The initial submission of this manuscript received critical general comments from both reviewers, before the reviewers would provide more detailed reviews.

Both reviewers share fundamental reservation with the way that the authors attempt to diagnose quasi-resonant amplification (QRA). Both reviewers question the use of temperature (T) anomalies as external forcing in the QRA analysis. Reviewer#1 does so with a very basic explanation of why using T anomalies leads to a circular argument. In my own words: The very strong null hypothesis based on our fundamental understanding of midlatitude dynamics is that T anomalies within Rossby waves form *as part of* the Rossby wave by meridional advection and vertical motion (in case of a baroclinic Rossby wave). In any case: T anomalies themselves do not(!) provide a Rossby wave source. In fact, T anomalies may constitute the Rossby wave itself, e.g., as Eady edge waves or following Bretherton (1966). It is diabatic heating that provides a Rossby wave source. A strong argument is thus needed why T anomalies here could be used as a proxy for diabatic heating. In particular with the lack of land-sea contrast in the southern hemisphere (SH), it is hard for me to see how such an argument could be constructed. This aspect is crucial because with the near absence of orographic forcing in the SH midlatitudes, it is hard to see where the forcing for QRA should originate from. The response of the authors (provided to reviewer#2, who has phrased a milder version of the issue) does not sufficiently address this issue.
Reviewer#1 further questions the validity of the authors' analysis of circumglobal waveguidability, a necessary condition for QRA. The reviewer's own previous work, referred to by the authors, has shown that the method applied by the authors yields spurious results, i.e., application of the method may not only be inconclusive but may actually lead to incorrect conclusions. The authors respond that their focus is on relating south-east Australian heatwaves "to existing diagnostics presented in the literature before" and that they "use the QRA metric for only an empirical study." The authors "will clarify that the paper does not aim at further exploring the QRA mechanism itself." It is hard for me to see, however, what the empirical use of the "QRA metric" would be if not to explore the relevance of the QRA mechanism.
Given the severity of the criticism shared by two reviewers, I do not encourage to resubmit a manuscript in that the QRA analysis has been revised *according to the line of argument given in the authors' responses*. I rather suggest considering omitting this contented part from a revised version of the manuscript.

At this point, I would like to add a more general comment: Newly introduced diagnostic methods - just as other scientific results - need to stand "the test of time", i.e., the scrutiny by expert colleagues before becoming generally accepted „standard" diagnostics. The "QRA metric" is currently undergoing such scrutiny. The expectation for any diagnostic is that its application yields insight that cannot be not gained by merely looking at the data at hand. The issue raised by Wirth and Polster (2021) is not that the „QRA metric" would be a weak metric in the sense that its signal-to-noise ratio were poor. The issue raised is that the „QRA metric" yields incorrect results. If a diagnostic faces such fundamental criticism, valid arguments that nullify the criticism seem to be in order before further use of the diagnostic could be encouraged – even if the diagnostic had been used in a number of previous studies.

A further issue raised by reviewer#1 is the well-known issue of correlation vs. causation. While the authors are certainly aware of this issue, I sense in their responses that they may dismiss the issue to easily. Reviewer#1 suggests a plausible causal pathway by which a long-lasting heat wave may promote the occurrence of recurrent Rossby wave packets. The authors are clear that they "ultimately only diagnose "co-variability" ", but then they go on to argue that based on physical understanding their results do suggest a causal relation. While I much appreciate that the authors relate their statistical analysis to physical hypotheses, they do not seem to appreciate

that reviewer#1 has proposed an equally valid physical hypothesis that the authors' statistical results would support also. In statistical terms, the authors would need to control for this alternative casual pathway; and they should be aware that in the presence of confounders and colliders a simplistic correlation analysis may yield misleading results. Testing statistical significance (last paragraph of authors' response to reviewer#1) does not help with the causation vs. correlation issue.
I'd like to point out that so-called causal inference provides a rigorous mathematical framework that allows to establish cause and effect based on correlations. While this framework is not yet commonly applied in atmospheric sciences, an increasing number of studies do so, and a recent review of causal inference in Earth system science can be found in Runge et al. (2019). I do not mean to imply the recommendation that the authors use causal inference in their study – although this would certainly be interesting. I do recommend, however, that the authors do not take lightly the critique by reviewer#1 and thus are clear about their underlying physical hypotheses, discuss alternative hypotheses, and carefully word their causal interpretation of their statistical results in a revised version.

References:

Bretherton, F. P. (1966). Critical layer instability in baroclinic flows. Quarterly Journal of the Royal Meteorological Society, 92(393), 325-334.
Runge, J., Bathiany, S., Bollt, E. *et al.* Inferring causation from time series in Earth system sciences. *Nat Commun* **10,** 2553 (2019). https://doi.org/10.1038/s41467-019-10105-3