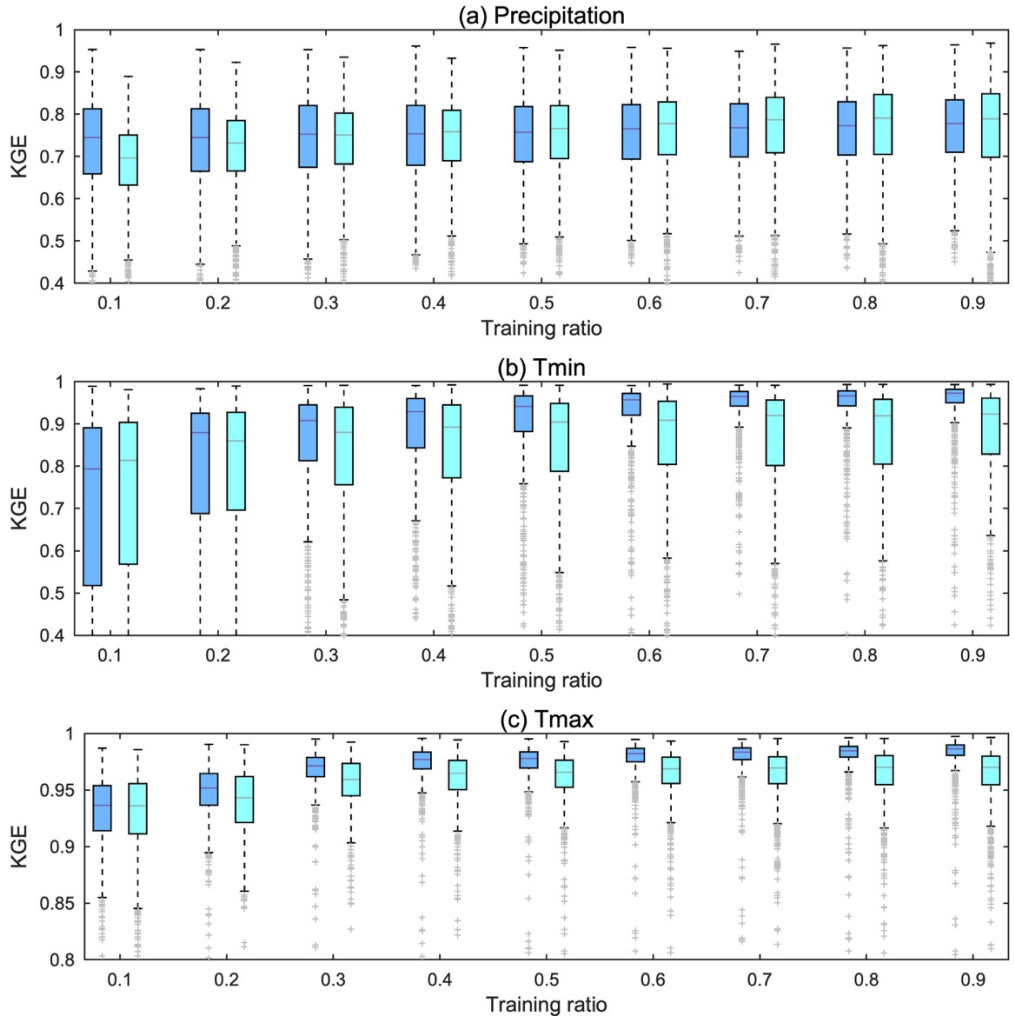

Figure S1. The accuracy of estimated precipitation, T_{min} and T_{max} for different training ratios of station observations. For example, if the training ratio is 20%, only the first 20% of observations from 1979 to 2018 (eight years if the station has no missing values during the 40-year period) are taken as known values; the remaining observations are taken as missing values, which can be used to independently evaluate estimates (see Sect. 3.3). KGE is used to represent the accuracy of precipitation and temperature estimates (see Sect. 3.4).

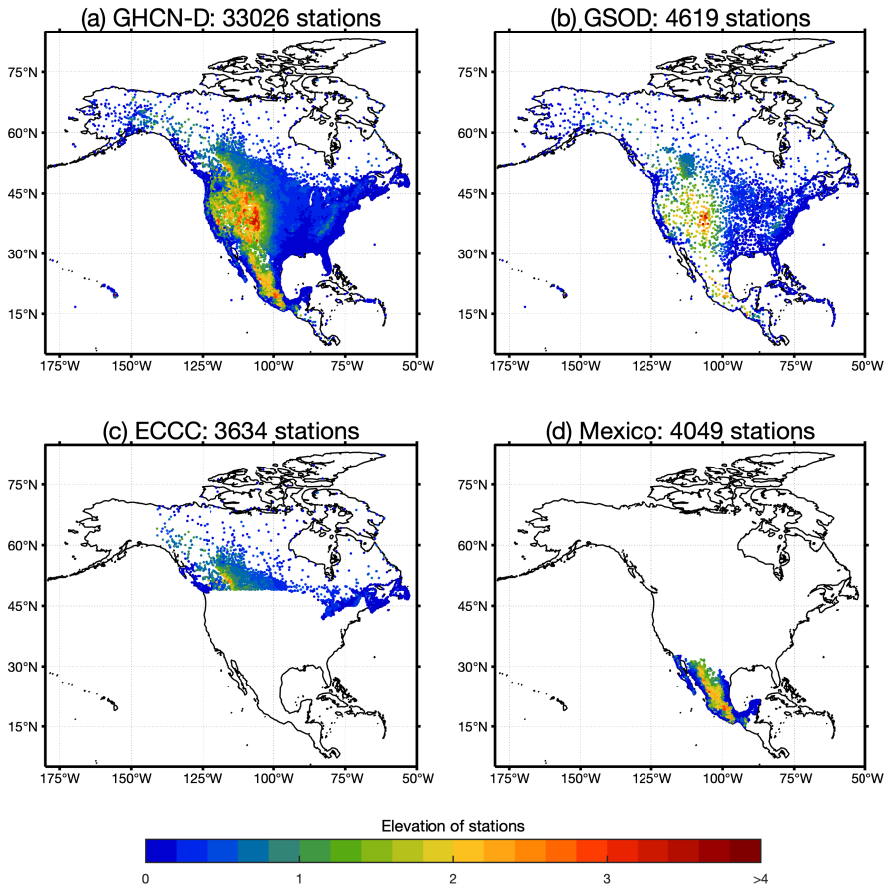

Figure S2. The spatial distributions of stations from (a) GHCN-D, (b) GSOD, (c) ECCC, and (d) Mexico.

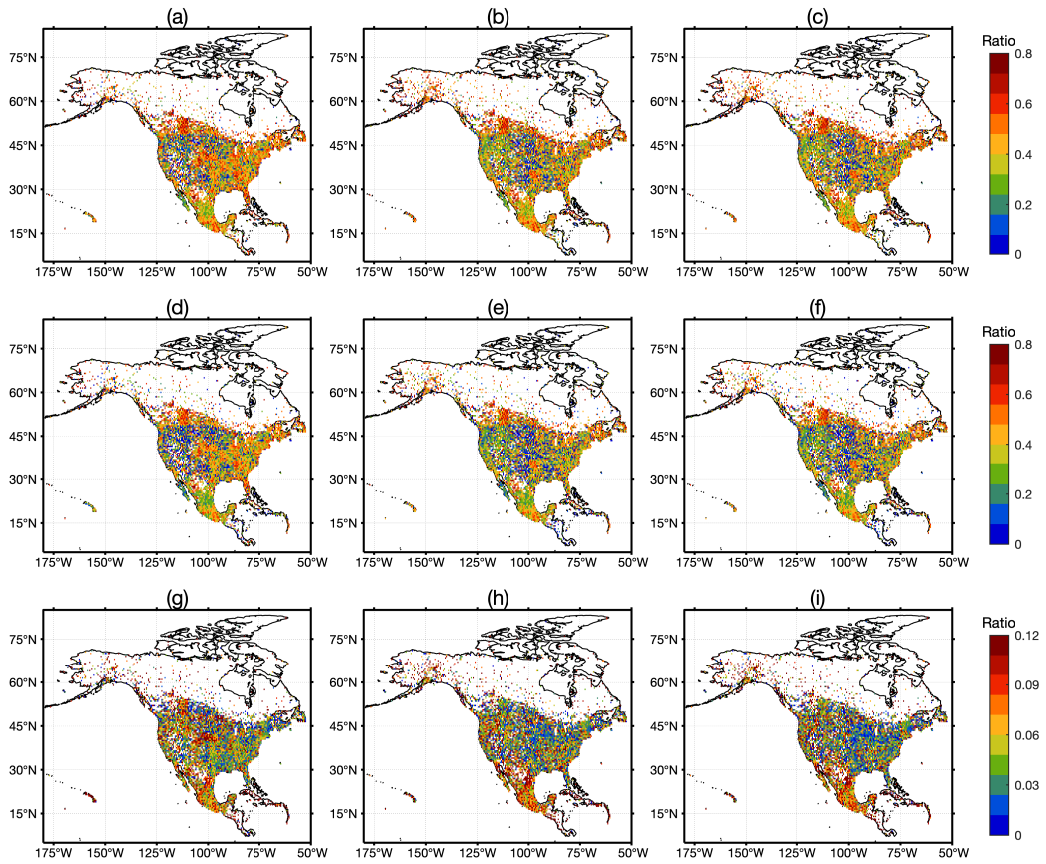

Figure S3. The distributions of (a-c) the ratio between the numbers of all missing values and all days from 1979 to 2018 (i.e., ratio-2 in Figure 3), (d-f) the ratio between the numbers of missing values in the reconstruction period and all days from 1979 to 2018, and (g-i) the ratio between the numbers of missing values in the observation period and all days in the observation period (i.e., ratio-1 in Figure 3). The three columns from left to right represent precipitation, T_{min}, and T_{max}, respectively. The maps are at the resolution of 0.5°. The ratio of each grid cell is the mean value of all stations within this grid cell.

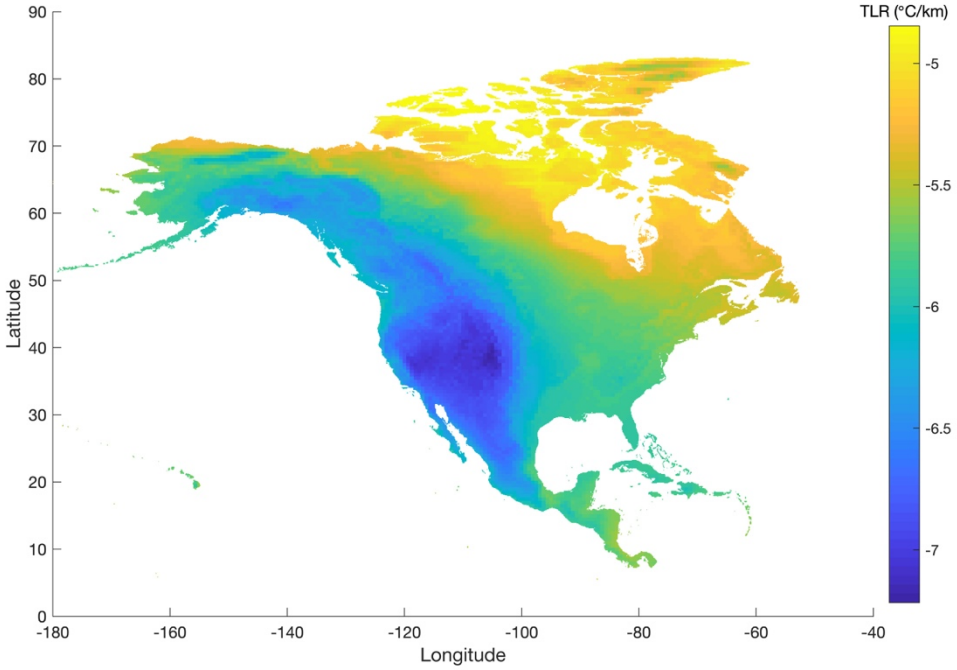

Figure S4. The mean monthly temperature lapse rate (TLR) over North America from 1980 to 2018 derived based on the linear regression between MERRA-2 geopotential heights and air temperature at pressure levels from 300 hPa to 1000 hPa.

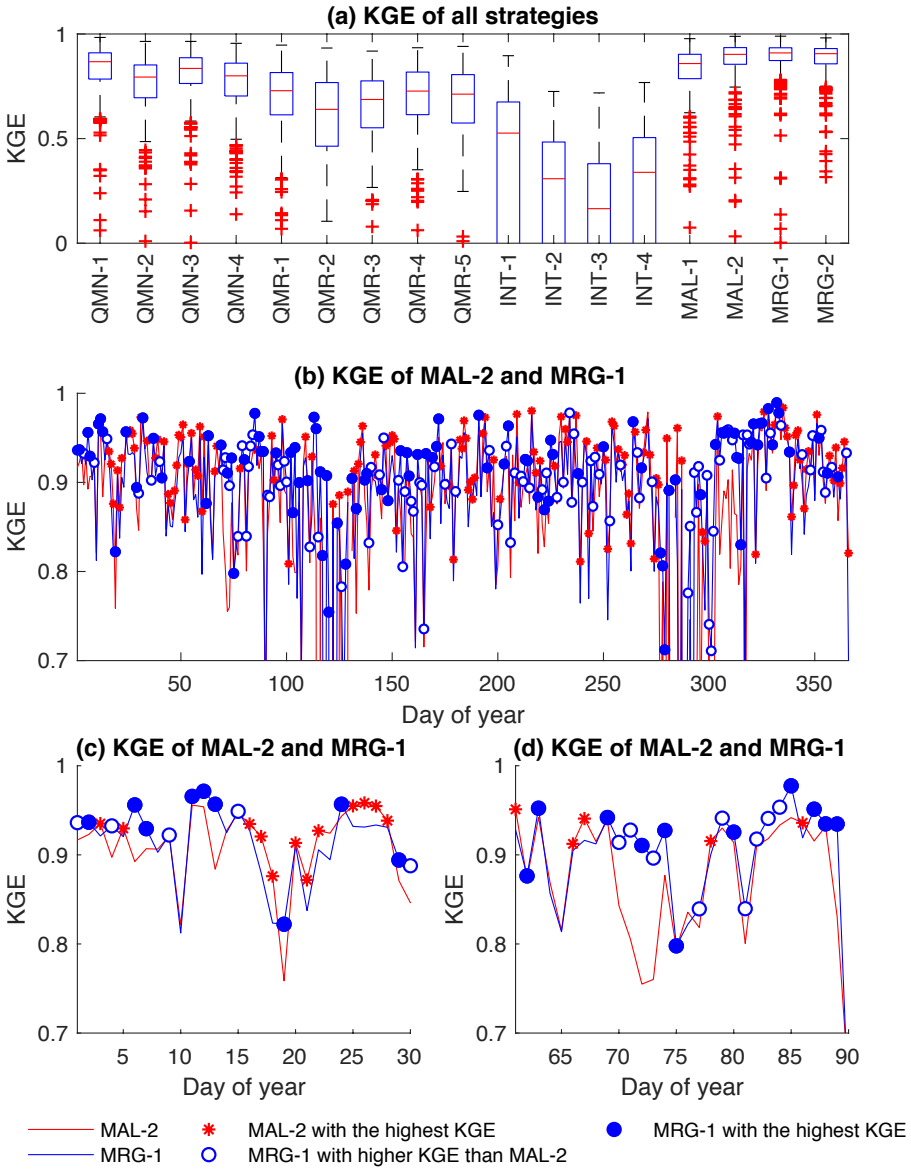

Figure S5. The KGE of T_{min} for each day of year (DOY: 1-366) for one typical station (ID: ECCA001123721) during the observation period. (a) is the KGE of 17 strategies for DOY 1-366. (b) is the KGE of MAL-2 and MRG-1 for DOY 1-366. (c) and (d) are subsets of (b), corresponding to DOY 1-30 and DOY 61-90, respectively. MRG-1 shows higher KGE than MAL-2 for 196 DOY, among which MRG-1 shows the highest KGE among all 17 strategies for only 98 DOY. In contrast, MAL-2 shows higher KGE than MRG-1 for 167 DOY, among which MAL-2 shows the highest KGE among all 17 strategies for 138 DOY. This is the reason why MRG-1 achieves higher KGE than MAL-2, but its contribution ratio is lower than MAL-2.

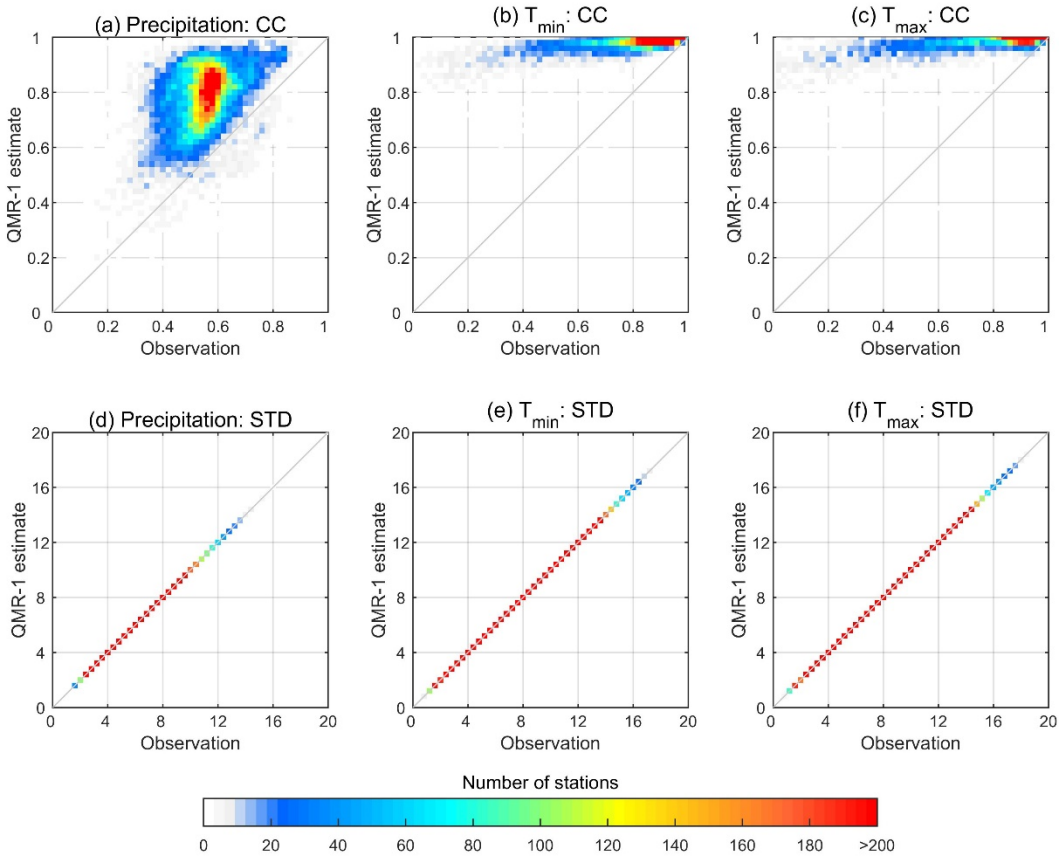

Figure S6. (a-c) correlation coefficient (CC) and (d-f) standard deviation (STD) between station observations/estimates and the concurrent reanalysis estimates. X-axis represents the CC/STD between station observations and reanalysis estimates in the observation period. Y-axis represents the CC/STD between QMR-1 estimates and reanalysis estimates in the observation period. QMR-1 is quantile mapping based on ERA5 data.

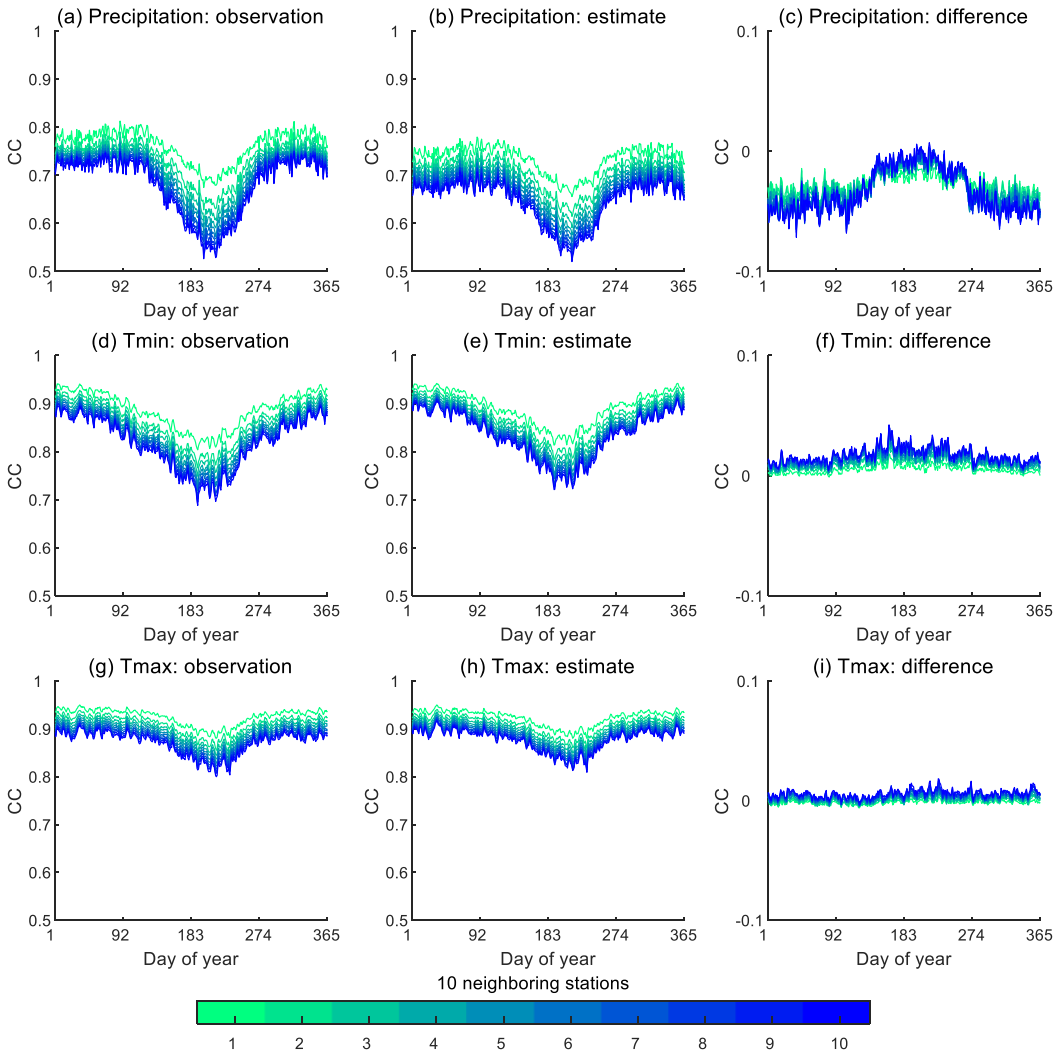

Figure S7. Similar with Figure 8, but the SCD estimates are replaced by observations whenever possible, and the CC calculation uses all data from 1979 to 2018.

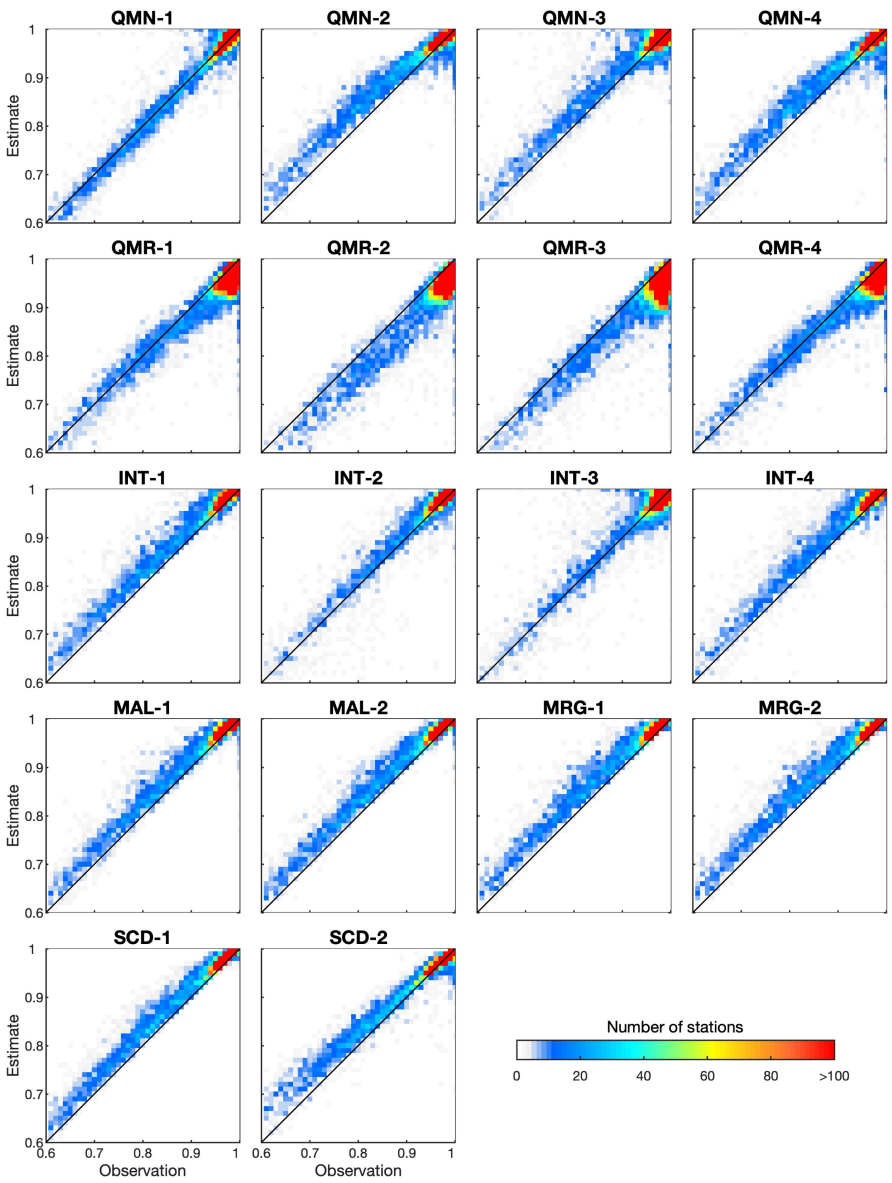

Figure S8. Similar with Figure 10, but for T_{max}.

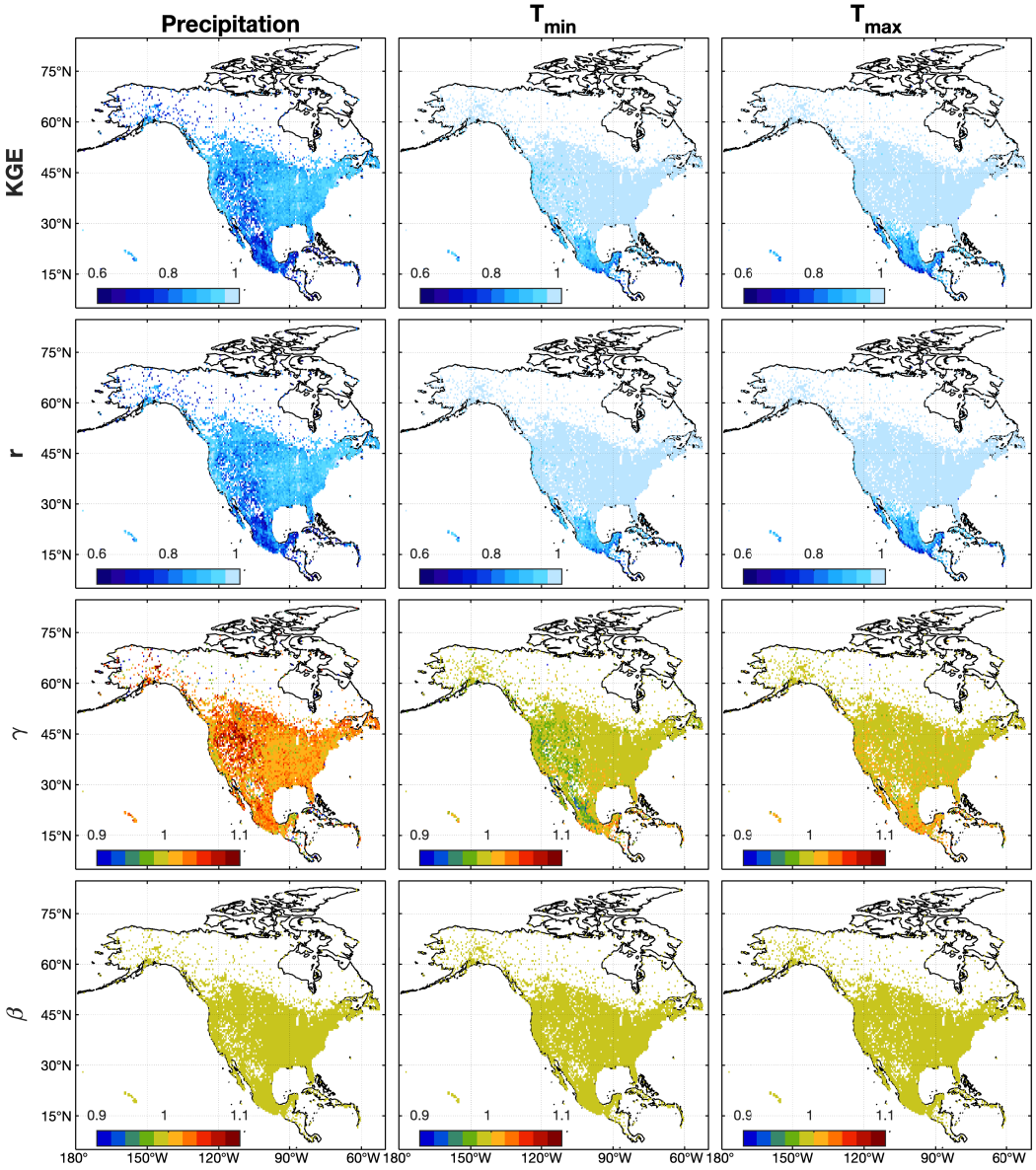

Figure S9. The spatial distributions of KGE and its three components (r is CC, β is the bias ratio, and γ is the variability ratio) for SCD estimates (after correction) over North America during the observation period. The maps are drawn at the 0.5° resolution to avoid stacked station points. For each 0.5° grid, the median values of all nested stations are used.

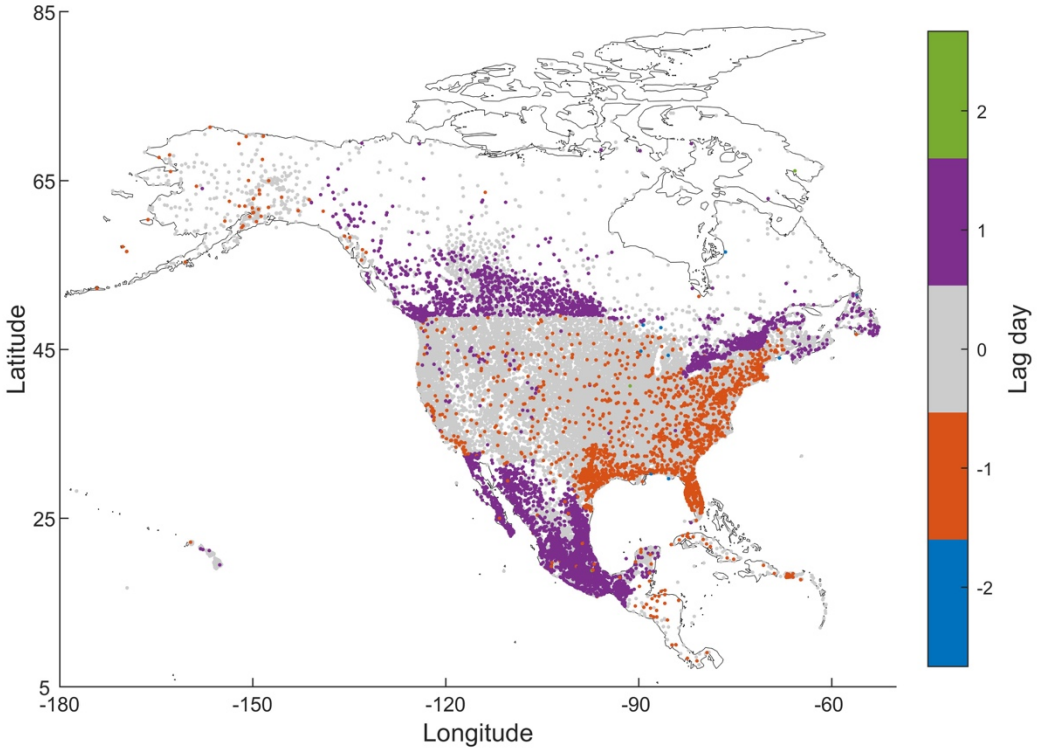

Figure S10. The time lag of daily precipitation from 32590 stations used in this study. First, daily reanalysis precipitation from ERA5, JRA-55, and MERRA-2 is calculated as the accumulation from 0 to 24 UTC. Then, Spearman correlation coefficients are calculated between station and reanalysis precipitation series. Station precipitation series are shifted by -2, -1, 0 (no time lag), 1, and 2 days. For each reanalysis product, the lag day with the highest Spearman correlation coefficient is adopted. If ERA5, JRA-55, and MERRA-2 have the same lag day estimation, the station precipitation series is regarded to have corresponding time lag. In this figure, 6418 stations show lag days that are not zero, indicating that their observations may not correspond to 0-24 UTC of their reporting date. Hourly-scale estimation of time lag can be achieved by shifting reanalysis series at 1-h or 3-h step.