LIST OF FIGURES

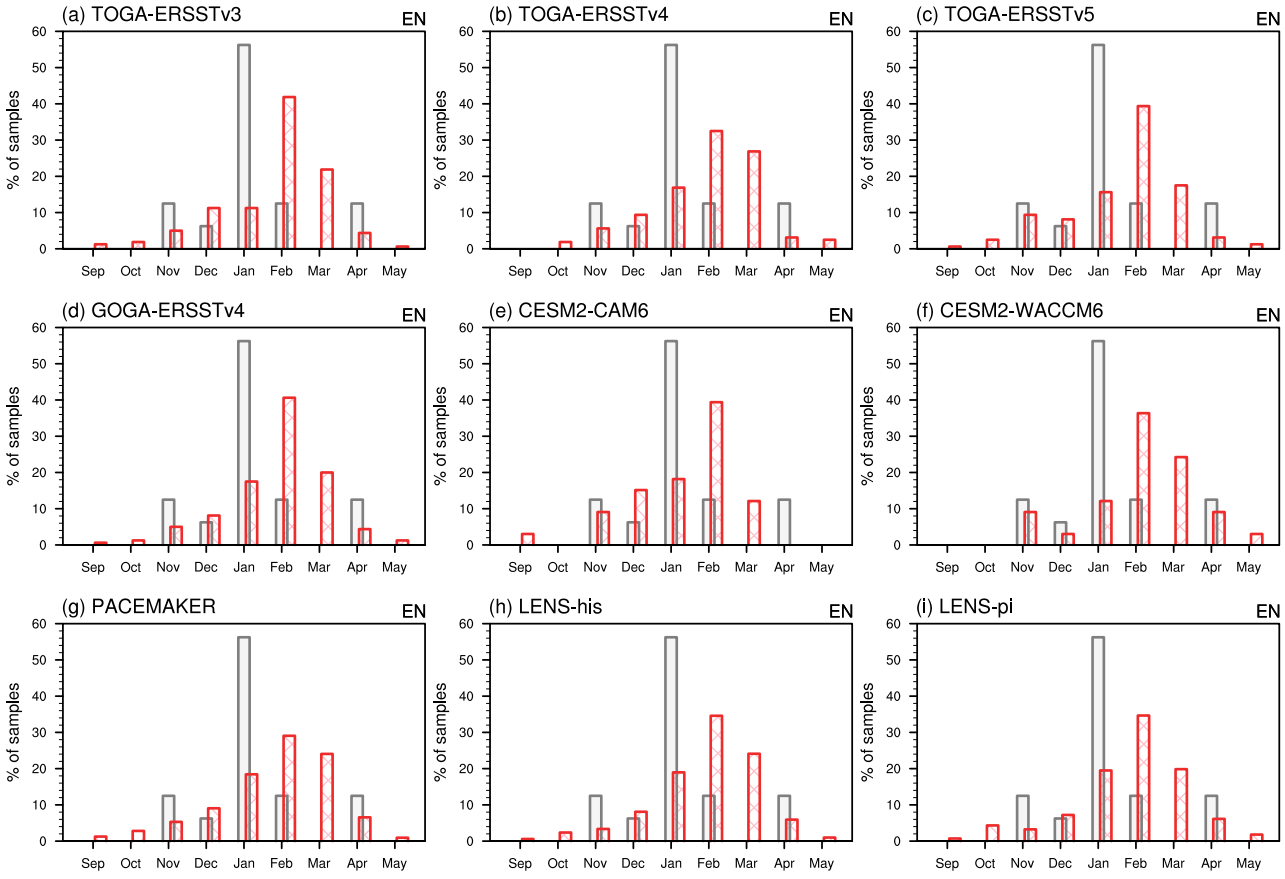

Fig. S1. Histograms of the month of the peak NPI during El Niño events in observations based on ERA20C (grey) and (a) TOGA-ERSSTv3b, (b)TOGA-ERSSTv4, (c) TOGA-ERSSTv5, (d) GOGA-ERSSTv4, (e) CESM2-CAM6, (f) CESM2-WACCM6, (g) PACEMAKER, (h) LENS-his, (i) LENS-pi simulations from CESM (red).

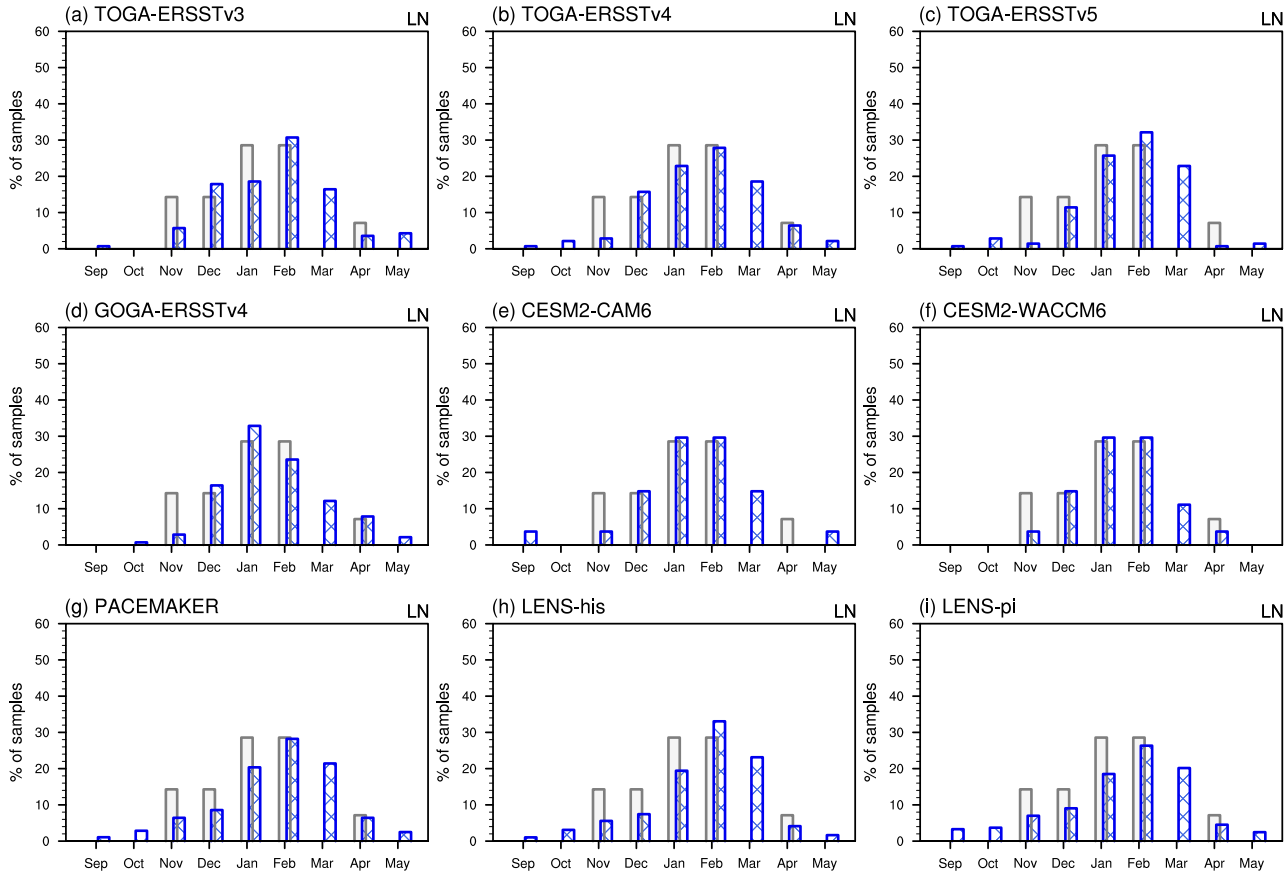

Fig. S2. Histograms of the month of the peak NPI during La Niña events in observations based on ERA20C (grey) and (a) TOGA-ERSSTv3b, (b) TOGA-ERSSTv4, (c) TOGA-ERSSTv5, (d) GOGA-ERSSTv4, (e) CESM2-CAM6, (f) CESM2-WACCM6, (g) PACEMAKER, (h) LENS-his, (i) LENS-pi simulations from CESM (blue).

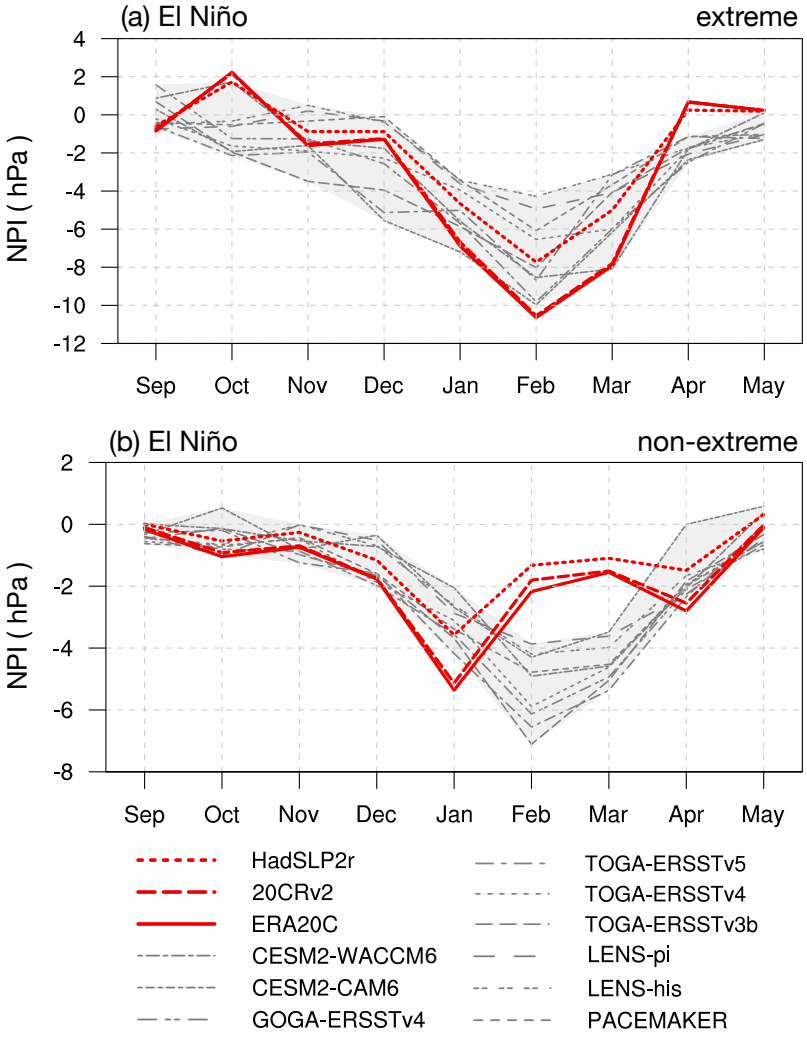

Fig. S3. Seasonal evolution of the NPI anomalies (hPa) for (a) extreme El Niño events, (b) non-extreme El Niño events. For observations and models based on observed SSTs, 1982/1983 and 1997/1998 are the only two extreme events defined during 1920–2010 with the DJF Niño3.4 exceeding two standard deviations. For the coupled runs, the extremes are defined by the same two standard deviations criteria but based on their own SST anomalies for calculating the Niño3.4 indices.Red shows the observation-based datasets and grey shows the 9 CESM simulations with the gray shading demarcating their range.

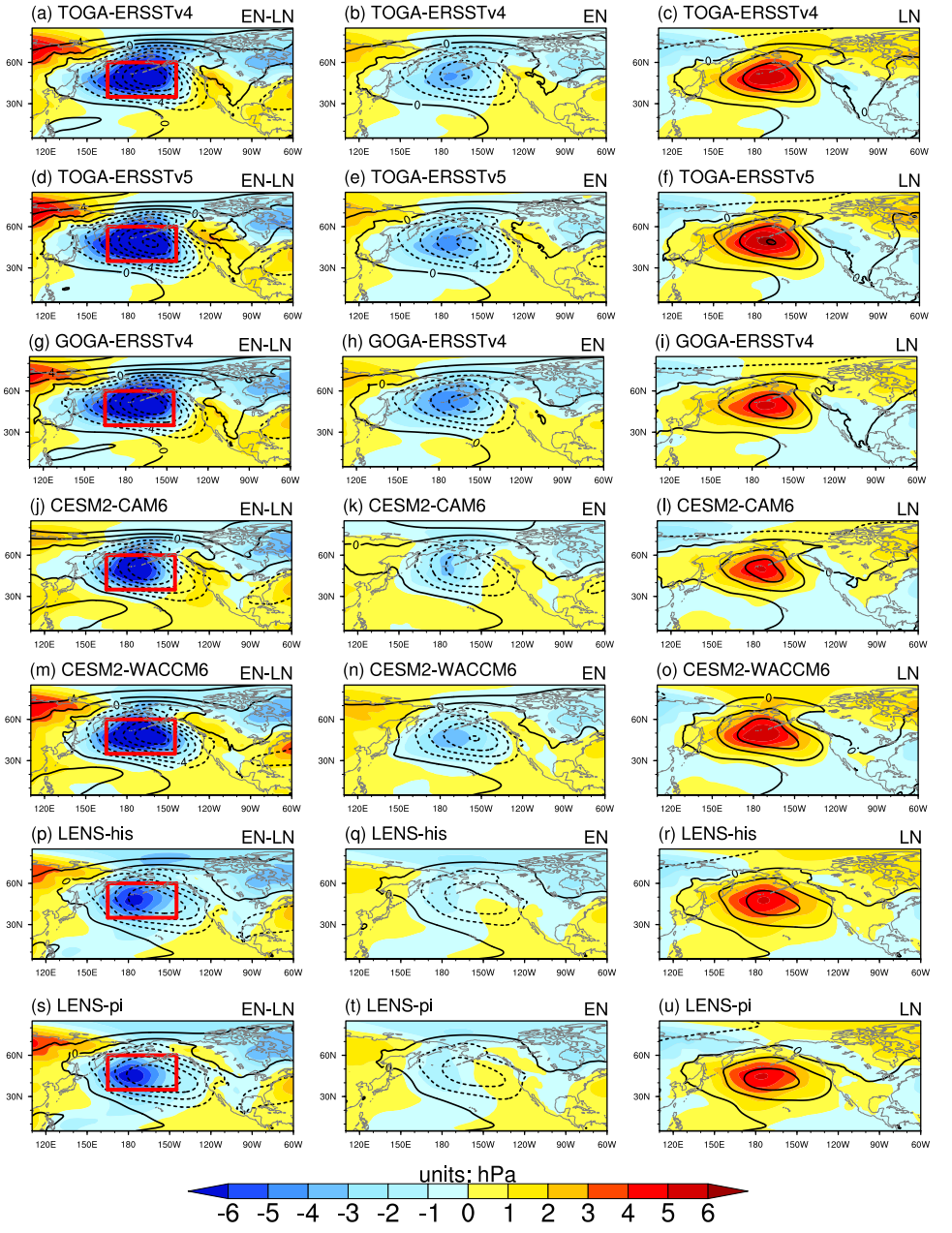

Fig. S4. Similar to Fig. 3, but for (a)-(c) TOGA-ERSSTv4, (d)-(f) TOGA-ERSSTv5, (g)-(i) GOGA-ERSSTv4, (j)-(l) CESM2-CAM6, (m)-(o) CESM2-WACCM6, (p)-(r) LENS-his, (s)-(u) LENS-pi simulations from CESM. The contour interval is 2 hPa with negative ones dashed. Shadings are the difference fields compared to ERA20C.

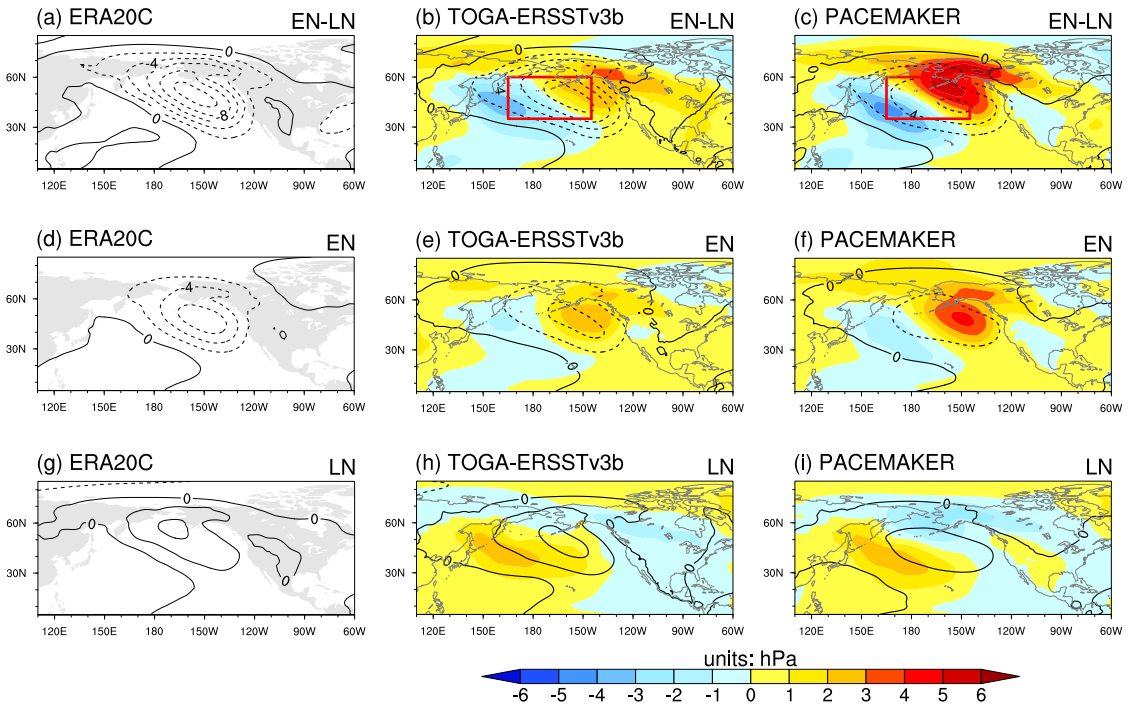

Fig. S5. Similar to Fig. 3, but over DJ.

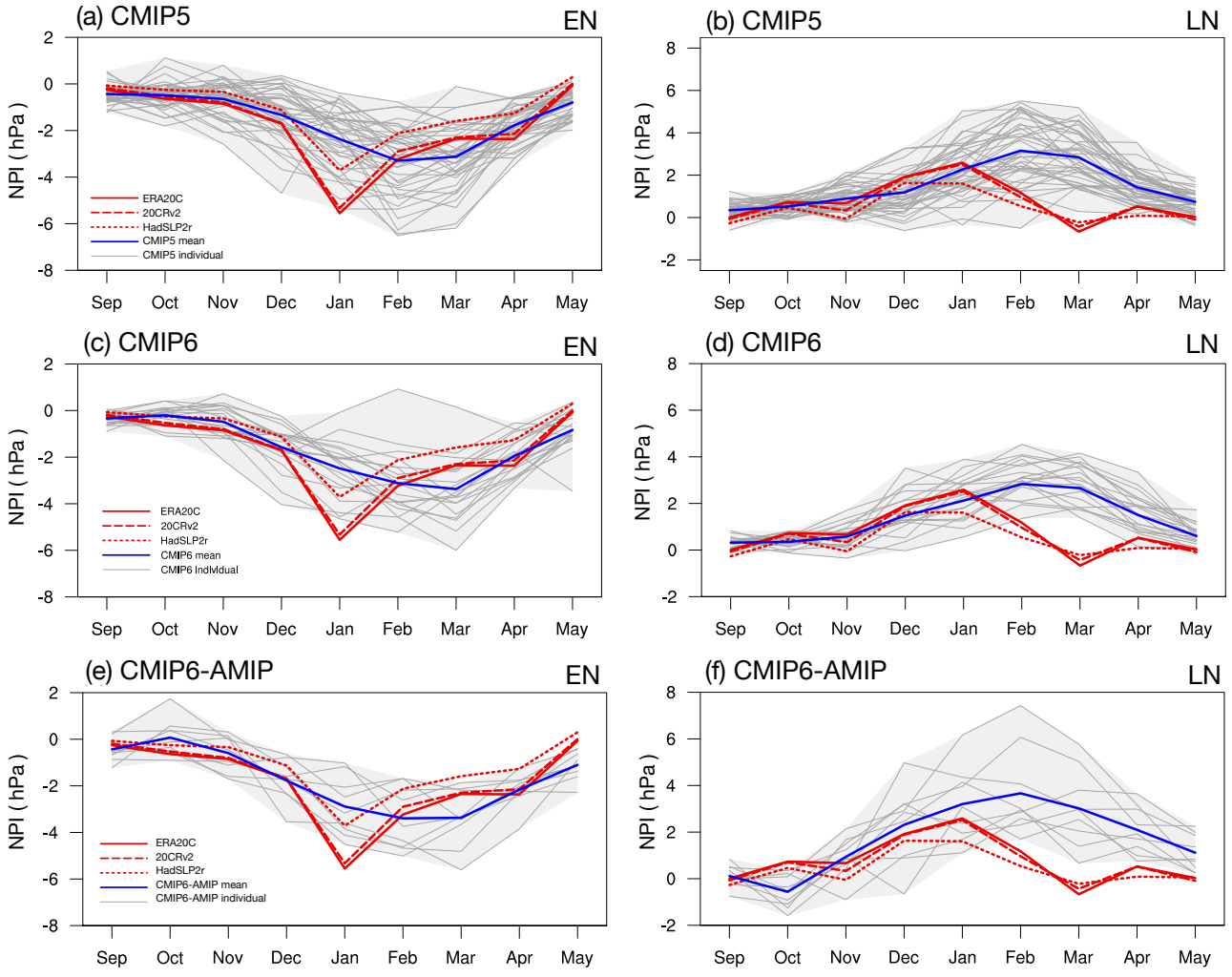

Fig. S6. Evolution of the NPI (hPa) in observations (red) and (a)-(b) each of the 43 CMIP5, (c)-(d) 20 CMIP6 models' piControl simulations, and (e)-(f) 9 CMIP6 models' AMIP simulations (grey) during El Niño events (left) and La Niña events (right). Blue is for the multi-model mean and grey shading depicts the model range.

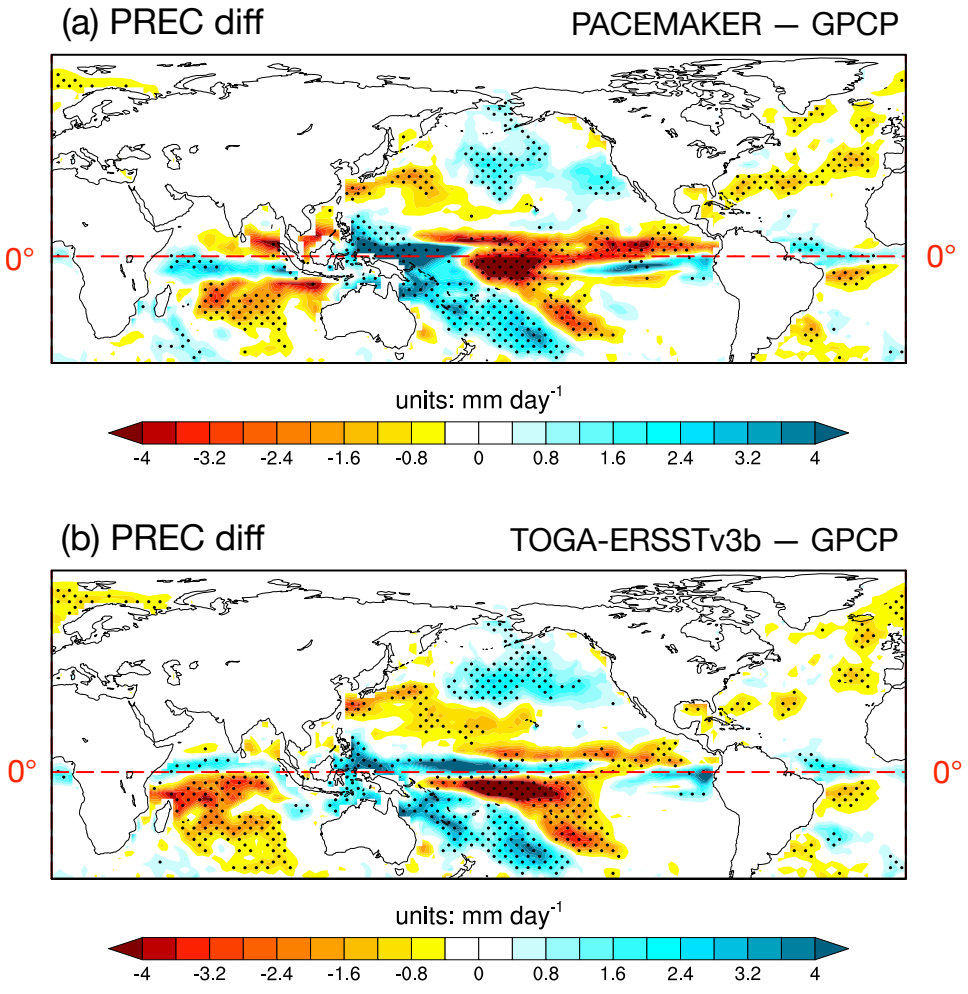

Fig. S7. Similar to Fig. 7(a) and 7(c), but based on the GPCP observational product during the period of 1979–2010. Stippling indicates that the composite difference exceeds 95% confidence level.