Response to Anonymous Referee #2 (egusphere-2026-4812)

We would like to express our gratitude to the referee for identifying three limits to the attribution: dependence on one model, one SST reconstruction, and an incomplete basin decomposition. We have completed additional experiments and analyses addressing all five comments. The responses below distinguish completed analyses from proposed manuscript changes, which will be incorporated when the editor requests the revision, together with those arising from Referee #1 (AC1). In line with the editorial office's instruction, the revised manuscript will be prepared and submitted when the editor requests it; the changes requested by Referee #1 and those below are therefore all described as proposals. Section references are to the submitted manuscript; figure and table numbering is provisional.

The additional experiments are (i) CAM4 in CESM 2.2, six experiments × 10 members, 1854–1907; (ii) T42 REST, forcing SST outside the three basin masks, 10 members, 1854–1920; and (iii) T42 with COBE-SST2, full-ocean (11 members) and three single-basin experiments (10 members each), 1854–1920. Their principal result: the full-forcing Korea-box Z850 decrease of the T42/ERSSTv5 experiments (−7.7 m, $p < 0.001$) is reproduced by CAM4 (−6.8 m, $p < 0.001$) and, at about half the amplitude, under COBE-SST2 (−3.8 m, $p = 0.03$), and the SST outside the three basins gives a height response indistinguishable from zero in both models; the basin partition is not reproduced, as no single basin is significant after Holm adjustment in more than one configuration. The attribution is therefore restricted to the circulation weakening, with the basin ranking presented as a property of the T42/ERSSTv5 experiments; neither model reproduces the Korean rain band. Comments 1–3 give the evidence and limitations; comments 4–5 address spatial extent and reference periods.

Unless stated otherwise, responses are Korea-box (34–43°N, 124–130°E) JJA (1881–1907) − (1855–1880), experiment minus the same change in CTL. Parenthetical p-values are unadjusted Welch p-values. Holm-adjusted p_H applies within each configuration's four-forcing Z850 family (TP, NA, IO, GOGA), as in AC1 comment 3, and separately over the three pairwise basin contrasts. For each new configuration the full-ocean test is the single confirmatory test of replication and the single-basin tests are exploratory; the Holm-adjusted values are also given so that the stricter four-test standard of AC1 can be applied throughout (under it the COBE-SST2 full-ocean response, $p = 0.03$, has $p_H = 0.12$ and is not significant). Maps use pointwise $p < 0.05$ together with Benjamini–Hochberg (BH) false-discovery-rate (FDR) control at 10 % over the plotted domain (Fig. 3b: over the global ocean 60°S–60°N; AC1, comment 3) (Wilks, 2016), independently of box-mean adjustment. Proposed figures and tables are shown once, under the comment they principally address, with cross-references elsewhere. A marked-up manuscript and Supplement will accompany the revised files.

Comment 1 — model dependence of the attribution; circulation versus moisture variables; the displaced rain band; attribution of the circulation weakening versus the drought

The conclusion that the Indian Ocean and North Atlantic were the primary drivers of monsoon circulation weakening is mainly based on a single T42 atmospheric model and the 850 hPa geopotential height response. However, the relative contributions of different ocean basins vary across circulation and moisture transport variables. Moreover, the model's summer rainband is displaced approximately 7° southward, limiting its ability to reproduce local

precipitation. The authors should further assess the robustness of the attribution, clarify its model dependence, and distinguish between the attribution of circulation weakening and that of the observed drought.

Response. We agree. The independent-model experiments reproduce the full-forcing circulation weakening but not the basin partition, and strengthen the distinction between circulation attribution and the observed rainfall deficit.

(a) **Second-model design (new Text S10)**. We repeated CTL, TP, NA, IO, GOGA and REST with CAM4 physics in CESM 2.2 (CIME compset long name 1850_CAM40_CLM50%SP_CICE5%PRES_DOCN%DOM_SROF_SGLC_SWAV_SIAC_SESP, as recorded in the case files; finite-volume core, 1.9° × 2.5°, 26 levels; CLM5 with prescribed phenology). Other forcings remain at 1850 values. SST is CAM4's climatology plus the same ERSSTv5 monthly anomalies (1854–1907 reference) and tanh-tapered masks as in T42, regridded to CAM4; sea ice is climatological and SST is clipped at freezing. Ten members start on 1–10 January 1854; 1854 is discarded. The CTL ensemble itself has a Korea-box Z850 epoch difference of +1.3 ± 0.7 m (mean ± standard error over its 10 members; p = 0.11 against zero, i.e. not distinguishable from internal variability), which the experiment-minus-CTL differencing removes.

(b) **Results (Table S10, Fig. S6)**. The Korea-box responses are given in Table S10(a).

Table S10(a). Table S10(a), proposed. T42/ERSSTv5 and CAM4/ERSSTv5 responses under the conventions above; 10 members each; unadjusted Welch p in parentheses and, for Z850, p_H, the Holm-adjusted p within the four-forcing family (TP, NA, IO, GOGA; REST is outside the family).

Forcing	Z850 (m): T42 / CAM4	V850 (m s^{-1}): T42 / CAM4	Sea-level pressure (hPa): CAM4	Precipitation (mm d^{-1}): T42 / CAM4
TP	−1.8 (0.25; p_H 0.25) / −2.8 (0.008; p_H 0.024)	−0.20 (0.057) / −0.25 (0.004)	−0.29 (0.018)	−0.16 (0.09) / −0.17 (0.051)
NA	−3.7 (0.019; p_H 0.039) / −0.9 (0.37; p_H 0.37)	−0.20 (0.20) / −0.02 (0.84)	−0.12 (0.31)	−0.15 (0.12) / −0.01 (0.90)
IO	−5.0 (0.005; p_H 0.014) / −2.3 (0.039; p_H 0.079)	−0.09 (0.42) / −0.06 (0.45)	−0.27 (0.032)	−0.01 (0.92) / −0.08 (0.41)
GOGA	−7.7 (<0.001; p_H <0.001) / −6.8 (<0.001; p_H <0.001)	−0.35 (0.023) / −0.26 (<0.001)	−0.69 (<0.001)	−0.26 (0.014) / −0.09 (0.16)
REST	−0.1 (0.96) / +0.8 (0.46)	−0.21 (0.075) / −0.10 (0.13)	+0.14 (0.23)	−0.24 (0.014) / −0.15 (0.17)

CAM4's full-forcing height, southerly-wind and sea-level-pressure decreases each have p < 0.001; its Z850 response is 88 % of T42's, with all ten member changes negative. Under CAM4, TP is the only basin significant after adjustment; IO has a similar point estimate but p_H = 0.079, and NA is small. No pairwise basin contrast is significant after adjustment (Table S10b). The T42 order of point estimates therefore does not establish a ranking across models. Under GOGA, TP and IO forcing, Fig. S6d–h shows the same height depression over the East China Sea–Korea–Japan region as in the T42 model (Fig. 4). After map FDR control, 15/15 Korea-

box grid points qualify for GOGA, 14/15 for TP and 5/15 for IO; NA and REST have none over or near Korea; the qualifying NA points lie south of 25°N and the REST points south of 25°N or over the Pacific east of 145°E. Map results do not alter the box-mean p-values.

Table S10(b). Pairwise Korea-box Z850 contrasts and decomposition (m; unadjusted p and Holm-adjusted p_H over the three contrasts). Residual uncertainty includes shared CTL covariance; comment 2 gives the REST definition and field-specific results.

Z850, Korea box (m)	T42 / ERSSTv5	T42 / COBE-SST2	CAM4 / ERSSTv5
Contrast IO − NA	−1.4 (0.30; p_H 0.31)	−0.2 (0.90; p_H 0.90)	−1.4 (0.19; p_H 0.39)
Contrast IO − TP	−3.2 (0.033; p_H 0.100)	−3.1 (0.028; p_H 0.084)	+0.5 (0.59; p_H 0.59)
Contrast NA − TP	−1.9 (0.15; p_H 0.31)	−2.9 (0.031; p_H 0.084)	+1.9 (0.056; p_H 0.17)
Sum of three (TP + NA + IO)	−10.5	−2.1	−5.9
Sum of four (+ REST)	−10.6	–	−5.1
GOGA	−7.7	−3.8	−6.8
Residual, GOGA − sum of three	+2.8	−1.7	−0.8
Residual, GOGA − sum of four	+2.9 (SE 4.0; 95 % CI [−5.6, +11.5]; p = 0.48)	–	−1.6 (SE 2.6; 95 % CI [−7.1, +3.9]; p = 0.55)

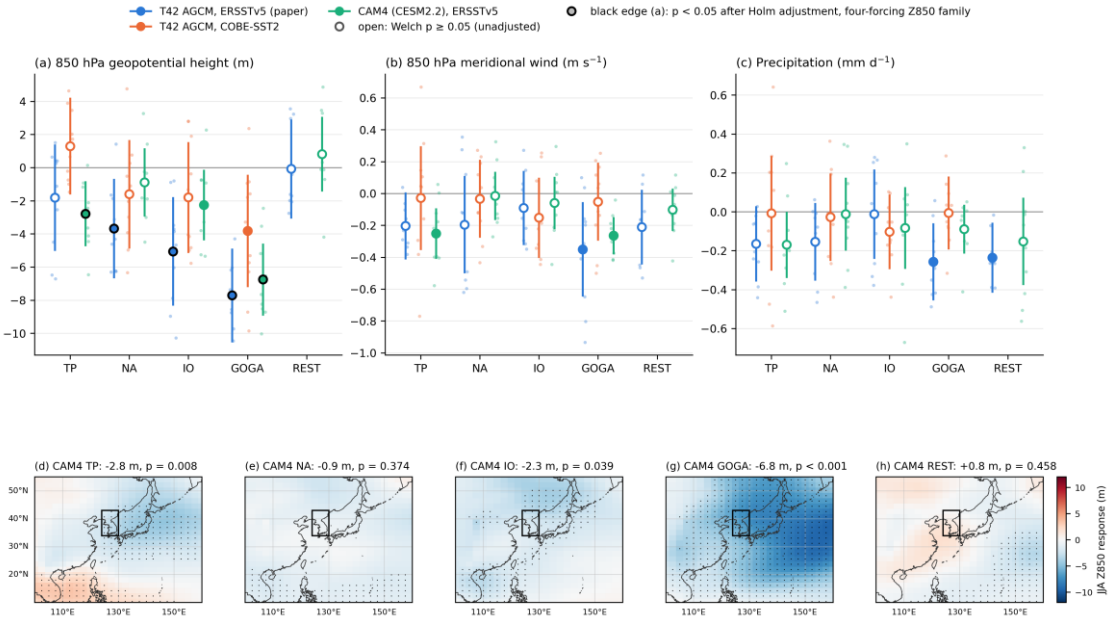


Figure S6 (new). Korea-box JJA responses, (1881–1907) − (1855–1880), experiment − CTL, in the three configurations: T42 AGCM with ERSSTv5 (the manuscript), T42 AGCM with COBE-SST2, and CAM4 (CESM 2.2) with ERSSTv5. (a) 850 hPa geopotential height, (b) 850 hPa meridional wind, (c) precipitation; small dots, members; large symbols, ensemble means with 95% Welch intervals; filled, Welch $p < 0.05$; open, $p \geq 0.05$; black edge in (a), $p < 0.05$ after Holm adjustment within the four-forcing Z850 family. (d–h) CAM4 JJA Z850 response to each forcing (shading over terrain above the 850 hPa surface interpolated for display); dots, significant pointwise ($p < 0.05$) and after control of the false discovery rate (FDR) at 10% (Wilks, 2016); black box, the Korea diagnostic region (34–43°N, 124–130°E). Panels (a–c) are proposed as a main-text figure.

(c) **What survives the additional tests**. The full-forcing Korea-box Z850 point estimate is negative in all three configurations: −7.7 m (T42/ERSSTv5), −6.8 m (CAM4/ERSSTv5) and −3.8 m (T42/COBE-SST2). The last has p = 0.03 as a single confirmatory test but p_H = 0.12 within the four-test family and no FDR-significant map points (comment 3). Under ERSSTv5, both models also weaken the southerly wind and the subtropical-high intensity index ((f) below). Under COBE-SST2 neither the wind, the precipitation nor the intensity index shows a statistically resolved response; only the eastward shift of the 500 hPa western edge is resolved under COBE-SST2 ((f) below).

The basin partition is unresolved across models and reconstructions: after adjustment, IO and NA are significant only in T42/ERSSTv5, TP only in CAM4/ERSSTv5, and no basin in T42/COBE-SST2. IO and NA height point estimates are negative in all three configurations, whereas TP changes sign under COBE-SST2; a shared sign is not a shared significant effect. Accordingly the manuscript would no longer present the Indian Ocean and North Atlantic lead over the Pacific as the result: the abstract and Conclusions would report the full-forcing circulation response, present the basin ranking as a property of the T42/ERSSTv5 experiments, and state the partition's model/reconstruction dependence. The T42 weak-Pacific finding is not reproduced by CAM4; the Late Victorian El Niño cannot be excluded as a contributor.

(d) **Field dependence**. Table S10(a) confirms the referee's point. In T42, TP and NA give the largest single-basin V850 point estimates (−0.20 m s^{-1} each, p = 0.057 and 0.20). In CAM4, TP's wind response is nearly GOGA's (−0.25 versus −0.26 m s^{-1}, p = 0.004), whereas IO lowers height with a small, unresolved wind response (−0.06 m s^{-1}, p = 0.45). As explained in AC1 comment 1, geostrophic wind depends on the height gradient: IO produces a broad depression with a weak gradient at Korea, while TP's depression lies farther east. Thus the Pacific has greater relative weight in wind than in height in both models, by point estimate.

(e) **Rain-band bias and scope of attribution**. T42's JJA rain-band maximum lies at 26.51°N against 33.75°N in GPCP (7.2°; Fig. S5). CAM4 has no distinct band along 124–130°E: rainfall has a minimum of 3.1 mm d^{-1} near 29°N and increases toward 41°N. Korea-box means are 3.6 mm d^{-1} (T42) and 4.6 (CAM4), against GPCP's 7.2. Neither model places the observed band over Korea; no single-basin precipitation response reaches p < 0.05 (CAM4 TP is closest at 0.051), and T42's full-forcing decrease is not reproduced by CAM4 (Table S10a).

We would separate (i) the observed Seoul drought, supported by the gauge and independent records; (ii) circulation weakening registered by 20CRv3 and reproduced under ERSSTv5 in two models, with a smaller height decrease under COBE-SST2; and (iii) the inferred rainfall link, supported by T42/ERSSTv5 moisture-flux diagnostics (Fig. 6) and consistency with the gauge, rather than adequate rainfall simulation. The proposed title is "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; AC1 comment 2).

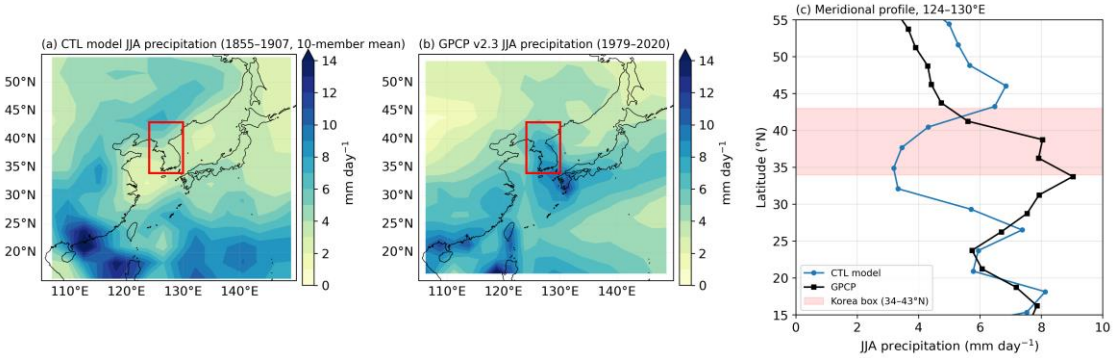


Figure S5 (new). T42 JJA rainfall climatology: (a) 10-member CTL, 1855–1907; (b) GPCP v2.3, 1979–2020 (Adler et al., 2018); red box, Korea diagnostic region. (c) Profiles along 124–130°E. The simulated rain-band maximum is 7.2° south of GPCP's; the Korea-box mean is half the observed value. This bias limits interpretation of simulated local rainfall.

(f) **Does Korea-box height diagnose the subtropical high?** We added the Wang et al. (2013) intensity index (JJA Z850 averaged over 15–25°N, 115–150°E), in full and eddy form; the latter removes the zonal mean at each latitude (He et al., 2015; Huang et al., 2016). Fig. S8 and Table S12 below give the results. Under full ERSSTv5 forcing, the index decreases by 6.0 m in T42 and 4.8 m in CAM4; the eddy components are −4.2 and −3.0 m, respectively (70% and 63% of the full responses; all four p_H < 0.001). The intensity index and its eddy form thus indicate a weaker western North Pacific subtropical high under ERSSTv5 in both models; the Korea-box decrease accompanies this broader circulation response, while including a substantial zonal-mean component (below). COBE-SST2 gives no statistically detectable weakening (full −0.7 m, p = 0.31; eddy approximately zero).

The single-basin index point estimates are IO and NA negative in all three configurations, and TP positive in both T42 configurations but near zero in CAM4, where it is not statistically resolved. Table S12 shows their adjustment explicitly; in CAM4 neither IO nor NA is significant (p_H = 0.15), so these are not established basin-specific effects across models. IO cooling (−0.26 K in the prescribed ERSSTv5 field, comment 3(b); −0.27 K in the raw ERSSTv5 box mean, Table S9a) has the sign expected through the capacitor pathway (Xie et al., 2009). T42's positive TP index response must not be identified with Wang et al.'s central-Pacific mechanism, which strengthens the high under central-Pacific cooling: our TP forcing is a spatially heterogeneous pattern with Niño3.4 warming (+0.24 K as prescribed), not an isolated central-Pacific perturbation. It is reported as a model result.

The Korea-box decrease includes a substantial zonal-mean component, −4.0 of −7.7 m in T42 and −4.3 of −6.8 m in CAM4; the eddy remainders, −3.7 m (p = 0.006, p_H = 0.025) and −2.5 m (p = 0.015, p_H = 0.060), are the part specific to East Asia. The zonal-mean lowering extends from 10°S to 55°N (T42) or 70°N (CAM4) with maxima at 34–43°N; on a common grid it combines a surface-pressure fall (Z1000 −4.8 m) with a thickness decrease (−1.3 m, layer cooling of about −0.25 K), and it is distinct from the high's zonal asymmetry (Text S11). For TP the Korea-box response is mostly zonal-mean (−1.4 of −1.8 m; −2.0 of −2.8 m), which helps explain why the Korea-box and index partitions differ. We would report eddy height alongside full Korea-box height.

Position indices are secondary. The CTL-anchored 500 hPa contour moves eastward by +6.7°, +2.9° and +8.6° (T42/ERSSTv5, T42/COBE-SST2, CAM4/ERSSTv5; p_H ≤ 0.017), but after

removal of the drought-epoch zonal-mean height change the shifts are +1.4°, +0.4° and −1.0° ($p \geq 0.09$), so they are not isolated displacements of the asymmetric circulation; changing the contour value (±5 m) or search band alters the magnitudes by about a factor of two without changing sign. Only T42/ERSSTv5 has a significant southward ridge shift (−1.6°, $p_H = 0.04$). The Korea-box decrease is part of a broad same-signed response across the East Asian sector (100–150°E: −7.9 m T42, −6.5 m CAM4) that strengthens eastward, from about −3.5 m at 110°E to −12.5 (T42) or −8.7 m (CAM4) at 150°E, a gradient consistent with the geostrophic wind weakening. The proposed text would identify the weakening of the intensity index, in full and eddy form, under ERSSTv5, distinguish the zonal-mean contribution, and retain position changes only as qualified diagnostics (Text S11, Table S12).

References: Wang, B., Xiang, B., and Lee, J.-Y., PNAS, 110, 2718–2722, 2013; Xie, S.-P., et al., J. Climate, 22, 730–747, 2009; He, C., et al., Sci. Rep., 5, 16771, 2015; Huang, Y., Li, X., and Wang, H., Clim. Dyn., 47, 567–577, 2016.

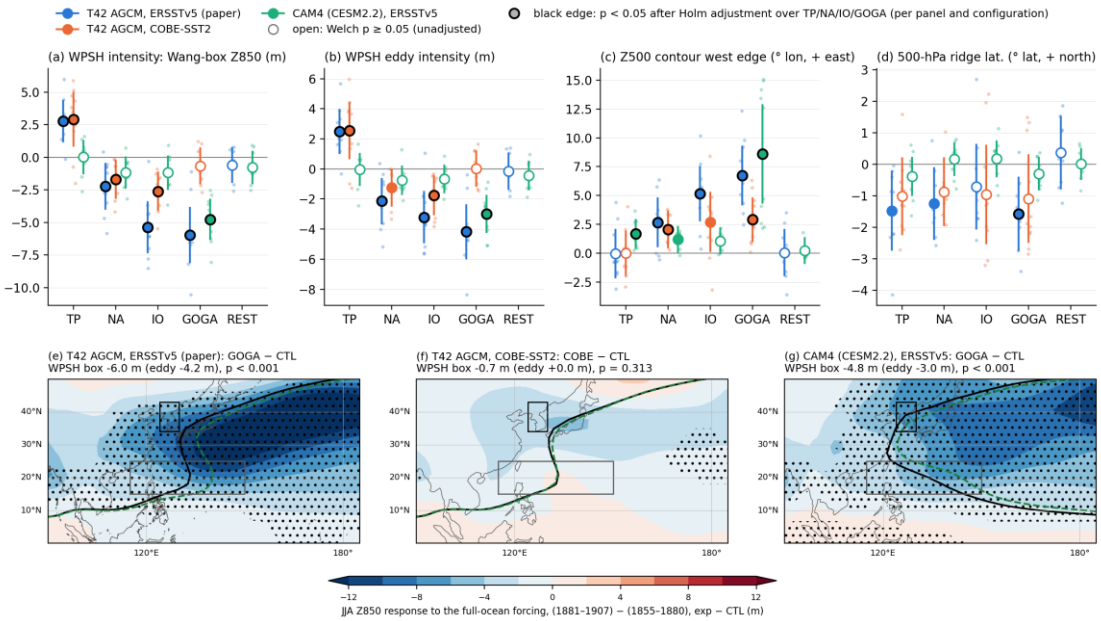


Figure S8 (new). Subtropical-high diagnostics under the response conventions above. (a) Wang et al. (2013) intensity index; (b) its eddy form; (c) westernmost longitude of the 500 hPa contour whose value is the CTL-climatology mean over 15–25°N, 115–150°E (search: 100–180°E, 10–40°N; positive eastward); (d) latitude of maximum 500 hPa height at each longitude 110–150°E, then averaged (positive northward). Dots, members; large symbols, means and 95 % Welch intervals; filled/open, unadjusted $p < / \geq 0.05$; black edge, Holm $p_H < 0.05$ within each panel/configuration's four-forcing family. (e–g) Full-ocean Z850 responses (shading interpolated for display over terrain above the 850 hPa surface): black contour, CTL index-box mean; green dashed, same value in CTL plus response; boxes, Korea and intensity-index regions; stippling, pointwise $p < 0.05$ and false-discovery-rate (FDR) 10%. Positions in (c–d) are definition-sensitive secondary diagnostics.

Table S12. Intensity, eddy-intensity, position and Korea-box eddy-height responses. Parentheses give p_H within each index/configuration's four-forcing family; REST values are unadjusted p. Full Korea-box height is in Table S10(a). Symbol fill in Fig. S8 conveys unadjusted significance; exact unadjusted p-values for quantities discussed in the text are given above. Position indices are secondary and definition-sensitive.

Index	Configuration	TP	NA	IO	GOGA	REST
Intensity (Wang et al., 2013 box; m)	T42 / ERSSTv5	+2.8 (0.003)	−2.2 (0.014)	−5.4 (<0.001)	−6.0 (<0.001)	−0.6 (0.35)
	T42 / COBE-SST2	+2.9 (0.024)	−1.7 (0.043)	−2.6 (0.005)	−0.7 (0.31)	–
	CAM4 / ERSSTv5	+0.0 (0.98)	−1.2 (0.15)	−1.2 (0.15)	−4.8 (<0.001)	−0.8 (0.19)
Eddy intensity (zonal mean removed; m)	T42 / ERSSTv5	+2.5 (0.004)	−2.1 (0.007)	−3.2 (0.003)	−4.2 (<0.001)	−0.1 (0.79)
	T42 / COBE-SST2	+2.5 (0.040)	−1.2 (0.084)	−1.8 (0.040)	+0.0 (0.97)	–
	CAM4 / ERSSTv5	−0.0 (0.95)	−0.7 (0.31)	−0.7 (0.31)	−3.0 (<0.001)	−0.4 (0.32)
500 hPa western edge (° lon; + = east)	T42 / ERSSTv5	−0.0 (0.96)	+2.7 (0.030)	+5.1 (<0.001)	+6.7 (<0.001)	+0.1 (0.95)
	T42 / COBE-SST2	−0.0 (0.99)	+2.1 (0.042)	+2.7 (0.075)	+2.9 (0.017)	–
	CAM4 / ERSSTv5	+1.7 (0.029)	+1.2 (0.064)	+1.0 (0.064)	+8.6 (0.004)	+0.2 (0.68)
500 hPa ridge latitude (° lat; + = north)	T42 / ERSSTv5	−1.5 (0.065)	−1.3 (0.065)	−0.7 (0.27)	−1.6 (0.040)	+0.4 (0.50)
	T42 / COBE-SST2	−1.0 (0.37)	−0.9 (0.37)	−1.0 (0.37)	−1.1 (0.37)	–
	CAM4 / ERSSTv5	−0.4 (0.75)	+0.2 (1.00)	+0.2 (1.00)	−0.3 (0.75)	+0.0 (0.98)
Korea-box eddy Z850 (m)	T42 / ERSSTv5	−0.4 (0.77)	−3.6 (0.026)	−2.3 (0.20)	−3.7 (0.025)	+0.8 (0.51)
	T42 / COBE-SST2	+2.0 (0.42)	−1.3 (0.71)	−0.5 (0.71)	−2.0 (0.54)	–
	CAM4 / ERSSTv5	−0.8 (0.86)	−0.7 (0.86)	−1.0 (0.86)	−2.5 (0.060)	+1.5 (0.20)

Proposed changes. Remove basin attribution from the title; revise the abstract and Conclusions to separate the observed drought, SST-forced circulation response and inferred rainfall link, and to state the basin partition's model/reconstruction dependence. The Conclusions would close by naming the steps that would resolve the partition instead of leaving it as a caveat: ensembles of about 30–40 members per experiment (with 10 members the pairwise basin contrasts have standard errors of 1.2–1.4 m, so a 2 m contrast is detectable at 80% power only with three to four times as many members), pairwise-forcing experiments (TP+IO, TP+NA, NA+IO; comment 2(b)) to test additivity directly, and repetition with a third atmospheric model and with reconstructions in which the Indian Ocean cooling is absent or nearly so (HadISST1, Kaplan v2; comment 3). Add CAM4 methods to Sect. 2.2/Text S10; add "Model and reconstruction dependence" and full/eddy subtropical-high diagnostics to Sect. 3.5/Text S11. Present Fig. S6(a–c) as a main-text comparison, with CAM4 maps, Fig. S8 and Tables S10/S12 in the Supplement; give eddy height alongside Korea-box full height and qualify position indices by their sensitivity to the zonal mean.

Comment 2 — the Fig. 5 residual: SST outside the three basins, basin interactions

Figure 5 compares the sum of individual ocean-basin responses with the global SST-forced response to evaluate nonlinear interactions. However, the global experiment also includes SST anomalies outside the three selected basins. Therefore, the difference cannot be attributed entirely to nonlinear interactions. The authors should clarify this limitation, further discuss possible interactions among ocean basins, and consider additional experiments or a more cautious interpretation.

Response. We agree that the original three-basin residual combined remote-SST effects with interactions and ensemble noise. We added REST in both models and redefined the decomposition with all four forcing regions. Its residual is not significant in either model; this does not establish absence of interactions or a difference in additivity between models.

(a) **Forcing and REST response.** The three mask rectangles cover 48 % of the global ocean, and the SST outside them carries 47 % of the area-weighted squared JJA forcing amplitude, chiefly South Atlantic (−0.16 K), North Pacific north of 35°N (−0.12 K) and South Pacific (−0.09 K) cooling (prescribed T42 fields, JJA 1881–1907 minus 1855–1880, ocean points; regions defined in Text S7, raw ERSSTv5 values in Table S9a). REST prescribes GOGA minus the three basin anomalies, with ten members in each model (comment 1); the four forcings close to GOGA within 0.03 % (T42) and 2.5 % (CAM4, after boundary-condition preprocessing) in RMS. Fig. S4 shows the forcing and Fig. 4 the response.

The REST responses (point estimates in Table S10a; intervals in the first table below) are not resolved for height: the T42 interval [−3.1, +2.9] m permits a remote contribution of about 40 % of GOGA in either direction, and CAM4's is [−1.4, +3.1] m, so negligible remote forcing is not established. In wind and rainfall the REST point estimates are comparable to GOGA's, and the T42 rainfall response is significant (−0.24 mm d^{-1}, p = 0.014), which further limits a basin attribution of moisture delivery.

REST and decomposition results for proposed Table S5. The first table below includes the REST intervals; the second extends Table S10(b)'s height decomposition to wind and rainfall. Responses use the conventions above. CAM4's height residual and interval are in Table S10(b); no REST experiment was run for COBE-SST2.

Field (T42, ERSSTv5)	TP	NA	IO	REST (p)	95 % CI of REST	GOGA
Z850 (m)	−1.8	−3.7	−5.0	−0.1 (0.96)	[−3.1, +2.9]	−7.7
V850 (m s^{-1})	−0.20	−0.20	−0.09	−0.21 (0.075)	[−0.45, +0.02]	−0.35
Precip. (mm d^{-1})	−0.16	−0.15	−0.01	−0.24 (0.014)	[−0.42, −0.06]	−0.26

Field (T42, ERSSTv5)	Sum of 3	Sum of 4	GOGA	GOGA − sum of 4	SE	95 % CI	p
Z850 (m)	−10.5	−10.6	−7.7	+2.9	4.0	[−5.6, +11.5]	0.48
V850 (m s^{-1})	−0.49	−0.70	−0.35	+0.35	0.33	[−0.33, +1.03]	0.30
Precip. (mm d^{-1})	−0.33	−0.57	−0.26	+0.31	0.26	[−0.24, +0.86]	0.25

Because every response shares the same CTL, the residual GOGA − TP − NA − IO − REST contains the CTL mean with coefficient 3 (variance coefficient 9); its degrees of freedom follow Welch–Satterthwaite (Text S7). The wind/rainfall sums exceed GOGA by about a factor of two, but their residuals are not significant and the decomposition is poorly constrained.

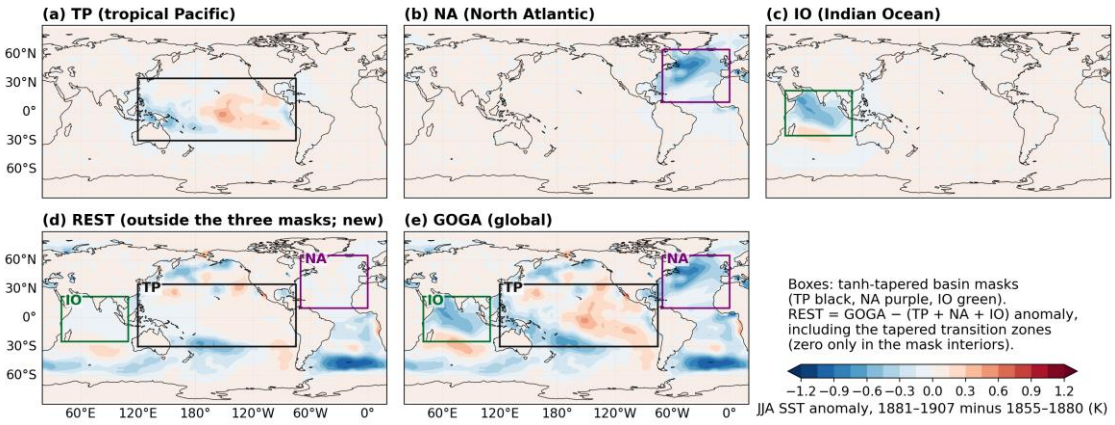


Figure S4 (proposed revision). Prescribed JJA SST anomaly of each pacemaker experiment (1881–1907 minus 1855–1880, K), as read by the model: (a) TP, (b) NA, (c) IO, (d) REST (new; the GOGA anomaly minus the three single-basin anomalies, i.e. the forcing outside the three masks including their tapered transition zones) and (e) GOGA. The anomalies are relative to the 1854–1907 monthly climatology (comment 5) and added to the model climatology. Boxes outline the tanh-tapered basin masks (TP black, NA purple, IO green). Changes from the submitted figure: panel (d) added; anomaly reference corrected.

In Fig. 4, REST height has no FDR-significant point and no pointwise-significant Korea-box point. Its weak South China/Indochina height rise and North Pacific decrease have amplitudes around 2 m. REST rainfall decreases over the East China Sea, Korea and southeastern China, with 10 % of the plotted domain and two of six Korea-box points qualifying after FDR control. These are distinct responses: a detectable rainfall effect does not require a detectable height effect in this ensemble.

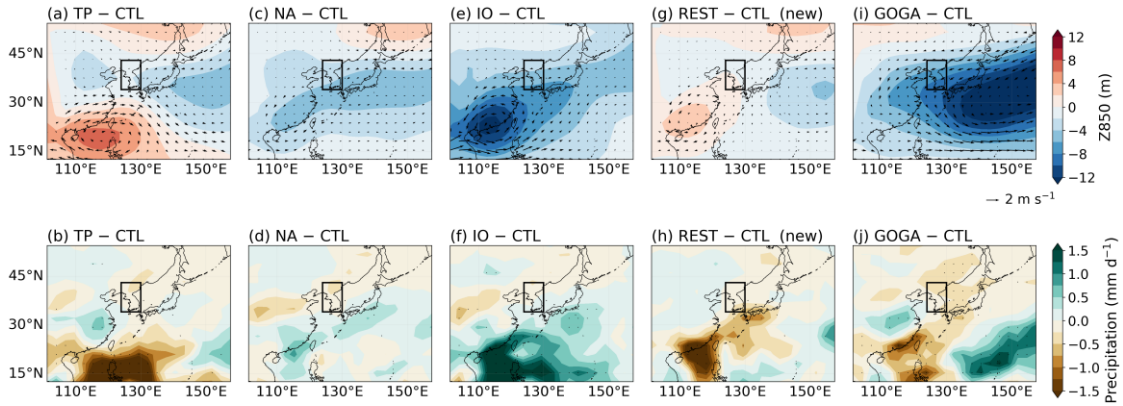


Figure 4 (proposed revision). T42/ERSSTv5 JJA responses, ten-member ensembles, under the response conventions above. Columns: TP, NA, IO, REST and GOGA. Top row: Z850 (m) and 850 hPa winds; bottom row: precipitation (mm d^{-1}). Over terrain above the 850 hPa surface (Tibetan Plateau) the Z850 shading is interpolated from neighboring 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 FDR 10 % over the plotted domain (grid points where 850 hPa is above ground); black box, Korea. Revisions are the REST column (this comment) and FDR-controlled stippling (AC1 comment 3). Exact Korea-box estimates are in Table S10(a).

(b) **Interactions and interpretation**. The complete residuals have opposite signs, but neither is significant over the Korea box, the T42 residual map has no field-significant point, and their difference is not resolved either (SEs 4.0 and 2.6 m). We would describe both as compatible with additivity within ensemble uncertainty, without establishing absence of interactions or labeling one model sub-additive and the other additive. The submitted interpretation of a saturating high response is only a possibility. Pairwise-forcing experiments (TP+IO, TP+NA, NA+IO), needed to isolate interactions, were not run and remain a next step. A single REST experiment also cannot establish that remote SST is unnecessary; the wide intervals explicitly allow a substantial contribution. The three-basin attribution would no longer be called complete.

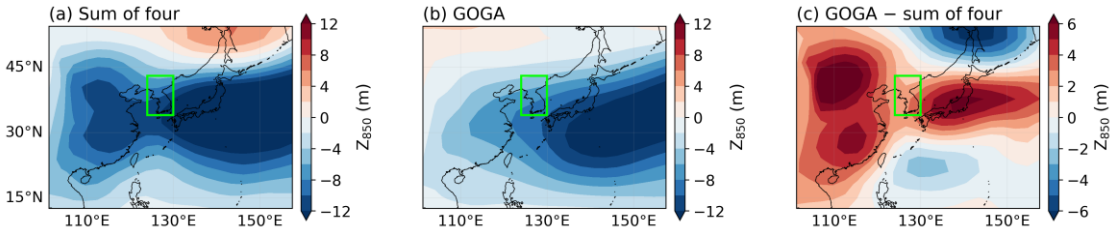


Figure 5 (proposed revision). T42/ERSSTv5 Z850 response: (a) TP + NA + IO + REST; (b) GOGA; (c) GOGA minus the four-term sum. Color scales are ±12 m in (a,b), ±6 m in (c); over terrain above the 850 hPa surface the shading is interpolated for display only. In (c), stippling requires pointwise $p < 0.05$ and FDR 10 %, including shared-control covariance: no point qualifies (0.3 % qualify pointwise). Korea-box estimates and uncertainty are in the decomposition tables above. REST and FDR control are the changes from the submitted figure.

Proposed changes. The revised Figs. 4, 5 and S4 displayed above already include REST and the complete four-term sum; these would replace the submitted figures in the revised manuscript. Define REST in Sect. 2.2; revise Sect. 3.5/Text S7 to report complete sums, residual uncertainty and the limits on interaction inference. Extend Tables S5/S10 with REST and shared-control residuals in both models and include CAM4 REST in Fig. S6. Confine basin

attribution to height, while noting unresolved remote-SST contributions and the need for pairwise experiments.

Comment 3 — uncertainty of the nineteenth-century SST reconstruction; provisional gauge values; extremes

Ocean observations during the nineteenth century were sparse, potentially introducing uncertainties into the ERSSTv5 reconstruction and the resulting attribution. Some historical precipitation records also contain provisional values. The authors should further assess how these uncertainties affect the SST-forced responses and estimated drought characteristics, particularly changes in extreme precipitation.

Response. SST reconstruction uncertainty materially affects the forcing amplitude, response and basin partition. Repeating the experiments with COBE-SST2 retains a negative full-forcing Korea-box height point estimate but does not establish a reconstruction-independent basin ranking. Separately, all candidate-value treatments of provisional gauge days retain the observed annual and heavy-rainfall deficits.

SST reconstruction

(a) **Reconstruction spread (new Text S9, Table S9a).** The four-dataset comparison below (ERSSTv5, HadISST1, Kaplan v2 and the uninterpolated HadSST4, median realization) uses JJA 1881–1907 minus 1856–1880. All datasets make the Indian Ocean and North Atlantic cooler relative to Niño3.4, but the contrast amplitudes differ by factors of two to three; absolute anomalies can change sign (HadISST1 has no Indian Ocean cooling). Box means here are from the raw datasets with the 1856–1880 baseline; the values 'as prescribed' in comments 2 and 3(b) are ocean-point means of the T42 fields read by the model with the 1855–1880 baseline, which agree with the raw means to within about 0.03 K.

Table S9(a), **proposed.** JJA SST change, 1881–1907 minus 1856–1880 (K), in four reconstructions: ERSSTv5, HadISST1, Kaplan v2 and the uninterpolated HadSST4 (median realization). Rows are box means of the raw datasets (boxes as in Fig. 3b and comment 2; South Atlantic 40°S–10°N, 290–360°E); the last two rows are differences from Niño3.4.

Box, JJA 1881–1907 minus 1856–1880 (K)	ERSSTv5	HadISST1	Kaplan v2	HadSST4
Niño3.4	+0.26	+0.25	+0.15	+0.07
Indian Ocean box (25°S–22°N, 40–110°E)	−0.27	+0.01	−0.04	−0.26
North Atlantic box (10–65°N)	−0.32	−0.07	−0.07	−0.20
South Atlantic	−0.19	+0.12	+0.11	−0.16
Global 60°S–60°N	−0.17	+0.03	−0.01	−0.10
Indian Ocean minus Niño3.4	−0.53	−0.23	−0.19	−0.33
North Atlantic minus Niño3.4	−0.58	−0.32	−0.23	−0.27

HadSST4 agrees in basin-anomaly sign with ERSSTv5 but does not validate its amplitudes: bias-adjustment and sampling uncertainties remain. HadSST4 coverage rises between epochs from 35% to 59% of 5°-month cells in the Indian Ocean, 48% to 7 % in the North Atlantic and 13% to 33% in the tropical Pacific. Restricting IO to cells observed in both epochs reduces its

change from −0.26 K to −0.11 to −0.19 K, depending on the minimum month count. Sampling ERSSTv5 with HadSST4 coverage reduces Niño3.4 from +0.26 to +0.09 K; Atlantic changes move by <0.03 K. These tests support a factor-of-two-or-more amplitude uncertainty, not a uniquely validated reconstruction. The redrawn Fig. 3b (AC1, comment 3) now stipples the ERSSTv5 epoch difference where it is field-significant (Welch test of the 27 against 27 JJA means; FDR 10% over the global ocean 60°S–60°N): the Indian Ocean and North Atlantic cooling qualify over 62% and 43 % of their boxes, the Niño3.4 warming over 10%; given the smoothness of the reconstruction these fractions are upper bounds, and they describe sampling variability within ERSSTv5 only – HadISST1 shows no Indian Ocean cooling and Kaplan v2 almost none (+0.01 and −0.04 K against −0.27 K; table above).

(b) **COBE-SST2 experiments (Text S9, Table S9b, Fig. S7).** COBE-SST2 samples the weak end of this forcing range. It uses optimal interpolation toward background fields in sparse regions; we have not established that this causes its weaker anomalies. In the files actually prescribed (ocean-point means), its Niño3.4/IO/NA/global epoch changes are +0.10/−0.09/−0.11/−0.08 K, versus ERSSTv5 +0.24/−0.26/−0.32/−0.14 K (34–56% of the ERSSTv5 amplitudes). The forcing patterns correlate at 0.42 over the 60°S–60°N ocean, with RMS ratio 0.52, so this tests a structurally different, weaker pattern rather than simple rescaling or the full uncertainty range. We ran the full-ocean COBE experiment and TP/NA/IO_COBE, otherwise as for T42/ERSSTv5.

Table S9(b), proposed. Korea-box JJA responses of the T42 model to COBE-SST2 forcing (experiment − CTL, 1881–1907 minus 1855–1880; unadjusted Welch p in parentheses; full-ocean experiment 11 members, single-basin experiments 10 each), with the ERSSTv5-forced responses of Table S10(a) in the last column for comparison.

Field	TP_COBE	NA_COBE	IO_COBE	COBE (full ocean)	ERSSTv5 set: TP / NA / IO / GOGA
Z850 (m)	+1.3 (0.36)	−1.6 (0.32)	−1.8 (0.27)	−3.8 (0.030; n = 11)	−1.8 / −3.7 / −5.0 / −7.7
V850 (m s^{-1})	−0.03 (0.86)	−0.03 (0.78)	−0.15 (0.22)	−0.05 (0.67)	−0.20 / −0.20 / −0.09 / −0.35
Precip. (mm d^{-1})	−0.01 (0.96)	−0.03 (0.80)	−0.10 (0.28)	−0.01 (0.95)	−0.16 / −0.15 / −0.01 / −0.26

The full-ocean height response is −3.8 m (p = 0.03 as one confirmatory test; p_H = 0.12 in the four-test family), about half T42/ERSSTv5. Its pattern (Fig. S7d) resembles the ERSSTv5 map but has no FDR-significant point, against 70% of the domain under ERSSTv5: the evidence under COBE-SST2 is the box-mean decrease, not a field-significant circulation anomaly. Wind and precipitation are not statistically resolved, so the circulation-to-moisture link is supported under ERSSTv5 only. No single basin is significant ($|Z850| \leq 1.8$ m, $p \geq 0.27$). The IO−TP and NA−TP contrasts are negative by point estimate (unadjusted p = 0.03 each, p_H = 0.084), IO−NA is not resolved, and the three-basin sum (−2.1 m) has the same sign as, and about half of, the full response (−3.8 m; Table S10b, comment 1). The IO−NA contrast narrows from −1.4 m under ERSSTv5 to −0.2 m under COBE-SST2, under which both basin anomalies are damped to about one-third (IO −0.09 against −0.26 K, NA −0.11 against −0.32 K, as prescribed), and is not statistically resolved in either reconstruction (p_H = 0.31 and 0.90); what the two reconstructions share, by point estimate, is a larger non-Pacific than Pacific contribution to the Korea-box height decrease. The revised Sect. 3.5 would state the basin

ranking as a T42/ERSSTv5 result and add that under COBE-SST2 no basin is significant and the Indian Ocean and North Atlantic are not separable, alongside the reconstruction spread.

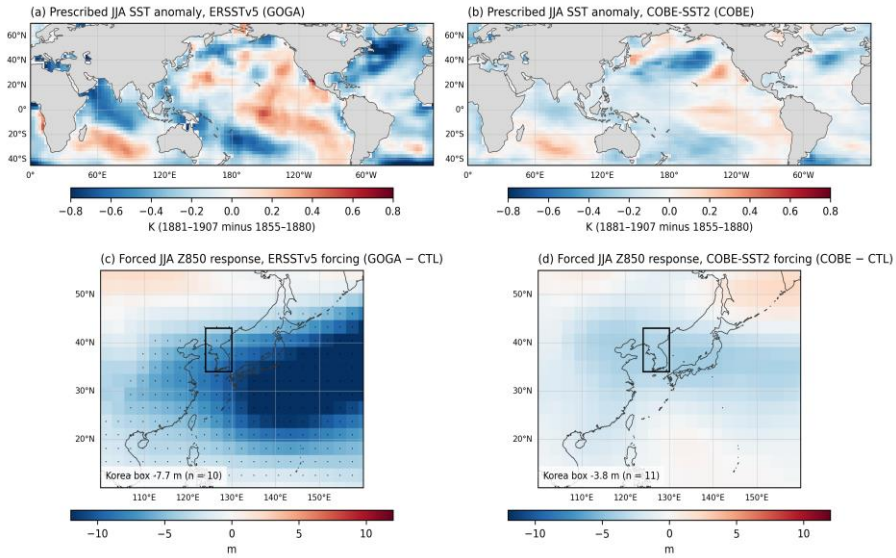


Figure S7 (new). ERSSTv5 and COBE-SST2 forcing and the T42 responses. (a, b) Prescribed JJA SST anomaly (1881–1907 minus 1855–1880, K) of the full-ocean experiment under ERSSTv5 (GOGA) and under COBE-SST2 (COBE), as read by the model. (c, d) Forced JJA Z850 response (experiment − CTL, m) of the T42 model to each forcing (shading over terrain above the 850 hPa surface interpolated for display); dots, significant pointwise ($p < 0.05$) and after control of the false discovery rate at 10 % (Wilks, 2016).

Provisional gauge values and extremes

(c) **Candidate-value bounds (extended Text S1)**. Of 47,845 reconciled days, 526 (1.1 %) are provisional because prior reconstructions disagree and the chronicle does not settle the amount or could not be retrieved. Candidate values are the chronicle-derived v2.1 value, Jhun–Moon, Lee et al. and, when available, the online database (chronicle totals that span midnight are event totals, not single-day candidates); they define the following sensitivity bounds, not a bound on all possible measurement errors.

Table. Sensitivity bounds for extended Text S1 (proposed). Seoul annual mean and ≥100 mm d^{-1} contribution, 1778–1880 versus 1881–1907, under alternative treatments of the 526 provisional days; candidate values from the chronicle, Jhun and Moon (1997), Lee et al. (2024) and the online database (v2.1 as released in the first row). In every treatment the heavy-rain contribution falls more than the annual total (its share of the annual total declines from about 15% to about 12%), the heavy-event dominance reported in Sect. 3.1.

Treatment of the 526 provisional days	Annual mean, 1778–1880 / 1881–1907 (mm)	Change	≥100 mm d^{-1} contribution (mm yr^{-1})	Change
v2.1 as released	1230 / 889	−28 %	182 / 107	−41 %
all at their maximum candidate	1261 / 953	−24 %	186 / 111	−40 %
all at their minimum candidate	1194 / 854	−28 %	182 / 100	−45 %
most anti-drought: minimum before 1881, maximum after	1194 / 953	−20 %	182 / 111	−39 %
most pro-drought: maximum before 1881, minimum after	1261 / 854	−32 %	186 / 100	−46 %
provisional days excluded (set to zero)	1190 / 850	−29 %	182 / 100	−45 %

Only one of the released ≥100 mm days is provisional (after 1881; counts are 142 before and 21 after), and only four provisional days have any candidate ≥100 mm. Across treatments, annual rainfall decreases by 20–32 %, ≥100 mm contribution by 39–46 %, and ≥100 mm days per year by 44–47 % (released values 1.38 to 0.78 d yr^{-1}). Seoul's subsequent instrumental record has 0.7 such days per year in 1908–1913 versus 0.8 in 1881–1907, consistent with an end-of-record deficit under different observers/gauge; its lack of overlap prevents validation of the chronicle's values (AC1, additional change 1).

(d) **Scope of extreme-rainfall inference.** These are observational sensitivity tests. The models are not used to attribute extremes, because neither simulates Korea's rain band adequately (comment 1e). The manuscript would retain the possible role of tropical cyclones, which a monthly-mean circulation response does not address.

Proposed changes. New Text S9 with Table S9 (a: four-reconstruction comparison; b: COBE-SST2 responses), the coverage test and Fig. S7; Fig. 3b (FDR-controlled stippling; AC1, comment 3); Sect. 2.2 (COBE-SST2 experiments defined); Sect. 3.4 (reconstruction spread; field significance of the basin changes); Sect. 3.5 (closing subsection; the basin ranking is stated as conditional on the reconstruction); abstract and Conclusions; Text S1 extended with the table above; Sect. 3.1 (one sentence: the heavy-rainfall deficit is insensitive to the provisional values, with the bounds).

Comment 4 — a single station

The drought analysis relies primarily on the Seoul Cheugugi record. A single station cannot establish whether the drought extended across the Korean Peninsula or the broader East Asian monsoon region.

Response. We agree: one station cannot establish the drought's peninsula-wide extent. Before 1904 the other quantitative Korean rainfall measurements known to us are the fragmentary provincial Cheugugi records recovered from the Gaksa Deungnok by Kim et al. (2012) for Wonju, Hamheung and Haeju (9, 13 and 18 years between 1831 and 1900, March–October) and the customs-station observations begun in the 1880s at the treaty ports (Jemulpo, Busan, Wonsan), which we have not yet obtained. Compared with each seat's own earlier values the provincial records lean dry, but they are too fragmentary to establish the spatial extent of the drought in 1881–1903 (see (c) below). The early instrumental stations document the later tail at Seoul and Incheon, with weaker evidence at Mokpo and no detected deficit at Busan; they cannot establish the earlier footprint.

(a) **Early stations (new Table S11a)**. Incheon, Busan and Mokpo observations begin in 1904; Seoul's begin in late 1907. The table compares complete calendar years of the 1904–1913 tail (AC1, additional change 1) with the first 27 recovery years, 1914–1940; the annual series therefore begin in 1905 at Incheon and Busan and in 1906 at Mokpo, whose 1904–1905 records are incomplete, whereas JJA is complete from 1904 at all

three ports. The eight station/season comparisons are exploratory and p-values are unadjusted Welch values.

Station (KMA id)	Annual: tail years	Tail mean (mm)	1914–1940 mean (mm)	Change (Welch p)	JJA change (Welch p)
Seoul (108)	1908–1913	1005	1322	−24 % (0.001)	−22 % (0.06)
Incheon (112)	1905–1913	878	1115	−21 % (0.019)	−27 % (0.021)
Mokpo (165)	1906–1913	948	1103	−14 % (0.12)	−24 % (0.07)
Busan (159)	1905–1913	1422	1413	+1 % (0.95)	+5 % (0.76)

Seoul and Incheon show annual deficits; Mokpo's same-signed summer deficit is uncertain, and Busan's estimates are slightly positive. This supports a deficit beyond Seoul during the tail, particularly on the western side of the peninsula, consistent with weakened southwesterly moisture supply over the Yellow Sea. It does not determine the extent of 1881–1903.

(b) **Regional circulation**. The 20CRv3 sea-level-pressure change (Fig. 3a) – a weakened monsoon low over Mongolia and North China (about +1 hPa, $p < 10^{-4}$, 96% of that region field-significant) with anomalous easterlies over the Yellow and East China Seas – is a regional-scale circulation anomaly, and the SST-forced height depression in both models extends over the East China Sea, Korea and Japan (Fig. 4, Fig. S6). The circulation evidence thus concerns the northern East Asian monsoon region, not Seoul alone, which is the reason the attribution is made for the circulation.

(c) **Other records and proxies (proposed Table S11b).** The provincial Cheugugi series of Kim et al. (2012) include drought-epoch summers at all three seats (Wonju 1888, 1889, 1892, 1893; Hamheung 1888–1890, 1897, 1898, 1900; Haeju 1888, 1889, 1897–1899). Because the seat totals are systematically lower than Seoul's in the same years (July means 149–321 mm against 353–447 mm; Kim et al., Table 6), which may reflect incomplete preservation of the daily entries as much as climate, we compare each seat's June–August values only with its own pre-1881 values of the same month (Table S11b). They lean dry but are not conclusive: 17 of the 25 drought-epoch seat-months with at least three reference values lie below the seat's pre-1881 mean (two-sided sign test $p = 0.11$; median anomaly −26%; pooled total −6%). The deficit is clearest at Wonju and Hamheung, whereas Haeju is dominated by the wet summer of 1899; in 1888, the severe drought year identified by Kim et al., June and August were far below normal at Seoul and at the seats alike. Month by month, however, the seats covary only weakly with Seoul (pooled correlations 0.30–0.39, $n = 28$ per month), and the seat and Seoul anomalies share their sign in only 12 of the 25 months. The provincial records are therefore compatible with, but do not establish, an extension of the Seoul deficit to the three seats; the revised Sect. 4 would say so, citing Kim et al. (2012) and Table S11b. MADA does not corroborate the Korea drought (mean JJA PDSI +0.11 against −0.21 in 1851–1880), but its Korean verification skill is low (R^2 0.12–0.22); we have not examined its grid points outside Korea or Chinese documentary dryness/wetness indices, and the revised Conclusions would retain that gap explicitly.

Table S11(b), proposed. June–August totals at the three provincial seats (Kim et al., 2012, Table 4) in the drought epoch (seat years 1888–1900) against each seat's pre-1881 values of the same month, and the Seoul totals of the same years against the Seoul 1778–1880 mean of that month. Asterisks mark reference samples of fewer than three years, which are excluded from the pooled count; a complete June–August row is possible only for Haeju (nine reference years).

Seat, month	Pre-1881 years (mean, mm)	Drought-epoch years (mean, mm)	Change	Years below the pre-1881 mean	Seoul, same years: change from 1778–1880 (years below)
Wonju, Jun	3 (93)	2 (71)	−24 %	2 of 2	−27 % (1 of 2)
Wonju, Jul*	2 (498)	3 (203)	−59 %	3 of 3	+15 % (1 of 3)
Wonju, Aug*	2 (412)	2 (116)	−72 %	2 of 2	−20 % (1 of 2)
Hamheung, Jun	3 (93)	6 (66)	−28 %	5 of 6	−22 % (4 of 6)
Hamheung, Jul	5 (158)	2 (127)	−19 %	1 of 2	+18 % (1 of 2)
Hamheung, Aug	5 (81)	5 (86)	+6 %	3 of 5	−65 % (5 of 5)
Haeju, Jun	11 (89)	3 (51)	−43 %	3 of 3	−2 % (1 of 3)
Haeju, Jul	12 (300)	4 (314)	+4 %	2 of 4	+2 % (1 of 4)
Haeju, Aug	11 (223)	3 (249)	+12 %	1 of 3	−37 % (3 of 3)
Haeju, JJA	9 (638)	2 (800)	+25 %	1 of 2	−21 % (2 of 2)

Reference: Kim, S.-W., Park, J.-S., Kim, J.-A., and Hong, Y.: Restoration of 19th-century Chugugi rainfall data for Wonju, Hamheung and Haeju, Korea, Atmosphere (Korean Meteorological Society), 22, 129–135, https://doi.org/10.14191/Atmos.2012.22.1.129, 2012 (in Korean with English abstract).

Proposed changes. Sect. 3.2 (one paragraph on the four early stations, with Table S11a); Sect. 4 (limitation sentence rewritten as above; the provincial Cheugugi comparison of Table S11b summarised with its caveats); new Table S11 (a: early stations; b: provincial Cheugugi); KMA (2026) reference extended to the four stations; Kim et al. (2012) added to the references.

Comment 5 — reference periods and documentation of the statistics

Different reference periods are used for the precipitation analysis, SST composites, and numerical experiments. The authors should clarify the rationale for these choices and examine whether the main results remain robust under alternative reference periods. The statistical treatment of the change-point detection and significance tests should also be clearly documented.

Response. The baselines differ for stated reasons: the gauge uses its whole pre-drought record (1778–1880); for 20CRv3, whose subset extends back to 1806, the submitted manuscript uses the three decades immediately preceding the change point (1851–1880), because the pressure network over East Asia is sparsest in the early decades (Sect. 3.3 shows that the weakened-low signal is, if anything, stronger over the best-observed later decades); ERSSTv5 begins in 1854; and the pacemaker integrations start in 1854 with the first year discarded (baseline 1855–1880). All use the same drought epoch (1881–1907) and the same end of the baseline (1880, the Pettitt change point). We would state this rationale in Sect. 2 and show in a new Table S8 that the

main results do not depend on the choice; for 20CRv3 the alternatives tested are shorter baselines within 1851–1880, not the sparsely observed pre-1851 decades.

(a) **Observational sensitivity.** Table S8a shows that the reanalysis/SST/model-era baselines give annual and JJA declines of about 30–32%, slightly larger than the full-record 28%, because 1851–1880 was a wet period; the adopted gauge baseline is the conservative choice. Bootstrap intervals remain positive and heavy-rainfall contribution decreases by 33–39% under the late baselines. Across all tested alternatives the weakened continental monsoon low remains +0.9 to +1.1 hPa ($p < 10^{-4}$), with IO/NA cooling and Niño3.4 warming retained. The small +0.05 K Niño3.4 contrast against the 1870s shows amplitude sensitivity.

Table S8 (proposed), part (a). Sensitivity of the observational results to the reference period (drought epoch 1881–1907 throughout; Welch p in brackets; block-bootstrap interval as in Sect. 3.1). Additional baselines computed: precipitation 1808–1880 −31.3%, 1831–1880 −30.2%; SLP 1866–1880 +0.99, 1851–1870 +0.78 hPa; SST 1866–1880 Niño3.4 +0.22, IO −0.29, NA −0.40 K.

Quantity (drought epoch 1881–1907 minus the baseline)	1778–1880 (paper)	1851–1880 (20CR)	1854–1880 (SST)	1855–1880 (model)	1861–1880	1871–1880
Seoul annual precipitation change (%) [Welch p]	−27.7 [8×10^{-7}]	−31.2 [2×10^{-5}]	−31.8 [2×10^{-5}]	−31.4 [5×10^{-5}]	−31.1 [7×10^{-4}]	–
Annual deficit, block-bootstrap 95 % CI (mm)	[226, 461]	[265, 504]	[263, 514]	[257, 513]	[262, 538]	–
Seoul JJA precipitation, change (%)	−27.8	−30.5	−31.7	−30.6	−32.0	–
≥100 mm d^{-1} contribution, change (%)	−41	−39	−35	−33	−38	–
20CRv3 JJA SLP change, Mongolia–N. China (35–50°N, 100–125°E) (hPa) [p]	–	+0.89 [1×10^{-6}]	–	–	+0.96 [7×10^{-7}]	+1.12 [4×10^{-6}]
ERSSTv5 JJA SST, Niño3.4 (K)	–	–	+0.24	–	+0.31	+0.05
ERSSTv5 JJA SST, Indian Ocean box (K)	–	–	−0.28	–	−0.27	−0.16
ERSSTv5 JJA SST, North Atlantic box (K)	–	–	−0.31	–	−0.37	−0.42

(b) **Experiment sensitivity**. Table S8b contains seven baseline/epoch combinations. In T42/ERSSTv5, GOGA height remains negative and significant (−7.3 to −9.6 m, $p < 0.001$); IO/NA remain the largest single-basin point estimates and TP the smallest. Their mutual order is not resolved. These tests qualify the T42/ERSSTv5 result and do not establish its partition across configurations. Wind sensitivity is included in the table caption: some alternatives do not reach $p < 0.05$.

Table S8 (proposed), part (b). Sensitivity of the pacemaker Korea-box Z850 response (m; experiment − CTL, 10 members each, unadjusted Welch p in parentheses) to the baseline and drought epoch. V850 (m s^{-1}), GOGA: −0.35 (0.02), −0.33 (0.03), −0.25 (0.11), −0.48 (0.01), −0.29 (0.07), −0.48 (<0.01), −0.26 (0.06) for the seven rows in order. The Indian Ocean value −5.1 m in the first row is the same response as the −5.0 m quoted elsewhere (−5.05 m before rounding).

Baseline → drought epoch	TP	NA	IO	GOGA
1855–1880 → 1881–1907 (paper)	−1.8 (0.25)	−3.7 (0.019)	−5.1 (0.005)	−7.7 (<0.001)
1861–1880 → 1881–1907	−2.5 (0.14)	−4.3 (0.008)	−4.7 (0.010)	−9.0 (<0.001)
1866–1880 → 1881–1907	−2.8 (0.10)	−4.5 (0.007)	−4.5 (0.020)	−9.3 (<0.001)
1855–1870 → 1881–1907	−1.0 (0.57)	−3.1 (0.080)	−6.7 (0.001)	−7.4 (<0.001)
1855–1880 → 1881–1900	−2.8 (0.08)	−3.9 (0.014)	−5.2 (0.002)	−7.3 (<0.001)
1855–1880 → 1887–1907	−1.1 (0.48)	−3.8 (0.030)	−5.3 (0.009)	−9.6 (<0.001)
1855–1880 → 1881–1913	−0.5 (0.69)	−3.8 (0.015)	−4.6 (0.004)	−7.9 (<0.001)

(c) **A correction**. The prescribed SST anomalies are relative to the 1854–1907 monthly mean of ERSSTv5, not 1854–1920 as written in the submitted Sect. 2.2, Text S5 and the Fig. S4 caption. This corrects the description, not the experiments: the boundary files actually used carry the 1854–1907 reference, so no result changes. (The reference adds a constant for each calendar month to the prescribed anomaly of every experiment and cancels in the epoch differences of the forcing; we do not claim that a model forced with a different absolute SST would respond identically, only that the experiments were run as now described.)

(d) **Statistical documentation**. Text S2 covers the two-sided Pettitt test (search-corrected asymptotic p and 5,000 AR(1) surrogates with observed lag-1 correlation), change-point stability under observational noise, and epoch contrasts (Student t, Mann–Whitney, 20,000 permutations and 5-year moving-block bootstrap, 10,000 resamples, with block-length sensitivity). Because epoch contrasts use the data-selected change point, they are interpreted as effect-size statements rather than independent confirmation of the break.

The new Text S8 (AC1, comment 3) covers Welch ensemble tests, Holm adjustment for primary Z850 and pairwise families, Benjamini–Hochberg q-values across box tests, map FDR control at 10% over the plotted domain (Wilks, 2016), and power. New tests follow the conventions in the introduction; decomposition uncertainty includes shared CTL covariance (comment 2). Sect. 2 would summarize which hypotheses and families these address, with pointers to the detailed methods.

During deposit checks, we corrected auxiliary v2.1 script v21_numbers.py: it drew epoch means from two permutations instead of splitting one and printed the Pettitt year one index early. The main script supplying manuscript values was correct. The corrected auxiliary script gives permutation $p < 10^{-4}$ and a break after 1880, and is the version included in the v2.1 deposit.

Proposed changes. Sect. 2 (rationale for the baselines; one-paragraph summary of the statistical treatment); Sect. 2.2, Text S5 and Fig. S4 caption (anomaly reference corrected to 1854–1907); new Table S8 (a, b); Text S8 (new; AC1) extended with the items above.

Supplement items to be added in response to this review

New: Text S9 (SST reconstructions, coverage test and COBE-SST2 experiments), Text S10 (CAM4 experiments), Text S11 (subtropical-high indices); Tables S8–S12; Figs. S6–S8. Fig. S5, Text S8 and Tables S6–S7 are proposed in AC1.

Revised: Figs. 3 (b stippled), 4, 5 and S4, Text S7 and Table S5 (REST; residual with standard error). Fig. S6(a–c) is proposed for the main text; numbering will be finalized when assembling the revision. Contents lists and in-text cross-references will be checked mechanically (AC1 technical correction 3). Boundary conditions, namelists, per-member Korea-box series and analysis scripts for the new simulations will enter a new Zenodo version.