Response to Reviewer #2

Corresponding author: Yusuke Hiraga
September 3, 2026

We sincerely thank the reviewer for the careful and constructive evaluation of our manuscript. We particularly appreciate the reviewer's positive assessment of the study's motivation, the precipitation-efficiency framework, the combination of the long-term AR database with high-resolution Radar–AMeDAS observations, the SOM classification, and the use of Huber-regression residuals to examine rainfall variability beyond IVT.

We considered all of the comments. As with the revisions made in response to Reviewer #1, we have focused on clarifying the scope of the present analysis and avoiding interpretations that go beyond what can be supported by the current methodology.

Several of the suggested analyses would certainly strengthen a future, more process-oriented version of this work, including alternative precipitation metrics, field-significance testing, multivariate predictor analysis, spatially constrained AR–QSLRB matching, and sensitivity tests using alternative AR detection criteria. We have not added these analyses in the present revision. Instead, we have explicitly acknowledged the corresponding limitations, moderated the relevant conclusions, clarified the statistical methods, and incorporated the references suggested by the reviewer.

General Comments

This manuscript addresses a useful gap in the literature. Most AR–precipitation studies in East Asia quantify the fraction of extreme events that coincide with ARs, which in effect treats the presence of an AR as a sufficient condition for heavy rainfall. The authors are correct to identify this limitation and to reframe the question as one of precipitation efficiency: given that an AR is present, what determines whether it produces damaging rainfall. The study combines an 84-year AR database with high-resolution radar–gauge observations and establishes an explicit link between synoptic-scale ARs and mesoscale quasi-stationary linear rainbands. The Self-Organizing Map classification is appropriate for the stated purpose, and the use of Huber-regression residuals to stratify events at comparable IVT is a reasonable way to isolate factors beyond moisture transport. I offer the following comments in the hope that they will help the authors further strengthen the manuscript.

Response:

We sincerely thank the reviewer for this positive assessment of the study framework. We agree that the central purpose of the manuscript is to move beyond the question of whether an AR is present and instead examine the variability in precipitation efficiency among AR events.

We have carefully revised the manuscript in response to the statistical and methodological concerns raised below. In particular, we have:

1. explicitly identified Welch's t-test as the statistical test used for the group comparisons;
2. clarified that the reported p-values are nominal and that the multiple-comparison results are exploratory;
3. removed statements implying that individual environmental variables represent independently established primary drivers;
4. clarified the limitations associated with the event-integrated, spatially averaged precipitation metric;
5. moderated the interpretation of the 1940–2023 AR-frequency trend;
6. clarified that the QSLRB association is based on temporal overlap and does not represent a spatially or randomly controlled attribution;
7. discussed the dependence of the results on the selected tARget v4 AR detection framework; and
8. clarified the temporal coverage of the QSLRB analysis.

These revisions are intended to preserve the main contribution of the study while ensuring that the conclusions are appropriately scaled to the analyses actually performed.

Major Comments

Major Comment 1. Multiple comparisons are not accounted for, which affects the precipitable water result.

Figure 4 reports ten significance tests, and Section 3.2.2 applies the same set of tests to the QSLRB comparison. The conclusion that precipitable water is the sole significant differentiator between AR events with and without QSLRBs (abstract; L260–263; L298) rests on a p-value of 0.02 obtained from approximately ten tests. This result would not remain significant under Bonferroni correction ($\alpha/10 = 0.005$) or under a moderate false discovery rate procedure. As this finding constitutes one of the four principal conclusions of the paper, the authors should either apply an appropriate correction and report the outcome, or state explicitly that the result is exploratory.

The manuscript does not identify which statistical test was used. Given the skewed distributions evident in Figure 4, the choice between a t-test, Welch's test, and the Mann–Whitney U test is material, and the omission prevents reproduction of the analysis. The authors should state the test used and report effect sizes, such as Cohen's d or the rank-biserial correlation, alongside the p-values.

Response:

We thank the reviewer for this important statistical comment. We agree that the multiple comparisons should be considered when interpreting the significance of individual variables.

We have not applied a Bonferroni or FDR correction in the revised manuscript, because the group comparisons were originally designed as an exploratory screening of environmental characteristics rather than as a formal multiple-hypothesis testing framework. We therefore agree that the original wording, particularly the statement that PW was the "sole significant differentiator," was too strong.

Accordingly, we now explicitly describe the p-values as nominal p-values and state that the group comparisons should be interpreted as exploratory associations. We have removed the expression "sole significant differentiator" from the Abstract, Results, Discussion, and Conclusion.

Regarding the statistical test, we confirm that the comparisons were performed using Welch's t-test. Welch's t-test was selected because it does not assume equal variances between the two groups. We have now explicitly stated the statistical test in the Methods and Results.

We have not added effect-size statistics in the present revision because the purpose of these comparisons is descriptive/exploratory rather than to establish an effect-size hierarchy among the variables. Instead, we have reduced the interpretation of the p-values and explicitly acknowledged the multiple-comparison issue.

Revisions in the Manuscript:

Original manuscript lines 219–238 — Figure 4

The interpretation has been revised to:

"Several environmental variables show nominally significant differences between the two residual groups, including moisture convergence, convective inhibition, mid-level humidity, mean latitude, and elevation. These differences indicate that rainfall efficiency varies with the dynamical, thermodynamic, and geographical environment in addition to the magnitude

of moisture transport. Because multiple comparisons were performed and the variables are physically coupled, these results should be interpreted as exploratory associations rather than as independently established effects."

Original manuscript lines 245–264 — Section 3.2.2

The original statement that PW was the "sole significant differentiator" and a "critical precursor" has been replaced by:

"PW showed a statistically distinguishable difference between QSLRB-associated and non-QSLRB events in the rainfall-matched comparison, suggesting that a moisture-rich atmospheric environment is an important characteristic of QSLRB-associated ARs. Because multiple variables were compared, this result should be interpreted as an exploratory association rather than as evidence that PW is uniquely responsible for QSLRB formation."

Statistical methodology

We have added:

"Differences in the environmental variables between the two groups were evaluated using Welch's t-test, which does not assume equal variances between the two groups."

Major Comment 2. The precipitation metric does not represent localized extremes.

Section 2.4 states that all fields are spatially averaged to yield a single scalar value per AR event (L100–101). Figure 3 accordingly shows event-total precipitation reaching approximately 25 mm. The stated motivation of the study is catastrophic localized rainfall (abstract, L18–19), and Section 3.2.2 concerns systems capable of producing more than 150 mm in three hours. A domain mean of order 10 mm is a substantially different quantity, and spatial averaging over the full intersection area attenuates precisely the localized signal the study seeks to explain.

The authors should repeat the residual analysis using a percentile-based or maximum-grid-cell metric, such as the 95th-percentile or areal-maximum event-total precipitation within the intersection, and report whether the discriminating variables change. Robustness across metrics would strengthen the conclusions considerably. Sensitivity to the choice of metric would itself be an important result.

Response:

We agree with the reviewer that the precipitation metric used in the residual analysis represents event-integrated, spatially averaged precipitation, rather than the localized precipitation maximum associated with the most damaging rainfall systems.

The purpose of this metric in the present study was to establish a consistent event-level relationship between AR moisture transport and precipitation over the AR–target-region intersection. This metric was deliberately selected to maintain consistency between the spatial domain used for the atmospheric variables and precipitation.

We acknowledge, however, that spatial averaging necessarily attenuates localized precipitation maxima. We have therefore revised the manuscript to make clear that the residual analysis concerns event-scale precipitation efficiency, rather than the intensity of the local rainfall maximum.

We have not repeated the residual analysis using the 95th-percentile or areal-maximum precipitation metric in the present revision. We agree that such a sensitivity analysis would be valuable for determining whether the environmental characteristics identified here are robust across precipitation metrics, and we now identify this explicitly as an important direction for future work.

Revisions in the Manuscript:

Original manuscript lines 96–101 — Section 2.4

The existing description of the event-level precipitation metric has been clarified.

Revised text:

"For each AR event, precipitation was accumulated over the event duration and spatially averaged over the intersection between the AR footprint and the target region. This metric represents event-scale precipitation associated with the AR–target-region intersection and is not intended to represent the local maximum rainfall intensity within the domain."

Discussion / limitations

We have added the following clarification to the consolidated limitations discussion:

"The precipitation metric used in the residual analysis is an event-integrated spatial average over the AR–target-region intersection. Although this provides a consistent event-scale measure for comparison with the spatially averaged atmospheric variables, it inevitably smooths localized precipitation maxima. Future work should examine whether

the identified environmental associations are robust when localized metrics, such as upper-percentile or areal-maximum precipitation, are used."

Major Comment 3. The trend analysis requires assessment of data homogeneity and field significance.

Two separate concerns apply to Figure 2c and the associated conclusion.

(a) The trend is computed over 1940–2023. ERA5 prior to 1979 relies on a considerably sparser observing system, and back-extension products are known to contain inhomogeneities associated with major observing-system transitions. A monotonic increase in detected AR frequency across 84 years may therefore reflect improving analysis capability in addition to any physical trend. The authors should repeat the Mann–Kendall analysis for the satellite era (1979–2023) and report whether the northern Japan signal persists. If the trend appears only in the full-period analysis, it should be presented with substantially greater caution.

(b) The test is applied independently at each 0.25° grid point across Japan, amounting to several hundred tests, without an assessment of field significance. The authors should apply a false discovery rate procedure (Wilks, 2016, BAMS) or a bootstrap field-significance test. In addition, annual AR frequency series may exhibit serial correlation. A modified Mann–Kendall test (Hamed and Rao) or prewhitening should be applied, or the absence of autocorrelation demonstrated.

Response:

We thank the reviewer for highlighting these important considerations concerning the long-term trend analysis.

We agree that the 1940–2023 AR database spans major changes in the observing system and that the statistical interpretation of a spatial field of trend estimates requires caution. The original manuscript presented the northern-Japan trend too strongly as an observational indication of a physical increase.

We have substantially moderated the interpretation of the trend. The revised manuscript now explicitly identifies the trend as a result of the ERA5/tARget historical dataset and acknowledges that the long-term trend may be affected by changes in the underlying observing system and by the multiple spatial testing problem.

We have also added the references suggested by the reviewer concerning field significance and autocorrelated Mann–Kendall testing: Wilks (2016) and Hamed and Rao (1998). Wilks (2016) emphasizes the need for appropriate interpretation of multiple spatial hypothesis tests, while Hamed and Rao (1998) explicitly address the effect of autocorrelation on Mann–Kendall trend significance.

Revisions in the Manuscript:

Original manuscript lines 143–150 / Figure 2c

The original statement of a "significant regional intensification" has been moderated.

Revised text:

"The Mann–Kendall analysis of annual AR frequency from 1940 to 2023 (Figure 2c) identifies a statistically significant positive trend over parts of northern Japan. Because the analysis spans the full ERA5 period, including years with substantially different observing-system coverage, and involves multiple spatial tests (Wilks, 2016), this long-term trend should be interpreted cautiously as an indication of a possible change in AR occurrence rather than as definitive evidence of a homogeneous physical trend."

Original manuscript lines 275–280 — Discussion

The discussion has similarly been revised to avoid treating the trend as unequivocal evidence of changing physical AR activity.

Revised text:

"The increasing trend in annual AR frequency identified over parts of northern Japan is noteworthy, but its interpretation requires caution. The 1940–2023 period spans substantial changes in the observing system underlying ERA5, and the grid-point trend analysis involves multiple simultaneous tests (Wilks, 2016). In addition, serial dependence can affect the significance of Mann–Kendall trend estimates (Hamed and Rao, 1998). Therefore, the detected trend should be regarded as an observational feature of the historical reanalysis-based AR dataset, while its robustness to the analysis period, serial dependence, and field-significance treatment remains to be examined."

Major Comment 4. Univariate comparisons cannot establish relative importance, and the predictors are likely collinear.

All conclusions regarding relative importance, including the description of MVIMC as a primary driver and of elevation as a decisive mechanism (L227–234), derive from ten separate two-group comparisons. These variables are unlikely to be independent. Lower AR-axis latitude (Fig. 4g) and higher landfalling elevation (Fig. 4j) may represent the same geometric effect, since an AR whose axis lies further south will intersect the mountainous interior of the target region rather than the coastal plain. Similarly, low CIN and high mid-level relative humidity are thermodynamically coupled.

A multivariate analysis is required before any variable can be described as primary. Suitable options include logistic regression on the positive/negative group label, or a random forest with permutation importance. In either case the pairwise correlation matrix among the ten variables should be reported. At minimum, the correlation matrix should be provided.

Response:

We agree with the reviewer that the ten two-group comparisons cannot establish the relative importance or independent contribution of individual predictors.

The underlying analysis included examination of pairwise correlations and variance inflation factors, which indicated substantial interrelationships among several predictor variables. These diagnostics reinforce the reviewer's point that variables such as latitude, elevation, CIN, and humidity should not be interpreted independently.

We have removed all wording that assigns primary or decisive importance to individual predictors.

We now describe MVIMC, CIN, RH500, latitude, and elevation as coupled environmental characteristics associated with enhanced rainfall efficiency. This is a more appropriate interpretation of the existing univariate comparison.

Revisions in the Manuscript:

Original manuscript lines 227–234 — Figure 4

The original statements describing MVIMC as a "primary driver" and elevation as a "decisive mechanism" have been removed.

Revised text:

"The positive residual group is characterized by stronger moisture convergence, lower CIN, higher mid-level humidity, a lower mean AR-axis latitude, and higher landfalling elevation. These differences indicate that enhanced rainfall efficiency is associated with a

combination of dynamical, thermodynamic, and geographical conditions. Because these variables are physically interrelated, the present comparisons do not establish their independent or relative contributions.”

Original manuscript lines 326–328 — Discussion

The previous “necessity factor” interpretation has been replaced by:

“The environmental variables examined here are not independent. Their physical coupling means that the observed differences between the residual groups should be interpreted as characteristics of the environmental configuration associated with enhanced rainfall efficiency rather than as evidence of individually dominant factors.”

Major Comment 5. The AR–QSLRB association is established on temporal overlap alone.

L126–128 states that a QSLRB is classified as AR-associated when its occurrence overlaps temporally with the AR lifetime. ARs are present over Japan very frequently during June to August, with 9,037 six-hourly detections in the record, and AR lifetimes are long relative to QSLRB durations. Matching on time alone will therefore produce a substantial number of coincidental associations and will inflate the reported figure of 19.34%.

The authors should add a spatial criterion, such as requiring the QSLRB centroid to lie within the AR footprint or within a specified distance of the AR axis, and report how the statistics change. A null expectation should also be provided: what fraction of QSLRBs would be classified as AR-associated if the AR catalogue were randomly shuffled in time. Without this baseline, the reader cannot assess whether 19.34% is high or low.

Response:

We agree that temporal overlap alone does not establish a physically attributable AR–QSLRB relationship and that a spatial criterion and randomized null expectation would provide a stronger basis for attribution.

The current analysis was designed to identify QSLRB events occurring during the lifetime of a landfalling AR, rather than to establish a statistically controlled attribution probability. Therefore, the reported 19.34% should be interpreted as the fraction of QSLRB-associated AR events under the temporal-overlap definition, not as the probability that an observed QSLRB is caused by an AR.

We have not added the proposed spatial-distance criterion or a randomized temporal null experiment in the present revision. However, Figure 6 already provides qualitative evidence that the identified simultaneous AR–QSLRB cases are spatially close: the QSLRB outlines are generally located near the AR axis.

We have therefore clarified the definition of the association and substantially moderated the causal interpretation. In particular, statements describing the AR as a "robust precursor" or QSLRBs as being "intrinsically linked" to the AR axis have been revised.

Revisions in the Manuscript:

Original manuscript lines 126–128 — QSLRB association definition

The definition has been clarified as follows:

"A QSLRB was classified as AR-associated when its occurrence temporally overlapped with the lifetime of a landfalling AR. This criterion identifies temporal co-occurrence and does not by itself establish spatial attribution or causality."

Original manuscript lines 261–264 — QSLRB results

The interpretation has been revised as described under Major Comment 1.

Original manuscript lines 268–283 — Figure 6 discussion

The original causal wording has been moderated.

Revised text:

"Figure 6 shows that the identified AR-associated QSLRB events generally occur in close spatial proximity to the AR axis. This spatial correspondence provides qualitative support for a physical connection between the synoptic moisture-transport corridor and the mesoscale rainbands. However, because the formal AR–QSLRB classification is based on temporal overlap, the present analysis does not quantify the statistical significance of this spatial correspondence or distinguish it from the frequency of coincidental temporal overlap."

Major Comment 6. Sensitivity to the AR detection algorithm is not examined.

The authors cite ARTMIP results demonstrating substantial disagreement among AR detection algorithms (Collow et al., 2022), but employ a single algorithm (tARget v4) throughout. Because the intensity criterion uses a location- and month-dependent 85th-percentile IVT threshold, the detected AR population during the Baiu season is sensitive to this choice. The authors should at minimum discuss how the conclusions may depend on the detection algorithm. A repetition of the key residual analysis using a fixed IVT threshold, for example 250 kg m^{-1} s^{-1}, would provide a useful robustness check.

Response:

We agree with the reviewer that the choice of AR detection algorithm is an important source of methodological uncertainty.

The present study uses tARget v4 consistently throughout the analysis. As the reviewer notes, this algorithm uses a location- and month-dependent 85th-percentile IVT threshold together with geometric and directional criteria. Consequently, the event population and the resulting precipitation-efficiency relationships may depend to some degree on the selected detection framework.

We have not repeated the residual analysis using a fixed 250 kg m^{-1} s^{-1} threshold, because changing the detection criterion would constitute a separate sensitivity experiment and could alter the definition of the AR population itself.

Instead, we have strengthened the discussion of algorithm dependence. We also retain the citation to Collow et al. (2022), which demonstrates substantial uncertainty among AR detection methods, and explicitly state that the conclusions should be understood within the tARget v4 framework.

Revisions in the Manuscript:

Original manuscript lines 51–53 / 30–50

The description of the tARget v4 database has been supplemented with a methodological caveat.

Added text:

"Because AR detection can vary among algorithms and threshold definitions, the event population identified in this study should be interpreted within the tARget v4 framework. Sensitivity of the precipitation-efficiency relationships to alternative AR detection algorithms and intensity thresholds remains an important subject for future work."

Discussion / limitations

We have added:

"The present results are also conditional on the tARget v4 AR detection framework. Differences among AR detection algorithms can alter the identified AR population and associated precipitation relationships (Collow et al., 2022). Future sensitivity experiments using alternative detection criteria would therefore be useful for assessing the robustness of the identified environmental associations."

Minor Comments

Minor Comment 7. SOM configuration.

L141–143 states that multiple configurations were tested and that a 2 × 2 structure was selected. The quantization error and topographic error for the tested configurations should be reported so that the choice can be evaluated. In addition, the use of a single representative timing per event (L85) discards the temporal evolution of the AR. This should be noted as a limitation, particularly because the discussion (L312–314) identifies sub-event evolution as an important direction for future work.

Response:

Thank you for this comment. We agree that quantization error and topographic error would provide quantitative information for evaluating alternative SOM configurations.

The 2 × 2 configuration was selected because it provided physically interpretable and distinct AR trajectory patterns while avoiding excessive fragmentation of the event population.

We have nevertheless clarified that this represents a physically motivated classification rather than a uniquely optimal statistical resolution.

We also agree that selecting one representative timing for each AR event removes information concerning the temporal evolution of the AR. This limitation is now explicitly stated in the Discussion.

Revisions in the Manuscript:

Original manuscript lines 141–143

Revised text:

"After testing multiple SOM configurations, a 2 × 2 node structure was selected because it provided physically interpretable and distinct AR trajectory patterns while avoiding excessive fragmentation of the event population. The selected configuration should therefore be regarded as a physically motivated classification rather than as a uniquely optimal SOM resolution."

The current methodology selects the representative timing based on maximum spatial overlap with Japan.

Original manuscript lines 312–314 — Discussion / limitations

Added text:

"In addition, the climatological SOM classification uses one representative timing for each AR event, selected according to its maximum spatial overlap with Japan. This approach provides a consistent representation of event-scale AR structure but does not capture the temporal evolution of individual ARs. Future analyses using the full temporal sequence of each event could provide additional information on how evolving AR structure influences precipitation efficiency."

Minor Comment 8. Analysis period in Section 3.2.2.

The analysis is restricted to 2006–2023 without explanation. This is presumably because the resolution of the RA dataset changes from 5 km to 1 km in 2006 (L76), which is a sound justification. The reason should be stated in the text.

Response:

Thank you for pointing this out. We agree that the reason for restricting the QSLRB analysis to 2006–2023 should be stated explicitly.

The Radar–AMeDAS dataset has 1-km spatial resolution from 2006 onward, whereas the earlier dataset has 5-km resolution. Because the QSLRB analysis concerns mesoscale precipitation structures and relies on the high-resolution rainfall observations, we restricted this part of the analysis to the 1-km period.

Revisions in the Manuscript:

Original manuscript lines 76 and 245–250

We have added the following clarification:

"Because the Radar–AMeDAS dataset has 1-km spatial resolution from 2006 onward, whereas the earlier dataset has 5-km resolution, the QSLRB analysis was restricted to the 2006–2023 period to maintain consistent spatial resolution for the identification and analysis of mesoscale precipitation structures."

Additional revisions to the manuscript

In addition to the point-by-point responses above, we have conducted a focused revision of the manuscript to ensure consistency between the statistical evidence and the strength of the conclusions.

1. Abstract

Original manuscript lines 17–23

The original Abstract stated:

"rainfall intensity is primarily modulated by a combination of strong moisture convergence (MVIMC), low convective inhibition (CIN), and orographic enhancement over high elevations"

and that PW was the "critical differentiator" for QSLRBs.

These statements have been revised to:

"IVT explains a substantial portion of the overall relationship (R = 0.71), but considerable variability remains among events with similar IVT. Within our event-based residual-group comparison, enhanced rainfall efficiency is associated with stronger moisture convergence, lower convective inhibition, higher mid-level humidity, and higher landfalling elevation. These results indicate that synoptic moisture transport alone does not fully determine rainfall efficiency and that the thermodynamic and geographical environment also contributes to event-to-event variability."

For QSLRBs:

"PW showed a statistically distinguishable difference between QSLRB-associated and non-QSLRB events in the rainfall-matched comparison, suggesting that a moisture-rich atmospheric environment is an important characteristic of QSLRB-associated ARs."

2. Figure 4 interpretation

The original Figure 4 discussion strongly interpreted individual variables as mechanisms. The manuscript currently states, for example, that MVIMC is a “primary driver” and that elevation is a “decisive mechanism.”

These statements have been replaced by the following more cautious interpretation:

“Several environmental variables show nominally significant differences between the two residual groups, including moisture convergence, convective inhibition, mid-level humidity, mean latitude, and elevation. These differences indicate that rainfall efficiency varies with the dynamical, thermodynamic, and geographical environment in addition to the magnitude of moisture transport. Because multiple comparisons were performed and the variables are physically coupled, these results should be interpreted as exploratory associations rather than as independently established effects.”

3. QSLRB interpretation

The original manuscript states that PW was the “sole significant differentiator” and a “critical precursor.”

This has been revised throughout the Results, Discussion, Abstract, and Conclusion to:

“PW showed a statistically distinguishable difference between QSLRB-associated and non-QSLRB events in the rainfall-matched comparison, suggesting that a moisture-rich atmospheric environment is an important characteristic of QSLRB-associated ARs.”

This preserves the observed statistical result while avoiding the interpretation that PW has been demonstrated to be an independent causal or temporal precursor.

4. Trend interpretation

The original manuscript described the 1940–2023 northern-Japan trend as a “significant regional intensification of AR activity.”

We now explicitly acknowledge that:

- the analysis covers the full ERA5 period;

- the observing system changes substantially during this period;
- the analysis consists of multiple grid-point tests;
- serial dependence has not been explicitly assessed; and
- therefore the result should be interpreted as a feature of the historical reanalysis-based AR dataset rather than as definitive evidence of a homogeneous physical trend.

We have added Wilks (2016) and Hamed and Rao (1998) to the reference list in connection with these statistical considerations. Wilks (2016) specifically discusses field significance and false discovery rate procedures, while Hamed and Rao (1998) addresses modified Mann–Kendall testing for autocorrelated data.

5. Consolidation of methodological limitations

To avoid repeating the same limitation in several sections of the manuscript, the relevant limitations are consolidated in the Discussion.

Revised limitations paragraph:

"Several limitations should be considered when interpreting the present results. First, the precipitation metric used for the residual analysis is an event-integrated spatial average over the AR–target-region intersection and therefore does not directly represent localized precipitation maxima. Second, the residual-group comparisons involve multiple, physically coupled environmental variables and are therefore exploratory rather than a basis for identifying independent dominant factors. Third, the long-term AR-frequency trend is based on the full 1940–2023 ERA5 period and involves multiple spatial tests; its robustness to observing-system changes, serial correlation, and field-significance treatment remains to be examined. Fourth, the formal AR–QSLRB association is defined by temporal overlap and does not by itself establish spatial attribution or causality. Fifth, the results are conditional on the tARget v4 AR detection framework. Finally, the SOM classification uses a single representative timing for each AR event and the event-scale analysis does not resolve the internal temporal evolution of ARs. Future work using alternative precipitation metrics, spatially constrained AR–QSLRB matching, multivariate analysis, alternative AR detection algorithms, and higher-temporal-resolution event tracking would provide useful robustness and process-oriented extensions of the present framework."

References added or clarified in response to Reviewer #2

The reviewer-suggested references are incorporated as follows:

- **Collow et al. (2022)** — already included in the manuscript and retained as support for uncertainty among AR detection algorithms.
- **Wilks (2016)** — added to discuss field significance and multiple spatial testing. The paper is *“The Stippling Shows Statistically Significant Grid Points”: How Research Results Are Routinely Overstated and Overinterpreted, and What to Do about It*, *Bulletin of the American Meteorological Society*, 97, 2263–2273.
- **Hamed and Rao (1998)** — added to discuss the effect of autocorrelation on Mann–Kendall trend testing. The paper is *A modified Mann-Kendall trend test for autocorrelated data*, *Journal of Hydrology*, 204, 182–196.