Reviewer One:

Bennet et al. compare different approaches to detect compound and seesaw events across New Zealand, particularly focusing on the potential value of using multivariate copula-based approaches. They recommend using multivariate indices (incl. soil moisture metrics) for assessing compound hot-dry events, whereas the selection of indices for seesaw events may depend on the hydrological domain (i.e., drought type or dry phase) investigated. The manuscript is a valuable contribution to research on compound events and is for the most part well written. Nevertheless, it would benefit from some corrections outlined below, which the authors may want to consider. Overall, I would recommend publication subject to major revisions.

- We thank the reviewer for their comments and are pleased that they agree that the manuscript makes a valuable contribution. The reviewer makes many good points, which we largely agree with. Direct responses to each reviewer comment follow.

General comments

1. In the Introduction, it would be helpful if the authors placed more emphasis on describing the specific knowledge gap they are addressing and the objectives they want to meet. Specifically, the description of the objectives in the last paragraph of the Introduction is somewhat unclear and would benefit from clear structuring, e.g., by using numbering (1), (2), … to indicate the specific objective.

> Thanks for this observation and suggestion. We have split the final paragraph in two and have re-written the remaining text

2. In the last paragraph of the Discussion, the authors briefly discuss the topic of selecting the indicators for hot-dry and seesaw event detection. As also shown in the Results, detection of hot-dry events is more sensitive to the indicators used compared to seesaw events, where no approach is somewhat superior to another. It would be beneficial if the authors would discuss more in detail the reasons for this discrepancy, possibly also in terms of the different copula indices (SMI and SBI) used for the two event types.

- The sensitivity of compound hot-dry events to dry indicators in comparison to seesaw events reflects the inclusion of temperature within the SMI (but not the SBI). Accordingly, the SMI captures a greater range of variables within the hydrological cycle, thus a single variable comparison will highlight a greater discrepancy. In contrast, single variable comparison to the bi-variate index highlights less of the hydrological cycle components, and less variation therein. A statement to this effect has been added to this final paragraph of

the Discussion – while retaining the emphasis that no one method appears superior in the identification of seesaw events.

The Discussion partly reads like an introduction (especially section 4.1) or like a summary of the results (e.g., lines 506-512). In this regard, the Discussion may be shortened and more specifically focus on discussing the objectives outlined in the Introduction (see also first comment).

- Thank-you for this comment – in retrospect, we agree that parts of the discussion can be shortened, and more direct reference to the objectives can be added at various points.
- In terms of the specific passages highlighted by the reviewer, for Section 4.1 lines 389-394 have been deleted, with the final statement of this section incorporated into the Conclusions.
- For the passage between lines 506-512 (and on re-reading, probably the wider section from lines 499-513) we still think some summary of the results is required for clarity, but agree that it could be shorter. We have re-written accordingly.

Although the authors briefly mention the role of evapotranspiration in the Discussion section 4.2, a more thorough discussion on how seasonality in evapotranspiration influences the indices investigated (related to both compound and seesaw events) would be valuable.

- Thank-you for the suggestion. We agree that evapotranspiration can be an important process with respect to seesaw events and appreciate the desire for thorough discussion. However, without further analysis it is difficult to move beyond speculation here (and adding such a further analytical dimension is beyond the scope of what could be included here) and we are mindful of overall manuscript length. Accordingly, we have added only brief further comment on this matter here.

Minor comments

Line 62: This sentence may rather fit into the paragraph talking about copulas (Line 50 and following)

- This will be implemented.

Line 79: “new insights” is a somewhat vague formulation and may be replaced with a more specific objective (see also first general comment)

- This section of the Introduction has been re-written as per the first general comment.

Line 107: It would be helpful for the reader if the authors would provide the equations for the respective calculated indices

- These have been added (brought over from supplementary material).

Line 159: When describing the Standardised Multivariate Index (SMI), it may be beneficial to shortly repeat the original variables (precipitation, …) contained in the copula data which are then represented by the index

- These have been added.

Line 193: The description of the averaging process at each grid cell is partly unclear (“taking the mean” of what exactly?) and a more thorough explanation would be valuable

- This was incorrectly stated. The correct method is: "The results at each grid cell were summarised by taking the mean across all days".

Line 206: The authors may provide a short explanation of why 14 days were chosen as a threshold length for the compound event duration

- A short explanation is already provided in the manuscript: to minimise the effects of minor (i.e. short duration) compound events.

Line 260: Here, the authors may briefly mention variations of the run theory metrics between the North and South Island (in terms of the coincident SSMI and STI approach), as is also displayed in Figure 3b and 3e

- We believe this discussion is covered by the subsequent lines (275-278), which discuss the variation across the North and South islands, including variation across the west coast of the South Island.

Line 363: “longer transition time” compared to which approaches? It becomes clear from Figure 8, but it would be beneficial to shortly mention the two other approaches in this sentence

- This has been changed to "Seesaw transitions (time from -1 to +1) show an overall longer transition time for SSMI based classifications compared to SPI and SBI based classifications (Fig. 8 b; transition time of 107 days)".

Line 423: This sentence is rather long and may be split

- This has been split into two.

Line 506: This paragraph may be moved to the Results section

- This has been moved.

Line 548: The authors may consider renaming this section “Summary and conclusions”, as the text up until line 572 reads like a summary

- The heading has been changed to "Summary and Conclusions".

Line 581: This sentence is rather long and may be split

- This idea has been split into two.

Technical points

The figures would benefit from increasing the font sizes on the x- and y-axis and legends, particularly figures 2, 3, 6, 7

- Agreed, thanks for the suggestion.

For an overview, a table with the indices (and their abbreviations) used and the respective approaches applied would be helpful for the reader

- This has been added.

Please ensure that headings follow a consecutive order (e.g., 2.5.2 appears two times) and do not repeat themselves (e.g., 2.3.1 Preliminary work on Vine Copula Structure and 2.3.2 Preliminary work on Vine Copula Structure)

- Thank you for identifying. This has been corrected.

In the Supplement, the authors may provide a reference list for the sources they cite, even if they already appear in the original manuscript's reference list

- Thank you for the suggestion. This has been added.

For better coherence, it would be helpful to not repeat content in the Supplement that has already been written in the manuscript (and vice versa); this particularly relates to the Data and Methods sections. In general, I would recommend to shorten the sections describing the methods and refer the reader to the Supplement for more in-depth information, e.g., on the preliminary work conducted (sections 2.3.1 and 2.3.2)

- Preliminary work has been placed into the supplementary section. The multivariate index construction has also been shortened following the inclusion of formula as noted above (which in this case will be shifted from the supplementary into the main).

Reviewer Two

The manuscript presents an interesting and methodologically rigorous analysis of compound and seesaw hydro-meteorological extremes using a copula framework. The study is potentially impactful as it demonstrates the value of multivariate statistical approaches (copulas) in better characterizing the frequency, severity, and spatial distribution of extremes compared to traditional coincident/consecutive approaches. This work is especially relevant for climate risk management in agriculture, water resources, and hydropower, and it provides a transferable framework that could be applied in other regions. However, the manuscript would be significantly strengthened by additional sensitivity analyses and clarifications to ensure the robustness and broader applicability of the results. My detailed comments are below, which the authors may want to consider. Overall, I recommend a major revision for the manuscript.

- We thank the reviewer for their comments and are pleased that they consider this work to be potentially impactful, with applied relevance across a number of domains. We agree with many of the reviewer suggestions and have conducted/included some new analyses accordingly. While interesting, we feel strongly that some other suggestions would require too much additional work and reporting to adequately present and explain within the length restrictions of a single journal paper. We explain our thinking further and provide detail of what new analyses have been conducted, in the point-by-point response that follows.

General Comments

The analysis focuses solely on New Zealand, a strong and relevant case study. However, to strengthen the generalizability of the findings, I suggest testing the methodology in at least one additional region with contrasting hydro-climatic conditions. Even a limited replication using ERA5-Land data would demonstrate whether the framework and event characteristics hold more broadly in the global context.

- Thank you for the suggestion. While we appreciate the added value that an additional region or limited step-up to the global-scale may bring to the manuscript, we also believe that the manuscript is at present quite extensive, and an additional region (or more) will make it difficult to draw out the main points relevant to NZ, and the methodology more broadly. Importantly, NZ actually offers just the contrasting hydro-climatic conditions that the

reviewer mentions, as we already explain in the Introduction (lines 65-70). We propose to further strengthen this justification by also noting that NZ also covers a range of hydrological regimes (e.g. mountain-fed, hill-fed and spring-fed; Biggs et al., 2008).

All indices are constructed at a 30-day scale (monthly focus). However, compound/"seesaw" behaviour may flip across scales; robustness across scales would boost confidence that results aren't an artifact of the 30-day choice. Please assess the sensitivity of event detection to different accumulation periods (e.g., 15, 45, 90 days).

- This is a good point raised by the reviewer. We had already performed the analysis for the 10 and 90 day periods as well, but were unable to include these for reasons of manuscript length. We have instead made brief mention of these sensitivity analyses in the Methods (lines 110-114), with the results presented as Supplementary material. Essentially, these results do not show anything that substantively calls into question the findings from our primary 30-day analyses. We will now acknowledge this in the main body of the manuscript.

Event definitions rely on −1/+1 thresholds and a 30-day buffer for transitions. Results could be sensitive to these choices. It would be useful to repeat the analyses with percentile-based thresholds or alternative buffers (e.g., 15 or 45 days) to demonstrate robustness.

- The use of +/- 1 ensures the relative transitions are captured across the broad climatic regimes of NZ (e.g. one standard deviation from the grid-cell specific climate). Buffer periods were set as the same temporal length as the index in use (e.g. SMI-30 becomes a 30 day buffer) to capture the rolling nature of the accumulation. As noted above, sensitivity testing on this accumulation period selection is captured under the use of differing accumulation periods (10 and 90 days), which are now to be included in the supplementary material.

Including an additional drought index that incorporates evaporative demand (e.g., SPEI or PET-based metric) would strengthen conclusions about hot–dry compounding, particularly in transitional regions.

- We feel that the inclusion of soil moisture, which directly incorporates the actual loss of moisture from the soil rather than a proxy metric offered by the

SPEI and PET based method of atmospheric evaporative demand, offers a more accurate and physically-based representation of the dry component of the hot-dry compound event. The use of PET-based methods would introduce new uncertainty to the analysis due to the calculation procedure chosen, as well as the increasingly limited physical relevance of evaporative demand as the land surface dries out. In contrast, the use of soil moisture (and by extension AET within the land model component) avoids these uncertainties. While a study of atmospheric evaporative demand during hot-dry compounding events may provide additional insight, particularly for transitional regimes, we believe that the statistical modelling of the direct relationship between actual evaporative fluxes, soil moisture and temperature (via the copula approach) performed here provides a more robust approach. Furthermore, we feel that introducing a different conceptual approach (such as PET) to the study would require extensive additions to the Methods, Results and Discussion sections of the manuscript, which would ultimately be to the detriment of manuscript readability as well as the core messages from the existing more physically-based analyses.

Validation is focused on statistical consistency. Demonstrating added value through alignment with real-world impact proxies (e.g., hydro power inflows, agricultural losses) would enhance the applied relevance of the results.

- Thanks for the suggestion. We agree that aligning the index with real-world proxies could add value; however, doing so would extend the study beyond its current scope. It would require sourcing and processing multiple datasets from different institutions across New Zealand, many of which are not readily or freely accessible. While publicly available alternatives exist, these are typically only available at broad (e.g. national) spatial scales, raising data quality concerns. A qualitative assessment using known events is another option, but this would shift the manuscript toward event-specific analysis—where definitions and classifications (e.g. seesaw versus compound events) remain uncertain—rather than the intended methodological comparison.

Results are presented per grid cell, but spatial clustering of extremes (event footprint, coherence across regions) is not explored. Such an analysis would enhance the understanding of regional impacts and risk management implications.

- A spatial analysis (cluster) was performed on the index values themselves analyses and was used to inform both our Methods and Discussion, with regions broadly conforming to existing understanding of regional climate

variation across New Zealand. The results of this cluster analysis are shown in Figure S8, with brief accompanying text in the main manuscript (lines 173-176 and 416-417).

Other Comments

"Hot and dry" vs. "compound hot-dry" – Please check the use of terminology with meaning and keep them consistent.

- This has been adjusted accordingly, thank you for identifying.

Typo: "Charaacteristics" should be Characteristics (section 2.5.2 heading).

- This has been corrected.

One reference (Miller, B. B. and Carter, C., 2015) appears to be a test citation - please check.

- Thank you for identifying, as this was not part of the original manuscript – we have corrected accordingly.

Line 400: placeholder Latin text remains in the Discussion; please revise or remove.

- Thank you for identifying – we have corrected accordingly.