Response to Anonymous Referee #1 (egusphere-2026-4812)

We would like to express our gratitude to the referee for their constructive and encouraging review. The comments in question are reproduced in italics below. The requested analyses have been completed; in accordance with the instructions issued by the editorial office, the revised manuscript will be prepared and submitted at the time of the editor's request. Consequently, all manuscript changes, including those resulting from the additional experiments reported in our response to Referee #2 (AC2), are hereby described as a proposal. Section references are to the submitted manuscript; Supplement numbering is provisional.

The additional experiments affect comments 1, 2 and 5. The full-forcing Korea-box Z850 decrease of the T42/ERSSTv5 experiments (−7.7 m, $p < 0.001$) is reproduced by a second atmospheric model (CAM4: −6.8 m, $p < 0.001$) and, at about half the amplitude, by a second SST reconstruction in the T42 model (COBE-SST2: −3.8 m, $p = 0.03$). Both models weaken the southerly inflow under ERSSTv5 (not resolved under COBE-SST2), but the basin partition is not reproduced across configurations. We therefore propose to remove the basin attribution from the title and present the T42/ERSSTv5 basin ordering as a conditional result. The numerical evidence and limitations are given under comments 1–2 below; experiment details and full figures are in AC2, comments 1–3.

The redrawn Figs. 3–6, the new Fig. S5 and Tables S6–S7, and Figs. S6 and S8 from AC2 are placed in the text next to the paragraphs that discuss them. Unless stated otherwise, model responses are Korea-box (34–43°N, 124–130°E) JJA (1881–1907) minus (1855–1880), experiment minus CTL, with unadjusted Welch p-values; p_H denotes the Holm-adjusted p-value within the four-forcing Z850 family (comment 3). FDR denotes the false-discovery-rate (Benjamini–Hochberg procedure): stippled maps control it at 10 % (Wilks, 2016), and the box-level tests report the corresponding q-values (comment 3). A marked-up manuscript and Supplement will accompany the revised files.

General comments

The three-channel argument … is a convincing way to rule out the most obvious alternative explanation… The use of 20CRv3 in Sect. 3.3 … is a defensible and instructive way to interrogate an early reanalysis. The pacemaker experimental design (Sect. 2.2, 3.5) is sound, and the additivity check … is a welcome addition. My main concerns are (i) an inconsistency between diagnostic fields in the basin attribution that is not fully reconciled, (ii) the strength of the title/abstract claim relative to the single-model basis of the attribution, and (iii) some statistical details that are currently confined to the Supplement and would strengthen the main text if summarized there.

Response. We address the three concerns directly: comment 1 explains the statistical and physical reasons for the diagnostic-field differences and limits the basin claim; comment 2 proposes revised title, abstract and Conclusions in light of the new configurations; comments 3–6 would bring multiplicity, domain sensitivity, ENSO-test power and serial-dependence treatment into the main text, with a new Text S8 and Tables S6–S7 documenting the checks. Figure S5 supports the circulation-based attribution by showing the model's rain-band bias. These checks use the existing experiments (the redrawn Figs. 4–5 add only the REST experiment); the additional simulations addressing Referee #2 are reported in AC2. The proposed manuscript changes are identified after each response below.

Specific comment 1 — basin attribution differs across diagnostic fields

The headline finding … is based on the 850-hPa geopotential height field … However, the same section reports that for the 850-hPa meridional wind and the vertically integrated moisture flux, no single basin was more than marginally significant and the tropical Pacific contributed comparably. Since these latter fields most directly govern moisture delivery to Korea, I would ask the author to discuss explicitly why the basin ranking differs across fields, rather than noting the discrepancy in passing.

Response. We agree. The difference has both statistical and physical causes, and the wind-based basin partition is unresolved with 10 members. We would explain this explicitly in Sect. 3.5 and present the T42 Z850 ordering as a ranking by point estimate, rather than a demonstrated separation among all basins.

(a) **Signal-to-noise**. The GOGA signal-to-noise ratios (mean response divided by its standard error) are 5.9 for Z850, 2.5 for V850, 2.8 for meridional moisture flux (VIMFy), and 1.6 for moisture-flux convergence. A single-basin signal of roughly one-third of the total would have a signal-to-noise ratio near 1 in the wind-based fields, below detection, so their basin ordering reflects sampling noise as much as physics and is no longer presented as a competing ranking. Even for Z850, the IO−TP difference of −3.2 m has unadjusted Welch $p = 0.03$ but Holm-adjusted $p = 0.10$ over the three pairwise contrasts; NA−TP and IO−NA have unadjusted $p = 0.15$ and 0.30. Individually detectable basin responses therefore do not establish the basin ranking statistically.

(b) **Geostrophic balance**. The meridional wind responds to the zonal gradient of height, not its magnitude. We evaluated the geostrophic V850 from the forced Z850 difference between the grid columns bounding the box (123.8°E and 132.2°E, 8.4° of longitude apart, averaged over the box latitudes, with f at 38°N): −0.12, −0.14, −0.08 and −0.26 m s^{-1} for TP, NA, IO and GOGA, against actual responses of −0.20, −0.20, −0.09 and −0.35 m s^{-1} (Table S6). The IO depression is largest but broad, centred south of Korea (−7.9 m over 25–34°N, 120–130°E versus −5.1 m over Korea); at the bounding columns it is −4.8 m (west) and −5.4 m (east), a difference of only 0.6 m over 8.4° of longitude, which is the −0.08 m s^{-1} geostrophic estimate above, so the local gradient is small. The smaller TP and NA depressions deepen eastward by about 1 m over the same span and yield a comparable wind weakening. VIMFy follows the low-level wind.

(c) **Moisture-delivery pattern**. IO forcing increases upstream southerly wind (+0.29 m s^{-1}, $p = 0.01$), moisture-flux convergence (+0.57 mm d^{-1}, $p = 0.01$) and precipitation (+0.49 mm d^{-1}, $p = 0.008$), consistent with a southward displacement of convergence. TP and NA instead reduce the inflow toward Korea with little upstream change. GOGA combines these effects (upstream V850 −0.50 m s^{-1}, $p < 0.001$; Korea −0.35 m s^{-1}). Its Z850, V850, VIMFy and precipitation responses are individually significant, but its moisture-flux convergence is not ($p = 0.14$), and its wind-based responses are marginal after multiplicity control (comment 3). Single-basin Z850 pattern correlations with GOGA over East Asia are 0.62 (IO), 0.51 (NA) and −0.38 (TP); these are supporting diagnostics of the same height field, not independent evidence.

(d) **Claim and uncertainty.** We restrict the T42/ERSSTv5 basin attribution to Z850, the field in which single-basin responses (IO and NA) are individually detectable in these

experiments; V850 and the meridional moisture flux support the full-forcing weakening in individual tests but are marginal after multiplicity control (comment 3) and do not resolve a basin partition with 10 members, and the moisture-flux convergence is not individually significant ($p = 0.14$). The T42 TP 95% interval, [−5.0, +1.4] m, permits a Pacific contribution of up to about two-thirds of the full response; the 80%-power minimum detectable effect is 4.5 m. Neither the observational null tests (comment 5) nor the T42 experiment excludes a substantial Pacific contribution: what suggests a minor Pacific role in this model is the point estimate, the IO−TP contrast (significant before, not after, adjustment) and the anti-correlated TP pattern. Moreover, CAM4 reproduces the full-forcing height decrease (−6.8 m, $p < 0.001$) and wind weakening, but TP is its only significant single-basin height response after Holm adjustment (−2.8 m, $p_H = 0.024$); IO has $p_H = 0.079$ (raw $p = 0.039$), and NA is small (−0.9 m, raw $p = 0.37$). Under COBE-SST2, no single basin is significant. The full-ocean height response there is −3.8 m (raw $p = 0.03$ as the single confirmatory test of that configuration; $p_H = 0.12$ if it is treated as one of four tests; see comment 3). Thus no basin ordering is established across models and reconstructions. Details are in AC2 (comments 1 and 3).

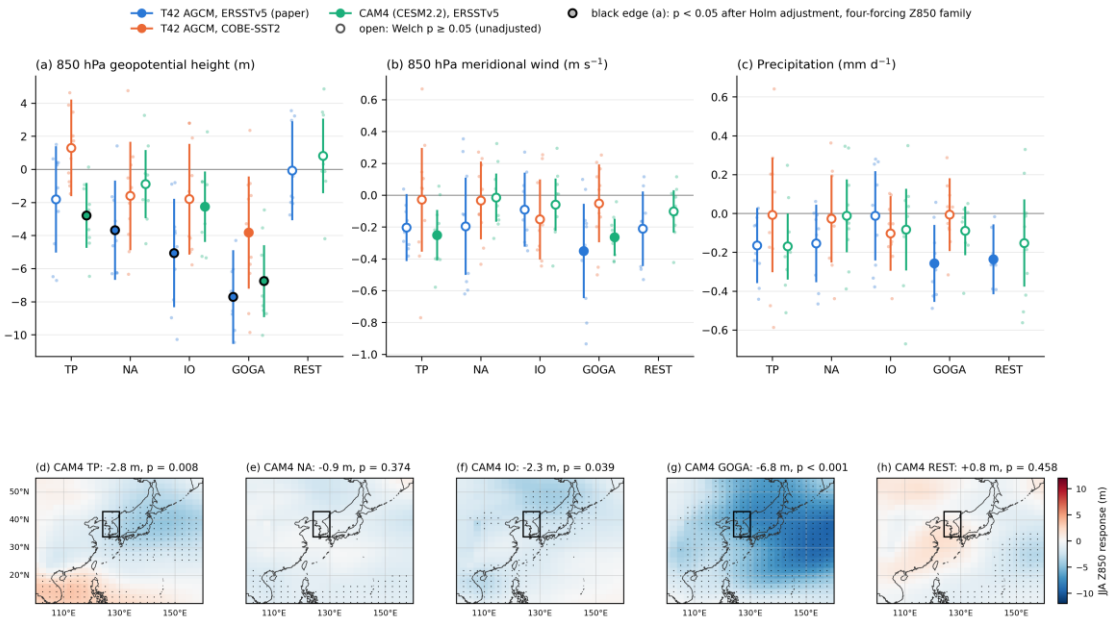


Figure S6 (new; AC2, comment 1). Korea-box responses in T42/ERSSTv5, T42/COBE-SST2 and CAM4/ERSSTv5: (a) Z850, (b) V850 and (c) precipitation. Small dots: members; large symbols: means with 95 % Welch intervals; filled/open: raw $p < / \geq 0.05$; black edges in (a): Holm $p < 0.05$ within the four-forcing Z850 family. (d–h) CAM4 Z850 response maps (shading over terrain above the 850 hPa surface interpolated for display); stippling requires pointwise $p < 0.05$ and false-discovery-rate (FDR) control at 10 %; black box: Korea (34–43°N, 124–130°E). Panels (a–c) are proposed for the main text.

A check of the subtropical high itself (AC2, comment 1(f)) supports stating the attribution for the mass field: the intensity index of Wang et al. (2013; JJA Z850 over 15–25°N, 115–150°E) decreases under the full forcing by 6.0 m in T42 and 4.8 m in CAM4, with eddy components of −4.2 and −3.0 m (all $p_H < 0.001$), supporting a weaker western North Pacific subtropical high under ERSSTv5 in both models; the Korea-box decrease accompanies this broader circulation response, while including the zonal-mean component noted below. The Indian Ocean and North Atlantic contributions to the index are negative by point estimate in all three

configurations, whereas the tropical-Pacific point estimate is positive in both T42 configurations (T42/ERSSTv5 +2.8 m, p_H = 0.003) and near zero in CAM4, where it is not statistically resolved; in CAM4 no single-basin effect on the high is significant after adjustment. About half the Korea-box height decrease is a zonal-mean lowering of the 850 hPa surface; the T42 eddy component is −3.7 m (raw p = 0.006), and in the eddy Korea-box height the order of the T42 single-basin estimates changes (NA −3.6 m, p_H = 0.026; IO −2.3 m, p_H = 0.20; Table S12 in AC2), a further reason to present the basin ranking as a point-estimate property of the full-height diagnostic. AC2, comment 1(f), and Table S12 give the full results and sensitivity of the secondary contour-position diagnostics. The revised text would report eddy alongside full height and distinguish circulation attribution from attribution of the observed rainfall deficit, which neither model reproduces reliably.

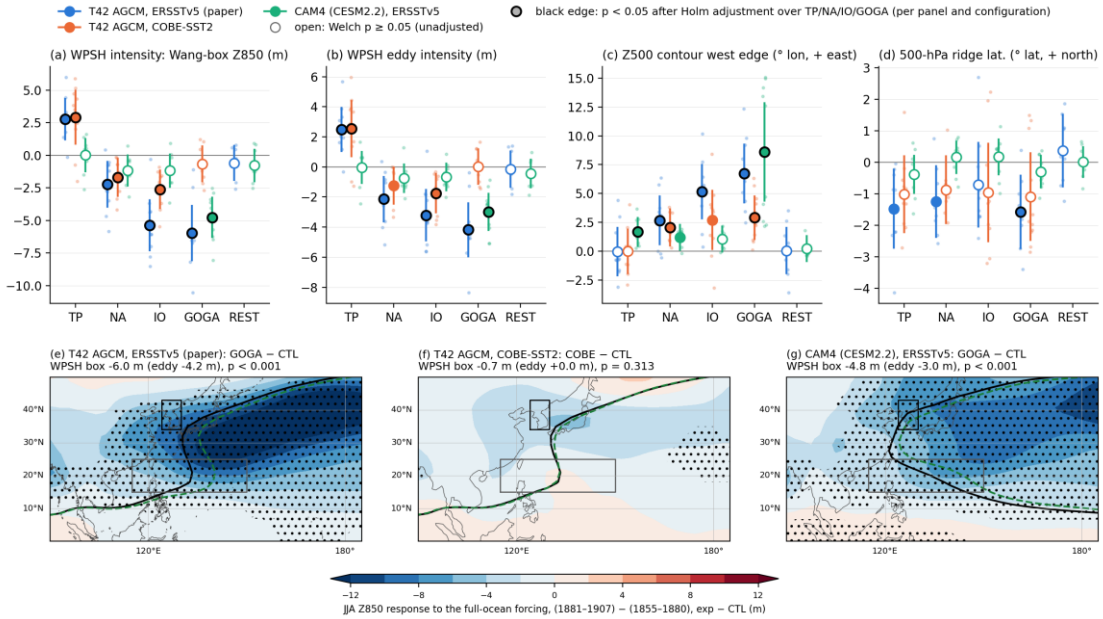


Figure S8 (new; AC2, comment 1(f)). (a) Wang et al. (2013) intensity: Z850 over 15–25°N, 115–150°E; (b) intensity with each latitude's zonal mean removed; (c) westernmost longitude (100–180°E, 10–40°N) of the Z500 contour whose value is the CTL-climatology mean over 15–25°N, 115–150°E (positive = eastward); (d) latitude of maximum Z500 at each longitude 110–150°E, then averaged (positive = northward). Symbols and intervals as in Fig. S6; Holm adjustment is within each index/configuration. (e–g) Full-forcing Z850 responses (shading; interpolated for display over terrain above the 850 hPa surface); black line: the CTL-climatology Z850 contour whose value equals the CTL index-box mean; green dashed line: the same Z850 value in the CTL climatology plus the response; boxes: Korea and index regions; stippling: pointwise $p < 0.05$ and false-discovery-rate (FDR) at 10 %. Position diagnostics (c–d) are secondary and definition-sensitive. (Wang, B., Xiang, B., and Lee, J.-Y., PNAS, 110, 2718–2722, 2013)

Proposed changes. Sect. 3.5 would replace the passing observation of field dependence with this explanation, confidence intervals and qualified basin claims; a new Text S8 and Table S6 would document the checks. The abstract and Conclusions would report the model and reconstruction dependence (comment 2).

Specific comment 2 — strength of the title/abstract claim

The title and abstract state the attribution in fairly unconditional terms, while the Conclusions appropriately note that the result "rests on a single T42 model…" I suggest bringing this caveat forward into the abstract, or softening the title/abstract phrasing…

Response. We agree and propose to remove the basin attribution from the title. With the additional configurations, the appropriate qualification is now the distinction between the reproduced full-forcing circulation response and the model-dependent basin partition, rather than a single-model caveat alone.

Proposed changes.

- Title: "Sustained Late-Nineteenth-Century Summer Drought in Seoul and the SST-Forced Weakening of the East Asian Monsoon Circulation" (wording to be settled with the revised manuscript).
- Abstract and Conclusions: confine the IO-led height response and weak TP response to T42/ERSSTv5; report that CAM4 reproduces height and southerly-wind weakening but has a different basin partition, with TP the only significant basin after adjustment; report the smaller height decrease under COBE-SST2 and its unresolved single-basin contributions. The attribution concerns circulation weakening: T42 misplaces the rain band, and CAM4 does not reproduce the significant full-forcing precipitation response. Quantitative evidence is in comment 1 above and AC2, comment 1. The Conclusions would close by naming what would resolve the basin partition – ensembles of about 30–40 members per experiment, pairwise-forcing experiments, and repetition with a third model and further SST reconstructions (AC2, comment 1) – so that the open question is stated as the next step rather than left as a caveat.
- Sect. 3.5: change the heading to "SST-forced attribution of the circulation weakening: basin contributions in the T42 model and their model dependence" and add the cross-model/reconstruction comparison (AC2, comments 1–3). Report full and eddy subtropical-high intensity as the primary high diagnostics and qualify the western edge and ridge latitude as secondary, definition-sensitive indicators (AC2, comment 1(f); Text S11, Table S12, Fig. S8).

Specific comment 3 — statistical multiplicity

Statistical multiplicity across the many significance tests … is not discussed. Please clarify whether multiple-comparison correction was applied across the basin × field Welch t-tests, and whether the $p < 0.05$ stippling in Figs. 3, 4, and 6 reflects pointwise or field significance.

Response. No correction had been applied, and the submitted stippling was pointwise. We have recomputed the tests and redrawn the maps, and the families and thresholds would be stated explicitly in Sect. 2.

Basin × field tests. The four Korea-box Z850 forcing tests form the primary family. Their Holm-adjusted p-values are 0.0002 (GOGA), 0.014 (IO), 0.039 (NA) and 0.25 (TP). For each new configuration reported in AC2 (CAM4, COBE-SST2) the full-forcing test is the single confirmatory test of replication and the single-basin tests are exploratory; both unadjusted and Holm-adjusted values are given there, and the two standards are stated wherever they differ. The three pairwise basin contrasts are separately Holm-adjusted; IO−TP changes from raw $p = 0.03$ to adjusted $p = 0.10$. Thus, the primary result is that the GOGA, IO and NA responses are

individually detectable, whereas TP is not; it does not establish a significant ordering among the basins.

Across all 16 box-level tests, the Benjamini–Hochberg (BH) FDR q-values are 0.001 (GOGA Z850) and 0.036 (IO Z850), the only results at $q < 0.05$, followed by about 0.07 for GOGA VIMFy, NA Z850 and GOGA V850 (Table S7a). Across all 16, Holm retains only GOGA Z850 at 0.05; IO Z850 has adjusted $p = 0.068$. The abstract and main text would therefore distinguish individual significance from significance after adjustment, rather than describing the wind/moisture responses as significant without qualification.

Table S7 (new), part (a). The 16 box-level Welch tests (Korea box, experiment − CTL) with raw p, Holm-adjusted p across the 16 tests and Benjamini–Hochberg false-discovery-rate (FDR) q.

(a) Test	Raw p	Holm-adjusted p	Benjamini–Hochberg q
Z850 GOGA	0.0001	0.001	0.001
Z850 IO	0.0045	0.068	0.036
VIMFy GOGA	0.013	0.18	0.069
Z850 NA	0.020	0.25	0.074
V850 GOGA	0.023	0.28	0.074
VIMFy TP	0.052	0.58	0.13
V850 TP	0.057	0.58	0.13
MFC TP	0.067	0.60	0.13
MFC NA	0.079	0.63	0.14
VIMFy NA	0.11	0.76	0.17
MFC GOGA	0.14	0.84	0.20
VIMFy IO	0.18	0.90	0.24
V850 NA	0.20	0.90	0.24
Z850 TP	0.25	0.90	0.29
V850 IO	0.42	0.90	0.45
MFC IO	0.68	0.90	0.68

Other tests. We have not pooled tests addressing distinct questions into one family: change-point detection (Pettitt), its effect size (bootstrap/permutation), observational artifact, reanalysis skill, ENSO association and volcanic forcing. The Methods would identify these roles. Because the epoch is data-selected, the epoch tests are explicitly effect-size statements, and the search-corrected Pettitt test supplies the headline significance. Multiplicity adjustment can only raise p-values, so it would not change the interpretation of the ENSO tests, which are already non-significant.

Maps. Figures 3–6, redrawn below, require both pointwise $p < 0.05$ and BH FDR control at 10 % over the plotted domain (Fig. 3b: over the global ocean 60°S–60°N), following Wilks (2016). The coherent height and pressure signals are essentially unaffected (20CRv3 SLP 48.1 % of the domain pointwise versus 47.8 % after FDR control; forced Z850 40–70 % under both criteria), whereas the scattered dots of the noisy fields disappear (NA precipitation 4 % → 0 %, NA moisture-flux convergence 8 % → 0 %, the Fig. 5c residual 12 % → 0 %; Table S7b). IO and GOGA retain all six Korea-box Z850 points, NA four and TP none; the ensemble members are independent integrations, so temporal autocorrelation does not enter these tests. Fig. 3b,

which had no significance test in the submitted manuscript, now carries the same stippling (Welch test of the 27 drought-epoch against the 27 baseline JJA means at each grid point; FDR 10 % over the global ocean 60°S–60°N): the Indian Ocean and North Atlantic cooling are field-significant within ERSSTv5 over 62 % and 43 % of their box areas, whereas the tropical-Pacific warming is not (10 % of the Niño3.4 box, mean +0.24 K against an interannual standard deviation of about 0.8 K). An effective-sample-size correction for lag-1 autocorrelation (median r = 0.14) leaves this contrast in place (57, 25 and 3 %). Because the nineteenth-century ERSSTv5 field is a smoothed reconstruction whose interannual variance is probably underestimated where observations are sparse, these fractions are upper bounds; the sign of the basin changes is checked against the non-infilled HadSST4 in AC2, comment 3. The submitted three-basin residual retains no points after FDR control; the proposed four-term residual likewise has none (0.3 % pointwise). Failure to detect a residual does not establish additivity. The shared-CTL uncertainty and the full revised residual diagnostic are reported in AC2, comment 2, Fig. 5.

Table S7 (new), part (b). Percentage of plotted grid points stippled under the pointwise criterion ($p < 0.05$) and under the false-discovery-rate (FDR) criterion used in the redrawn figures ($p < 0.05$ and below the Wilks (2016) threshold at $\alpha_FDR = 0.10$), and the number of the six Korea-box grid points stippled. The Fig. 5c row refers to the three-basin residual of the submitted figure; the four-term residual of the proposed Fig. 5 (below) has 0.3 % of grid points significant pointwise and none after FDR control.

(b) Map	Pointwise (%)	FDR (%)	Korea-box points
Fig. 3a SLP (20CRv3)	48.1	47.8	– (Mongolia–N. China region 96 → 96 %)
Fig. 3b SST (ERSSTv5; global ocean 60°S–60°N)	38.8	35.4	– (Indian Ocean box 66 → 62 %; North Atlantic box 47 → 43 %; Niño3.4 13 → 10 %)
Fig. 4 Z850 TP	41.7	39.9	0 → 0
Fig. 4 Z850 NA	56.8	56.8	4 → 4
Fig. 4 Z850 IO	66.7	66.7	6 → 6
Fig. 4 Z850 GOGA	69.6	69.6	6 → 6
Fig. 4 precipitation TP	27.7	17.0	0 → 0
Fig. 4 precipitation NA	4.2	0.0	0 → 0
Fig. 4 precipitation IO	30.1	22.9	0 → 0
Fig. 4 precipitation GOGA	35.4	25.9	2 → 1
Fig. 5c residual Z850	11.9	0.0	0 → 0
Fig. 6 MFC TP	23.5	18.5	0 → 0
Fig. 6 MFC NA	7.7	0.0	3 → 0
Fig. 6 MFC IO	28.0	19.6	0 → 0
Fig. 6 MFC GOGA	35.4	30.1	0 → 0

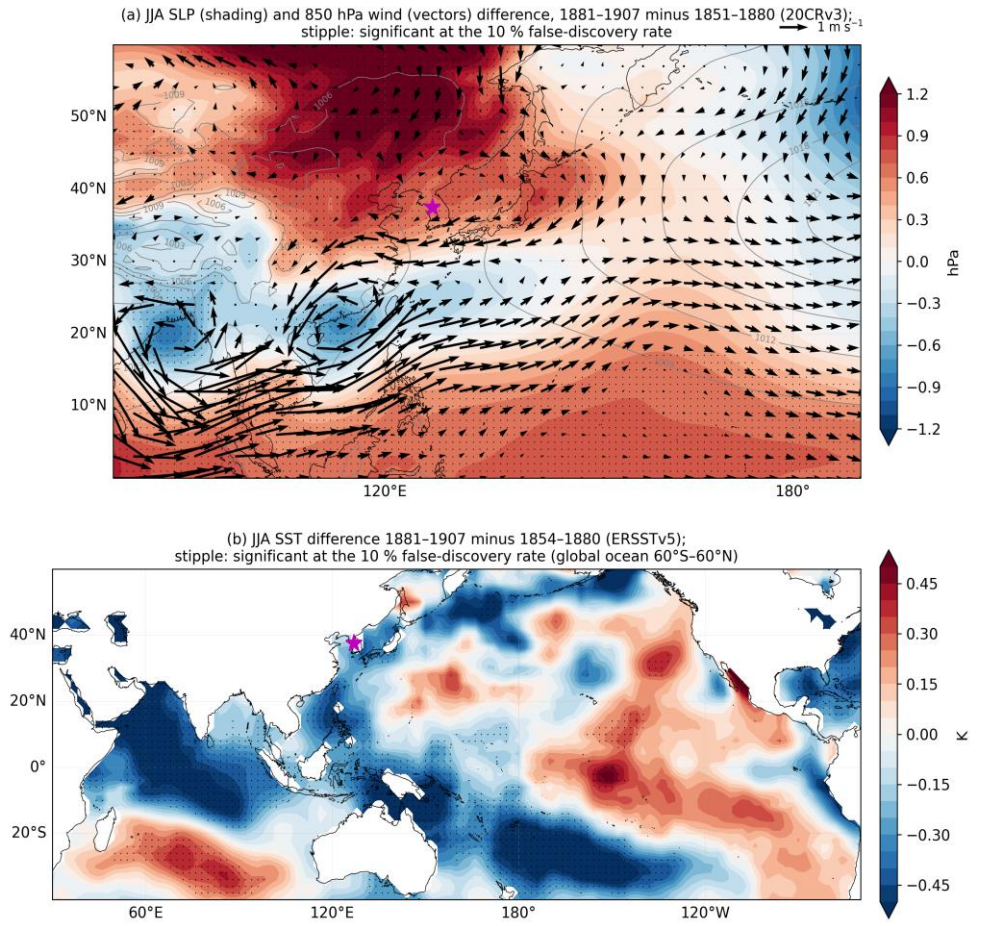


Figure 3 (redrawn). (a) 20CRv3 JJA SLP, 1881–1907 minus 1851–1880; contours: reference climatology; vectors: reanalysis-derived 850 hPa wind; star: Seoul. Stippling requires pointwise $p < 0.05$ and false-discovery-rate (FDR) control at 10 % (Wilks, 2016). (b) ERSSTv5 JJA SST, 1881–1907 minus 1854–1880 (27 against 27 JJA means); stippling as in (a), with the FDR controlled over the global ocean 60°S–60°N (new; the submitted panel carried no significance test). The SLP baseline is the three decades preceding the change point and the SST baseline begins with ERSSTv5 (1854); baseline sensitivity is reported in AC2, comment 5.

Figures 4–5 also incorporate REST, the forcing outside the three original basins (AC2, comment 2). The four-term Korea-box residual is +2.9 m (standard error 4.0 m, $p = 0.48$), with no FDR-retained grid points. It contains unresolved interactions and ensemble noise and establishes neither the absence of interactions nor a difference in additivity between models. Text S7 and the captions would state this limitation.

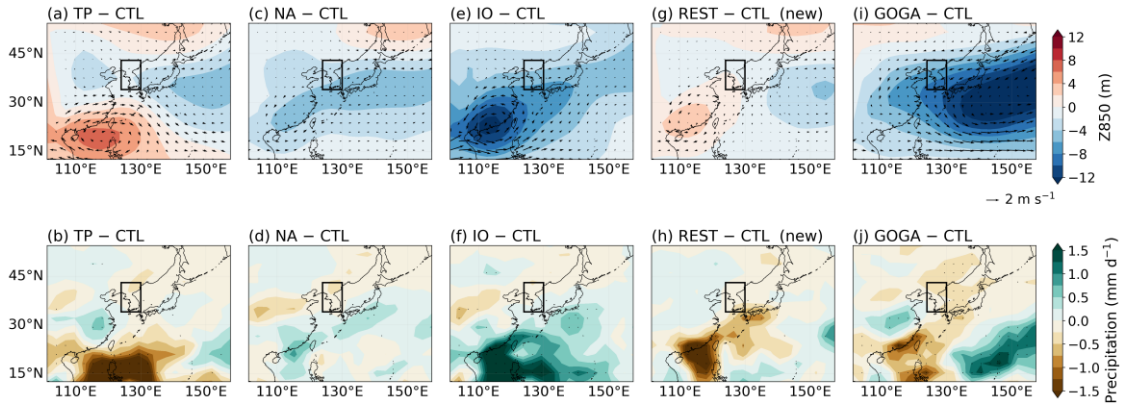


Figure 4 (redrawn; REST added). T42/ERSSTv5 JJA responses to TP, NA, IO, REST (new) and GOGA forcing (columns; 10 members each): upper row (a, c, e, g, i), Z850 (shading, m) and 850 hPa wind vectors; lower row (b, d, f, h, j), precipitation (mm d^{-1}). Over terrain above the 850 hPa surface (Tibetan Plateau) the Z850 shading is interpolated from neighbouring grid points for display only (the submitted figure showed the model's below-ground values there, median-filtered), and winds are not drawn there; dots: pointwise $p < 0.05$ and false-discovery-rate (FDR) control at 10 % over the plotted domain (grid points where 850 hPa is above ground); black box: Korea. In this model the Korea height decrease is largest under IO and NA forcing by point estimate, with a non-significant TP response. REST has no field-significant height point but field-significant precipitation over 10 % of the domain, including two Korea-box points. The weak Korean precipitation responses reflect the southward-displaced rain band. Changes from the submitted figure: FDR-controlled stippling (this comment) and the REST column (AC2, comment 2).

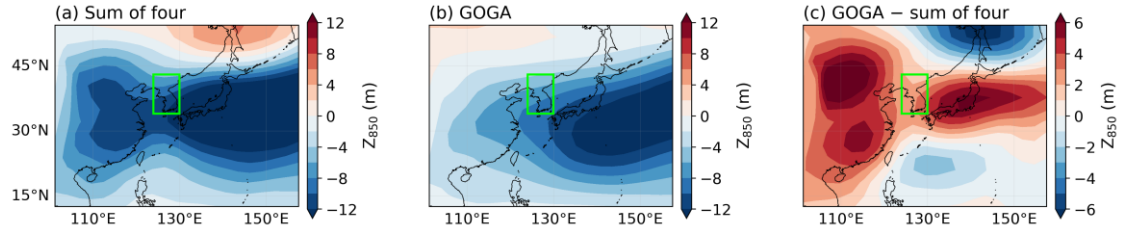


Figure 5 (redrawn; four-term decomposition). T42/ERSSTv5 Z850: (a) TP + NA + IO + REST, (b) GOGA, (c) GOGA minus the sum. Colour scales are ±12 m in (a–b), ±6 m in (c); over terrain above the 850 hPa surface the shading is interpolated for display only. The residual standard error includes the shared CTL ensemble. In (c), stippling requires both pointwise $p < 0.05$ and FDR at 10 %: no point qualifies (0.3 % pass pointwise alone). The Korea-box sum is −10.6 m against GOGA −7.7 m; its residual is consistent with additivity within ensemble uncertainty, without establishing absence of interactions. Details: AC2, comment 2.

Figure 6 shows the FDR-controlled moisture-supply response; the GOGA meridional-flux decrease is individually significant but marginal across the 16-test family ($q \approx 0.07$).

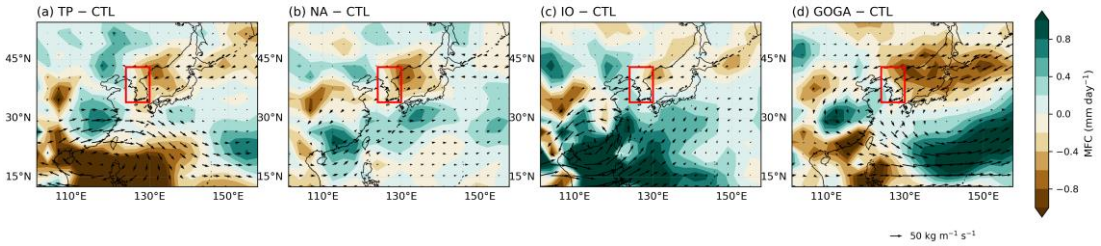


Figure 6 (redrawn). Forced JJA moisture-flux convergence (mm d^{-1}) and flux vectors (Tibet-masked), T42/ERSSTv5: (a) TP, (b) NA, (c) IO, (d) GOGA (10 members each). Stippling requires pointwise $p < 0.05$ and FDR at 10 %; box: Korea. Under GOGA the southerly moisture flux into the Korea box changes by −17.5 kg m^{-1} s^{-1} (individually $p < 0.05$; $q = 0.07$ across the 16 box-level tests), linking the circulation weakening to reduced moisture supply.

Proposed changes. Sect. 2 would define the primary family and multiplicity treatment; Sect. 3.5, the abstract and the Table S3/S4 notes would qualify significance. Figures 3–6 would be replaced by the redrawn versions shown here (Fig. 3b newly stippled), Table S7 added and Wilks (2016) cited; Sect. 3.4 would note that, within ERSSTv5, the Indian Ocean and North Atlantic cooling, but not the tropical-Pacific warming, are field-significant epoch changes, stated together with the reconstruction dependence of the basin anomalies (AC2, comment 3 : HadISST1 shows no Indian Ocean cooling and Kaplan v2 almost none, +0.01 and −0.04 K against −0.27 K in ERSSTv5) and with the smoothed-reconstruction caveat above. The original pointwise and corrected percentages are retained above so the impact of correction can be inspected directly.

Specific comment 4 — size of the Korea box

The attribution diagnostics are computed over a fairly large "Korea box" (34–43N, 124–130E) relative to the single point-station record … Some justification for this domain choice, or a sensitivity check with a tighter box, would strengthen the link…

Response. The box contains only 3 × 2 T42 grid points (34.9°, 37.7°, 40.5°N × 126.6°, 129.4°E). It was chosen to cover the Peninsula at this coarse resolution; the height response is a synoptic-scale field. We now test that the box represents Seoul's circulation.

Table S6 compares the original and tighter boxes, the nearest Seoul grid point, bilinear interpolation to Seoul, and a larger box. Z850 responses and their significance pattern change little: IO and NA have raw $p < 0.03$, TP $p > 0.2$ and GOGA $p < 0.001$ throughout. This is robustness of the point-estimate ordering; it does not resolve the adjusted pairwise differences discussed in comment 1. Wind and moisture-flux responses retain comparable signs and magnitudes, but individual-grid-point tests have larger p-values. The upstream rows document the distinct IO structure discussed in comment 1.

Table S6 (new). Sensitivity of the forced JJA response (experiment − CTL, 10 members each; value with Welch p in parentheses) to the diagnostic domain. Z850 in m, V850 in m s^{-1}, VIMFy (meridional vertically integrated moisture flux) in kg m^{-1} s^{-1}.

Field	Domain	TP	NA	IO	GOGA
Z850	Korea box (34–43°N, 124–130°E; 3 × 2 pts)	−1.8 (0.25)	−3.7 (0.019)	−5.0 (0.005)	−7.7 (<0.001)
Z850	Tight box (34–40°N, 124–130°E; 2 × 2 pts)	−1.9 (0.26)	−4.2 (0.016)	−5.7 (0.002)	−8.5 (<0.001)
Z850	Seoul grid point (37.7°N, 126.6°E)	−1.9 (0.26)	−3.5 (0.028)	−5.0 (0.007)	−7.6 (<0.001)
Z850	Seoul, bilinear (37.57°N, 126.98°E)	−1.9 (0.25)	−3.6 (0.025)	−5.0 (0.006)	−7.7 (<0.001)
Z850	Large box (30–45°N, 120–135°E; 5 × 6 pts)	−1.7 (0.21)	−3.7 (0.009)	−5.1 (0.002)	−7.5 (<0.001)
Z850	Upstream seas (25–34°N, 120–130°E)	+1.2 (0.46)	−4.3 (0.014)	−7.9 (<0.001)	−8.5 (<0.001)
V850	Korea box	−0.20 (0.057)	−0.20 (0.20)	−0.09 (0.42)	−0.35 (0.023)
V850	Tight box	−0.24 (0.046)	−0.22 (0.16)	−0.09 (0.48)	−0.39 (0.018)
V850	Seoul grid point	−0.15 (0.27)	−0.20 (0.28)	−0.10 (0.46)	−0.34 (0.064)
V850	Large box	−0.13 (0.11)	−0.11 (0.35)	−0.02 (0.82)	−0.34 (0.004)
V850	Upstream seas	−0.15 (0.14)	−0.01 (0.96)	+0.29 (0.012)	−0.50 (<0.001)
VIMFy	Korea box	−9.4 (0.052)	−10.2 (0.11)	−6.7 (0.18)	−17.5 (0.013)
VIMFy	Tight box	−10.8 (0.041)	−11.5 (0.080)	−6.4 (0.25)	−19.3 (0.009)
VIMFy	Seoul grid point	−7.8 (0.17)	−10.7 (0.15)	−7.9 (0.18)	−17.1 (0.033)
VIMFy	Large box	−5.9 (0.12)	−5.6 (0.26)	−2.7 (0.50)	−15.8 (0.004)
VIMFy	Upstream seas	−6.3 (0.16)	+0.0 (1.00)	+12.5 (0.022)	−21.3 (0.001)

Proposed changes. Sect. 2.2 would give the domain rationale and cite Table S6; Sect. 3.5 would summarize the sensitivity. The record remains a single-station observation: the domain test concerns the circulation diagnostic, not evidence for the Peninsula-wide rainfall deficit (AC2, comment 4).

Specific comment 5 — statistical power of the observational ENSO tests

The reversal of the conventional El Niño-driven explanation … rests on a short (27-year) epoch (r = −0.07 …; p ~ 0.19 for the Fisher's exact test). A brief discussion of statistical power at this sample size would help readers weigh how strongly "not significant" argues against an El Niño role.

Response. We agree: non-significance does not exclude an El Niño role. The correlation uses the full ERSSTv5 overlap, 1854–1907 (n = 54), not only the 27-year drought epoch. For r = −0.07, the Fisher-z 95 % interval is [−0.33, +0.20]; 80% power at $\alpha = 0.05$ requires $|r| = 0.37$ (0.52 for n = 27). The data therefore remain compatible with an interannual association of moderate size. The epoch-mean Niño3.4 warming itself (+0.24 K) is not field-significant within ERSSTv5 in the redrawn Fig. 3b (comment 3).

The Fisher exact test compares 5/27 versus 1/27 El Niño-like summers. Its power for that observed contrast is only 0.20; power rises to 0.62 for a 30%-versus-4% contrast and 0.89 for

40%-versus-4% (4,000 Monte Carlo trials). It does not exclude a change in frequency either. The weak-Pacific inference of the submitted paper rests on the TP pacemaker experiment, not on these tests, which are reported only to show that the epoch-mean El Niño-like state has no detectable concurrent interannual link with Seoul rainfall; Sect. 3.4 would be reworded to make this clear.

The T42 TP experiment is also limited: its minimum detectable Korea-box Z850 effect is 4.5 m at 80 % power, and its interval permits a substantial contribution (comment 1). CAM4's TP response is significant after adjustment. The revised text would therefore confine the weak-Pacific interpretation to T42, without using the observational nulls to exclude El Niño.

Proposed changes. Sect. 3.4 would give n, the correlation interval and Fisher-test power and state these limitations; Text S8 would document the power calculations. The abstract and Conclusions would reflect the model dependence (comment 2; AC2, comment 1).

Specific comment 6 — block bootstrap and serial autocorrelation

The block-bootstrap methodology and treatment of serial autocorrelation (Text S2) materially affect the robustness claims … Please summarize the approach briefly in the main text.

Response. Sect. 3.1 would summarize the moving-block bootstrap: 10,000 resamples with 5-year blocks give a 95 % interval of [226, 461] mm for the 1778–1880 minus 1881–1907 annual difference. Lag-1 autocorrelation is 0.07 for annual and −0.06 for JJA totals. Intervals for 1-, 3- and 10-year blocks are [221, 462], [212, 467] and [236, 453] mm, respectively, indicating little block-length sensitivity.

For Pettitt, AR(1) surrogates with the observed autocorrelation give $p = 0.011$ versus asymptotic $p = 0.010$; an inflated autocorrelation of 0.3 gives $p = 0.07$. The main text would also state that epoch tests use the data-selected change point and describe effect sizes; the search-corrected Pettitt value supplies the headline significance. These are the dataset v2.1 values, described under additional change 2 below.

Proposed changes. Sect. 3.1 (summary and qualification); Text S2 (block-length and AR(1) sensitivity).

Technical correction 1 — "~7 degrees too far south"

The claim that the model's summer rain band sits "~7 degrees too far south" has no citation or supporting figure. Please clarify whether this is an established bias from prior work or a diagnostic finding of this study.

Response. This is a diagnostic finding of this study, shown in the new Fig. S5. The 10-member control rain-band maximum along 124–130°E lies at 26.51°N against 33.75°N in GPCP v2.3 (1979–2020), a 7.2° southward bias. Korea-box precipitation is 3.6 versus 7.2 mm d^{-1}. The phrase "a known T42 limitation" would be removed and this diagnostic cited instead; Text S5 would report the 10-member value of 3.6 mm d^{-1} rather than 3.9 from two members.

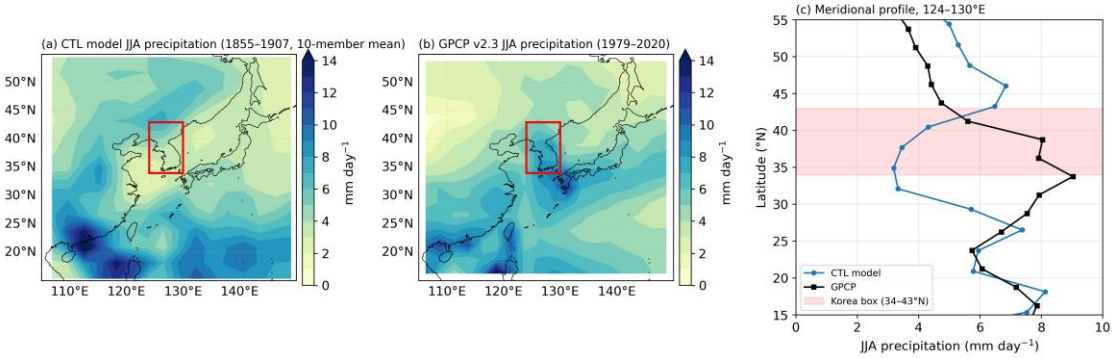


Figure S5 (new). JJA precipitation climatology: (a) 10-member CTL, 1855–1907; (b) GPCP v2.3, 1979–2020 (Adler et al., 2018); red box: Korea. (c) Profiles along 124–130°E. The maxima lie at 26.51°N and 33.75°N (7.2° apart); Korean means are 3.6 and 7.2 mm d^{-1}. This diagnosed bias motivates circulation-based attribution.

Proposed changes. Sect. 2.2, Text S5 and new Fig. S5.

Technical correction 2 — tree-ring drought atlases (Cook et al., 2010)

Please clarify whether the comparison to tree-ring drought atlases (Cook et al., 2010) was checked directly against the atlas data, or is cited from the literature.

Response. We checked the MADA data directly at its three Korean grid points (126.25°E/36.25°N, 126.25°E/38.75°N, 128.75°E/36.25°N). Mean JJA PDSI is −0.21 in 1851–1880 and +0.11 in 1881–1907 (Welch p = 0.32), without a corresponding drought signal. Correlation with Seoul JJA rainfall is r = 0.07 interannually (n = 130, p = 0.45) and −0.27 for 11-year means. Calibration R^2 is 0.32–0.38, verification R^2 0.12–0.22 and reduction of error 0.02–0.18. These values document the "limited verification skill" statement.

Proposed changes. Sect. 4 would replace "to our knowledge" with the direct comparison, epoch PDSI values, rainfall correlation and verification R^2, citing Cook et al. (2010).

Technical correction 3 — cross-references to the Supplement

Please confirm that all cross-references to Supplementary Tables/Text (S1–S8, Table S1–S5) are consistent with the final Supplement.

Response. We checked the submitted cross-references mechanically: the main text cites Texts S1–S7, Tables S1–S5 and Figs. S1–S4, all defined in the Supplement. Text S8 would be new. The revised Supplement would contain Texts S1–S8, Tables S1–S7 and Figs. S1–S5 from this response together with the items listed in AC2. Before submitting the assembled revision, we will check its contents list and all in-text references again against the final Supplement.

Additional changes made on our own initiative (not requested by the referee)

We report two further changes for transparency. Both would be documented in the revised manuscript and in the updated deposit (v2.1), which will be made public with the revised manuscript.

1. The drought outlasts the chronicle (Sect. 3.2). We extended the Seoul series with the Korea Meteorological Administration record at Seoul (station 108; KMA, 2026): daily data taken directly from the KMA portal for 1908–1925 and 1979–2024 and, for 1926–1978 and 2025, from the KMA-derived modern segment of the Seoul daily series compiled by Wang et al. (2026; dataset https://doi.org/10.12197/2026GA010), which is identical to the portal data on every overlapping day and month. Wang et al. (2026), merging a Cheugugi restoration with the KMA record, find no consistent discontinuity at the 1907 junction in June–September totals, R95p intensity or Rx5day, though some daily-intensity indices (RII, R20mm, Rx1day) shift there, so the comparison here rests mainly on annual totals. Under different observers and a different gauge, annual precipitation remained below the pre-drought mean for six more years (1908–1913: 1005 mm yr^{-1} and 0.7 days yr^{-1} at ≥100 mm, against 1230 mm yr^{-1} before 1881 and 889 mm with 0.8 days yr^{-1} in 1881–1907). The merged 1778–2025 series has its strongest Pettitt change at the 1913/1914 recovery ($p < 10^{-4}$), and its driest 27-year window is 1887–1913. This six-year extension weakens, without excluding, observational-practice decay as an explanation for earlier decades. Sect. 3.2 would add it as a fourth line of evidence, with the sources in Sect. 2 and Wang et al. (2026) added to the references (other early stations: AC2, comment 4).

Reference: Wang, Y., Wei, Y., Yao, Z., Yang, S., Shi, F., and Lin, W.: A 256-year daily rainfall record reveals multiscale rainy-season variability and emerging hydroclimatic hazard over northern East Asia, Sci. Adv., 12, eaeh6716, https://doi.org/10.1126/sciadv.aeh6716, 2026 (daily dataset: https://doi.org/10.12197/2026GA010).

2. Dataset refinement (Cheugugi-Seoul v2.1). Since submission we have retrieved the complete chronicle entries for the adjudicated days wherever the page is available (database excerpts had truncated long entries and sometimes omitted a second reading) and reprocessed them at the event level, with an independent audit that no fall is counted twice. Version 2.1 changes 490 days, about 1 % of the record (360 from the complete chronicle entries or the author's direct readings, 130 neighboring days receiving or losing a night share); the day-by-day change log in the deposit documents each change against the chronicle entry and both earlier reconstructions. The results are unchanged: annual means 1230 versus 889 mm (−27.7%; previously 1251 versus 904, also −27.7%), Pettitt change point 1880 ($p = 0.010$; previously 0.011) and ≥100 mm contribution −41% (previously −43%); the attribution experiments do not use the daily data. All values in this response are v2.1 values. The revised manuscript and the Zenodo deposit would use v2.1 throughout, with the day-by-day change log, scripts and audit alongside v2.

Proposed changes. Sect. 2.1 and Text S1 would describe the processing. Figs. 1, S1–S2 and Tables S1–S2 would be regenerated, and the revised manuscript would use v2.1 throughout. The day-by-day change log, scripts and audit accompany v2.1 alongside v2 in Zenodo. The auxiliary permutation-test and printed-change-point errors found during code checking were corrected and logged; manuscript values were unaffected (AC2, comment 5), and the robustness bounds for the heavy-rainfall component are in AC2, comment 3.