

## **Supplementary Information**

This document presents supplementary figures showing bias in the annual number of heavy precipitation days (>10mm).

## Calibration task

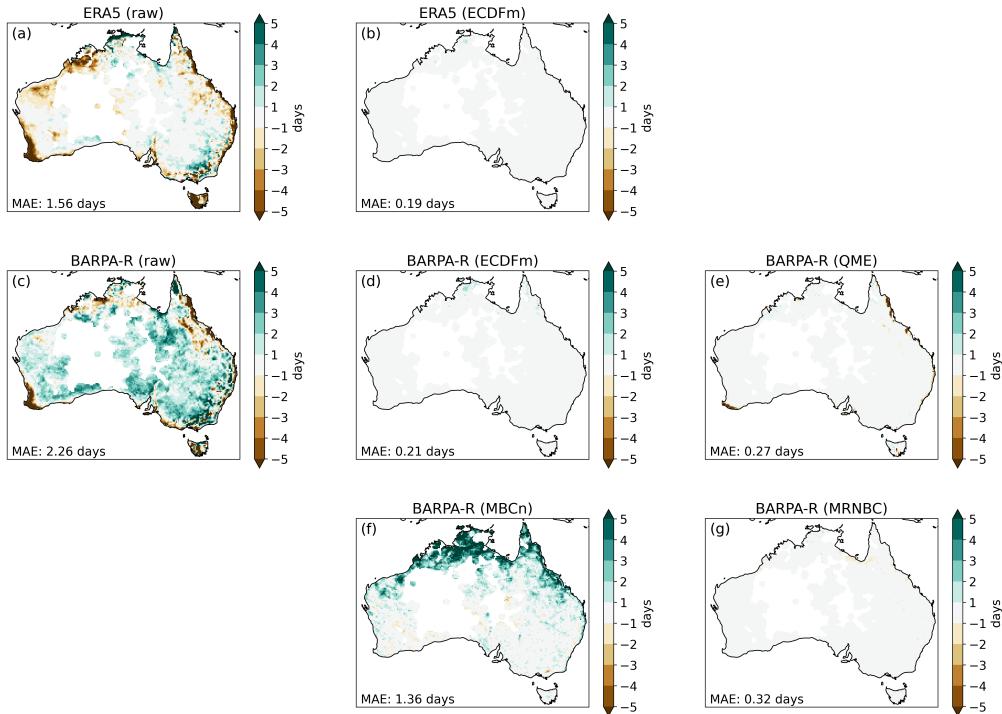


Figure S1: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ERA5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

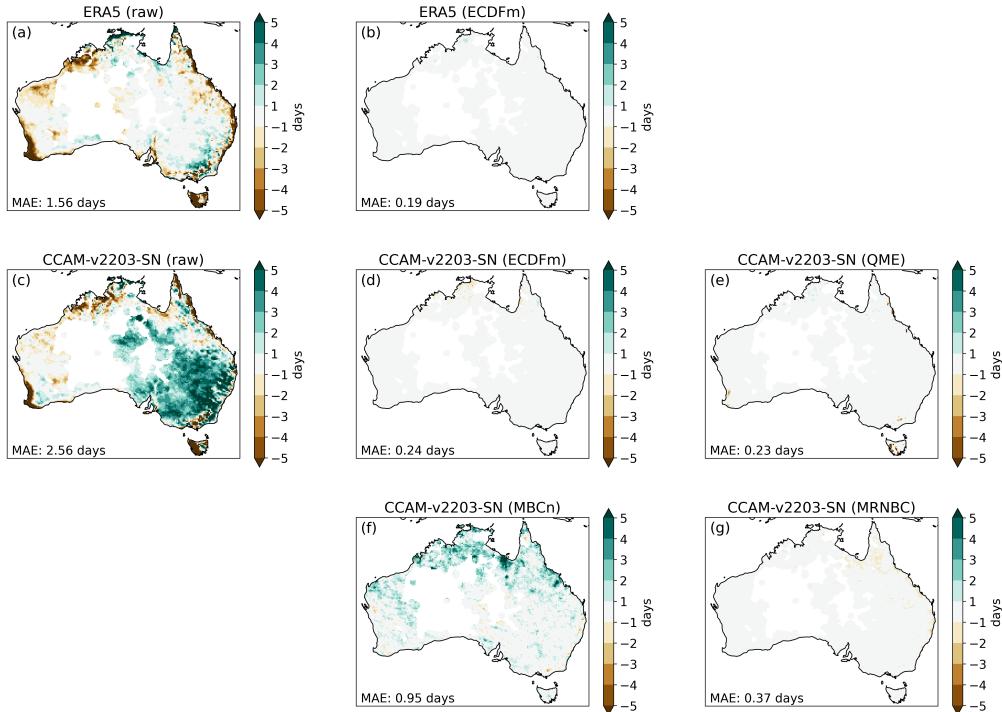


Figure S2: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ERA5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

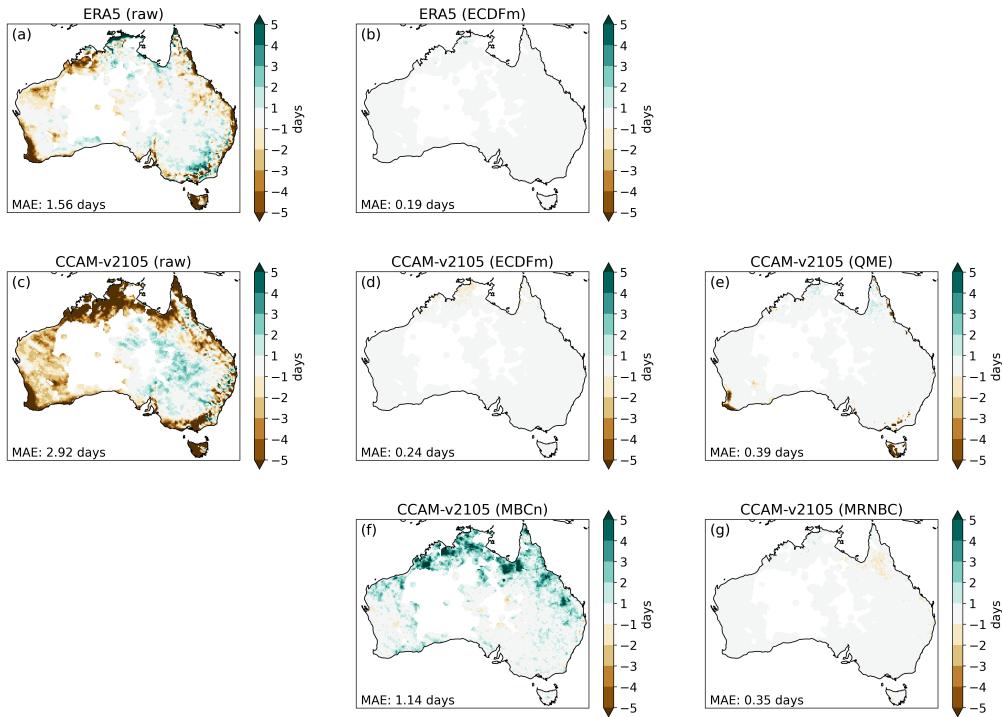


Figure S3: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ERA5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

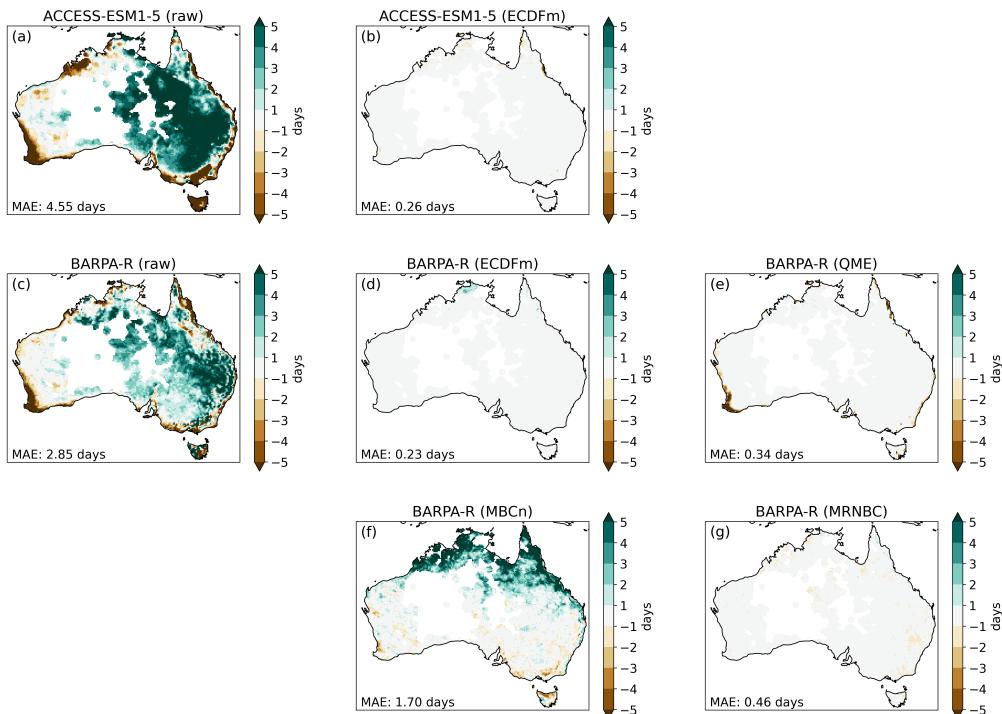


Figure S4: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

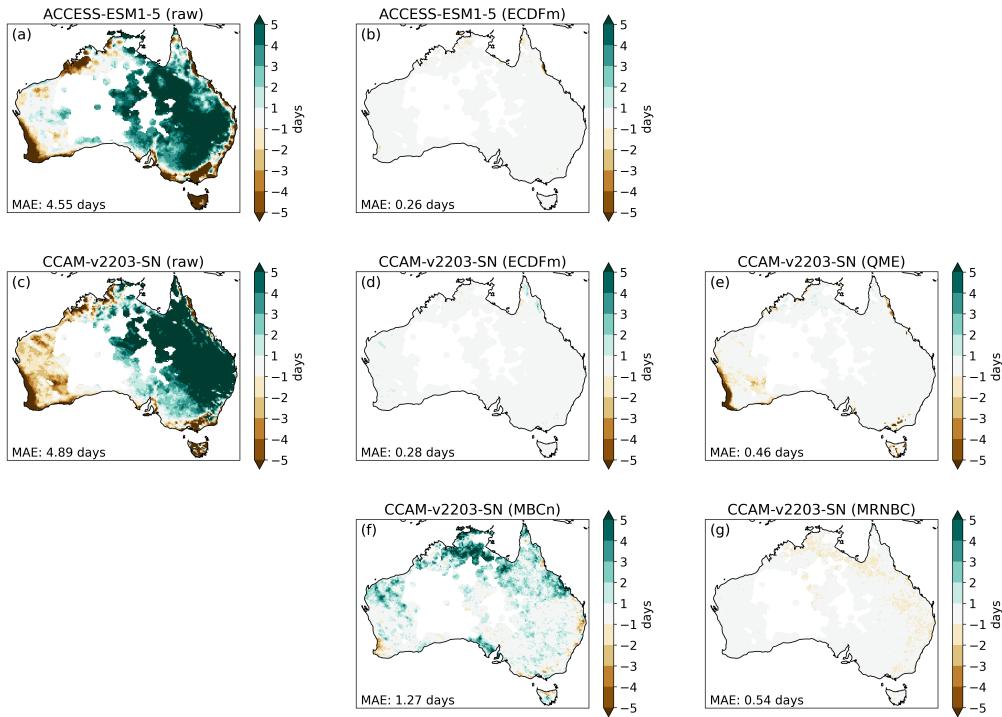


Figure S5: Bias in the annual number of heavy precipitation days ( $>10\text{mm}$ ) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

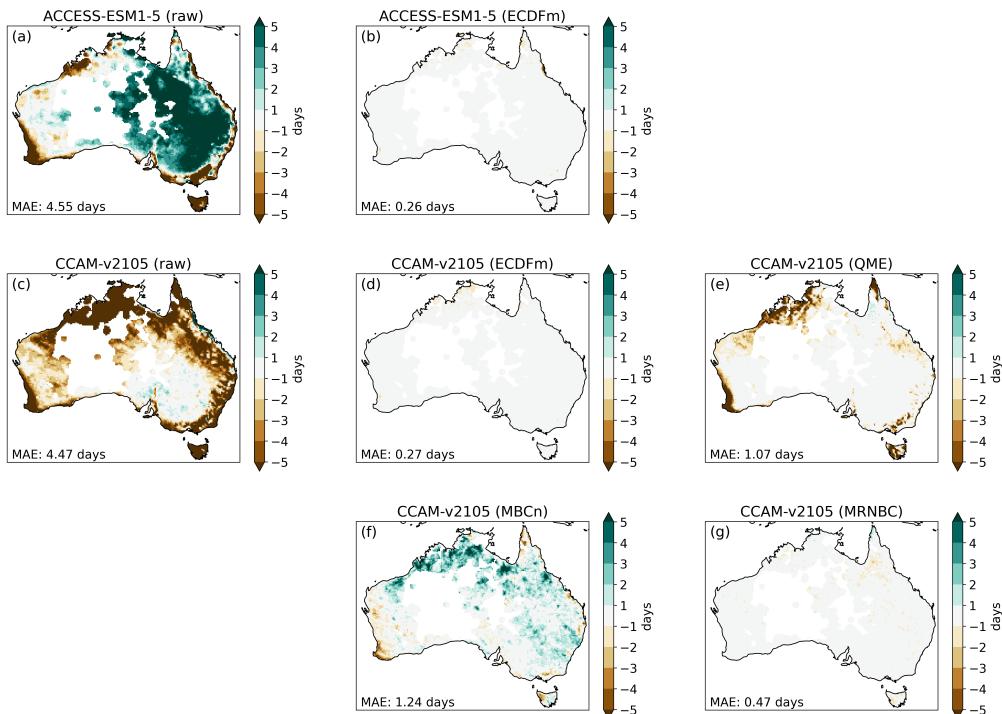


Figure S6: Bias in the annual number of heavy precipitation days ( $>10\text{mm}$ ) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

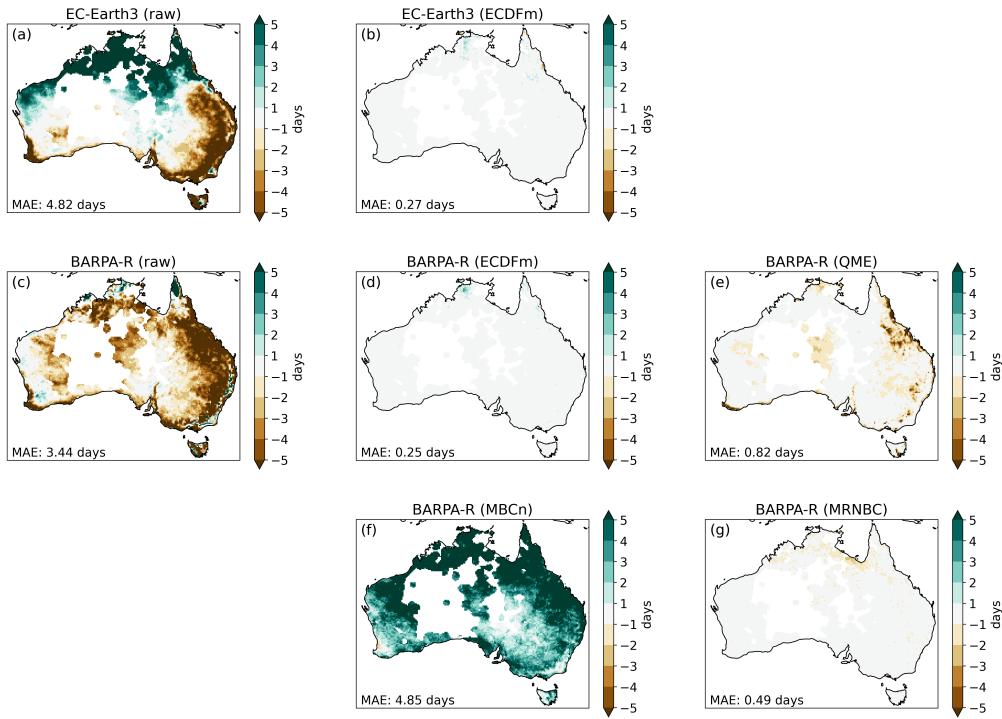


Figure S7: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the EC-Earth3 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

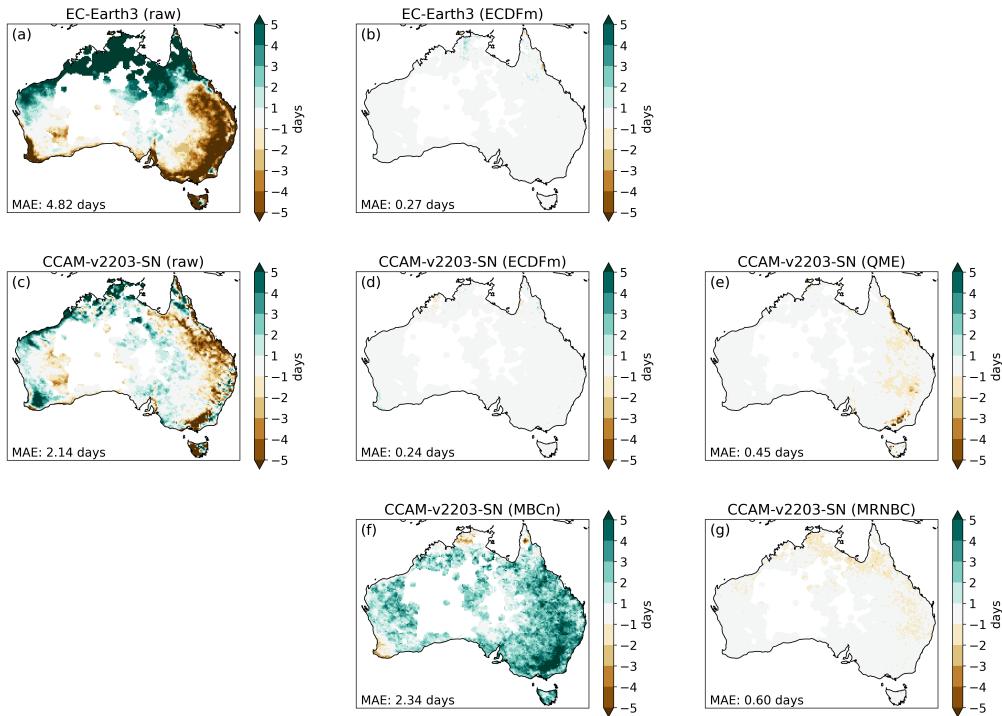


Figure S8: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the EC-Earth3 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

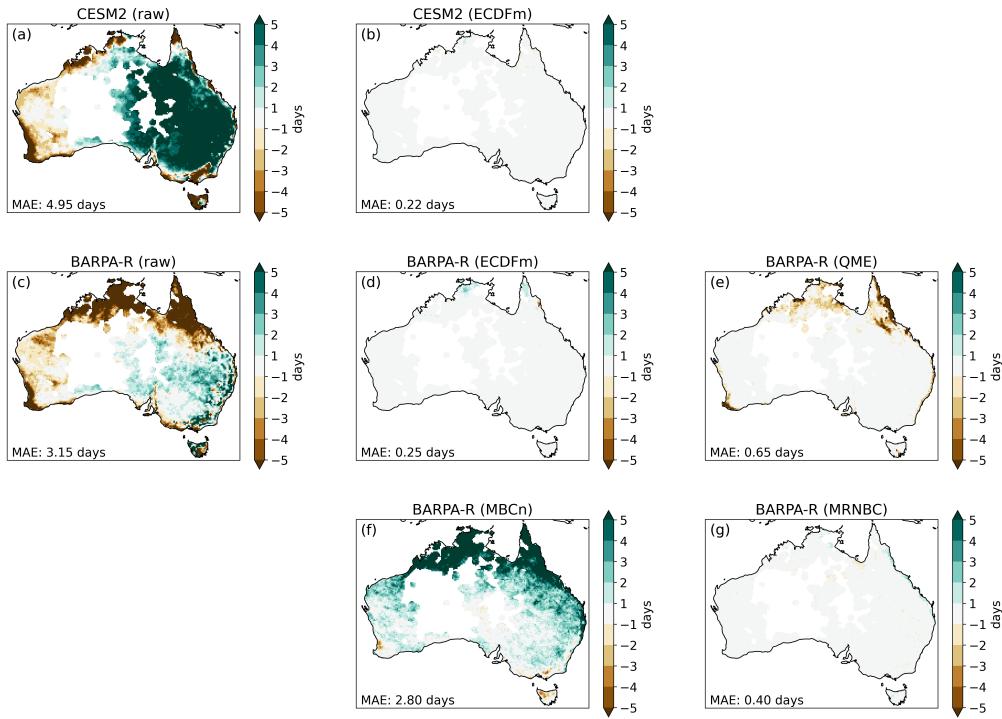


Figure S9: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the CESM2 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

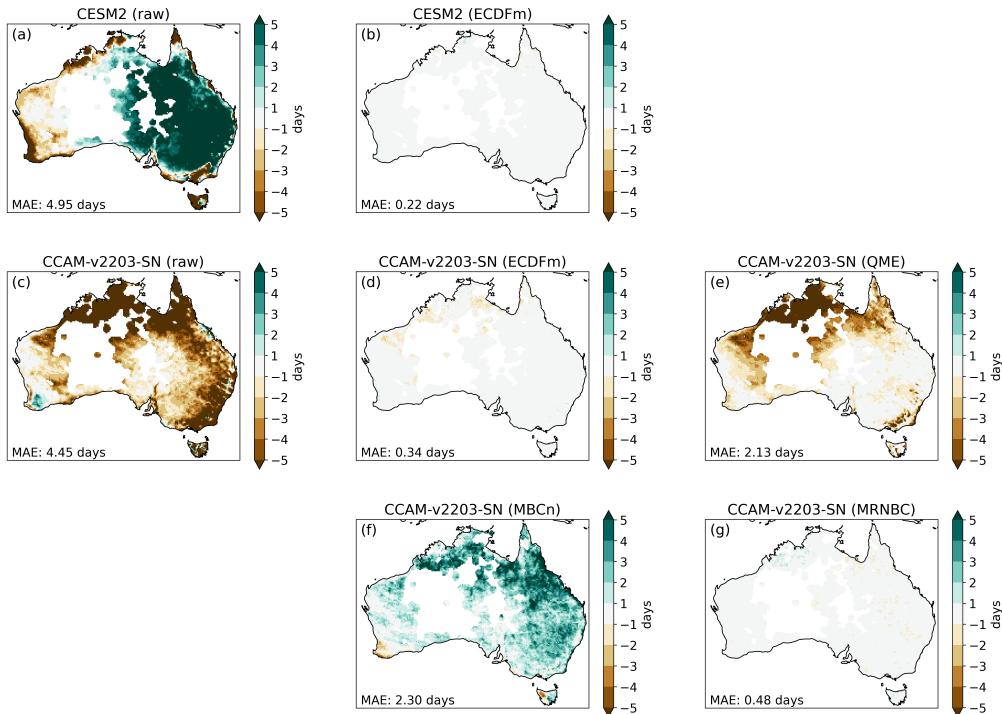


Figure S10: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the CESM2 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

## Cross validation task

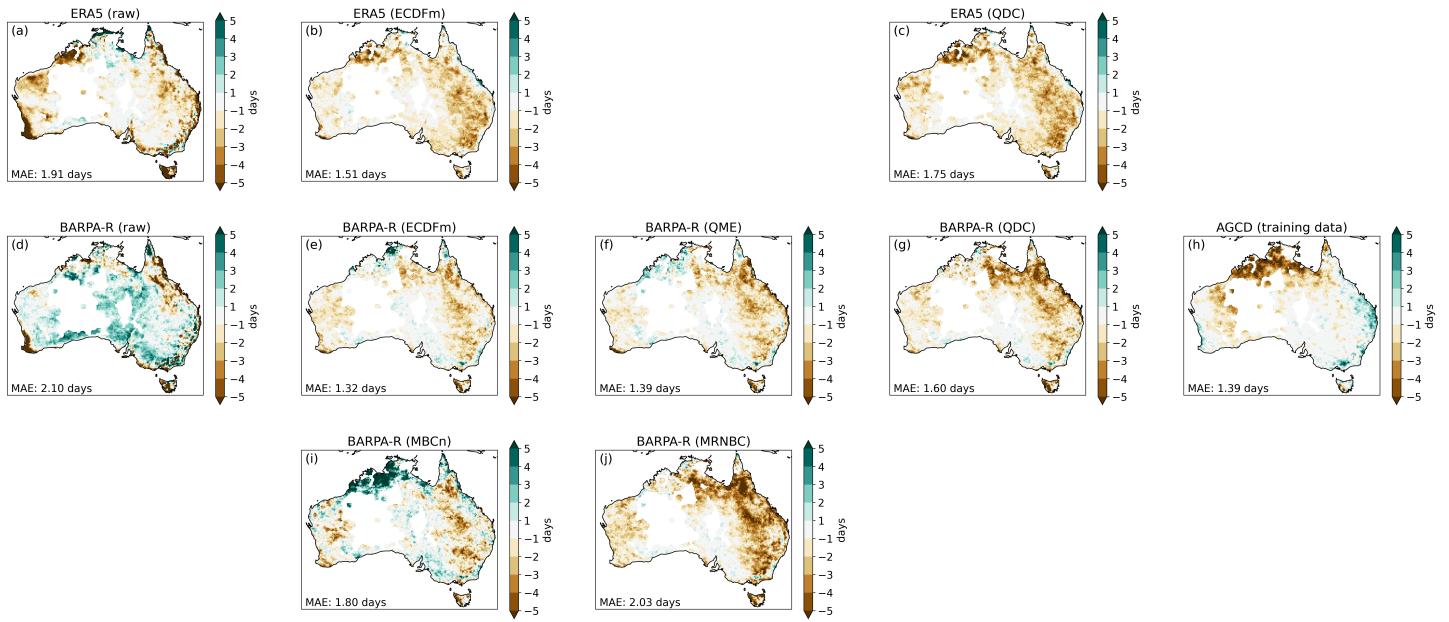


Figure S11: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ERA5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

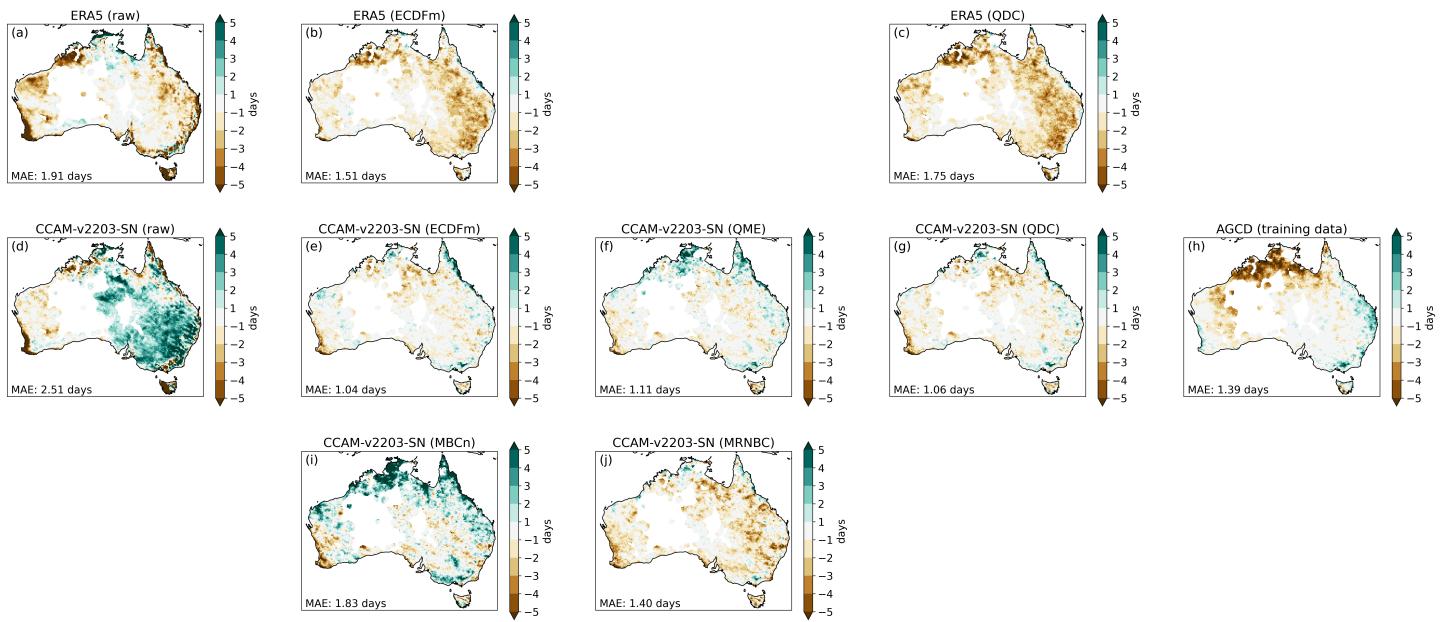


Figure S12: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ERA5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

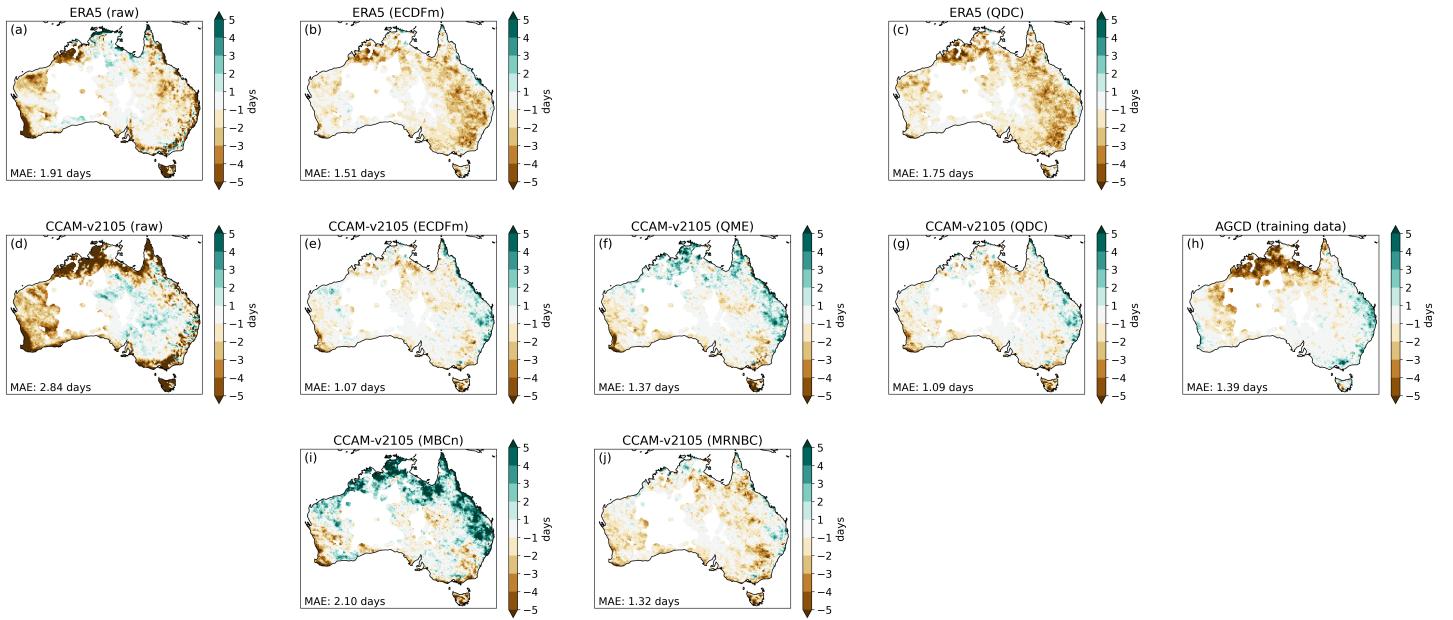


Figure S13: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ERA5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

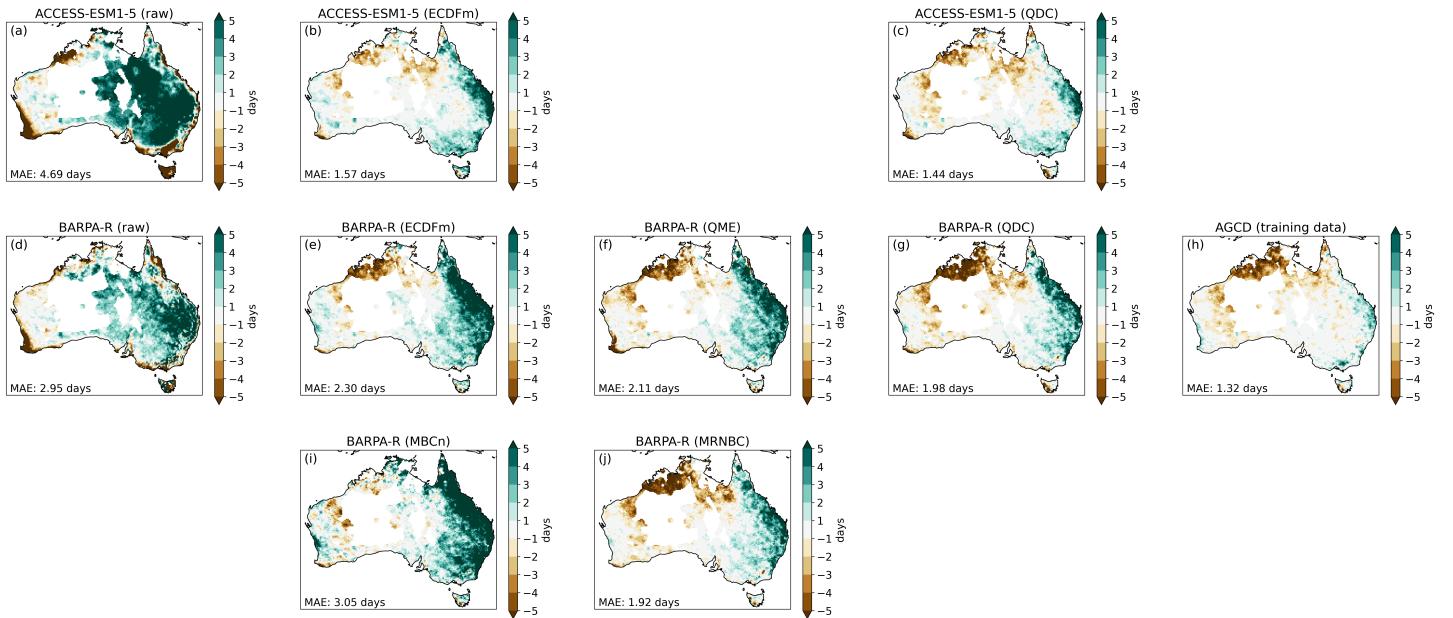


Figure S14: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

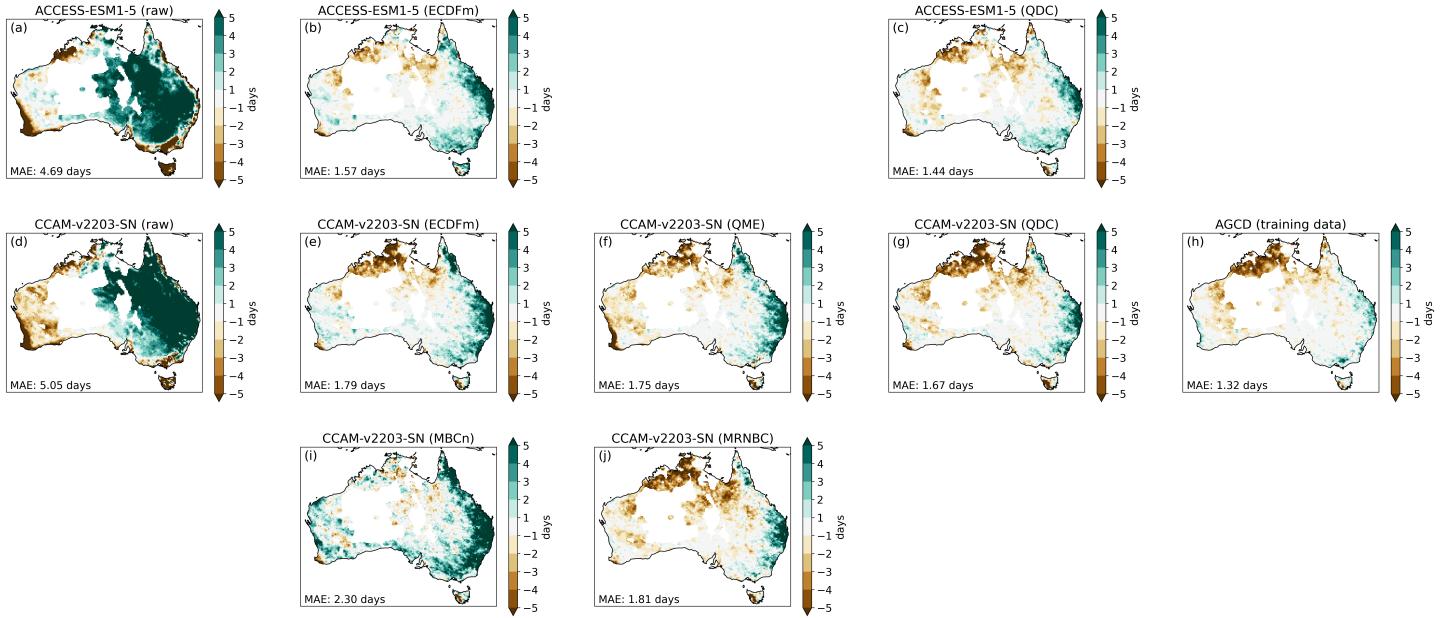


Figure S15: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

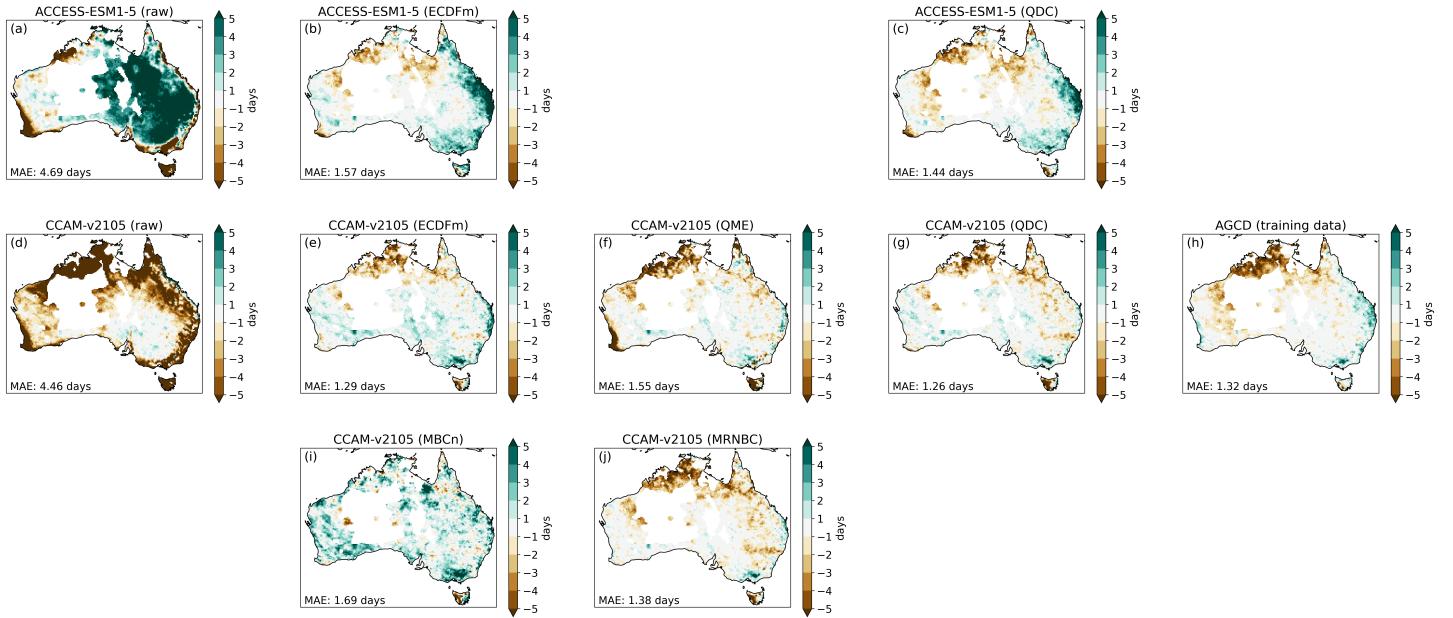


Figure S16: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

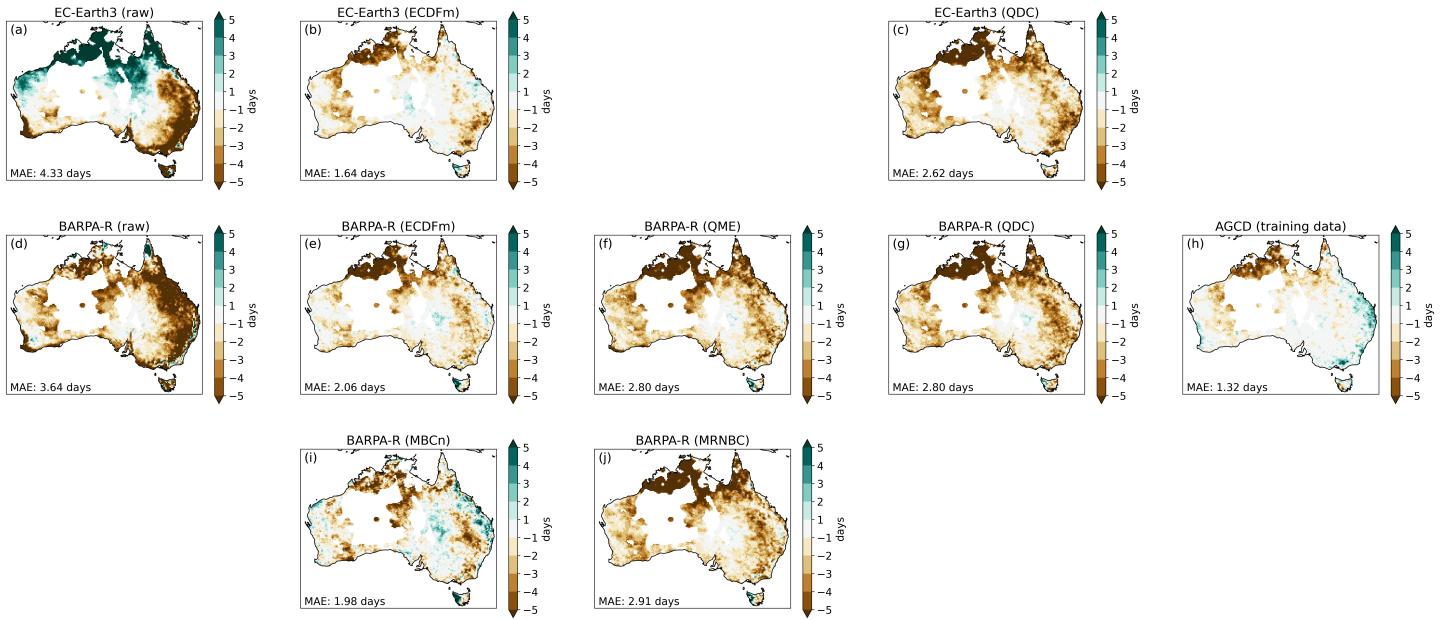


Figure S17: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the EC-Earth3 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

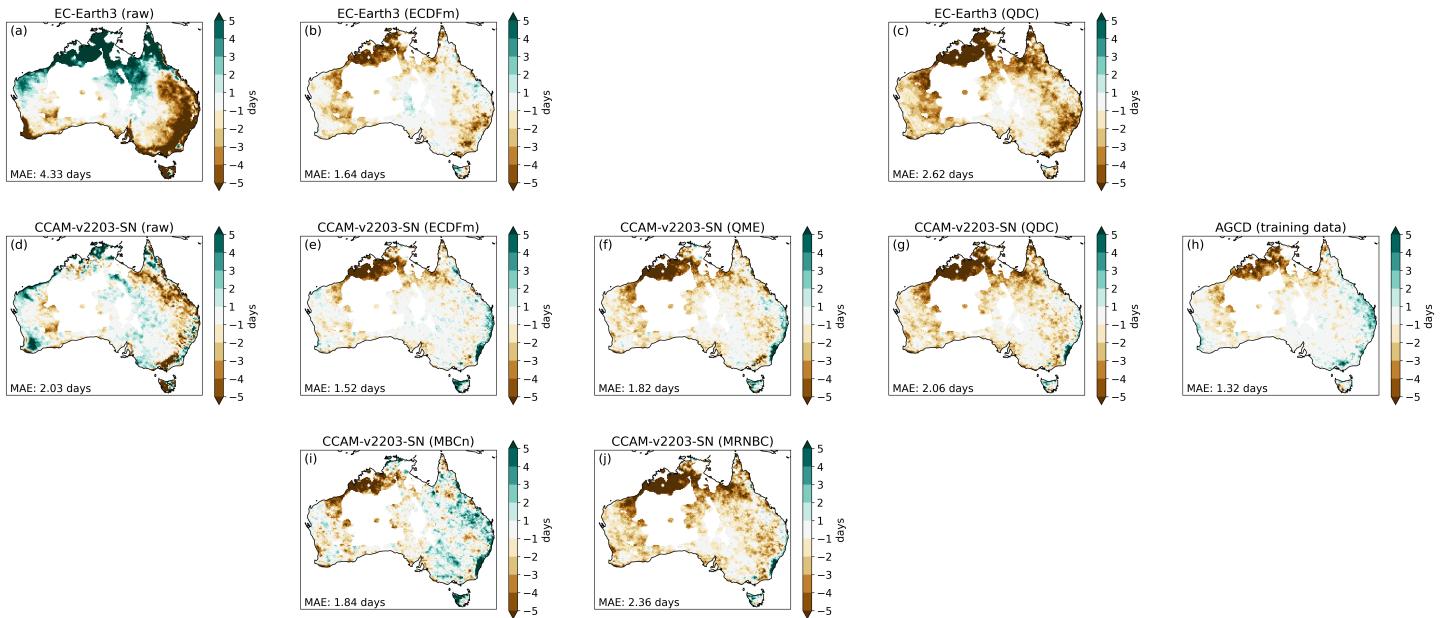


Figure S18: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the EC-Earth3 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

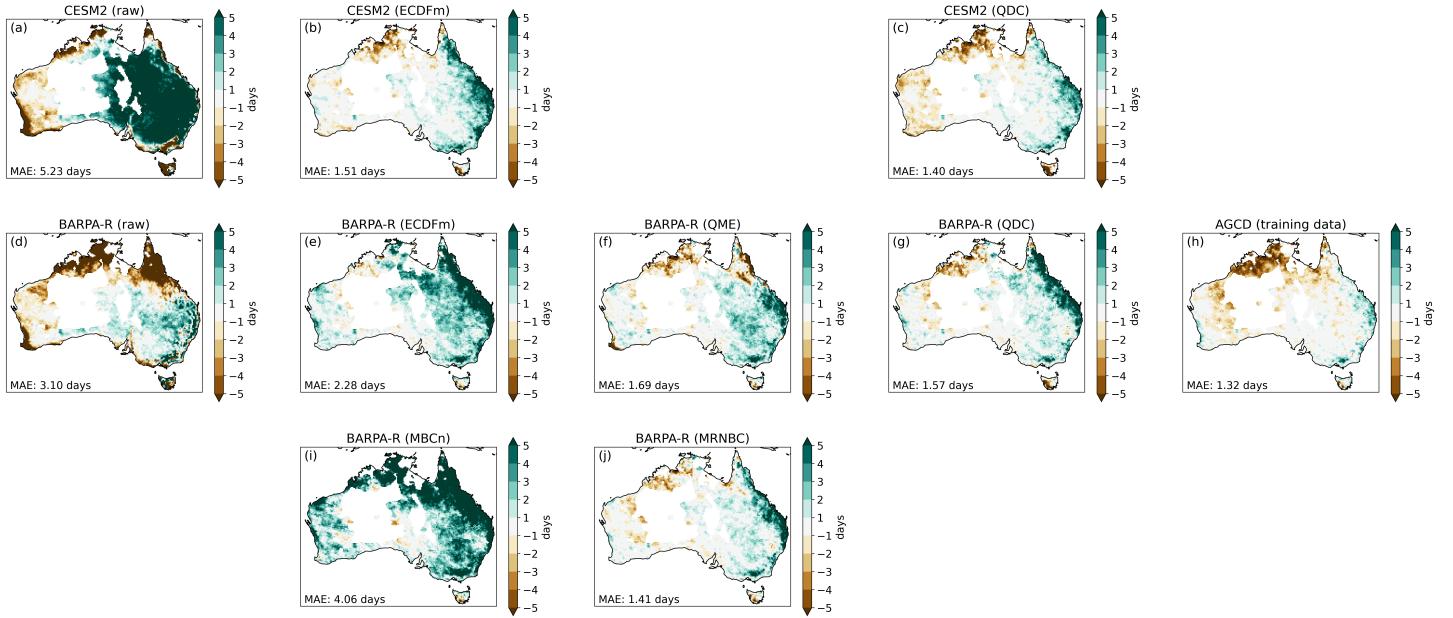


Figure S19: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the CESM2 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)

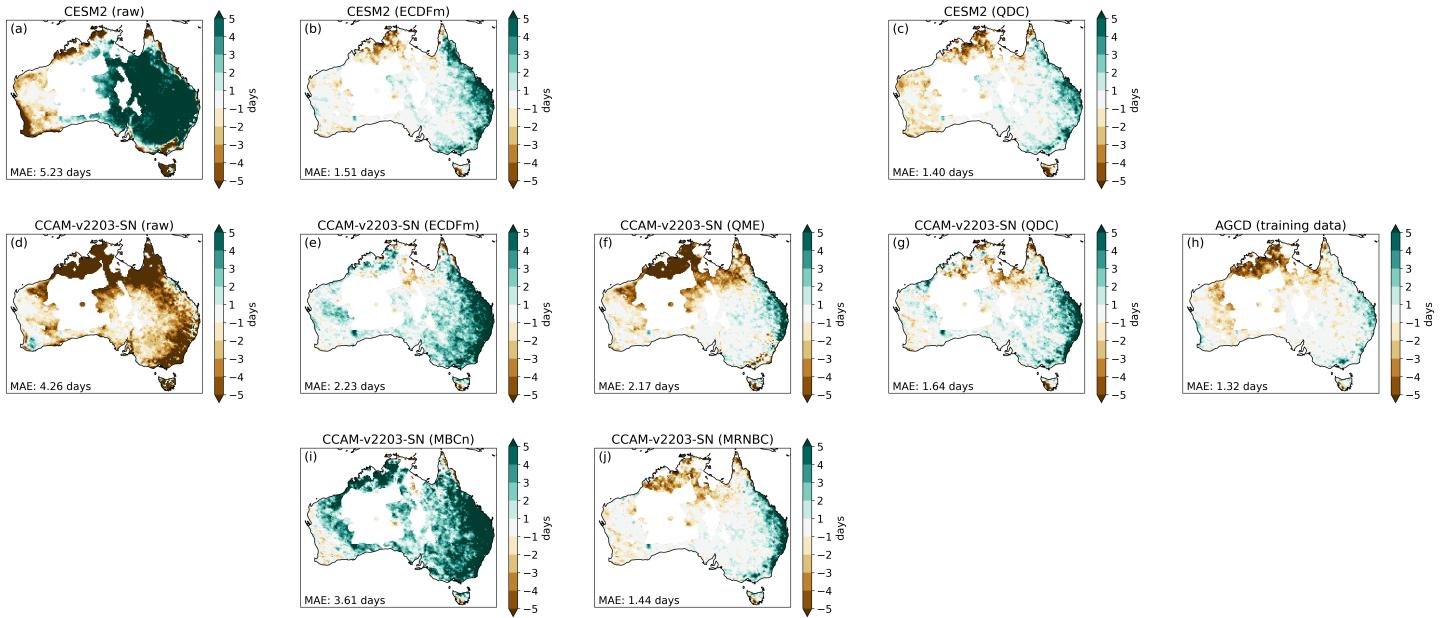


Figure S20: Bias in the annual number of heavy precipitation days (>10mm) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the CESM2 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) and RCM (panels e, f, g, i and j) data. A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel h). Land areas where the AGCD data are unreliable due to weather station sparsity have been masked in white. (MAE = mean absolute error.)