown in Fig. 1).

3. Regarding Fig.1. A legend in Fig. 1a would be useful. I also suggest using a different plot type (line plot) for observations and keep the bar plot for model data in order to allow a more straightforward comparison with the model. Additionally, the caption states "The gold bar (black line)..." and there is no black line at the moment.

4. Please check units in Fig. 4. If this is a regression map, the time-series index used (DMI in this case) is standardised and therefore has no units. The resulting regression units are mm/day. Otherwise, if the index is not standardised, the map shown is a covariance map, not a regression map. If that is the case, the units are indeed mm/day/ºC but the figure caption needs to be changed.

5. Lines 452-453. This claim should be stated more cautiously, since the authors are only showing regression patterns and not all the mechanisms involved in that teleconnection. Thus it should be something similar to "The model at both L0 and L1 appears to successfully capture this Atlantic teleconnection".

6. Regarding Fig. 6: I understand the authors' clarification regarding how the model metrics have been computed member by member and then the mean value is used. However, the model data shown in Fig. 6b and Fig. 6c are ensemble-mean values (43 data points). If all the ensemble members were shown, the number of points shown in Figs. 6b and 6c would be much larger (25 members x 43 years = 1,075 data points). I understand that in order to compute the correlations shown in the upper-left corner, the authors did not use the ensemble-mean data shown in Figs. 6b and 6c. Instead, they computed the correlation in each member separately, and then computed the mean correlation value. Could you confirm this is the case? If it is, please clarify in the figure caption that the data points shown are ensemble-mean values. Additionally, while discussing Fig. 6, the authors state that the model overestimates the correlations. Since the correlation in the model is supposed to be an average of 25 correlation values, it would make the result much more robust if the authors give a confidence interval, and/or the number of ensemble members featuring a higher correlation than the observations. If all (or most of) the 25 ensemble members have higher correlation than the observations, that is really robust and worth highlighting in the discussion.

7. The composites of SY and WY feature relatively small sample sizes (n=7 and n=6, respectively), and assessing statistical significance of these composites with the Student's t-test is not appropriate because the estimation of the sample variance is noisy and the number of degrees of freedom is small. In these cases with small sample sizes, it is strongly encouraged to evaluate statistical significance with resampling methods (e.g. Monte Carlo) instead of the t-test.

8. Fig. 8 shows some measurable differences in the precipitation composites between CHIRPS and ERA5, especially over the central part of EA, which are worth discussing. Could this be influenced by a reduced number of observation points in this area? Given these important differences, I think it is relevant to compute the skill

(correlation coefficient and bias) and teleconnection metrics taking ERA5 as observational reference and include those results in the supplementary material. I understand that this may involve a sizeable amount of work, but I believe that it is important to confirm whether using ERA5 or CHIRPS as observational reference significantly alters the results. Addressing this will make the results even more robust.

9. The authors mention the underestimation or overestimation of zonal moisture flux and wind in the discussion of Fig. 11. Please, could you specify whether you are referring to overestimation/underestimation of the magnitude of the anomalies or to how much the anomalies extend in a particular direction?

10. The differences in the AEJ-N and AEJ-S between the climatology, strong and weak composites appear to be very small. For instance, in ERA5, the differences are of the order of 1 m/s. Hence I wonder whether differences of 1 m/s are meaningful or if the AEJ may not be playing an important role here. I would appreciate it if you could discuss this point.

11. Lines 674-676. The authors might have accidentally written "(divergence)", because the sentence would read: "... leading to decreases in moisture divergence over western EA favouring dry conditions". I think that the authors mean a decrease in moisture convergence, not divergence.

12. Line 203: "From November 2022 onwards, the updated version SEAS5.1 is used". This statement might be misleading because the reader may understand that the data from 1981-2021 is from SEAS5 and from 2022 is from SEAS5.1. I recommend rewriting it so that it is clear that all the data used in the study is from SEAS5.1.

13. In the Methods subsection, the authors indicate that the horizontal resolution from SEAS5.1 data is 1º. Please indicate that this is the resolution available from the Copernicus CDS, because the model's native grid for the atmospheric component of SEAS5 has a spectral resolution of TCo319 (which is of about 36 km). The model analysis is an important part of this study, so it would be worth devoting just a few lines in the Data subsection to introducing the SEAS5 model: the different model components (atmosphere, ocean and land surface) and where do the initial conditions come from.

14. The discussion of Fig. 12 is very descriptive and, in my opinion, lacks a bit of discussion. I think it is worth discussing what these differences could signify in terms of forecast skill. In particular: could the differences in the westward/eastward extension of the zonal moisture flux and wind anomalies explain why SEAS5 lacks skill in particular regions? This would also further improve the comprehensiveness of the manuscript, as it links the analysis of mechanisms with the preceding analysis of forecast skill.

Minor technical or language comments
1. Line 161 I suggest adding "over EA" after "SON precipitation".
2. Line 427: remove the "d" from "overestimated"
3. Line 442: I think you mean "observations" and not "observed".

4. Line 449: "A recent study …" (add "A")
5. Line 485: use "computed" instead of "performed"
6. Line 492: "resumes" does not sound correct, I suggest using "summarises" instead
7. Line 668: Replace "the both" by either "both" or "the two"
8. Line 673: "..., the AEJ-S and AEJ-N core speeds decrease compared …." (add the "s" in speed and remove the "s" in decreases).
9. Line 674: "... increases (decreases) in moisture convergence…" (add "in")
10. Lines 633-634: Replace "overestimate" and "underestimate" with "overestimation" and "underestimation" respectively.
11. There are still some words in American spelling (e.g. "favoring" in Lines 641 and 675, "centered" in line 205, "analyzed" in line 136, "analyzes" in line 175, "initialized" in lines 146, 158, 186, 188, 380, 758 and 759, "characterize" in line 474… Please check all the manuscript to ensure consistency in the spelling.
12. Line 41 in the supporting information "for the SON period" (add "the")