

## **Supplementary Information**

This document presents supplementary figures showing bias in the Cold Spell Duration Index (CSDI).

## Calibration task

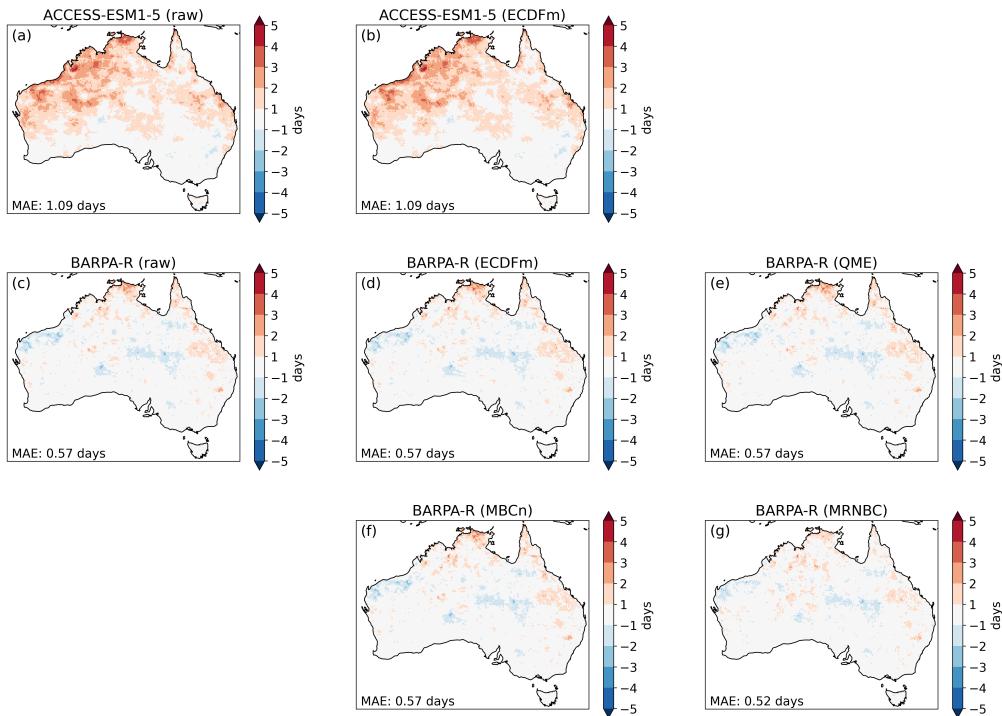


Figure S1: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. (MAE = mean absolute error.)

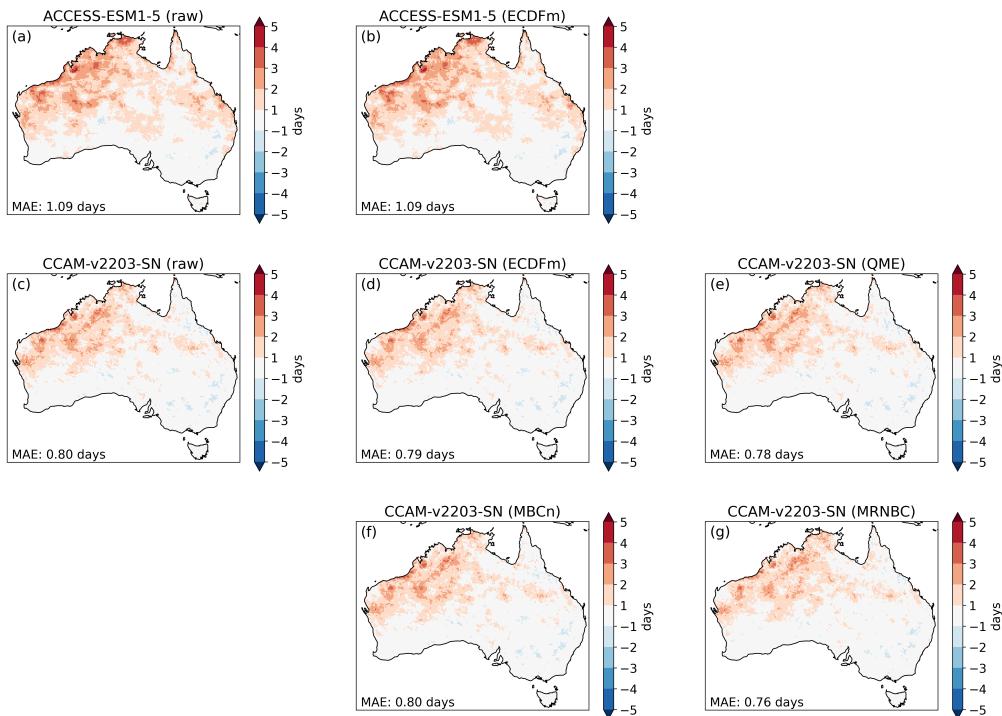


Figure S2: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. (MAE = mean absolute error.)

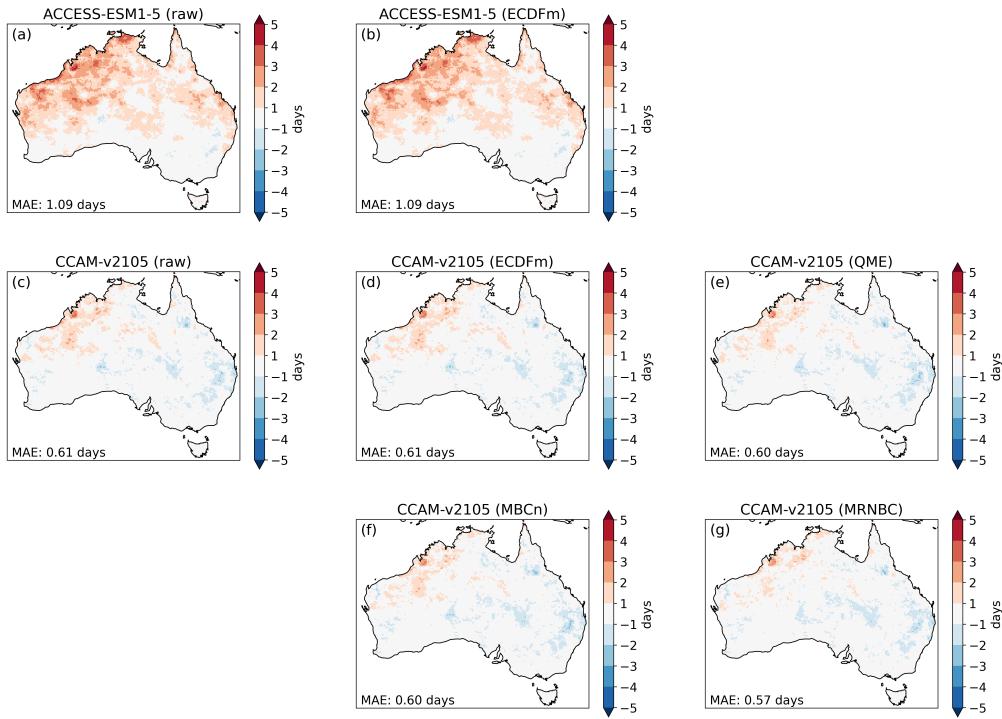


Figure S3: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. (MAE = mean absolute error.)

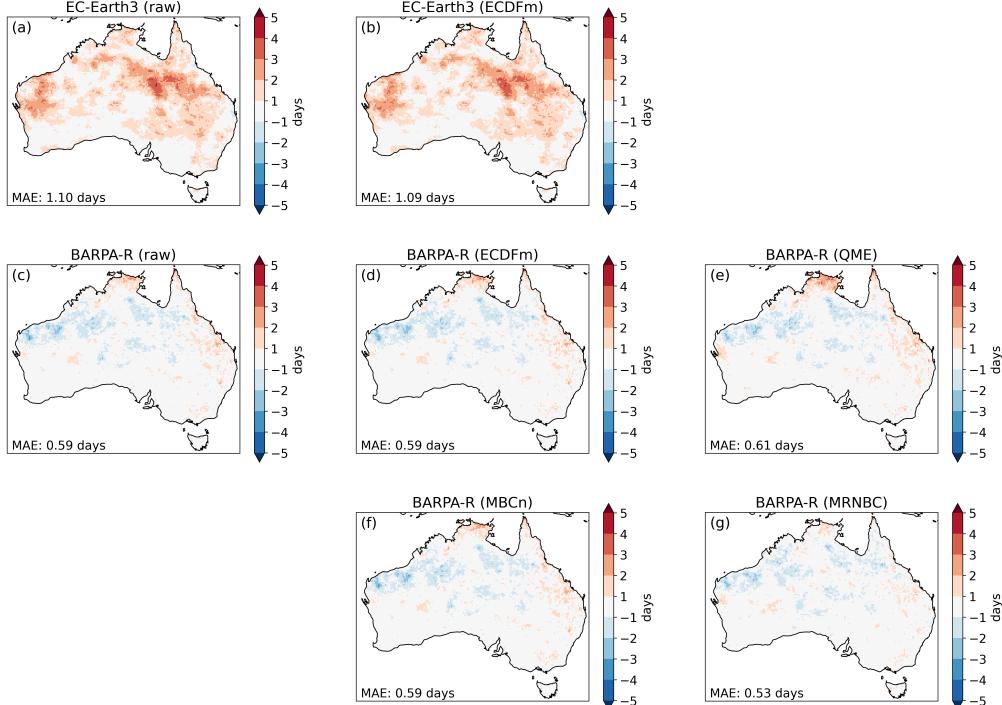


Figure S4: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the EC-Earth3 GCM (panel a), the BARPA-R RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. (MAE = mean absolute error.)

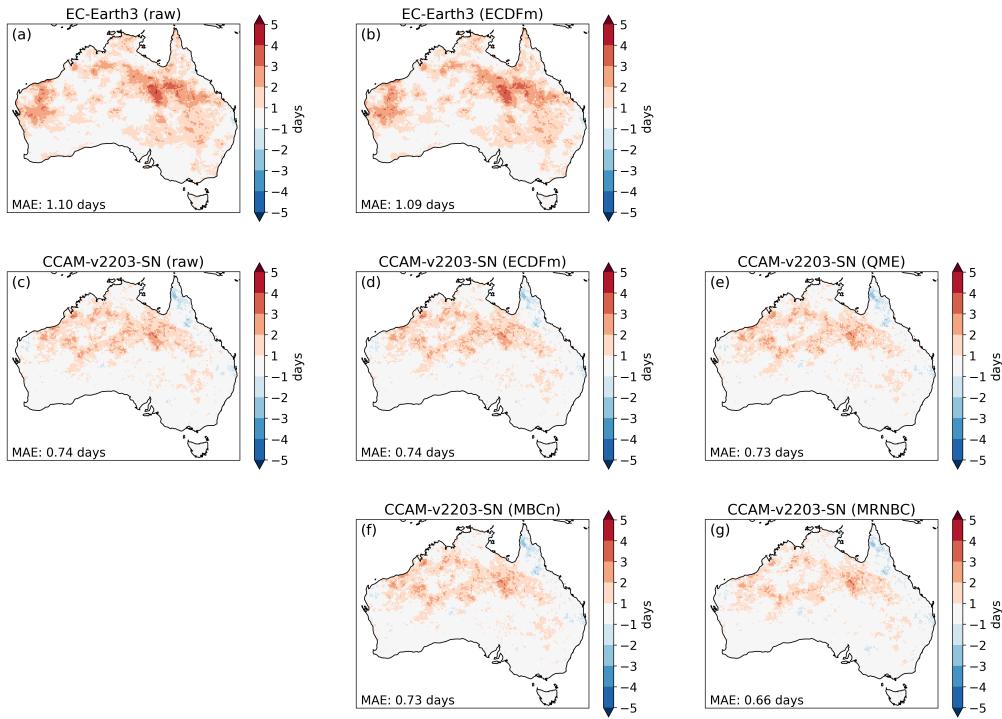


Figure S5: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the EC-Earth3 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel c), and various bias correction methods applied to those GCM (panel b) and RCM (panels d-g) data. (MAE = mean absolute error.)

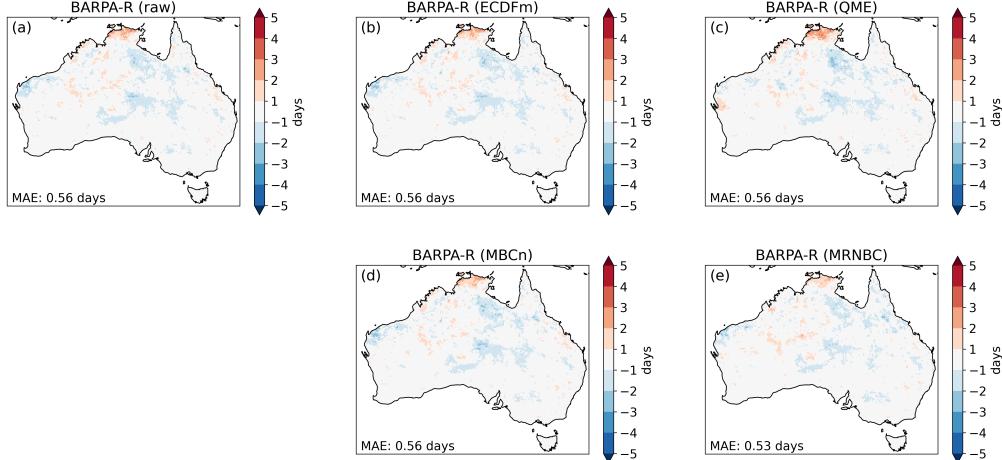


Figure S6: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the BARPA-R RCM forced by the CESM2 GCM (panel a) and various bias correction methods applied to those RCM data (panels b-e) data. Unlike the other GCMs, no raw CESM2 data were available. (MAE = mean absolute error.)

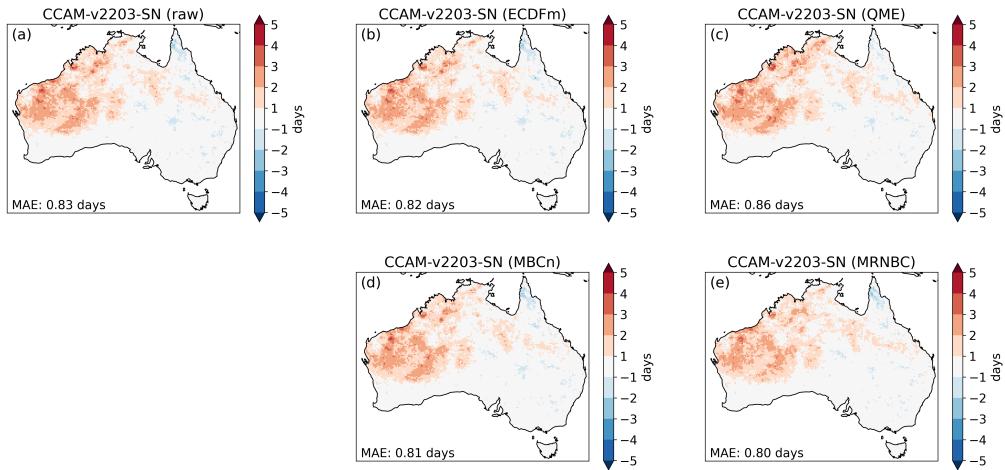


Figure S7: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the calibration assessment task. Results are shown for the CCAM-v2203-SN RCM forced by the CESM2 GCM (panel a) and various bias correction methods applied to those RCM data (panels b-e) data. Unlike the other GCMs, no raw CESM2 data were available. (MAE = mean absolute error.)

## Cross validation task

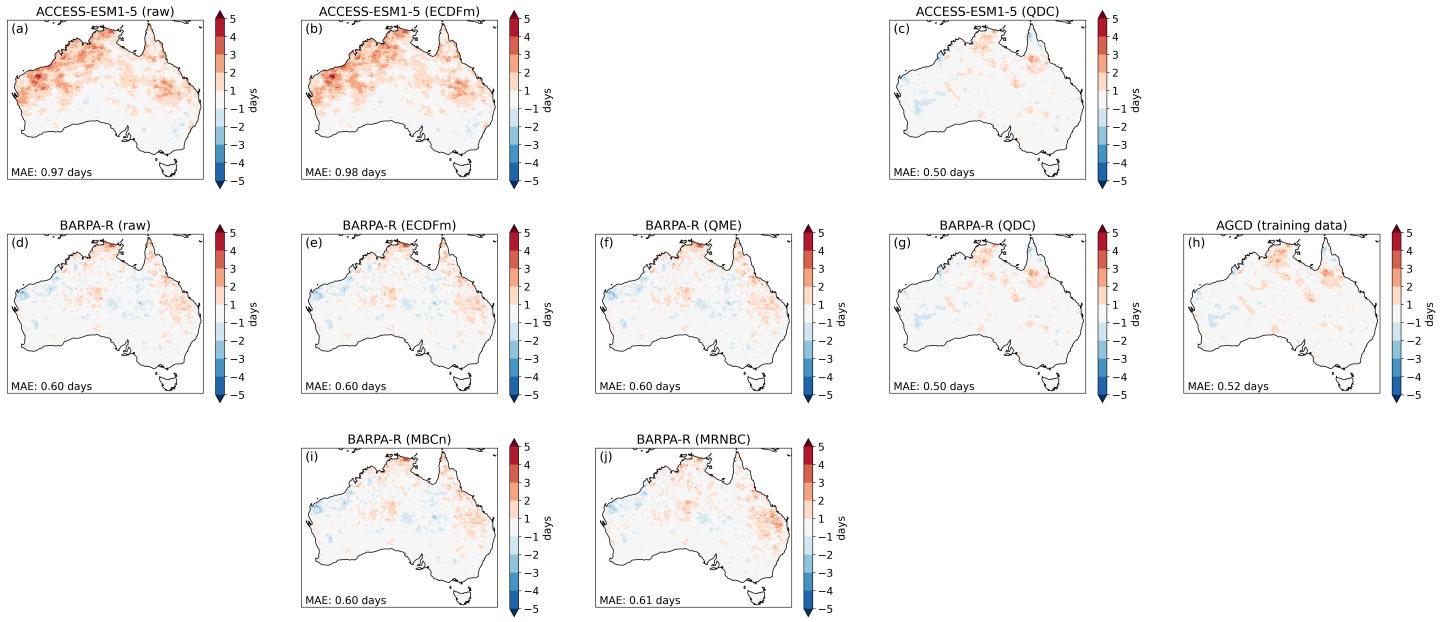


Figure S8: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and 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). (MAE = mean absolute error.)

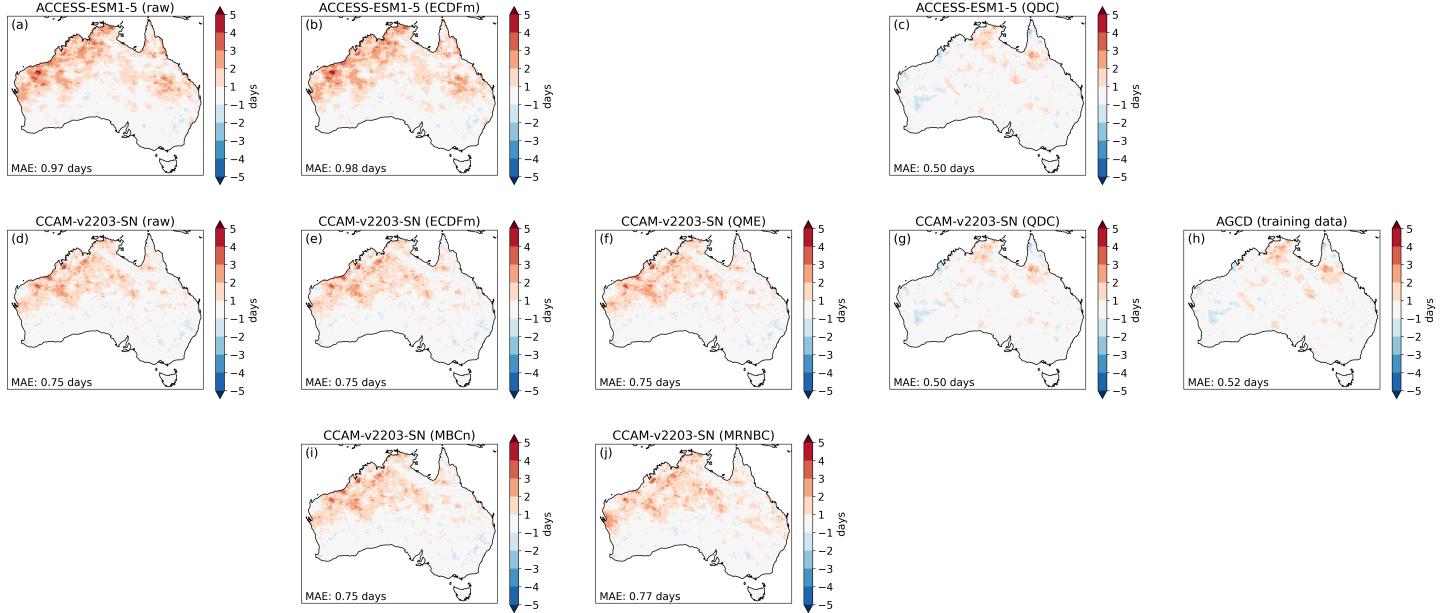


Figure S9: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels 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). (MAE = mean absolute error.)

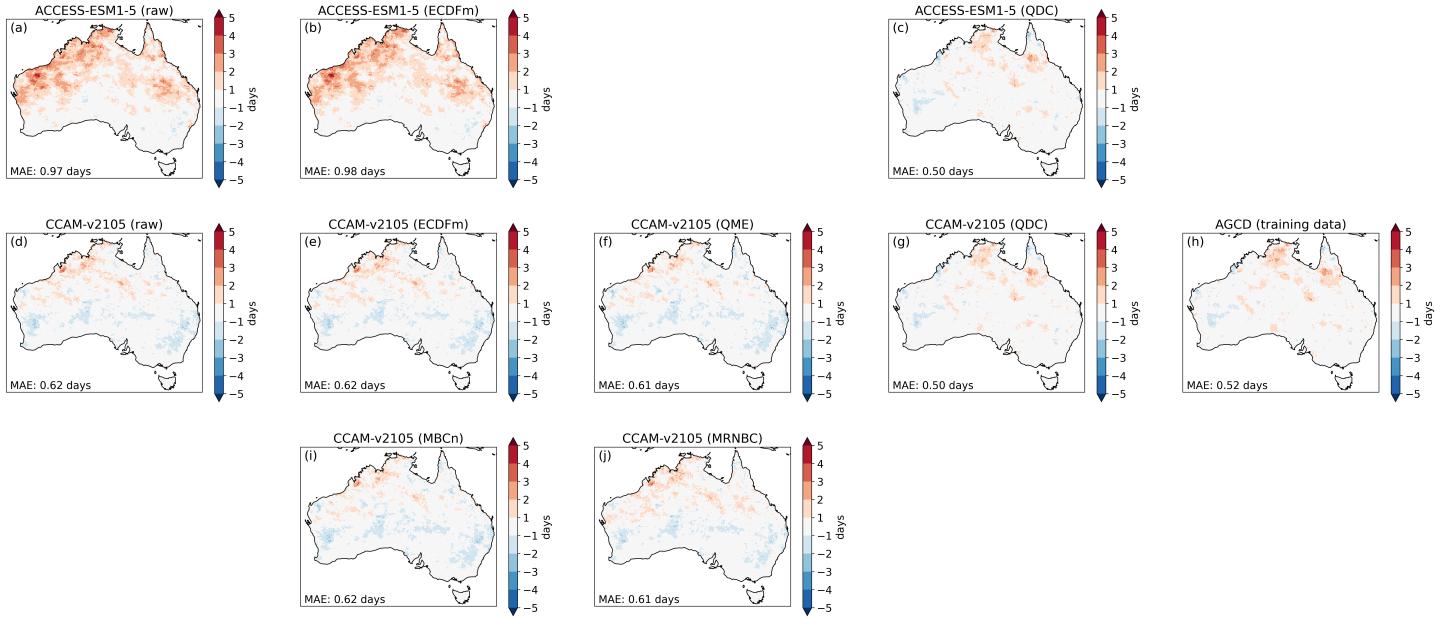


Figure S10: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the ACCESS-ESM1-5 GCM (panel a), the CCAM-v2105 RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b 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). (MAE = mean absolute error.)

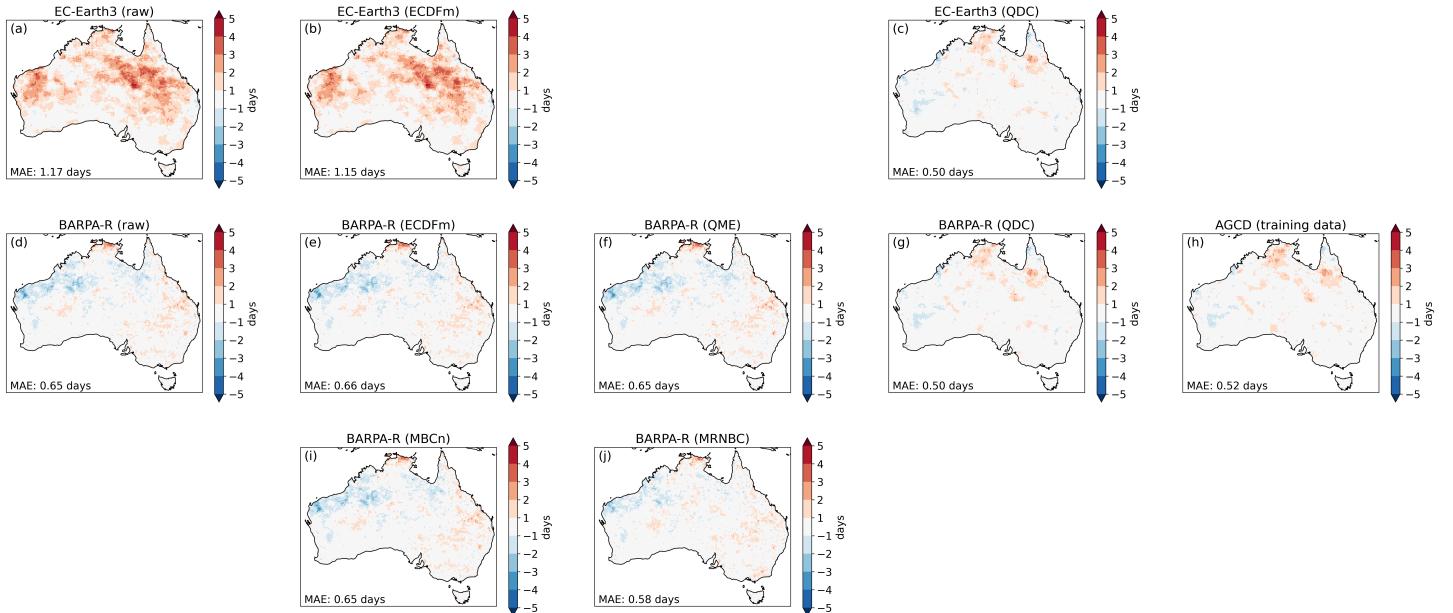


Figure S11: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the EC-Earth3 GCM (panel a), the BARPA-R RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and c) 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). (MAE = mean absolute error.)

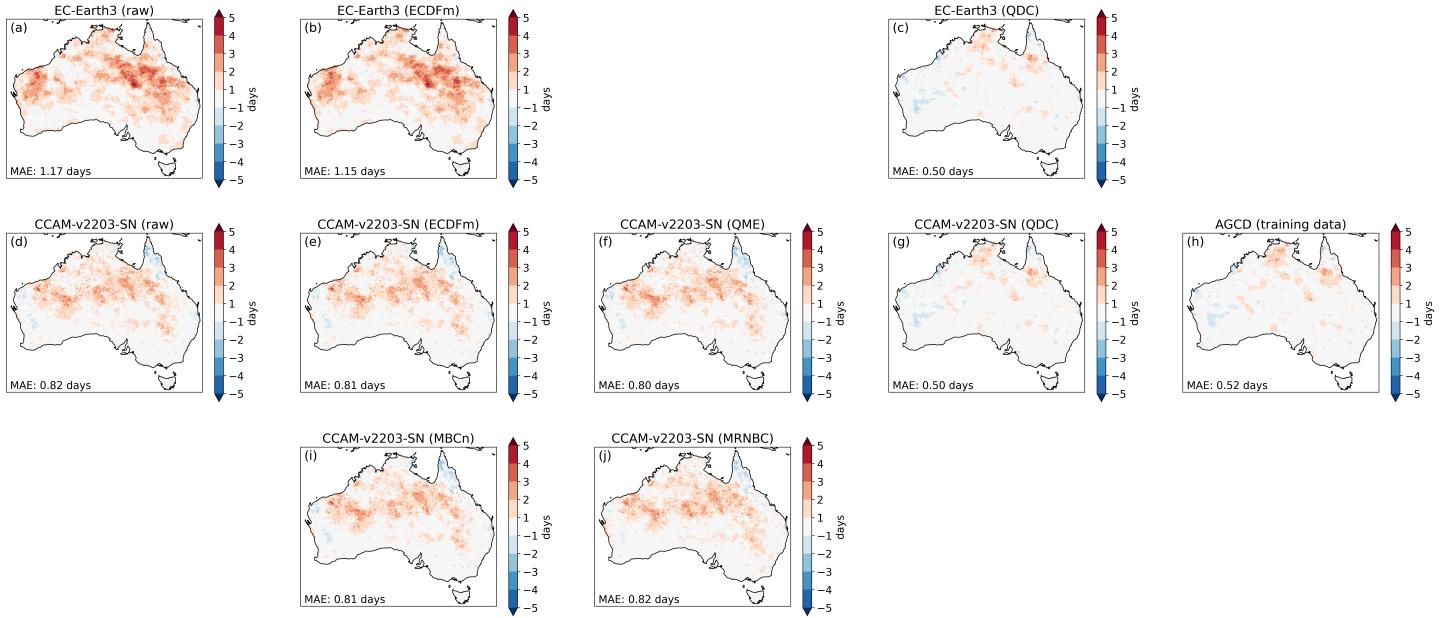


Figure S12: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the EC-Earth3 GCM (panel a), the CCAM-v2203-SN RCM forced by that GCM (panel d), and various bias correction methods applied to those GCM (panels b and 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). (MAE = mean absolute error.)

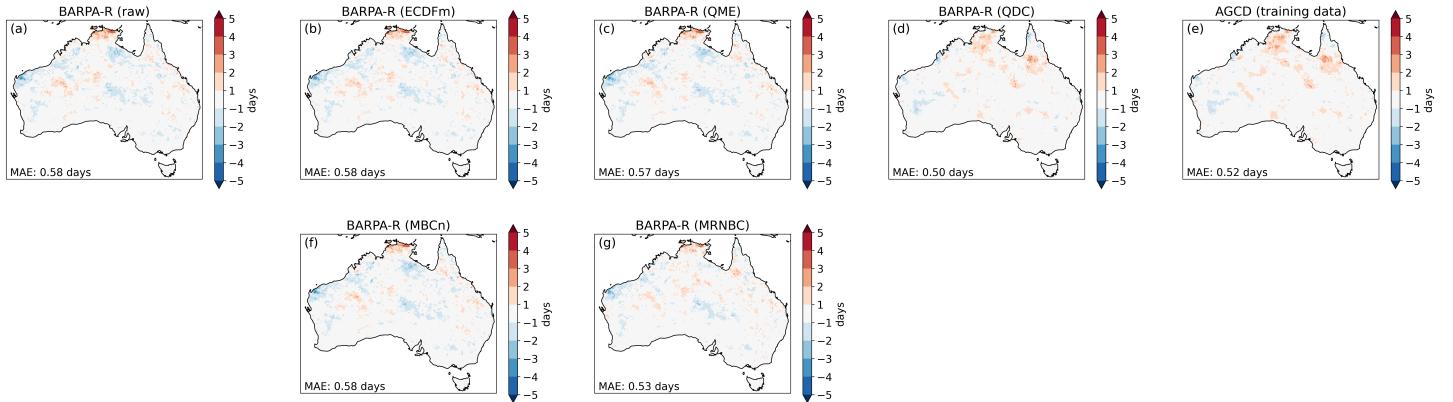


Figure S13: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the BARPA-R RCM forced by the CESM2 GCM (panel a) and various bias correction methods applied to those RCM data (panels b, c, d, f and g). A reference case where the AGCD training data (1960–1989) was simply duplicated for the assessment period (1990–2019) is also shown (panel e). Unlike the other GCMs, no raw CESM2 data were available. (MAE = mean absolute error.)

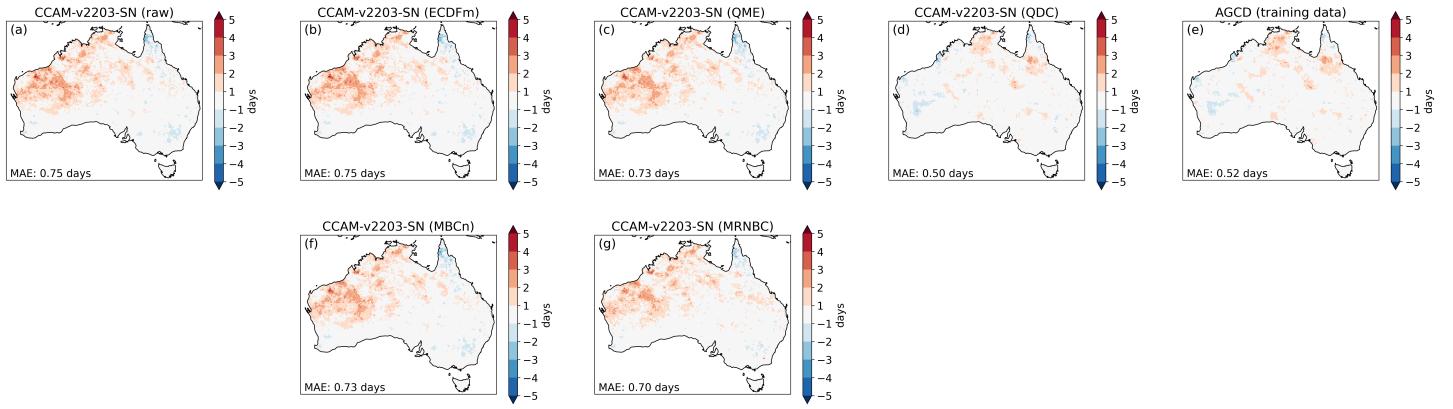


Figure S14: Bias in the Cold Spell Duration Index (CSDI) (relative to the AGCD dataset) for the cross validation assessment task. Results are shown for the CCAM-v2203-SN RCM forced by the CESM2 GCM (panel a) and various bias correction methods applied to those RCM data (panels b, c, d, f and g). A reference case where the AGCD training data (1960-1989) was simply duplicated for the assessment period (1990-2019) is also shown (panel e). Unlike the other GCMs, no raw CESM2 data were available. (MAE = mean absolute error.)