Responses to referee 2 comments about the article

“Spatial structures of emerging hot & dry compound events over Europe from 1950 to 2023”

By Schmutz, Vrac, François, Bulut

This manuscript by Schmutz et al introduces a new framework within which to consider the emergence of hot and dry compound events over Europe and North Africa. Building on the extant concept of Time of Emergence (ToE), the authors introduce a Period of Emergence (PoE) which generalizes the ToE to include emergence [above/below the natural variability] that does not last until the end of the analyzed time period. The authors claim that this allows better characterization of the resulting signal by analyzing the number and duration(s) of PoE events in the time record. Further development of the framework allows the authors to introduce a budget-type analysis examining the role that the individual components (temperature and drought) play in this emergence, as well as their combination and interactions. To demonstrate this framework, the authors use the ERA5 reanalysis dataset to look for changes in compound events in Europe and North Africa between 1950 and 2023.

Response:

We appreciate the time and effort that the referee dedicated to providing feedback on our manuscript and are grateful for the insightful comments on our paper. We have incorporated most of the referee’s suggestions. Those changes are highlighted within the manuscript. The reviewer can find below, in blue, a point-by-point response to the comments and concerns.

General Comments

Overall, this manuscript lays out the foundations for the scientific gap addressed well and rightly highlights the myriad of ways in which heatwaves, droughts, and their co-occurrence affect human life, society, and the environment at large. More detail could be included about the data used, which I have outlined further below. As the main novelty within this manuscript, the methods and underlying reasoning are laid out in a comprehensive manner, including reference to finding the analysis code online. Areas where the authors include slightly less technical language to summarize the steps they've made are appreciated and add accessibility to the manuscript beyond those who are primarily statistical modelers. The resulting analysis is also quite comprehensive, although there might be a slight dependence on supplemental figures through this section (although I appreciate the authors transparency in sharing all figures associated with the analysis!). I have some reservations, mainly about the underlying data and assumptions of some of the analysis, outlined below in my specific comments, and followed by technical corrections. While a full exploration of the issues raised in the specific comments would entail additional reanalysis that the authors could choose to pursue, I believe that a more complete explanation of their choices as to the data in the manuscript would also be sufficient.

Response:

We fully understand the concerns raised by the reviewer regarding the data. Below, we address each of the comments individually, including justifications for our choice of variables.

Additionally, we have included a supplementary analysis to support our choice of reference period. The reviewer can find our detailed responses to each comment below.

Specific Comments

1.) When attempting to understand the framework and analysis methods in the paper, perhaps I overlooked it, but it is unclear to me how exactly the monthly temperature and SPEI data is handled within the framework - how the three data points become one for the year, for example. More clarification on this process would help to make the transition from the data introduction to the analysis products clearer.

Response:

We have three data points per year, corresponding to monthly data for the months of June, July, and August. The probability signal is calculated using a 20-year sliding window, with the central year (more precisely the 11th year) of each window assigned as the representative year. For example, the first window spans 1950 to 1969, giving 60 data points in total. The probability of a compound event (CE) is computed over this period, and the year 1960—being the representative year—is used to represent it. The next point, for 1961, corresponds to the probability computed over the window from 1951 to 1970, and so on.

We have addressed it by adding a clarification in Section 2.2 (lines 148-152 of the revised article):

> *[Compound event probability] is estimated at each grid point within the study area. For a given threshold (x_e,y_e), the probability is computed over a 20-year sliding window, which advances by one year at each step. Each 20-year window (comprising 20 years × 3 months = 60 data points) is then associated with its central year for analysis purposes. Between 1960 and 2014, this results in 55 annual probability values, each estimated from its corresponding 60-point window.*

2.) Further explanation for the choice of using monthly maximum temperatures would help to provide context for the results. Although maximum temperatures are certainly important and crucial to identifying impactful heat wave events, increases in minimum temperatures can also play a significant role, especially in thinking about mortality applications of this type of analysis. Are similar results achieved for substituting in the monthly maximum of the daily average temperature for example? Additionally, how is the analysis affected by use of monthly aggregated temperatures versus weekly for example, especially in terms of using a single point in time to represent the month?

Response:

The use of monthly temperature is guided by the choice of the drought index. We used the 6-month Standardized Precipitation Evapotranspiration Index (SPEI6) to represent drought, as it captures the water balance over six months and is well-suited for identifying agricultural drought (Ionita and Nagavciuc, 2021). We have added this clarification on lines 120-125 of the revised paper:

The [SPEI] is based on the difference between precipitation and evapotranspiration, reflecting the water balance. The SPEI indicator, considered to be one of the best for drought monitoring (e.g., Ionita and Nagavciuc, 2021; Blauhut et al., 2016), is used for the present study as it combines the advantages of SPI, with its variety of timescales, and those of PDSI with the consideration of temperature evolution. Depending on the chosen accumulation period, it can capture the different types of drought (Ionita and Nagavciuc, 2021). Then the analysis is performed at a monthly time scale to match the temporal resolution of the SPEI.

The method developed in this study is generic and can be applied to any bivariate compound event. Modifying the variables, spatial scale, or temporal resolution will naturally lead to different compound event probabilities, and consequently, different outcomes in terms of emergence and component contributions. The reviewer highlighted an interesting avenue for future work: exploring how changing variables (like minimum temperature) affects the results. We have incorporated this perspective on lines 474-479 of the revised article :

"The method developed in this study is generic and applicable to any bivariate compound event. Changing the variables, spatial scale, or temporal resolution will naturally change the event probabilities and, in turn, highlight different patterns of emergence and contribution. Further analysis of other bivariate events, such as extreme precipitation and wind, or alternative representations of hot and dry conditions (e.g., minimum temperature with precipitation, or the number of hot days with SPEI3), across different spatial and temporal scales, could yield valuable insights for future research tailored to the needs of societal stakeholders."

3.) While understanding that the authors are somewhat boxed in given the ERA5 dataset limitations in time, I'm concerned about the reference period for emergence only being from 1950-1969. How does this choice affect the results afterwards? Use of only a two-decade period may lend itself to an underestimation or transposition of where the natural variability in the system lies, depending on if/how that period was itself anomalous - which then cascades downstream by affecting the ToE and PoE thresholds. Have the authors considered longer reference periods or other time slices as their periods of reference?

Regardless, along this line, the detection and definition of multiyear to multi-decadal oscillations strikes me as both an issue that plagues the reference period as well as an intriguing potential application of the PoE method to climate data analysis.

Response:

The choice of the reference period is a key factor in interpreting results. We selected a 20-year period (1950–1969) as it is the earliest available, it offers greater stationarity, and is less affected by long-term global warming trends. It is essential to keep this choice in mind when analyzing the results. We tested the sensitivity of outcomes to changes in the length and timing of the reference period.

We compared two experiments: one using a 20-year sliding window with 1950–1969 as the reference for natural variability (as in the submitted article), and another using a 30-year sliding window with 1950–1979 as the reference. The results are shown below on Fig. 1. The signal from the 20-year window is colored in black; the 30-year, in blue. Natural variability is shown in red and green for the 20-year and 30-year reference period, respectively. The figure below illustrates the differences for the five locations (similarly to Figure 6 of the paper). Results show that the choice of reference period affects both variability and Time of Emergence (ToE): it can increase variability (e.g., Fig. 1d), decrease it (e.g., Fig. 1e), delay ToE (Fig. 1e), advance it (Fig. 1b), or have little effect (Fig. 1a). Hence, changing the length of the reference period used for estimating natural variability can affect the detection results. However, where a ToE-up is detected, it remains present regardless of the reference used, reinforcing the robustness of the findings. Regarding PoE, PoE-low is preserved (Fig. 1b), while PoE-up disappears in Fig. 1e. The interpretation of PoE should always be made in light of the chosen reference period. Depending on the goal, one may prefer a longer or shorter baseline.

We also tested the sensitivity to the choice of baseline period for natural variability, keeping the same signal as in the article (20-year sliding window). The results are shown below on Fig. 2. We compared two baselines: 1950–1969 (as in the article, in red) and 1960–1979 (in green), across the same five locations. From the 1960s to 1980s, Europe experienced particularly low compound event (CE) probabilities, reducing natural variability (Fig. 2a-c). In Fig. 2d, the signal drops around 1970, and its confidence interval falls below the original natural variability (1950-1969 in red). In contrast, Fig. 2e shows an increasing signal with an uncertainty in the 1970s lying above the red natural variability. This demonstrates that the effect of changing the reference period depends on the shape of the signal. Choosing a more recent reference is also possible, but less meaningful in terms of emergence.

This flexibility should be seen as an opportunity: the methodology can be adapted to suit different analytical needs, offering broader perspectives. We thank the reviewer for highlighting this important point and have added the following in the revised manuscript (in the Perspectives section, lines 480-490):

> *The method is adaptable to any reference period. In this study, we used the 1950–1969 period for its early timing, stationarity, and limited global warming influence. Reference periods are often around 30 years (with reanalyses), and up to 50 years (with simulations). However, several studies have used shorter baseline periods, such as the 20-year window from 1980 to 1999 (Giorgi et Bi., 2009, Diffenbaugh,et Sherer., 2011, Vrac et al., 2023). Diffenbaugh,et Sherer., (2011) found similar results for the global emergence of permanent, unprecedented heat in the 20th and 21st centuries when using either the 1980–1999 or 1951–1999 reference period. Zscheischler et Lehner., (2022) even adopts a 10-year baseline period (1950–1969) for a compound event attribution study. Then other baseline periods (whether longer or set at different times) can be used to analyze the emergence of compound event probabilities. The choice of the reference period is crucial when interpreting detection results, as detection is inherently relative to this baseline. Changing the reference period can alter the estimation of natural variability, either narrowing or widening it, which in turn affects Periods of Emergence (PoE). This flexibility allows the analysis to be adapted to different research objectives.*

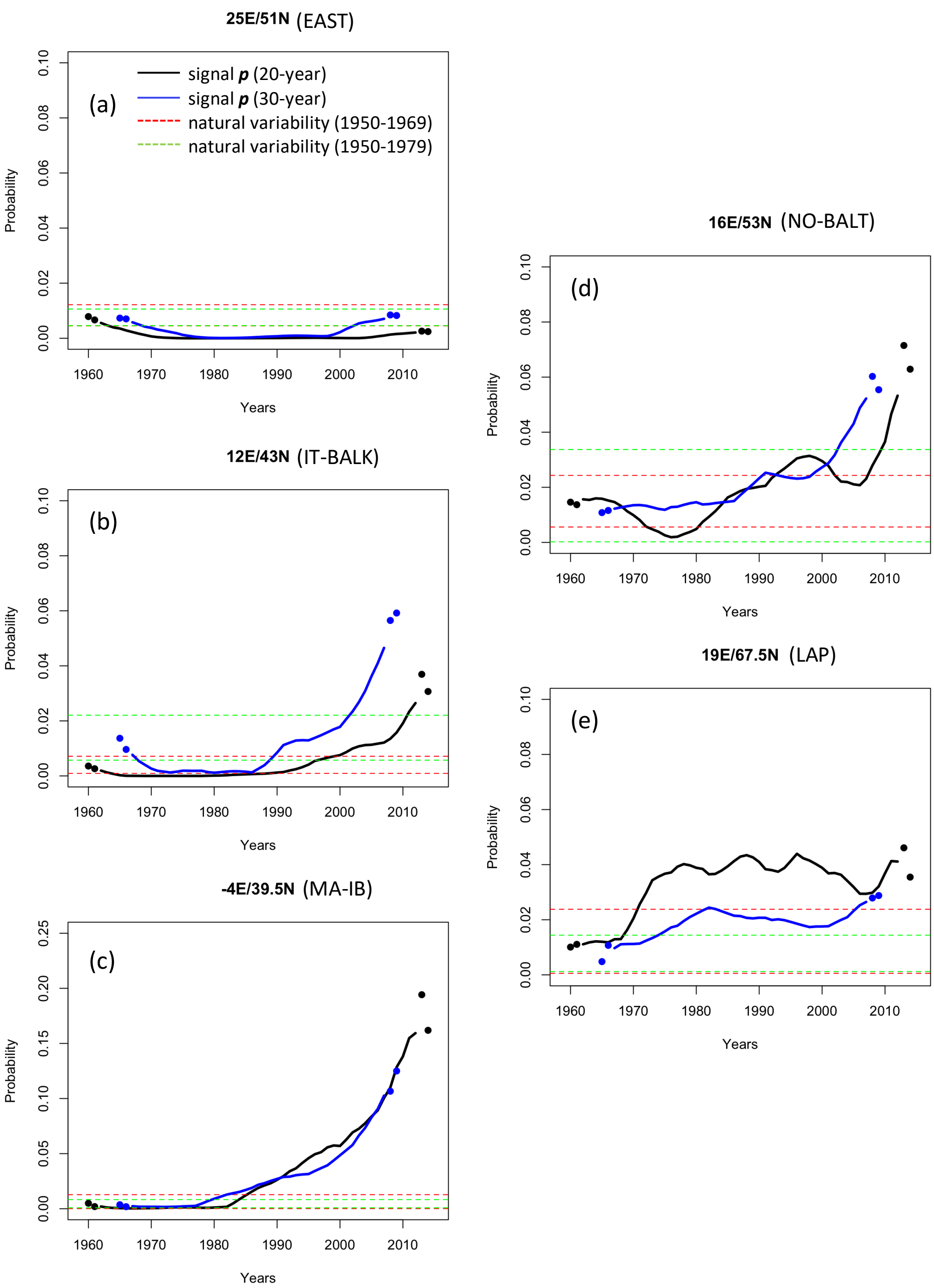

Figure 1 : Evolution of hot-dry compound event probability computed on a 20-year and 30-year sliding window, for 5 points located in the 5 areas under study.

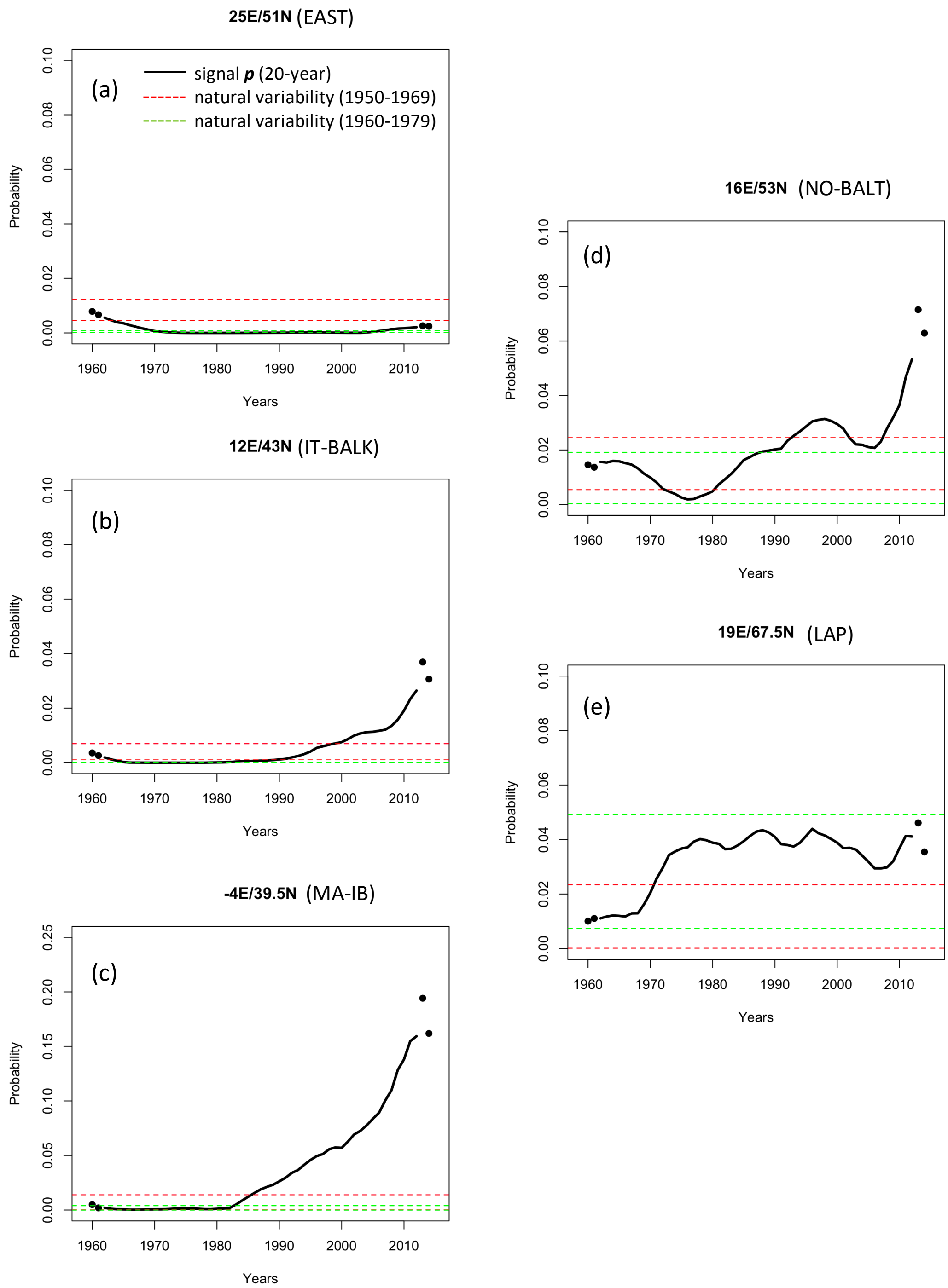

Figure 2 : Evolution of hot-dry compound event probability computed on a 20-year for 5 points located in the 5 areas under study.

4.) Similarly, does a different selection for the preceding time period of SPEI (other than 6 months, perhaps 3 or 4) have a significant effect on the results? If so, this could also enlighten differences as to the effect that winter versus spring moisture amounts have on these CEs.

Response:

We selected SPEI6 to better capture winter climate conditions, as both Russo et al. (2019) and Oikonomou (2020) highlight the relevance of a 6-month timescale compared to shorter or longer durations. As noted in our response to referee's comment 2, changing the heat or drought index would influence the results.

For instance, Russo et al. (2019) analyzed the correlation between the number of hot days and nights and SPEI at 3-, 6-, and 9-month timescales, finding that the strongest negative correlations occurred at 3 or 6 months, depending on the region. We include this example as a future perspective to emphasize that different SPEI timescales may reflect distinct physical mechanisms behind hot and dry events. We have added this point on lines 474-479 of the revised manuscript:

> *"The method developed in this study is generic and applicable to any bivariate compound event. Changing the variables, spatial scale, or temporal resolution will naturally change the event probabilities and, in turn, highlight different patterns of emergence and contribution. Further analysis of other bivariate events, such as extreme precipitation and wind, or alternative representations of hot and dry conditions (e.g., minimum temperature with precipitation, or the number of hot days with SPEI3), across different spatial and temporal scales, could yield valuable insights for future research tailored to the needs of societal stakeholders."*

Generally, SPEI3 represents meteorological drought, while SPEI12 is more indicative of hydrological drought. For a comprehensive comparison of SPEI with other drought indices across timescales and regions, we recommend to read Vicente-Serrano et al., (2010). High correlations between SPEI3 and SPEI6 have been observed (e.g., ~0.7 in the USA; Poudel et al., 2024), and SPEI3 tends to show higher variability (e.g. in China; Mokhtar et al., 2024 and in Ethiopia; Mulualem et al., 2024). Therefore, using SPEI3 instead of SPEI6 may reveal similar patterns of emergence and contribution, but with shorter and less frequent periods of emergence due to the increased variability.

Technical Comments

Generally, there are some verb tense disagreements, but these are simple enough to fix up.

Response:

We corrected the verb tense inconsistencies, with modifications shown below: green strikethrough for deletions, and bold blue underline for additions.

- Lines 22-25 of the revised manuscript:

*For several years, Europe **has** faced severe hot and dry events, corresponding to a combination of drought and heatwave. This climate ~~phenomena~~ **phenomenon**, such* *as during summer 2018, has significantly affected various sectors of society. It **has** impacted biodiversity with an increased fire risk during this period (San-Miguel-Ayanz et al., 2018), agriculture with yields falling by up to 50% (Toreti et al., 2019), and human health with an increased number of deaths (Pascal et al., 2021).*

- Lines 52-53 of the revised manuscript:

*The question of multivariate emergence **has** recently **arisen** ~~arises~~.*

- Lines 420-421 of the revised manuscript:

*Compound events are the most impactful phenomena. Changes in ~~bivariate~~ compound hot & dry event probability (signal) relative to the natural variability (noise) ~~was~~ **were** detected in terms of timing and location.*

- Line 437 of the revised manuscript:

*Five areas were studied in ~~details~~ **detail**, each characterized by their own specificity in terms of ToE, PoE, contributions*

- Line 441-442 of the revised manuscript:

*Finally, the dependence change ~~impacts differently~~ **impacted** Baltic states and Lapland **differently**.*

- Line 446 of the revised manuscript:

*The study ~~showed~~ **shows** that the developed approach can be adapted to the analysis of a specific impactful CE.*

115 - ressources --> resources

Response:

We have taken this spelling error into account.

Lines 115-116 of the revised manuscript:

*Three other types of drought have major impacts on society: hydrological drought, related to low surface and subsurface water ~~ressources~~ **resources**;*

117 - Additional "Drought" included in the PDSI acronym

Response:

We have corrected this error.

Lines 118-120 of the revised manuscript:

This classification implies the use of several drought indices, such as: the Palmer Drought Severity ~~Drought~~ Index, PDSI (Palmer, 1965), the Standardised Precipitation Index, SPI (McKee et al., 1993) and the Standardised Precipitation-Evapotranspiration Index, SPEI (Vicente-Serrano et al., 2010).

119 - It is made clear late that "this indicator" is the SPEI, but it might be good to specifically mention it here, as it is indirectly referenced twice in a row.

Response:

We added “SPEI” for more clarity.

Lines 121-124 of the revised article:

*~~This~~ **The SPEI** indicator, considered to be one of the best for drought monitoring (e.g., Ionita and Nagavciuc, 2021; Blauhut et al., 2016), is used for the present study as it combines the advantages of SPI, with its variety of timescales, and those of PDSI with the consideration of temperature evolution.*

262 - an other --> another

Response:

We have corrected this error.

- Line 268 of the revised manuscript :

***Another** difference relies on the sign of the contribution. If both signals p and pZ decrease, the Bevacqua et al. (2019)’s contribution metric would be negative, while our contribution value would be positive as p behaves like pZ .*

- Lines 352-353 of the revised manuscript:

*Spatial patterns of their ToE and PoE are shown in Figs. S5 and S6 in Supplementary and provide **another** overview of the contributions.*