

Response to Reviewer #2's Comments

This manuscript presents a five-year (2021–2025), station-based objective framework that separates warm-sector heavy rainfall (WSHR) from other heavy rainfall over the eastern China plains, using rain-gauge, ERA5, and radar wind profiler (RWP) observations. The identification scheme (automated front detection, warm-sector mask, station-specific 99th-percentile threshold) is methodologically sound and well documented, and the spatiotemporal characterization of WSHR (Section 3.1–3.2) is clear and well supported by the data. The RWP-based precursor analysis (Section 3.4) is, in my view, the most valuable contribution of the paper, as it extends the pre-convective picture beyond what ERA5 or radiosondes can resolve. The manuscript is well organized, the figures are clear, and the writing is precise. I recommend minor revision, focusing on the points below.

Reply: Thank you for your careful reading and positive comments, and we also appreciate it tremendously for recognizing the value of the high-frequency RWP analysis in our manuscript. In response to your suggestions, we clarify the complementary roles and limitations of ERA5 and RWP data in this revised manuscript. Also, a comparison of warm-sector non-heavy rainfall is performed against warm-sector heavy rainfall, with the dependence among station-based events addressed. The specific changes are documented below. For clarity, reviewer comments are listed below in italics, our responses in plain blue font, and the corresponding changes to the manuscript in blue italics.

Major Comments

(1) Vertical thermodynamic structure vs. RWP kinematic structure (Section 3.3–3.4, 4.1)

*The manuscript characterizes the pre-onset environment using two complementary but separate sources: ERA5-based thermodynamic composites (specific humidity, θ_e , VIMFD; Figs. 7–8) and RWP-based kinematic composites (wind, KE, shear, ascent; Figs. 9–12). The authors correctly note (L538–541, Sect. 4.1) that the RWP's 6-min resolution captures pre-convective flow evolution that hourly reanalyses and radiosondes cannot resolve—this justification is reasonable and **I do not ask the authors to substitute RWP with radiosonde data.***

However, because ERA5 has only hourly resolution, the thermodynamic evolution shown in Figs. 7–8 is comparatively coarse relative to the RWP kinematic evolution, which is resolved at 6-min steps down to the final 30 min before onset. This creates an asymmetry: the paper can show precisely when, kinematically, the pre-convective environment intensifies, but not precisely when the thermodynamic destabilization (e.g.,

a rapid increase in low-level moisture or a sharp decrease in stability) occurs relative to that kinematic evolution.

*→ I would suggest **briefly discussing this resolution asymmetry as a limitation** (in addition to the limitations already listed in Sect. 4.1).*

Reply: Thank you for pointing out the difference in temporal resolution between the two datasets. Sect. 4.1 now states that hourly ERA5 fields cannot resolve moisture and stability changes on the 6-min RWP timescale, particularly during the final 30 min before local rainfall onset. Their timing relative to the observed kinematic intensification therefore remains unresolved.

This revised manuscript uses the RWP observations to characterize rapid wind evolution and ERA5 to describe the broader moisture and thermodynamic background. Sect. 2.4 now also makes the sampling distinction explicit: the ERA5 composites use all 477 station-based WSHR events, whereas the RWP composites use the subset with available profiler observations. The two composites are consequently complementary rather than a fully matched thermodynamic–kinematic composite.

In addition, Sect. 2.4 clarifies that the reference time is local rainfall onset at the co-located gauge, which does not necessarily coincide with convective initiation. These clarifications delimit the temporal interpretation.

The specific changes made in this revised manuscript are listed below:

In the 1st paragraph of Sect. 2.4: *“Here, onset refers to the first detected rainfall at the co-located gauge and does not necessarily coincide with convective initiation, because convection may develop upstream before reaching the station.”*

In the 2nd paragraph of Sect. 2.4: *“The ERA5 composites use the full sample of 477 station-based WSHR events, whereas the RWP composites use the subset with available profiler observations. The two datasets therefore characterize complementary aspects of the pre-onset environment and do not provide a fully matched thermodynamic–kinematic composite.”*

In the 3rd paragraph of Sect. 4.1: *“The hourly ERA5 fields cannot resolve thermodynamic changes on the same timescale as the 6-min RWP observations, particularly during the final 30 min before rainfall onset. The timing of moisture and stability changes relative to the observed kinematic intensification therefore remains unresolved.”*

(2) Specificity and predictive skill of the precursor signals (Section 3.4, 4.1, 4.3)

The authors appropriately flag, in Section 4.1 (L548–551), that the current analysis has no non-heavy or non-warm-sector baseline, so the specificity of the RWP precursor signals to WSHR (rather than to warm-season rainfall in general, or to non-precipitating warm-sector conditions) remains untested. This is an important and honest limitation, but I believe it is central enough to the manuscript's main claim —

that these signals "may serve as candidate reference indicators" for imminent WSHR (L514–516, L651–653) — that it merits at least a partial quantitative treatment rather than being deferred entirely to future work.

Specifically, I would encourage the authors to: compute the same RWP diagnostics (LLJ occurrence frequency, low-level KE, shear, ascent frequency) for a comparison sample—e.g., warm-sector non-heavy rainfall events, or non-raining warm-sector periods at the same RWP stations — over an equivalent pre-onset window.

Reply: We agree that a quantitative comparison is needed to assess the specificity of the signals. To provide a warm-sector baseline, we added 271 WS non-HR events from the same 37 co-located stations that contributed the 62 WSHR events. Both samples use the same RWP quality-control criteria, onset alignment and 180-min analysis window. The shared station set, sample definition and processing are now stated explicitly in Sect. 2.4.

This revised manuscript now provides comparison composites for all four diagnostics identified by the referee: LLJ occurrence frequency, low-level KE, vertical wind shear and ascending-motion frequency. The revised Fig. 11 compares maximum low-level wind speed, event-mean jet-core characteristics and LLJ occurrence frequency. The new Fig. S6 presents the WS non-HR KE composite for comparison with the WSHR composite in Fig. 9c. Fig. 12 compares ascending-motion frequency and layer-maximum ascent-speed distributions. WS non-HR shear composites are provided in Fig. S7, while hodographs for WSHR and WS non-HR are shown in Figs. S8 and S9, respectively. The quantitative LLJ metrics are summarized below.

The comparison shows shared features as well as differences. Low-level shear, clockwise turning and ascent occur in both categories. KE generally increases with height in both composites, but WSHR shows a more pronounced enhancement at approximately 1–3 km AGL during the final hour before local rainfall onset. The late increase in LLJ occurrence and the upper tail of the ascent-speed distribution near 1–2 km are also more pronounced in WSHR, although the LLJ core characteristics and ascent-speed distributions overlap substantially between the samples. Moreover, the 95% episode-cluster bootstrap interval for the WSHR early-to-late LLJ increase includes zero. We therefore report differences in composite evolution without claiming WSHR-specific signals or validated event-level predictive skill.

The Abstract, Sect. 3.4, the limitations in Sect. 4.1 and the Conclusions have been revised accordingly. Sect. 4.3 now calls for event-level validation of specificity, useful lead time and false-alarm characteristics. It distinguishes the observed precursor information supported by Guo et al. (2023) from the potential assimilation benefits investigated by Zhao et al. (2025), without claiming that establishes forecast skill for the WSHR sample analyzed here.

The specific changes made accordingly in the revised manuscript are listed below:

In the 1st paragraph of Sect. 2.4: “A comparison sample consisted of 271 warm-sector non-heavy rainfall (WS non-HR) events from the same 37 co-located stations

that contributed the 62 WSHR events. This sample was used to assess which features of the pre-onset kinematic evolution distinguish WSHR from non-heavy rainfall within the warm sector. WS non-HR events were defined as rainfall events that satisfied the warm-sector criteria at onset but never reached the station-specific heavy-rainfall threshold. The two samples were processed using the same RWP quality-control criteria, onset alignment and 180-min analysis window.”

In the paragraph following Eq. (6). *“The same KE diagnostic was calculated for both WSHR and WS non-HR events, with the corresponding time–height composites shown in Fig. 9c and Fig. S6, respectively.”*

In the 3rd paragraph of Sect. 3.4: *“The wind kinetic energy generally increases with height in both the WSHR and WS non-HR composites (Fig. 9c and Fig. S6). The WSHR composite shows a more pronounced enhancement at approximately 1–3 km AGL during the final hour before local rainfall onset, whereas the corresponding evolution in WS non-HR is comparatively gradual.”*

In the 8th paragraph of Sect. 3.4: *“The comparison with WS non-HR events indicates that low-level shear, clockwise wind turning and ascending motion occur in both rainfall categories. The more pronounced late increase in LLJ occurrence and upper-tail ascent at 1–2 km in WSHR provides partial evidence of differing pre-onset evolution. However, substantial overlap between the samples remains, and neither a causal jet–ascent pathway nor event-level predictive skill is established by these composites.”*

In the 1st paragraph of Sect. 4.3: *“In particular, the late increase in LLJ occurrence and enhanced low-level ascent motivate further examination, although their specificity, useful lead time and false-alarm characteristics remain to be established through event-level validation.”*

Fig. 12. Comparison of ascending-motion frequency and layer-maximum ascent speeds for WSHR and WS non-HR. The two samples share low-level ascent, while the late upper-tail enhancement at 1–2 km is more pronounced in WSHR.

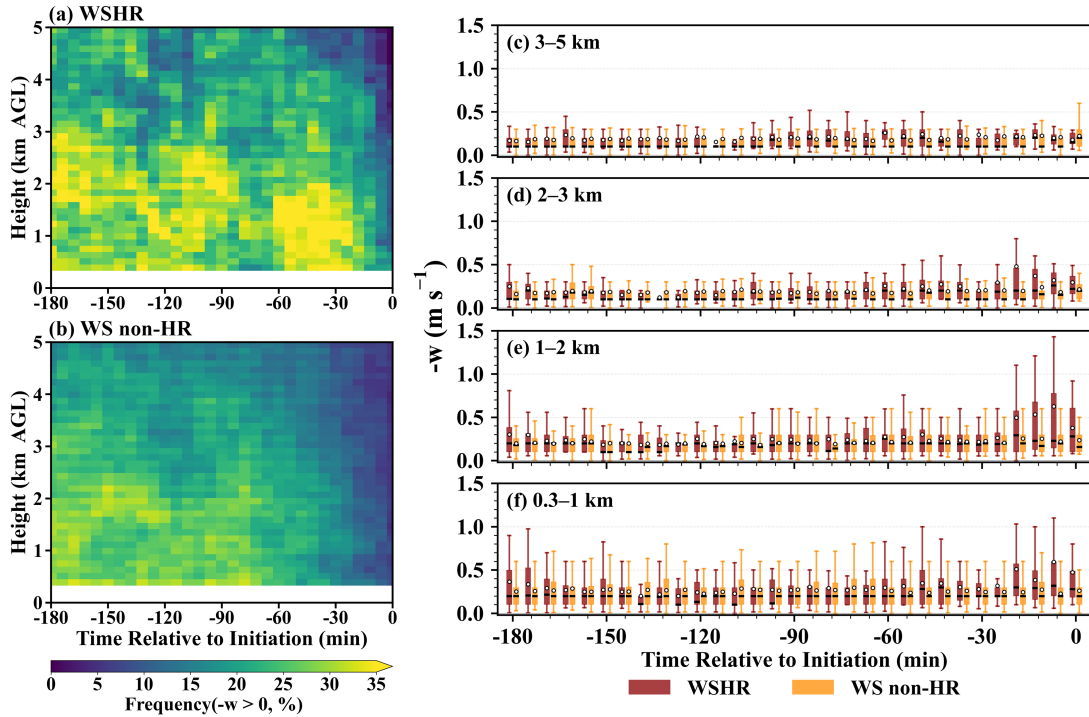


Figure 12. Composite ascending motion preceding the onset of WSHR events and WS non-HR events. (a, b) Frequency of ascending motion ($-w > 0$; %) as a function of height (km AGL) and time relative to the rainfall onset (min) for (a) WSHR events and (b) WS non-HR events. (c–f) Box plots of the across-event distributions of layer-maximum ascent speed ($-w$; m s^{-1}) at each 6-min lead time for the WSHR (dark red) and WS non-HR (orange) events in the (c) 3–5 km, (d) 2–3 km, (e) 1–2 km, and (f) 0.3–1 km AGL layers.

Minor / Specific Comment

(1) L37–38 (Abstract): "contributed 13.8% of the total heavy-rainfall accumulation, and locally 40–50%" — please clarify in the abstract (as is done in the text, Fig. 6e) that the 40–50% figure refers to station-level maxima rather than a regional mean, to avoid ambiguity for readers who read only the abstract.

Reply: Thank you for noting this ambiguity. The Abstract now replaces "locally 40–50%" with the maximum station-level value of 48.1%. We have also clarified the averaging population: 13.8% is the mean station-level contribution among the 297 stations that recorded at least one WSHR event, not the rainfall fraction obtained by pooling all stations. This distinction is now stated in the Abstract and Sect. 3.2 and retained in the Conclusions.

(2) L91–95 (Introduction): The rationale for extending the objective warm-sector scheme to the eastern China plains is clear; consider briefly stating up front (rather than only in 4.1) that the 200–400 km distance band and percentile thresholds are adopted from prior studies rather than independently tuned for this region, so readers

are aware of this choice earlier.

Reply: We have added an explicit clarification in the Introduction. To distinguish the origins of the two choices accurately, the revised text states that the 200–400 km distance band follows prior studies, whereas the 99th-percentile criterion was chosen for this study; neither was specifically optimized for this region.

Sect. 2.2 further specifies that each station's threshold is calculated from its non-zero hourly rainfall totals pooled over June–September of 2021–2025, excluding missing observations. The limitations concerning untested parameter sensitivity remain in Sect. 4.1.

(3) Eq. (7), L289–293: Please confirm the sign convention and units are consistent with the axis labels in Fig. 10 (10^{-3} s^{-1}); a brief note on how bulk shear direction (not just magnitude) might relate to the veering profile discussed in Fig. 9b would help connect Section 3.4's shear results to the warm-air-advection interpretation given for the wind profile.

Reply: The sign convention and units are now clarified immediately after Eq. (7). The bulk wind difference is defined as the top-level vector minus the bottom-level vector. Reversing this order changes the vector direction but not its magnitude. VWS is the magnitude divided by layer depth in meters, so it is non-negative and has units of s^{-1} . Fig. 10 displays 10^3 times VWS and labels the scale in 10^{-3} s^{-1} .

To show wind-direction changes that the scalar VWS does not capture, we have added composite hodographs for WSHR and WS non-HR in Figs. S8 and S9. Sect. 2.4 distinguishes the wind-vector difference, expressed in m s^{-1} , from the depth-normalized scalar in Eq. (7). Sect. 3.4 relates the WSHR hodograph to the veering profile in Fig. 9b: the preceding 3-h mean turns clockwise by about 80° between the lowest and highest sampled levels. Clockwise turning is also present in the WS non-HR composite, so this structure is not presented as unique to WSHR.

The specific changes made accordingly in the revised manuscript are listed below:

In the paragraph following Eq. (7). *“The top-minus-bottom convention defines the orientation of the bulk-shear vector; reversing the order would reverse the vector but would not change its magnitude. The scalar VWS is therefore non-negative and has units of s^{-1} .”*

“To retain directional information, composite hodographs are also presented for WSHR and WS non-HR events in Figs. S8 and S9.”

In the 2nd paragraph of Sect. 3.4: *“The composite hodograph also shows clockwise turning between the lowest and highest sampled levels, with a turning angle of about 80° for the preceding 3-h mean (Fig. S8a). Clockwise turning is also present in the WS non-HR composite (Fig. S9).”*

Fig. S8. Composite WSHR hodographs. The colored points identify height, and

the orange arrows show the bulk wind-vector difference.

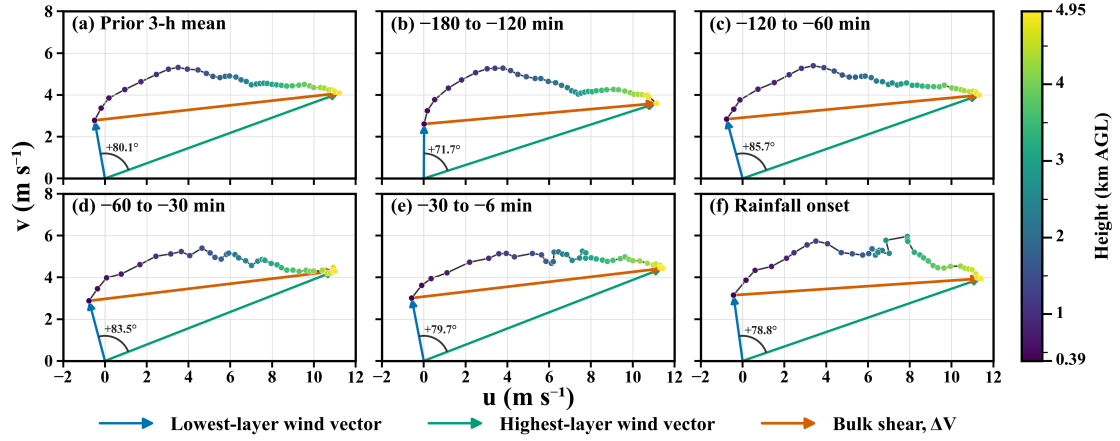


Figure S8. Composite hodographs for WSHR events: (a) the 3 h preceding rainfall onset, (b–e) the indicated pre-onset intervals, and (f) rainfall onset. Colors indicate height above ground level. Blue and green arrows denote the lowest- and highest-layer wind vectors, respectively; orange arrows indicate the top-minus-bottom wind-vector difference (ΔV , m s^{-1}). Annotated angles show the angular separation between the two wind vectors.

Fig. S9. Corresponding WS non-HR hodographs, showing that clockwise turning also occurs in the comparison sample.

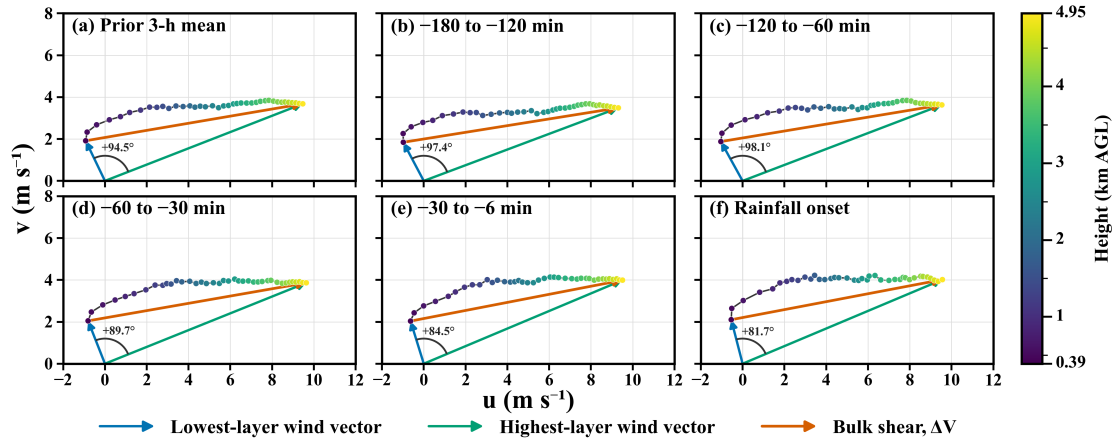


Figure S9. Same as Fig. S8, but for WS non-HR events.

(4) L429–430: "62 WSHR events across 48 stations" — please clarify whether events from the same station are treated as independent samples in the composite and bootstrap analyses (Sect. 3 of the Supplement), given potential spatial/temporal autocorrelation among nearby stations during the same synoptic episode.

Reply: The original bootstrap used station-based events as the resampling units. Resampling complete event sequences preserves within-event temporal dependence but does not by itself account for dependence among station events associated with a shared weather episode. We now state this distinction explicitly and have expanded Supplement, Sect. 3 to describe sample grouping, composite weighting and both

bootstrap approaches.

The station count has been corrected from 48 to 37. The sample still contains 62 WSHR events. Under the 500-km/24-h grouping, these events belong to 47 episode clusters. Fig. S4 shows the distributions of events per station and per cluster. Fig. S5 compares equal-episode and station-event composites, and Fig. S10 compares event and episode-cluster bootstrap intervals. In the latter procedure, each selected cluster contributes all of its member event sequences together.

An episode-cluster bootstrap with 5,000 resamples was also used to reassess the WSHR early-to-late LLJ change. For the -180 to -120 min and -60 to 0 min windows, including endpoints, the increase is 7.04 percentage points and its 95% interval is -0.25 to 14.18 percentage points. Excluding onset gives a similar mean increase of 6.94 percentage points and an interval of -0.52 to 14.22. Neither interval excludes zero. Accordingly, the revised manuscript describes the increase as an uncertain composite mean tendency. The operational cluster count is not treated as a verified number of independent systems.

The specific changes made accordingly in the revised manuscript are listed below:

In the 1st paragraph of Sect. 3.1: *“Following the procedure described in Section 2.4, 62 station-based WSHR events at 37 co-located stations were retained for the precursor analysis. Most stations contributed a single event (Fig. S4b). Using spatial and onset-time difference thresholds of 500 km and 24 h, respectively, these events were assigned to 47 episode clusters (Fig. S4d). These clusters are operational groupings of potentially related events; their statistical independence is not established. RWP profiles were aligned with each event’s local rainfall onset and composited over the preceding 180 min.”*

In the 2nd paragraph of Sect. 3.2 in the Supplementary materials: *“To assess sensitivity to dependence among station events, event and episode-cluster bootstrap intervals are compared in Fig. S10 using 5000 replicates. In the episode-cluster bootstrap, clusters are sampled with replacement and all member event sequences of each selected cluster are retained together. Frequencies are then recalculated from the sampled events. These intervals quantify uncertainty at individual lead times and are not simultaneous confidence bands for the entire time series.”*

In the 2nd paragraph of Sect. 3.3 in the Supplementary materials: *“The mean paired change is 7.04 percentage points. Resampling the 47 episode clusters with replacement 5000 times, with all member events retained and a random seed of 42, gives a percentile-based 95% confidence interval of -0.25 to 14.18 percentage points. A sensitivity calculation uses the windows from -180 to -126 min and from -60 to -6 min, excluding rainfall onset. Each window contains 10 observations. The mean change is 6.94 percentage points, with a 95% confidence interval of -0.52 to 14.22 percentage points. Both estimates indicate an increase in the sample mean, but neither interval excludes zero.”*

Fig. S10. Pointwise LLJ-frequency intervals from event and episode-cluster

bootstrap resampling. The figure assesses sensitivity to the resampling unit; the early-to-late paired changes are tested separately in Supplement, Sect. 3.3.

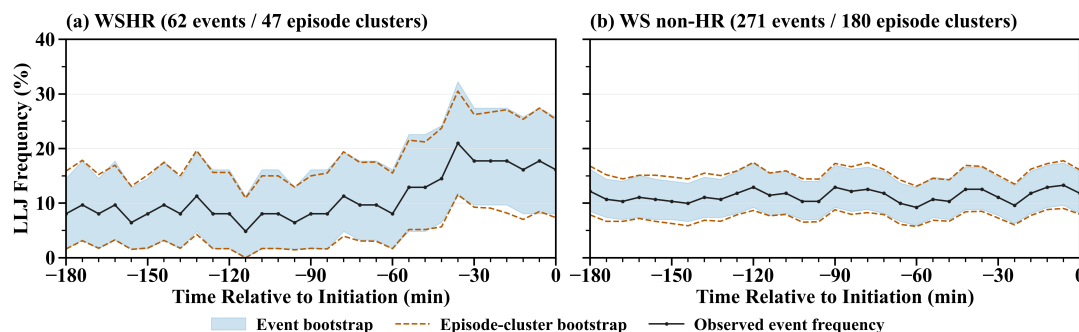


Figure S10. Event and episode-cluster bootstrap comparisons of LLJ frequency for (a) WSHR and (b) WS non-HR. Black curves show observed frequencies at 6-min intervals relative to local rainfall onset. Blue shading and orange dashed boundaries denote pointwise 95% confidence intervals from 5000 event and episode-cluster bootstrap replicates, respectively. Episode clusters use 500-km and 24-h thresholds.

(5) Fig. 11b: "Num = 29" — please state explicitly in the caption or text what fraction of the 62 WSHR events contained an identifiable LLJ at any point in the 180-min window, versus the ~20% figure quoted for individual time steps (L465–473), to avoid reader confusion between event-level and time-step-level frequencies.

Reply: We have explicitly distinguished the two statistics in Sect. 3.4 and the revised Fig. 11 caption. An LLJ occurs at least once within –180 to 0 min in 29 of the 62 WSHR events, giving an event-level proportion of 46.8%. The corresponding WS non-HR value is 107 of 271 events, or 39.5%.

The frequency curve in the revised Fig. 11d instead gives the fraction of events with an LLJ at each individual 6-min lead time. For WSHR, it peaks at 21.0% at –36 min and is 16.1% at onset. Thus, the event-level 46.8% is not the frequency at a single time step. The caption also states that the scatterplot averages are calculated over LLJ-containing profiles.

(6) L546–547 (4.1): The point that "compositing 62 events smooths the large case-to-case variability... best read as robust mean tendencies, not event-level certainties" is well taken and should be echoed, even briefly, in the Conclusions (Section 5), since the conclusions currently state the precursor findings somewhat more categorically than the discussion does.

Reply: The final paragraph of the Conclusions states that the composites describe mean tendencies that need not occur in every event. It also notes the overlap with WS non-HR and the unverified causal links. Specificity and event-level predictive skill remain to be evaluated.

Recommendation

***I recommend minor revision.** The manuscript makes a solid and novel observational contribution, and my comments are intended to strengthen the causal/thermodynamic linkage (M1) and to move the precursor-signal claim from a plausible hypothesis toward a more quantitatively supported one (M2), which I believe is well within reach given the dataset already assembled.*

Reply: Thank you for your positive assessment and focused guidance. The revised manuscript now provides the additional comparison and dependence assessment while stating the limits of the thermodynamic timing, causal interpretation and predictive application more explicitly.