

Supplementary Materials

for

Seasonal Predictability of Baroclinic Wave Activity

Gan Zhang^{1,2*}, Hiroyuki Murakami^{2,3}, William F. Cooke^{2,3},

Zhuo Wang⁴, Liwei Jia^{2,3}, Feiyu Lu^{1,2}, Xiaosong Yang², Thomas L. Delworth²,

Andrew T. Wittenberg², Matthew J. Harrison², Mitchell Bushuk^{2,3}, Colleen McHugh^{2,5},

Nathaniel C. Johnson², Sarah B. Kapnick^{2†}, Kai-Chih Tseng^{1,2}, and Liping Zhang^{2,3}

¹Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey

²Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton,
New Jersey

³University Corporation for Atmospheric Research, Boulder, Colorado

⁴Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois

⁵SAIC, Science Applications International Corporation, Reston, Virginia

*Now at Citadel Enterprise Americas, LLC.

† Now at J.P. Morgan Chase & Co.

Supplementary Note 1

The extreme events in Figure 1 had been reported by media and/or examined by past studies. These events were selected to represent events of different nature and at diverse geographic locations. We highlight their associations with baroclinic waves and breaking Rossby waves but refrain from conducting any extensive analyses. Interested readers can refer to Supplementary Table 1, which points to more details of these extreme events.

Supplementary Note 2

Here we compare the LC1 and LC2 events in the observation and climate model simulations. As our simulation datasets are exceptionally large, we randomly selected an ensemble member and analyzed composites of Rossby wave breaking related to LC1 and LC2 events in January–March 2001. The event composites in the reanalysis and simulations show similar spatial scales and patterns, suggesting the models can reasonably simulate the atmospheric flow during the life cycle of baroclinic waves. This feature is also key for simulating and predicting LC1- and LC2-related weather extremes. Meanwhile, the events in the reanalysis and simulations also show interesting differences. For example, the event frequency in the d4PDF ensemble member is lower than that in the ERA-Interim reanalysis and the SPEAR prediction. The difference remains even after we consider the difference in the data temporal resolution (6-hourly vs 12-hourly). The undercount in this d4PDF ensemble member appears to be consistent with the negative biases of this particular large ensemble (Fig. 2). Whether the issue is related to the lack of atmosphere-ocean coupling and potential SST-RWB feedback (e.g., Zhang and Wang 2019) remains to be determined. Meanwhile, the event frequency in the SPEAR prediction is similar to that in ERA-Interim.

However, the high-latitude events in the SPEAR prediction appear to be associated with more extreme PV values.

Further analyses suggest that the composite patterns of wave breaking do not strongly depend on the ensemble member or the time period. However, the frequency of LC1 and LC2 events is subject to influences of unforced variability in the model climate system. The consistency of event frequency between ERA-Interim and model simulations depends on seasons and regions (e.g., Supplementary Figure 4).

Supplementary Note 3

We extend the discussion of predictable components (PCs) in the d4PDF large ensemble simulation by examining more individual regions (Supplementary Figures 2–3). For brevity, the regions and the corresponding acronyms are denoted in the figures and their captions. When all 100 ensemble members are considered, the PCs (red) of CN-PAC LC1 and N-ATL LC1 range between ~ 0.5 and ~ 0.8 throughout the year, suggesting $\sim 25\%$ to $\sim 64\%$ of year-to-year variations could be simulated with the observed SST forcings. The PCs of NE-PAC LC1 demonstrate a strong time dependence, with an extremely high value (~ 0.9) during January–March. Given a perfect prediction of SST, the high PC value would then imply high predictability of LC1-related atmospheric river events on the West Coast of the United States. In comparison, AUS-NZ LC1 events show much lower PC values (~ 0.2 to 0.3), suggesting much less predictability. Nonetheless, LC1 events in the subtropics of the Southern Hemisphere (e.g., South America and South Africa) are highly predictable when the SST is known (Fig. 2e–h).

The PCs of LC2 events are generally below 0.5, indicating that the high-latitude baroclinic wave activity is not strongly regulated by SST forcings. Besides N-PAC LC2 events discussed in the main text, AN-SO LC2 events also show relatively high values during October–December. More interestingly, the AN-SO LC2 events also show characteristics of “the signal-to-noise paradox”, namely the PC value being lower than the prediction skill (Eade et al. 2014). However, both the PC value and the prediction skill of AN-SO LC2 events are strongly associated with long-term trends. Detrending the event counts reduces the PC value of AN-SO LC2 events to ≤ 0.35 throughout all the seasons. The detrending treatment also reduces the prediction skill of AN-SO LC2 during January–March to ~ 0.1 , well below new PC values (not shown). The trend in the high-latitude Southern Hemisphere might be particularly pronounced because of changes in the ozone forcing and the related circulation responses. Overall, AN-SO LC2 suggests the PC-based findings are sensitive to long-term trends. This sensitivity has interesting implications for predictability and skill assessments, as well as projections of future risks associated with baroclinic wave activity.

After accounting for the case of AN-SO LC2, long-term trends do not have substantial impacts on the estimates of PC values and prediction skills in the other analyzed cases. Moreover, PC values are generally below or comparable to the prediction skills shown in Supplementary Figures 2–3. It is particularly encouraging that the model-indicated predictability translates to high prediction skills, especially for the subtropical LC1 events.

Supplementary Note 4

Here we add comments on the SST forcings related to baroclinic wave activity in the d4PDF simulation. COBE-SST2 used in the d4PDF simulation show some SST differences from HadISST

in the midlatitudes (Supplementary Figure 4a–d). These midlatitude differences sometimes exceed the ensemble spread of the SST forcing that drives the large ensemble simulation. This can contribute to some uncertainties in the long-term SST trend (Supplementary Figure 4c). Nonetheless, the year-to-year variations of the SST are mostly consistent between HadISST and COBE-SST2, so most findings in this section should not be too sensitive to the choice of the SST dataset. We recognize that our model-based estimates of the S/N ratio and the PC metric may have uncertainties, as the observed and simulated year-to-year variations of N-PAC LC2 show somewhat different magnitudes.

The year-to-year variations of LC1 and LC2 events are correlated with SST in various regions (Supplementary Figures 5 and 6). In almost all of the analyzed regions, the linear regression analysis suggests an influence of the El Niño–Southern Oscillation (ENSO). Nonetheless, the ENSO-related variations of baroclinic wave activity show strong seasonality and can even have opposite signs in adjacent months (e.g., N-ATL LC2). The linear regression analysis also highlights other SST forcings, such as patterns that resemble the Atlantic Multidecadal Variability (AMV) and the basin mode of the tropical Indian Ocean. Specifically, the AMV-like patterns are correlated with N-ATL LC1, N-ATL LC2, and AN-SO LC2, while the Indian basin mode is correlated with N-ATL LC1, N-ATL LC2, AN-SO LC2, and AUS-NZ LC2. It is also striking that the Pacific SSTs near 30°N and 30°S are correlated with baroclinic wave activity in many regions, especially during January–March. The above SST forcings sometimes co-exist and might synchronize LC1- or LC2-related weather extremes in different regions.

Interestingly, the previously reported relationships between SST variations and LC1/LC2 events concentrate on the circumstances with high PC values (i.e., S/N ratios) (e.g., Martius et al. 2007; Ryoo et al. 2013; Zhang and Wang 2019). As suggested by Fig. 3, the SST-LC1/LC2

relationships vary across ensemble members and are subject to the influence of unforced variability. For relationships with low S/N ratios, it is likely difficult to identify them using relatively short climate records. This might explain why most of the relationships that involve non-ENSO SST forcings have not been widely recognized or investigated.

Supplementary Note 5

Despite the high predictability and encouraging prediction skill discussed in the main text, the prediction of LC1 and LC2 events by SPEAR still has several notable issues. On the scale of a 5-degree grid, the prediction skill of year-to-year variations is much lower than what the PC analysis suggests (cf. Fig. 4 and Supplementary Figure 10). Besides the limited predictability of SST, we also note that SPEAR has some deficiency in predicting extratropical SSTs near the western boundary currents and in the Southern Hemisphere (Supplementary Figure 7). We speculate that model biases reduce the prediction skill by contributing to subtle spatial shifts of baroclinic wave activity. On the regional scale, the prediction system appears to be overconfident in many cases and/or fails to fully capitalize the model-indicated predictability, as suggested by the differences between the PC values and the prediction skills (Supplementary Figure 11).

A comparison of d4PDF (Supplementary Figures 5 and 6) and SPEAR (Supplementary Figures 13 and 14) suggests general consistency related to the ENSO responses. However, the relationship between the SST and LC1/LC2 events is distorted or absent in SPEAR predictions. For example, SPEAR's N-ATL LC1 events are poorly correlated with TNA and NP40 in July–September; the poor correlations differ from the observation and the SST-forced large ensemble simulation (Fig. 3), likely undermining the prediction skill of N-ATL LC1 events (Fig. 4). This inconsistency persists even after controlling for the differences in the ensemble size and analysis period (not

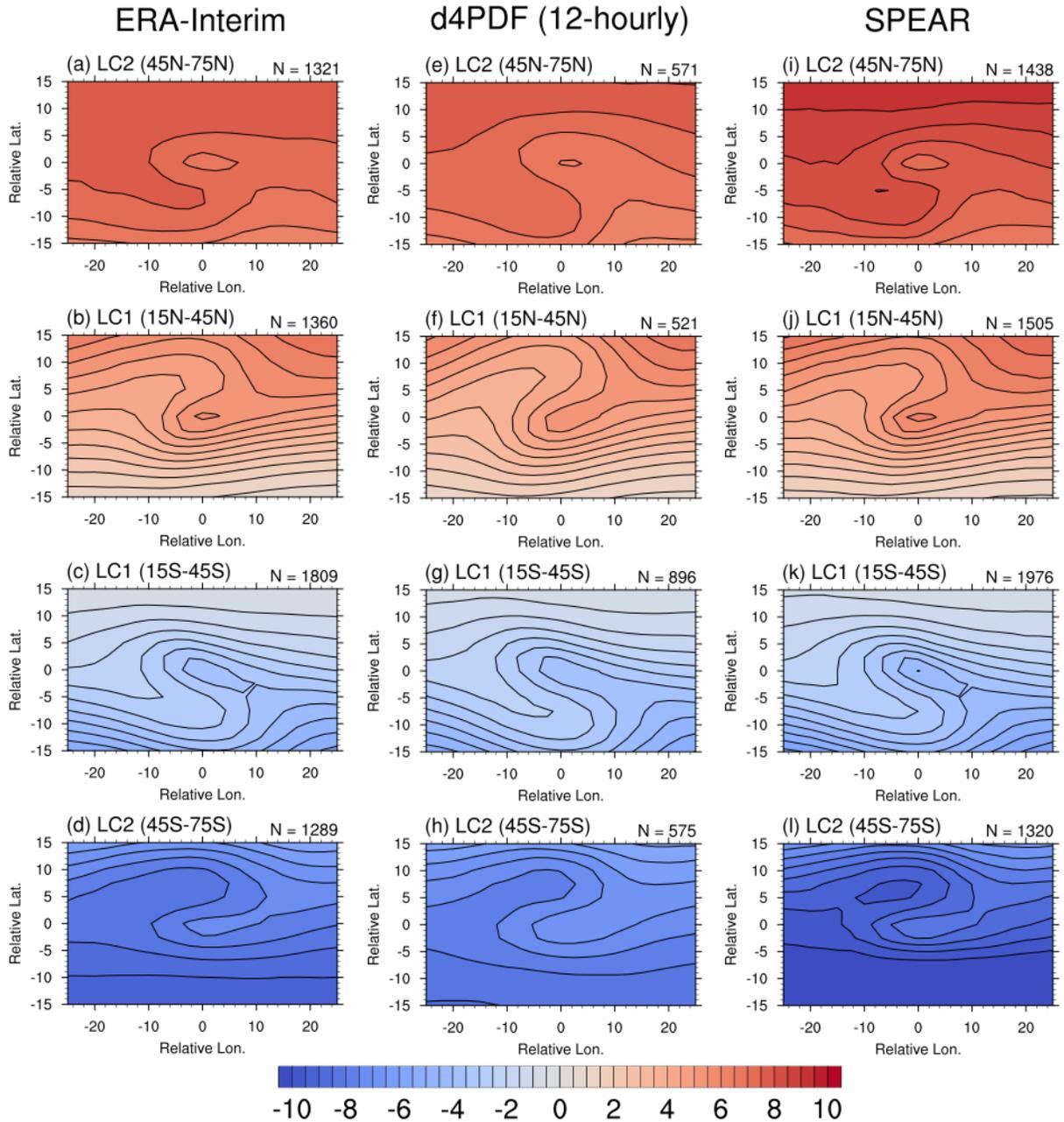
shown). We are exploring whether these issues could be (partially) mitigated by reducing some climatological biases of SPEAR (Delworth et al. 2020), using online correction techniques like flux adjustments. In addition, process-based studies will likely consolidate existing understandings of LC1 and LC2 events and offer valuable insights into how to improve their long-range predictions.

Supplementary Tables

Supplementary Table 1 Reports and studies of the extreme events in Figure 1. The web links of media reports are accessed on July 28, 2020.

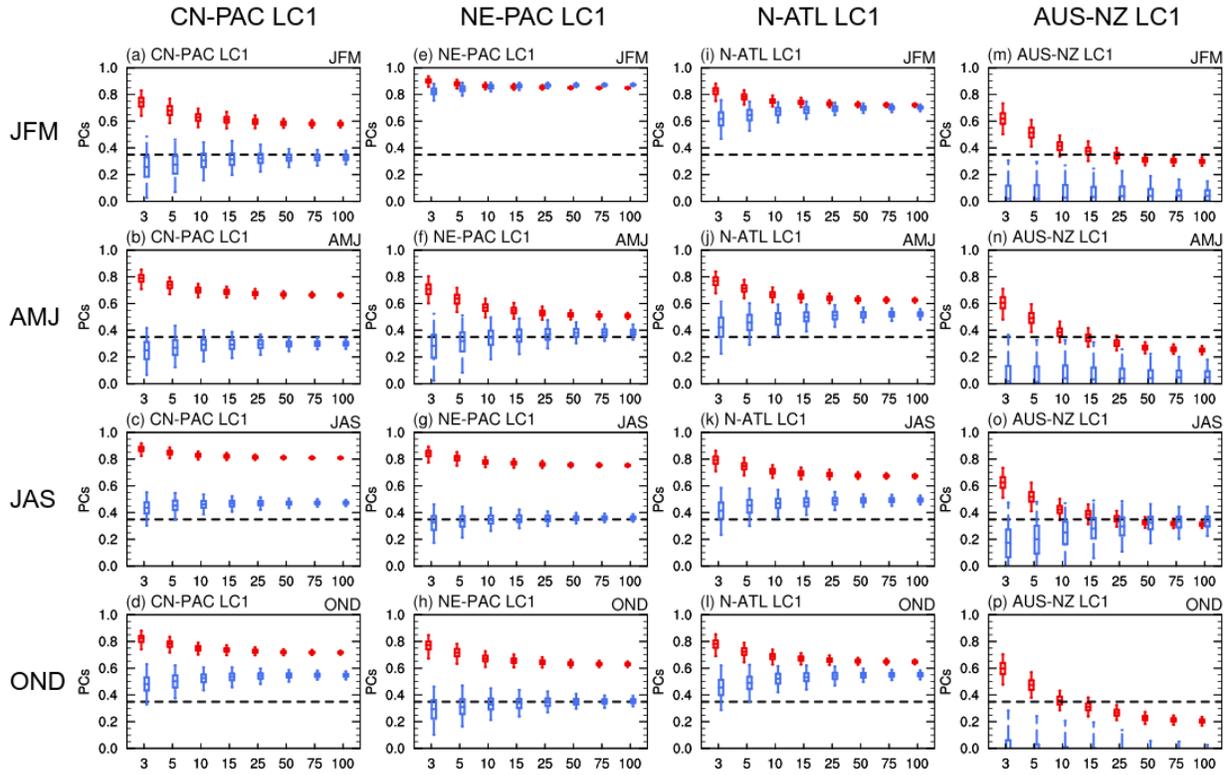
Extreme Event	Media Reports / Academic Studies
Atmospheric River (Fig. 1e–1f)	+ https://www.mercurynews.com/2017/01/09/california-storms-fill-drought-parched-reservoirs/ + Wen, Y, Behrangi, A, Chen, H, Lambriksen, B. How well were the early 2017 California Atmospheric River precipitation events captured by satellite products and ground-based radars? Q. J. R. Meteorol. Soc. 144, 344– 359 (2018).
Hurricane (Fig. 1g–1h)	+ https://phys.org/news/2018-09-hurricane-florence-unusual-dangerous.html + https://weather.com/storms/hurricane/news/2018-09-07-florence-unusual-track-united-states-east-coast
Heat Wave (Fig. 1i–1j)	+ https://www.wunderground.com/cat6/historic-heat-wave-sweeps-asia-middle-east-and-europe + https://www.washingtonpost.com/weather/2019/06/18/recent-scorching-temperatures-kuwait-pakistan-confirmed-third-fourth-hottest-measured-earth/
Heat Wave (Fig. 1k–1l)	+ https://www.theguardian.com/australia-news/2019/jan/24/australia-heatwave-adelaide-breaks-its-all-time-heat-record-hitting-466c + https://www.npr.org/2019/01/25/688755024/australias-heatwave-is-taking-a-toll-on-people-animals-infrastructure-and-land
Polar Extreme (Fig. 1m–1n)	+ https://mashable.com/2018/02/26/arctic-heat-wave-north-pole-february-sea-ice/ + Lee, S. H., Charlton-Perez, A. J., Furtado, J. C., & Woolnough, S. J.. Abrupt stratospheric vortex weakening associated with North Atlantic anticyclonic wave breaking. J. Geophys. Res. Atmos., 124(15), 8563-8575 (2019).
Polar Extreme (Fig. 1o–1p)	+ Wille, J. D., Favier, V., Dufour, A., Gorodetskaya, I. V., Turner, J., Agosta, C., & Codron, F.. West Antarctic surface melt triggered by atmospheric rivers. Nat. Geosci., 12(11), 911-916 (2019). + Hu, X., Sejas, S. A., Cai, M., Li, Z., & Yang, S.. Atmospheric dynamics footprint on the January 2016 ice sheet melting in West Antarctica. Geophys. Res. Lett., 46, 2829– 2835 (2019).

Supplementary Figures



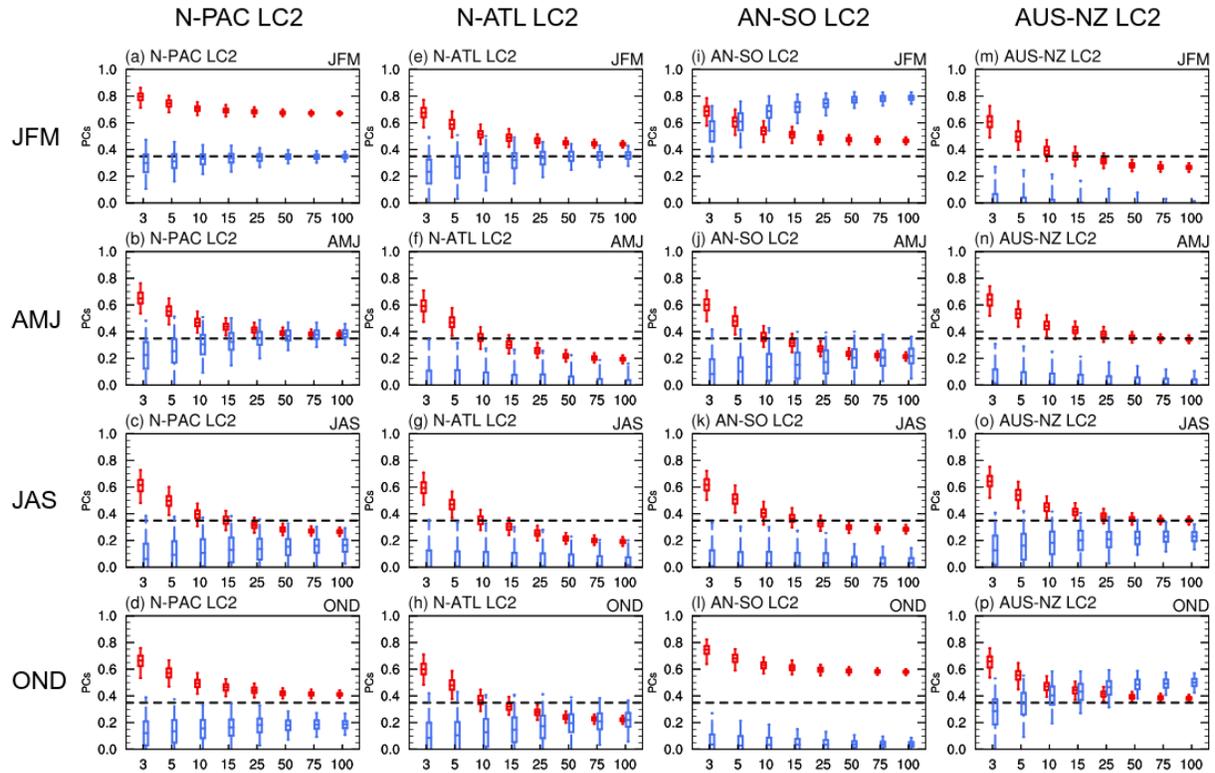
Supplementary Figure 1 Potential vorticity composites of LC1 and LC2 events in relative coordinate systems. Potential vorticity is plotted with color shading and contour lines with the

unit of PVU ($10^{-6} \text{ K kg}^{-1} \text{ m}^2 \text{ s}^{-1}$). (a–d) ERA-Interim reanalysis. (e–h) d4PDF large ensemble. (i–l) SPEAR prediction experiment. From top to bottom, the composites are LC2 events between 45–75°N, LC1 events between 15–45°N, LC1 events between 15–45°S, and LC2 events between 45–75°S. The centers of the relative coordinates correspond to high-PV tongues of LC1 events and low-PV tongues of LC2 events, respectively. The composites show LC1 and LC2 events in January–March 2001, with the number of individual events denoted in the upper right corner of subplots. The event number in the d4PDF column is substantially smaller, partly because we analyzed 12-hourly instead of 6-hourly data. For both d4PDF and SPEAR, only one ensemble member was analyzed here.



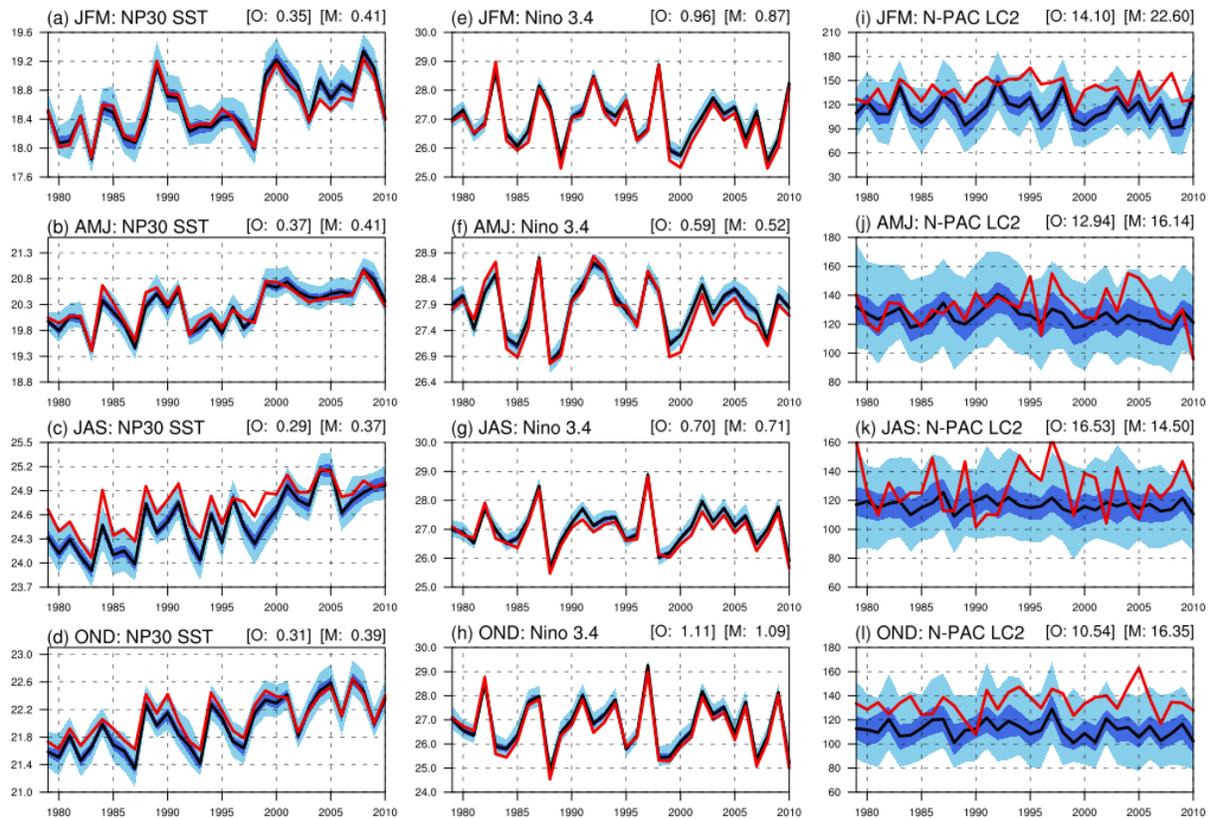
Supplementary Figure 2 Predictable components (PCs) of LC1 events in d4PDF simulation.

The PCs are estimated for (a–d) CN-PAC LC1 (10–30°N, 130°E–170°W), (e–h) NE-PAC LC1 (10–45°N, 100–160°W), (j–l) N-ATL LC1 (10–30°N, 45–85°W), and (m–p) AUS-NZ LC1 (30–55°S, 120–180°E). From top to bottom, the subplots show January–March, April–June, July–September, and October–December of (1979–2010). Horizontal axes show the ensemble size. The boxplots indicate 2.5th, 25th, 50th, 75th, and 97.5th percentiles of metrics, determined from the bootstrapping technique. Red shows the estimated model PCs, and blue shows the estimated simulation skills (or the PC in observations). The black dashed line indicates the 5% significance level of the correlation coefficient.

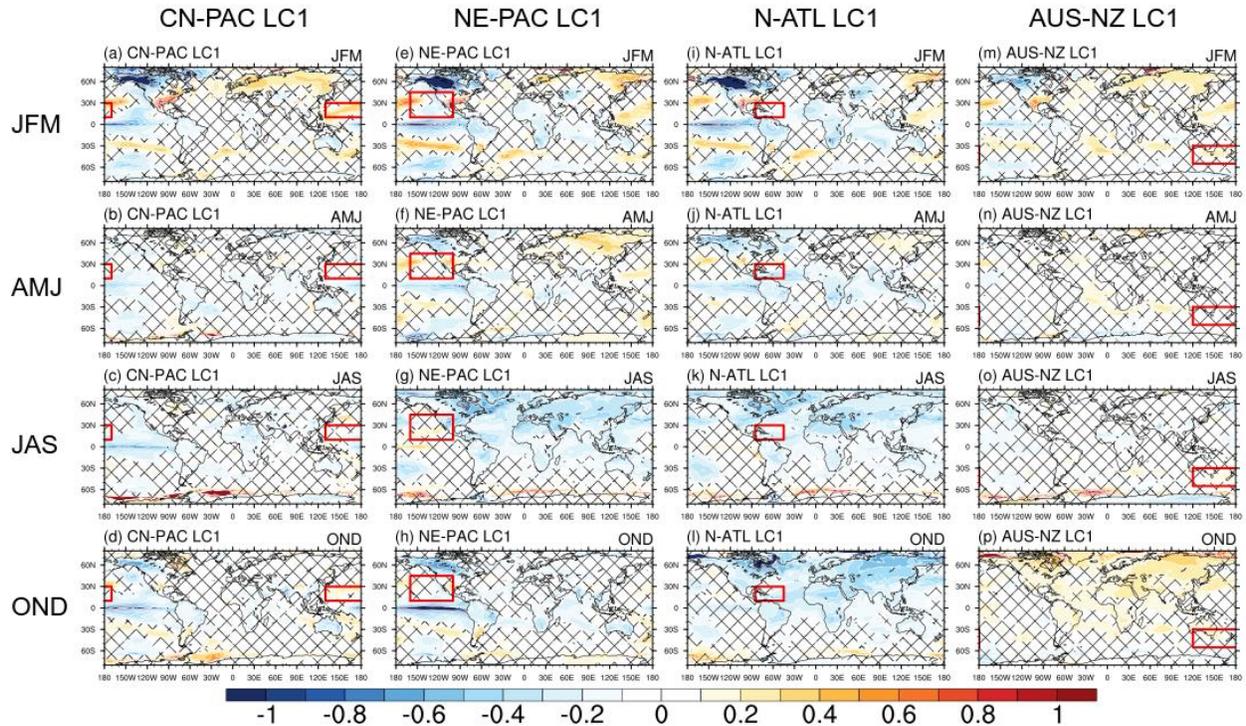


Supplementary Figure 3 Predictable components (PCs) of LC2 events in d4PDF simulation.

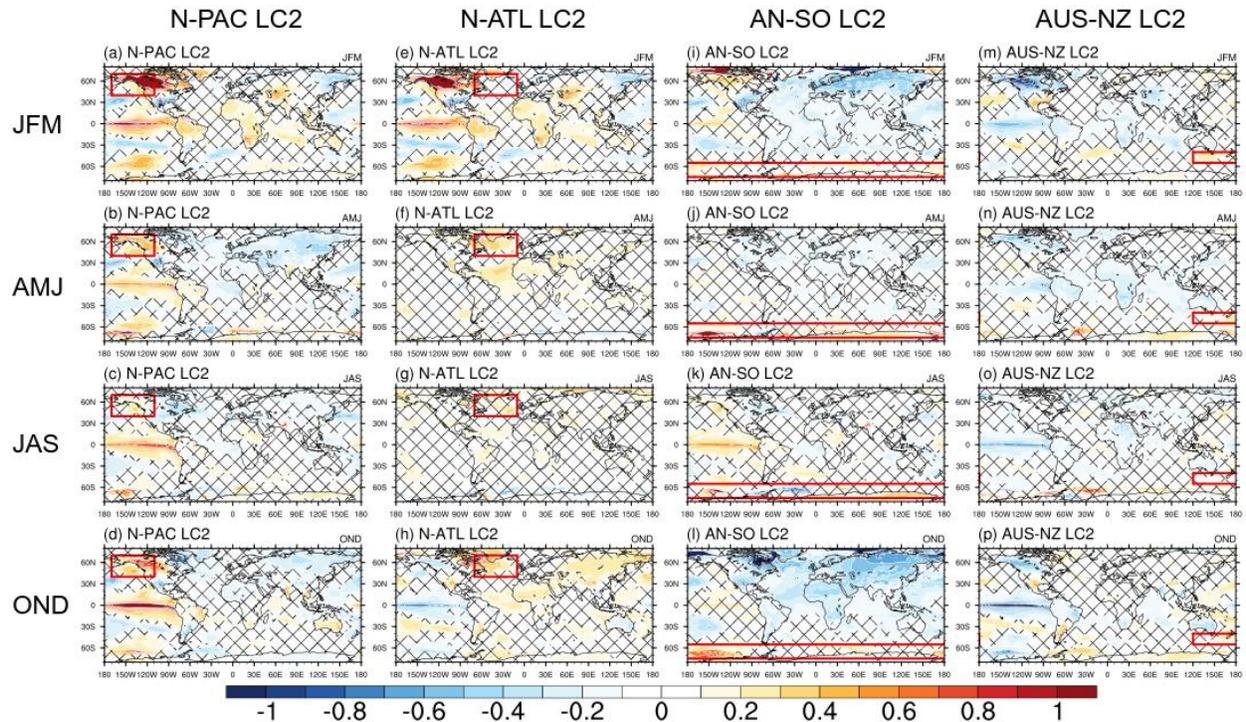
The PCs are estimated for (a–d) N-PAC LC2 (40–70°N, 110–170°W), (e–h) N-ATL LC2 (40–70°N, 10–70°W), (j–l) AN-SO LC2 (55–75°S, circum-global), and (m–p) AUS-NZ LC2 (40–55°S, 120–180°E). From top to bottom, the subplots show January–March, April–June, July–September, and October–December of 1979–2010. Horizontal axes show the ensemble size. The boxplots indicate 2.5th, 25th, 50th, 75th, and 97.5th percentiles of metrics, determined from the bootstrapping technique. Red shows the estimated model PCs, and blue shows the estimated simulation skills (or the PC in observations). The black dashed line indicates the 5% significance level of the correlation coefficient.



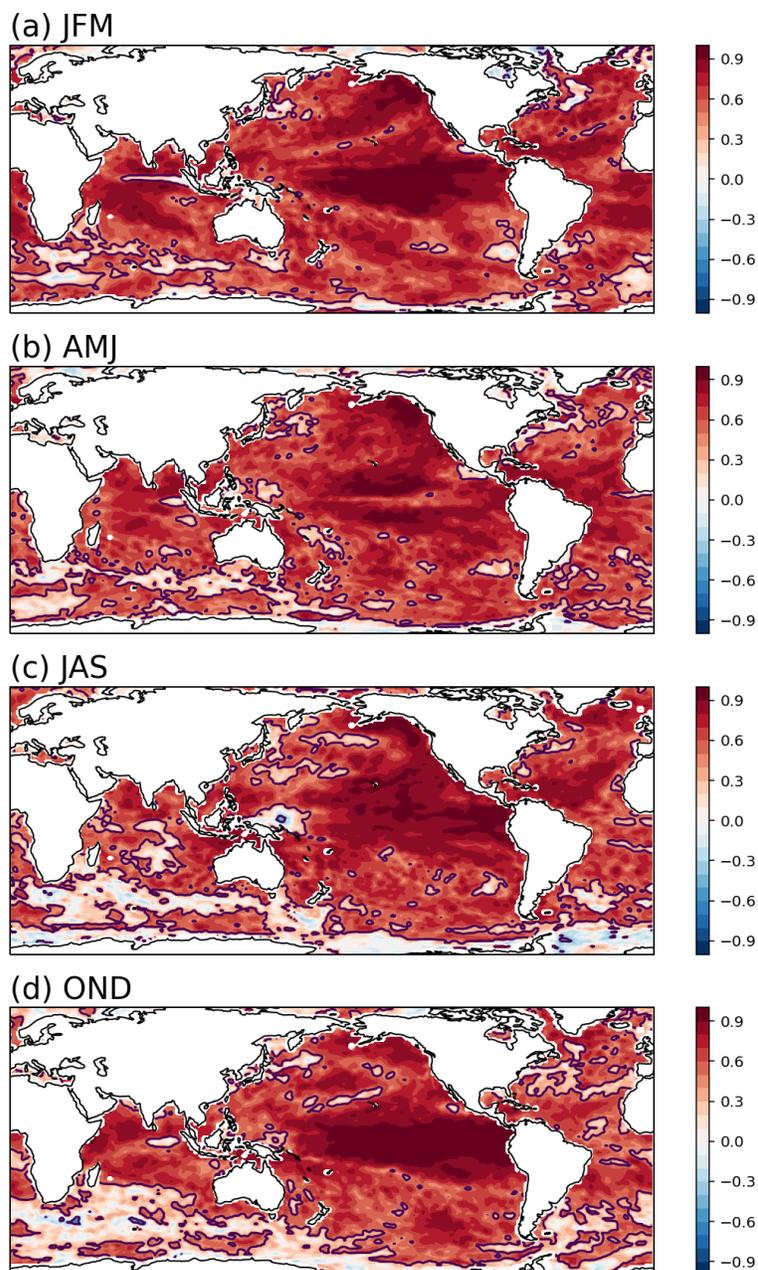
Supplementary Figure 4 Time series of regional SST and N-PAC LC2 events. (a–d) SST in the midlatitude North Pacific (NP30; 20°–40°N, 170°E–140°W). (e–h) SST in Niño 3.4 region (120–170°W, 5°S–5°N). (i–l) Sums of N-PAC LC2 events. From top to bottom, the subplots show January–March, April–June, July–September, and October–December. The SST time series are seasonal means (unit: °C), while the LC2 series are sums over three-month periods. Light blue shading shows the 95% confidence interval, and dark blue shading shows the 50% confidence interval. Black lines show the d4PDF ensemble median, and red lines show the observation. The standard deviations of the observation and ensemble members are denoted with “M” and “O” in the upper right corner of subplots, respectively.



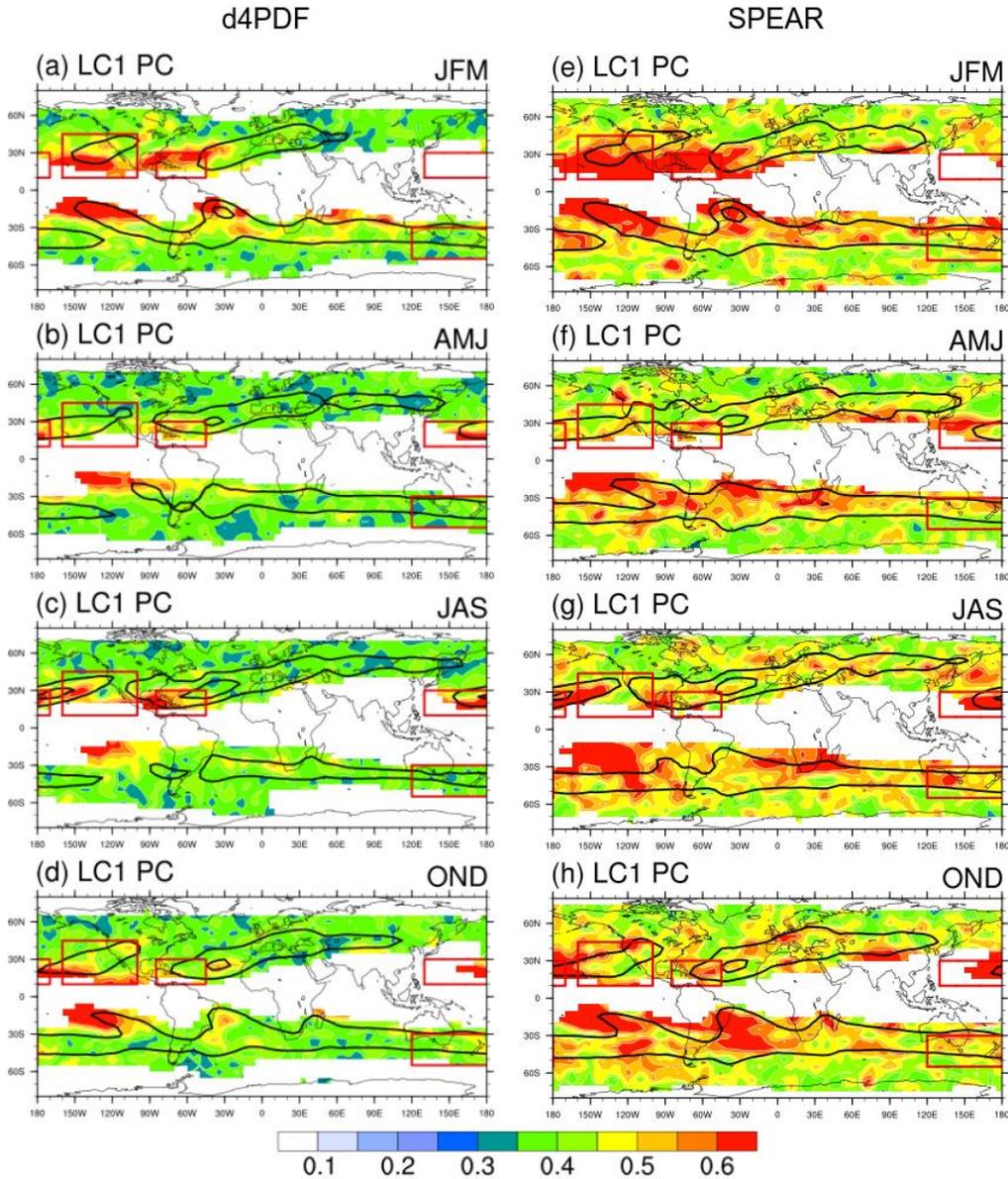
Supplementary Figure 5 Relationships between year-to-year variations of LC1 events and surface temperature. Color shading shows the linear regression of ensemble-mean surface temperature (unit: K) onto ensemble-mean LC1 activity during 1979-2010. The baroclinic wave activity is represented with the event counts of (a–d) CN-PAC LC1, (e–h) NE-PAC LC1, (j–l) N-ATL LC1, and (m–p) AUS-NZ LC1. These ensemble means are standardized based on regional year-to-year variations, and the analyzed regions are denoted with red boxes in the subplots. Hatching suggests the linear relationship is below 95% confidence. From top to bottom, the subplots show January–March, April–June, July–September, and October–December.



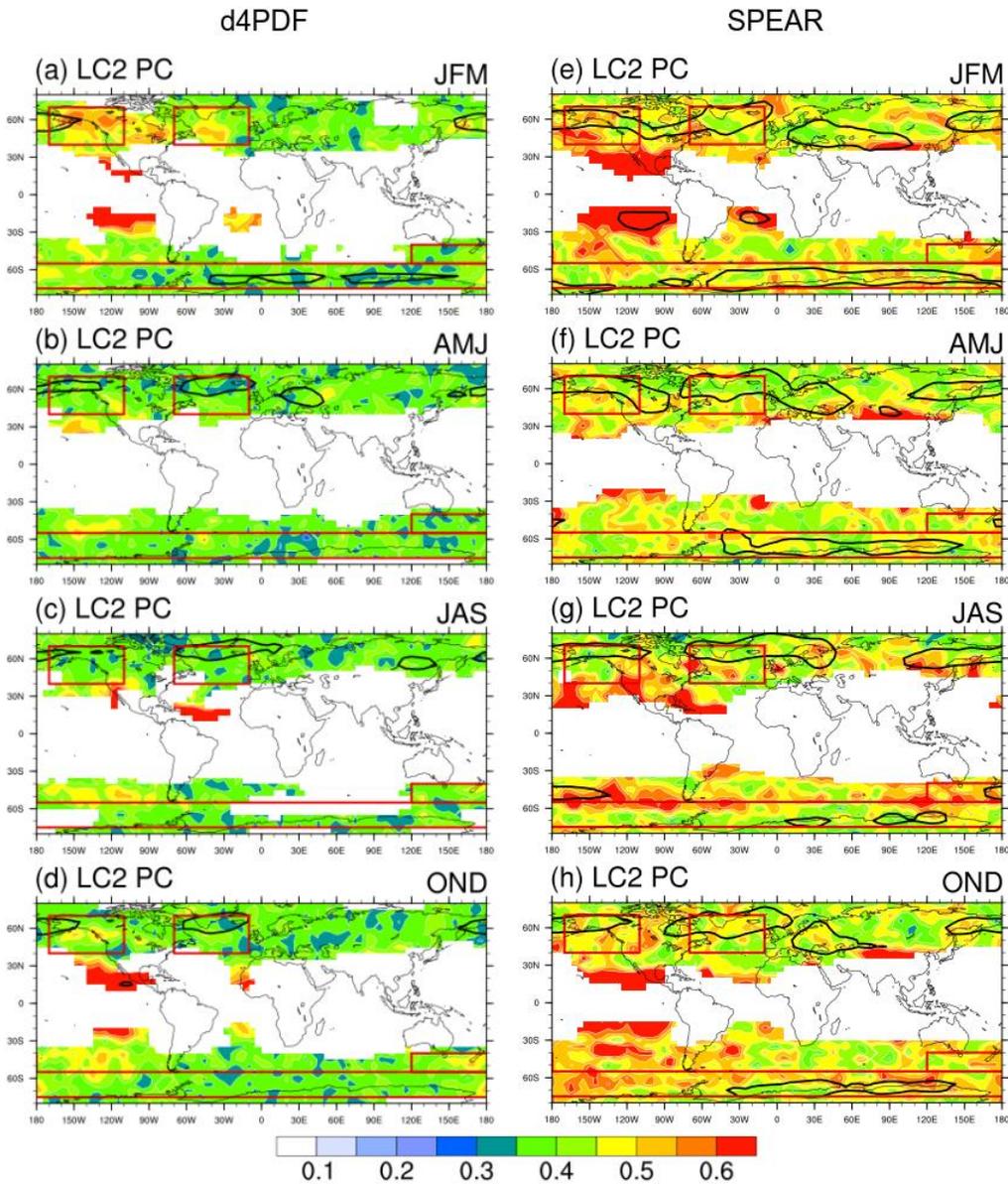
Supplementary Figure 6 Relationships between year-to-year variations of LC2 events and surface temperature. Color shading shows the linear regression of ensemble-mean surface temperature (unit: K) onto ensemble-mean LC2 activity during 1979-2010. The baroclinic wave activity is represented with the event counts of (a–d) N-PAC LC2, (e–h) N-ATL LC2, (j–l) AN-SO LC2, and (m–p) AUS-NZ LC2. These ensemble means are standardized based on regional year-to-year variations, and the analyzed regions are denoted with red boxes in the subplots. Hatching suggests the linear relationship is below 95% confidence. From top to bottom, the subplots show January–March, April–June, July–September, and October–December.



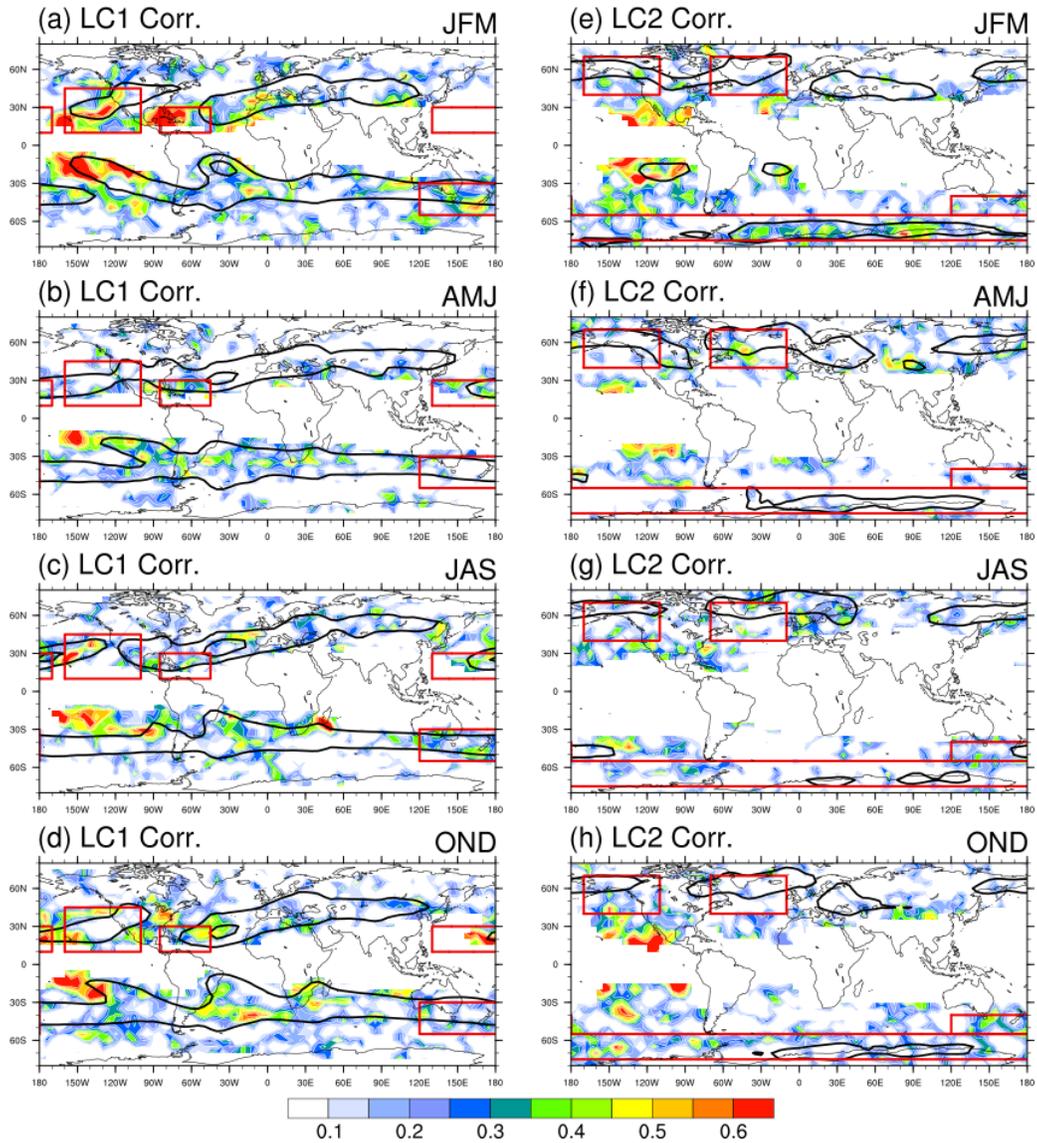
Supplementary Figure 7 Anomaly correlation coefficients between the SST observations and SPEAR predictions. The examined SPEAR predictions are Month 1–3 predictions initialized in (a) January, (b) April, (c) July, and (d) October during 1995–2018. The data have been detrended before the correlation coefficients are calculated. Black solid lines over the ocean indicate the 95% confidence level.



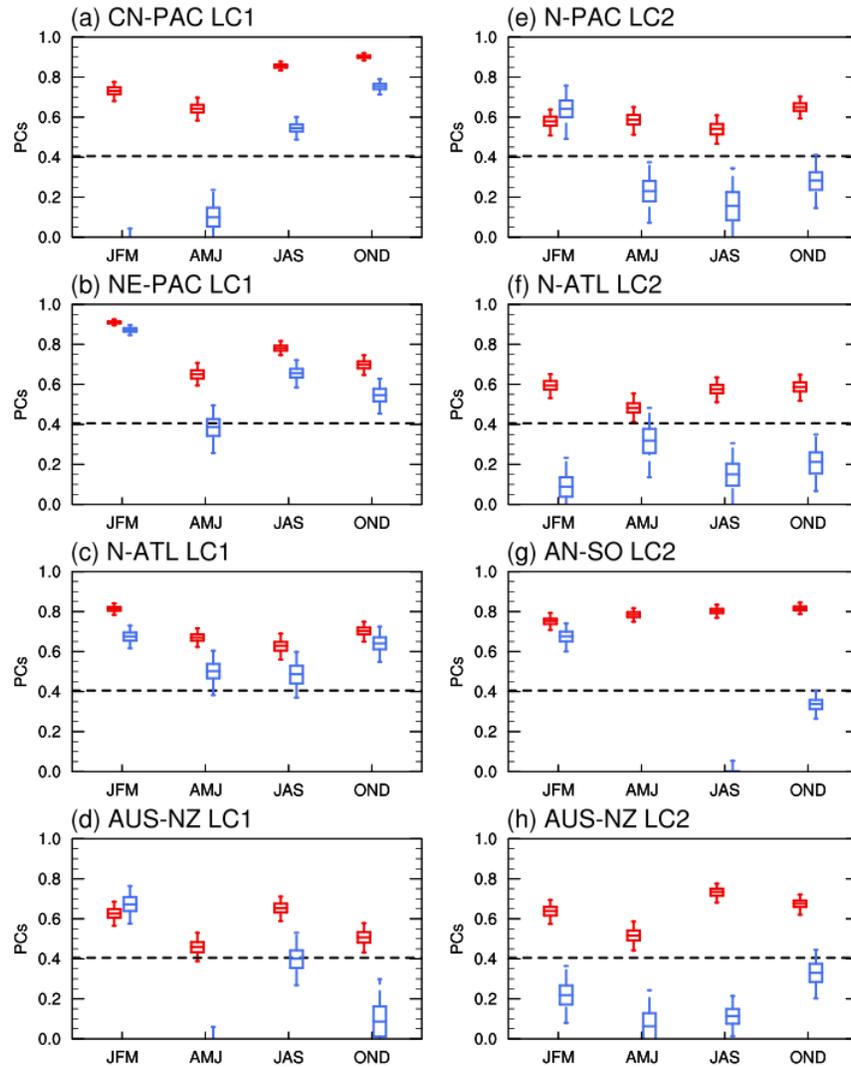
Supplementary Figure 8 Predictable components (PCs) of LC1 events (1995–2010) in the d4PDF and SPEAR. (a–d) PCs in the d4PDF large ensemble that are estimated when the ensemble size is set at fifteen. (e–h) PCs in SPEAR’s three-month predictions that are initialized in January, April, July, and October. The black contours show the climatology of the simulations, with the same contour settings as Figure 1m–1p. From top to bottom, the subplots show January–March, April–June, July–September, and October–December.



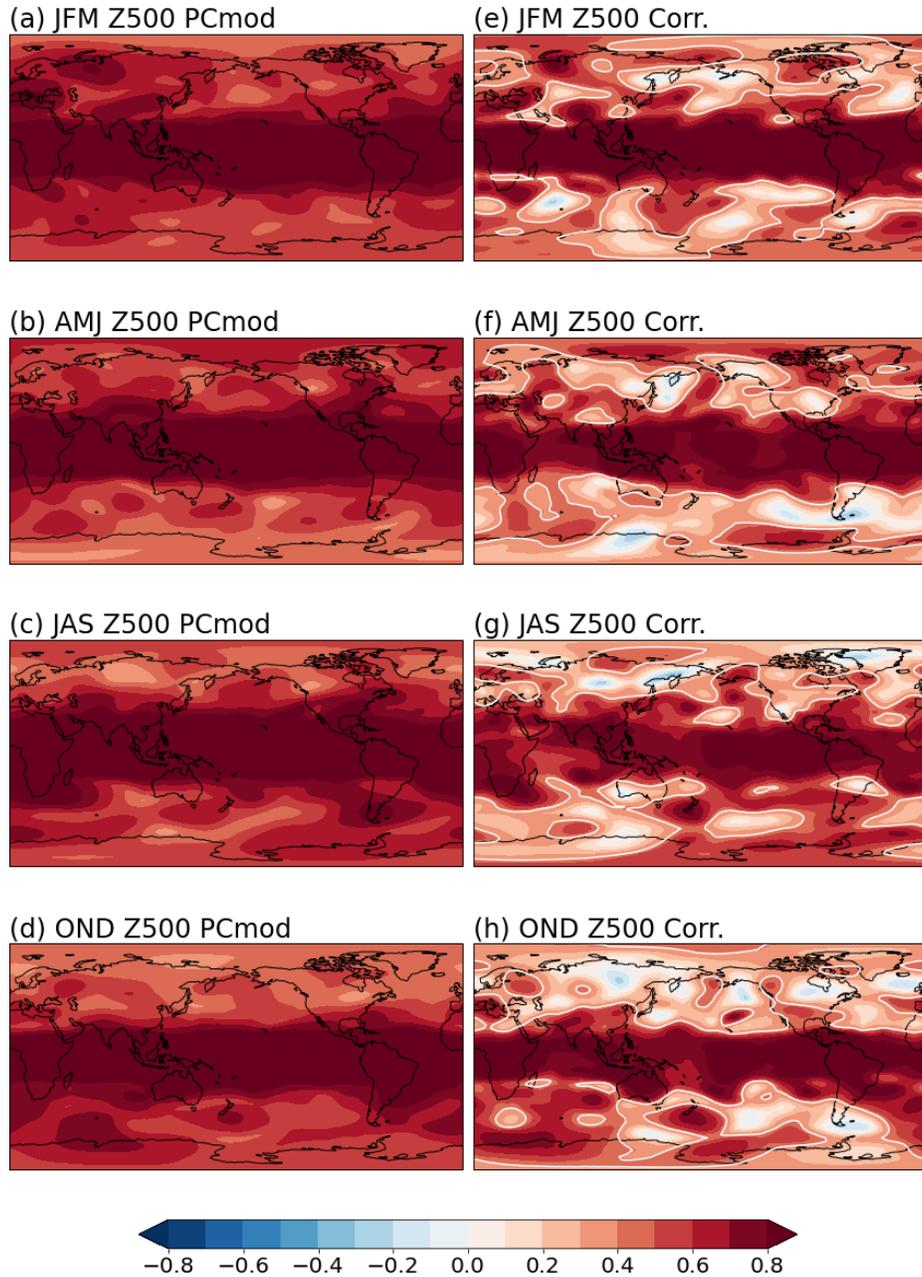
Supplementary Figure 9 Predictable components (PCs) of LC2 events (1995–2010) in the d4PDF and SPEAR. (a–d) PCs in the d4PDF large ensemble, estimated with the ensemble size is set at fifteen. (e–h) PCs in SPEAR’s three-month predictions that are initialized in January, April, July, and October. The black contours show the climatology of the simulations, with the same contour settings as Figure 1m–1p. From top to bottom, the subplots show January–March, April–June, July–September, and October–December.



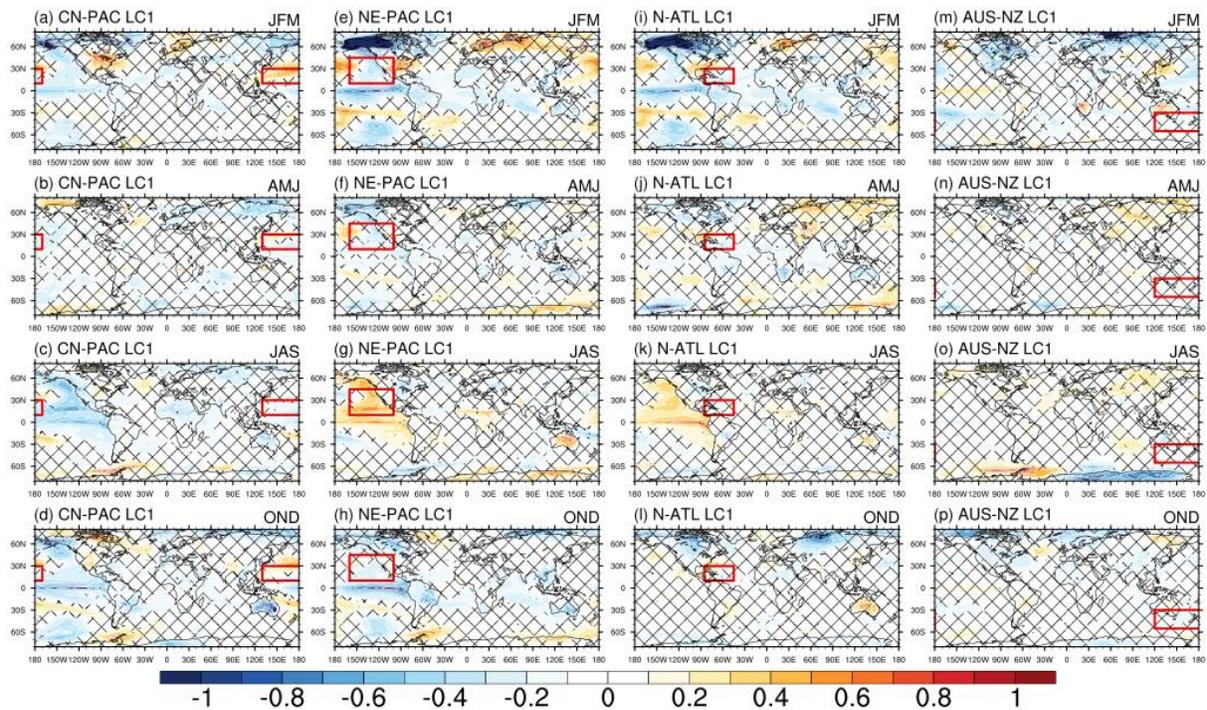
Supplementary Figure 10 Prediction skill of LC1 & LC2 events (1995–2018) of SPEAR. (a–d) Correlation skill of LC1 events in three-month predictions that are initialized in January, April, July, and October. (d–g) Same as (a–d), but for LC2 events. The black contours show the climatology of the SPEAR retrospective prediction, with the same contour settings as Figure 1m–1p. The red boxes indicate the individual regions examined in the main text and Supplementary Materials.



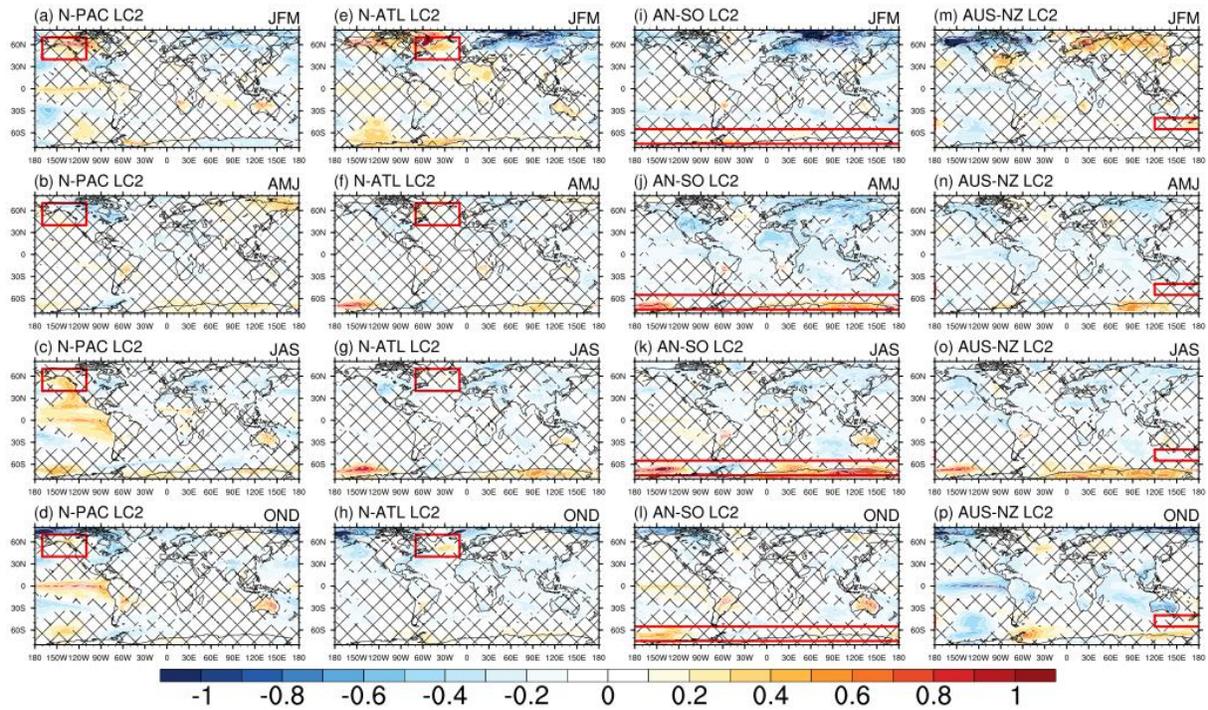
Supplementary Figure 11 Predictable components and prediction skills (1995–2018) of BWA in SPEAR. The predictions are three-month ones that are initialized in January, April, July, and October. The left column shows the results of (a) CN-PAC LC1, (b) NE-PAC LC1, (c) N-ATL LC1, and (d) AUS-NZ LC1. The right column shows (e) N-PAC LC2, (f) N-ATL LC2, (g) AN-SO LC2, and (h) AUS-NZ LC2. Red shows the estimated model PCs, and blue shows the estimated simulation skills (or the PC in observations). The boxplots indicate 2.5th, 25th, 50th, 75th, and 97.5th percentiles of metrics, determined from the bootstrapping technique. The black dashed line indicates the 5% significance level of the correlation coefficient.



Supplementary Figure 12 Predictable components and prediction skills of 500-hPa geopotential height in SPEAR. (a–d) Model predictable components (color shading) of 3-month predictions initialized on the firsts of January, April, June, and October of 1995–2018. (e–h) Same as (a–d), but for the correlation between SPEAR predictions and the ERA-Interim reanalysis. White solid lines in (e–h) indicate correlation coefficients at the 95% confidence level.



Supplementary Figure 13 Relationships between year-to-year variations of LC1 events and surface air temperature in SPEAR predictions (1995–2010). The analysis is similar to Supplementary Figure 5 except that SPEAR has shorter records, a smaller ensemble size, and a different model temperature variable.



Supplementary Figure 14 Relationships between year-to-year variations of LC2 events and surface air temperature in SPEAR predictions (1995–2010). The analysis is similar to Supplementary Figure 6 except that SPEAR has shorter records, a smaller ensemble size, and a different model temperature variable.