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 present a combination of Age-ofAir and Transformed Eulerian Mean (TEM) analyses based on CMIP6 simulations and supporting water vapor observations by the Aura Microwave Limb Sounder (MLS) to investigate the effect of increased greenhouse gases on the Brewer-DobsonCirculation (BDC) and compare more recent developements of the BDC to historical trends, as wellas propose a vision of the future based on the 1ptCO2 (annual increase of atmospheric CO2 by 1%) scenario. In its current form, the manuscript has several significant issues, which would demand additional attention to varying degrees.

Authors' reply

We thank the reviewer for this overall assessment. We agree that the original manuscript required substantial revision in both its scientific framing and methodological presentation. In response, we have restructured the study so that the different diagnostics are no longer presented as components of a single "transport-efficiency" framework, but instead as complementary measures of distinct aspects of the stratospheric transport system: AoA for integrated transit characteristics, w* for residual-circulation strength, objectively diagnosed upwelling boundaries for circulation geometry, wave-related zonal-momentum tendencies for dynamical context, and Aura MLS H_2O lags for observed tracer propagation.

Several major methodological changes were also introduced. The historical and SSP3-7.0 analyses are now expressed as within-model period-mean differences rather than temporal trends, and model/configuration identities are retained rather than combined into a single grand multimodel mean. The 1pctCO2 analysis is now defined entirely in experiment years and interpreted as idealised dynamical context rather than as a historical or future calendar projection. The MLS analysis was substantially redesigned to use an unconstrained 0–365 d lag search, unsmoothed daily anomalies, moving-block bootstrap uncertainty, and a four-method autocorrelation-based robustness audit.

The interpretation of the results has also been moderated. The revised manuscript no longer argues for a single uniform BDC acceleration or direct observational validation of greenhouse-driven circulation changes. Instead, it emphasizes that stronger tropical residual upwelling can coexist with a narrower lower-stratospheric upwelling region, that wave-related momentum forcing is vertically differentiated, and that MLS tracer propagation is strongly pathway dependent.

We have also revised the figures, tables, terminology, and discussion throughout the manuscript to improve readability and to ensure that the conclusions remain proportional to the evidence presented. The detailed revisions are addressed individually in our responses to the reviewer's specific and major comments below.

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; Discussion and Conclusions; revised main and supplementary figures/tables.

Regarding its structural and linguistic form, the manuscript suffers from a few spelling mistakes and bad punctuation, which do make understanding the potentially novel ideas more difficult. Such issues should have been addressed prior to submission more rigorously. Parts of the manuscript, in particular the section„ Disscussion & Conclusions", and in part the section „Results", appear to unneccessarily repeat certain passages, without providing new information. From their discussion, the significance to repeat them in this form does not become clear to me. The usage of „codes" used to describe certain simulation configurations or variables should be revised, as these are likely opaque to anyone not directly familiar with the simulation results. For example, „utendepfd" does not help a first-time reader to understand what it is referring to, similarly to „i1p1f1", which the authors have provided the means to interpret as „initialization method 1, physics 1 (whatever that means) and forcings 1", but do not make clear what any of these configurations actually mean. Figures in the manuscript

suffer from tiny and often redundant labels, which cannot be deciphered without greatly zooming in. The size of these figures also varies greatly, which is not critical, but could be improved upon.

Authors' reply

We thank the reviewer for these comments on presentation and accessibility. We have revised the manuscript extensively to improve linguistic clarity, reduce repetition, make technical notation more transparent, and improve figure readability.

The Results and Discussion have been substantially reorganized so that each section has a clearer and more distinct purpose. The revised Discussion is now structured into four focused subsections: 4.1 A stronger but geometrically narrower residual circulation, 4.2 Vertically structured wave forcing and its relationship to the circulation response, 4.3 Pathway-dependent tracer transport from MLS observations, and 4.4 Implications, limitations and synthesis. This restructuring reduces repetition of the Results and shifts the Discussion toward interpretation, limitations, and synthesis rather than restating numerical findings.

We also revised the treatment of technical variable names. In the main text, the CMIP6 archive names are now introduced together with their physical meanings and thereafter discussed using the physical terms. For example, wtem is introduced as the transformed Eulerian mean residual vertical velocity, hereafter w*, while utendepfd, utendogw, and utendnogw are introduced respectively as the EP-flux-divergence tendency, orographic gravity-wave tendency, and non-orographic gravity-wave tendency. These terms are then interpreted explicitly as zonal-momentum tendencies rather than left as unexplained variable codes.

For CMIP6 configuration labels such as i1p1f1, we have retained the standard CMIP6 notation because different configurations are analysed separately and should remain traceable to the source simulations. However, the revised manuscript now explains that the i, p, and f indices denote CMIP6 initialization, physics, and forcing indices, while realization indices identify individual ensemble members. We also clarify that different configurations are retained separately and are not assumed to be interchangeable. The Table 1 notes repeat this definition so that readers do not need to infer the meaning of the labels from the archive notation alone. \

Table 1. Inventory and temporal coverage of the CMIP6 age-of-air and residual-circulation datasets available for this study. Historical and SSP3–7.0 simulations are listed separately by model, variable and configuration. N denotes the number of available ensemble members; the two period-specific columns in each panel give the numbers of members with complete temporal coverage of the indicated 30-year intervals. Coverage denotes the earliest and latest years represented across the available files.

(a) Historical simulations						
Model	**Variable**	**Configuration**	**Coverage**	**N**	**1871–1900**	**1985–2014**
CNRM-ESM2-1	AoA	i1p1f2	1850–2014	9	9	9
CanESM5	w*	i1p2f1	1850–2014	25	25	25
HadGEM3-GC31-LL	w*	i1p1f3	1850–2014	5	5	5
HadGEM3-GC31-MM	w*	i1p1f3	1850–2014	2	2	2
MIROC6	w*	i1p1f1	1850–2014	1	1	1
MRI-ESM2-0	AoA	i1p1f1	1870–2014	10	10	10
MRI-ESM2-0	AoA	i2p1f1	1870–2014	1	1	1
MRI-ESM2-0	w*	i1p1f1	1850–2014	10	10	10
UKESM1-0-LL	AoA	i1p1f2	1850–2014	16	16	16
UKESM1-0-LL	AoA	i1p1f3	1850–2014	3	3	3
UKESM1-0-LL	w*	i1p1f2	1850–2014	15	15	15
UKESM1-0-LL	w*	i1p1f3	1850–2014	3	3	3

(b) SSP3–7.0 simulations						
Model	**Variable**	**Configuration**	**Coverage**	**N**	**2041–2070**	**2071–2100**
CESM2	w*	i1p1f1	2015–2100	3	1	3
CESM2-WACCM	w*	i1p1f1	2015–2100	3	1	1
CNRM-ESM2-1	AoA	i1p1f2	2015–2100	5	5	5
CanESM5	w*	i1p1f1	2015–2100	15	15	15
CanESM5	w*	i1p2f1	2015–2100	25	25	25
CanESM5-1	w*	i1p1f1	2015–2100	10	10	10
CanESM5-1	w*	i1p2f1	2015–2100	10	10	10
CanESM5-CanOE	w*	i1p2f1	2015–2100	3	3	3
UKESM1-0-LL	AoA	i1p1f2	2015–2100	19	19	16
UKESM1-0-LL	w*	i1p1f2	2015–2100	19	16	16

Notes. (a) AoA: mean age of stratospheric air (meanage); w: transformed Eulerian mean residual vertical velocity (wtem); Ψ*: transformed Eulerian mean residual mass streamfunction (psitem). (b) Configuration labels retain the CMIP6 initialization (i), physics (p) and forcing (f) indices; realization (r) indices vary among ensemble members. Configurations are reported separately. (c) Complete temporal coverage does not establish tracer equilibration, field-value validity, historical–scenario parent compatibility or availability of a matching circulation/AoA variable. Counts therefore describe the inventory and must not be interpreted as final sample sizes for paired analyses. (d) The UKESM1-0-LL SSP3 wtem member r13i1p1f2 contains 300 missing monthly records within its available time span and is excluded from both future-period counts. Other incomplete members are also omitted from the corresponding period counts. Member counts across variables are not additive because the same simulation may supply more than one variable.*

The figures have also been redesigned for clearer quantitative interpretation. Several previous climatological plots were replaced by period-mean change figures, panel content was simplified, and common or zero-centred scales are used where direct comparison is intended. For example, Fig. 2 now presents model-specific historical w* changes using one common zero-centred colour scale across all panels. Figure 5 likewise uses physically descriptive panel headings—EP-flux divergence, orographic gravity waves, and non-orographic gravity waves—rather than raw CMIP6 variable codes.

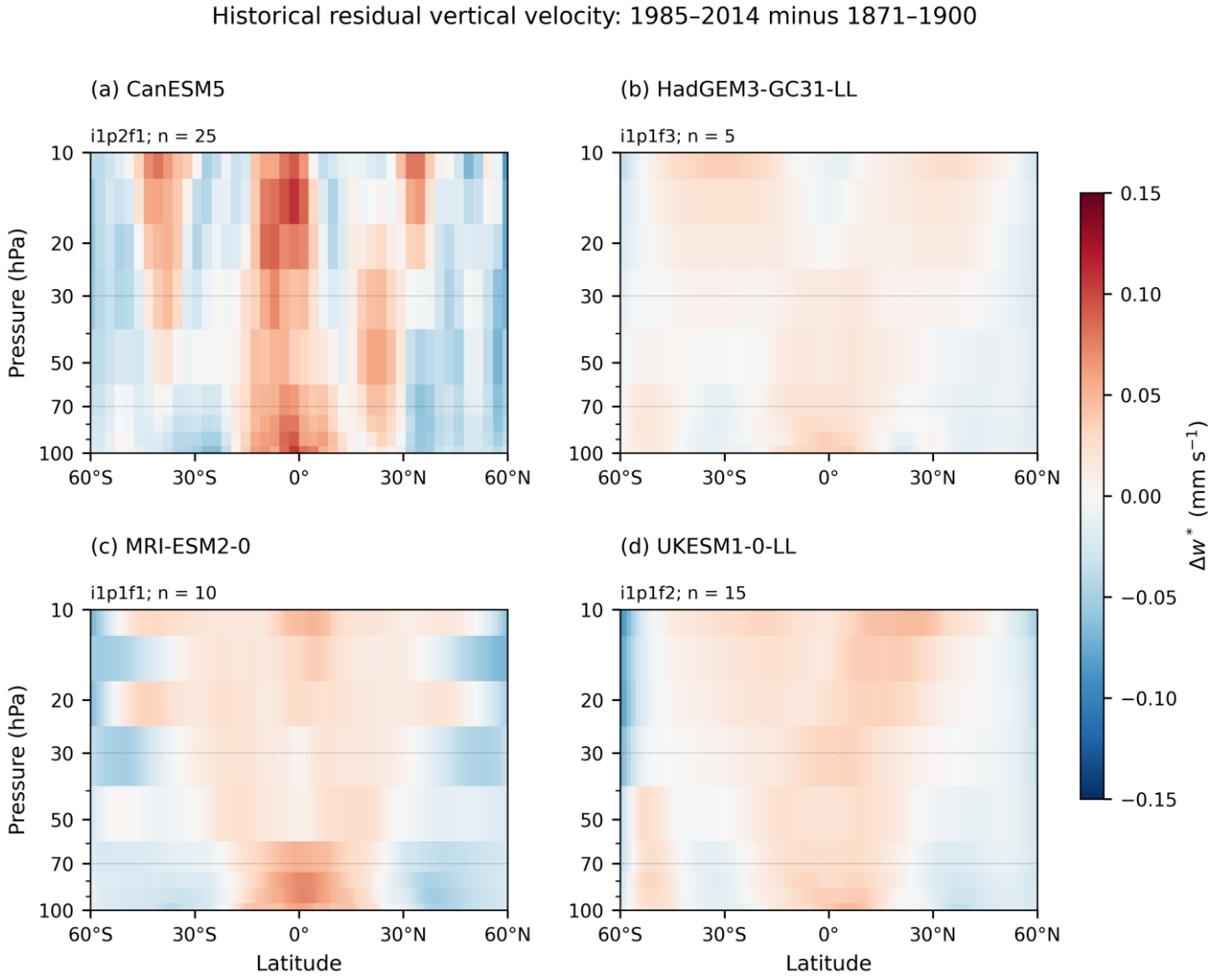


Figure 2. Model dependence of historical residual vertical velocity changes: (a) CanESM5 i1p2f1 (n = 25); (b) HadGEM3-GC31-LL i1p1f3 (n = 5); (c) MRI-ESM2-0 i1p1f1 (n = 10); (d) UKESM1-0-LL i1p1f2 (n = 15). Panels show equal-member annual period-mean changes, 1985–2014 minus 1871–1900, in mm s^{-1}, 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.

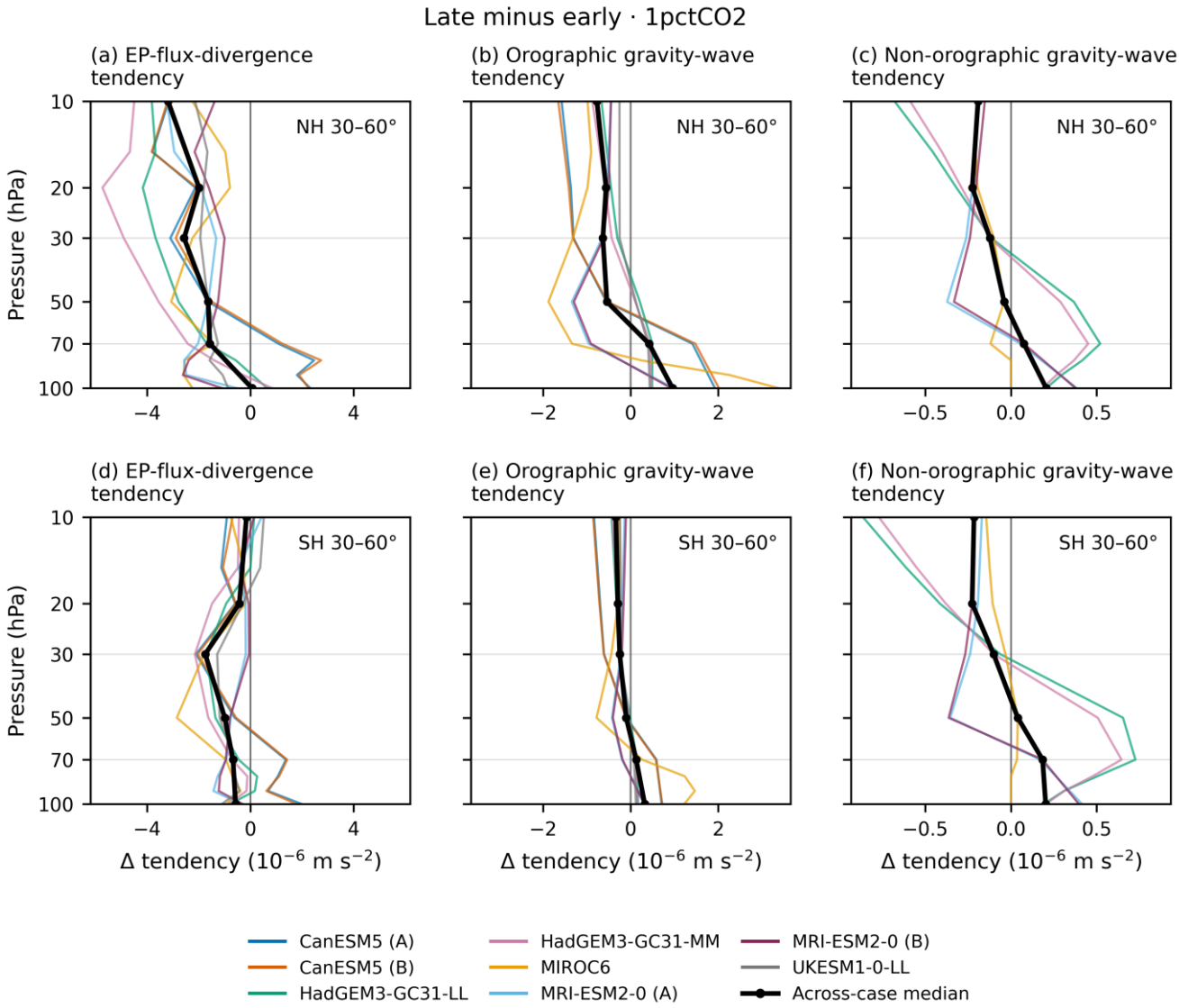


Figure 5. Vertical distribution of wave-related zonal-momentum tendency changes under idealised CO_2 forcing. Rows show NH 30–60° and SH 30–60° area-weighted bands; columns show EP-flux divergence, orographic gravity waves and non-orographic gravity waves. Values are late (experiment years 111–140) minus 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).

Finally, we carried out a general language and formatting revision throughout the manuscript to correct spelling, punctuation, and awkward phrasing and to improve consistency of terminology, section structure, figure captions, and citations.

These changes were intended to make the revised manuscript more readable to readers who are not already familiar with CMIP6 archive conventions while preserving the traceability of the underlying simulations.

Key revised locations: throughout manuscript; Sect. 2.1; Sect. 2.4; Sects. 3.1–3.4; Sects. 4.1–4.4; Table 1 notes; revised Figs. 1–7 and captions.

Regarding its content, the authors present their analysis as a form to utilize many forms of analyses in a unified framework as a proposed framework for future analyses, but ultimately present a collection of several, established analyses, whose synergy mostly lies on the qualitative side. While I do not critizise each analyis itself, the initial framing of the work might have been a bit unfortunate. The analyses themselves (naturally) contain the same results, since most models are indeed consistent with themselves across different properties. For example, the distribution of Age-of-Air does depend on the transport scheme used in a model, and therefore any quantity derived from such a scheme cannot directly contradict the scheme itself, to some extend. The argument for this particular approach is therefore relatively weak.

Authors' reply

We thank the reviewer for this important point. We agree that the original manuscript placed too much emphasis on the novelty of an "integrated framework" and risked implying that agreement among the individual diagnostics itself constituted the main scientific advance. We have therefore revised the framing substantially.

The revised manuscript no longer presents the study as a new methodological framework in which several established diagnostics are expected to confirm one another. Instead, the diagnostics are now treated explicitly as complementary but non-equivalent measures of different aspects of the transport system. AoA characterizes integrated transit-time behaviour, w* characterizes residual-circulation strength, objectively diagnosed w*=0 boundaries characterize circulation geometry, wave-related zonal-momentum tendencies provide dynamical context, and MLS H_2O lags characterize observed tracer propagation.

We also agree with the reviewer that internal consistency among model-derived quantities should not be overinterpreted as independent confirmation. For this reason, the revised manuscript no longer uses agreement among AoA, w*, and other model diagnostics as proof of a single underlying "transport-efficiency" response. Instead, the analysis focuses on whether the diagnostics reveal different structural properties of the circulation response. In particular, the revised results show that stronger tropical residual upwelling can coexist with a narrower lower-stratospheric upwelling region, demonstrating that circulation strength and circulation geometry are not interchangeable properties.

The same principle is applied to the other diagnostics. The wave-related momentum tendencies are interpreted as a vertically structured dynamical context rather than as a direct local measure of w*, and the MLS analysis is treated as an observational characterization of tracer propagation rather than as an independent validation of the modelled circulation changes.

Accordingly, we have revised the manuscript so that the value of combining the diagnostics lies not in demonstrating mutual agreement, but in distinguishing how circulation strength, geometry, dynamical forcing, and tracer propagation can evolve differently while remaining physically consistent. This revised framing is now stated explicitly in the Introduction and Discussion.

We believe this change addresses the reviewer's concern by removing the earlier overstatement of methodological synergy and by focusing the manuscript on the physical information gained from the differences among the diagnostics rather than their expected internal consistency.

Key revised locations: final paragraphs of the Introduction; Sects. 3.1–3.4; Sects. 4.1–4.4.

Regarding the novelty of the results, I have to question if there are insights gained from this study that were unknown to me previously. An acceleration of the BDC under increased greenhouse gas forcing has been hypothesized by Rind et al. 1990. Streamline functions have been used by to investgiate the trends of the BDC e.g. Okamoto et al. 2011, which do attribute the acceleration to increased wave activity. The analysis of using water vapor anomalies to quantity the BDC is loosely based on Flurry et al. 2013, which the authors acknowledge. Indeed, while the authors do reference previous work adequately in the respective „Results“ subsections, the novelty of their results does come into question.I found that some results presented in this manuscript, mostly through the various figures, are not necessarily well supported. Overall, unfortunately I cannot recomment the manuscript in its current form for publication. Parts of the analyses should be revised and it should be made sure that they allsupport the overarching claim of the manuscript. The style of the manuscript has a lot of potential for improvement, and maybe this work could be rephrased as a comprehensive study on BDC dynamics under greenhouse gas forcings rather than a new analysis framework, if the combination of these different analysis methods yields some benefit.

Authors' reply

We thank the reviewer for this important and constructive assessment. We agree that the original manuscript placed too much emphasis on the novelty of a new "analysis framework" and on the general finding of BDC acceleration, both of which are already well established in the literature. In response, we have substantially revised the framing, analysis, figures, and interpretation of the manuscript so that the contribution now lies in resolving the structured nature of the BDC response, rather than in presenting the individual diagnostic methods themselves as novel.

The revised paper is no longer framed primarily as a new methodological framework. Instead, it is presented as a comprehensive diagnostic study of BDC dynamics under greenhouse forcing, in which established but complementary diagnostics are used to distinguish circulation strength, circulation geometry, wave-related forcing, and observed tracer propagation. The revised Introduction now makes clear that these quantities are related but non-interchangeable and are not combined into a single "transport-efficiency" metric.

The first major advance of the revised analysis is the objective diagnosis of circulation geometry, which was not part of the original framing. The new analysis shows that stronger tropical residual upwelling can coexist with a narrower lower-stratospheric upwelling region. By late century, the diagnosed width at 70 hPa decreases by approximately 6.5° in both CanESM5 and UKESM1-0-LL, while tropical ascent and extratropical descent strengthen. This demonstrates that circulation strength and circulation geometry are distinct properties and need not evolve in the same direction. The result is quantified directly in Fig. 4 and Table 3, including bootstrap uncertainty for the diagnosed boundary shifts and width

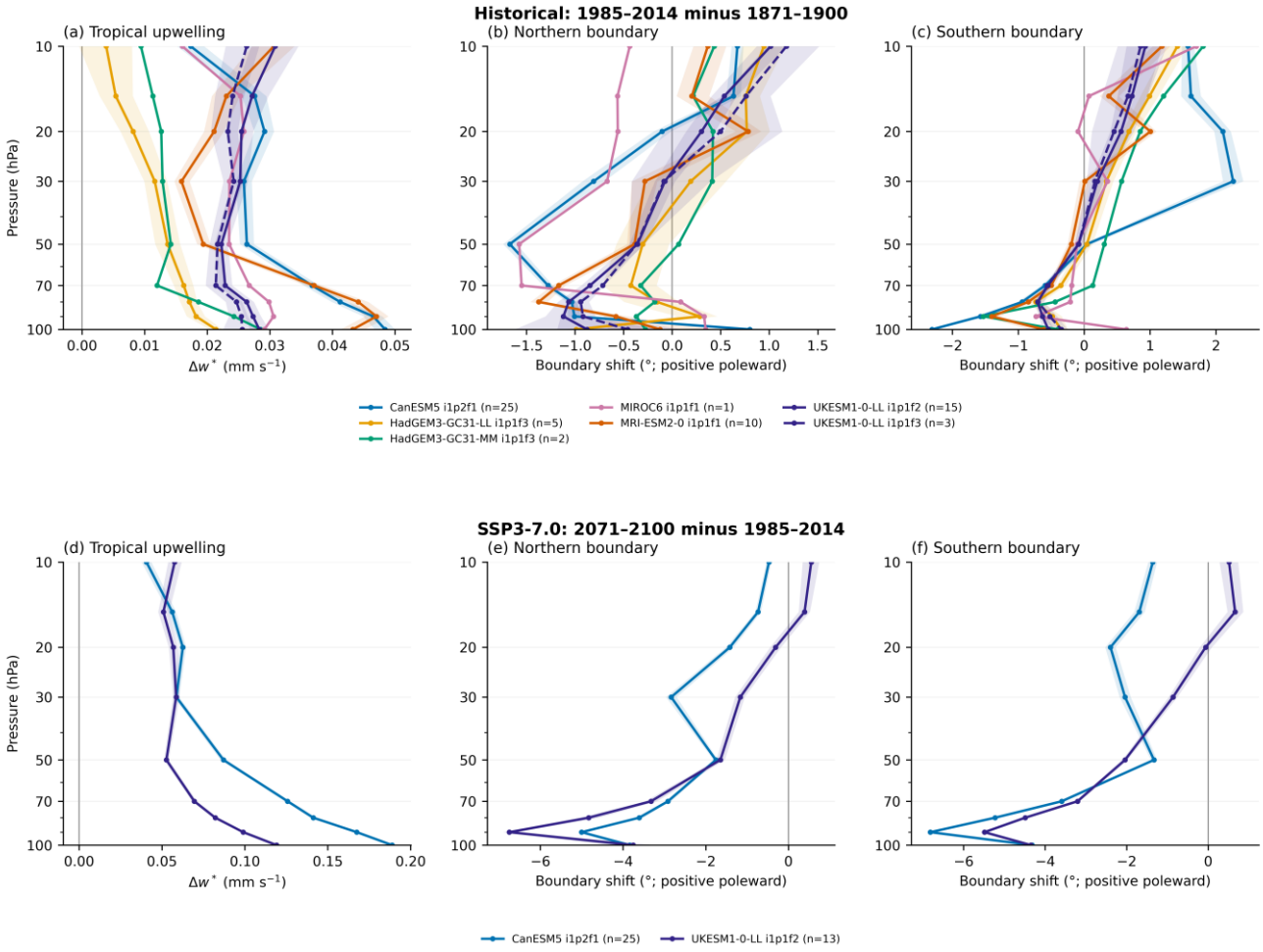


Figure 4. Changes in tropical residual vertical velocity (20°S–20°N) and NH/SH upwelling boundaries. Top: 1985–2014 minus 1871–1900. Bottom: SSP3-7.0 2071–2100 minus historical 1985–2014. Lines show equal-member mean period changes; shaded bands are pointwise 95% percentile member-bootstrap intervals 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]

The second major contribution is the revised analysis of wave-related zonal-momentum forcing. Rather than interpreting the wave response as a vertically uniform enhancement associated directly with stronger upwelling, the revised analysis shows a strongly height-dependent response. The clearest agreement among the available model/configuration cases occurs around 10–30 hPa, whereas the response near 70 hPa is substantially more heterogeneous. This is now quantified in Fig. 5 and Table 4, which separate EP-flux-divergence, orographic gravity-wave, and non-orographic gravity-wave tendencies and report the experiment-stage changes, sign agreement, and member-bootstrap uncertainty.

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

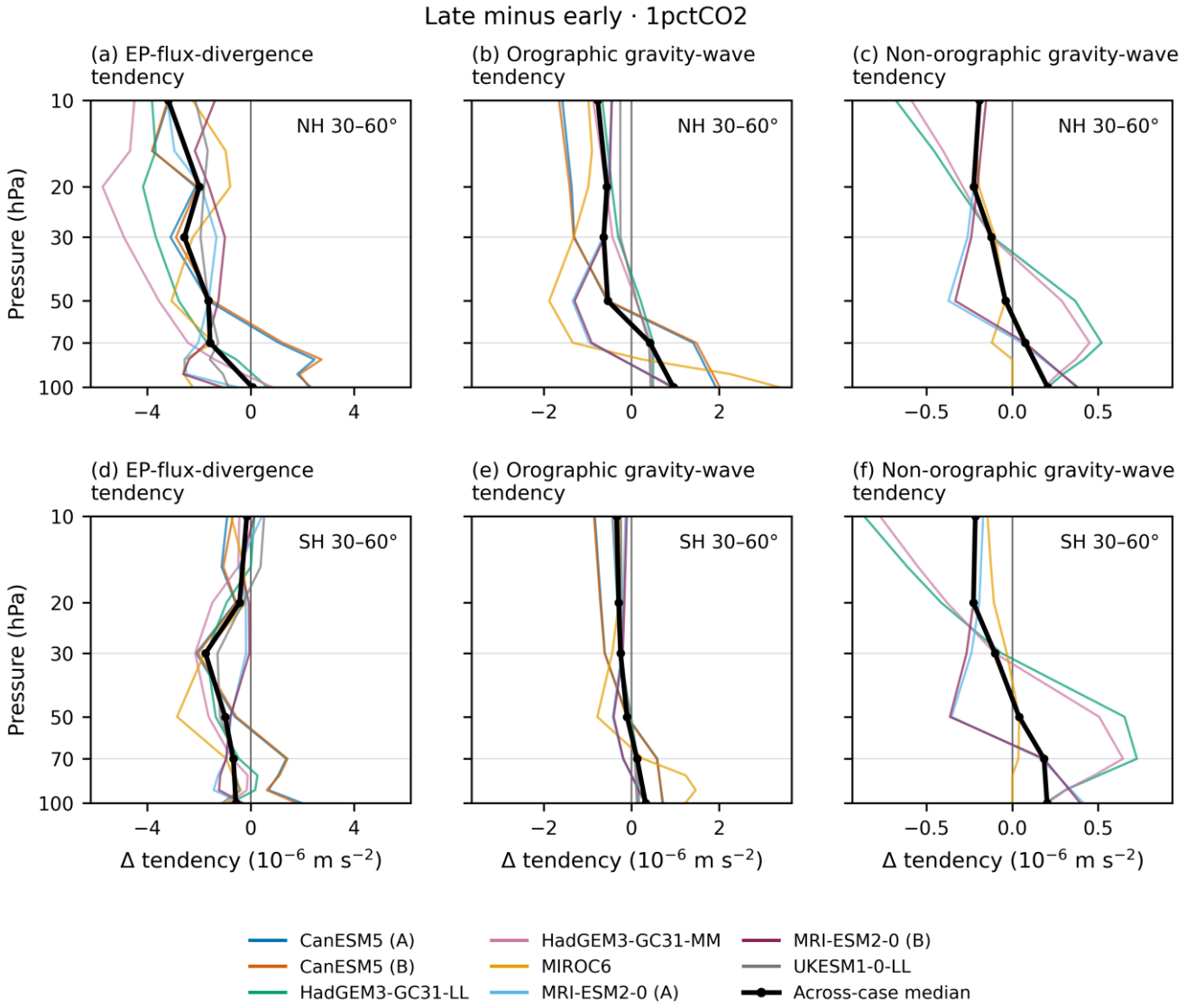


Figure 5. Vertical distribution of wave-related zonal-momentum tendency changes under idealised CO_2 forcing. Rows show NH 30–60° and SH 30–60° area-weighted bands; columns show EP-flux divergence, orographic gravity waves and non-orographic gravity waves. Values are late (experiment years 111–140) minus 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).

This result strengthens the revised physical interpretation because it shows that the altitude of the clearest wave-forcing response does not simply coincide with the altitude of the strongest lower-stratospheric w* response. The manuscript therefore now interprets the wave diagnostics as vertically structured dynamical context within the broader TEM/downward-control framework, rather than as a one-to-one local explanation of BDC acceleration.

The third major revision concerns the MLS tracer-propagation analysis. This part has been reanalysed extensively rather than being retained as a qualitative extension of Flury et al. (2013). The revised analysis removes the prescribed 45- and 90-d lag thresholds, uses an unconstrained 0–365 d search on unsmoothed daily anomalies, applies moving-block bootstrap uncertainty, and subjects every initially stable pathway to a four-method autocorrelation-based robustness audit. Of 365 tested meridional and vertical pathways, only 39 retain a uniquely defensible lag. The resulting pathway dependence is summarized in Fig. 7 and Table 5, which show that robust propagation timescales vary strongly with latitude, altitude, and hemisphere and therefore cannot be represented by one characteristic BDC timescale.

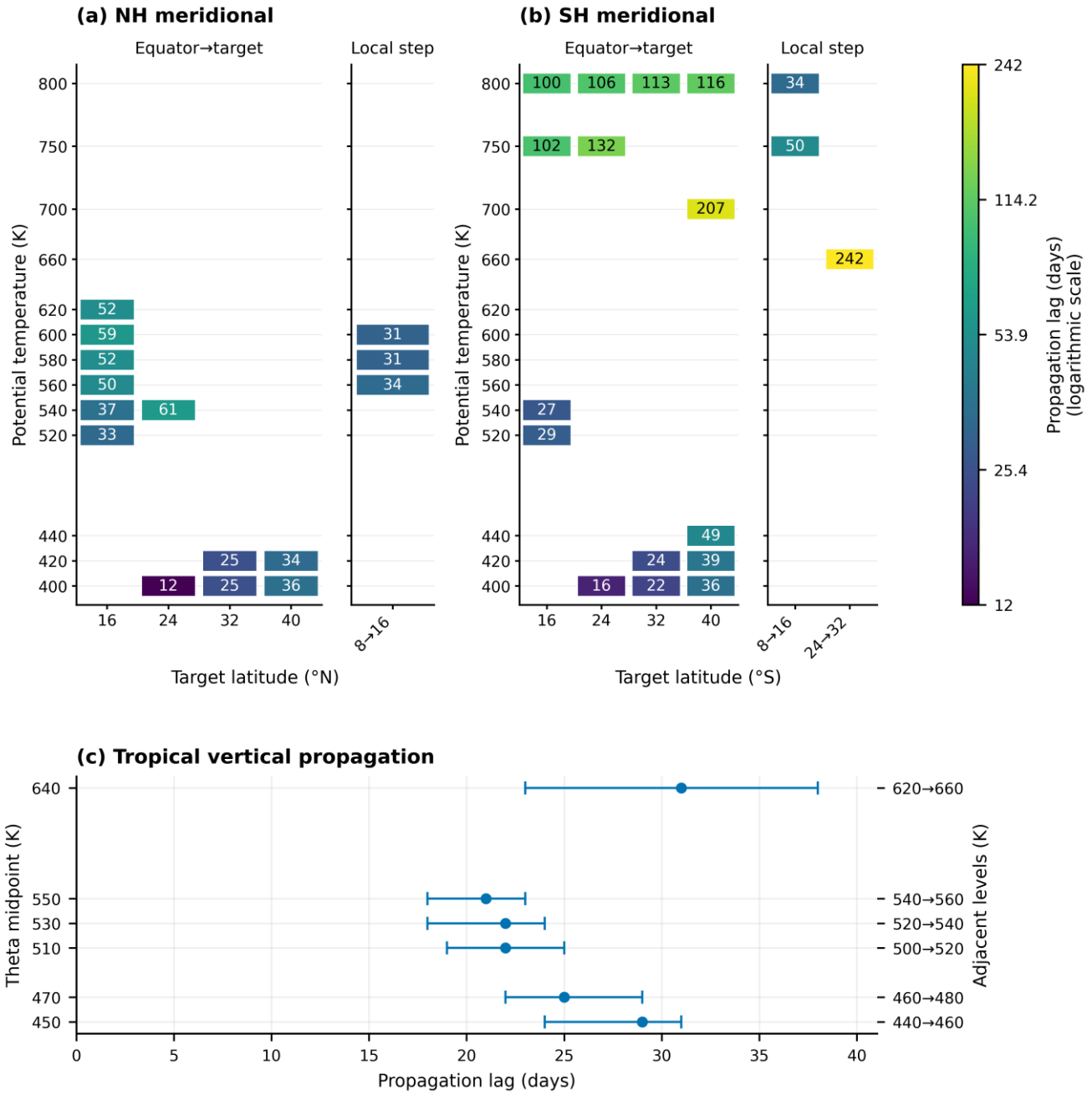


Figure 7. Robust Aura MLS H_2O propagation pathways from the full uncertainty audit. Only final ROBUST_STABLE pathways are shown: initially stable unconstrained lag peaks that remained STABLE_SINGLE_PEAK under all four predefined block-length definitions (first ACF zero crossing, e-folding, truncated 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

Together, these three revised result sets demonstrate that the contribution of the study is no longer simply that the BDC strengthens under greenhouse forcing. Rather, the revised manuscript shows that the response is structured in several distinct

dimensions: the residual circulation becomes stronger but geometrically narrower in the lower stratosphere, the wave-related forcing response is vertically differentiated, and observed tracer propagation occurs only along selected latitude–height pathways with strongly varying timescales.

We have therefore followed the reviewer's suggestion and reframed the manuscript as a comprehensive study of BDC dynamics under greenhouse forcing rather than as a fundamentally new analysis framework. The revised novelty lies in the physical synthesis produced by the coordinated analyses, particularly in identifying where circulation strength, geometry, forcing, and tracer propagation diverge from a simple uniform-acceleration picture. The Discussion now states explicitly that the value of combining the diagnostics lies not in requiring them to produce the same signal, but in showing that they describe different components of the transport system.

We believe that the revised Fig. 4/Table 3, Fig. 5/Table 4, and Fig. 7/Table 5 together demonstrate that the manuscript now has a substantially stronger analytical basis and a clearer scientific contribution than the original submission.

Key revised locations: Introduction; Sect. 3.2; Fig. 4/Table 3; Sect. 3.3; Fig. 5/Table 4; Sect. 3.4; Fig. 7/Table 5; Sects. 4.1–4.4.

Minor comments:
Overall spelling and punctuation must be improved. If requested, I am provide a list of the errors I found, but they are numerous and trivial to fix. Figures need to be adjusted/redone to ensure visibility. The linguisitc style of the manuscript could be adjusted to more closely resemble standard English. Equations should be numbered, unless trivial (such as the conversion from volume mixing rate to part per million). The lag correlation coefficient should be expressed here for a discrete time series, the summation limits need to be corrected.

Authors' reply

We thank the reviewer for these detailed presentation and notation comments. The manuscript has undergone extensive linguistic and formatting revision, including correction of spelling, punctuation, grammar, terminology, and sentence structure throughout the main text and Supplement. The figures have also been substantially redesigned to improve readability, with simplified panel content, clearer physical labels, enlarged annotations, and more consistent presentation across the manuscript.

We have also reviewed the mathematical notation and equation presentation. Non-trivial diagnostic definitions are now presented more clearly in the revised Methods, while simple unit conversions are retained in the text rather than treated as formal equations.

The lag-correlation methodology has been substantially revised. Rather than retaining the original continuous-form correlation expression with problematic summation limits, the revised Sect. 2.5 now defines the calculation explicitly for the daily discrete source and target time series, with correlations evaluated at integer lags from 0 to 365 d. Positive lag is defined as the source anomaly leading the target anomaly, and each lag must satisfy the prescribed minimum number of valid source–target pairs and temporal-overlap criteria.

In addition, the statistical treatment of the lag correlations has been strengthened substantially. The revised analysis no longer relies on conventional Pearson significance calculated from the nominal daily sample size. Instead, temporal dependence is accounted for using a moving-block bootstrap of the paired lag-moment series, with the complete 0–365 d correlation curve reconstructed and the candidate peak redetected in each of 1000 bootstrap realizations.

We therefore believe that the revised manuscript addresses both the reviewer's presentation concerns and the more specific concern regarding the mathematical definition and statistical treatment of the lag-correlation analysis.

Key revised locations: throughout manuscript and Supplement; revised Figs. 1–7; Sect. 2.3 for circulation-geometry definitions; Sect. 2.5 for the discrete daily lag-correlation analysis and uncertainty treatment.

Abbreviations should be revised to allow a more intuitive understanding of the abbreviation without needing to look up the meaning from a few pages ago. Citations sometimes are missing brackets or unnecessarily have them. Citations need to actually support the statements they are mentioned with. ACP conventions should be upheld. A summary at the end of the manuscript would be very useful here.

Authors' reply

Thank you for these helpful suggestions. We have revised the manuscript accordingly. First, abbreviations were reviewed throughout the text to improve readability, with key terms reintroduced or defined where they first appear in major sections and figure/table captions so that readers do not need to refer back several pages for their meaning. For example, Sect. 2.1 explicitly defines age of air (AoA) and residual vertical velocity (w*), while Sect. 2.5 reintroduces the Aura Microwave Limb Sounder (MLS) and the H_2O tracer dataset used in the observational analysis.

Second, we systematically rechecked all in-text citations for bracket placement, formatting, and scientific relevance. Citations that did not directly support the associated statements were removed or replaced, and the remaining references were checked against the revised bibliography to ensure consistency with ACP conventions. We also revised the reference list to remove uncited works and to provide complete bibliographic information for all retained citations.

Finally, we agree that a synthesis at the end of the manuscript improves accessibility. We therefore expanded Sect. 4.4, **"Implications, limitations and synthesis,"** to summarize the principal conclusions and limitations of the study. A summary paragraph is now added at the final paragraph to reconcile the overall finding of the work. The final paragraph synthesizes the distinct responses of circulation strength, circulation geometry, wave-related forcing, and MLS tracer propagation rather than reducing them to a single transport-efficiency metric. We have also added a conceptual summary figure (Fig. 8) that visually integrates the main findings and their physical interpretation.

Major comments:

- What are the weights used for your low-pass filter? If I'm reading the manuscript correctly, these weights are never mentioned. They are an essential part of your methodology. Every choice of parameters will have an influence on the results.

Authors' reply

We thank the reviewer for raising this important methodological point. We agree that the original manuscript did not document the filtering weights sufficiently clearly and that smoothing choices can influence lag-correlation results.

In the revised analysis, however, the low-pass filter is no longer used in the primary MLS propagation analysis. The primary analysis now uses unsmoothed daily H_2O anomalies at individual potential-temperature levels over 2004–2021, so the main reported propagation lags do not depend on any low-pass-filter weighting choice.

Smoothing is retained only as a sensitivity test. For these tests, we use centred uniform running means of 15 and 31 d, with equal weights of 1/15 and 1/31, respectively, and require complete windows rather than partial-window averages. The analysis settings explicitly record these weights as 15 equal weights of 1/15 and 31 equal weights of 1/31.

This sensitivity analysis also demonstrates why smoothing was removed from the primary methodology. The revised Results show that 15- and 31-d running means can generate apparently stable lag peaks in pathways that are not stable in the unsmoothed daily data. Consequently, all primary propagation estimates are based on the unsmoothed series, while the smoothed calculations are used only to evaluate methodological sensitivity.

We have therefore addressed the reviewer's concern in two ways: the smoothing weights are now explicitly defined for the sensitivity experiments, and, more importantly, the principal scientific conclusions no longer rely on a low-pass-filtered time series.

Key revised locations: Sect. 2.5; Sect. 3.4; Fig. S6; Table S8; Supplementary Methods.

- Line 121: „...necessitating ensemble averaging to obtain reliable conclusion…"
I take it that you average mean Age-of-Air and vertical wind speeds. As mentioned in section „Limitations and Future Perspectives" as a third point, generally you cannot trivially average mean Age-of-Air distributions for the reasons mentioned in this section. There are effects such as multimodel mean dilution, biased in absolute values of Age-of-Air and such. I am willing to assume that this has been considered in this analysis, but it must be addressed in the manuscript as well.

Authors' reply

We thank the reviewer for highlighting this important issue. We agree that mean age of air should not be combined indiscriminately across models because absolute AoA climatologies, transport characteristics, ensemble sizes, and diagnostic availability differ substantially among CMIP6 models. A simple multimodel average could therefore obscure physically meaningful model differences or introduce multimodel-mean dilution.

We have revised the analysis specifically to avoid this problem. The revised manuscript no longer pools AoA or w* across models into a single grand multimodel mean. Instead, all primary diagnostics are calculated as within-model, within-configuration period-mean differences. For each model/configuration, individual members are first compared between the relevant baseline and target periods, and retained members are then equally weighted only within that model/configuration. Model and configuration identities are preserved throughout the analysis.

The manuscript now states explicitly that, because model climatologies, ensemble sizes, and diagnostic availability differ among the participating CMIP6 models, no grand multimodel mean is used as the primary estimate of the circulation response. This is also why the analysis is based on period differences rather than direct comparison of absolute AoA values between models. The historical and SSP3-7.0 changes are therefore interpreted as within-model changes, not as differences among absolute model climatologies.

We have further strengthened this aspect by retaining CMIP6 configurations separately rather than assuming that different physics or forcing configurations of the same model are interchangeable. Where AoA and w* are both available for the same model, configuration, and realization, we additionally use an exactly matched-member analysis at tropical 70 hPa, as shown in Fig. 3. This allows co-occurring AoA and w* changes to be examined within the same simulations rather than by comparing different model populations.

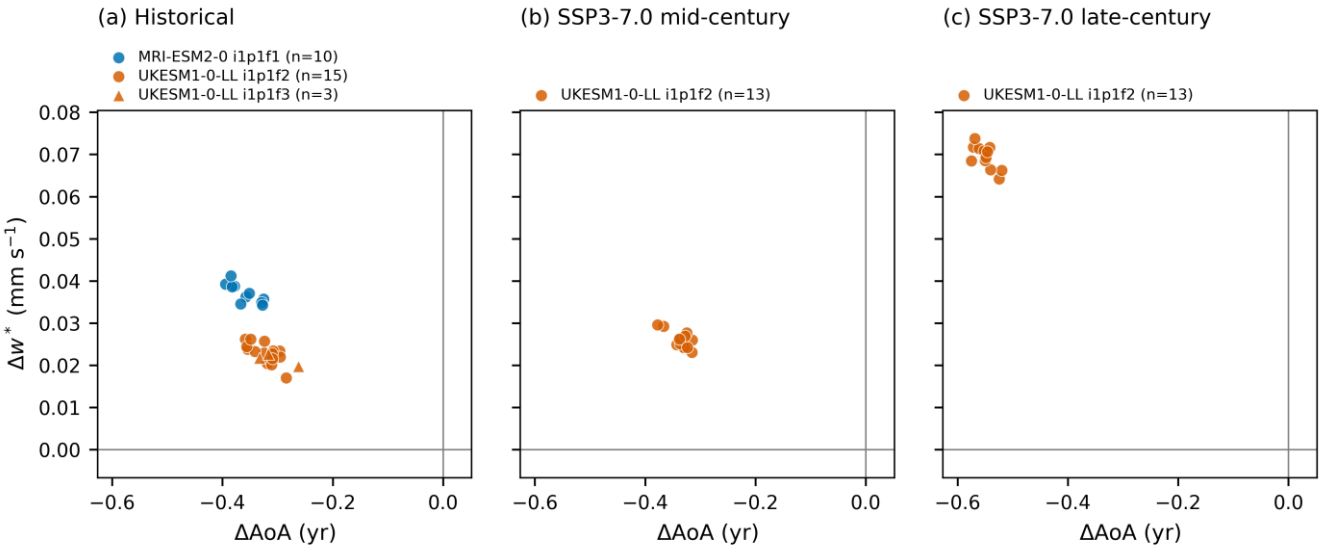


Figure 3. Paired-member annual tropical (20°S–20°N) changes in AoA and residual vertical velocity at 70 hPa. Each point is explicitly matched model/configuration/member. Panels compare (a) 1985–2014 with 1871–1900, (b) SSP3-7.0 2041–2070 with 1985–2014 and (c) SSP3-7.0 2071–2100 with 1985–2014 (target minus baseline). AoA is in years and w* in mm s^{-1}; the source w* changes are multiplied by exactly 1000 for unit conversion (m s^{-1} to mm s^{-1}). Colour identifies model and symbol identifies configuration within model. Both axes have common limits. Zero lines separate joint-sign quadrants.

Accordingly, the original wording suggesting that general "ensemble averaging" was required has been revised. Ensemble averaging in the revised manuscript refers only to averaging retained members within an individual model/configuration, with uncertainty evaluated using member bootstrap and leave-one-member-out tests; it does not refer to averaging AoA distributions across different models.

This revision directly addresses the reviewer's concern regarding AoA climatological biases and multimodel-mean dilution, while preserving the benefit of using ensembles to estimate within-model forced changes and associated uncertainty.

Key revised locations: Sect. 2.1; Sect. 2.2; Table 1; Table 2; Tables S1–S2.

- State what „normalized" and „standardized" mean The term „normalization" is generally used quite broadly, but in my field it refers to scaling to the interval (0, 1). Standardization usually refers to scaling to zero-mean, unit variance. If you do this – motivate why you do it. You must differentiate between scaling between multiple models and within multiple models. For the later, the resulting scaled cannot be trivially compared to each other. As of now, this raises a lot of unnecessary, technical questions.

Authors' reply

We thank the reviewer for pointing out this ambiguity. We agree that the terms "normalized" and "standardized" can imply specific statistical transformations and should not be used unless their definitions and purposes are stated explicitly.

In the revised manuscript, these terms have therefore been removed from the methodological description because no min–max normalization or z-score standardization is applied to the CMIP6 AoA or w* diagnostics. The model results are retained in their physical units and analysed as within-model, within-configuration period-mean differences. AoA is reported in years and w* in mm s^{-1}, and no scaling is applied either across models or within individual models before the primary comparisons.

This revision is consistent with our broader treatment of inter-model differences. Because model climatologies and ensemble characteristics differ, the revised analysis preserves model/configuration identity and does not construct a grand multimodel mean from rescaled fields.

The MLS analysis likewise does not standardize the anomaly time series to zero mean and unit variance. Seasonal variability is removed using fitted annual harmonics, after which the lag analysis is performed on the resulting physical H_2O anomalies.

Accordingly, the revised manuscript avoids the unnecessary ambiguity raised by the reviewer by removing statistical "normalization" and "standardization" terminology where such transformations are not actually performed.

Key revised locations: Sects. 2.1–2.2 and 2.5.

- Usage of Pearson correlation coefficients in table 5
The Pearson correlation coefficient requires the samples to be statistically independent. Applying a smoothing filter does violate this assumption, which could explain these very low values. This metric is not suitable here.

Authors' reply
We thank the reviewer for this important statistical concern. We agree that the original treatment did not adequately account for temporal dependence and that conventional Pearson-correlation significance based on the nominal daily sample size would be inappropriate for these persistent stratospheric time series.

The MLS analysis has therefore been substantially revised. First, the primary analysis no longer uses smoothed time series. It is based on unsmoothed, deseasonalized daily H_2O anomalies for 2004–2021, with smoothing retained only as a sensitivity test. The revised analysis explicitly shows that 15- and 31-d smoothing can generate apparently stable peaks in pathways that are not stable in the unsmoothed data, which is why smoothing is not used for the primary propagation estimates.

Second, we no longer use conventional Pearson p-values or confidence intervals that assume independent daily samples. Instead, temporal dependence is handled using a moving-block bootstrap that preserves contiguous source–target temporal structure. For each pathway, the full 0–365 d lag-correlation curve is reconstructed in 1000 bootstrap realizations and candidate peaks are redetected using the same selection rules. The resulting lag intervals are therefore conditional bootstrap intervals, not ordinary Pearson confidence intervals.

Third, every initially stable pathway is subjected to an additional four-method block-length audit based on the autocorrelation structure of the source and target series. Only pathways that remain a stable single peak under all four block-length definitions are retained as robust. This explicitly addresses the reviewer's concern that serial dependence can create misleading apparent correlations.

In the revised Table 5, the column labelled "Peak *r*" is retained only as a descriptive measure of the correlation amplitude at the selected lag peak; it is not used as an inferential significance statistic. Statistical robustness is instead determined by bootstrap peak recurrence, conditional lag uncertainty, and stability under the four autocorrelation-based block-length definitions. Table 5 now makes this distinction explicit and reports only pathways that pass the full robustness audit.

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

We therefore agree with the reviewer that a naïve Pearson significance test would not be suitable here. The revised manuscript avoids that use entirely: Pearson *r* is retained only descriptively, while the actual inference is based on autocorrelation-aware moving-block bootstrap and peak-stability criteria.

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

- Core statement of the manuscript What are the actual benefits of using many analysis methods in tandem? I am missing some form of synergy. As presented right now, the manuscript presents the findings to simply agree with each other, implying the opposite claim that a single analysis method would have been sufficient. However, some results appear to actually not agree with each other, actually contradicting the initial statement. Ideally, these apparent contradictions would be resolved through the interplay of the different methods, which they do not appear to do so. This needs to be revised.

Authors' reply

We thank the reviewer for identifying this central weakness in the original framing. We agree that the value of combining multiple diagnostics should not be that they merely reproduce the same result. We have therefore revised the manuscript so that the apparent differences among the diagnostics are now treated as physically informative rather than as inconsistencies that must disappear.

The revised core statement is that circulation strength, circulation geometry, wave-related forcing, and tracer propagation diagnose different aspects of the BDC and therefore need not exhibit identical spatial or vertical responses. The purpose of using them together is to determine how these different components of the transport system relate to one another and where a single diagnostic would give an incomplete interpretation. This distinction is now stated explicitly in the Introduction and Discussion.

The clearest example is the relationship between residual-circulation strength and geometry. The w* analysis indicates stronger tropical residual upwelling, whereas the new objective boundary analysis shows that the lower-stratospheric upwelling region becomes narrower rather than broader. These results initially appear contradictory only if circulation strength and circulation extent are assumed to be interchangeable. The revised Fig. 4 and Table 3 show that they are not: stronger ascent can be concentrated within a smaller tropical latitude range. This has become one of the central physical results of the revised manuscript.

A second apparent discrepancy concerns the wave-forcing diagnostics. The largest lower-stratospheric increase in tropical w* occurs near 70 hPa, whereas the clearest cross-model agreement in the wave-related zonal-momentum tendencies occurs higher, around 10–30 hPa. Rather than requiring these quantities to peak at the same level, the revised interpretation recognizes that the residual circulation at a given level reflects the vertically integrated effect of momentum forcing above, consistent with the TEM/downward-control framework. The revised Fig. 5 and Table 4 therefore show why heterogeneous local forcing near 70 hPa can coexist with a strengthened lower-stratospheric residual circulation.

The MLS analysis provides a third complementary perspective. It is no longer used as direct observational validation of the simulated greenhouse-forced circulation response. Instead, the revised Fig. 7 and Table 5 show that observed H_2O propagation is highly pathway dependent and that only a subset of latitude–height pathways support a unique robust propagation lag. This demonstrates that tracer redistribution cannot be inferred directly from w*, AoA, or wave forcing alone, because it also reflects mixing, tracer gradients, and tracer-specific processes.

Thus, the revised synergy of the study lies specifically in resolving these apparent contradictions:

- stronger residual upwelling does not imply a broader upwelling region;
- stronger lower-stratospheric circulation does not require the strongest local momentum-tendency response at the same pressure level;
- and modelled circulation changes do not uniquely determine observed tracer-propagation timescales.

The Discussion now states explicitly that the benefit of combining the diagnostics lies not in requiring them to produce the same signal, but in showing that circulation strength, circulation geometry, dynamical forcing, and tracer propagation describe different components of the same transport system.

We believe this revision directly addresses the reviewer's concern. The manuscript no longer uses agreement among multiple analyses as its central justification; instead, their differing responses provide the physical synthesis and explain why no single diagnostic is sufficient to characterize the greenhouse-associated BDC response.

Key revised locations: final paragraph of the Introduction; Sect. 3.2 and Fig. 4/Table 3; Sect. 3.3 and Fig. 5/Table 4; Sect. 3.4 and Fig. 7/Table 5; Sects. 4.1–4.4.