

General comments:

The authors investigated spin-up behaviour in four single-column models within the MUMIP framework, with a particular focus on convective diagnostics. The same external forcing from the ICON-DYAMOND experiment is applied to all four SCMs, and the evolution of precipitation and CAPE over the Indian ocean domain are examined by looking at correlation and EOFs. The study offers two potentially novel contributions, one is a quantification of spin-up variability across different models, and the finding that CAPE change is more strongly correlated with mixed-layer drying, which is generally expected to correlate with precipitation in the mid latitude.

The underlying scientific question is timely and interesting, and the MUMIP dataset is genuinely novel. I believe the study contains publishable scientific findings. However, I found the manuscript difficult to follow, making it challenging to clearly identify the main scientific message in its current form.

For this reason, I would encourage the author to seek additional internal feedback on the manuscript structure and readability prior to resubmission. I have listed my major concerns below, focusing primarily on issues of presentation and organisation. I would welcome the opportunity to evaluate the scientific contents of the work in greater depth once the manuscript has been revised and its central conclusions are more clearly articulated.

Specific comments:

1. Missing no spin-up setting

To separate spin-up effects from the evolution of the background state, it would be valuable to include a baseline experiment that is not expected to exhibit spin-up effects (e.g., ICON SCM forced by ICON output). The authors instead include the ICON-DYAMOND output as a benchmark, which I consider a reasonable substitute if used consistently throughout the analysis. However, this benchmark is only shown in Figs. 1 and 2 and is absent from the subsequent analyses. At the same time, the manuscript continues to discuss the ICON benchmark results (e.g. L260, L273, L394), including the statement “the ICON benchmark tendencies would be negligible”. It was not always clear to me whether this statement is supported by analyses not shown in the manuscript or is inferred from the experimental design. I therefore encourage the authors to ensure that all discussion points are clearly linked to presented results and that the ICON benchmark is treated consistently throughout the paper.

2. Reliance on preprint and in preparation.

The manuscript currently depends heavily on preprint by Groot et al. (2026b) and in-preparation studies. Including additional methodological and scientific context within the manuscript itself will make the manuscript more self-contained and would broaden accessibility.

3. Missing concluding remark.

The authors state their aim in this paper as “Our overall aim is to quantify spin-up and physically meaningful variability with the statistical analyses. Thus, we here identify which spin-up plays an important role in the MUMIP and whether physics tendency patterns can be generalised across multiple models.” These objectives provide a useful framework for the study. However, I found it difficult to identify where they are explicitly revisited and answered in the later sections of the manuscript. The Summary section restates individual analysis results, but does not clearly synthesise them into conclusions. Consequently, I found it challenging to assess the extent to which the analyses support the conclusions corresponding to the objectives implied by the Introduction. I encourage the authors to include a dedicated Conclusions section that explicitly revisits the original objectives and summarises the main findings in terms of the posed questions.

4. Structural and stylistic problems.

The manuscript contains confluences of materials that would conventionally belong in different

sections and paragraphs. For example, the paragraph starting at L48 in Introduction appears more appropriate for the method section. Similarly, the EOF methodology would be easier for readers to follow if introduced in the method section rather than in the result section. Another example appears in Section 3.3. The first three paragraphs seem introducing the scientific background, which might fit more naturally in introduction. These are followed by correlation analysis results and then by a discussion of the Hadley-cell dynamics underlying the results. Those are all within one subsection of Results. Similar mixture repeatedly appears throughout the paper. I therefore recommend a substantial restructuring of the manuscript so that readers can more easily follow the scientific narratives.

I also encourage the author to have a thorough grammar, and consistency review. Examples include inconsistent line-break styles (e.g., L66, L317, and L323) and very short paragraphs (e.g., L426 and L477) that interrupt the flow of the text. Addressing these issues would ensure that readers can focus on the paper's inherently valuable scientific messages.

Minor specific comments:

1. Fig. 5: I recommend to include a plot for DCAPE.
2. L161: It is stated that "The precipitation rate is broadly similar for five configurations. That means no systematic bias between the convection-permitting benchmark and our physics suites appears, although this might be expected."
I do not agree with this statement. They seem similar because of the logarithmic scale. Precipitation rate is the most known biased variable among models and it is one of the largest reasons why we develop expensive convection-permitting models. If a SCM forced by coarse external forcing can produce the same precipitation as finer-resolution models, we don't need to invest the development. I found that showing a raw PDF with a logarithmic axis and concluding no bias is deliberately misleading.