General Comments:

The paper is a good follow-on paper to the UK Met Office’s regional forecasting system RAL1 science configuration published by the same author in 2020. The case studies covered a wide range of phenomena and sufficient sample size to make convincing conclusions on model performance. The paper also highlights the specialised and wide range of physics research which can improve model performance over and above the dynamical core of NWP systems. Worth noting is that the proposed microphysics changes do not improve performance in all model outputs; in some aspects, RAL1 outperforms RAL2. This information is always valuable for operational centres to consider when planning model upgrades. Furthermore, the paper highlights the benefits of consortia in NWP development.

Specific Comments:

To fully understand some references and model configurations, one needs to have a good understanding and knowledge of the contents and results of the 2020 paper. Therefore, some technical references or specific comments might be seen as unnecessary.

- During the explanation of rotated v. unrotated grids, it is mentioned that Australia is at lower latitudes (p 3 line 7). However, the continent lies between 10-43 degrees south. Therefore, it might be worthwhile to give a short explanation/reasoning why regional domains for parts of Australia does/does not apply a rotated grid. Alternatively, since the domain is not yet defined y p3, mention that the domain of interest in this paper lies within the tropics; hence the model is unrotated.
- Remove “%” in all table headings
- P2, L13: Do forecasters also conduct subjective assessments at NCHWRF and BoM, or is
this done only at the Met Office?
- P5, Table 2: Define \( b_{\text{LEM}} \) and \( c_{\text{LEM}} \).
- P6 Equation 1: Define \( \Gamma \).
- P6,L24: Briefly explain on how/why changes in surface snow settings affect graupel in precipitation.
- P7,L20: Explain why scorecards use 10.5 km (7 grid-lengths); possibly related to the scale of the synoptic observations?
- In Section 3.2: It is overall not clear which case studies are evaluated and presented in Figures 4-8. It is stated that RAL1-M were evaluated for UKV and Darwin, but not the number of cases for Darwin.
  - Which case studies are included in the results shown in the scorecards, and how many cases from each partner were included? Apart from the 100 cases at the Met Office, it is not stated how many cases were conducted by NCMWRF and BoM.
- In the scorecards, does “precipitation” include snowfall?
- Figure 6: Does the possible weakness of GPM capturing higher rain rates contribute to this result? Or is this not applicable in higher latitudes?
- Figure 6: Was the model resolution upscaled to the GPM resolution?
- P8,L26: To which Figure does this paragraph refer? How did you distinguish between Met Office and Darwin cases?
- P9,L1: No plausible explanation is given for the performance increase in winter and not summer months.
- P9,L23: Define the grid resolution of MOGREPS-UK since, in Figure 11, 7 grid-lengths now equal 15 km as opposed to 10.5 km in previous scorecards.
- P10,L4: Is the more extensive fog in the RAL-M configs more accurate? Also, is it more extensive in both temporal and spatial scales compared to the observed?
- P10,L13: “Figure 2 shows…” – is the results in 3.7/Figure 12 only for the circle (radar coverage) in Figure 2 or for the whole domain as shown in Figure 2? Please clarify.
- P11: Define GWP, LMI, ENLS, QCF
- Table A2 is not referred to in the text

**Technical corrections:**

- Figures 1 and 2: use either “orography” or “height” in both figures for consistency.
- P1, L18: It is suggested that the term “National Hydrological and Meteorological Services” (NHMS) be used to align with international practice (also used by the WMO).
- P1, L19: Correct the reference syntax; include all references in one bracket.
- P2, L6: Typing error “trhe”
- P2, L8: Suggest shortening the sentence: “…10 years. This strategy includes…”
- P2, L21: “The systems run in variable…”
- P2: L23: Suggest defining UK (might be pedantic)
- P3,L5: “…centre of the regional model domain…”
- P3,L10: Swap RAL2 and RAL1 in the sentence to agree with the order as shown in Table 1.
- P3,L17-8: Add a reference for this statement
- P4,L6: The link is to an internal site at the Met Office and is inaccessible to the public. Suggest using the University of Leeds site for the PDF document (http://homepages.see.leeds.ac.uk/lecsjed/winscpruse/socrates_techguide.pdf) website OR the SOCRATES github page (https://execlim.github.io/Isca/modules/socrates.html) similar.
- P4,L10: Give an example of how the single moment microphysics has been “extensively modified.”
- P4,L20: Provide a reference or briefly explain why a limit is applied and the possible ranges between liquid water and ice phases.
- P5,L3: Reference for statement ending with “... kilometre scale models”
- P5,L5: Suggestion for consistency, use “s” in both parameterisation” and “parameterise”, although “z” is an acceptable British spelling.
- P6,L8: Has GL been defined?
- P7,L31: Again remove “%” in table title
- P7,L9: Suggestion for consistency, use “s” in both parameterisation” and “parameterise”, although “z” is an acceptable British spelling.
- P7,L10: Define RMED earlier since it refers not only to the Toolbox but also to the larger collaboration in the UM Partnership on regional model evaluation.
- P7,L18: Define “GPM”.
- P8,L20: “observations” and not “obs”.
- P8,L28: Typing error “together”.
- P9,L1: “reveals that almost all parameters improved during the winter months, as opposed to summer”.
- P9,L8, 4th
- P9,L6, L11 and L12: Use dates syntax consistently and throughout the rest of the paper.
- P9,L11: What were the results for the winter months, or was the sample size too small?
- P10,L24: Suggest changing to “model for a large domain covering ...”, indicated in Figure 14.
- P10,L28: It might be worth mentioning why the results from t+72 onwards are not of significance or not described in the results.
- P11,L7 and 17: Where are these regional areas located? Possibly show on a map to support the reader in interpreting the results.
- P11,L26: U.K to UK
- P11,L19: “The maximum flash counts...”
- P31: Figure 11, 13 – Reduce the space between the title and graph.
- P34: Figure 14: Panels a) to i) is not defined or referred to in the text/Figure title. Suggest using the panel references when discussing the results for ease of reference to the maps. Conversely, Figure 15 for the second case study does not indicate panel numbering.