Response to the reviewers of egusphere-2026-2465: Subseasonal prediction of compound heat and drought events

Dear Editor Kai Schröter,

On behalf of all authors I would like to submit the revised version of the original article "Subseasonal predictability of compound heat and drought events" including an annotated version of the modified manuscript.

Best regards,
Rachel Wu

General response

We thank both reviewers for their very helpful and constructive comments. These comments have motivated the revision of the manuscript to significantly improve the flow and the presentation of the main messages of the manuscript. The principal findings and overall message remain unchanged.

Major changes to figures:

- We moved the BSS_{CHD_ind} figure (originally Fig. 3) to the Supplementary Material (Fig. D1), as the analysis is now interpreted as a sensitivity of BSS to the climatological reference and provides limited additional information beyond the COR analysis.
- We repeated the hindcast verification analyses using GPCP. The GPCP-based results are qualitatively consistent with the ERA5-based results over Europe and are now provided in the Supplementary Material (Fig. E1-E3).
- We revised figure layouts by reducing excess whitespace, increasing the font size of key labels and correlation values, and reversing the FAR colour scale (Fig. 3, originally Fig. 4) for a more intuitive comparison with the hit rate.

Major changes to text and interpretation:

- We reorganised Sect. 3.2 to interpret the BSS together with the hit rate and false alarm ratio. This clarifies why compound event skill remains relatively high in some regions despite biases in mean event frequency, and why skill is low in Scandinavia and Eastern Europe for different reasons.
- We clarified the temporal aggregation and climatology procedure in Sect. 2.1. The 31-day window selects consecutive initialization dates to construct a smoothed climatology; all target values and climatological samples use the same weeks 1–4, 28-day aggregation window.
- We have refined the abstract and conclusions according to the reviewer comments. We have removed the results inferred from *BSS_ind* and corrected the statements describing the land-atmosphere coupling biases in the conclusions.

1 Reviewer 1

The manuscript **"Subseasonal predictability of compound heat and drought events in Europe"** by Wu and co-authors presents the predictability of compound heat and drought events in summer by the ECMWF models. The article is concise and dedicates most of the space to introduce clearly the methods used to evaluate the predictability and the role of the atmosphere-land coupling in the skill obtained. I think the article is well structured, clear, and fits in the scope of the journal. I only have minor comments to the authors:

- **Appendices:** Could you please reorder the appendices (first C, second B, third A, fourth D) so the first one introduced is Appendix A?

  Response: Thank you for the suggestion. We have reordered the appendices so that they are introduced sequentially in the same order as they are introduced in the manuscript.

- **Figures C1–C3:** Could you please reduce the blank spaces between figures? (This also applies to Figures 1, 4, and 5.)

  Response: Thanks. We have updated all figures with reduced spacing between panels.

- **Line 131:** I'd rather use "forecast" as the past participle, instead of "forecasted".

  Response: Thanks. We have replaced 'forecasted' with 'forecast' at Line 136.

- **Lines 209–210:** Interestingly, this only happens for positive biases. Regions with negative biases show fair skill. Any comment on this?

  Response: Thank you for this comment. Indeed, in regions where compound-event frequency is underestimated, including Central Europe, Southeastern Europe, the British Isles, and western Russia, the hindcasts nevertheless retain useful discrimination between compound events and non-events, as indicated by relatively high hit rates and low false-alarm ratios (Fig. 3c,f). We have added text in Lines 212-224 to clarify this interpretation.

- **Figure 6:** Could you put the legend outside panel a and increase its font size? Could you also increase the font size of correlations in all panels? I struggle to read the values there. This also holds for Figure D1.

  Response: Thank you. We have moved the legend to the bottom of the panels and increased font sizes for both the legend and correlation values in Figure 5 (originally Figure 6) and in Fig. F1 (originally Figure D1).

2 Reviewer 2

Major Comments

1. The conclusion suggests that the soil moisture–temperature (SM–T) coupling is underestimated, while the soil moisture–precipitation (SM–P) coupling is overestimated. Given that temperature and precipitation are not independent variables, I would appreciate a more in-depth discussion of the physical mechanisms that could lead to such contrasting biases. In addition, it would be helpful if the manuscript explicitly defined what processes are represented by the SM–T and SM–P coupling metrics. Relevant discussions in Santanello et al. (2018) may provide useful context.

   Response: Thank you very much identifying this inconsistency. Our original conclusion incorrectly described the SM-T bias. The hindcasts exhibit a negative SM-T coupling bias and positive SM-P bias. Given the signs of the bias, these biases indicate an overestimation of coupling strength along both pathways. We have added texts to guide the readers the interpretation of the SM-coupling metrics in Lines 181-184 in Sect. 2.6 and Lines 253-258 Sect 3.3. We have also corrected and clarified in Lines 295-297 in Sect 3.3. and in Lines 316-322 in the conclusions. We also thank the reviewer for suggesting the helpful reference (Santanello et al, 2018), which we have now cited in Line 257.

2. **Line 230:** The manuscript states that the high predictability of heat contributes to improved prediction of compound events. Could the authors clarify the mechanism behind this statement? Also, does the "heat" category include compound heat–drought events, or does it refer only to heat-only events? This distinction is important and should be explicitly stated in the manuscript.

   Response: Thank you for this comment. We agree that the original sentence was unclear and potentially misleading, and have now removed the sentence. Heat and drought events include events that are also classified as compound events. We now state this explicitly in Lines 96-98.

3. Would the results shown in Figures 1–4 remain qualitatively unchanged if GPCP precipitation were used instead of ERA5 precipitation? Since Figure 5 is linked to the interpretation of precipitation-related land–atmosphere coupling processes, using GPCP consistently throughout the analysis may provide a more coherent framework.

   Response: Thank you for this valuable suggestion. We repeated the hindcast verification analyses using GPCP. The resulting spatial patterns are qualitatively similar to the ERA5-based results. These analyses using GPCP are now included as Figs. E1-E3 as supplementary material. We retain the ERA5-based figures in the main text because they provide the higher-resolution reference (retrieved on a $4^\circ \times 4^\circ$ grid, same as the hindcasts) used in the primary verification framework, whereas GPCP is available on a coarser grid ($1^\circ \times 1^\circ$). We have clarified in the revised manuscript that the GPCP analysis provides a robustness check. The relevant sentences are in Lines 90-92, Lines 244-247.

4. The discussion focuses on regions where compound-event predictability is degraded by misrepresented land–atmosphere coupling. Conversely, do regions with relatively high predictive skill also exhibit more realistic representations of the land–atmosphere coupling processes shown in Figure 5? Additional discussion on this aspect would strengthen the manuscript. For example, it may be useful to examine regions east of Eastern Europe and parts of Western Europe where skill is comparatively high.

   Response: Thank you for this suggestion. Land-atmosphere coupling biases are indeed relatively small over these high skill regions, particularly for SM-T coupling (Fig. 4). Yet, these small biases do not provide direct evidence for their high compound event skill, since land-atmosphere coupling is originally weak or absent in the observations over these regions. We therefore hypothesize that the high skill over these regions, the relatively high hit rates and low false alarm ratios, might be contributed by other factors, such as the representation of large-scale circulation, but investigating this is beyond the scope of this study. We have added the relevant discussion of the potential sources for the high skill of the compound hotspots in Lines 325-328.

Minor Comments

- **Section 2.1:** I am having difficulty understanding the centered 30-day approach used to construct the climatology. Is this methodology different from that used in previous subseasonal forecast skill studies? The analysis focuses on Weeks 1–4, which implies that only lead days 1–28 are considered, whereas the climatology appears to be calculated using a centered 30-day window. If the monthly-mean anomaly is computed from a 28-day forecast period but referenced against a 30-day climatology, the comparison may not be entirely consistent. Please clarify this methodology and consider using matching periods if necessary. In addition, please clarify whether "aggregation" refers specifically to averaging, or whether it may also involve summation.

  Response: Thank you for pointing out that our original description was unclear. We have rewritten this section to clarify that all target values are aggregated over the same lead time window (weeks 1-4, i.e. lead days 1-28). The centered 31-day window refers only to the selection of consecutive initialization dates used to construct a smoothed climatology. We have clarified this in Lines 59-67.

- **Lines 40–44:** While I understand the point being made, I am not sure why this discussion is needed in the current context. The connection to the main objectives of the manuscript is not immediately clear.

  Response: Thank you for the comment. Indeed this might be too much in detail and might have disrupted the flow of the introduction. We have removed the relevant lines in the introduction since the motivation of temporal aggregation is already discussed in Lines 108-113 in the methods section.

- **Lines 78–79:** This sentence appears to fit more naturally at the end of the first paragraph. Unless the methodology is highly specialized, it may be more reader-friendly to briefly describe the approach (2–3 sentences) rather than only referring the reader to previous studies.

  Response: Thank you for your comment. We have now moved the sentence to the first paragraph and elaborated details from Büeler et al (2026). These can be found in Lines 59-67.

- **Equation 1:** Please define the indices i, j, n, and y explicitly in the text.

  Response: Thank you for the comment. We have added the text in Lines 116 - 118.

- **Lines 110–117:** This paragraph may fit more naturally immediately after Line 95 or Line 101.

  Response: Thank you for the suggestion. We have moved the paragraph to after the first paragraph in Section 2.3 (Lines 99-107).

- **Equation 3:** The formulation appears different from that used in previous studies. Could the authors explain why the additional term following $(p_k - o_k)^2$ is included?

  Response: Thank you for pointing this out. We use the fair Brier score rather than the standard Brier score. The additional term corrects for the effect of finite ensemble size on the Brier score, following Ferro et al. (2008). We have clarified this in Lines 138-139.

- **Lines 142–151:** This appears to be one of the most novel aspects of the manuscript, but I found it difficult to follow. Please provide additional explanation regarding the physical and statistical interpretation of $\mathrm{BSS}_{\mathrm{CHD\_ind}}$ and how it should be interpreted relative to the standard $\mathrm{BSS}_{\mathrm{CHD}}$.

  Response: Thank you for highlighting this issue. After reconsideration, we concluded that the comparison between $\mathrm{BSS}_{\mathrm{CHD}}$ and $\mathrm{BSS}_{\mathrm{CHD\_ind}}$ is primarily a sensitivity test to the climatological reference, rather than a direct measure of the contribution of physical coupling to forecast skill. The original interpretation overstated its physical meaning. We have therefore moved the figure to the Supplementary Material (Fig. D1). The relevant section is in Lines 225-231.

- **Figure 4:** It may be helpful to reverse the FAR color scale so that the visual interpretation is consistent with HR (i.e., colors associated with better performance correspond to the same direction in both panels).
  Response: Thank you for the suggestion. We have reversed the colorbar in Fig. 3 (originally Fig. 4) for better consistency.