An integrated diagnostic framework for evaluating Brewer–Dobson circulation transport efficiency in a warming climate using CMIP6 and Aura MLS

Jackson Hian-Wui Chang[1,2,*], Ahmad Fairudz bin Jamaluddin[3], Sharifah Faridah binti Syed Mahbar[3], Yong Jie Wong[4], Maggie Chel-Gee Ooi[5], Juneng Liew[6], Fuei Pien Chee[7], Jedol Dayou[8,9], Justin Sentian[7], Raul R. Cordero[10], Neng-Huei Lin[2]

[1]Preparatory Center for Science and Technology, University Malaysia Sabah (UMS), Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia
[2]Department of Atmospheric Sciences, National Central University (NCU), Taoyuan, 32001, Taiwan
[3]National Climate Centre, Malaysian Meteorological Department, Petaling Jaya, Selangor, Malaysia
[4]Department of Bioenvironmental Design, Faculty of Bioenvironmental Sciences, Kyoto University of Advanced Science, Kyoto, 606-8501, Japan
[5]Centre of Tropical Climate Change System, Institute of Climate Change, National University of Malaysia (UKM), Bangi, 43600, Malaysia
[6]Department of Earth Sciences and Environment, Faculty of Sciences and Technology, National University of Malaysia (UKM), Bangi, 43600, Malaysia
[7]Faculty of Science and Technology, University Malaysia Sabah (UMS), Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia
[8]Energy, Vibration and Sound Research Laboratory (e-VIBS), Faculty of Science and Technology, University Malaysia Sabah (UMS), Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia
[9]Centre for Instrumentation and Science Services (CISS), University Malaysia Sabah (UMS), Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia
[10]Department of Physics, Universidad de Santiago de Chile, Santiago, CA, USA

Correspondence to: Jackson Hian-Wui Chang (jacksonchw@ums.edu.my)

General comments

The authors aim to investigate the response of stratospheric transport to climate change by developing an integrated diagnostic framework that combines CMIP6 mean AoA, TEM diagnostics, zonal-wind tendency terms, and Aura MLS water-vapour observations. However,, in its present form, the manuscript contains several methodological and interpretative issues that affect its central conclusions. In particular, I have concerns regarding the interpretation of the idealised 1pctCO2 experiment; the construction of the multi-model means is not sufficiently documented; period-mean climatologies are repeatedly interpreted as statistically significant trends; and the MLS correlation analysis does not provide an independent observational constraint on greenhouse-gas-driven circulation changes. Furthermore, the assessment of statistical significance does not appear to account adequately for temporal autocorrelation, smoothing, lag selection, or multiple testing.I therefore consider that the main claims concerning enhanced lower-stratospheric "transport efficiency" and its greenhouse-gas-driven mechanism are not yet sufficiently supported by the evidence presented. Substantial reanalysis, clearer quantification of uncertainties, and a more cautious interpretation of the results are required before the manuscript can be considered for publication.

Authors' reply

We thank the reviewer for this comprehensive assessment. We agree that the original manuscript combined several diagnostics under an interpretation that was too broad and, in several places, stronger than warranted by the analyses. In response, we substantially revised both the methodology and the central framing of the manuscript. The revised paper no longer treats age of air, residual circulation, wave-related momentum tendencies, and MLS tracer propagation as interchangeable measures of a single "transport-efficiency" quantity. Instead, these diagnostics are used to distinguish complementary aspects of the system: circulation strength, circulation geometry, dynamical forcing, and observed tracer propagation. This revised framing is now stated explicitly in the Introduction and Abstract.

First, the 1pctCO2 analysis has been fully reformulated in terms of experiment year rather than historical calendar time. We now define early, middle, and late 30-year stages as experiment years 1–30, 56–85, and 111–140, respectively, and analyse EP-flux-divergence, orographic gravity-wave, and non-orographic gravity-wave tendencies separately as zonal-momentum tendencies. We further added alternative-window sensitivity tests. Importantly, because these diagnostics come from the idealised 1pctCO2 experiment, the revised manuscript now interprets their relationship to the historical and SSP3-7.0 circulation changes as dynamical consistency rather than direct causal attribution.

Second, we removed the previous reliance on an insufficiently documented grand multimodel mean. Models, configurations, and realizations are now retained separately throughout the analysis, and ensemble summaries are calculated only within each model/configuration using equal member weights. The manuscript explicitly states that no grand multimodel mean is used as the primary estimate because model climatologies, ensemble sizes, and diagnostic availability differ. We also introduced matched-member AoA–w* comparisons so that younger air and enhanced residual upwelling are evaluated within the same simulations rather than across different model populations.

Third, the historical and SSP3-7.0 diagnostics are now described consistently as within-model period-mean differences rather than temporal trends. The revised analysis uses fixed 30-year baseline and target periods and reports member-bootstrap uncertainty, leave-one-member-out robustness, and temporal-block sensitivity separately. This distinction is also made explicitly in the Supplementary Tables.

Fourth, we replaced the former qualitative statement that circulation geometry remains largely unchanged with an objective diagnosis of the poleward boundaries of tropical upwelling. The boundary is defined from the persistent w*=0 transition, and its uncertainty is quantified by member bootstrap. Sensitivity to 3°, 5°, and 7° persistence criteria is also tested. The revised result is that the lower-stratospheric upwelling region narrows rather than remaining geometrically unchanged; by late century the diagnosed width at 70 hPa decreases by approximately 6.5° in both CanESM5 and UKESM1-0-LL.

Fifth, the MLS analysis was substantially reworked. The primary analysis now uses unsmoothed, non-detrended daily H_2O anomalies on individual native potential-temperature levels, with an unconstrained 0–365 d lag search and no imposed 45- or

90-d minimum lag. Candidate peaks are identified using predefined prominence, separation, recurrence, and ambiguity criteria. Temporal dependence is handled using 1000 moving-block bootstrap realizations in which the complete lag-correlation curve is reconstructed and peaks are redetected.

To address uncertainty in the choice of bootstrap block length, every initially stable pathway is additionally subjected to a four-method autocorrelation-based block-length audit using the first ACF zero crossing, e-folding time, truncated integrated autocorrelation time, and the original conservative formulation. A pathway is retained as ROBUST_STABLE only if it remains a unique stable peak under all four definitions. The Supplementary Methods provide the mathematical definitions and stability criteria in full.

We also added sensitivity tests for 15- and 31-d smoothing, broad versus individual potential-temperature levels, equivalent latitude, pressure coordinates, detrending, and extension beyond 2021. Importantly, these sensitivity treatments are not substituted for the primary unsmoothed analysis. The Supplementary Figures now show representative full lag-correlation curves, smoothing sensitivity, and block-length sensitivity.

Finally, we substantially moderated the interpretation of the observational component. The revised manuscript explicitly states that MLS H_2O propagation is not a direct validation of greenhouse-driven BDC acceleration. The MLS analysis is now presented as an observational characterization of tracer-propagation pathways that reflect residual advection together with mixing and tracer-specific processes. Similarly, the paper no longer presents the combined diagnostics as proof of a single greenhouse-driven "transport efficiency" mechanism. The revised conclusion instead emphasizes that circulation strength, circulation geometry, wave forcing, and tracer propagation provide complementary but non-equivalent information about the stratospheric transport system.

We believe these changes directly address the reviewer's general concerns by replacing the original broad framework with a more quantitatively constrained and cautious interpretation, supported by explicit uncertainty analysis and a substantially revised observational methodology.

Because the revision involved substantial restructuring of the manuscript, extensive reanalysis, replacement of several figures and tables, and addition of new supplementary analyses, we do not reproduce every revised passage verbatim in this response. Instead, for each comment we describe the changes made and identify the corresponding revised sections, figures, tables, and supplementary material so that the revisions can be located and evaluated directly.

Key revised locations: Abstract; Introduction; Sects. 2.1–2.5; Sects. 3.1–3.4; Figs. 1–7; Tables 1–5; Figs. S1–S6; Tables S1–S8; Supplementary Methods.

Specific comments

1. Interpretation and temporal coverage of the 1pctCO2 experiment

Regarding the description on page 6, Table 3 lists the simulation period as 1850–1999, whereas Fig. 6 and the associated discussion refer to the period 1975–2014. The authors should clarify how the years 2000–2014, which fall outside the stated simulation period, could be included in the analysis as currently described.More fundamentally, 1pctCO2 is an idealised experiment in which atmospheric CO2 concentrations increase by 1% yr−1 from a pre-industrial control state, while other forcing agents do not follow their historical evolution. Consequently, the time coordinate of this experiment should not be interpreted as actual historical calendar time. The 1pctCO2 results should instead be presented in terms of experiment year or CO2 level, for example years 1–50, 51–100, and 101–140, or conditions corresponding approximately to 1 × CO2, 2 × CO2, and 4 × CO2.

Authors' reply

We thank the reviewer for identifying this important issue. We agree that the original manuscript incorrectly treated the 1pctCO2 time coordinate as if it represented historical calendar years, which led to an internally inconsistent description of the experiment period. We have therefore completely revised this part of the analysis.

In the revised manuscript, the 1pctCO2 experiment is now analysed exclusively in experiment years, with no historical calendar interpretation. We define three 30-year stages: early (years 1–30), middle (years 56–85), and late (years 111–140), and calculate middle-minus-early and late-minus-early changes.

We also added two alternative stage definitions to test whether the results depend on the exact choice of experiment windows: Sensitivity A uses years 1–30, 51–80, and 101–130, while Sensitivity B uses years 11–40, 61–90, and 111–140. The same retained member cohort is used for all alternatives, and robustness is assessed from sign preservation and fractional magnitude differences relative to the primary definition.

The revised analysis is now presented directly in Fig. 5 and Table 4. Figure 5 shows the vertical distribution of late-minus-early wave-related zonal-momentum tendency changes, where late corresponds to experiment years 111–140 and early to years 1–30. The figure therefore uses only experiment-year stages and no longer implies any correspondence with historical calendar dates. Table 4 reports both middle-minus-early and late-minus-early changes and explicitly defines early, middle, and late as experiment years 1–30, 56–85, and 111–140, respectively.

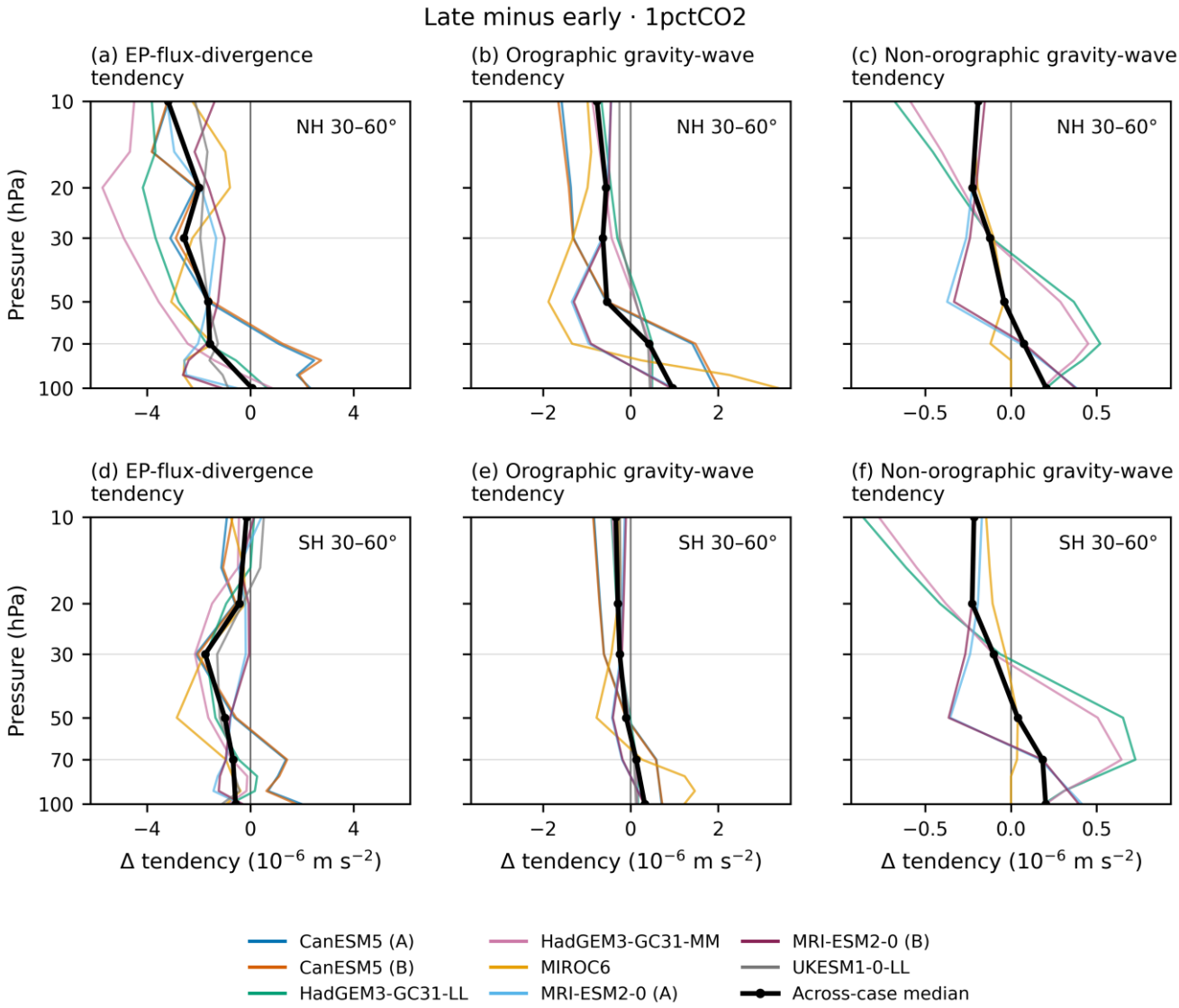


Figure 5. Vertical distribution of wave-related zonal-momentum tendency changes under idealised CO_2 forcing. Rows show NH 30–60° and SH 30–60° area-weighted bands; columns show EP-flux divergence, orographic gravity waves and non-orographic gravity waves. Values are late (experiment years 111–140) minus early (years 1–30) in 1pctCO2. The middle minus early stage is years 56–85 and is reported in Table 4 and Table S5. Negative values denote more westward tendencies. Thin coloured lines show within-case member means at the pressure levels already present in the validated output, joined without smoothing. The thick black line is the equally weighted median across available model/configuration cases at six shared levels (100, 70, 50, 30, 20 and 10 hPa), joined by straight segments. There are eight cases for EP-flux divergence and orographic drag and five for non-orographic drag; missing components are excluded, never treated as zero. Member coverage varies from one to four (Table S5).

Table 4. Experiment-stage changes in wave-related zonal-momentum tendencies under 1pctCO2 forcing. Values in the readable table are in 10^{-6} m s^{-2}. Medians are computed across within-case member means, giving each available model/configuration case equal weight. Negative changes denote a shift toward a more westward zonal-momentum tendency under the verified eastward-positive CMIP6 standard names. Values are experiment-stage period differences, not temporal

trends. Early = years 1–30, middle = 56–85 and late = 111–140. Model/configuration cases are counted separately for sign agreement but are not interpreted as statistically independent models. CI counts give eligible n ≥ 3 cohorts whose pointwise 95% member-bootstrap interval excludes zero / number tested, separately for each stage. They describe uncertainty conditional on retained members, not temporal or model-structural uncertainty.

Component	Band	hPa	Cases	Middle median	Middle signs	CI excludes 0	Late median	Late signs	CI excludes 0
EP-flux divergence	NH 30–60°	70	8	−0.721	8/8 negative	2/2	−1.570	6/8 negative; 2/8 positive	2/2
EP-flux divergence	NH 30–60°	30	8	−1.369	8/8 negative	2/2	−2.569	8/8 negative	2/2
EP-flux divergence	NH 30–60°	10	8	−2.111	8/8 negative	2/2	−3.191	8/8 negative	2/2
EP-flux divergence	SH 30–60°	70	8	−0.274	6/8 negative; 2/8 positive	2/2	−0.674	6/8 negative; 2/8 positive	2/2
EP-flux divergence	SH 30–60°	30	8	−0.798	8/8 negative	2/2	−1.752	8/8 negative	2/2
EP-flux divergence	SH 30–60°	10	8	−0.123	5/8 negative; 3/8 positive	1/2	−0.154	4/8 negative; 4/8 positive	1/2
Orographic gravity waves	NH 30–60°	70	8	0.160	3/8 negative; 5/8 positive	2/2	0.418	3/8 negative; 5/8 positive	2/2
Orographic gravity waves	NH 30–60°	30	8	−0.260	8/8 negative	2/2	−0.636	8/8 negative	2/2
Orographic gravity waves	NH 30–60°	10	8	−0.286	8/8 negative	2/2	−0.772	8/8 negative	2/2
Orographic gravity waves	SH 30–60°	70	8	0.033	3/8 negative; 5/8 positive	2/2	0.126	2/8 negative; 6/8 positive	2/2
Orographic gravity waves	SH 30–60°	30	8	−0.119	8/8 negative	2/2	−0.251	8/8 negative	2/2
Orographic gravity waves	SH 30–60°	10	8	−0.160	8/8 negative	2/2	−0.339	8/8 negative	2/2
Non-orographic gravity waves	NH 30–60°	70	5	−0.020	3/5 negative; 2/5 positive	1/1	0.075	1/5 negative; 4/5 positive	1/1
Non-orographic gravity waves	NH 30–60°	30	5	−0.103	5/5 negative	1/1	−0.123	5/5 negative	1/1
Non-orographic gravity waves	NH 30–60°	10	5	−0.096	5/5 negative	1/1	−0.193	5/5 negative	1/1
Non-orographic gravity waves	SH 30–60°	70	5	0.033	5/5 positive	1/1	0.184	5/5 positive	1/1
Non-orographic gravity waves	SH 30–60°	30	5	−0.106	5/5 negative	1/1	−0.102	5/5 negative	1/1
Non-orographic gravity waves	SH 30–60°	10	5	−0.105	5/5 negative	1/1	−0.216	5/5 negative	1/1

The Supplementary material has been revised consistently. For example, Fig. S3 now presents middle-minus-early and late-minus-early changes using the same experiment-year definitions and explicitly states that these refer to experiment years 1–30, 56–85, and 111–140 of 1pctCO2.

We have also revised the interpretation throughout the manuscript to avoid direct equivalence between the idealised 1pctCO2 response and the historical/SSP3-7.0 simulations. The revised Methods now state that the relationship between the idealised forcing diagnostics and the historical/scenario circulation changes is interpreted as dynamical consistency rather than direct causal attribution.

This revision removes the previous calendar-year inconsistency and aligns the analysis with the physical meaning of the 1pctCO2 experiment.

Key revised locations: Sect. 2.4; Sect. 3.3; Fig. 5; Table 4; Fig. S3; Tables S5–S6.

2. Period means should not be interpreted as trends

Regarding the descriptions on pages 10–12, Figs. 1–4 mainly present climatological means for selected periods. Nevertheless, the manuscript repeatedly refers to these results as “trends”. Comparisons among several period-mean contour plots are not sufficient to establish temporal trends or statistical significance. The statement that the turnaround latitude remains unchanged also requires quantitative support. The turnaround latitude should be objectively diagnosed for each model and period, and its changes and associated uncertainties should be reported.

Authors’ reply

We thank the reviewer for this important clarification. We agree that the original manuscript incorrectly used the term “trend” for differences between selected 30-year climatological periods. We have revised this terminology throughout the

manuscript. The historical and SSP3-7.0 diagnostics are now consistently defined and interpreted as within-model period-mean differences rather than temporal trends. The revised Methods define the historical comparison as 1985–2014 minus 1871–1900, and the SSP3-7.0 mid- and late-century comparisons as 2041–2070 minus 1985–2014 and 2071–2100 minus 1985–2014, respectively.

We also substantially revised the treatment of circulation geometry. Rather than retaining the original qualitative statement that the turnaround latitude remains approximately unchanged, we now diagnose the poleward boundary of tropical upwelling objectively from each member's 30-year mean w* field. For each hemisphere, the boundary is defined as the first persistent transition from positive to negative w* moving poleward from the tropical upwelling maximum. The zero crossing is linearly interpolated between the bracketing latitude points, and a transition is retained only when the sign change persists for at least 5° on both sides and is supported by at least two grid points; ambiguous or insufficiently persistent crossings are excluded.

Uncertainty in the diagnosed geometry is now quantified using 5000 member-bootstrap resamples for cohorts with $n \geq 3$, with paired-member resampling used for upwelling-width changes so that the NH and SH boundaries remain physically linked. We additionally perform leave-one-member-out tests and repeat the geometry diagnosis using 3°, 5°, and 7° persistence criteria to assess sensitivity to the boundary definition.

The results of this new quantitative diagnosis are presented in Fig. 4 and Table 3. Figure 4 shows the vertical profiles of changes in tropical residual velocity together with the objectively diagnosed NH and SH upwelling-boundary displacements. The shaded bands represent pointwise 95% member-bootstrap intervals for the ensemble mean where $n \geq 3$. Table 3 provides the corresponding numerical boundary shifts and changes in total tropical-upwelling width, together with their 95% confidence intervals.

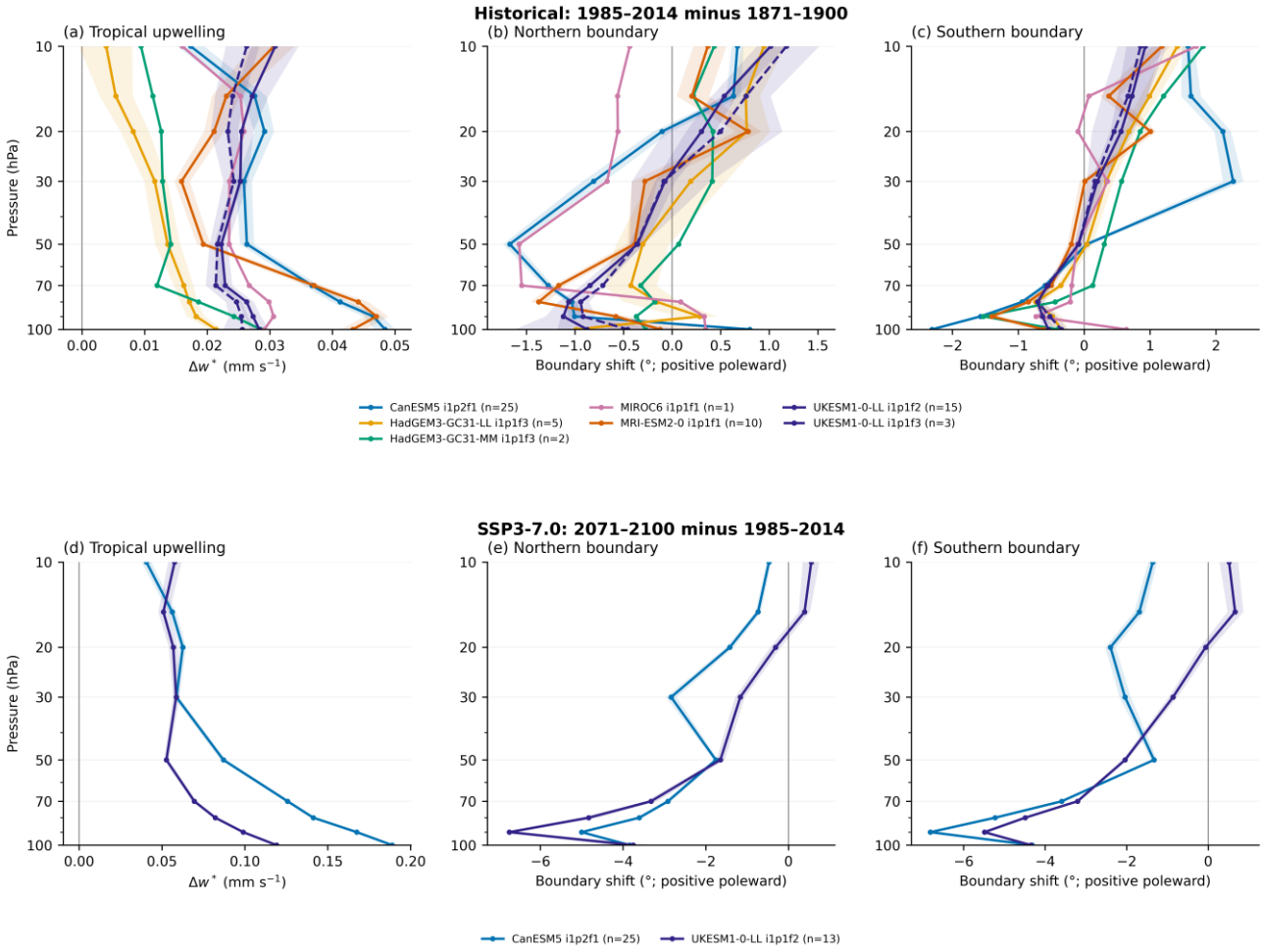


Figure 4. Changes in tropical residual vertical velocity (20°S–20°N) and NH/SH upwelling boundaries. Top: 1985–2014 minus 1871–1900. Bottom: SSP3-7.0 2071–2100 minus historical 1985–2014. Lines show equal-member mean period changes; shaded bands are pointwise 95% percentile member-bootstrap intervals for the ensemble mean. No band is estimated for n < 3. Geometry is calculated from each member's 30-year mean field with 5° persistence. Positive boundary shifts are poleward in both hemispheres.

Table 3. Upwelling boundaries and width. Positive NH/SH shifts mean poleward; negative width changes mean narrowing. Width = NH latitude minus SH latitude. Intervals describe ensemble-mean uncertainty, not member spread; no interval for n < 3. Width intervals resample paired member width changes.

Model/configuration	Pressure (hPa)	Members	NH shift (deg) [95% CI]	SH shift (deg) [95% CI]	Width change (deg) [95% CI]
Historical: 1985–2014 minus 1871–1900					
CanESM5 i1p2f1	70	25	-1.28 [-1.33, -1.23]	-0.59 [-0.65, -0.54]	-1.87 [-1.94, -1.80]
CanESM5 i1p2f1	30	25	-0.81 [-0.92, -0.70]	+2.26 [+2.11, +2.41]	+1.45 [+1.29, +1.61]
CanESM5 i1p2f1	10	25	+0.67 [+0.61, +0.73]	+1.58 [+1.48, +1.67]	+2.25 [+2.14, +2.35]
HadGEM3-GC31-LL i1p1f3	70	5	-0.43 [-0.72, -0.16]	-0.35 [-0.42, -0.28]	-0.78 [-1.07, -0.50]
HadGEM3-GC31-LL i1p1f3	30	5	+0.19 [-0.24, +0.55]	+0.34 [+0.22, +0.44]	+0.53 [+0.15, +0.88]

HadGEM3-GC31-LL i1p1f3	10	5	+0.94 [+0.72, +1.16]	+1.41 [+1.26, +1.56]	+2.36 [+2.17, +2.57]
HadGEM3-GC31-MM i1p1f3	70	2	-0.32 [not estimated]	+0.13 [not estimated]	-0.19 [not estimated]
HadGEM3-GC31-MM i1p1f3	30	2	+0.41 [not estimated]	+0.57 [not estimated]	+0.98 [not estimated]
HadGEM3-GC31-MM i1p1f3	10	2	+0.43 [not estimated]	+1.81 [not estimated]	+2.24 [not estimated]
MIROC6 i1p1f1	70	1	-1.55 [not estimated]	-0.19 [not estimated]	-1.74 [not estimated]
MIROC6 i1p1f1	30	1	-0.67 [not estimated]	+0.35 [not estimated]	-0.32 [not estimated]
MIROC6 i1p1f1	10	1	-0.44 [not estimated]	+1.70 [not estimated]	+1.27 [not estimated]
MRI-ESM2-0 i1p1f1	70	10	-1.17 [-1.24, -1.10]	-0.50 [-0.54, -0.44]	-1.67 [-1.75, -1.57]
MRI-ESM2-0 i1p1f1	30	10	-0.28 [-0.42, -0.15]	+0.01 [-0.07, +0.09]	-0.27 [-0.38, -0.17]
MRI-ESM2-0 i1p1f1	10	10	+0.37 [+0.22, +0.56]	+1.18 [+1.04, +1.31]	+1.55 [+1.32, +1.79]
UKESM1-0-LL i1p1f2	70	15	-0.85 [-0.94, -0.76]	-0.55 [-0.59, -0.50]	-1.39 [-1.50, -1.28]
UKESM1-0-LL i1p1f2	30	15	-0.07 [-0.18, +0.03]	+0.20 [+0.10, +0.29]	+0.13 [-0.04, +0.28]
UKESM1-0-LL i1p1f2	10	15	+1.01 [+0.87, +1.17]	+0.93 [+0.64, +1.19]	+1.94 [+1.66, +2.22]
UKESM1-0-LL i1p1f3	70	3	-0.72 [-0.72, -0.70]	-0.56 [-0.60, -0.49]	-1.28 [-1.33, -1.19]
UKESM1-0-LL i1p1f3	30	3	-0.08 [-0.42, +0.39]	+0.16 [+0.04, +0.31]	+0.08 [-0.19, +0.53]
UKESM1-0-LL i1p1f3	10	3	+1.18 [+0.98, +1.51]	+0.86 [+0.51, +1.25]	+2.03 [+1.85, +2.23]
SSP3-7.0: 2071–2100 minus 1985–2014					
Model/configuration	**Pressure (hPa)**	**Members**	**NH shift (deg) [95% CI]**	**SH shift (deg) [95% CI]**	**Width change (deg) [95% CI]**
CanESM5 i1p2f1	70	25	-2.91 [-2.98, -2.85]	-3.59 [-3.65, -3.54]	-6.51 [-6.59, -6.41]
CanESM5 i1p2f1	30	25	-2.84 [-2.96, -2.71]	-2.04 [-2.17, -1.90]	-4.88 [-5.05, -4.70]
CanESM5 i1p2f1	10	25	-0.47 [-0.53, -0.42]	-1.36 [-1.47, -1.25]	-1.83 [-1.96, -1.70]
UKESM1-0-LL i1p1f2	70	13	-3.32 [-3.46, -3.18]	-3.20 [-3.26, -3.15]	-6.53 [-6.67, -6.36]
UKESM1-0-LL i1p1f2	30	13	-1.17 [-1.29, -1.06]	-0.86 [-0.97, -0.76]	-2.03 [-2.15, -1.90]
UKESM1-0-LL i1p1f2	10	13	+0.55 [+0.34, +0.74]	+0.51 [+0.28, +0.73]	+1.06 [+0.88, +1.23]

Importantly, this reanalysis changes the original interpretation. The lower-stratospheric upwelling region does not remain unchanged. During the historical comparison, most models show equatorward displacement and narrowing of the diagnosed upwelling region at 70 hPa, while higher levels generally exhibit a different response. Under SSP3-7.0, the lower-stratospheric contraction becomes more pronounced. By late century, the diagnosed 70 hPa upwelling width decreases by 6.51° in CanESM5 and 6.53° in UKESM1-0-LL. At 30 hPa, the corresponding narrowing is 4.88° and 2.03°, respectively, whereas the 10 hPa response is model dependent.

The persistence-threshold sensitivity test further shows that repeating the calculation with 3°, 5°, and 7° criteria does not change the valid-member classifications or diagnosed boundary locations across the historical, mid-century, and late-century comparisons (Table S4).

These revisions directly address both aspects of the reviewer's concern: the period comparisons are no longer described as temporal trends, and the previous qualitative statement about an unchanged turnaround latitude has been replaced by an objective and uncertainty-quantified diagnosis of upwelling-boundary position and width.

Key revised locations: Sects. 2.1–2.3; Sect. 3.2; Fig. 4; Table 3; Figs. S1–S2; Tables S1, S3–S4.

3. Definition of "transport efficiency"

Section 2.3 states that transport efficiency is not measured directly but is inferred collectively from AoA, TEM circulation diagnostics, wave forcing, and MLS lag correlations. As currently formulated, "transport efficiency" therefore appears to be an interpretative concept rather than a clearly defined diagnostic quantity or quantitative metric. The authors should either provide a clear, quantitative, and testable definition of transport efficiency or moderate the associated claims throughout the manuscript.

Authors' reply

We thank the reviewer for this important point. We agree that the original manuscript used the term "transport efficiency" too broadly and risked implying the existence of a single quantitative metric that was not actually defined or calculated.

Rather than introducing a new composite metric retrospectively, we have revised the manuscript to moderate this terminology and to separate the individual diagnostics according to the specific physical quantity each represents.

In the revised manuscript, AoA is interpreted as an integrated measure of transit-time characteristics, w* as a measure of residual-circulation strength, objectively diagnosed w*=0 boundaries as a measure of circulation geometry, wave-related zonal-momentum tendencies as dynamical context, and MLS H_2O lags as tracer-propagation timescales. The revised Introduction now states explicitly that these quantities are not treated as interchangeable measures of a single transport-efficiency metric, but are instead used to distinguish changes in circulation strength, geometry, dynamical forcing, and observed tracer propagation.

We also revised the interpretation of the matched AoA–w* analysis. The manuscript now states that, although younger tropical air and stronger residual upwelling co-occur within the same simulations, this joint response is not interpreted as a direct quantitative measure of transport efficiency, because AoA also reflects mixing and the distribution of transit times.

The same change was made for the observational component. MLS H_2O lag estimates are now described as effective tracer-propagation timescales rather than direct measures of circulation efficiency, parcel transit time, or w*. The Methods explicitly note that H_2O propagation reflects residual advection together with mixing, tracer gradients, and tracer-specific processes.

Accordingly, the central interpretation of the paper has been changed from a single “enhanced transport efficiency” framework to a structured transport response in which different diagnostics provide complementary but non-equivalent information. The revised results show, for example, that stronger tropical residual upwelling can coexist with a narrower lower-stratospheric upwelling region, while wave-related momentum tendencies and MLS propagation exhibit distinct vertical and pathway dependence. This is now reflected in the Abstract and throughout the Discussion.

We therefore chose the reviewer’s second suggested option: rather than defining an artificial composite transport-efficiency metric, we have substantially moderated the associated claims and now report each diagnostic according to the physical property it directly represents.

Key revised locations: Abstract; final paragraph of Introduction; Sect. 2.2; Sect. 2.5; Sects. 3.1–3.4; Sects. 4.1–4.4.

4. Relationship to the method of Flury et al. (2013)

The MLS analysis differs substantially from the method used by Flury et al. (2013). That study correlated one-year H2O time series between selected, physically adjacent pressure levels and between the Equator and approximately 40° latitude at 100 hPa. In contrast, the present manuscript averages data over broad 30° latitude bands and broad isentropic layers, uses an approximately 18-year anomaly record, and selects correlation peaks only after imposing minimum lags of 45 or 90 days.These methodological modifications may be reasonable, but they require independent justification and validation. In particular, selecting the first or largest correlation peak only after imposing a prescribed minimum lag creates a risk that the inferred result depends strongly on the selected threshold. Short-lag maxima should not simply be classified as spurious because they are inconsistent with an expected transport timescale.At a minimum, the authors should:

--show the unconstrained lag-correlation functions and all relevant correlation peaks;

-specify the complete lag-search interval and peak-selection algorithm;

-provide uncertainty intervals for the selected lags;

-test the sensitivity of the results to different smoothing windows and latitude-band definitions;

-compare results based on broad-layer averages with those obtained at individual isentropic levels; and

-validate the inferred lags using an independent tracer, model experiment, trajectory calculation, or age-spectrum diagnostic.

Authors’ reply

We thank the reviewer for this detailed and constructive comment. We agree that the original MLS analysis departed substantially from Flury et al. (2013) and that the use of prescribed minimum lags, broad averaging, and limited uncertainty treatment made the inferred propagation timescales too dependent on methodological choices. We therefore substantially redesigned the MLS analysis rather than attempting to reproduce the Flury et al. methodology exactly. The revised approach uses Flury et al. as conceptual motivation for lag-based tracer propagation, but the method is now defined independently and documented in full.

Most importantly, the previous 45- and 90-d minimum-lag thresholds have been removed. The revised analysis searches the complete predefined interval of 0–365 d, and no short-lag maximum is rejected merely because it is shorter than an expected transport timescale. The primary analysis uses unsmoothed, deseasonalized daily H_2O anomalies for 2004–2021 on individual native potential-temperature levels. Candidate positive-correlation peaks are identified directly from the full lag curve using a minimum prominence of 0.03 and a minimum separation of 15 d. Peaks within 15 d are grouped into the same peak family; a family must recur in at least 60% of bootstrap realizations to be considered stable, and competing recurrent families are treated as ambiguous when the secondary recurrence-weighted prominence score reaches at least 80% of the leading score. Boundary maxima are retained diagnostically but cannot be used to justify selection of a weaker interior peak.

Uncertainty in the selected lag is now estimated using a moving-block bootstrap rather than a nominal Pearson significance test. For every pathway, 1000 bootstrap realizations are generated; the complete 0–365 d lag-correlation function is reconstructed in every realization and the peak-selection algorithm is rerun. The reported 95% intervals are therefore conditional on repeated selection of the observed peak family. In addition, every initially stable pathway is subjected to a four-method autocorrelation-based block-length audit using the first ACF zero crossing, ACF e-folding time, truncated integrated autocorrelation time, and the original conservative formulation. A pathway is retained as ROBUST_STABLE only when all four definitions independently return a unique stable peak.

The final outcome of this revised procedure is presented in Fig. 7 and Table 5. Of the 365 tested meridional and vertical pathways, 46 initially yielded a stable single peak, but only 39 remained robust after the complete four-method block-length audit. Figure 7 shows only these final ROBUST_STABLE pathways and therefore provides a spatial overview of where a defensible propagation lag can actually be assigned. Table 5 reports the corresponding bootstrap median lags, conditional 95% intervals, peak correlations, and peak-recurrence fractions. All pathways listed in Table 5 remained stable under all four block-length definitions.

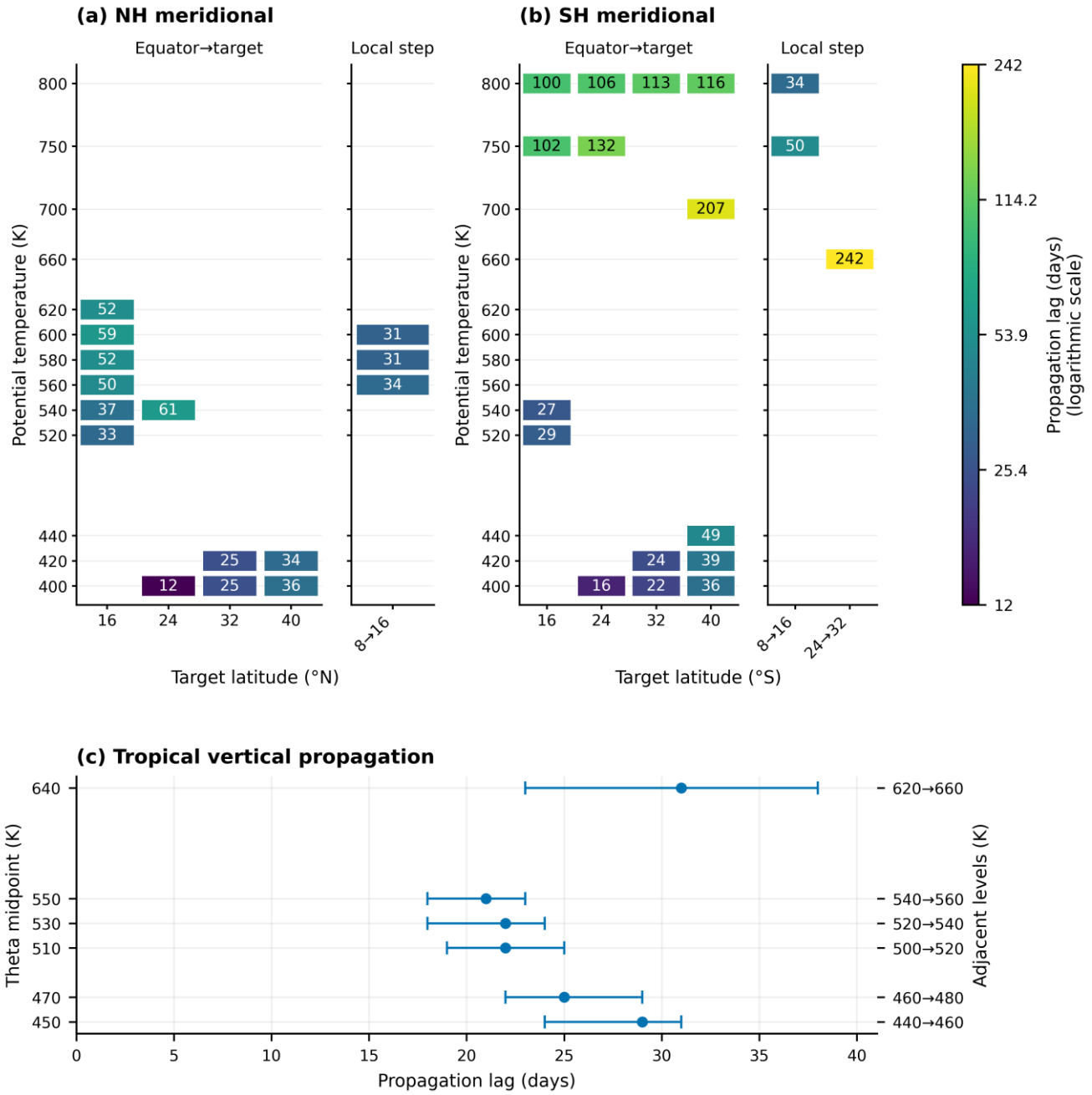


Figure 7. Robust Aura MLS H_2O propagation pathways from the full uncertainty audit. Only final ROBUST_STABLE pathways are shown: initially stable unconstrained lag peaks that remained STABLE_SINGLE_PEAK under all four predefined block-length definitions (first ACF zero crossing, e-folding, truncated integrated ACF, and original/current method), subject to the fixed data-quality criteria. Primary observations are unsmoothed Aura MLS H_2O anomalies for 2004–2021 on individual ThetaZM levels. (a,b) NH and SH latitude–theta matrices show current-method bootstrap median lags as numbers and colors. Equator-to-

target and local-step pathways occupy separate sections; local ticks identify source→target absolute latitude in the indicated hemisphere. Blank positions are not shown as robust; non-robust and untested positions are distinguished in Figure S4. Theta positions preserve physical spacing, and labeled levels are the union with robust meridional support. Both hemispheres share one unclipped logarithmic lag color normalization. (c) Equatorial adjacent-level pathways show current-method bootstrap medians and horizontal 95% intervals at theta midpoints; the right axis identifies each source→target pair.

Table 5. Selected robust Aura MLS H_2O propagation pathways during 2004–2021. Bootstrap median lags and conditional 95% confidence intervals are reported using the original block-length method. All listed pathways remained stable under all four block-length definitions. Analyses use unsmoothed, deseasonalized H_2O anomalies. Meridional pathways connect latitude bins at the same potential temperature; vertical pathways connect the indicated source and target levels at the equator. Positive lag indicates that the source anomaly leads the target anomaly. Confidence intervals are conditional bootstrap 2.5th–97.5th percentiles for the unambiguously selected observed peak family. Peak recurrence denotes the bootstrap recurrence fraction of that family, expressed as a percentage; it is not a confidence level. These lags characterize tracer propagation and should not be interpreted as exact transit times or TEM residual velocities. Complete pathway results are provided in Table S7.

Latitude: source → target	Potential temperature (K)	Median lag (95% CI), days	Peak *r*	Peak recurrence (%)
Equator-to-target meridional propagation				
0° → 40°S	400	36 (25.0–47.0)	0.716	98.9
0° → 40°N	400	36 (31.0–45.0)	0.700	84.2
0° → 40°S	420	39 (32.0–51.1)	0.763	99.6
0° → 40°N	420	34 (28.0–39.0)	0.784	100.0
0° → 40°S	440	49 (34.0–53.0)	0.763	84.0
0° → 16°S	520	29 (20.0–34.0)	0.820	91.1
0° → 16°N	520	33 (27.0–45.9)	0.792	76.5
0° → 16°S	540	27 (20.0–34.0)	0.799	85.9
0° → 24°N	540	61 (44.0–72.0)	0.616	86.2
0° → 16°N	560	50 (36.0–65.0)	0.633	81.2
0° → 16°N	580	52 (37.0–65.0)	0.620	74.3
0° → 16°N	600	59 (45.0–74.0)	0.595	77.8
0° → 16°N	620	52 (38.0–63.0)	0.605	79.4
0° → 40°S	700	207 (197.0–222.0)	0.560	61.6
0° → 24°S	750	132 (127.0–150.2)	0.604	63.4
0° → 16°S	750	102 (91.0–116.0)	0.689	70.3
0° → 40°S	800	116 (101.0–129.0)	0.432	86.6
0° → 16°S	800	100 (86.0–108.0)	0.609	86.1
Local meridional propagation				
8°N → 16°N	560	34 (24.0–41.0)	0.750	92.4
8°N → 16°N	580	31 (24.0–46.0)	0.753	87.1
8°N → 16°N	600	31 (22.0–46.0)	0.749	83.0
24°S → 32°S	660	242 (239.0–254.0)	0.317	95.3
8°S → 16°S	750	50 (37.0–65.0)	0.782	73.3
8°S → 16°S	800	34 (24.0–50.0)	0.808	89.2
Adjacent-level vertical propagation				
0° → 0°	440 → 460	29 (24.0–31.0)	0.924	95.5
0° → 0°	460 → 480	25 (22.0–29.0)	0.948	90.7
0° → 0°	500 → 520	22 (19.0–25.0)	0.960	99.7
0° → 0°	520 → 540	22 (18.0–24.0)	0.954	99.9
0° → 0°	540 → 560	21 (18.0–23.0)	0.962	76.4

These revised results also demonstrate why a single characteristic BDC lag is not appropriate. Near 400–440 K, robust equator-to-extratropical pathways occur in both hemispheres, whereas the higher-level pathways become much more

selective. In the NH, robust pathways between about 520 and 620 K are concentrated mainly in the subtropics, while substantially longer meridional propagation times occur in the SH between about 700 and 800 K. Six robust adjacent-level equatorial pathways are also retained between 440 and 660 K, with typical lags of approximately 20–30 d.

The complete lag-correlation structure is retained in Fig. S5, which shows representative unconstrained lag curves including detector-listed peaks and boundary candidates. The full inventory of all 365 tested pathways is provided in Table S7, including cases classified as no stable peak, data-limited, or block-length sensitive. This avoids presenting only successful pathways and makes the selection process transparent.

We also performed the requested sensitivity tests. The primary analysis deliberately does not use smoothing; 15- and 31-d running means are instead treated as sensitivity experiments. These tests show that smoothing can create apparently stable peaks that are absent in the unsmoothed data, supporting the use of unsmoothed daily anomalies as the primary analysis. Broad 400–500, 500–600, and 600–800 K layer averages were also tested explicitly against individual potential-temperature levels and were generally poor representations of their constituent levels. The primary analysis therefore uses individual native potential-temperature levels, while broad-layer averages are retained only as sensitivity tests.

The spatial representation was also revised. Rather than relying on broad 30° latitude-band averages as the primary lag geometry, the analysis now uses predefined equator-to-target meridional pathways, local meridional steps between successive sampled latitude bins, and adjacent-level vertical pathways on the native MLS grid. Equivalent-latitude and pressure-coordinate analyses are treated separately as sensitivity checks.

Regarding the reviewer's final request for independent validation using another tracer, model experiment, trajectory calculation, or age-spectrum diagnostic, we did not add such a validation because the revised study does not contain a sufficiently independent tracer or trajectory product that could provide a defensible validation of the MLS lag estimates. Rather than claiming validation that was not performed, we have moderated the interpretation accordingly. The revised manuscript now treats the inferred lags as H_2O tracer-propagation timescales rather than exact parcel transit times, TEM residual velocities, or a direct observational validation of greenhouse-forced BDC changes.

The revised analysis therefore addresses the underlying concern by removing the prescribed lag thresholds, exposing the full lag structure, defining an explicit and reproducible peak-selection algorithm, quantifying lag uncertainty with autocorrelation-aware resampling, testing smoothing and spatial averaging choices, and restricting the main quantitative interpretation to the robust pathways shown in Fig. 7 and Table 5. We have also deliberately refrained from claiming independent validation where none was performed.

Key revised locations: Sect. 2.5; Sect. 3.4; Fig. 7; Table 5; Figs. S4–S6; Tables S7–S8; Supplementary Methods.

5. Statistical significance of the lag correlations

Table 5 reports $p < 0.001$ even for correlation coefficients as small as 0.14–0.18. The manuscript appears to apply a conventional Pearson-correlation significance test. Such a test assumes that the samples are approximately independent. This assumption is clearly violated by the daily stratospheric tracer records, the application of a 31-point smoothing filter, overlapping lagged samples, and strong seasonal and interannual persistence.The effective number of independent observations is therefore likely to be substantially smaller than the nominal sample size. The statistical significance should be reassessed using a method that accounts for temporal autocorrelation, such as an effective-sample-size correction, a moving-block bootstrap, or autocorrelation-preserving surrogate time series. The effects of searching across multiple lags and selecting the maximum correlation should also be taken into account.

Authors' reply

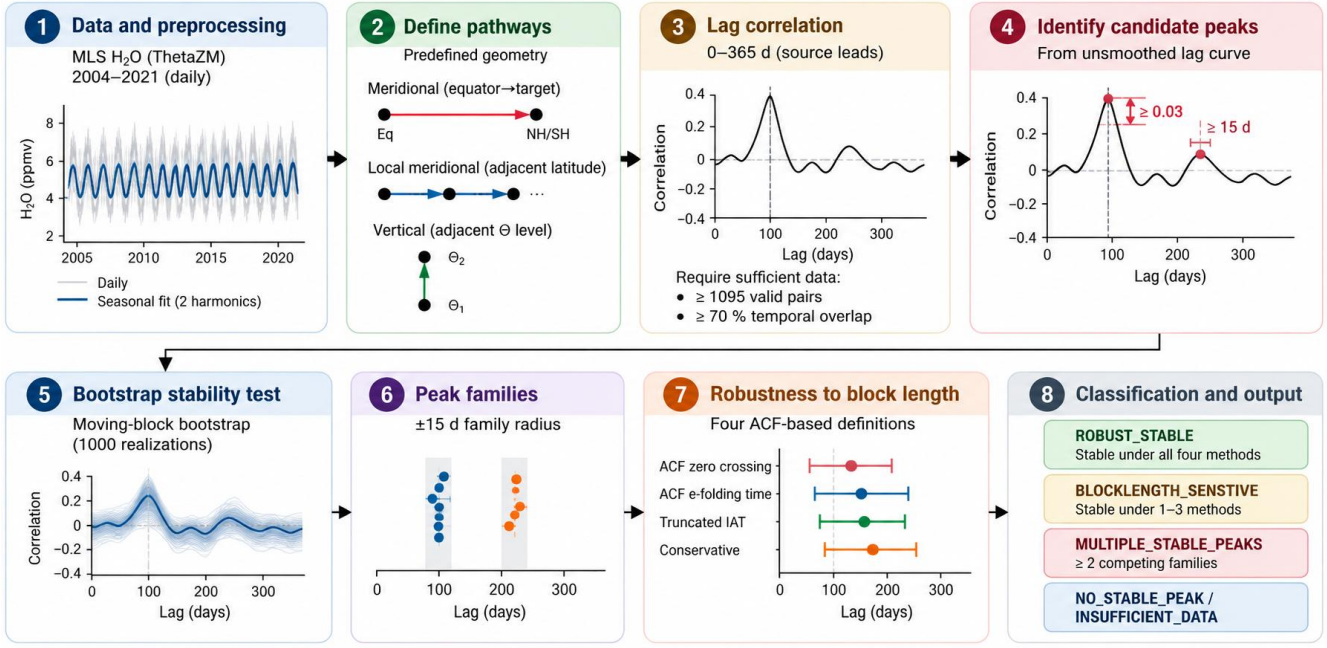
We thank the reviewer for identifying this important statistical issue. We agree that conventional Pearson p-values are not appropriate for the daily MLS tracer series because the observations are strongly serially dependent and the lag search itself

introduces a multiple-selection problem. We have therefore removed the original Pearson significance testing from the revised analysis and replaced it with an autocorrelation-aware moving-block bootstrap framework.

The revised primary analysis uses unsmoothed daily H_2O anomalies and evaluates the complete predefined lag interval from 0 to 365 d. For each pathway, candidate peaks are identified from the observed lag-correlation curve using fixed prominence, separation, recurrence, and ambiguity criteria. Temporal dependence is then handled by resampling contiguous lag-moment blocks rather than individual days. For each pathway, 1000 bootstrap realizations are generated, the entire 0–365 d lag-correlation curve is reconstructed in every realization, and all candidate peaks are redetected using the same algorithm applied to the observed data.

This procedure directly addresses the reviewer's concern about both serial dependence and searching across multiple lags. The uncertainty interval is not based on a single selected correlation coefficient or on an assumption of independent daily samples. Instead, it is conditional on repeated selection of the same observed peak family across bootstrap realizations. The revised manuscript now states explicitly that the reported interval is not a conventional Pearson-correlation confidence interval.

We further added a four-method block-length robustness audit because the effective decorrelation timescale is itself uncertain for persistent stratospheric H_2O anomalies. The four definitions use the first ACF zero crossing, ACF e-folding time, truncated integrated autocorrelation time, and the original conservative lag-search-aware formulation. The complete bootstrap peak-selection procedure is repeated independently under each block-length definition, and a pathway is retained as ROBUST_STABLE only if all four methods independently return a unique stable peak.



Response Figure R1 (not added in text). Workflow used to derive robust Aura MLS H_2O tracer-propagation lags. Daily 2004–2021 ThetaZM H_2O anomalies are deseasonalized and evaluated along predefined cumulative meridional, local meridional, and adjacent-level vertical pathways. Lag correlations are calculated over 0–365 d subject to minimum data-coverage requirements, after which candidate positive-correlation peaks are identified using predefined prominence and separation criteria. Peak stability is assessed with 1000 moving-block bootstrap realizations, recurring peaks are grouped into lag families, and initially stable pathways are further tested using four autocorrelation-based block-length definitions. Only

pathways retaining a unique stable peak under all four treatments are classified as ROBUST_STABLE and retained for quantitative propagation-lag interpretation.

The revised analysis also imposes additional safeguards against inflated apparent significance from dense daily sampling. Each lag requires at least 1095 valid source–target pairs and at least 70% of the possible temporal overlap, and a stable inference is not accepted when the chosen block length yields fewer than five effective blocks. The supplementary robustness procedure also requires at least 95% of the requested bootstrap realizations to produce a valid lag-correlation curve.
Accordingly, the original $p<0.001$ entries have been removed from the revised quantitative interpretation. The revised tables instead report the selected lag, conditional 95% bootstrap interval, peak correlation, and recurrence fraction only for pathways that survive the full robustness audit. Table S7 preserves the complete set of tested pathways, including those that fail to yield a defensible unique lag.

These changes replace nominal Pearson significance with a procedure that explicitly accounts for temporal autocorrelation, repeated lag searching, peak redetection, and uncertainty in the dependence scale itself.

Key revised locations: Sect. 2.5; Sect. 3.4; Table 5; Tables S7–S8; Figs. S5–S6; Supplementary Methods.

6. Figures and support for the conclusions
Several figures contain very small axis labels, tick labels, colour bars, and panel annotations, making quantitative evaluation difficult. The period-mean figures should be supplemented or replaced by change maps. Consistent colour scales should be used, and stippling or hatching should be added to indicate statistical significance and inter-model agreement.The abstract and conclusions should also be revised. Statements such as “robust multi-model evidence”, “direct observational evidence”, “observational constraint”, and “greenhouse-driven changes are expressed primarily” are currently stronger than warranted by the analyses presented.

Authors’ reply
We thank the reviewer for this important comment. We agree that the original figures were difficult to evaluate quantitatively and that several statements in the Abstract and Conclusions were stronger than warranted by the analyses. We have therefore revised both the graphical presentation and the interpretation of the main conclusions.

The main CMIP6 figures were redesigned to emphasize period-mean changes rather than climatological states. Figure 1 now shows within-model changes in AoA and w* for the historical, SSP3-7.0 mid-century, and SSP3-7.0 late-century comparisons using explicitly matched member subsets. Figure 2 presents model-specific historical w* change maps on each model’s native grid, and all panels use a common zero-centred colour scale to facilitate direct comparison.

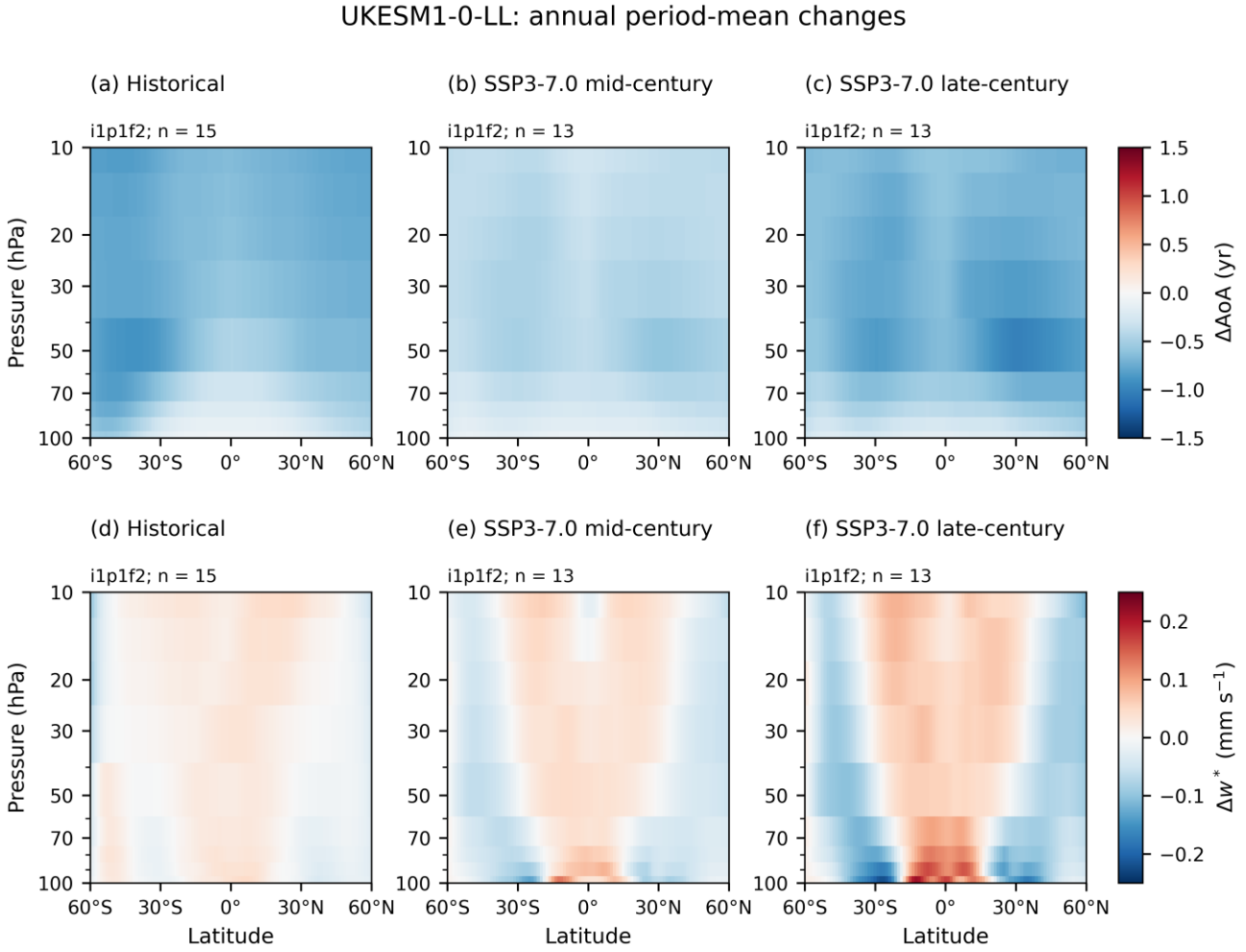


Figure 1. Within-model changes in stratospheric mean age of air (AoA; a–c, years) and residual vertical velocity (w*; d–f, mm s−1) in UKESM1-0-LL, configuration i1p1f2. Columns show annual 1985–2014 minus 1871–1900, SSP3-7.0 2041–2070 minus 1985–2014, and SSP3-7.0 2071–2100 minus 1985–2014 period means. The identical AoA/w* member subset is used within each column (n = 15, 13, 13, respectively); members receive equal weight. Negative ΔAoA indicates younger air; positive Δw* indicates a more upward residual velocity, which denotes enhanced upwelling where the baseline velocity is upward.

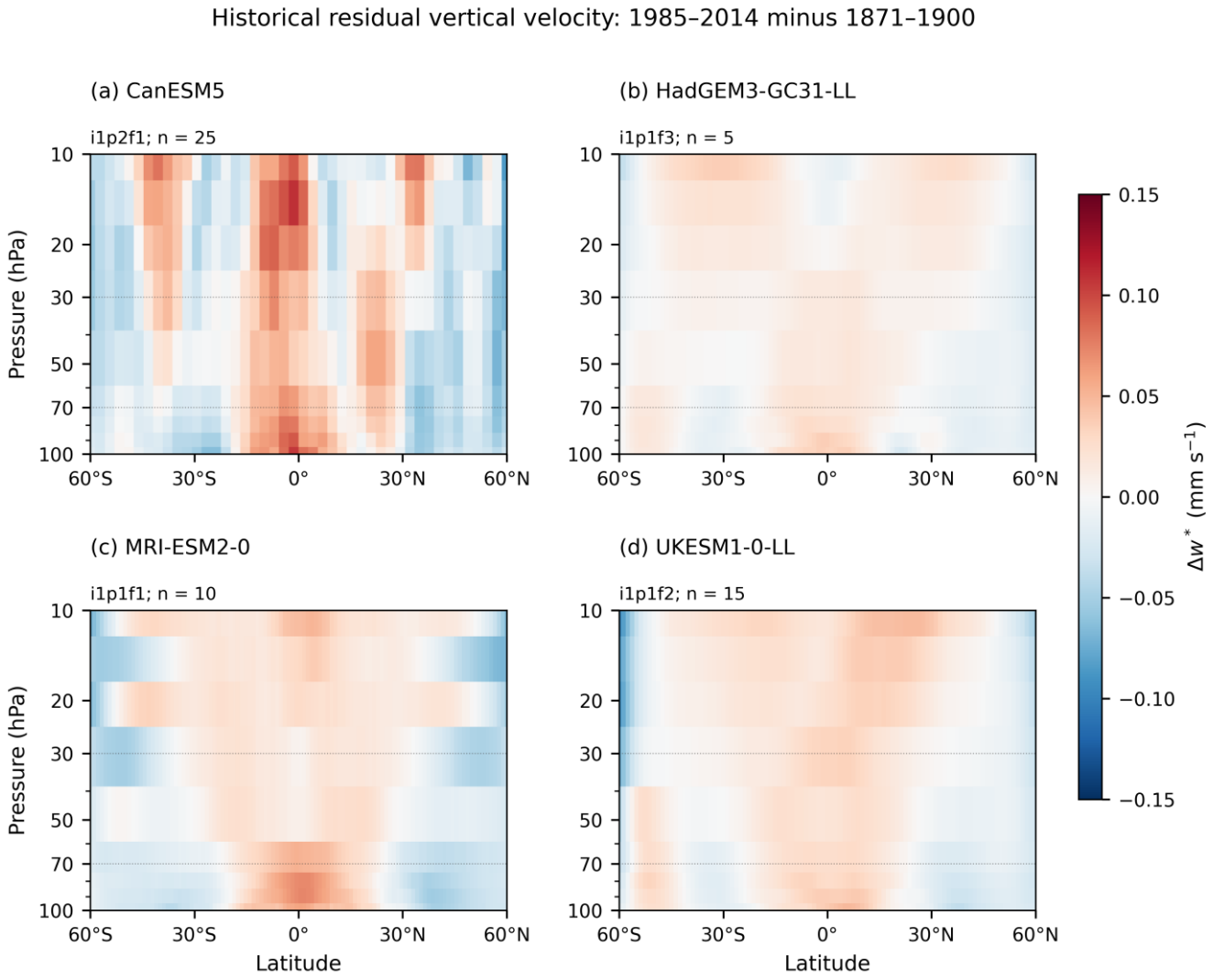


Figure 2. Model dependence of historical residual vertical velocity changes: (a) CanESM5 i1p2f1 (n = 25); (b) HadGEM3-GC31-LL i1p1f3 (n = 5); (c) MRI-ESM2-0 i1p1f1 (n = 10); (d) UKESM1-0-LL i1p1f2 (n = 15). Panels show equal-member annual period-mean changes, 1985–2014 minus 1871–1900, in mm s^{-1}, on each model's native grid. All panels share one zero-centred colour scale. Positive/negative changes denote more upward/downward residual velocity; a positive change in a descending region can represent weaker descent rather than stronger circulation. Dotted lines mark 70, 30 and 10 hPa. Grey denotes missing or incomplete-cohort cells.

For the circulation-strength and geometry diagnostics, Fig. 4 now presents period-mean changes directly, with pointwise 95% percentile member-bootstrap intervals shown as shaded bands where the ensemble size permits estimation. The associated Table 3 reports the diagnosed NH and SH boundary shifts and changes in tropical-upwelling width with 95% confidence intervals. The Supplementary Figures likewise show the corresponding historical, mid-century, and late-century changes with uncertainty bands rather than relying on visual comparison of climatological fields.

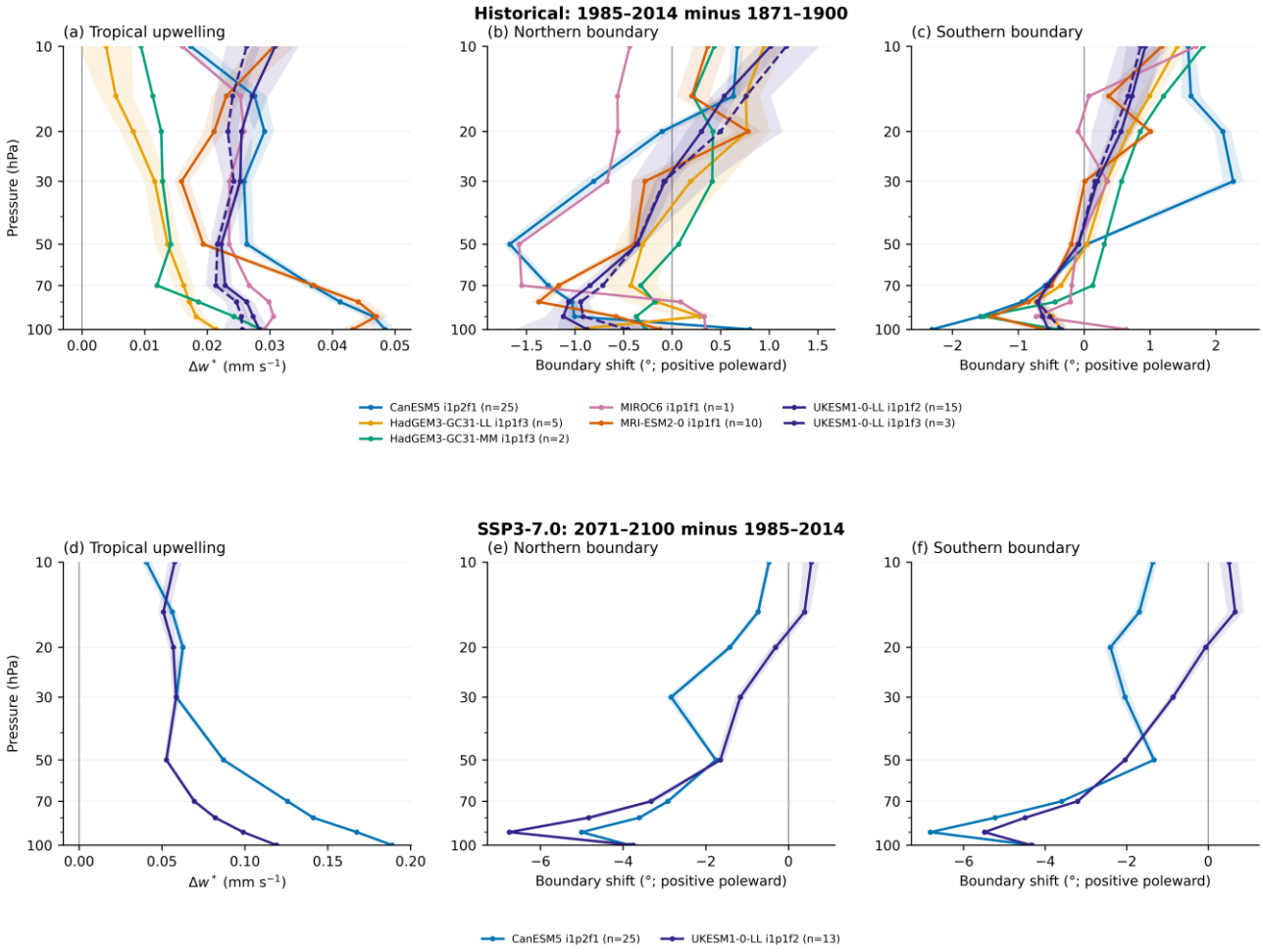


Figure 4. Changes in tropical residual vertical velocity (20°S–20°N) and NH/SH upwelling boundaries. Top: 1985–2014 minus 1871–1900. Bottom: SSP3-7.0 2071–2100 minus historical 1985–2014. Lines show equal-member mean period changes; shaded bands are pointwise 95% percentile member-bootstrap intervals for the ensemble mean. No band is estimated for n < 3. Geometry is calculated from each member's 30-year mean field with 5° persistence. Positive boundary shifts are poleward in both hemispheres.

We did not add stippling or hatching to indicate a single multimodel significance field, because the revised analysis deliberately does not construct a grand multimodel mean or treat different model/configuration cases as statistically interchangeable. Instead, uncertainty is represented at the within-model/configuration level using member-bootstrap intervals and leave-one-member-out robustness tests. For the wave-related momentum tendencies, model/configuration cases are retained separately, and cross-case agreement is summarized through sign consistency and case medians rather than a pooled significance mask. This approach is more consistent with the revised treatment of unequal model coverage and ensemble size than applying a common stippling criterion across heterogeneous model populations.

We also substantially revised the Abstract and Discussion/Conclusions to moderate the strength of the claims. The revised Abstract no longer presents the results as "robust multi-model evidence" for a single enhanced transport-efficiency mechanism or as "direct observational evidence" of greenhouse-driven acceleration. Instead, it states that the diagnostics distinguish changes in AoA, residual-circulation strength and geometry, wave-related momentum forcing, and observed tracer propagation, and concludes that these represent distinct but complementary aspects of the BDC response rather than a single uniform acceleration metric. Similarly, the MLS contribution is now explicitly described as an observational characterization of tracer propagation rather than a direct validation of the simulated greenhouse-forced circulation response.

The revised conclusions likewise emphasize that younger AoA, stronger tropical residual ascent, a narrower lower-stratospheric upwelling region, vertically differentiated wave forcing, and pathway-dependent MLS tracer propagation describe different components of the transport system rather than mutually redundant evidence for one mechanism.

These revisions improve both the quantitative readability of the figures and the correspondence between the evidence presented and the strength of the conclusions.

Key revised locations: Abstract; Sects. 3.1–3.4; Figs. 1–5; Table 3; Table 4; Figs. S1–S3; Discussion and Conclusions.

Technical corrections

7. Use the standard experiment name "1pctCO2" consistently. The term "1ptCO2" in Table 3 is incorrect.

Authors' reply

We thank the reviewer for noting this typographical inconsistency. We have corrected the experiment name to the standard CMIP6 notation "1pctCO2" throughout the revised manuscript, including the relevant table, figure captions, Methods, Results, and Discussion. The revised manuscript now uses "1pctCO2" consistently when referring to the idealised 1% yr^{-1} CO_2-increase experiment.

8. The caption of Fig. 6 gives the period as 1850–1990, whereas the corresponding period is defined elsewhere as 1850–1900.

Authors' reply

We thank the reviewer for noting this inconsistency. The original figure and caption have been removed as part of the substantial restructuring of the manuscript, so the incorrect 1850–1990 period no longer appears in the revised version.

In the revised manuscript, Fig. 6 now presents Aura MLS H_2O time–potential temperature cross sections for 2004–2024, and its caption is fully unrelated to the historical CMIP6 period comparison. The historical CMIP6 comparison is now defined consistently throughout as 1985–2014 minus 1871–1900. The revised Fig. 1-4 explicitly states that historical period comparison as 1985–2014 minus 1871–1900.

9. Revise the legend and caption of Fig. 6. The EPFD, NOGW, and OGW terms represent zonal-momentum tendencies rather than "total upwelling".

Authors' reply

We thank the reviewer for this correction. We agree that the original wording was physically inaccurate. EP-flux-divergence (EPFD), orographic gravity-wave (OGW), and non-orographic gravity-wave (NOGW) terms are now described consistently as zonal-momentum tendencies, not as components of "total upwelling" or as direct measures of residual vertical velocity.

This point has been revised throughout the Methods, Results, figure captions, and discussion of the idealised 1pctCO2 analysis. The revised Methods explicitly state that utendepfd, utendogw, and utendnogw are analysed separately as zonal-momentum tendencies and are not combined into a total forcing. Negative changes are interpreted as shifts toward a more westward tendency under the CMIP6 eastward-positive convention.

The original wave-forcing figure has also been replaced/restructured in the revised manuscript. The current wave-forcing figure and associated supplementary figure now use legends and captions referring explicitly to EP-flux divergence, orographic gravity-wave tendency, and non-orographic gravity-wave tendency. The Supplementary Fig. S3 caption likewise states that the rows show EP-flux divergence, orographic drag, and non-orographic drag, with negative values denoting more westward zonal-momentum tendencies.

We have also moderated the interpretation accordingly: these terms are used to provide dynamical context for the circulation response and are not treated as a direct local measure of tropical upwelling.

Key revised locations: Sect. 2.4; Sect. 3.3; revised wave-forcing figure and caption; Fig. S3.

10. Carr et al. (1995) and Clark et al. (1998) refer to the earlier UARS MLS instrument. The Aura MLS Level-3 product should be supported by appropriate references to the Aura MLS product description and data-quality documentation.

Authors' reply

Thank you for pointing this out. We agree that Carr et al. (1995) and Clark et al. (1998) refer to the earlier UARS MLS instrument and are therefore not appropriate references for the Aura MLS product used in this study. We have removed these references and replaced them with citations specific to Aura MLS, including the instrument description by Waters et al. (2006), the Aura MLS H_2O validation study by Lambert et al. (2007), and the Version 5 Level-2/Level-3 data-quality documentation by Livesey et al. (2022). The corresponding description in Sect. 2.5 has been revised accordingly. The manuscript now states that Aura MLS provides vertically resolved global measurements of stratospheric water vapour and other trace species suitable for diagnosing large-scale stratospheric transport, supported by these Aura-specific references.

11. Replace "pre-industry" with "pre-industrial" throughout the manuscript.

Authors' reply

We thank the reviewer for noting this terminology issue. We have replaced "pre-industry" with the standard term "pre-industrial" throughout the revised manuscript and associated supplementary material.