We thank Reviewer 1 for a thorough and constructive review. The comments have helped us strengthen both the methodological transparency and the scientific discussion of the manuscript. We address each point below (red), with specific changes to the revised manuscript indicated in each response.

General comments:

The paper is well structured and readable, and addresses a region that is not commonly discussed in atmospheric blocking papers. While I have some clarifying questions about the methodology below, I think that the overall work is quite robust and a useful contribution to the literature.

We appreciate the positive assessment.

Specific comments:

Line 93: How does ERA5 improve representation of the Southern Hemisphere, relative to ERA-Interim, in a manner that is relevant to the topic at hand? Better representation of wind, temperature, moisture transport?

ERA5 improves on ERA-Interim primarily through higher horizontal resolution (0.25° vs 0.75°), hourly output, and an updated data assimilation scheme (4D-Var with a more complete observation record). For the Southern Hemisphere circulation, these translate into a better representation of synoptic-scale features, including blocking anticyclones, the polar vortex edge, and PV gradients at upper levels. Studies such as Ploeger et al. (2021) and Hoffmann et al. (2012) and already cited in the manuscript, document improvements in dynamical consistency and stratospheric representation. Hersbach et al. (2020) also note improved wind and temperature fields in the extratropics due to better use of satellite-era observations.

We expanded the sentence at Line 93 to read: *"ERA5 provides improved representation of Southern Hemisphere circulation relative to ERA-Interim through higher horizontal resolution (0.25°), hourly temporal resolution, and an updated 4D-Var assimilation system that better constrains upper-level wind, temperature, and potential vorticity fields, which can directly affect blocking detection (Hersbach et al., 2020; Ploeger et al., 2021; Hoffmann et al., 2022)."*

Line 127: If you had hourly data, why not use the full hourly resolution rather than 6hr resolution? Was it resource constraints or something else, like trying to maintain consistency with past papers?

Both reasons apply, but the most important is data quality. ERA5 produces fully assimilated analysis fields at 00, 06, 12, and 18 UTC only, while the intervening hourly fields are short-range forecasts that are less reliable and subject to more errors and uncertainties.and therefore less constrained by observations. Using 6-hourly data ensures that the blocking detection is based on the most reliable ERA5 fields available. Consistency with Pinheiro et al. (2019) and reduced computational cost are additional motivations. Atmospheric blocking is, in any case, a slowly evolving phenomenon with lifetimes of days to weeks, so 6-hourly sampling captures its evolution without loss of relevant information.

We added a brief clarifying sentence at Line 127: *"Although ERA5 provides hourly output, we use 6-hourly fields (00, 06, 12, and 18 UTC) because these correspond to fully assimilated analysis time steps, whereas intervening hourly fields are short-range forecasts and thus less observationally constrained. This choice also maintains consistency with Pinheiro et al. (2019) and is well suited to blocking, which evolves on timescales of days to weeks."*

Line 128: What is the vertical resolution? 50hPA, something else?

We used all pressure levels available in ERA5 between 500 and 150 hPa for PV methods, i.e., a vertical resolution of 25 hPa between 150 and 250 hPa and 50 hPa between 250 and 500 hPa.

Sections 2.3.2 and 2.3.3: Initially I struggled to understand the difference between the two methods just based on the text/Table 1. It took looking at the Figure 1 schematic for me to finally understand it . They both sounds like they're anomaly-based from the descriptions, but one is called "Absolute threshold" and the other is called "standard deviation-based anomaly". Would suggest revising the text of these two sections to more clearly outline the calculation steps, as Figure 1 does a very nice job of explaining it visually.

Both reviewers identify the same issue, and we thank them for it. The naming is potentially confusing because both methods use climatological reference fields, but they differ fundamentally in what field the threshold is applied to. In Z500-Abs, the threshold is applied to the raw field: a grid point is blocked when the actual Z500 value exceeds the climatological mean plus a variability term. In Z500-SD, the threshold is applied to the anomaly field: a grid point is classified as blocked when its Z500 anomaly exceeds the larger of 100 m and 1.5 times the anomaly standard deviation calculated at that grid point and time of year. Thus, the method thresholds anomaly magnitude rather than the absolute Z500 field. Figure 1 (as Reviewer 1 notes) already illustrates this well visually, and we now make it equally clear in the text.

Beyond this conceptual difference, two further differences cause the methods to produce distinct results even when the algebraic forms appear equivalent.

<u>Standard deviations from different fields</u>. In Z500-Abs, σ is computed from the absolute Z500 field and therefore includes variance from the full seasonal cycle. In Z500-SD, σ is computed from the smoothed anomaly field and does not include seasonal cycle variance, making it systematically smaller. This means the effective detection threshold differs between the two methods even when the blocking condition is written in the same form.

<u>Latitude-dependent amplitude scaling</u> (in Z500-SD only). Before thresholding in Z500-SD, the anomaly is rescaled by a latitude-dependent factor following Pinheiro et al. (2019) to account for meridional differences in planetary wave amplitude. No such scaling is applied in Z500-Abs.

Explicit step-by-step summaries are added at the start of Sections 2.3.2 and 2.3.3. For Section 2.3.2: *"In this method, a grid point at time t is classified as blocked when the raw field value (Z500 or PV) equals or exceeds a spatially and temporally varying threshold, defined as the smoothed daily climatological mean plus a variability term. Formally:*

$$X(\lambda, \phi, t) \geq \overline{X}_{\mathrm{clim}(\lambda,\phi,d(t))} + \max\left[\delta_{\min}, 1.5\,\sigma_{\mathrm{clim}(\lambda,\phi,d(t))}\right]$$

where X denotes Z500 or PV, $\overline{X}$ is the smoothed daily climatological mean, σ_{clim} is the smoothed daily standard deviation of the absolute field, δ_{min} is the fixed minimum threshold value, and $d(t)$ is the day of year corresponding to time step t. The fixed minimum value is 100 m for Z500 and 1.1 PVU for PV."

A parallel formulation is added to Section 2.3.3 making explicit that the same inequality applies to the anomaly field rather than the absolute value.

Section 2.3.4: The data has 6hr resolution, but the thresholds are established on a daily basis? (my assumption based on line 246) Does that mean that the full day has a single threshold rather than one per time step? If so, why that single daily threshold and not per time step?

We thank the reviewer for raising this point. In the revised analysis, we have updated the percentile-threshold calculation to use a centred 15-day running window (±7 days) around each calendar day, pooled across all years of the 1991–2020 reference period. This provides a larger and more stable sample for estimating the 90th percentile while retaining the gradual seasonal evolution of anomaly variability. A single threshold is then applied to all four 6-hourly time steps within that calendar day. This is appropriate because the anomaly fields are based on a 3-day running mean and therefore represent variability on daily-to-synoptic timescales. Using separate thresholds for each 6-hourly time step would introduce unnecessary sub-daily threshold variability that is not consistent with the temporal scale of blocking.

Figure 4: Would like to see the polygons in 4b) shown in all items, not just b), so that it is clear which areas are being summarized in Table 2.

Modified as suggested.

Line 429 and Section 4.1/4.2 in general: Would we expect to see much blocking over Antarctica at all for the Z500 gradient method given the constraints of the latitudinal range? Trying to compare the frequencies of the Z500 gradient method and the other methods in the higher latitudes is rather misleading given the different latitudinal extents. Discussion should account for this fact.

This is a valid and important point. Thank you for this. The Z500-Grad method is explicitly restricted to 25°-75°S due to the geometric requirement of sampling at ±15° latitude, which makes it mathematically incapable of detecting blocking poleward of 75°S (since the southward reference would fall outside the detection domain). Any comparison of blocking frequencies over the Antarctic continent or at latitudes near 75°S between Z500-Grad and the full-domain methods is therefore not a fair like-for-like comparison and can mislead interpretation. We already noted this domain restriction in the methodology, but the discussion in Sections 4.1 and 4.2 did not consistently account for it when contrasting spatial patterns. Our new sensitivity experiments confirm this directly. Extending the gradient method south of 75°S using GHGS values of 0 and 10 m $°lat^{-1}$ yields very low blocking frequencies even at the most permissive settings, consistent with the absence of a strong background poleward geopotential height gradient over the Antarctic continental interior. A spatially coherent signal only appears when GHGS is simultaneously reduced to 0 m $°lat^{-1}$ and the area filter is relaxed to 0.5×10^6 km^2, and even then it is confined to the circumpolar belt south of approximately 70°S rather than the continent itself. We have provided these sensitivity analyses as supplemental material in the revised version. These results confirm that the near-zero Z500-Grad frequencies over Antarctica are not a detection failure but reflect a genuine dynamical property of the polar circulation. We have revised Sections 4.1 and 4.2 to consistently exclude Z500-Grad from frequency comparisons at latitudes poleward of 60°S, and added a sentence at the start of Section 4.1 noting that the Z500-Grad domain restriction makes direct frequency comparisons with full-domain methods at high latitudes inappropriate.

Figure 5: as with Figure 4, would like to see the polygons shown so that it's clear which areas are being summarized in the table

Modified as suggested.

Table 2: Perhaps would be more appropriate as a figure? It's very hard to easily digest this information in its current format. Suggest something like grouped boxplots.

Modified as suggested.

Figure 9: Would it be possible to provide a background color of a relevant field, such as 500hPA geopotential height, to provide some context to what exactly is being captured by the various methods? It would be interesting to see how "accurate" or "inaccurate" the various methods are or what they are specifically capturing (i.e., a ridge vs a rex block vs an omega block). Perhaps not feasible if it obscures the contours, but maybe worth trying?

Thank you for the suggestion. In the revised version, we have incorporated a background color of the corresponding field.

Section 4.5: I found this discussion to be rather unsatisfying with respect to the events being shown in Figure 9. It discusses the results in broad strokes rather than referring to what happened during the specific selected events. For example, in 9d, what is going on with those contours off the eastern coast of South America? Did they merge with their respective counterparts in 9e, or was that the detection of a local anomaly rather than a persistent blocking feature?

Agreed. We recognize that the original discussion of Figure 9, now Figure 10, was too general and did not sufficiently explain the specific features shown in the selected cases. We have therefore revised Section 4.5 to add event-specific descriptions while retaining the primary purpose of the figure: to illustrate how the different detection methods delineate the same broad circulation episode.

For the February 2020 event (Figs. 10d-f), the contours off the eastern coast of South America Fig. 10d reflect a separate, short-lived anticyclonic ridge over the South Atlantic, detected by the threshold-based Z500 methods (Z500-Abs and Z500-SD). This feature does not merge with the main blocking structure over the Antarctic Peninsula, and it weakens in the subsequent panels while the Peninsula feature becomes more consolidated, consistent with the documented February 2020 warm and melt event (Clem et al., 2021; Xu et al., 2021). The March 2022 event (Fig. 10g-i) captures the East Antarctic heat wave documented by Wille et al. (2024a), with the PV-based methods identifying a compact, dynamically coherent block near 90°–120°E while the Z500 threshold methods capture a broader ridge signature. In the revised version, we only showed the events with ID's assigned to atmospheric blocking of interest.

We rewrote Section 4.5 to briefly add event-specific descriptions for each of the three cases, referencing the contemporaneous literature and clarifying how the different detection methods capture distinct features within each event. We also note that this is a methods/dataset paper, rather than a paper studying extreme weather events. The questions the reviewer raises are an excellent example of precisely why we feel producing and making public this new comprehensive dataset will serve as an important resource to the scientific community, and we believe the dataset will support many new studies in which questions like these can be investigated. This is also why we present these events: to demonstrate how different methods capture different features, while recognizing that studying these events in detail is outside the scope of this current paper.

We thank Reviewer 2 for a thorough and constructive review. The comments have helped us strengthen both the methodological transparency and the scientific discussion of the manuscript. We address each point below (red), with specific changes to the revised manuscript indicated in each response.

In the article "**A Multi-Method Antarctic Atmospheric Blocking Dataset (1979-2024)**", submitted to Earth System Science Data, the authors apply a selection of seven of blocking detection methods developed in the midlatitude meteorology literature, and apply them to the Southern hemisphere excluding the tropics but including Antarctica. The authors then compare the climatological properties of blocking according to each metric (frequency and persistence) in great detail and also demonstrate the differences in instantaneous detection using a case study. Importantly, the dataset is made public and is described in enough details as to let researchers use it for their own work without onboarding pain, which is commendable.

The authors focus their discussion on the polar region where there is still a methodological gap in blocking detection. Their solution to help fill this gap is to gather many different methods for midlatitude blocking detection and to apply them all to this polar region without modification, to the thresholds most crucially, and then compare the results. While this provides a good starting point and the full dataset is a useful tool, I believe a critical assessment of the threshold values in this very different region could have made this work a lot more significant.

We agree that a critical assessment justifying our methodology and thresholds, particularly over the Antarctic, was missing, which would offer important new insights and ensure the dataset and methods are robust. We have now addressed this gap through a set of dedicated sensitivity analyses presented in supplementary material. The analysis covers three sets of experiments. For the Z500 and PV threshold methods, we tested both the fixed minimum threshold (Z500: 50, 100, and 150 m; PV: 0.6, 1.1, and 1.6 PVU) and the variability factor (1.0, 1.5, and 2.0 times the standard deviation). For the gradient method (Z500-Grad), we tested the minimum area filter (0, 0.5, and 1.0 × 10^6 km^2), the latitudinal spacing used to compute gradients (Δlat = 10° and 15°), and the southern gradient threshold (GHGS = 0, 5, and 10 m $°lat^{-1}$). The GHGS experiment was run at both area filter levels (0.5 and 1.0 × 10^6 km^2) to evaluate their combined effect.

For the threshold-based methods, the 50 m and 100 m criteria produce broadly similar blocking patterns. The 50 m threshold produces more frequent detections over the Antarctic interior, while the 150 m threshold is overly restrictive and misses substantial blocking over continental East Antarctica (Fig. R1). For the PV thresholds, the 0.6 PVU criterion yields comparatively widespread and potentially overdetected blocking (Fig. R2). The 1.6 PVU criterion is overly restrictive and captures little blocking over continental East Antarctic. The 100 m and 1.1 PVU thresholds provide a physically coherent balance.

For the variability factor, 1.0 picks up transient ridging and over-estimates blocking frequency, while 2.0 captures only the most extreme events (Figs. R3 and R4). The 1.5 value is consistent with Pinheiro et al. (2019) and produces the most physically reasonable results. Regarding which threshold component dominates spatially: the fixed minimum threshold controls detection primarily over the Antarctic interior and other low-variability regions, while the 1.5 standard deviation term dominates over the Southern Ocean storm track where variability is large (Figs. R3 and R4).

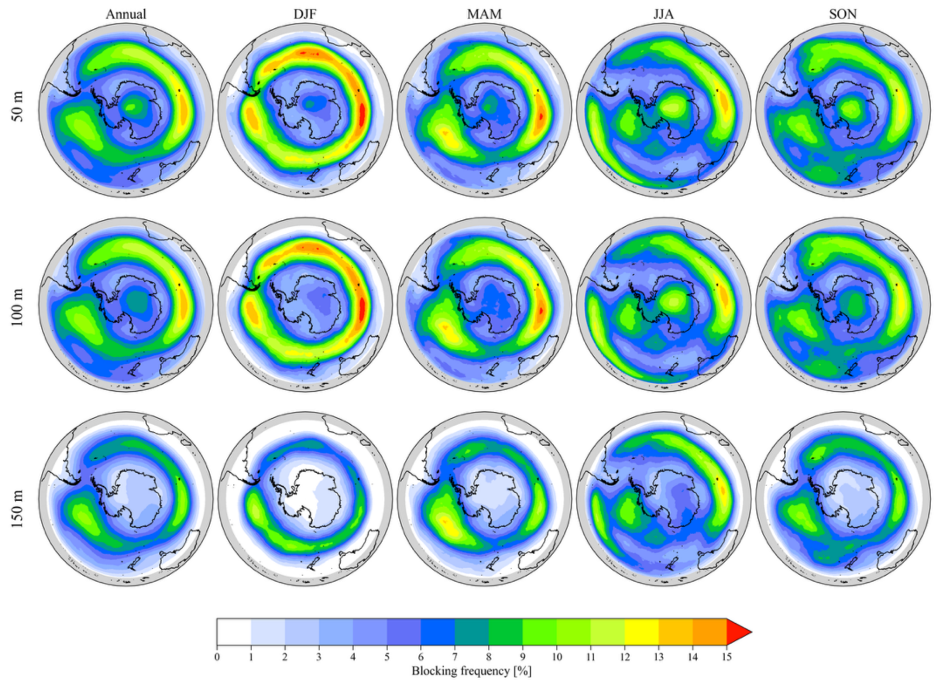

Fig. R1. Annual and seasonal blocking frequency (%) from the Z500-SD method using three minimum threshold values: 50 m (top), 100 m (middle, selected), and 150 m (bottom). Columns show annual, DJF, MAM, JJA, and SON blocking frequency for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

For the gradient method, the results are largely insensitive to the choice of Δlat. The GHGS threshold has a more pronounced effect: relaxing it from 10 to 0 m $°\text{lat}^{-1}$ at the standard area filter of 1.0×10^6 km^2 increases detection notably along the mid-latitude storm track belt, particularly in DJF and MAM, and also increases detection over eastern Antarctica (Figs. R5). The increase over eastern Antarctica becomes more pronounced when GHGS is relaxed to 0 m $°\text{lat}^{-1}$ and the area filter is simultaneously reduced to 0.5×10^6 km^2 (Fig. R6). We retain GHGS = 10 m $°\text{lat}^{-1}$ and an area filter of 1.0×10^6 km^2 as the settings for Z500-Grad for three reasons: (1) they are consistent with Pinheiro et al. (2019); (2) the continental signal at the more permissive setting is at frequencies where dynamically ambiguous quasi-stationary ridges cannot be confidently distinguished from classical blocks; and (3) the gradient method's role in the dataset is explicitly that of a conservative flow-geometry baseline; relaxing it toward the threshold methods reduces its diagnostic value. The sensitivity bounds are documented in the supplementary material.

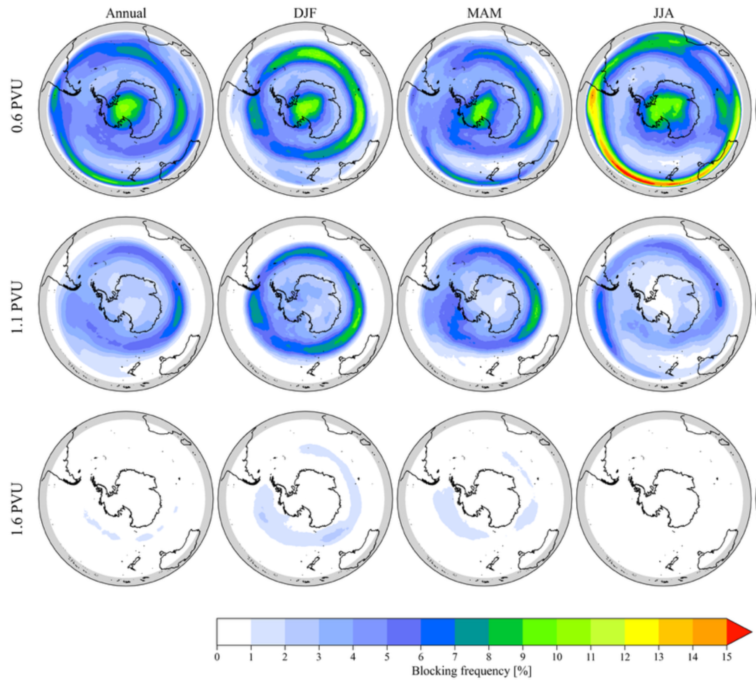

Fig. R2. Annual and seasonal blocking frequency (%) from the PV method using three minimum threshold values: 0.6 PVU (top), 1.1 PVU (middle, selected), and 1.6 PVU (bottom). Columns show annual, DJF, MAM, JJA, and SON blocking frequency for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

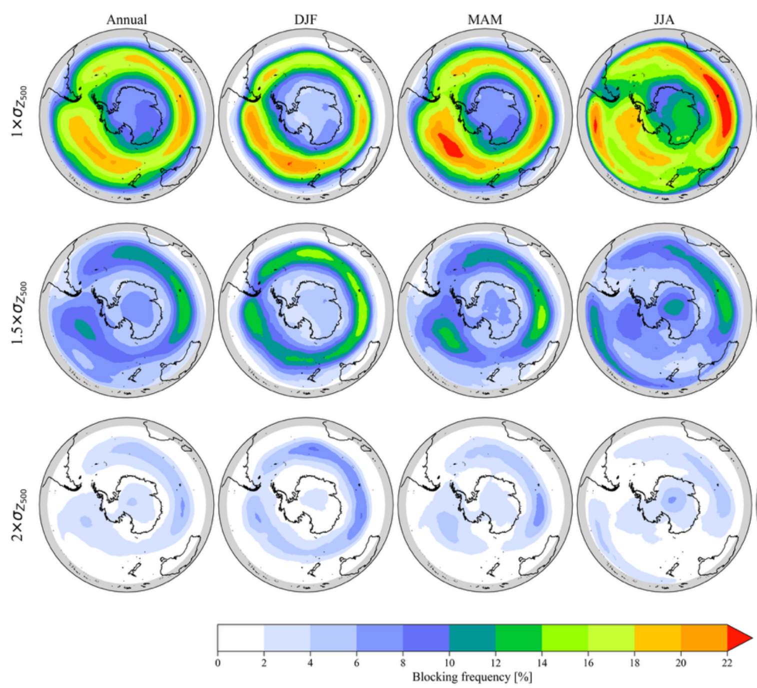

Fig. R3. Annual and seasonal blocking frequency (%) from the Z500-SD method using three variability factors: 1.0 (top), 1.5 (middle, selected), and 2.0 (bottom) times the local standard deviation. Columns show annual, DJF, MAM, JJA, and SON means for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

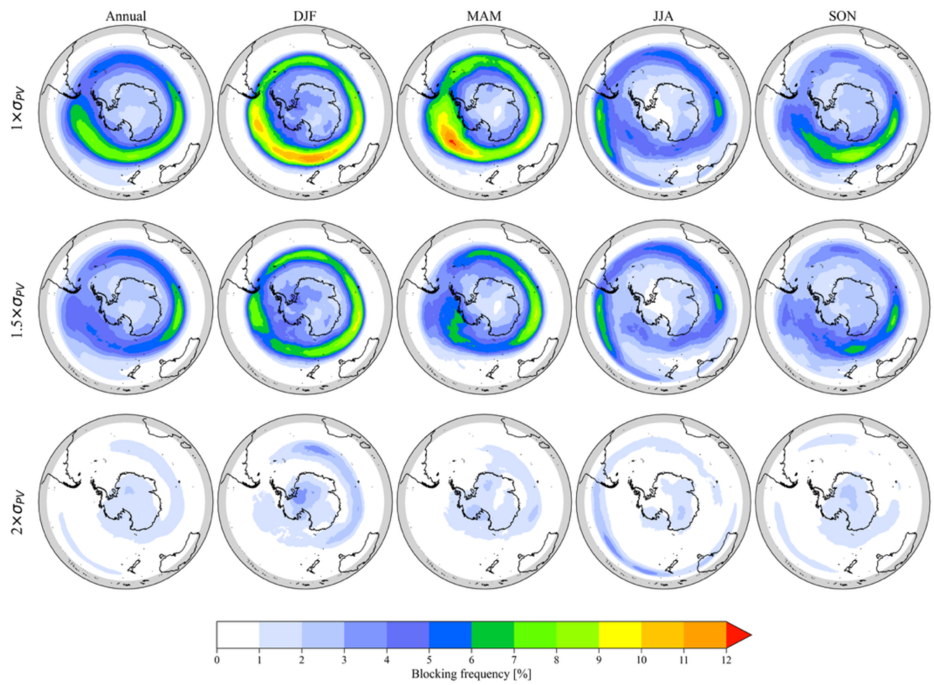

Fig. R4. Annual and seasonal blocking frequency (%) from the PV-SD method using three variability factors: 1.0 (top), 1.5 (middle, selected), and 2.0 (bottom) times the local standard deviation. Columns show annual, DJF, MAM, JJA, and SON means for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

Overall, the main spatial structures and inter-method contrasts are robust across these sensitivity experiments, and the conclusions do not depend on any single parameter choice.

We present the full sensitivity analysis as a supplementary material, including both text and figures, which is the appropriate location for this material. It provides complete documentation for users who wish to understand or adjust the threshold choices without disrupting the flow of the main manuscript.

Summaries of the sensitivity results are now incorporated into the relevant subsections of Section 2.3, with full documentation provided in the supplementary material. Section 2.3.1 summarizes the sensitivity of Z500-Grad to the latitudinal sampling distance, GHGS threshold, and area filter; Section 2.3.3 summarizes the sensitivity of the Z500-SD and PV-SD methods to the fixed minimum threshold and variability factor; and Section 2.3.5 discusses the sensitivity to the common area filter and the minimum overlap criterion. Together, these tests confirm that the selected parameter values, which follow Pinheiro et al. (2019), provide a physically coherent and conservative detection framework, and that the main spatial structures and inter-method contrasts are robust to plausible parameter variations.

Compounding with this issue, the authors do not provide the analysis code that created this dataset, such that the users cannot easily reproduce the dataset with the same methodology but with more carefully chosen thresholds in the region of interest. I will let the editor decide whether more work on the thresholds in the polar region should be done before resubmission or if this extra work is beyond the scope of this journal and the dataset as-is is sufficient to justify the publication.

We agree. The preprocessing and detection scripts are now archived in the same Zenodo repository as the dataset. The repository includes the full workflow from ERA5 preprocessing through TempestExtremes detection and tracking, with parameter choices documented at each step. Users can modify threshold values within the provided scripts to reproduce the dataset using different alternatives. Section 5 ("Data and Code Availability") was updated to explicitly state that the analysis code is archived alongside the dataset in the Zenodo repository.

There are other, smaller methodological points I would like to raise.

Is a fixed area threshold of 1million km2 the best choice? This is bound to create biases in number of PV and Z500 events because the fields inherently have different spatial coherence characteristics.

Indeed, this was an important threshold that we checked with sensitivity analyses in the early stages of constructing the dataset. Firstly, vertically integrated PV fields tend to be spatially more compact than Z500 fields, so a fixed-area threshold removes a larger fraction of PV-detected blocking features. For the gradient method specifically, the sensitivity analysis shows that reducing the area threshold from 1.0 to 0.5 × 10^6 km^2 increases the blocking event count, particularly through additional detection of smaller coherent features over the Antarctic continent and circumpolar latitudes, but does not substantially change the mid-latitude blocking climatology. For Z500-SD and PV-SD, reducing the area filter from 1.0 to 0.5 × 10^6 km^2 produces no significant change in the spatial blocking frequency pattern (Figs. S9 and S10 in the supplementary material). This confirms that the fixed 1.0 × 10^6 km^2 threshold does not introduce a substantial bias for these methods. Applying method-specific area thresholds would introduce an additional free parameter that complicates the intercomparison, which is the primary purpose of the dataset. The fixed threshold is therefore a deliberate design choice to isolate the effect of variable and threshold formulation from areal filtering.

The results of these sensitivity analyses are now documented in the supplementary material. Section 2.3.5 has been revised immediately after the description of the common 1.0 × 10^6 km^2 area requirement to acknowledge that Z500, PV, and gradient-based fields have different spatial coherence, and that the common area threshold can therefore contribute to inter-method differences in event counts. The revised text also refers readers to the supplementary material, where the area-filter sensitivity tests are documented.

To go back to the threshold problem, the paper the authors cite the most and seem to take most of their methods from is Pinheiro et al. 2019. In that article, the authors justify their choices of thresholds (maximum between 1.5 sigma and 100m, e.g.) by quantifying the standard deviation of the fields in the regions they are interested in. No such justification is provided for the polar region in this present work, and I think it should be there.

Related to the previous point: Between 1.5 std and 100m or 1.1 PV, which threshold is triggered more often and where ?

The sensitivity analysis now addresses this question directly. The fixed minimum threshold (100 m for Z500, 1.1 PVU for PV) affects blocking detection primarily over the Antarctic interior and other regions of weak climatological variability, where the standard deviation of the anomaly field is small enough that 1.5 times its value does not meet the threshold (Figs. R1 to R4). The 1.5 standard deviation term dominates over the Southern Ocean and mid-latitude storm tracks, where synoptic variability of the height field is large. This spatial partitioning is physically consistent with the rationale for using a combined threshold: the fixed floor prevents non-detection in low-variability polar regions, while the variability-scaled term provides appropriate sensitivity in the more dynamically active subpolar belt.

This is now stated briefly in Section 2.3.3 and fully documented in the supplementary material.

Was is the most judicious choice to apply the reversal index as is? It seems mostly unable to detect reversals over Antarctica, but with all thresholds kept identical to those used for the midlatitudes how might the reader interpret these results? Would a modified reversal index be more sensitive in this region? Or are there actually no reversals there?

The sensitivity analysis for Z500-Grad helps assess this question within the established 25°–75°S application domain. With an area filter of 1.0×10^6 km^2 and a GHGS threshold of 10 m °lat^{-1}, blocking detection over the Antarctic continent is essentially absent across all seasons. Relaxing the GHGS threshold to 0 m °lat^{-1} with the same area filter substantially increases detection along the mid-latitude storm track, particularly in DJF and MAM, but does not produce a meaningful blocking signal over the continent (Fig. R5). A coherent, spatially continuous signal over the Antarctic continent and circumpolar latitudes south of approximately 70°S only emerges when GHGS = 0 m °lat^{-1} and the area filter is simultaneously reduced to 0.5×10^6 km^2 (Fig. R6). Even then, annual blocking frequencies are only 1 to 4%, and the DJF peak is confined mainly to a narrow circumpolar band rather than the continental interior.

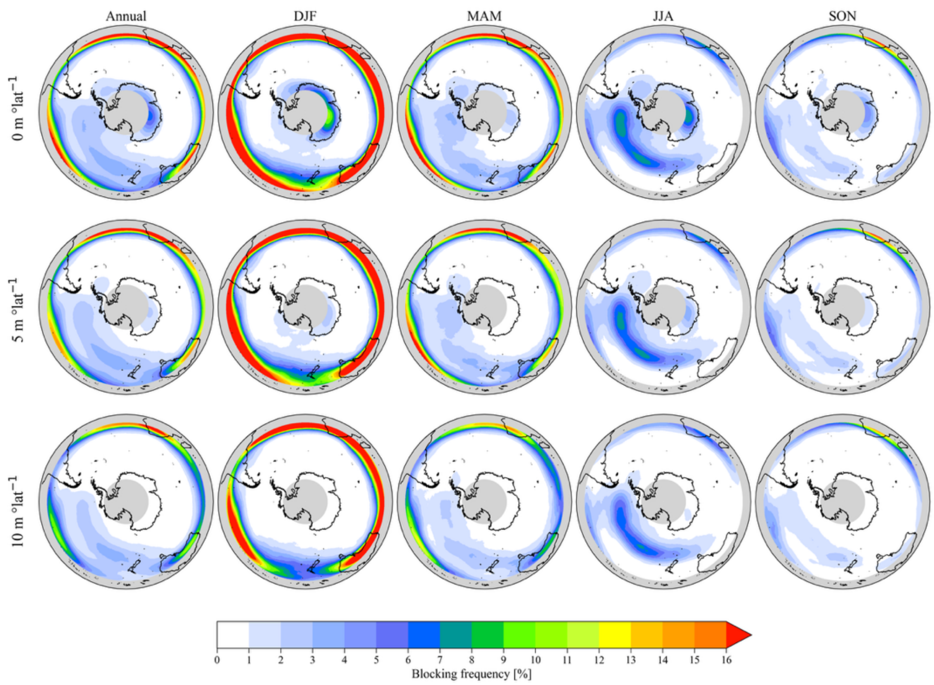

Fig. R5. Annual and seasonal blocking frequency (%) from the Z500-Grad method for three GHGS threshold values (0, 5, and 10 m °lat^{-1}, rows) using area filters of 1.0×10^6 km^2. Columns show annual, DJF, MAM, JJA, and SON means for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

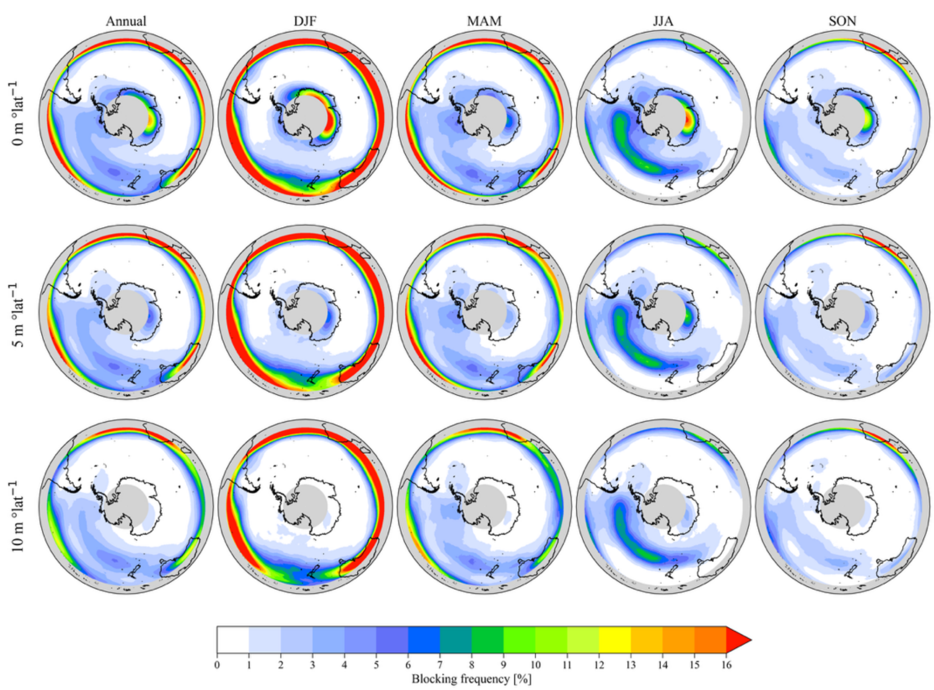

Fig. R6. Annual and seasonal blocking frequency (%) from the Z500-Grad method for three GHGS threshold values (0, 5, and 10 m °lat^{-1}, rows) using area filters of 0.5×10^6 km^2. Columns show annual, DJF, MAM, JJA, and SON means for 1979 to 2024. Regions shown in gray correspond to latitudes where blocking detection is not applied.

We additionally tested an extension of the gradient method south of 75°S under relaxed GHGS settings. The results showed high blocking frequencies over the Antarctic interior, but also a spatial discontinuity at the 75°S boundary. This indicates that the standard meridional-gradient formulation may not be robust near the pole, where geometric and numerical effects in the gradient calculation cannot be excluded. We therefore do not include these exploratory results or use them to infer the prevalence of polar flow reversals.

Accordingly, these tests cannot be interpreted as evidence either that classical blocking is absent over Antarctica or that the detected interior features represent genuine polar flow reversals. Although similar high interior frequencies have been reported by Sousa et al. (2021), a dedicated assessment of the physical nature of these features is beyond the scope of the present dataset paper. We retain the documented sensitivity experiments within the established 25°–75°S application domain and identify the development and validation of a polar-adapted reversal diagnostic as a priority for future work.

These results have two implications. First, within the established 25°–75°S application domain, gradient reversals in the classical Tibaldi–Molteni (1990) sense are rare over the Antarctic continental interior, consistent with the weak background poleward geopotential-height gradient at these latitudes. Second, when the domain is extended south of 75°S and the GHGS threshold is relaxed, higher apparent blocking frequencies emerge over the Antarctic interior; however, the discontinuity at the 75°S boundary means that these exploratory results should be interpreted with caution and are not used to infer the prevalence or nature of polar flow reversals. We retain the configuration GHGS > 10 m °lat^{-1} with a minimum area filter of 1.0 × 10^6 km^2 because it provides a conservative flow-geometry baseline consistent with Pinheiro et al. (2019). Its contrast with the threshold-based methods is an informative feature of the dataset. The sensitivity bounds for the tested configurations are presented in the supplementary material, and the full code is provided so that users can explore alternative threshold criteria and detection domains.

The following text has been added to Section 2.3.1: *"Sensitivity experiments varying the latitudinal sampling distance, the threshold applied to the southward meridional Z500 gradient (GHGS), and the minimum area filter are presented in Supplement S1 (Text S1.1; Figs. S1-S4). Over the Southern Ocean and Antarctic domain, the results are broadly insensitive to the latitudinal sampling distance (Fig. S2), whereas a discernible continental signal emerges only when both the GHGS threshold and the area filter are relaxed (Fig. S4). We retain the criterion GHGS > 10 m °lat^{-1} and a minimum area filter of 1.0 × 10^6 km^2. because these settings are consistent with Pinheiro et al. (2019) and preserve the role of Z500-Grad as a conservative flow-geometry baseline."*

A brief note is added in Section 6 that polar extension of the gradient method and adaptive formulations of the reversal index are topics of future development.

Tracking is beased on grid point overlap, but grid points represent diminishing areas with increased absolute latitude. Is persistence at the pole therefore comparable to persistence at 60°S?

This is a genuine limitation. Near the pole, a one-grid-cell overlap corresponds to a much smaller physical area than at 60°S, making the tracking criterion effectively less demanding at very high latitudes. This could inflate persistence statistics near the pole relative to lower latitudes. Because all methods share this limitation, it does not affect inter-method comparisons, but it is relevant when interpreting absolute persistence values in the high latitudes. We tested the sensitivity of blocking detection to the minimum overlap fraction used in StitchBlobs (min_overlap_prev), varying it from 0% (one grid cell, as used in the published dataset) to higher values for Z500-Grad, Z500-SD, and PV-SD. No significant change was found in the resulting blocking frequency patterns (Figs. S11 to S13).

We have noted this limitation in Section 2.3.5 and added that, because grid-cell area decreases toward the pole, the one-grid-cell overlap criterion is less restrictive at high latitudes and may affect absolute persistence estimates. We also report that sensitivity tests varying the minimum overlap fraction show no significant effect on blocking-frequency patterns for the tested Z500-Grad, Z500-SD, and PV-SD methods.

Why even use the gradient reversal method if by definition (without changing the latitudinal bands over which the gradient is calculated which could have been an option) it cannot detect blocks over Antarctica?

Thank you for this point. As shown in the sensitivity analysis discussed in previous comments, the near-absence of gradient-reversal detections over the Antarctic continent is not purely a methodological artefact that a simple parameter change would fix. Even at the most permissive settings tested (GHGS = 0 m $°lat^{-1}$, area filter = 0.5 × 10^6 km^2), the blocking signal over Antarctica tends to remain weak and largely confined to the circumpolar belt rather than the interior. This indicates that, under the published settings (GHGS = 10 m $°lat^{-1}$, area filter = 1.0 × 10^6 km^2), the gradient method detects a physically meaningful blocking result. Classical flow geometry blocking, defined by a reversal of the poleward geopotential height gradient, appears to be genuinely rare over the Antarctic continent.

The value of retaining this method is precisely in that contrast. The dataset includes four methods spanning two conceptual frameworks: flow-geometry reversal (Z500-Grad) and anomaly-based threshold exceedance. The divergence between these frameworks over Antarctica is itself a key scientific result of the dataset. It quantifies how much of what is labelled Antarctic blocking is a function of the detection concept rather than the circulation itself. A dataset that omitted the gradient method would lose this diagnostic dimension and would present only a partial picture of methodological uncertainty. Several studies of Southern Hemisphere blocking use gradient-based methods (e.g., Tibaldi et al. 1994; Pinheiro et al., 2019; Wang et al., 2024) and users of this dataset may wish to compare directly with those results. The inclusion of Z500-Grad makes that possible in a consistent framework.

We acknowledge the reviewer's implicit suggestion that modifying the latitudinal bands over which the gradient is calculated could be an option. Our sensitivity analysis tested Δlat (10° and 15°) and GHGS, which are the two most direct levers. Changing the latitudinal reference bands more fundamentally (e.g., redefining the polar cap boundary) would constitute a different method rather than a calibration of the existing one, and is beyond the scope of this dataset. We note it as a direction for future methodological work in Section 6.

We have mentioned this in Section 4.1 in the revised manuscript: *"These low Antarctic frequencies reflect the specific flow-geometry criterion targeted by Z500-Grad, namely reversal of the meridional Z500 gradient, rather than broad persistent anticyclonic anomalies. The contrast with the threshold-based methods is therefore an informative result of the dataset, indicating how diagnosed Antarctic blocking depends on the underlying detection concept. Z500-Grad is retained as a conservative flow-geometry baseline and to enable direct comparison with published studies using gradient-based diagnostics (e.g., Pinheiro et al., 2019)."*

Smaller points:

Introduction:

Line 1: Maybe half a sentence to define what is a block? General enough that it agrees with any detection method.

The sentence has been revised to more precisely reflect the AMS Glossary of Meteorology definition of a blocking high, distinguishing the persistent anticyclonic anomaly from the broader blocking situation it generates, and noting that blocking interrupts the eastward progression of weather systems rather than simply the westerly flow. The revised sentence ended up as: *"Atmospheric blocking refers to a persistent, quasi-stationary anticyclonic anomaly in the mid-to-upper troposphere that interrupts the normal eastward progression of weather systems and diverts the prevailing westerly flow for periods of days to weeks (Rex, 1950; Woollings et al., 2018; AMS, 2024)."*

In intro, when citing papers that discuss blocking in Antarctica, maybe briefly mention how they diagnose it; or dedicate another paragraph to this

A few clarifications were added in the second paragraph of the Introduction noting the diagnostic approaches used in cited studies: *"This dynamical environment challenges conventional approaches to blocking detection and has led to substantial methodological diversity in Antarctic and Southern Hemisphere studies. Existing approaches include gradient-reversal indices based on 500 hPa geopotential height (Pinheiro et al., 2019; Wang et al., 2024), a combined blocking–subtropical-ridge algorithm (Sousa et al., 2021), anomaly diagnostics based on 500 hPa geopotential height and potential vorticity (Pinheiro et al., 2019), and absolute-threshold methods implemented within objective feature-tracking frameworks (Ullrich et al., 2021). This diversity motivates a coordinated multi-method assessment of Antarctic blocking."*

maybe worth mentioning in the intro; PV framework also allows natural decomposition of conservative and non-conservative (cross-isentropic PV fluxes) processes affecting evolution of blocks

The following sentence was added in the paragraph introducing PV-based methods: *"The PV framework additionally supports decomposition of conservative versus diabatic contributions to block onset and maintenance, offering mechanistic insight beyond detection alone (Hoskins et al., 1985; Hauser et al., 2024; Hauser et al. 2026)."*

Line 86-88: "[...] its use for multi-method Antarctic blocking has not previously been integrated within a single dataset". Please rephrase

Rephrased to: "*TempestExtremes has not previously been used as a unified tracking framework for a multi-method Antarctic blocking dataset."*

Paragraph starting line 91 should have one sentence stating what you are doing, for those who have not read Pinheiro et al. 2019. I propose something like
"Earlier work by Pinheiro et al. (2019) provided a dataset of blocking characteristics derived from the ERA-Interim reanalysis ...."

The paragraph now opens with: *"Earlier work by Pinheiro et al. (2019) provided a dataset of Southern Hemisphere blocking characteristics derived from ERA-Interim from 1979 to 2018, based on multiple blocking diagnostics applied over the latitudinal band 25°S to 75°S within a unified threshold and tracking framework."*

Methods

Do the authors use the full spatial resolution of ERA5 (0.25°) or do they regrid to a coarser grid?

A clarifying sentence is added in Section 2.1*: "ERA5 data were used at the 0.25° × 0.25° horizontal resolution without regridding."*

(typesetting): cleaner maths typesetting; units and integral's d should be upright, indices as subscripts

Done.

Line 159: 3-day smoothing of anomalies: is this standard in blocking detection? How much does this choice affect the persistence requirements?

One sentence was added in Section 2.2.2: *"This smoothing follows Pinheiro et al. (2019) and reduces the amplitude of short-lived anomalies, preventing transient ridges to satisfy the subsequent 3-day persistence criterion."*

Line 173: "All methods share a common framework for spatial coherence, ..." please rephrase to a more precise statement. After each method, the the same area threshold is used to filter out too small events.

Rephrased to: *"After each detection step, the same minimum area threshold and temporal persistence criterion are applied to each method via TempestExtremes DetectBlobs and StitchBlobs, so that differences in blocking statistics across methods reflect the detection logic rather than post-processing inconsistencies."*

This might be me but I struggle to understant the difference between method 2 and 3. I see later in the results section that they do not detect the same amount of blocking so there must be a difference, but the description in the methods section makes me think they do the same thing.

I will summarise what I understood: At each grid point and time step, we can define a field x, its climatological value smoothed with five harmonics, x_clim, and the anomaly compared to this climatology x' = x-xclim, itself smoothed using a 3-day window. A smoothed daily standard deviation, sigma, is also computed and smoothed like the climatology. Finally a fixed threshold of 100m or 1.1 PVU is also defined, let's call it x. Method 2 defines a blocked grid point as respecting x > x_clim + max(1.5 sigma, x), while method 3 defines it as respecting x' > max(1.5 sigma, x), i.e. x-x_clim > max(1.5 sigma, x) which is the same inequality as the one of method 2. The only differences I can find are slight changes in smoothing (number of harmonics in lon, lat, time) for the climatology and sigma, and the fact that x in method 2 is not smoothed in time while x' is in method 3. What am I missing? Being able to see the code might make this a lot clearer.

Both reviewers identify the same issue, and we thank them for it. The naming is potentially confusing because both methods use climatological reference fields, but they differ fundamentally in what field the threshold is applied to. In Z500-Abs, the threshold is applied to the raw field: a grid point is blocked when the actual Z500 value exceeds the climatological mean plus a variability term. In Z500-SD, the threshold is applied to the anomaly field: a grid point is classified as blocked when its Z500 anomaly exceeds the larger of 100 m and 1.5 times the anomaly standard deviation calculated at that grid point and time of year. Thus, the method thresholds anomaly magnitude rather than the absolute Z500 field. Figure 1 (as Reviewer 1 notes) already illustrates this well visually, and we now make it equally clear in the text.

Beyond this conceptual difference, two further differences cause the methods to produce distinct results even when the algebraic forms appear equivalent.

Standard deviations from different fields. In Z500-Abs, σ is computed from the absolute Z500 field and therefore includes variance from the full seasonal cycle. In Z500-SD, σ is computed from the smoothed anomaly field and does not include seasonal cycle variance, making it systematically smaller. This means the effective detection threshold differs between the two methods even when the blocking condition is written in the same form.

Latitude-dependent amplitude scaling (in Z500-SD only). Before thresholding in Z500-SD, the anomaly is rescaled by a latitude-dependent factor following Pinheiro et al. (2019) to account for meridional differences in planetary wave amplitude. No such scaling is applied in Z500-Abs.

Explicit step-by-step summaries are added at the start of Sections 2.3.2 and 2.3.3. For Section 2.3.2: *"In this method, a grid point at time t is classified as blocked when the raw field value (Z500 or PV) equals or exceeds a spatially and temporally varying threshold, defined as the smoothed daily climatological mean plus a variability term. Formally:*

$$X(\lambda,\phi,t) \geq \overline{X}_{\mathrm{clim}(\lambda,\phi,d(t))} + \max\left[\delta_{\mathrm{min}}, 1.5\,\sigma_{\mathrm{clim}(\lambda,\phi,d(t))}\right]$$

where X denotes Z500 or PV, $\overline{X}$ is the smoothed daily climatological mean, σ_{clim} is the smoothed daily standard deviation of the absolute field, δ_{min} is the fixed minimum threshold value, and $d(t)$ is the day of year corresponding to time step t. The fixed minimum value is 100 m for Z500 and 1.1 PVU for PV."

A parallel formulation is added to Section 2.3.3 making explicit that the same inequality applies to the anomaly field rather than the absolute value.

Table 1:

row "threshold" type: use "or" instead of the confusing character "/" in "100 m / 1.1 PVU"

Threshold type for gradient method should read "fixed thresholds of north and south gradients" or something, currently it is the same text as the following row. suggestion: Use one large row spanning multiple columns to emphasize when all columns share the same entry especially for last two rows

Three changes are made to Table 1. First, all instances of "/" in "100 m / 1.1 PVU" are replaced with "or". Second, the threshold type entry for Z500-Grad is rewritten to read: "Fixed gradient reversal criteria applied to meridional Z500 gradients: GHGN < 0 m $°lat^{-1}$ and GHGS > 10 m $°lat^{-1}$." Third, the "Areal filter" and "Tracking method" rows are reformatted as single merged rows spanning all method columns, since the entry is identical for all methods: "Yes (minimum area 10^6 km^2)" and "Spatially contiguous blocked regions are linked across successive time steps when they overlap by at least 1 grid cell. Events are retained only if they persist for at least 3 consecutive days.", respectively.

Dataset description:

I appreciate the user friendliness of this section. The right amount of details for the users to start using the dataset.

Good also to provide the summary statistics (especially mean length)

We thank the reviewer for the positive assessment of Section 3. Regarding the summary statistics, we note that Section 3 describes the dataset structure and contents rather than presenting climatological results. Summary statistics such as mean event duration and annual event counts are presented in the Results section (Section 4).

3.2: if i understand correctly, in the case of splitting, the child with the most overlap with the parent retains the ID? and similarly for merging, the child retains the ID of the parent with the most overlap? how about the other one(s)? is there a dataset of all such merging and splitting? it would be very valuable for more advanced feature-based analysis.

StichBlobs from TempestExtremes retains the same ID for all blocking regions that split from a parent blocking region or that merge into a single child blocking region. This means StitchBlobs only considers whether regions satisfy the minimum area of overlap. In our data, we used one grid cell of overlap between two consecutive times. A dedicated merge and split catalogue is not provided in the current version of the dataset. The blocking_id fields do contain the full temporal sequence of IDs at each grid point, from which splitting and merging events can be reconstructed by tracking ID discontinuities across time steps. We agree that a pre-computed catalogue of such events would be valuable for feature-based analysis and note this as a priority direction for future dataset extensions.

Results

Figure 4: "Regions shown in white correspond to latitudes where blocking detection is not applied": not quite true, it also corresponds to regions with very low values of blocking frequency, which makes it hard to see at a glance where the reversal index stops applying and is in my opinion very important to visualise. Hatching maybe? or a light gray shading? or no white in the colormap of blocking frequency?

We agree. The figure is revised so that regions where detection is not applied (north of 25°S for all methods; south of 75°S for Z500-Grad) are indicated with light gray shading.

Line 416; "The PV-based methods exhibit weaker spatial contrasts than the Z500-based method": Isn't that purely a matter of choice of thresholds in intensity and area? The authors could compare with the midlatitude literature: what is the expected difference in blocking detection frequency between PV and z500 methods, with these thresholds?

Both contribute. Our sensitivity analysis shows that lowering the PV minimum threshold from 1.1 to 0.6 PVU substantially increases detection frequency, confirming a threshold effect. However, the qualitative contrast between PV and Z500 methods persists across all threshold configurations tested, pointing to a dynamical component as well. Vertically integrated PV anomalies associated with blocking are spatially more compact than Z500 anomalies because PV responds to sharp upper-tropospheric gradients near the tropopause, and vertical integration retains only anomalies coherent through a substantial layer depth (Schwierz et al., 2004; Hauser et al., 2024; Hauser et al., 2026). PV methods are therefore inherently more selective than Z500 methods, as also shown for the Southern Hemisphere by Pinheiro et al. (2019) under comparable threshold settings.

We have therefore revised the full paragraph in Section 4.1 describing the PV-based methods to clarify that the observed contrast reflects both threshold choices and the intrinsically more selective dynamical character of vertically integrated PV diagnostics.

Line 434: "In June-July-August (JJA), the spatial distribution of blocking over high latitudes changes markedly. The Z500-Grad method shows localized enhancements in blocking over the Amundsen and Bellingshausen Sea sectors, where the mid-latitude jet weakens and becomes less zonally continuous over the South Pacific" this last statement deserves a citation

Citations are added: *"...where the mid-latitude jet weakens and becomes less zonally continuous over the South Pacific (Trenberth and Mo, 1985; Renwick and Revell, 1999)."*

Line 483. In this paragraph, are the authors referring to the results of the previous subsections or to what one can see on figure 6? If the latter, it is important that the figure be referenced early on.

We clarified that the opening discussion in Section 4.3 synthesizes the annual and seasonal results presented in Sections 4.1 and 4.2. The corresponding figure is now referenced in the subsequent paragraphs, where the dominance-ratio analysis is introduced.

Line 497: "The persistence of some winter blocking signals at high latitudes..." I do not understand this sentence. At this point the authors have not presented any persistence results, and even when they do, they do not separate by season. Is this section out of order?

We agree. The original sentence referred to seasonal persistence, whereas Section 4.3 discusses annual and seasonal blocking-frequency patterns. Persistence is presented separately in the following subsection as an annual climatology and is not analysed by season. We have therefore removed the sentence. The subsequent sentence was revised to introduce the potential dynamical mechanisms directly.

Line 505: I appreciate this, this is a good idea.

Thank you.

FIgure 6: the computation of sigma_metric and sigma_variable should be in the main text and more detailed. The description in the figure caption is confusingly written.

Agreed. The definitions of σ_metric, σ_variable, and the dominance ratio are now provided in Section 4.3 before the corresponding figure is introduced, with the order of operations clarified. The figure caption has been simplified.

Line 557: "consistent with the smoother structure of vertically integrated PV fields.". Is that so? This smoothness should be quantified somehow, it is not obvious it should be the case. PV on single levels is very spatially variable, and it is not obvious to me that it should be completely smoothed out by vertical integration.

Single-level PV on isentropic surfaces can vary strongly in space because of sharp dynamical features such as tropopause folds, PV streamers, and cutoffs. Vertical integration over 150–500 hPa does not remove this variability completely, but it smooths out features limited to a narrow vertical layer. As a result, the integrated field mainly captures anomalies that extend through a deeper part of the upper troposphere. The statement has been revised to make this clearer. Instead of saying *"consistent with the smoother structure of vertically integrated PV fields"* we now wrote: *"consistent with the reduced small-scale spatial variance of vertically integrated PV compared with single-level PV fields. Vertical integration over 150–500 hPa averages out PV anomalies confined to narrow layers, such as tropopause folds and PV streamers, while retaining anomalies that are coherent through a substantial depth of the upper troposphere and lower stratosphere."*

Code availability:

As a reviewer, I would have liked to be able to inspect the code to make sure the methods applied really correspond to their descriptions, mostly to try and catch potential errors that we all make. For complex methods such as these, the code might provide another angle to the explanation that can clarify some points like the order of operations (smoothing, climatology computing, time-averaging...). I think just pointing out one of the tools used is not enough. Since the authors made a lot of choices not everyone might agree with with respect to threshold values, it would also allow the users re-create a similar dataset with their own choices of thresholds, but otherwise applying the same methodology.

We agree. The preprocessing and detection scripts are now archived in the same Zenodo repository as the dataset. The repository includes the full workflow from ERA5 preprocessing through TempestExtremes detection and tracking, with parameter choices documented at each step. Users can modify threshold values within the provided scripts to reproduce the dataset with alternative choices. Section 5 ("Data and Code Availability") is updated to state explicitly that the analysis code is archived alongside the dataset in the Zenodo repository.