**ANONYMOUS REVIEWER #2:**

GENERAL COMMENT

This manuscript examines the vertical structure of persistent warm anomalies over Rome using ERA5 data and a clustering approach. I find the effort to look beyond surface temperature interesting, and the combination of climatological analysis, composites, precursor evolution, and case studies has potential. I have, however, several questions about the motivation for the framework, some methodological choices, and the interpretation of the results. Addressing these would make the paper clearer and more convincing.

We thank the Reviewer for her/his constructive overview and for recognizing the value of moving beyond surface temperature to analyse the vertical structure of extreme warm events over Rome. We appreciate the positive assessment of our multi-faceted approach. We have carefully addressed all questions regarding our methodological framework, pressure-level selection, statistical trend robustness, and synoptic/dynamical interpretations. Please find below the point-by-point responses to each specific comment.

Please find below a point-by-point reply to the Reviewer's comments. The complete, original Reviewer's comments are given in black, and our responses are listed in blue. The revised text is shown in italics and the Line number refers to the revised version of the manuscript.

In the revised version of the manuscript, modifications are marked in red. Please note that, in the revised manuscript, the original figures (Figure 1, Figure 2, Figure 3, Figure 5, Figure 6, Figure 7) have been updated and replaced with expanded figures in direct response to the Reviewer's requests.

MAJOR COMMENTS

**Q1**: Motivation for the vertical framework. I think the manuscript would benefit from a clearer explanation of why it is useful to include events confined to the middle or high troposphere that have no clear surface signature. For surface heatwaves, the relevance is immediate, but it is less clear to me why a persistent warm anomaly at 300 hPa should be considered an "Extreme Warm Event" in the same sense. What is the physical or practical motivation for focusing on these high-level events, and what do they add to our understanding of surface heat extremes?

The introduction should also discuss more fully the growing literature on the vertical structure of heat extremes. In particular, the recent study by Hotz et al. (2026), Global characterisation of the vertical temperature anomaly structure of heat extremes over land in ERA5, seems highly relevant.

**A1:** We thank the Reviewer for this question. The physical motivation for including events confined to the middle and high troposphere (even without a surface heatwave signature) lies in understanding the complete three-dimensional structure of tropospheric anomalies. High- and mid-level persistent warm anomalies are often driven by upper-level dynamics that impose thermodynamic forcing on the column. Focusing exclusively on surface HWs filters out atmospheric events where strong upper-level thermodynamic forcing is present but remains decoupled from the near-surface atmospheric layers. On the contrary, by tracking all EWEs at selected isobaric levels regardless of surface occurrance, we can directly investigate the conditions that allow an upper-level anomaly to couple dynamically with the PBL versus cases where lower-tropospheric mechanisms (e.g., marine boundary layer stratification or cold-air advection) decouple the surface. In addition, upper-tropospheric warm anomalies often serve as dynamical indicators or precursors of atmospheric blocking and Rossby wave breaking before full-column vertical coupling occurs.

Regarding the literature, we have expanded the introduction and discussion including references to recent studies on the vertical structure of extreme temperatures, such as Hotz et al. (2026), and framing our regional, multi-level clustering approach within this growing body of research.

The text has been modified as follows:

LL. 76-89: *"Recently, a growing body of literature has emphasized that heat extremes are inherently three-dimensional atmospheric features driven by complex vertical structures and circulation patterns (e.g., Miralles et al., 2014; Sousa et al., 2018; Hotz et al., 2026). Hotz et al. (2026) also provided a comprehensive global baseline of vertical temperature anomaly profiles during HWs, showing that extreme thermal anomalies exhibit substantial vertical variability and are often decoupled between the PBL and the free troposphere. While upper-tropospheric thermal extremes do not directly impact terrestrial ecosystems or human health at the surface, their diagnostic value lies in their role as dynamical indicators of large-scale atmospheric patterns (Pfahl et al., 2015; Zschenderlein et al., 2019). In fact, analysing high-level events that do not reach the surface allows us to isolate the near surface and lower-tropospheric mechanisms (such as cold-air advection, marine boundary layer stratification, or sea-breeze regime) that decouple the near surface atmospheric layer from intense mid-tropospheric ridging. In addition, high-tropospheric anomalies frequently act as dynamic precursors to synoptic-scale blockings, offering a diagnostic window into the vertical evolution and downward propagation of thermodynamic anomalies prior to full-column coupling (Sousa et al., 2018). Analysing the entire vertical column allows for a comprehensive classification of warm events, distinguishing deep, surface-impacting HWs from dynamically isolated upper-tropospheric thermal features (Lhotka and Kyselý, 2024; Hassoun et al., 2025)."*

LL. 226-228: *"In doing so, upper-level EWEs are treated not as 'failed' surface HWs, but as distinct thermodynamic regimes governed by upper-tropospheric wave dynamics and adiabatic compression, providing a clear benchmark for when and why synoptic forcing successfully penetrates the boundary layer."*

**Q2:** Choice of pressure levels and seasonality. The dataset contains 15 pressure levels, but the event definition and clustering use only 1000, 500, and 300 hPa. I would like to see a stronger justification for this choice. Why not use the full vertical profile for the classification, especially given the emphasis on three-dimensional structure and vertical coupling? With only three levels, it is difficult to know how well the identified categories represent the intervening atmosphere.

I also wonder how strongly the results depend on season. The vertical structure and physical mechanisms of warm anomalies are likely to differ substantially between winter, summer, and the transition seasons. This could affect the clustering and the composites. It would be helpful to show the seasonal distribution of events in each regime and discuss this issue more directly.

**A2:** We thank the Reviewer for addressing these two important points. With respect to the choice of the pressure levels, the decision to rely on 3 key pressure levels for the event identification and the clustering was driven by methodological robustness. In fact, including all 15 ERA5 pressure levels introduces high vertical autocorrelation and redundant information (as adjacent vertical levels in ERA5 are strongly correlated, Hersbach, H., et al., 2020, https://doi.org/10.1002/qj.3803), possibly resulting in smoothing of cluster boundaries and increased overfitting noise rather than isolating distinct dynamical regimes. As detailed in the text, the 3 selected levels serve as physically-based references: 1000 hPa represents the near-surface atmospheric level; 500 hPa represents the mid-tropospheric steering level and synoptic patterns; and 300 hPa captures upper-tropospheric dynamics. Additionally, although the EWEs were investigated from only 3 levels, the three-dimensional structure and vertical coupling of the events are analysed in detail by considering all 15 ERA5 pressure levels in the case studies and regime cross-sections. This confirms that these 3 levels effectively act as key proxies and accurately capture the thermodynamic behaviour of the intervening layers. We have expanded Section 2 to clarify this aspect as follows:

LL. 176-179: *"This selection is based on physical and statistical considerations. In fact, adjacent pressure levels in high-resolution reanalyses exhibit strong vertical collinearity. Including all 15 levels in the clustering algorithm's distance matrix may introduce redundant variance, which tends to degrade cluster separation and excessively flatten regime boundaries. In addition, the three selected levels strategically sample three distinct functional layers of the atmosphere."*

With respect to the seasonality of events, we fully agree with the Reviewer that the physical drivers of warm anomalies differ substantially across seasons. To address this comment and also a similar one from another Reviewer, we have conducted a comprehensive seasonal analysis by partitioning all identified EWE days into the four standard meteorological seasons: Winter (DJF), Spring (MAM), Summer (JJA), and Autumn (SON). First of all, we independently evaluated the optimal number of clusters (K) for each seasonal subset using the Elbow Method and Silhouette Analysis. The dual-metric evaluation yielded an optimal configuration of K = 5 for Spring, Autumn, and Winter, perfectly matching our all-year baseline. For Summer, the metrics indicated a primary local optimum at K = 4. However, to maintain cross-temporal consistency and allow direct physical comparison across all seasons without masking shallow lower-tropospheric processes, K = 5 was retained for all seasonal clusterings. This analysis confirms that the five identified EWRs represent robust, recurring atmospheric modes rather than statistical artifacts of seasonal aggregation. To integrate the seasonal analysis, we updated the Methods section to describe the seasonal partitioning methodology, the independent validation framework (Elbow + Silhouette) for each season, and the rationale for retaining K = 5 for cross-seasonal comparability. In the Results, we added a dedicated paragraph detailing the seasonal dynamics, vertical structures, and spatial stability of both coupled and decoupled EWRs across the four seasons. Finally, in Appendix A we included four new seasonal composite figures (Figs. A3-A6) depicting $Z_{500}$ and $\Delta T_{850}$ for Spring (MAM), Summer (JJA), Winter (DJF), and Autumn (SON).

LL. 300-307: *"Moreover, to evaluate potential seasonal dependencies in the vertical structure and physical mechanisms of EWEs, a seasonal analysis was also performed. Identified EWE days were partitioned into the four standard meteorological seasons: Winter (DJF), Spring (MAM), Summer (JJA), and Autumn (SON). The dual-metric validation framework (Elbow Method and Silhouette Analysis) was applied independently to each seasonal subset. The optimal cluster count yielded K=5 for Spring, Autumn, and Winter, whereas Summer exhibited an optimal metric configuration at K=4. To ensure methodological consistency across all temporal domains, facilitate direct comparison between seasonal regimes and the all-year baseline, and avoid masking distinct lower-tropospheric processes, K=5 was retained for all seasonal clustering analyses."*

LL. 542-561: *"To further assess whether the driving mechanisms and spatial characteristics of EWRs depend on the seasonal cycle, independent seasonal composite maps were constructed for the four meteorological seasons (Figs. A3-A6). This seasonal analysis reveals that the core synoptic configurations and vertical decoupling mechanisms identified in the annual baseline remain physically robust, confirming that the identified EWRs represent recurring atmospheric patterns rather than artifacts of seasonal aggregation. In particular:*

- *Coupled Regimes (Extreme Deep and Standard Deep EWRs): across all seasons, both coupled regimes are consistently governed by high-amplitude upper-level ridges driving strong lower-tropospheric warm advection ($\Delta T_{850} > 4\,°C$) and subsidence heating. During winter (Figs. A6a and A6b) and spring (Figs. A3a and A3b), these regimes manifest as prominent blocking patterns or high-amplitude subtropical ridges steering tropical air masses towards Central Europe. In summer (Figs. A4a and A4b), geopotential gradients are broader and less intense, but positive thermal anomalies remain strong and geographically focused over the Central Mediterranean.*
- *Decoupled Regimes (Mid Troposphere, Low Troposphere, and Upper Troposphere): the intensity and characteristics of vertical decoupling exhibit distinct seasonal signatures. During summer and autumn, the Mid Troposphere EWR (Figs. A4c and A5c) shows marked thermal decoupling, where elevated warm anomalies coincide with near-neutral or slightly negative lower-tropospheric anomalies over the Mediterranean basin, linked to transient troughs and baroclinic activity. Conversely, the Low Troposphere EWR (Figs. A4d and A5d) consistently retains strong surface-bound warming ($\Delta T_{850}$ > 2 °C) under relatively zonal or weak upper-level flow, emphasizing boundary-layer thermodynamic processes irrespective of season. Finally, the Upper Troposphere regime (Figs. A3a-A6e) is persistently associated with high-tropospheric ridge axes and tropopause displacements, reaching its strongest spatial coherence during summer and autumn (Figs. A4e and A5e).”*

**Q3:** Physical interpretation. Some of the dynamical interpretations currently feel more suggestive than demonstrated. For example, it is not clear to me how the authors distinguish a ridge from a blocking event in the composites. Similarly, (\Delta T_{850}) alone does not identify warm-air advection, and warm-air advection can also occur aloft. I think the paper would benefit from a more careful explanation of these interpretations, and possibly from additional diagnostics or anomaly fields.

More generally, I would like to see more discussion of the variability among events within each regime. The proposed mechanisms are plausible, but it would be useful to know how robust they are across events and seasons, rather than primarily relying on a single selected case study.

**A3:** We thank the Reviewer for these insightful remarks on the dynamical interpretations. In the revised manuscript, we have clarified the criteria used to distinguish amplified atmospheric ridges from blocking events in the synoptic composites. An amplified ridge is typically characterized by an open, northward-extending ridge of geopotential height contours without localized gradient reversals or closed cyclonic/anticyclonic pairs, while atmospheric blocking is identified by closed Z500 contours and a reversal of the meridional geopotential height gradient, characteristic of Omega-type blocking structures. In the revised version of the paper, we introduce these terms in Section 2.5, ensuring that composite patterns are described using precise synoptic terminology.

LL. 314-320: *“Specifically, mid-tropospheric ridges are defined as open, high-amplitude poleward extensions of the $Z_{500}$ geopotential height contours accompanied by a coherent thermal ridge, maintaining a continuous eastward phase speed without closed isolines or meridional flow reversals over the Central Mediterranean. Atmospheric blockings are characterized by persistent, quasi stationary vortex (closed $Z_{500}$ contours) that induce a local reversal of the meridional geopotential height gradient, forming classic Omega patterns that obstruct the westerly flow. This distinction ensures that composite circulation features are categorized accurately according to their kinematic and dynamical properties.”*

We agree that $\Delta T_{850}$ alone reflects a thermal anomaly rather than direct advection, and that warm-air advection can occur at higher atmospheric levels. To clarify this aspect, we have revised the terms used in Section 3 to avoid equating positive $\Delta T_{850}$ directly with horizontal advection.

LL. 323-329: *"Lower-tropospheric $\Delta T_{850}$ must be interpreted as the net thermodynamic response to both horizontal thermal advection and vertical air mass alteration (e.g., adiabatic subsidence heating beneath high-pressure systems). An elevated $\Delta T_{850}$ alone does not uniquely isolate horizontal warm-air advection from North Africa, as warm anomalies at 850 hPa can equally result from large-scale adiabatic compression (subsidence heating) beneath synoptic anticyclones, or from strong surface-driven convective entrainment. In regimes where $\Delta T_{850}$ anomalies coincide with high $Z_{500}$ but weak horizontal advection, adiabatic subsidence under the ridge axis and local solar radiative heating act as the primary drivers of column warming."*

Finally, regarding the request for a more detailed discussion of the variability among events within each regime, we have chosen not to include additional maps of spatial spread to avoid overcomplicating the figure layout, which we already expanded with seasonal diagnostics. Instead, we have addressed this issue by further emphasizing the statistical robustness of our clustering framework in the text. The optimal number of regimes (K=5) was determined through the application of the Elbow Method and Silhouette Analysis (Sect. 2.4). By definition, this optimization minimizes intra-cluster variance, ensuring that all individual days assigned to a specific EWR share a high degree of structural similarity. This statistical stability guarantees that the resulting composite maps provide a robust representation of recurring physical mechanisms, rather than averages of disparate or outlier events. The consistency of these patterns across seasons, now detailed in Figs. A3-A6, further confirms that these five configurations identify recurring atmospheric modes that physically exist year-round.

LL. 25-26: *"Spatial composite analysis and seasonal analysis confirm that these five configurations reflect physically robust, structurally stable, and recurring atmospheric modes across all seasons rather than artifacts of seasonal aggregation."*

LL. 845-846: *"Furthermore, seasonal analysis demonstrates that the identified EWRs represent persistent, structurally coherent atmospheric modes rather than artifacts of seasonal aggregation or outlier-driven composites."*

**Q4:** Trends and recent changes. I was not fully convinced by the interpretation of the recent increase in coupled regimes. With a 35-year record, how can the authors distinguish a long-term trend from decadal variability? It would be helpful to quantify uncertainty in the changes in regime frequency and to discuss this limitation more explicitly.

Figure 6 also does not seem to show a clear systematic increase in CHI for the deep regimes, despite the text suggesting that it does. Similarly, the conclusion that blocking or synoptic control has become more important would need additional support. Do blocking events themselves become more frequent or persistent, or is this inferred only indirectly from the temperature structure?

**A4:** We thank the Reviewer for this observation about the interpretation of Figure 6. We acknowledge that the text description of Figure 6 over-emphasized a systematic increase in CHI for deep regimes that is not uniformly present across the entire 35-year record. We have revised Section 3.4 accordingly.

LL. 591-596: *"The CHI time series reveals pronounced multi-decadal variability. While the cumulative annual number of days belonging to deep, vertically coupled regimes (e.g., Extreme Deep) has increased post-2000, the intrinsic intensity (CHI) of individual events within these regimes exhibits strong decadal fluctuations rather than a uniform upward trend. This non-monotonic behaviour underscores that recent shifts in integrated thermal load reflect the combined effect of warming trends and internal atmospheric variability, such as low-frequency synoptic circulation patterns and decadal blocking persistence, rather than a simple linear acceleration in event severity."*

Working with a 35-year record like ERA5 naturally makes it difficult to disentangle a long-term anthropogenic signal from natural atmospheric fluctuations. In the revised manuscript, we have explicitly addressed this limitation in Section 4, clarifying that CHI displays strong decadal fluctuations rather than a linear trend.

Finally, we agree that inferring changes in large-scale atmospheric blocking solely from local thermodynamic profiles is an overstatement. We have revised our discussion and conclusions accordingly: rather than claiming that blocking events have become more frequent or persistent, we now state that when persistent ridges occur, lower-tropospheric warming is more frequently coupled throughout the vertical column.

LL. 807-813: *"Moreover, a key limitation of utilizing a 35-year reanalysis dataset is the difficulty in isolating anthropogenic climate trends from natural, low-frequency atmospheric fluctuations, such as shifts in the Atlantic Multidecadal Variability (AMV) or the North Atlantic Oscillation (NAO). Consequently, the increase in vertically coupled regimes observed after 2000 should be interpreted with caution. Moreover, a higher occurrence of coupled thermodynamic profiles does not directly translate to an increase in the frequency or duration of atmospheric blocking events. Verifying whether blocking structures are becoming inherently more persistent would require dedicated kinematic blocking detection algorithms across the broader Euro-Atlantic domain, which falls outside the scope of local vertical temperature analysis."*

LL. 854-855: *"Rather than reflecting an intrinsic change in atmospheric blocking behaviour, these findings point to a higher probability that persistent ridge patterns become vertically coupled across the column."*

**SPECIFIC COMMENTS**

**Q5:** Abstract, lines 8–10: Why are only three levels used to describe the vertical structure? This choice should be justified already in the abstract or more clearly in the methods.

A5: We have updated the Abstract to briefly explain the physical and structural rationale behind choosing 1000 hPa, 500 hPa, and 300 hPa. Furthermore, as mentioned in our response to Q2, we have expanded Section 2.2 in the revised manuscript to provide a detailed physical and methodological justification for selecting these three specific pressure levels over the full 15-level profile. The Abstract has been modified as follows:

LL. 11-14: *"This study investigates the three-dimensional patterns of EWEs over Rome during the 1991-2025 period using ERA5 temperature data across three strategic pressure levels representing distinct functional layers: 1000 hPa (near-surface), 500 hPa (middle troposphere), and 300 hPa (high troposphere)."*

**Q6:** Introduction, lines 54–60: The statement that most studies treat heatwaves as exclusively surface phenomena seems too broad. A number of recent studies examine the vertical structure of heat extremes; these should be discussed more fully.

**A6:** As suggested by the Reviewer, we have expanded the Introduction to explicitly recognize the growing body of literature focused on the three-dimensional and vertical structure of atmospheric thermal extremes (e.g., Miralles et al., 2014; Sousa et al., 2018; Lhotka and Kyselý, 2024; Hotz et al., 2026). We now clarify that conventional HW metrics (surface-focused) primarily capture surface impacts, whereas a multi-level framework isolates the vertical coupling or decoupling mechanisms between the near-surface atmospheric layer and the free troposphere.

LL. 76-80: *"Recently, a growing body of literature has emphasized that heat extremes are inherently three-dimensional atmospheric features driven by complex vertical structures and circulation patterns (e.g., Miralles et al., 2014; Sousa et al., 2018; Hotz et al., 2026). Hotz et al. (2026) also provided a comprehensive global baseline of vertical temperature anomaly profiles during land HWs, showing that extreme thermal anomalies exhibit substantial vertical variability and are often decoupled between the PBL and the free troposphere."*

**Q7:** Lines 64–79: The distinction among heatwaves, EWEs, and EWRs is useful, but I still found the motivation for high- and mid-tropospheric EWEs unclear. Could the authors explain more directly why these events should be considered alongside surface heat extremes?

**A7:** The core objective of our framework is not to equate an upper-tropospheric warm anomaly with a surface heatwave, but to diagnose the vertical structure of thermodynamic forcing. High- and mid-tropospheric EWEs represent cases where intense synoptic forcing (e.g., ridge amplification or adiabatic compression) is present aloft, but surface-layer processes (such as marine boundary layer stratification, cold-air advection, or sea-breeze buffering) prevent this warming from penetrating the boundary layer. Treating high-tropospheric events alongside surface-impacting ones provides a crucial diagnostic benchmark: it allows us to identify the specific atmospheric conditions required for an upper-level anomaly to vertically couple down to the surface, versus cases where local PBL mechanisms successfully buffer the near surface atmosphere. This aspect is now clarified in the Introduction and in the Methodology as follows:

LL. 80-89: "*While upper-tropospheric thermal extremes do not directly impact terrestrial ecosystems or human health at the surface, their diagnostic value lies in their role as dynamical indicators of large-scale atmospheric patterns (Pfahl et al., 2015; Zschenderlein et al., 2019). In fact, analysing high-level events that do not reach the surface allows us to isolate the near surface and lower-tropospheric mechanisms (such as cold-air advection, marine boundary layer stratification, or sea-breeze regime) that decouple the near surface atmospheric layer from intense mid-tropospheric ridging. In addition, high-tropospheric anomalies frequently act as dynamic precursors to synoptic-scale blockings, offering a diagnostic window into the vertical evolution and downward propagation of thermodynamic anomalies prior to full-column coupling (Sousa et al., 2018). Analysing the entire vertical column allows for a comprehensive classification of warm events, distinguishing deep, surface-impacting HWs from dynamically isolated upper-tropospheric thermal features (Lhotka and Kyselý, 2024; Hassoun et al., 2025)."*

LL. 226-228: *"In doing so, upper-level EWEs are treated not as 'failed' surface HWs, but as distinct thermodynamic regimes governed by upper-tropospheric wave dynamics and adiabatic compression, providing a clear benchmark for when and why synoptic forcing successfully penetrates the boundary layer."*

**Q8:** Lines 108–110: Which season is being discussed here? The role of the sea breeze and urban heat island is likely to be quite season dependent.

**A8:** We have updated the text to specify that while the interplay between sea-breeze buffering and the UHI effect reaches its maximum intensity and thermal impact during summer, driven by peak solar irradiance and strong land-sea thermal contrasts, both phenomena are well-documented over Rome throughout the entire year (e.g., Mastrantonio et al., 1994, https://doi.org/10.1007/BF00709220; Ferretti et al., 2003, https://doi.org/10.1029/2003JD003424; Ciardini et al., 2019, https://doi.org/10.1016/j.uclim.2019.100493). In fact, in winter and transition seasons, although the sea breeze is weaker and less frequent due to reduced thermal gradients, mesoscale coastal flows and urban thermal signatures remain active and continue to modulate low-level temperatures.

LL. 148-160: *"The vertical structure of the atmosphere over Rome is further modulated by the interaction between the urban footprint and the Tyrrhenian sea-breeze circulation. During summer near-surface HWs, high solar irradiance and strong land-sea thermal contrasts drive a robust daytime sea-breeze front that penetrates inland, establishing a stable thermal stratification that can temporarily decouple the near-surface layer from the free troposphere (Di Bernardino et al., 2021; Di Bernardino et al., 2023). However, under intense synoptic subsidence and weak background ventilation, this daytime coastal buffering is often overwhelmed, while nocturnal cooling is strongly suppressed by the summer UHI (Cecilia et al., 2023; Di Bernardino et al., 2023). In transition and winter seasons, although land-sea thermal gradients are weaker and coastal breezes are less vigorous, local urban microclimates and mesoscale flows remain active, though lower-tropospheric thermal anomalies are governed more directly by synoptic-scale advection."*

**Q9:** Lines 125–130: The paper uses the full 15-level profile for case-study cross-sections, but only three levels for identification and clustering. Why not use the full column for the classification itself?

**A9:** As addressed in our response to Q2, the decision to restrict the classification and clustering algorithm to three primary pressure levels (1000, 500, and 300 hPa) rather than the full 15-level profile is rooted in both statistical and physical considerations. In fact, including all 15 pressure levels introduces high vertical autocorrelation and redundant information (as adjacent vertical levels in ERA5 are strongly correlated), possibly resulting in smoothing of cluster boundaries and increased overfitting noise rather than isolating distinct dynamical regimes. As detailed in the text, the 3 selected levels serve as physically-based references: 1000 hPa represents the near surface atmospheric layer; 500 hPa represents the mid-tropospheric steering level and synoptic ridge core; and 300 hPa captures upper-tropospheric dynamics. Additionally, although the EWEs were determined from only 3 levels, the three-dimensional structure and vertical coupling of the events are analysed in detail by considering all 15 levels in the case studies and regime cross-sections. This confirms that these 3 levels effectively act as key proxies and accurately capture the thermodynamic behaviour of the intervening layers. We have expanded Section 2 to clarify this aspect as follows:

LL. 176-179: *"This selection is based on physical and statistical considerations. In fact, adjacent pressure levels in high-resolution reanalyses exhibit strong vertical collinearity. Including all 15 levels in the clustering algorithm's distance matrix may introduce redundant variance, which tends to degrade cluster separation and excessively flatten regime boundaries. In addition, the three selected levels strategically sample three distinct functional layers of the atmosphere."*

**Q10:** Lines 132–140: Please clarify the use of 1000 hPa over this region, given topography and the fact that it is not always equivalent to a near-surface or PBL level. Also, the phrase "advancing subtropical air masses" is not fully clear to me. Is this meant to refer specifically to lower-tropospheric air masses? Warm-air advection can occur aloft as well.

**A10:** We thank the Reviewer for requesting these two important clarifications. For what concern the selection of the 1000 hPa level, the study domain focuses on the coastal and low-lying urban area of Rome, where mean surface elevation ranges from sea level to less than 100 m a.s.l. In this topographic condition, the 1000 hPa surface represents a real, non-extrapolated atmospheric layer residing directly within the PBL. It is not subject to sub-surface extrapolation artifacts that affect mountainous or elevated topography. This aspect has been clarified in the revised manuscript as follows:

LL. 181-183: *"Because the target domain over the metropolitan area of Rome lies in low-elevation coastal terrain (surface elevation < 100 m a.s.l.), the 1000 hPa level consistently resides above the* *physical surface, avoiding artificial sub-surface reanalysis extrapolations common over elevated topography."*

For what concerns the phrase "advancing subtropical air masses", we agree that the sentence imprecise and could cause confusion. We have revised the text in Section 2 to specify that we are referring to lower-to-mid tropospheric warm-air advection (specifically in the 850-700 hPa layer) associated with the northward expansion of North African anticyclones, as follows:

LL. 183-187: *"The 850 hPa level serves as a standard diagnostic for air-mass characterization. Being located above the PBL, it is largely decoupled from local surface influences, including UHI effects. It can therefore be used to assess whether lower-tropospheric thermal extremes are driven by regional warm-air advection (associated with the poleward transport of subtropical air masses within the 850-700 hPa layer) or by local processes such as large-scale adiabatic compression."*

**Q11:** Lines 150,164: A brief plain-language description of the Mann-Kendall test and Sen's slope would help the reader understand what each diagnostic adds.

**A11:** We thank the Reviewer for this suggestion. We have added a concise, plain-language explanation in Section 2.3 to clarify the distinct roles of the Mann-Kendall test and Sen's slope estimator.

LL. 212-218: *"Therefore, the original MK test was applied (rather than its seasonal version) as the prior removal of the monthly mean ensures that the test detects whether a statistically significant, monotonic upward or downward trend exists in the residuals, without making assumptions about data normality. Complementarily, the magnitude of the temporal trend was quantified using Sen's Slope estimator (Hirsch and Slack, 1984), providing the rate of change per decade. By computing the median of all pairwise slopes, Sen's slope yields a robust linear rate of change that is highly resilient to single-year extreme outliers."*

**Q12:** Lines 186: Please explain the motivation for the four-day persistence criterion. Many heatwave definitions use three consecutive days; why was four days selected here?

**A12:** Adopting a four-consecutive-day persistence threshold, rather than the traditional three-day criterion, is motivated by both synoptic dynamics and statistical classification constraints. Fast-moving, high-frequency atmospheric waves and transient ridges over the Central Mediterranean typically exhibit lifetimes of 2-3 days, so selecting a four-day threshold effectively filters out these short-lived thermal fluctuations and isolates quasi-stationary synoptic patterns. Furthermore, in a coastal-urban environment like Rome, the full vertical coupling of thermodynamic anomalies requires sustained synoptic forcing across multiple diurnal cycles. From a clustering perspective, requiring four consecutive days reduces high-frequency noise and transient boundary events in the input time series, resulting in a clearer separation between clusters and distinct EWRs. We have expanded Section 2.2 in the revised manuscript to point out this rationale.

LL. 252-258: *"While many HW indices employ a three-day threshold, adopting here a four-day persistence requirement serves specific synoptic and methodological purposes. Synoptically, short-lived transient ridges and fast-propagating Rossby waves over the Central Mediterranean typically persist for two to three days, meaning that a four-day threshold ensures that identified events reflect quasi-stationary, high-amplitude synoptic structures rather than transient thermal fluctuations. Methodologically, requiring sustained four-day persistence allows sufficient time for multi-day thermodynamic feedbacks to fully manifest throughout the vertical column, while simultaneously improving signal-to-noise separation in the subsequent clustering analysis."*

**Q13:** Table 1: I would be cautious about interpreting “Surface Only” events as “typical surface heatwaves.” Many conventional heatwaves do have a clear free-tropospheric and circulation signature. Likewise, the physical interpretations assigned to the Mid Troposphere and High Troposphere categories seem somewhat speculative and would benefit from further explanation.

**A13:** We fully agree that labelling "Surface Only" events as "typical surface heatwaves" was misleading. To avoid over-speculative physical labelling and maintain scientific rigor, we have thoroughly revised the text in Section 2.2 to describe these events as shallow thermal anomalies governed primarily by localized surface energy fluxes, planetary boundary layer processes, or shallow low-level advection in the absence of strong free-tropospheric forcing. Similarly, mid- and high-tropospheric categories are now framed strictly as upper-level thermodynamic indicators reflecting passing synoptic ridges or upper-tropospheric wave dynamics that remain decoupled from the boundary layer. Furthermore, in response to comments from another reviewer, we have entirely removed the "Meteorological Interpretation" column from Table 1 to eliminate overly prescriptive assumptions and keep the classification focused purely on structural and vertical anomaly properties.

LL. 265-269: *“The functional characterizations assigned to each class reflect the vertical anomaly structure of the event rather than prescriptive physical outcomes. Events confined to lower levels represent shallow warm anomalies governed primarily by surface energy fluxes, PBL dynamics, or shallow advection under weak synoptic forcing, rather than conventional HWs. Conversely, mid- and high-tropospheric EWEs serve as upper-level thermodynamic indicators of passing ridges or wave dynamics aloft that remain decoupled from the PBL.”*

**Q14:** Lines 215–225: I did not fully follow the interpretation of (\Delta T_{850}). Why does a lower (\Delta T_{850}) in the presence of high (Z_{500}) imply that local vertical dynamics are more important? Please explain this reasoning more carefully.

**A14:** As addressed in detail in our response to Q3, high mid-tropospheric geopotential height anomalies represent the core of amplified synoptic ridges, which directly drive lower-tropospheric thermal anomalies through large-scale adiabatic subsidence. The high geopotential heights force sustained downward motion in the mid-to-lower troposphere, where the descending air undergoes compression heating, thereby elevating lower-tropospheric temperatures and producing strong positive $\Delta T_{850}$ anomalies. Furthermore, this lower-tropospheric warming stabilizes the upper boundary layer while clear-sky conditions maximize surface radiative heating, facilitating progressive heat accumulation within the planetary boundary layer. Thus, $\Delta T_{850}$ during high $Z_{500}$ episodes acts as a direct thermodynamic fingerprint of synoptic-scale adiabatic compression, establishing a clear physical bridge between upper-level ridge strength and lower-tropospheric heat buildup.

**Q15:** Section 3.2 / Figs. 2–3: The titles should be clearly distinguished from those in the methods, since this section presents results rather than the identification procedure. The captions for Figs. 2 and 3 are also too brief. Please define the colours and categories clearly and provide enough information for the figures to stand alone. I found it difficult to see directly from Fig. 3 the claimed seasonal expansion and transition from high-level precursor stages to full-column events.

**A15:** As suggested, we have updated the titles of Sections 3.2 and 3.3 to clearly differentiate the presentation of observational findings from the methodological framework introduced in Section 2. Additionally, we have expanded the figure captions for Figures 2 and 3 so that they stand completely independently. Finally, we have revised Section 3.2 to guide the reader through Figure 3, describing how high-tropospheric anomalies in early summer can act as upper-level precursors before transitioning into deep, full-column events during peak summer.

Title, Sect. 3.2: *"Temporal frequency and multidecadal evolution of Extreme Warm Events (EWEs)"*

Title, Sect 3.3: *"Vertical structure and multi-level classification of Extreme Warm Regimes (EWRs)"*

LL. 440-444, caption, Fig. 2: *"Annual frequency and vertical category distribution of EWE days over Rome from 1991 to 2025. Stacked bars represent the total annual count of EWE days, color-coded by the five categories defined in Table 1: Surface Only (yellow), Mid Troposphere (green), High Troposphere (cyan), Full Column (purple), and Transition (red)."*

LL. 475-178, caption, Fig. 3: *"Multi-decadal distribution and occurrence of EWE days over Rome from 1991 to 2025. The x-axis shows the calendar timeline across all 12 months (Jan–Dec) versus years on the vertical axis. Coloured segments denote individual EWE days, color-coded by the five categories defined in Table 1: Surface Only (yellow), Mid Troposphere (green), High Troposphere (cyan), Full Column (purple), and Transition (red)."*

LL. 461-472: *"The temporal distribution of individual EWE days from 1991 to 2025 throughout the year (Fig. 3) illustrates both the intra-annual seasonality and a pronounced multi-decadal intensification of warm extremes throughout the tropospheric column. Seasonally, EWEs are distributed across the entire year, with high-tropospheric anomalies (High Troposphere category) frequently developing in spring and early summer as elevated thermodynamic precursors associated with upper-level ridge building. During the core summer months of July and August, these upper-level features increasingly transition into and co-occur with persistent Full Column and Transition events, representing vertically coupled warm structures sustained by large-scale subsidence and surface radiative heating. Multi-decadally, Fig. 3 highlights an increase in event density, duration, and structural complexity during the most recent decade (particularly from 2015 to 2025). This recent period is characterized not only by a higher occurrence of peak-summer Full Column and Transition events, but also by a clear seasonal expansion, with Surface Only and High Troposphere EWEs occurring with unprecedented frequency during May, September, and October, and extending into the winter months."*

**Q16:** Fig. 4: Would it be more appropriate to show standardized anomalies, or at least discuss how differences in temperature variance with height affect the profiles and clustering? It would also be useful to discuss the fairly large spread within some regimes.

**A16:** We thank the Reviewer for this comment. We decided to present physical temperature anomalies because they directly quantify the actual thermal magnitude and thermodynamic energy content throughout the tropospheric column, which is essential for physically interpreting heating processes. Furthermore, the height-dependent differences in temperature variance are already accounted for during the initial identification stage, as EWEs are detected using level-specific 90th percentile thresholds calculated independently at each pressure level. Thus, the clustering input inherently reflects level-adapted extreme states, making absolute temperature profiles the most direct and physically meaningful metric for comparative analysis.

Regarding the spread shown in Figure 4, we agree that discussing this internal variability adds valuable context. The larger spread observed in certain regimes (most notably Upper Troposphere) stems from natural synoptic variability across seasons, where the precise vertical core and amplitude of ridge amplification or wave-breaking features fluctuate between individual events. Conversely, deep coupled regimes exhibit a more compact spread, reflecting strong, vertically coherent synoptic forcing. We have expanded the text in Section 3.3 to discuss these aspects as follows:

LL. 486-492: *"Expressing these profiles in terms of ΔT provides a direct measure of thermal energy content throughout the column, while altitude-dependent differences in temperature variance are implicitly controlled for during event identification through level-specific percentile thresholds. Furthermore, the spread depicted by the shaded regions (±1 standard deviation $\sigma_{Tp}$) reflects the natural intra-regime variability: while deep coupled regimes show strong structural consistency due to dominant synoptic subsidence, elevated and decoupled classes display wider dispersion owing to fluctuations in the exact core altitude and amplitude of upper-level features across seasons."*

**Q17: Fig. 5 and associated text:** Would geopotential-height anomalies be more informative than absolute (Z_{500})? As shown, it is difficult to assess whether the patterns represent anomalous ridging or blocking. The figure caption should also be expanded, and the basis for distinguishing a ridge from a block should be clarified.

**A17:** We thank the Reviewer for this constructive comment. We chose to present absolute geopotential height (Z500) rather than standardized anomalies because absolute contours provide a direct visual representation of the synoptic flow pattern, the spatial positioning of the ridge axis, and the presence of closed anticyclonic centers over the Mediterranean domain. To address the Reviewer's concern regarding the distinction between ridges and blocking structures, as detailed in the answer to Q3, we have clarified the criteria used to distinguish amplified atmospheric ridges from blocking events in the synoptic composites. An amplified ridge is typically characterized by an open, northward-extending meander of geopotential height contours without localized gradient reversal or closed cyclonic/anticyclonic pairs, while an atmospheric blocking is identified by closed Z500 contours and a reversal of the meridional geopotential height gradient, characteristic of Omega-type blocking structures.

**Q18:** Fig. 6 and lines 419: I do not clearly see the stated increase in CHI in the recent period. Please revisit this interpretation. It would also be useful to discuss uncertainty, variability, and whether the apparent changes could reflect decadal variability rather than a robust trend.

**A18:** We fully agree that attributing recent CHI variations solely to a clear upward trend would be oversimplifying the complex signals present in the data. As detailed answer to Q4, we have revised the text accompanying Figure 6 in Section 3.4 to offer a more nuanced and cautious interpretation.

LL. 591-596: *"The CHI time series reveals pronounced multi-decadal variability. While the cumulative annual number of days belonging to deep, vertically coupled regimes (e.g., Extreme Deep) has increased post-2000, the intrinsic intensity (CHI) of individual events within these regimes exhibits strong decadal fluctuations rather than a uniform upward trend. This non-monotonic behaviour underscores that recent shifts in integrated thermal load reflect the combined effect of background warming trends and internal atmospheric variability, such as low-frequency synoptic circulation patterns and decadal blocking persistence, rather than a simple linear acceleration in event severity."*

**Q19:** Fig. 7: The precursor analysis is interesting, but the physical explanation for the apparent cooling aloft in the Low Troposphere regime and at 500 hPa in the High Troposphere regime is not clear to me. Why do these layers cool as the event onset is approached? Showing the spread among events, rather than only the mean evolution, would also be helpful.

**A19:** We thank the Reviewer for these constructive comments. Regarding the apparent cooling aloft, the mechanisms driving these trajectory divergences prior to event onset stem from distinct synoptic and dynamic processes. The mid-to-upper tropospheric cooling (500 and 300 hPa) near $t_0$ reflects the eastward displacement of upper-level ridge support and the arrival of weaker baroclinic flow aloft. While the free atmosphere cools, extreme warming becomes isolated and amplified exclusively near the surface through boundary-layer thermodynamic processes, local sensible heat fluxes, and shallow warm advection. The prominent drop at 500 hPa paired with strong warming at 300 hPa represents the classic dynamical signature of Rossby wave-breaking and local tropopause displacement. Here, the extreme thermal anomaly remains rigidly confined to the upper troposphere, while an influx of cooler air in the mid-troposphere prevents any top-down coupling toward the surface. Importantly, as depicted in Figure 7, these layer-specific thermal fluctuations and cooling trends remain well within the inter-event variability range (standard deviation bounds), confirming that they represent subtle, physically consistent regime modulations rather than erratic outlier behaviour.

In line with your suggestion, we have updated Figure 7 by adding vertical error bars that display ±1 inter-event standard deviation across all time steps and levels. While individual episodes show natural variance in amplitude, the overarching precursory trajectories and layer-decoupling trends remain remarkably robust across the entire event population. Section 3.3 and the figure caption have been updated accordingly, as follows:

LL. 605-609: *"To investigate the dynamic mechanisms and vertical structures driving the onset of the different atmospheric configurations and quantify their inter-event spread, Fig. 7 depicts the temporal evolution of cluster-averaged temperature anomalies $\overline{\Delta T_p(t)}$ at 1000, 500, and 300 hPa from three days prior to the event ($t_{-3}$) up to the onset day ($t_0$), with vertical error bars representing ± 1 standard deviation across events within each EWR. This lagging framework highlights the distinct precursory pathways, vertical coupling modes, and intra-regime variability characterizing each extreme regime."*

LL. 631-634: *"This upper-level cooling reflects the eastward displacement of upper-level ridge support and the passage of weak baroclinic perturbations aloft, which effectively strip the free atmosphere of warm anomalies while trapping sensible heat and local advective fluxes within a shallow, PBL-confined layer near the surface."*

LL. 641-647: *"Dynamically, this contrasting vertical dip at 500 hPa paired with intense heating at 300 hPa is characteristic of upper-level wave-breaking and local tropopause folding. The extreme thermal anomaly remains rigidly confined to the upper troposphere, while an influx of cooler baroclinic air in the mid-troposphere prevents vertical propagation toward the surface. The vertical spread depicted by the error bars demonstrates that while individual events vary in absolute magnitude (≈ 1.5-2.5 °C), the mean temperature evolutions consistently lie well within ±1 standard deviation of the event population, confirming that the overarching precursory trajectories and layer-specific decoupling trends remain structurally consistent across individual episodes."*

**Q20:** Case studies: The case studies are useful illustrations, but it would help to explain more clearly why the selected cases are representative of each regime. For example, what is the variability among other events in the same regime? The language of "validation" may be too strong for a single case per cluster.

**A20:** We fully agree that a single case study cannot "validate" a statistical cluster and that the term "validation" is too strong. In response, we have revised Sections 2.7 and 3.4 to explicitly frame the case studies as observational exemplars designed to illustrate the physical evolution, thermodynamic pathways, and synoptic drivers of each regime under peak conditions, rather than as a statistical validation. Furthermore, we have clarified the exact criterion used for case selection in Section 2.7. Cases were objectively identified as the chronologically continuous events exhibiting the maximum CHI across the reference levels (1000, 500, and 300 hPa), prioritizing episodes where the dominant daily cluster ID matched the prevailing vertical category of the entire event (as detailed in Table 4). Selecting the highest-CHI continuous event guarantees that the key physical processes governing each regime are captured at their clearest, most persistent expression.

LL. 353-361: *"Rather than serving as a formal statistical validation of the clusters, these selected episodes act as observational exemplars designed to illustrate the detailed thermodynamic pathways and synoptic drivers characterizing each regime during extreme conditions. It is worth noticing that, since EWEs are transient phenomena, their vertical structure can dynamically evolve through different categories over time. To isolate the most representative and persistent fingerprint for each EWR, the chronologically continuous event exhibiting the maximum CHI value across the reference levels (1000, 500, and 300 hPa) was selected, prioritizing episodes where the dominant daily cluster ID matched the prevailing vertical category of the entire event. While individual events within the same EWR naturally exhibit variability in peak intensity, exact duration, and fine-scale trajectory pathways, selecting the maximum-CHI continuous episode ensures that the physical mechanisms governing the build-up and decay of the regime are captured at their peak expression."*

**Q21:** Fig. 8: Please indicate (t_0) directly in each panel. It would also help to explain the contouring and make clearer how the inferred downward or upward development is identified.

**A21:** The onset time $t_0$ corresponds to the origin of the time axis (x-axis) in each panel. We have explicitly stated this in the updated caption of Fig. 8. Moreover, we specified in the caption of Fig. 8 that the black solid contours in panels (a, c, e, g, i) highlight the 0 °C temperature anomaly boundary. Finally, we expanded the revised text to clarify how vertical progression is inferred.

LL. 737-743, caption, Fig. 8: *"Time-height cross-sections of the selected case studies. Left panels: temperature anomalies ($\Delta T$, °C) calculated relative to the 21-day local climatological window, with solid black contours denoting the 0.0 °C anomaly line. Right panels: equivalent potential temperature ($\theta_e, K$). From top to bottom, rows correspond to the identified EWRs: Extreme Deep, Standard Deep, Mid Troposphere, Low Troposphere, and High troposphere. Horizontal, black dashed line refers to the 850 hPa level shown in the composite maps in Fig. 9. Note that in each panel the onset time of the event ($t_0$) corresponds to the start of the time axis in each panel and that the duration displayed on the x-axis corresponds to the full lifecycle of each specific event, ranging from 4 days (panels a, b, c, and d) up to 10 days (panels e and f)."*

LL. 662-664: *"In Fig. 8, the vertical progression of the signals is identified through the temporal evolution and time-height tilt of these anomaly contours and $\theta_e$ isolines across pressure levels; specifically, a contour slope extending downward over time indicates downward development, whereas an opposite slope highlights upward propagation."*

**Q22:** Fig. 9: Please identify which case each row represents. The time direction in the trajectory panels is somewhat confusing relative to the spatial tracks, and I found it difficult to see the claimed differences among the trajectories. A clearer presentation would help.

**A22:** We have thoroughly revised the caption of Figure 9 to address all these points and provide a clear, unambiguous guide for the reader.

LL. 746-754, caption, Fig. 9: *"Spatial synoptic characterization and 144-hour HYSPLIT backward trajectories for the selected case studies, organized by EWRs from top to bottom: (a, b) Extreme Deep, (c, d) Standard Deep, (e, f) Mid Troposphere, (g, h) Low Troposphere, and (i, j) High Troposphere. Left panels: ERA5 composite maps of geopotential height at 500 hPa ($Z_{500}$, gpm, black solid lines)* *and temperature anomaly at 850 hPa ($T_{850}$, °C, color shading) averaged over the respective EWE duration. Right panels: 144-hour HYSPLIT backward trajectories arriving over Rome at 12:00 UTC on the day of maximum thermal anomaly. Spatial tracks illustrate air mass movement from remote origins toward Rome, with arrival altitudes set at 500 m (red lines), 1500 m (blue lines), and 5500 m (green lines) above ground level (agl). Bottom sub-panels display the corresponding vertical displacement profiles backward in time from the arrival date ($t_0$) on the left back to $t_{-144}$ on the right."*

**Q23:** Discussion and conclusions: The conclusions about increasingly synoptically controlled events, increased blocking, and reduced sea-breeze effectiveness seem stronger than what is directly shown. Could the authors either provide additional evidence for these interpretations or state them more cautiously?

**A23:** We agree that these interpretations, while supported by our synoptic composite analyses and trajectory results, should be framed more cautiously to avoid overstatement. Therefore, as detailed in the answer to Q4, we have revised the Conclusions section to adopt a more conservative approach.

LL. 847-855: *"The long-term analysis reveals a marked increase in the frequency of EWE over recent decades, with a significant acceleration observed after 2015. This trend is primarily driven by a higher occurrence of vertically coupled events, specifically the Full Column category corresponding to the Extreme Deep and Standard Deep EWRs, characterized by significant warming extending to a large portion of the tropospheric column. Conversely, the average duration and mean thermal intensity of individual events have remained relatively stable over the 35-year period. This implies that the recent growth in regional thermal stress is primarily driven by a higher recurrence frequency of deep-column anomalies rather than by the systematic prolongation or amplification of single episodes. Rather than reflecting an intrinsic change in atmospheric blocking behaviour, these findings point to a higher probability that persistent ridge patterns become vertically coupled across the column."*

Overall, I think the manuscript would be strengthened by a clearer motivation for the high-level events, a more fully justified treatment of vertical structure and seasonality, and a more cautious interpretation of the dynamical mechanisms and recent trends.