Supplementary Information

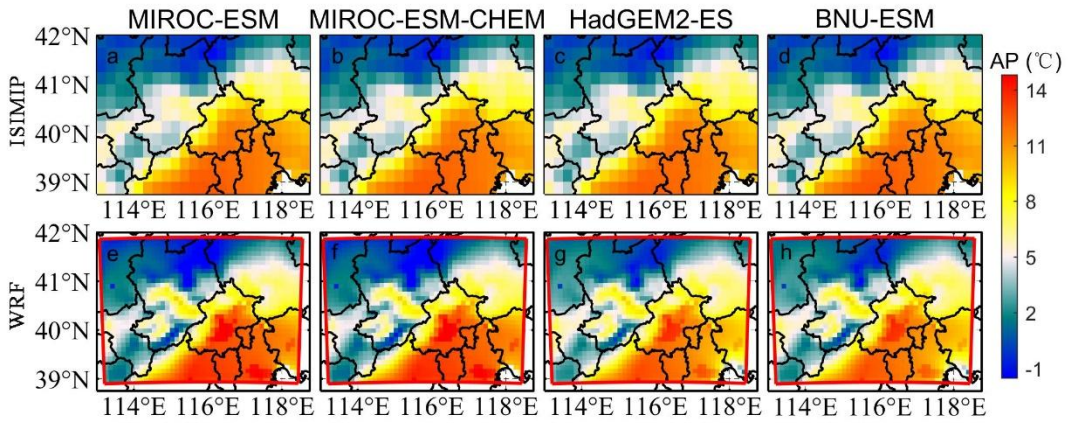

Figure S1. The spatial distribution of mean apparent temperature of MIROC-ESM **(a, e),** MIROC-ESM-CHEM **(b, f),** HadGEM2-ES **(c, g)** and BNU-ESM **(d, h)** during 2008-2017. The first row are results from ISIMIP method, and the second row are results from WRF.

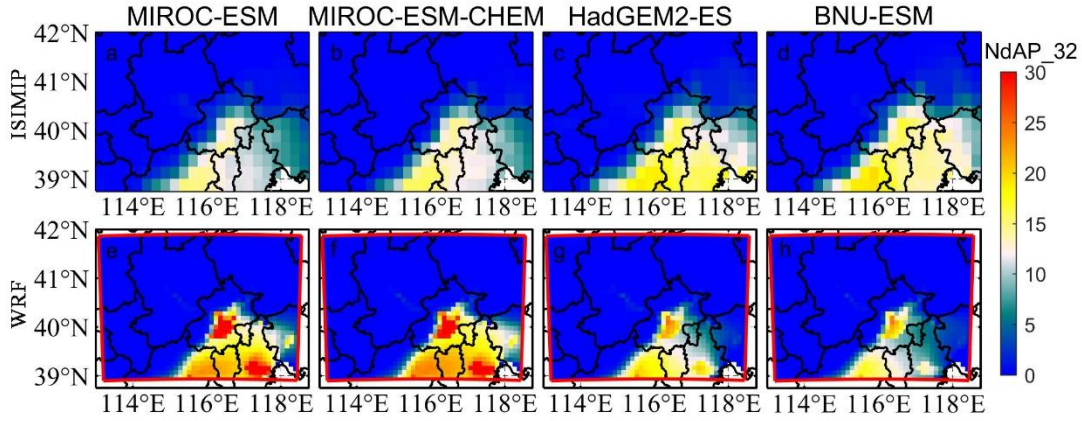

Figure S2. Same as Fig. S1 but for annual mean NdAP_32.

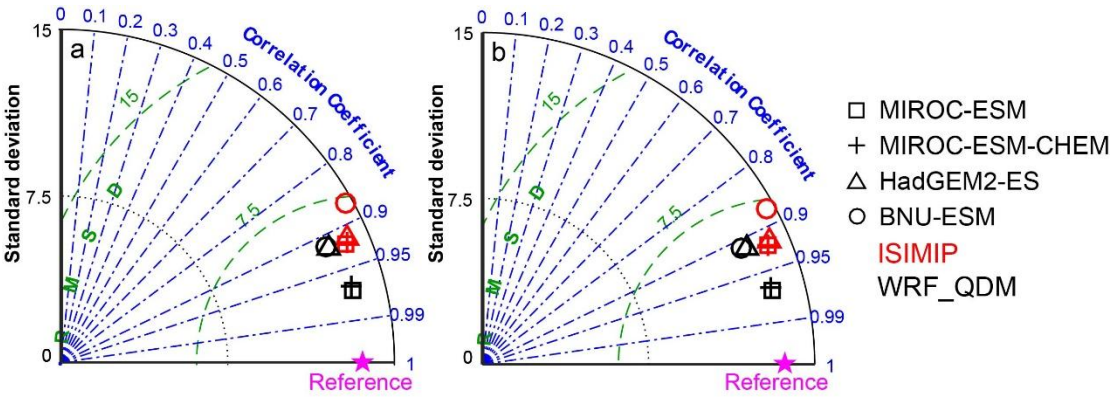

Figure S3. Taylor diagram for daily apparent temperature of four ESMs using two downscaling methods, i.e., ISIMIP (red) and WRF (black) compared to observed data in Beijing-Tianjin provinces **(a)** and Beijing-Tianjin urban areas **(b)**. The skill of downscaling methods is reflected by the distance from each symbol to the point labelled "Reference", the ERA5 reanalysis data. The blue lines are correlation coefficient which represents the similarity between each downscaling data and reanalysis data. The green contours are root mean standard deviation (RMSD), and black contours are standard deviation.

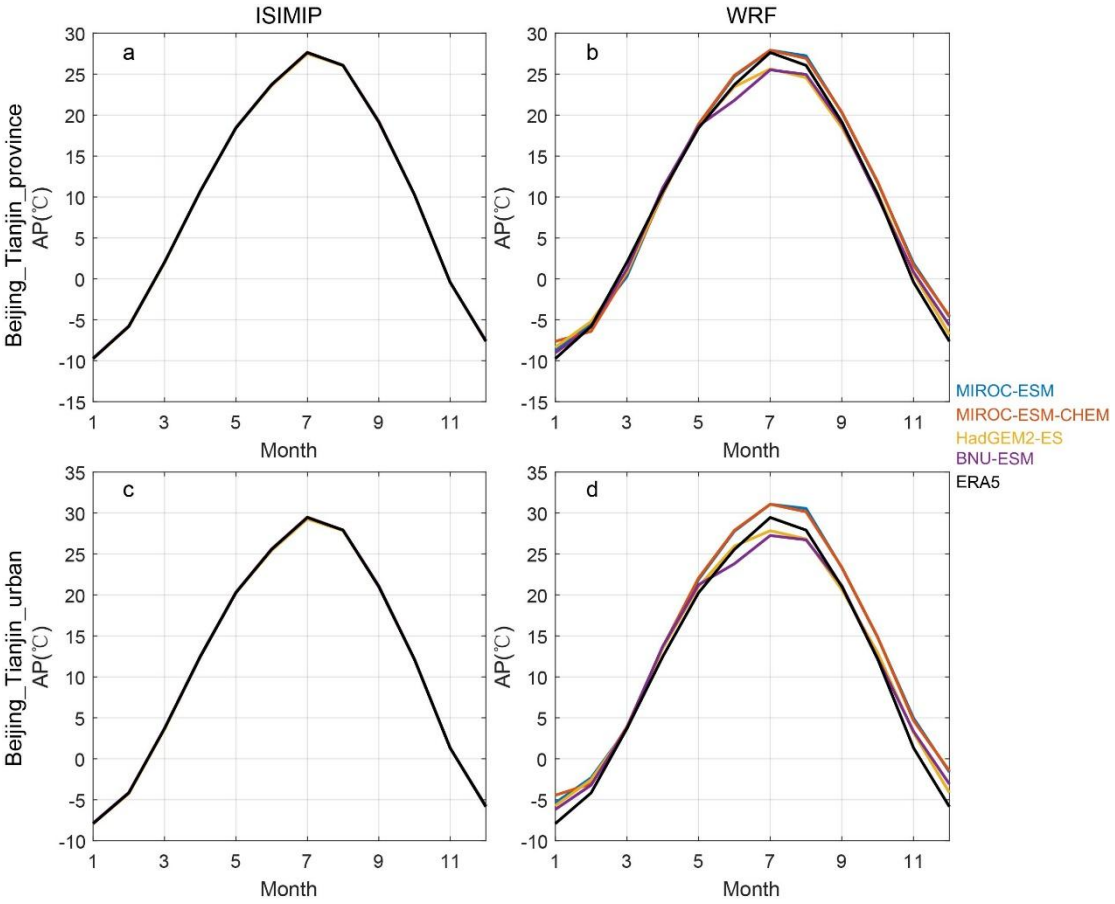

Figure S4. Seasonal cycle of average AP of 4 ESMs under ISIMIP **(a, c)** and WRF **(b, d)** in Beijing-Tianjin province **(a, b)** and Beijing-Tianjin urban areas **(c, d)** during 2008-2017.

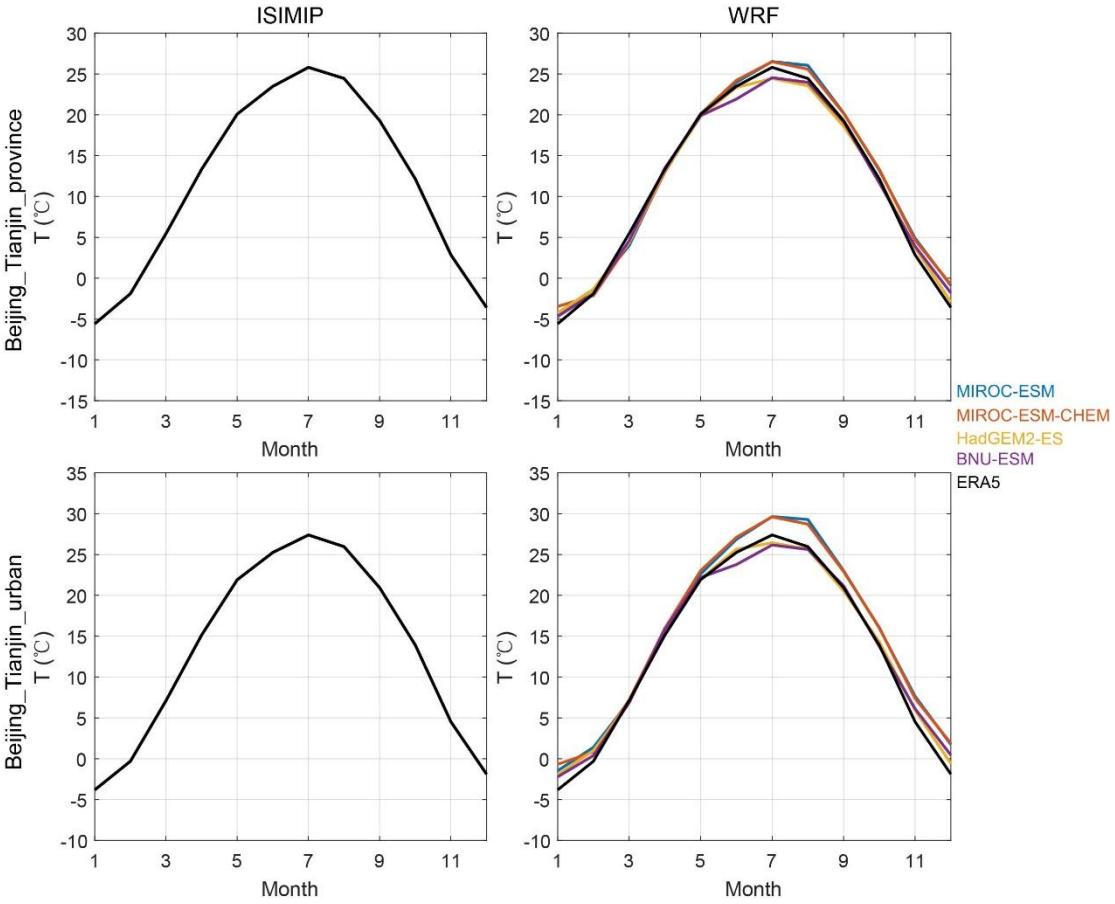

Figure S5. Seasonal cycle of average 2m temperature of 4 ESMs under ISIMIP **(a, c)** and WRF **(b, d)** in Beijing-Tianjin province **(a, b)** and Beijing-Tianjin urban areas **(c, d)** during 2008-2017.

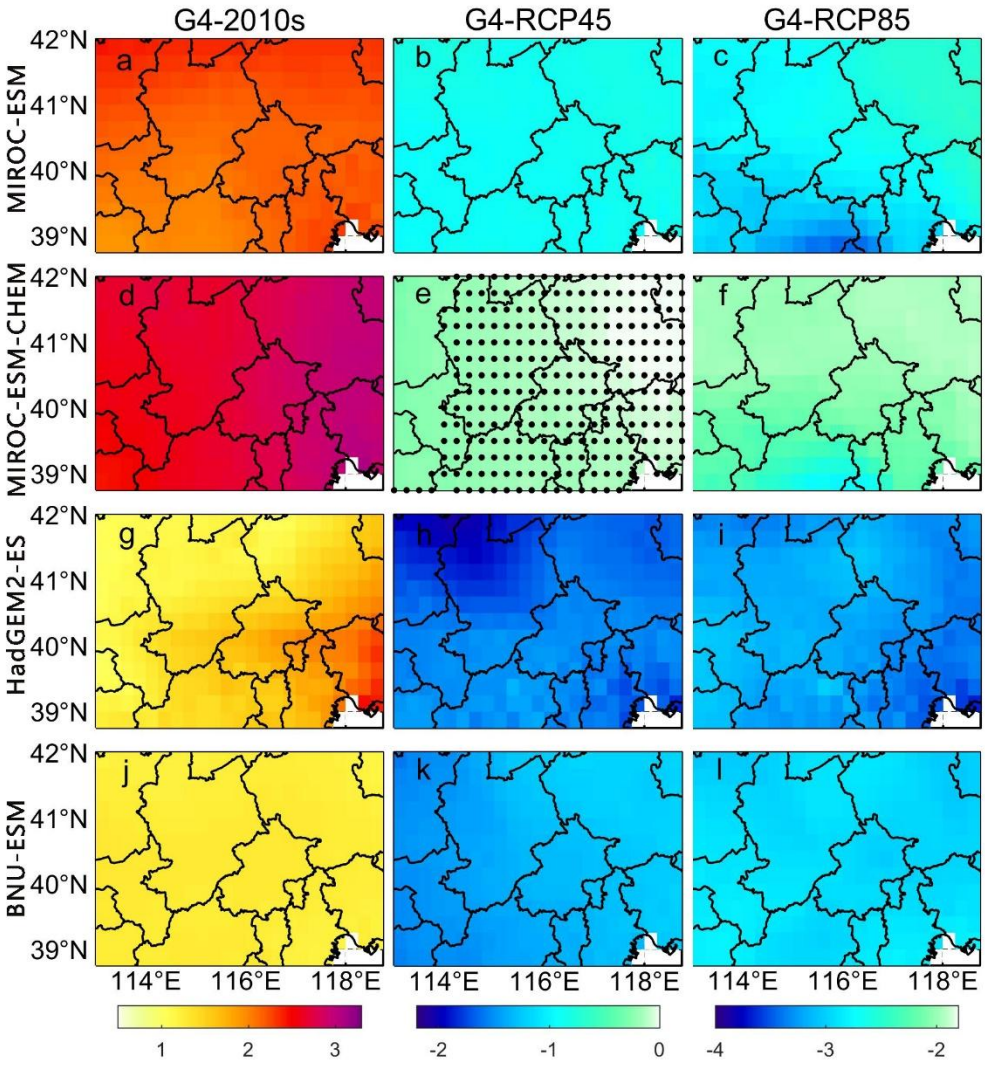

Figure S6. Spatial pattern of apparent temperature difference (℃) under different scenarios over 2060-2069: G4-2010s (left column), G4-rcp4.5 (second column) and G4-rcp8.5 (right column) based on ISIMIP method. 2010s means the results simulated during 2008-2017. From top to bottom are MIROC-ESM **(a-c)**, MIROC-ESM-CHEM **(d-f)**, HadGEM2-ES **(g-i)** and BNU-ESM **(j-l)**, respectively. Stippling indicates grid points where differences or changes are not significant at the 5% level according to the Wilcoxon signed rank test.

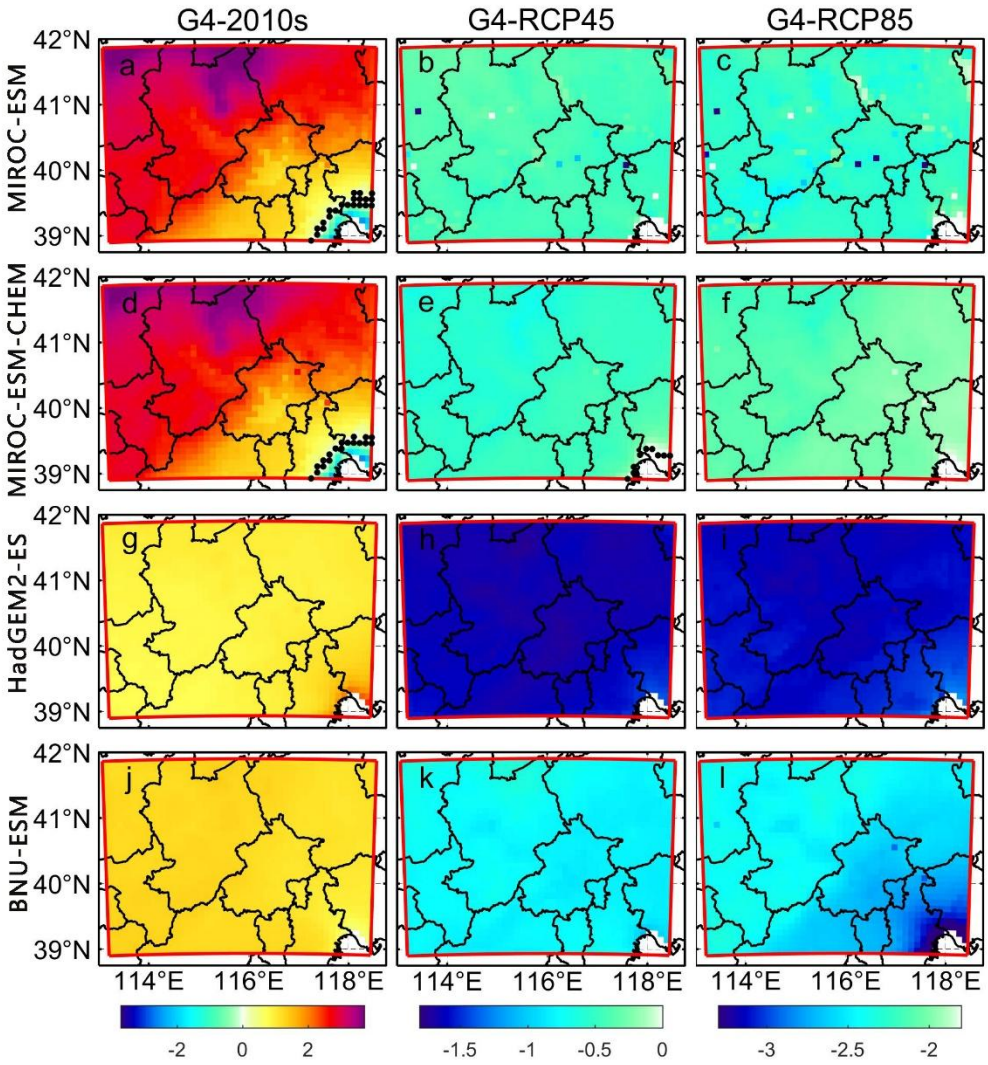

Figure S7. Spatial pattern for apparent temperature difference (℃) under different scenarios over 2060-2069: G4-2010s (left column), G4-rcp4.5 (second column) and G4-rcp8.5 (right column) based on WRF_QDM results. 2010s means the results simulated during 2008-2017. From top to bottom are MIROC-ESM (a-c), MIROC-ESM-CHEM (d-f), HadGEM2-ES (g-i) and BNU-ESM (j-l), respectively. Stippling indicates grid points where differences or changes are not significant at the 5% level according to the Wilcoxon signed rank test.

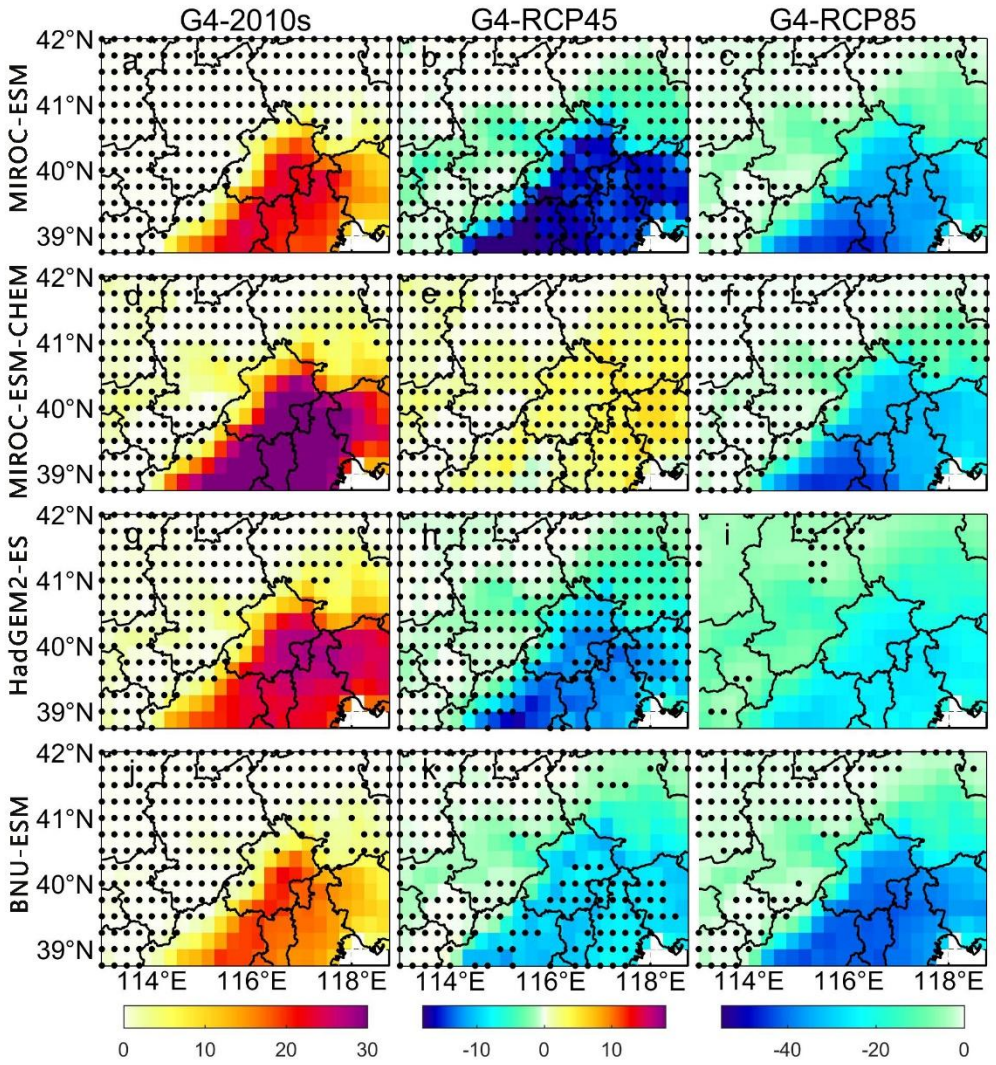

Figure S8. Number of days with AP > 32℃ differences under different scenarios over 2060-2069: G4-2010s (left column), G4-rcp4.5 (second column) and G4-rcp8.5 (right column) based on ISIMIP method. 2010s means the results simulated during 2008-2017. From top to bottom are MIROC-ESM (a-c), MIROC-ESM-CHEM (d-f), HadGEM2-ES (g-i) and BNU-ESM (j-l), respectively. Stippling indicates grid points where differences or changes are not significant at the 5% level according to the Wilcoxon signed rank test.

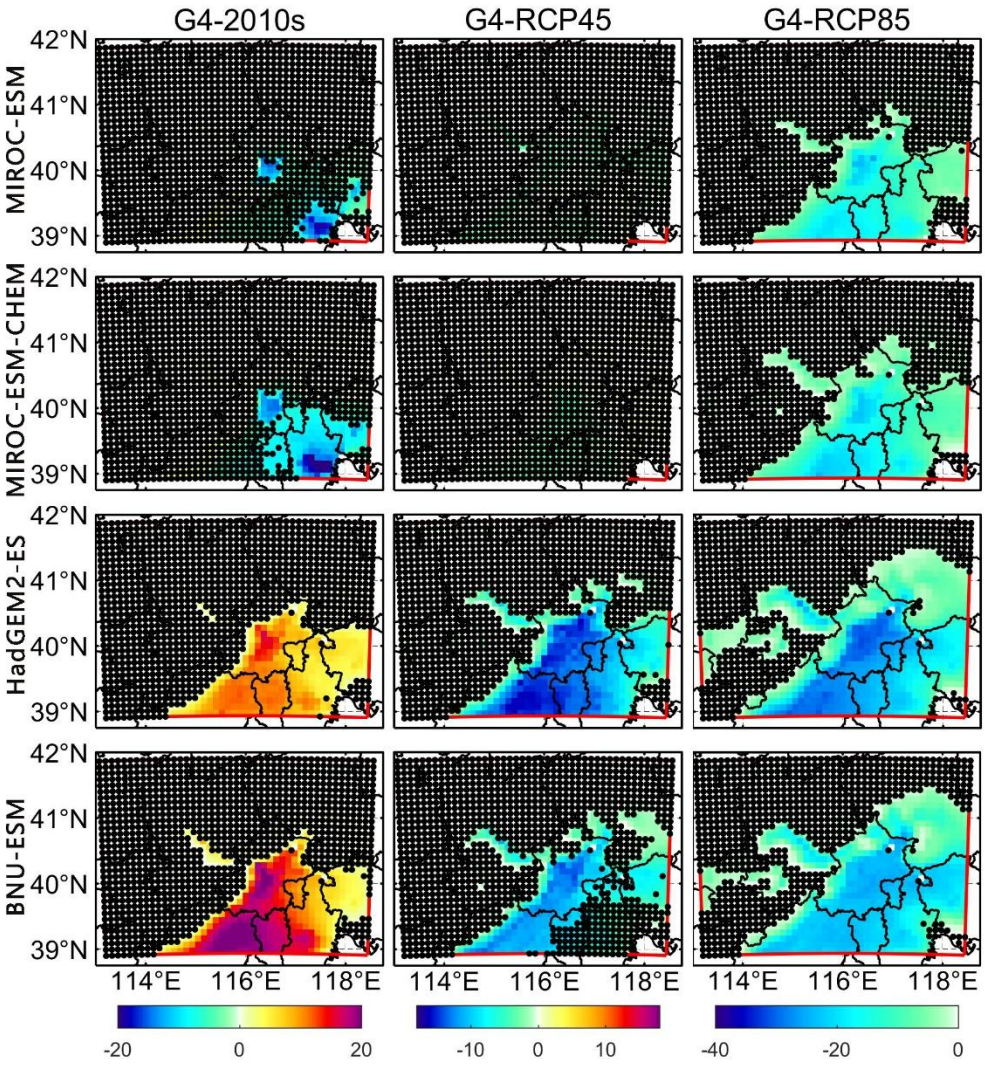

Figure S9. Number of days with AP > 32℃ differences under different scenarios over 2060-2069: G4-2010s (left column), G4-rcp4.5 (second column) and G4-rcp8.5 (right column) based on WRF_QDM method. 2010s means the results simulated during 2008-2017. From top to bottom are MIROC-ESM (a-c), MIROC-ESM-CHEM (d-f), HadGEM2-ES (g-i) and BNU-ESM (j-l), respectively. Stippling indicates grid points where differences or changes are not significant at the 5% level according to the Wilcoxon signed rank test.