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.