Supplement of Earth Syst. Dynam., 17, 1365–1380, 2026
https://doi.org/10.5194/esd-17-1365-2026-supplement

Earth System Dynamics
Open Access

Supplement of

Everyday weather in a warmer world

Rhidian Thomas et al.

Correspondence to: Rhidian Thomas (r.h.thomas@reading.ac.uk)

S1 Surface heat budget

The surface heat budget is calculated as:

$$Q_{net} = DLWRF - ULWRF_{sfc} + DSWRF_{sfc} - USWRF_{sfc} - LHTFL - SHTFL, \quad \text{(S1)}$$

where $DLWRF_{sfc}$ and $ULWRF_{sfc}$ are the downward and upward longwave fluxes at the surface, respectively; $DSWRF_{sfc}$ and $USWRF_{sfc}$ are the downward and upward shortwave fluxes at the surface, respectively; and $LHTFL$ and $SHTFL$ are the surface latent and sensible heat fluxes, respectively.

The budget is calculated locally for each (lat, lon, day) in a given region. Then, paired differences in the local budget terms are calculated between the +2K and 20CRv3$^+$ experiments. The differences are binned according to the 2m-temperature (T2m) percentile in 20CRv3$^+$ of that (lat, lon, day). Note that the sign convention of the differences shown in Fig. S5 is consistent with equation S1: for example, a positive value for the $DLWRF_{sfc}$ difference term in Fig. S5 indicates greater surface *warming* due to downward longwave radiation in the +2K experiment. Similarly, a negative value in the $LHTFL$ difference in Fig. S5 indicates greater surface *cooling* from evaporation in the +2K experiment.

Flux terms were not archived for the +2K+CO2 experiment, hence only differences in the +2K experiment are shown.

S2 Validation of 20CRv3$^+$ precipitation

In this section we validate the precipitation statistics of 20CRv3$^+$ against observations. The fairest comparison is to a daily gridded observational dataset, as in Fig. 8, but this is more difficult outside the UK. We therefore validate 20CRv3$^+$ against daily station precipitation observations for a small number of sub-regions within each of the four well-observed regions in Fig. 1. The sub-regions are chosen to sample potentially different climatological conditions within each region (e.g. mountains or plains, coast or inland). Observations are taken from the Global Land and Marine Observations Database (GLAMOD) v8.1 (Noone et al., 2021). The sub-regions are shown in Fig. S7, the daily time series in Fig. S8, and mean and daily precipitation variability are shown in Fig. S9.

In most sub-regions, the largest localised peaks at individual stations are not captured by the reanalysis. This is not a surprise, as 20CRv3$^+$ values represent averages over the sub-regions — when compared to the sub-regional average from the gridded observational dataset over Central England, the reanalysis performs much better (compare Figs. 8e and S8c). The reanalysis also struggles with orographic precipitation, for example at MONT-AIGOUAL GSN station (1567m elevation) compared to the other French stations where it performs much better (below 100m elevation). This is consistent with the poor performance seen over Scotland in Fig. 8b compared to the gridded observations in Fig. 8a. Otherwise, the mean and variability in 20CRv3$^+$

is broadly within the range of values at individual stations for most sub-regions. This builds confidence in the capability of the reanalysis system to represent precipitation variability over the larger well-observed regions.

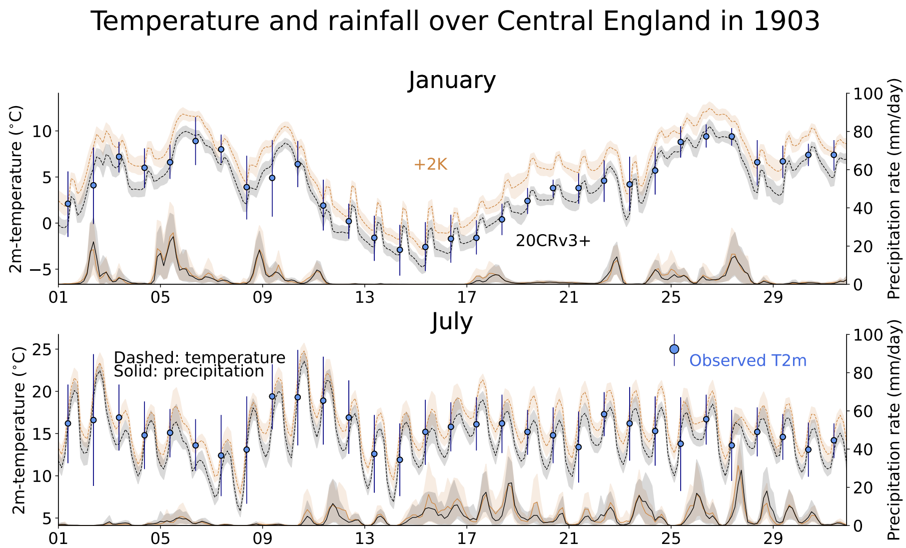

Figure S1. As Fig. 2, but showing the +2K experiment. Time series of 3-hourly 2m-temperature (dashed line) and precipitation rate (solid line) averaged over central England (3-0°W, 51.5-54.0°N, see boxed region in Fig. 8). The 20CRv3^{+} experiment is in black and the +2K experiment is in yellow. Lines show the ensemble means and shading denotes the 10-90% range across the ensemble. Blue markers show daily mean temperatures from the Central England Temperature (CET) series (Legg et al., 2025), plotted at 09:00am UTC each day. Vertical bars on each marker indicate the CET daily maximum and minimum temperatures.

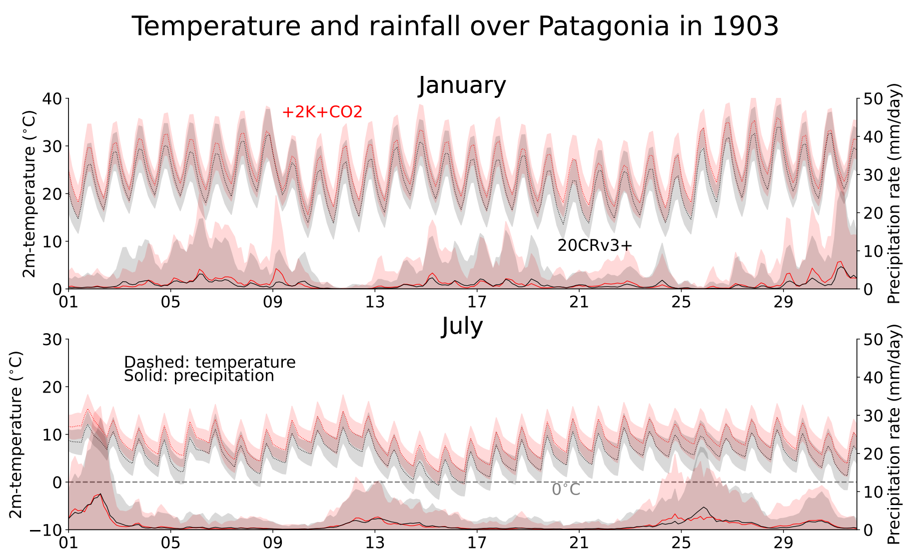

Figure S2. As Fig. 2, but for a region with fewer nearby pressure observations. The northern Patagonia region is shown by the green box in Fig. 1. Lines and shading are as in Fig. 2.

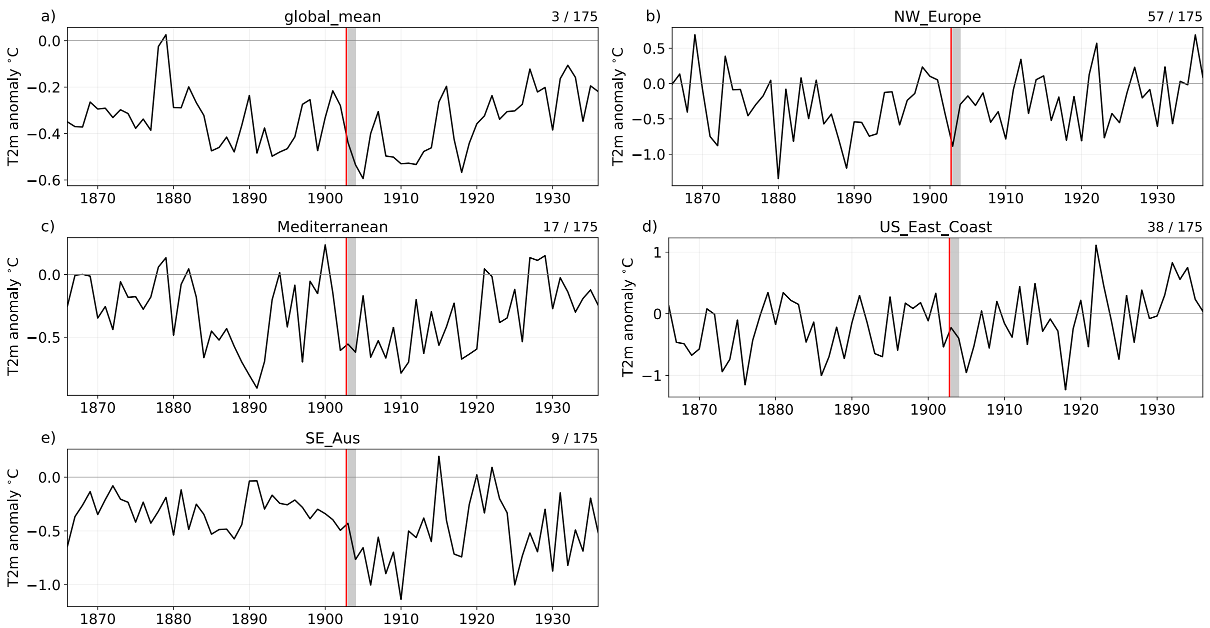

Figure S3. Annual-mean surface temperature anomalies in HadCRUT5.1 (Morice et al., 2021) for (a) the global mean and (b–e) the four well-observed regions in Fig. 1. Anomalies are with respect to the 1961-1990 reference period and include temperatures over both land and ocean. Grey shading shows the year 1903 and the red line shows the Santa Maria eruption in October 1902. The rank of 1903 within the 175-year HadCRUT5.1 dataset for each region is shown in the top right of the plot (rank 1 being the coldest and 175 the warmest).

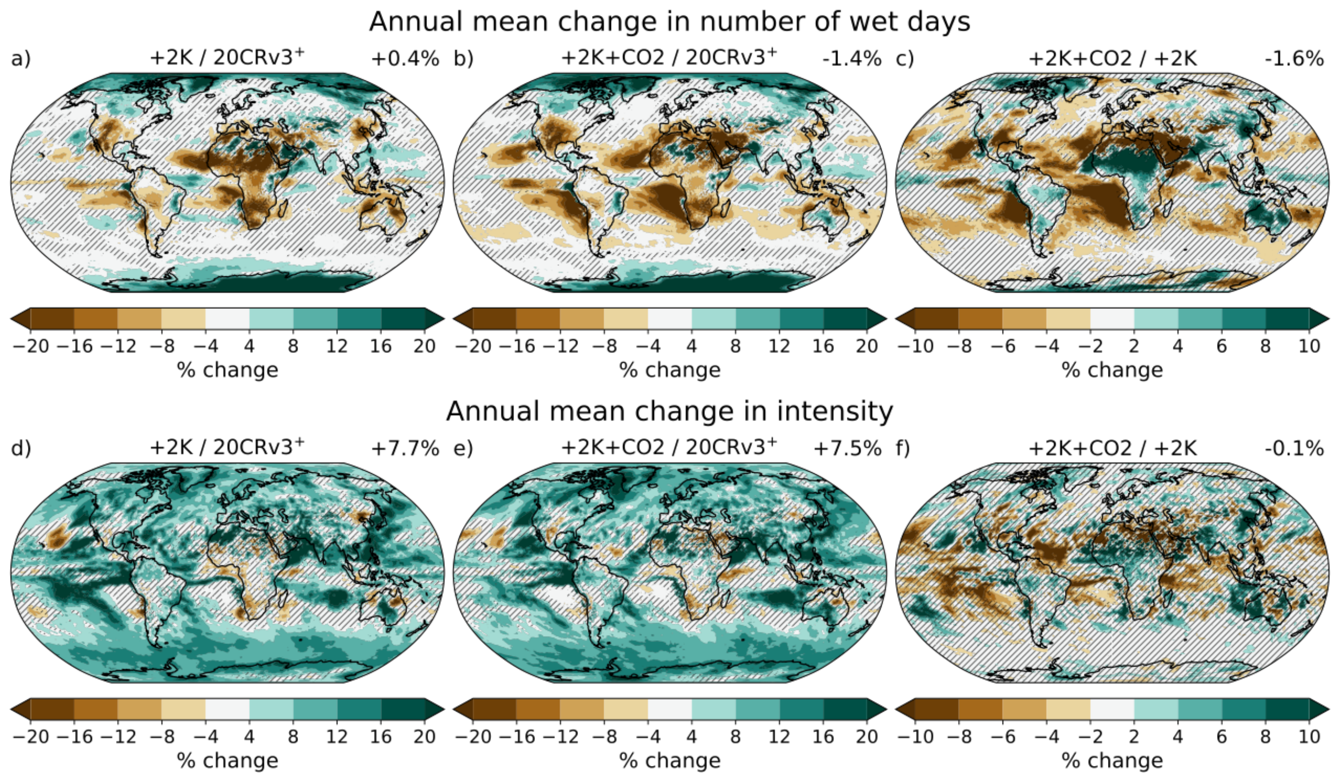

Figure S4. Percentage changes in precipitation (a-c) frequency (number of wet days) and (d-f) intensity in the warmer world experiments. The wet day threshold is 1mm/day. Precipitation intensity is the mean precipitation on wet days. The value at the top right of each plot is the global (area-weighted) average of the changes at each gridpoint, as in Fig. 4. Hatching hides non-significant changes at the 5% level using a two-tailed Student's t-test.

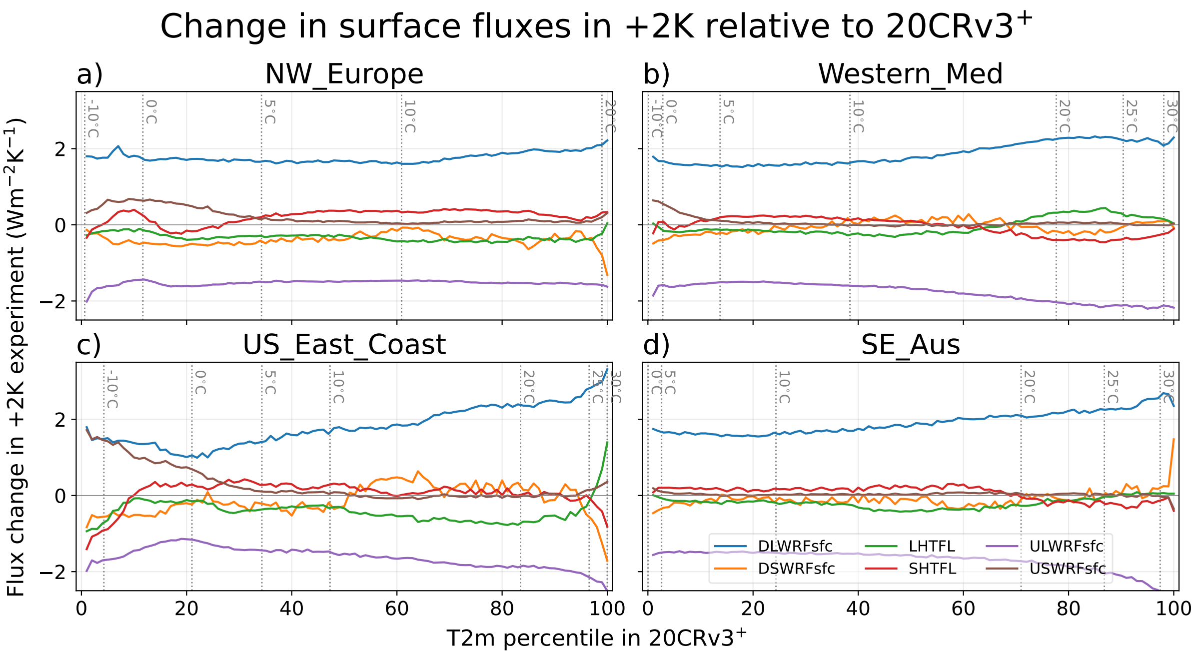

Figure S5. Change in surface fluxes in +2K experiment vs 20CRv3$^+$, as a function of 20CRv3$^+$ temperature percentile. Flux changes are calculated in the same way as the temperature differences in Fig. 5, i.e. paired differences (+2K minus 20CRv3$^+$) at each (lat, lon, day) are binned according to the 20CRv3$^+$ temperature at that (lat, lon, day). Fluxes are defined as downward positive, so that a positive value indicates increased surface heating in the +2K experiment.

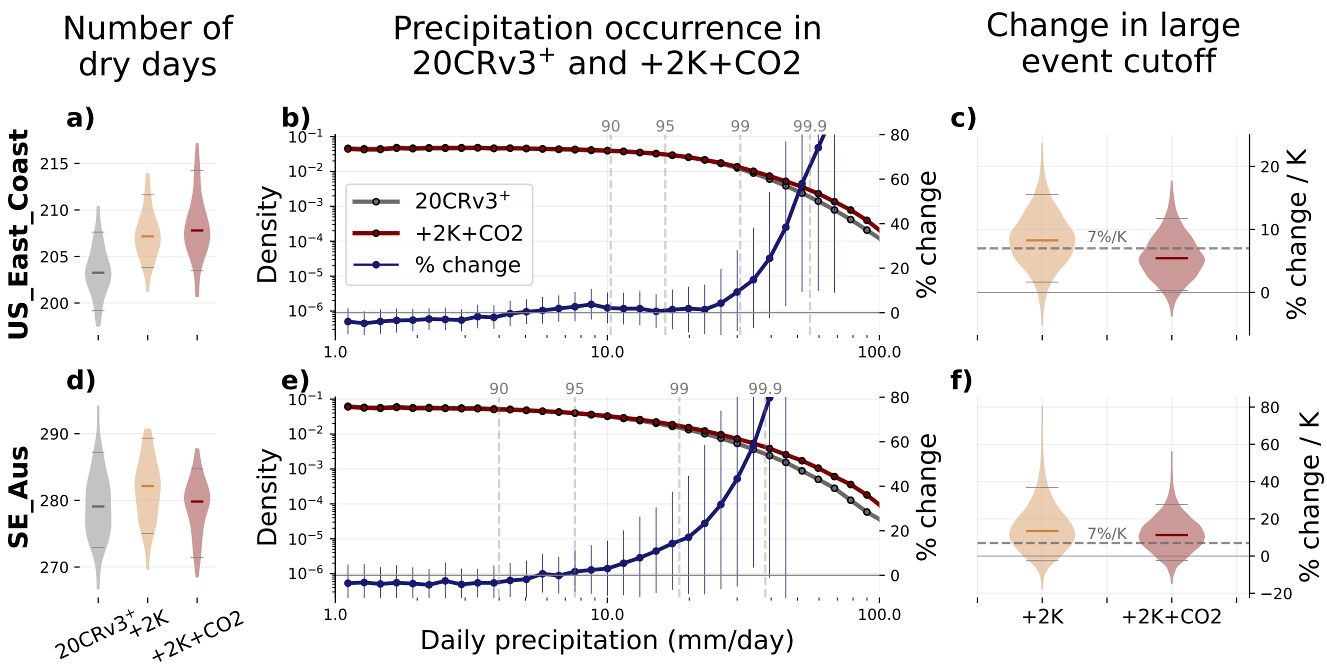

Figure S6. As Fig. 7, but for the US_East_Coast and SE_Aus regions. (a, d): Number of dry days in each experiment (daily precipitation less than 1.04mm). The median is shown by the horizontal line and the violin shows the distribution of ensemble members. (b, e) Daily precipitation occurrence in 20CRv3$^+$ and +2K+CO2. The blue line shows the percentage change in each bin in +2K+CO2. Error bars show the 10-90% range through sampling all combinations of ensemble members, and vertical dashed lines show percentiles of the precipitation distribution in 20CRv3$^+$. (c, f) Ratio of large event cutoff s_L in +2K+CO2 vs 20CRv3$^+$ (see main text). Only land gridcells are used for all calculations. Bins in (b) and (e) are logarithmically distributed with the smallest nonzero bin centred at 1.11 mm/day and successive bin widths increasing by 14.7%.

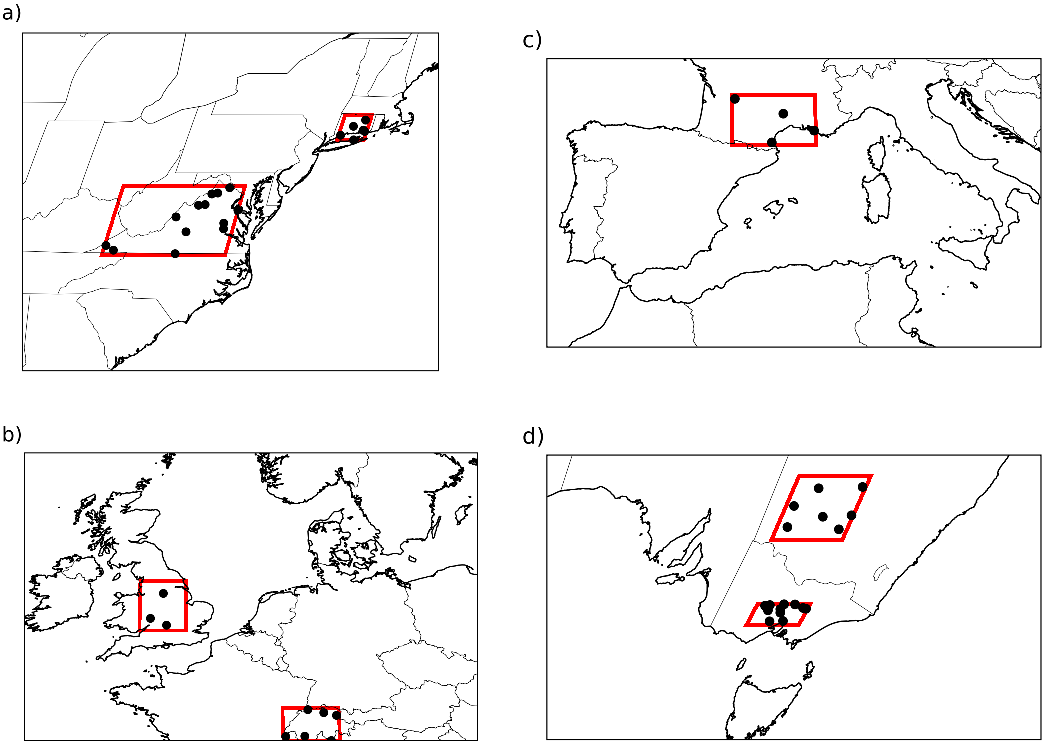

Figure S7. Subregions used for validating 20CRv3^{+} precipitation in Figs. S8 and S9. Subregions are chosen within each of the main well-observed regions in Fig. 1: (a) US_East_coast (subregions: CT and VA), (b) NW_Europe (subregions: Central England and Switzerland), c) Western_Med (subregion: France), and (d) SE_Aus (subregions: South_Victoria and SW_NSW). Markers show the GLAMOD v8.1 (Noone et al., 2021) stations within each subregion for which daily precipitation records from 1903 are available. Only one subregion was chosen in Western_Med due to lower data availability.

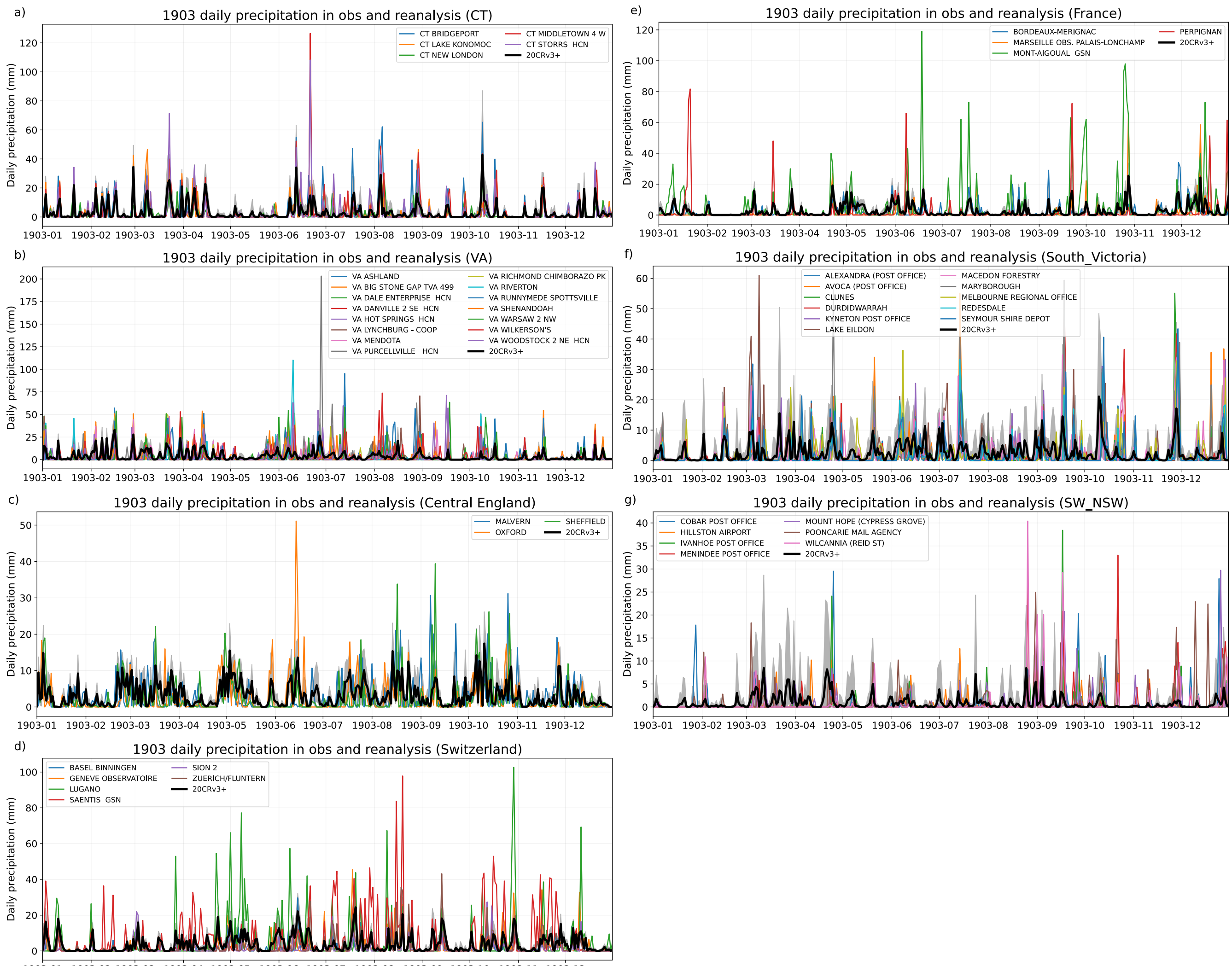

Figure S8. Daily precipitation time series for the validation stations in Fig. S7 and for 20CRv3$^+$. The dark black line shows the ensemble-mean 20CRv3$^+$ precipitation averaged over the whole subregion. The grey shading indicates the 5-95% range across the 20CRv3$^+$ ensemble.

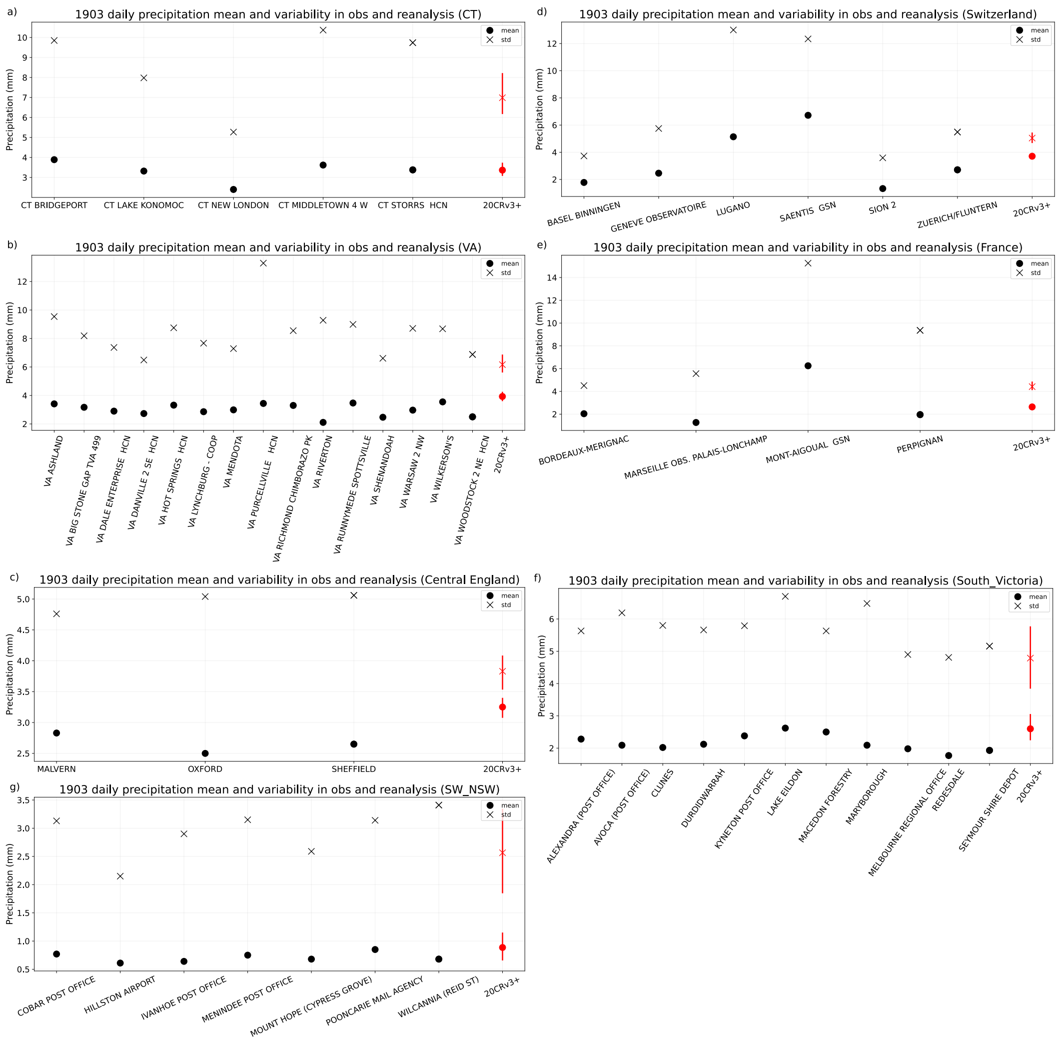

Figure S9. Mean (circles) and daily variability (crosses) in precipitation for the validation stations in Fig. S7 and for 20CRv3$^+$ (red). Statistics are calculated over the full 12 months of 1903. As in Fig. S8, the 20CRv3$^+$ values are averages over the whole subregion. Vertical bars on 20CRv3$^+$ indicate the 5–95% range across the ensemble.

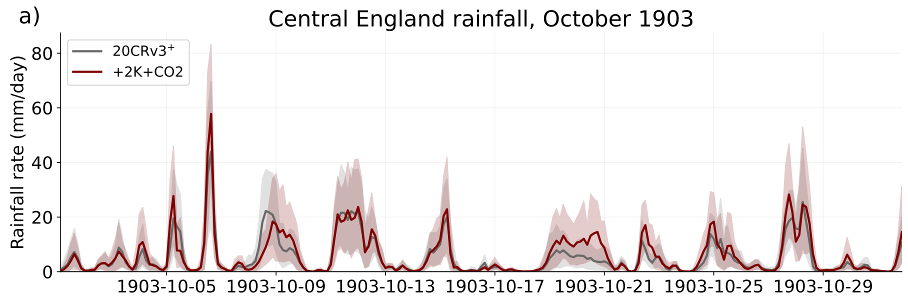

Figure S10. As Fig. 8e, but showing 3hr rainfall. Rainfall over Central England (see the red boxed region in Fig. 8). Shading shows the 10-90% range across the ensemble.

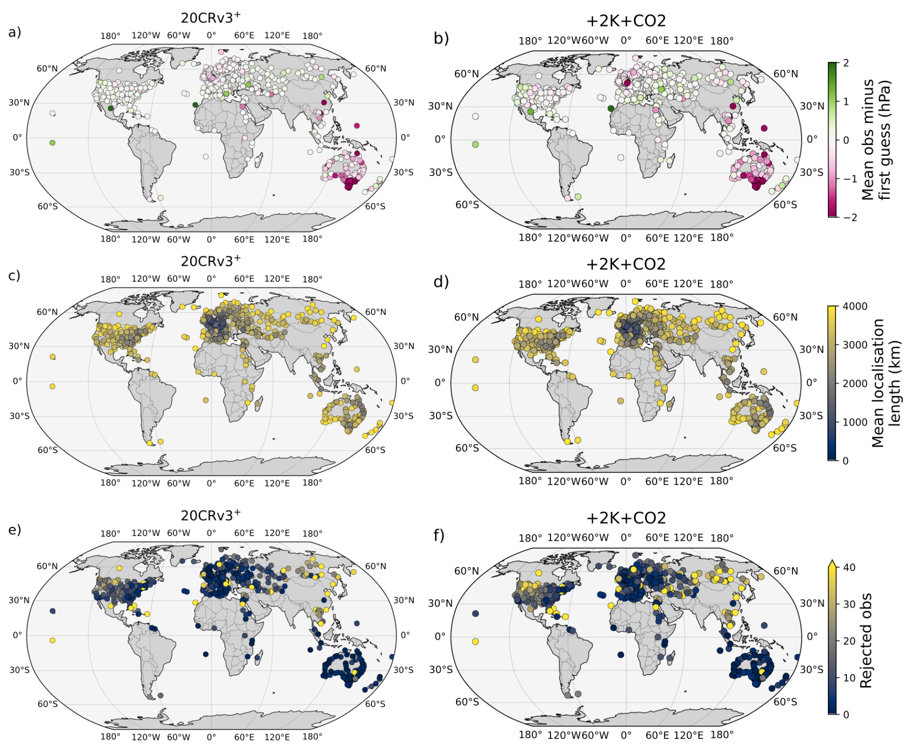

Figure S11. Maps of innovation (observation minus first guess, accounting for observation bias) (a, b), localisation length (c, d), and number of rejected observations (e, f) for land stations in the 20CRv3$^+$ and +2K+CO2 experiments. Innovation and localisation length are average values over 1903 while (c, d) show the total number of rejections across the year.

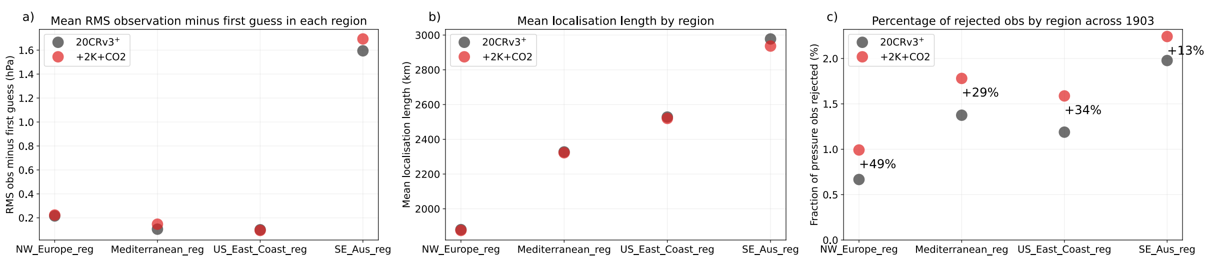

Figure S12. Regional statistics for the innovation (a), localisation length (b), and fraction of rejected observations (c) shown in Fig. S11. In each case, the values denote the average of all land stations within each of the well-observed regions in Fig. 1. Percentage changes relative to 20CRv3^{+} are shown next to the markers in (c).