

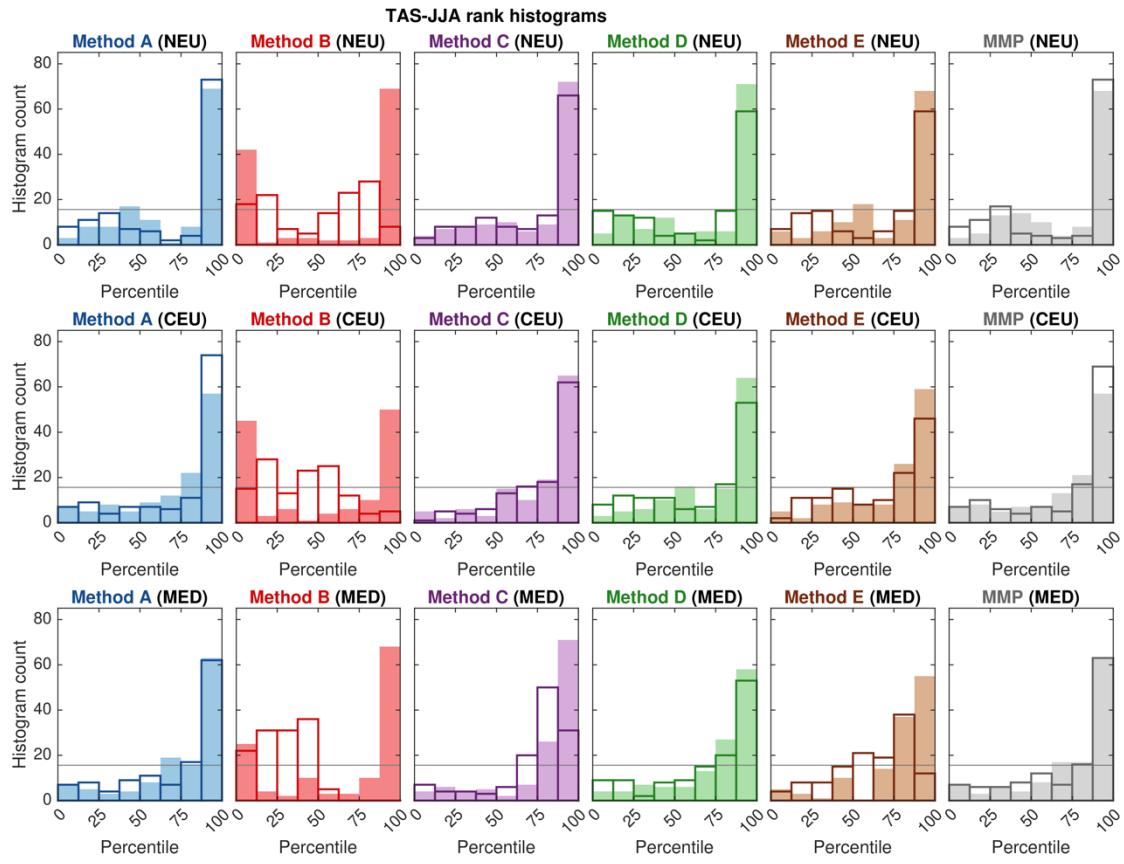
Assessing observational constraints on future European climate in an out-of-sample framework

Christopher H. O'Reilly*, Lukas Brunner, Said Qasmi, Rita Nogherotto, Andrew P. Ballinger, Ben Booth, Daniel Befort, Reto Knutti, Andrew Schurer, Aurelien Ribes, Antje Weisheimer, Erika Coppola, and Carol McSweeney.

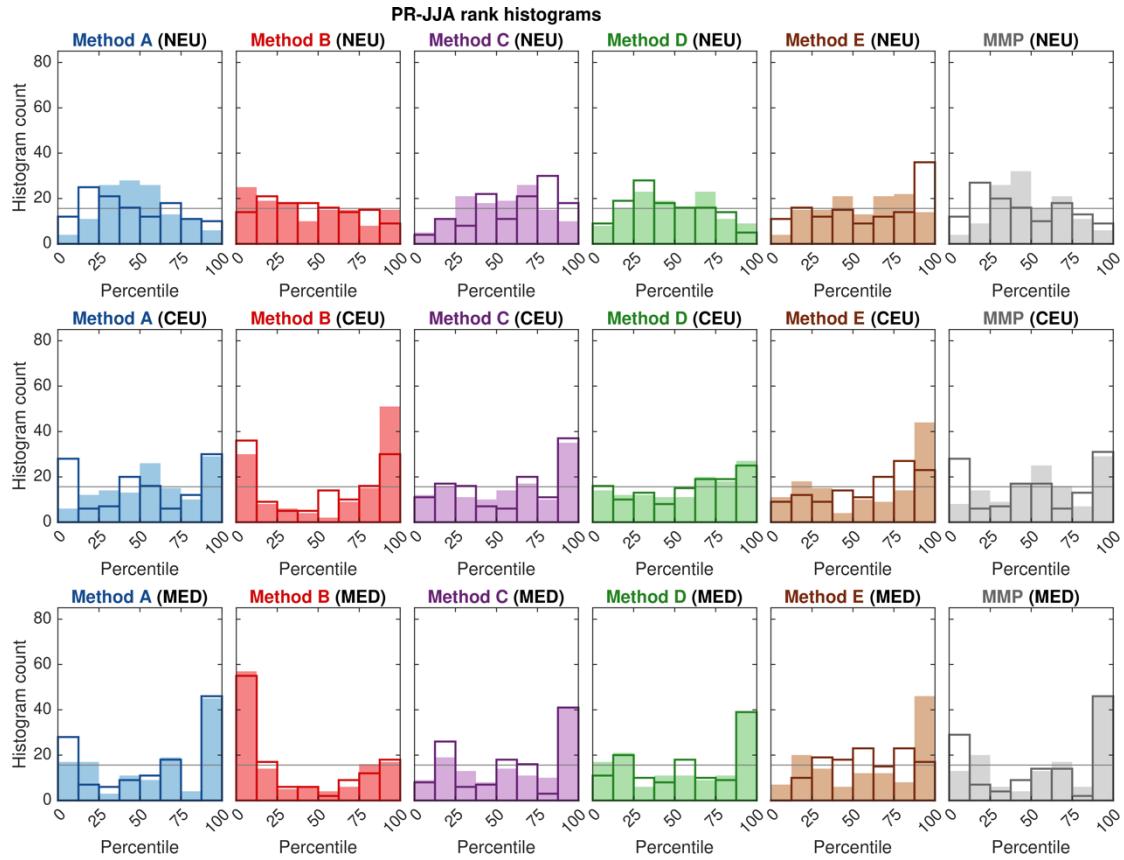
*Email: c.h.oreilly@reading.ac.uk

This PDF file includes:

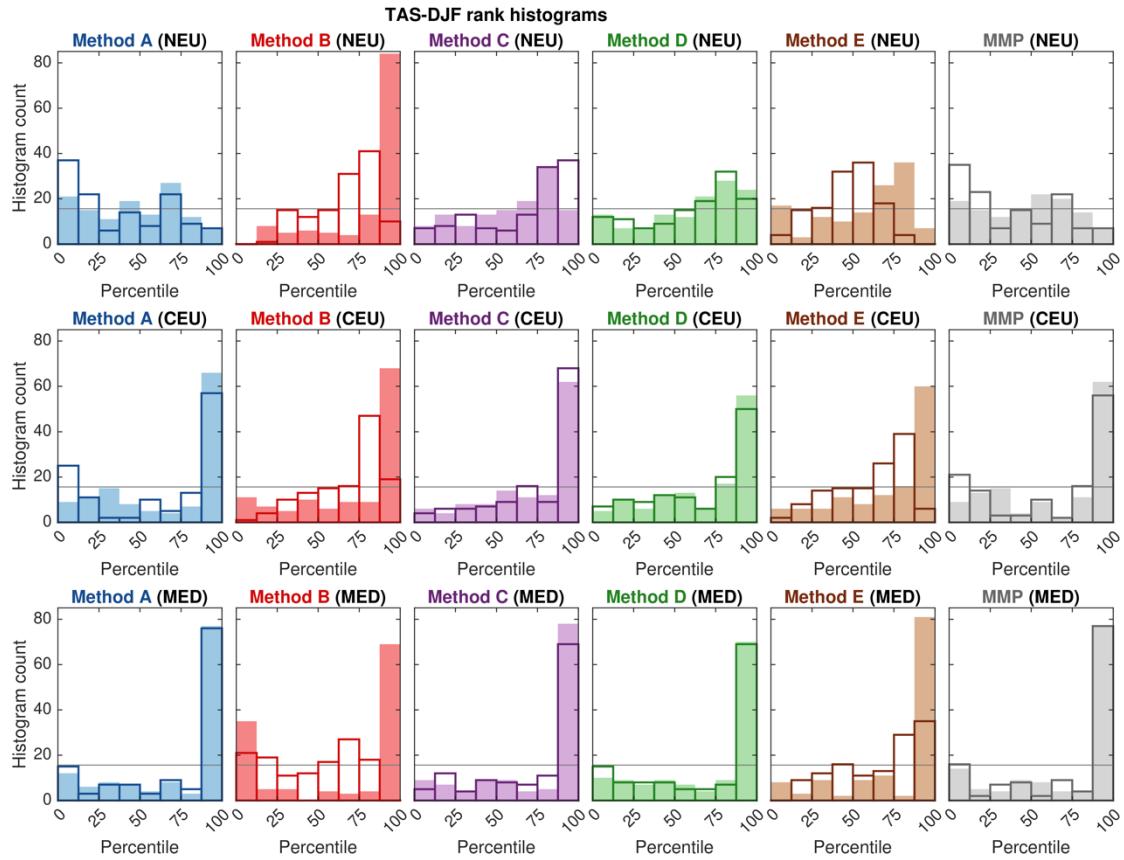
Supplementary Figures 1-25
Supplementary Tables 1 & 2



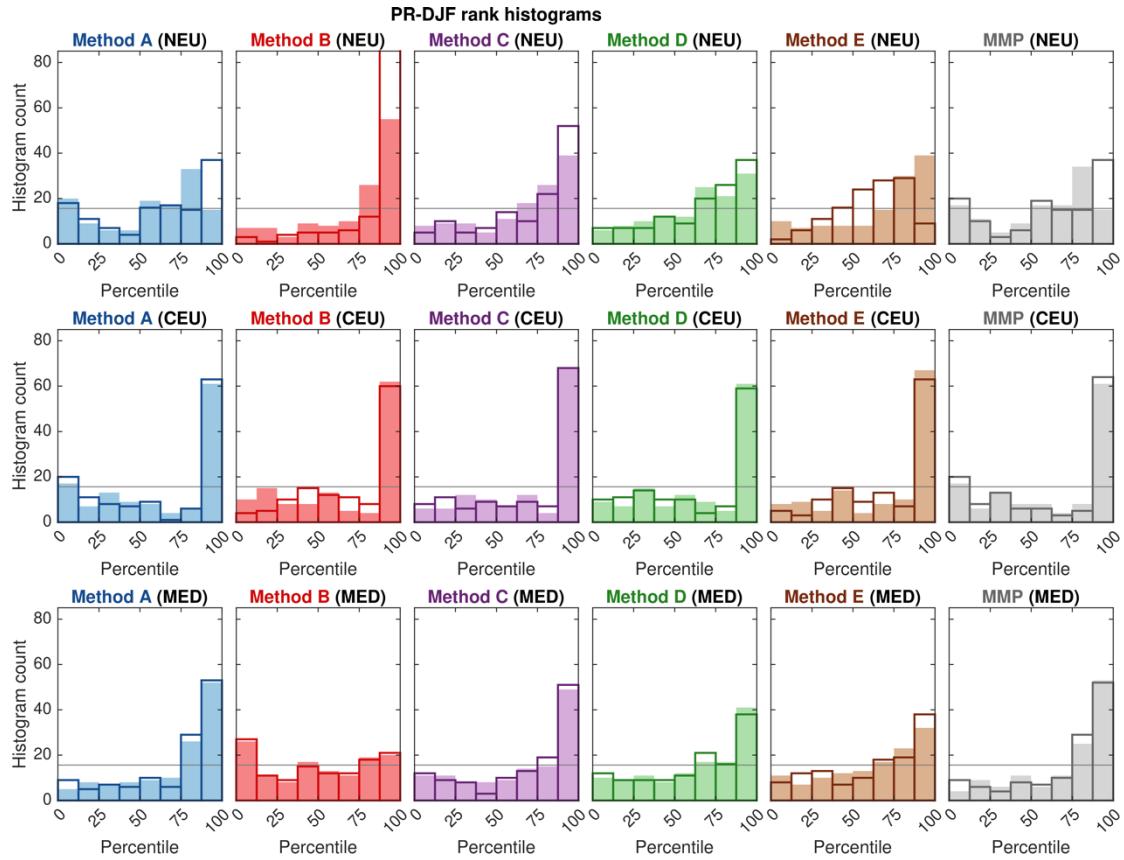
Supplementary Figure 1. Rank histograms of the future change of the summer temperature for all 125 of the pseudo-observational datasets. The shaded bars show the counts for the unconstrained projection and the outlined white bars show the counts for the constrained projection. The horizontal line shows the count for each bin if the counts were equally distributed.



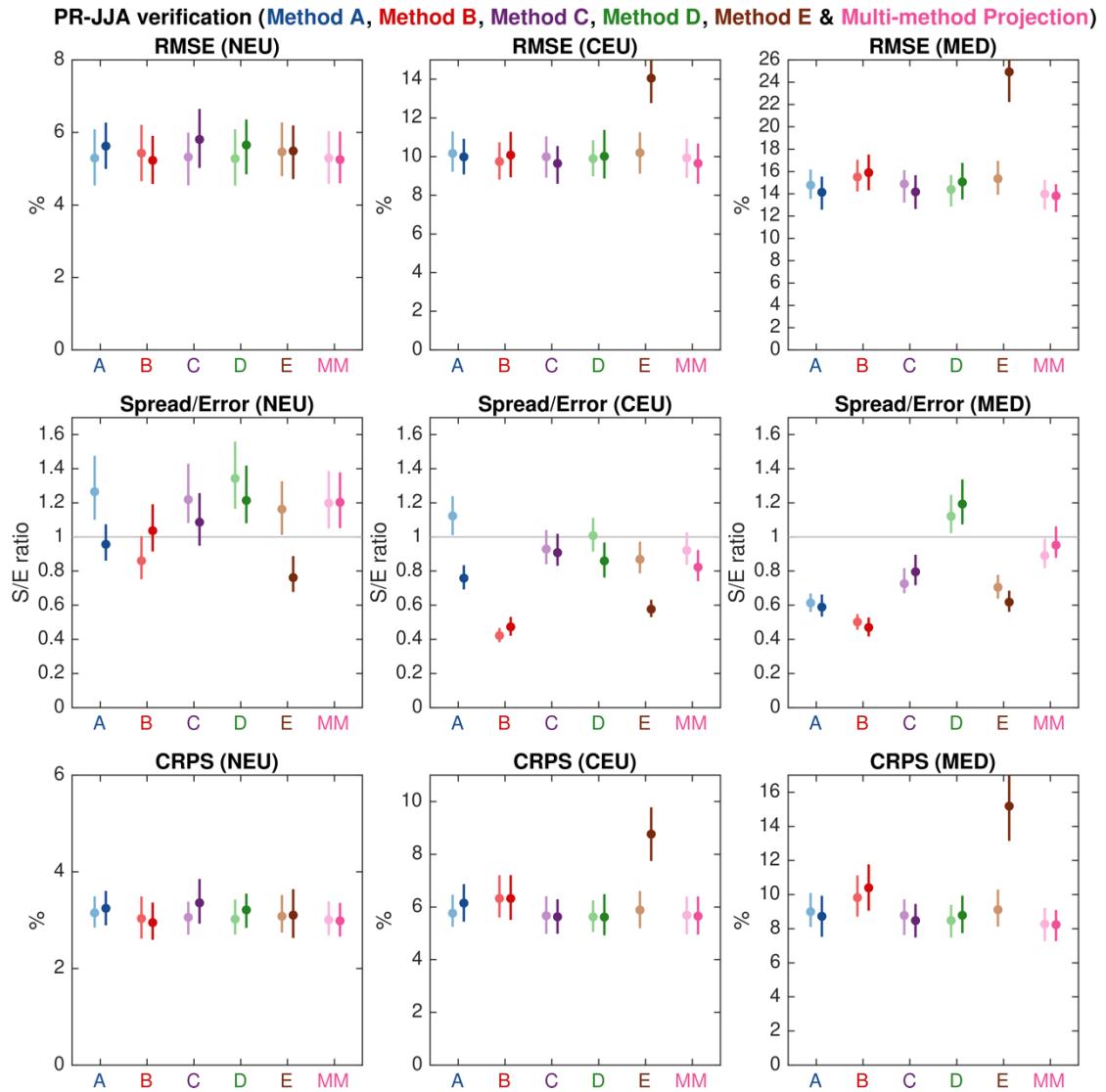
Supplementary Figure 2. Rank histograms of the future change of the summer precipitation for all 125 of the pseudo-observational datasets. The shaded bars show the counts for the unconstrained projection and the outlined white bars show the counts for the constrained projection. The horizontal line shows the count for each bin if the counts were equally distributed.



Supplementary Figure 3. Rank histograms of the future change of the winter temperature for all 125 of the pseudo-observational datasets. The shaded bars show the counts for the unconstrained projection and the outlined white bars show the counts for the constrained projection. The horizontal line shows the count for each bin if the counts were equally distributed.

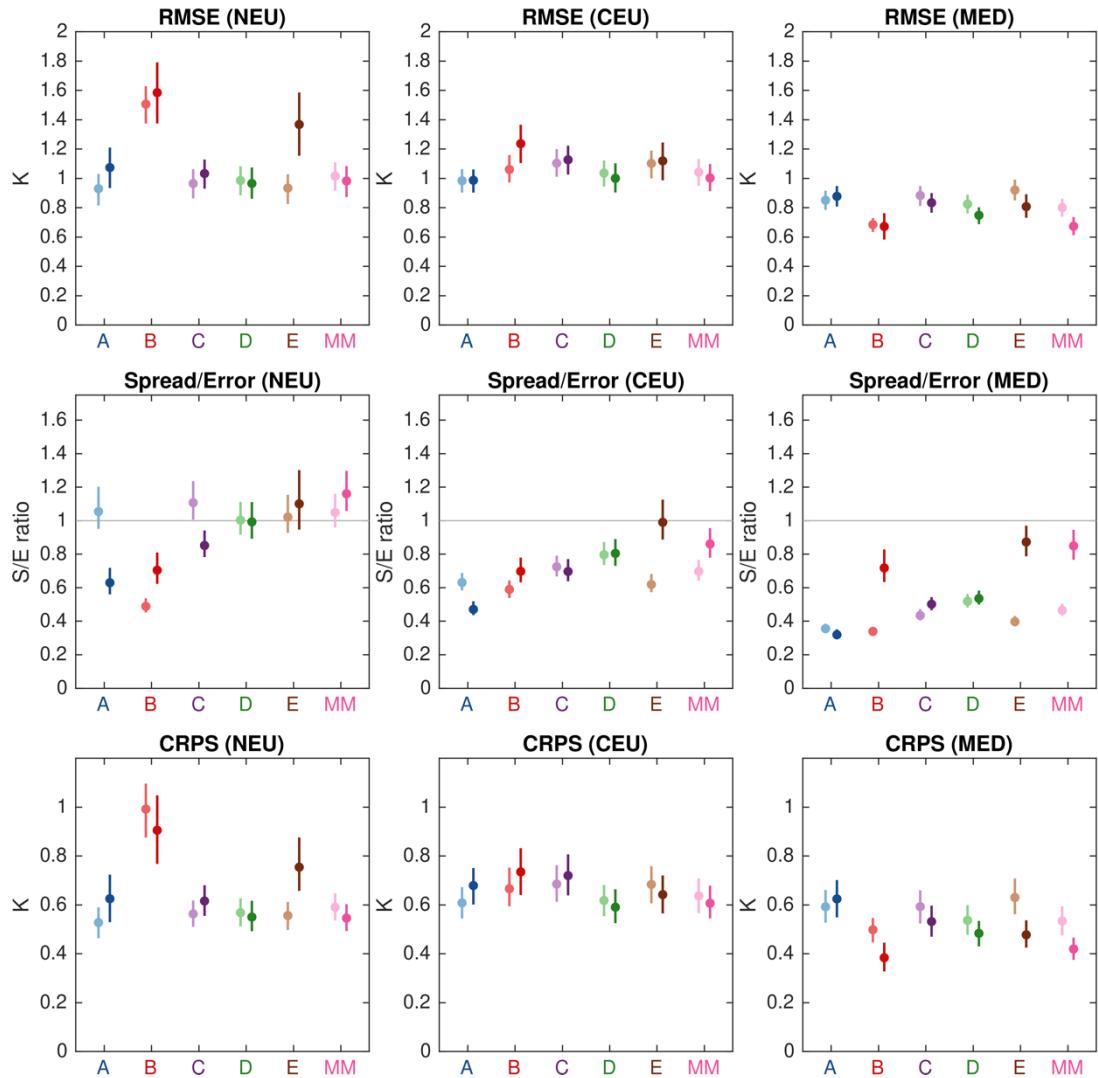


Supplementary Figure 4. Rank histograms of the future change of the winter precipitation for all 125 of the pseudo-observational datasets. The shaded bars show the counts for the unconstrained projection and the outlined white bars show the counts for the constrained projection. The horizontal line shows the count for each bin if the counts were equally distributed.



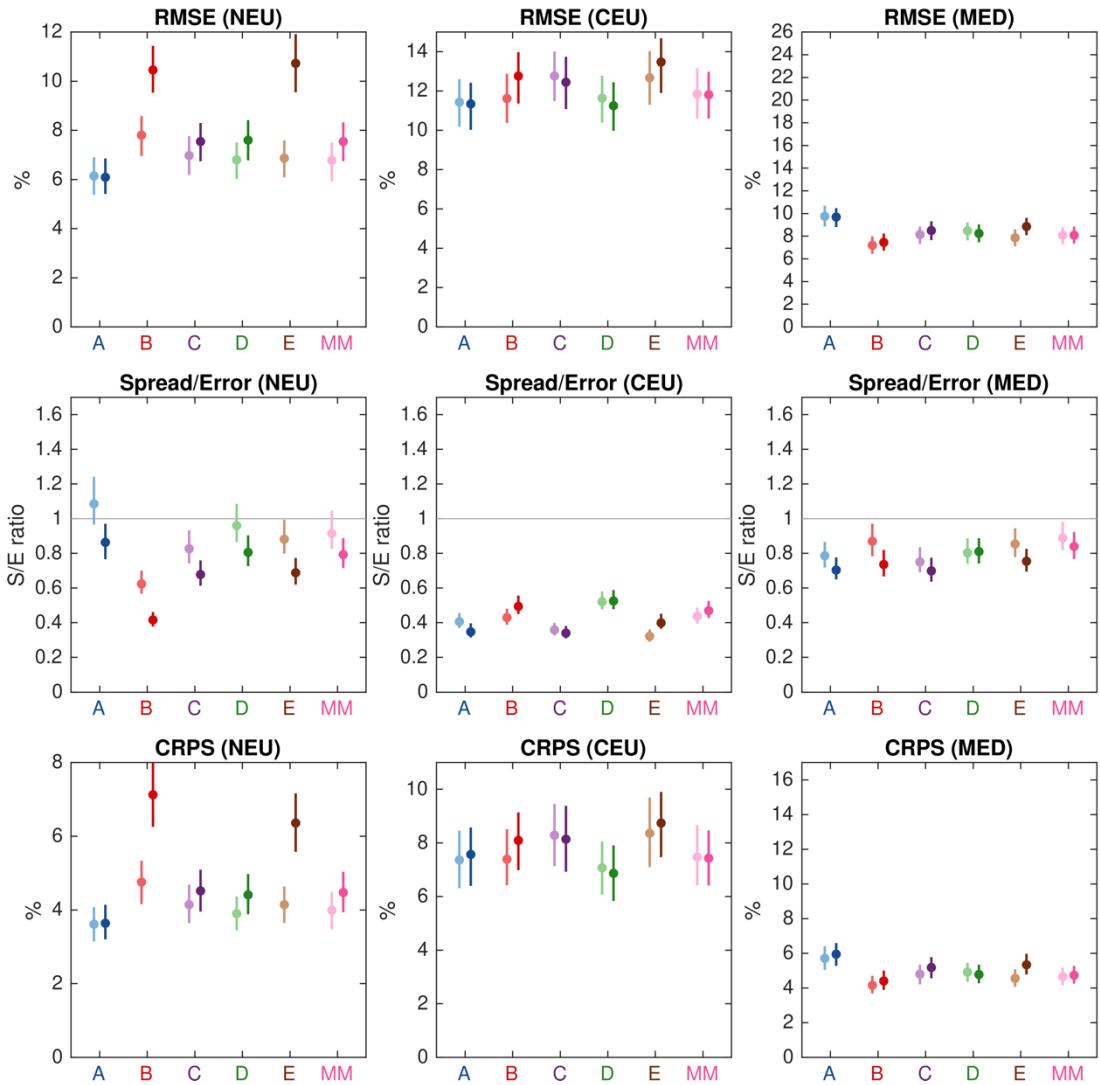
Supplementary Figure 5. Verification of the unconstrained and constrained projections across all 125 pseudo-observations for the 2041-2060 projected summer (JJA) precipitation changes in each of the European SREX regions. (a-c) Root-mean square error (RMSE); (d-e) Spread/Error; (g-i) Continuous ranked probability score (CRPS). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

TAS-DJF verification (Method A, Method B, Method C, Method D, Method E & Multi-method Projection)

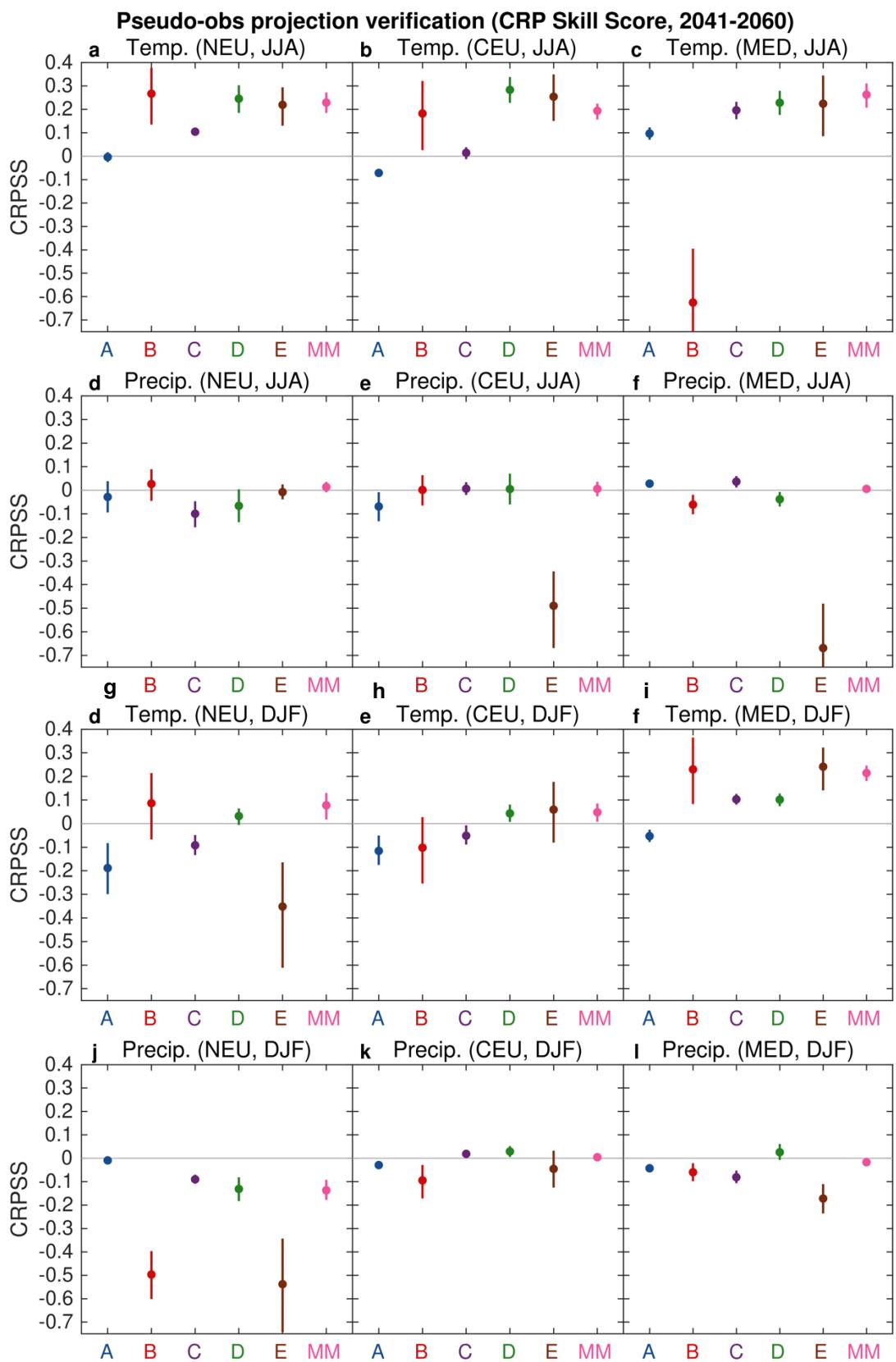


Supplementary Figure 6. Verification of the unconstrained and constrained projections across all 125 pseudo-observations for the 2041-2060 projected winter (DJF) temperature changes in each of the European SREX regions. (a-c) Root-mean square error (RMSE); (d-e) Spread/Error; (g-i) Continuous ranked probability score (CRPS). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

PR-DJF verification (Method A, Method B, Method C, Method D, Method E & Multi-method Projection)

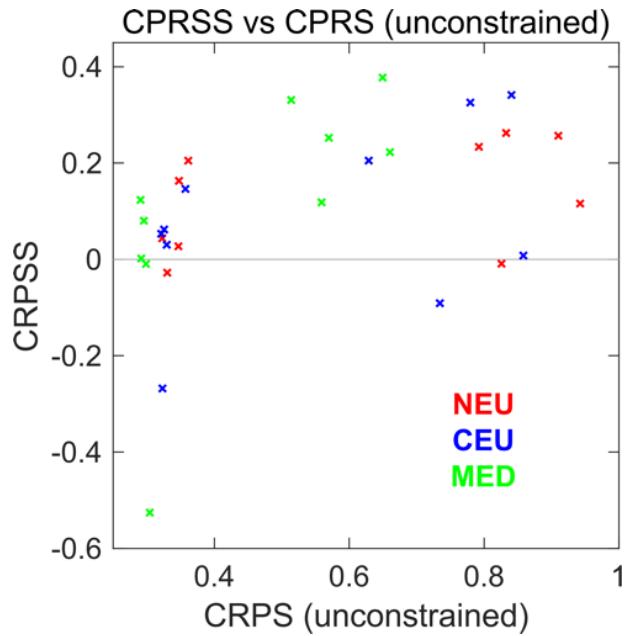


Supplementary Figure 7. Verification of the unconstrained and constrained projections across all 125 pseudo-observations for the 2041-2060 projected winter (DJF) precipitation changes in each of the European SREX regions. (a-c) Root-mean square error (RMSE); (d-e) Spread/Error; (g-i) Continuous ranked probability score (CRPS). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

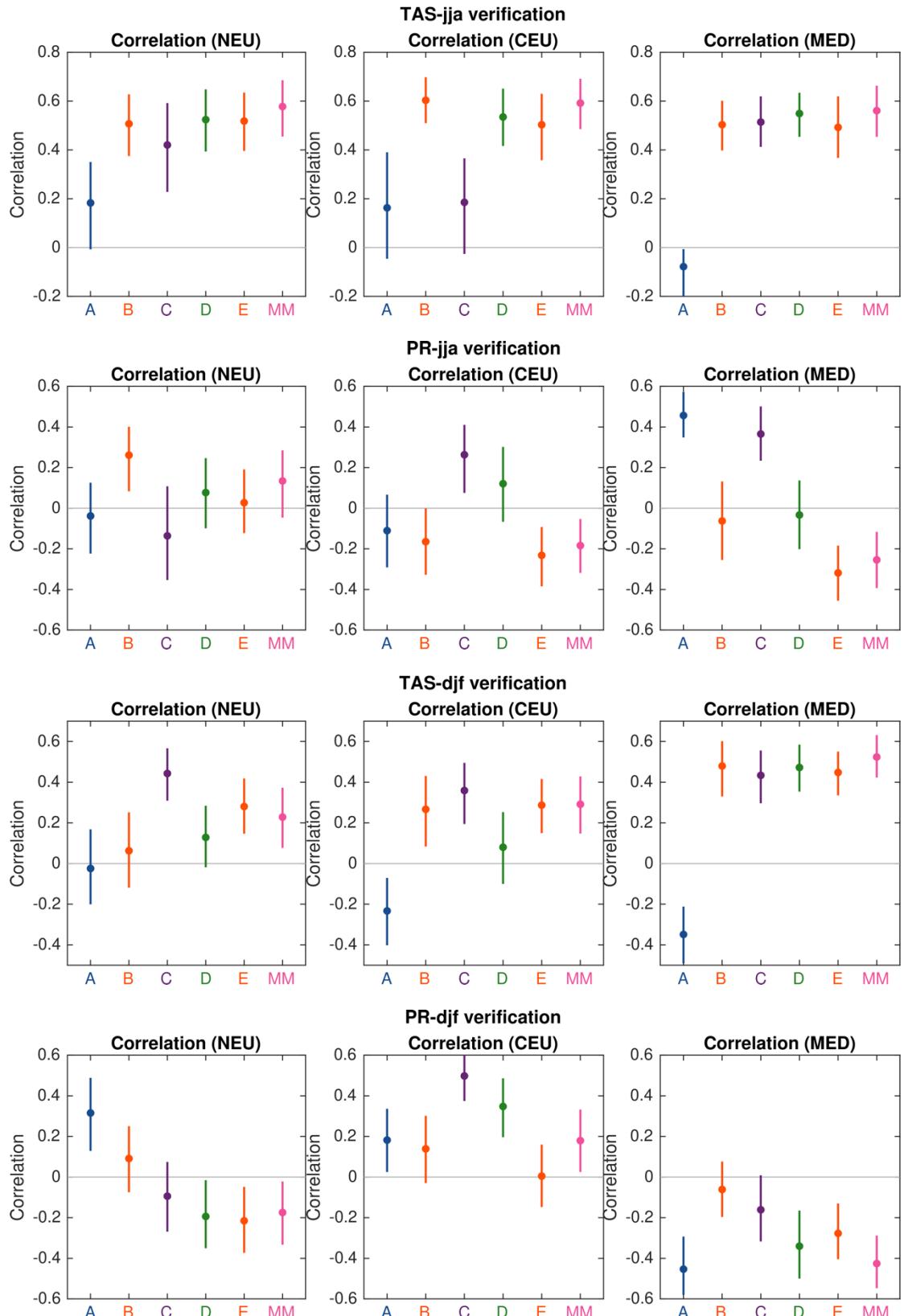


Supplementary Figure 8. Verification of the unconstrained and constrained projections across all 125 pseudo-observations, in terms of continuous ranked probability skill score (CRPSS; see Methods), for the 2041-2060 projected changes in each of the European SREX regions. (a-c) Summer temperature changes; (d-f) summer precipitation changes; (g-i) winter

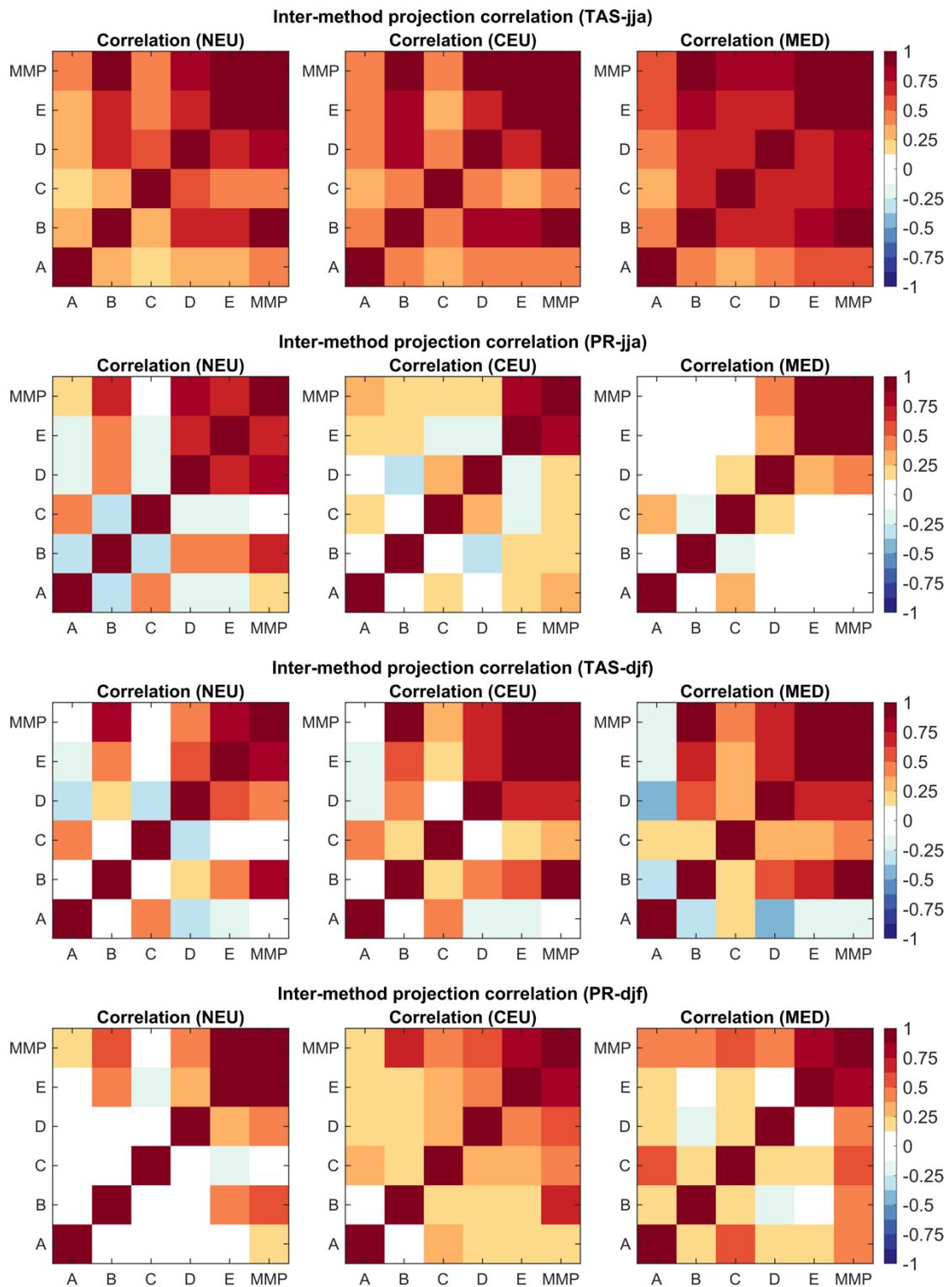
temperature changes; (j-l) winter precipitation changes. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



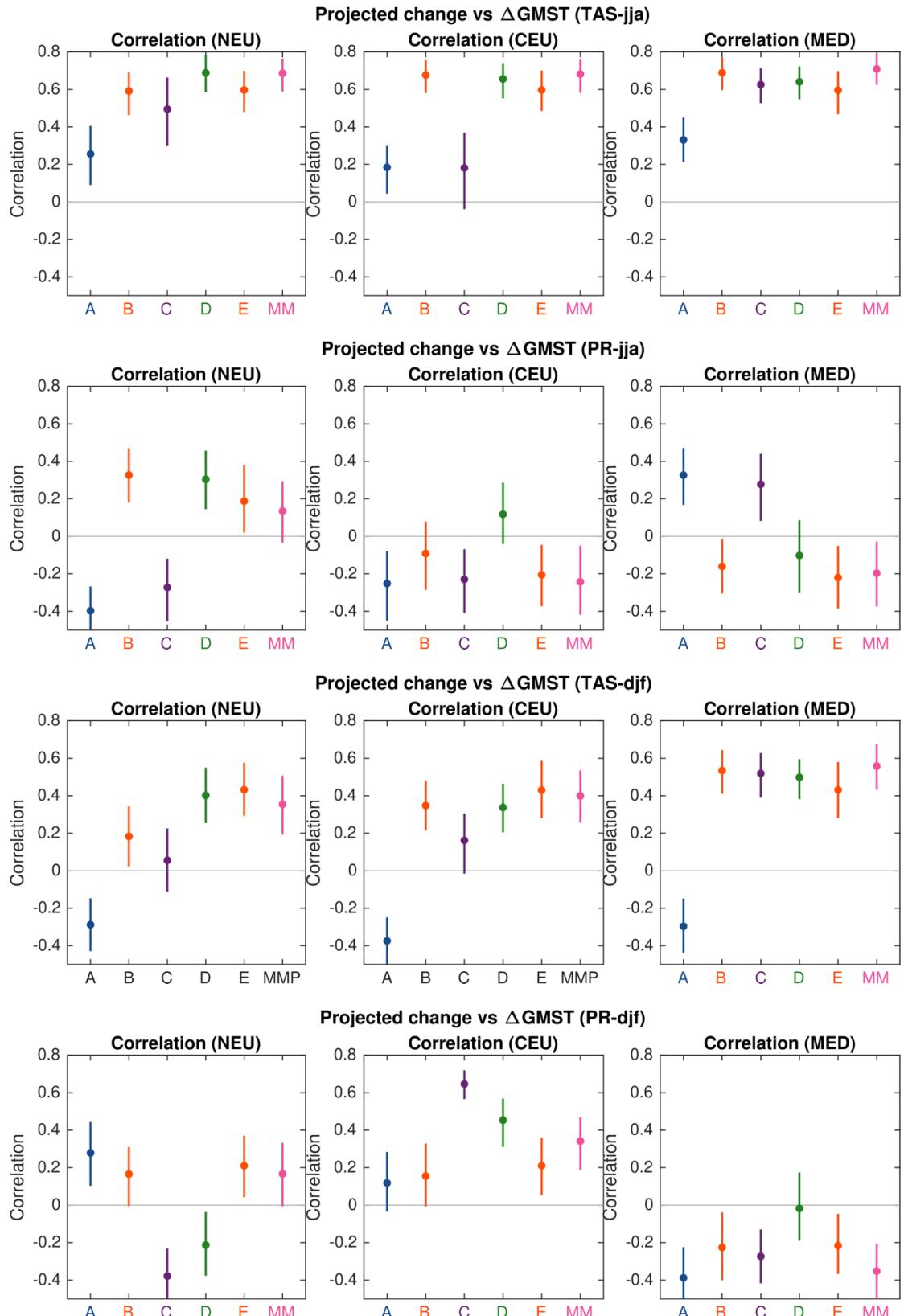
Supplementary Figure 9. Relationship between CRPSS and the CRPS from the equivalent unconstrained ensemble for summer temperature projections for each region and for each method. Here we have plotted the CRPSS and CRPS from the high and low climate sensitivity subsets (see Methods and Figures S11-S14). For all regions there is a positive correlation between the CRPSS and unconstrained CRPS, indicating that higher improvements in projection accuracy (i.e. CRPSS) are associated with the largest errors in the underlying unconstrained projection.



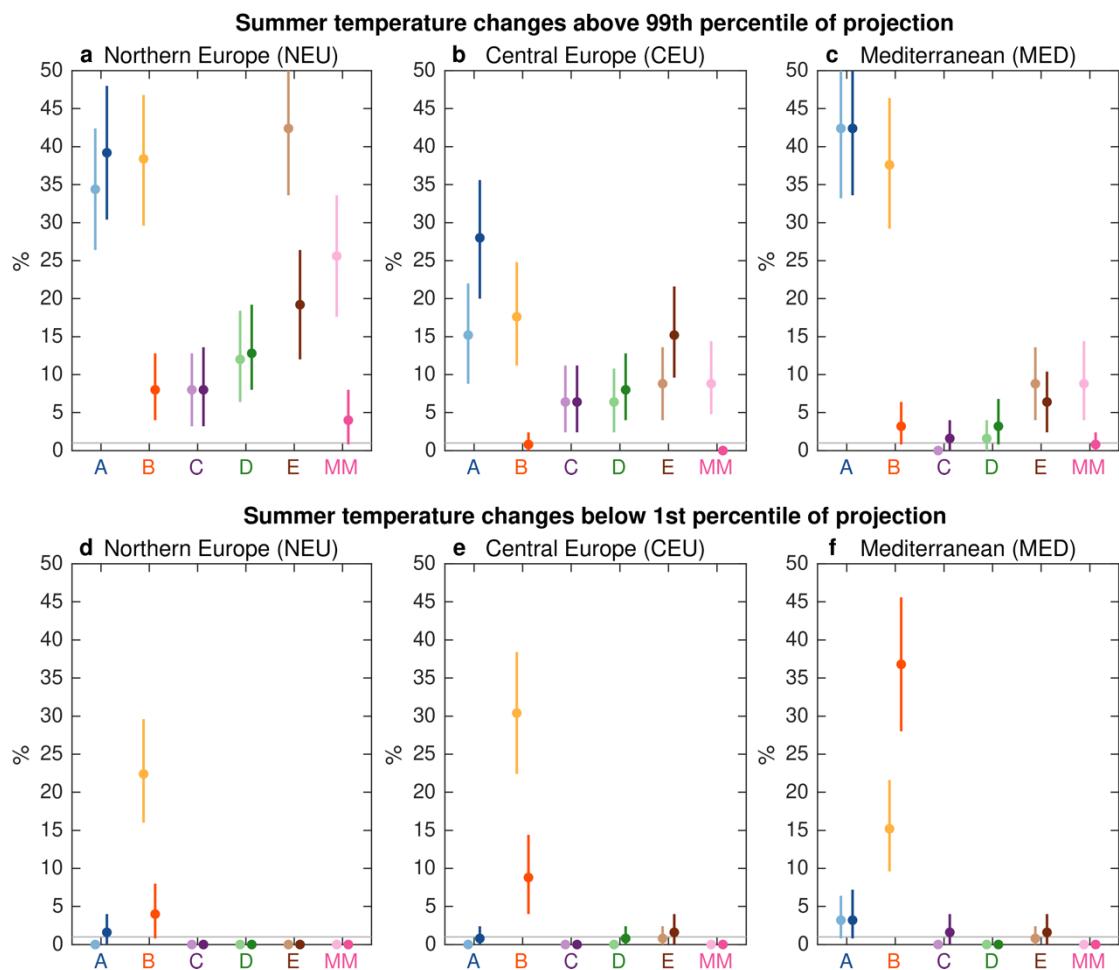
Supplementary Figure 10. Correlation of the different constrained ensemble mean projected change with actual change across all 125 pseudo-observations for the 2041-2060 period in each of the European SREX regions. Shown for temperature and precipitation for both the summer (JJA) and winter (DJF) seasons. The dots show the measured values and the lines indicate the 5-95% confidence limit based on a bootstrap resampling (see Methods).



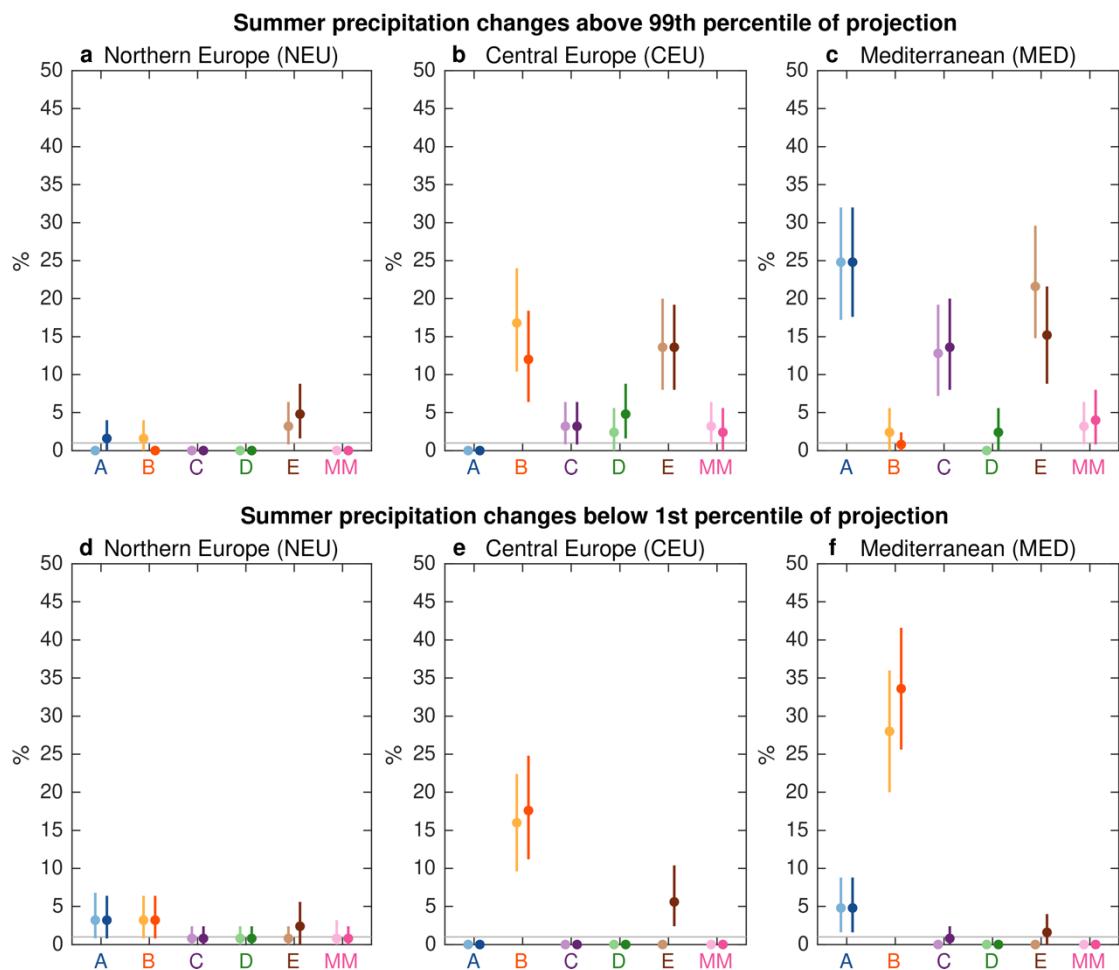
Supplementary Figure 11. Correlation of the different constrained ensemble mean projected change with each other across all 125 pseudo-observations for the 2041-2060 period in each of the European SREX regions.



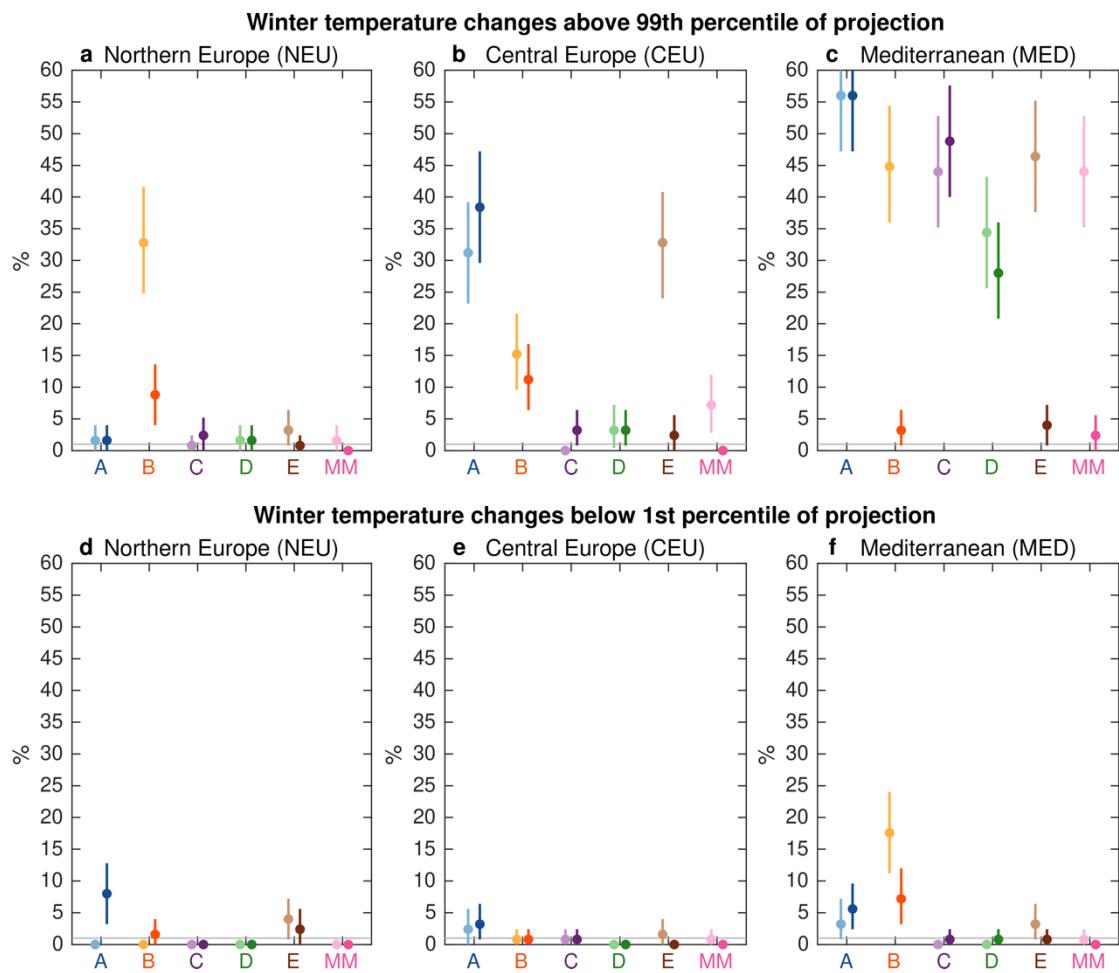
Supplementary Figure 12. Correlation of the different constrained ensemble mean projected change with (annual) global-mean surface-air temperature (GMST) change across all 125 pseudo-observations for the 2041–2060 period in each of the European SREX regions. Shown for temperature and precipitation for both the summer (JJA) and winter (DJF) seasons. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



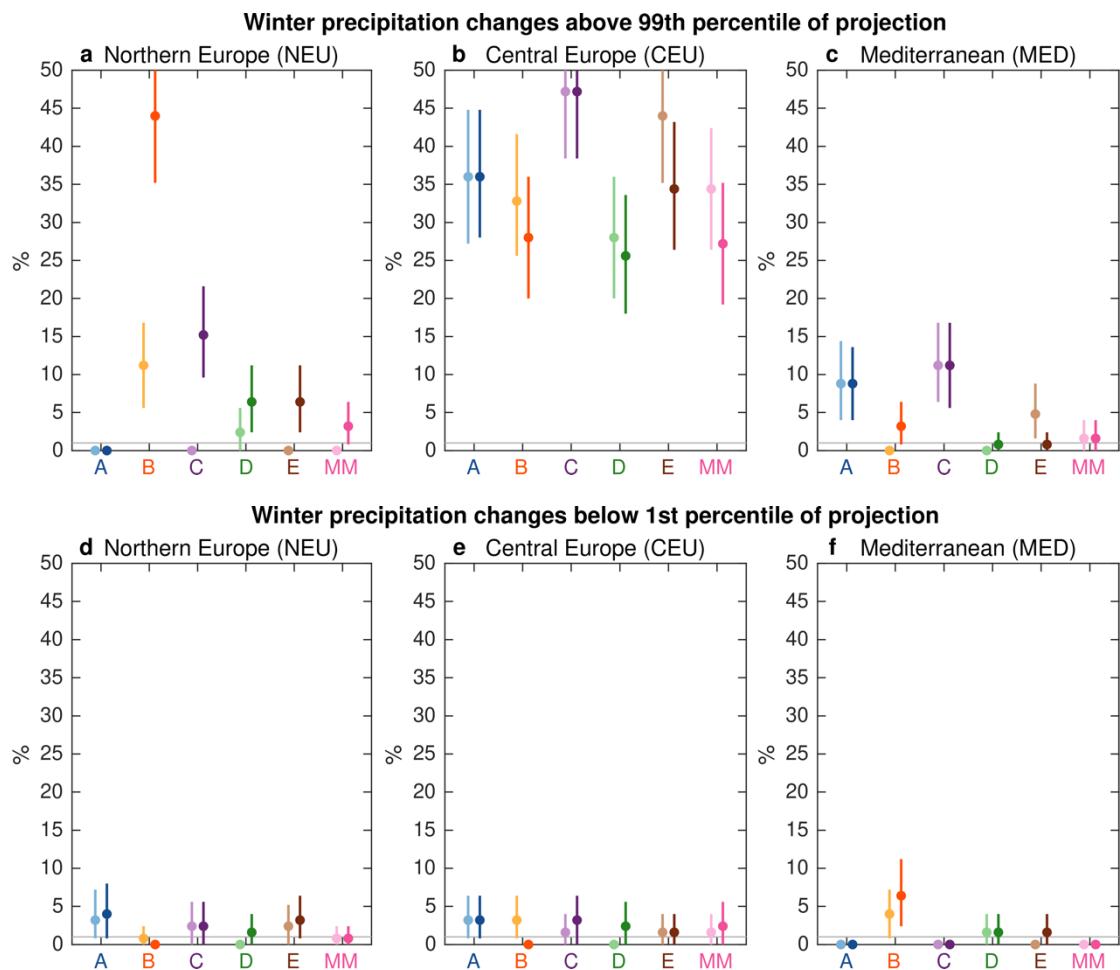
Supplementary Figure 13. Percentage of pseudo-observations outside the 1st and 99th percentiles of the unconstrained and constrained projections of future summer temperature changes (2041-2060). The grey horizontal line shows the 1% level expected for a perfectly reliable projection. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



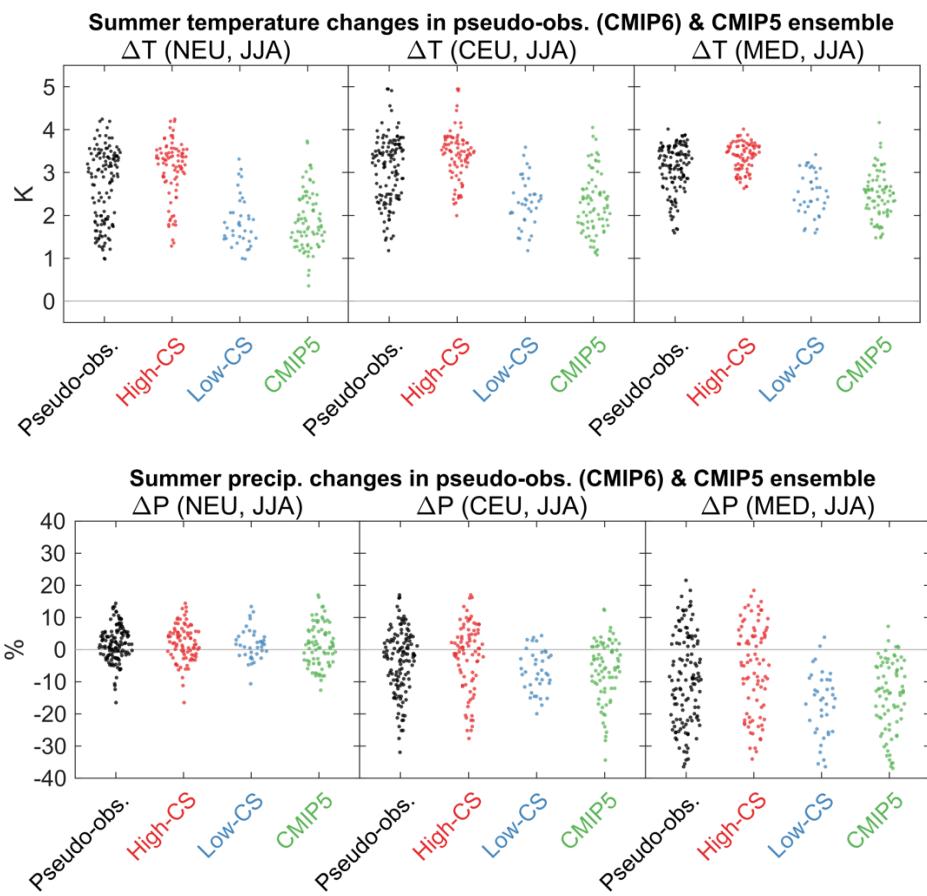
Supplementary Figure 14. Percentage of pseudo-observations outside the 1st and 99th percentiles of the unconstrained and constrained projections of future summer precipitation changes (2041–2060). The grey horizontal line shows the 1% level expected for a perfectly reliable projection. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



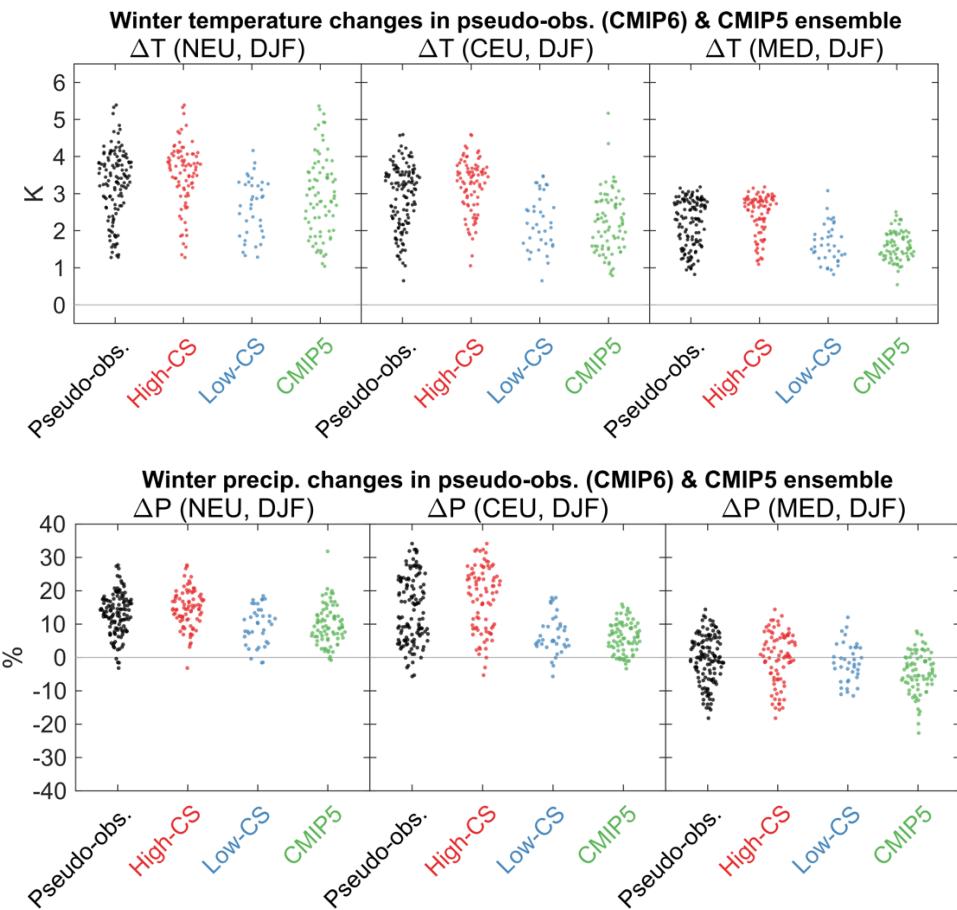
Supplementary Figure 15. Percentage of pseudo-observations outside the 1st and 99th percentiles of the unconstrained and constrained projections of future winter temperature changes (2041-2060). The grey horizontal line shows the 1% level expected for a perfectly reliable projection. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



Supplementary Figure 16. Percentage of pseudo-observations outside the 1st and 99th percentiles of the unconstrained and constrained projections of future winter precipitation changes (2041–2060). The grey horizontal line shows the 1% level expected for a perfectly reliable projection. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

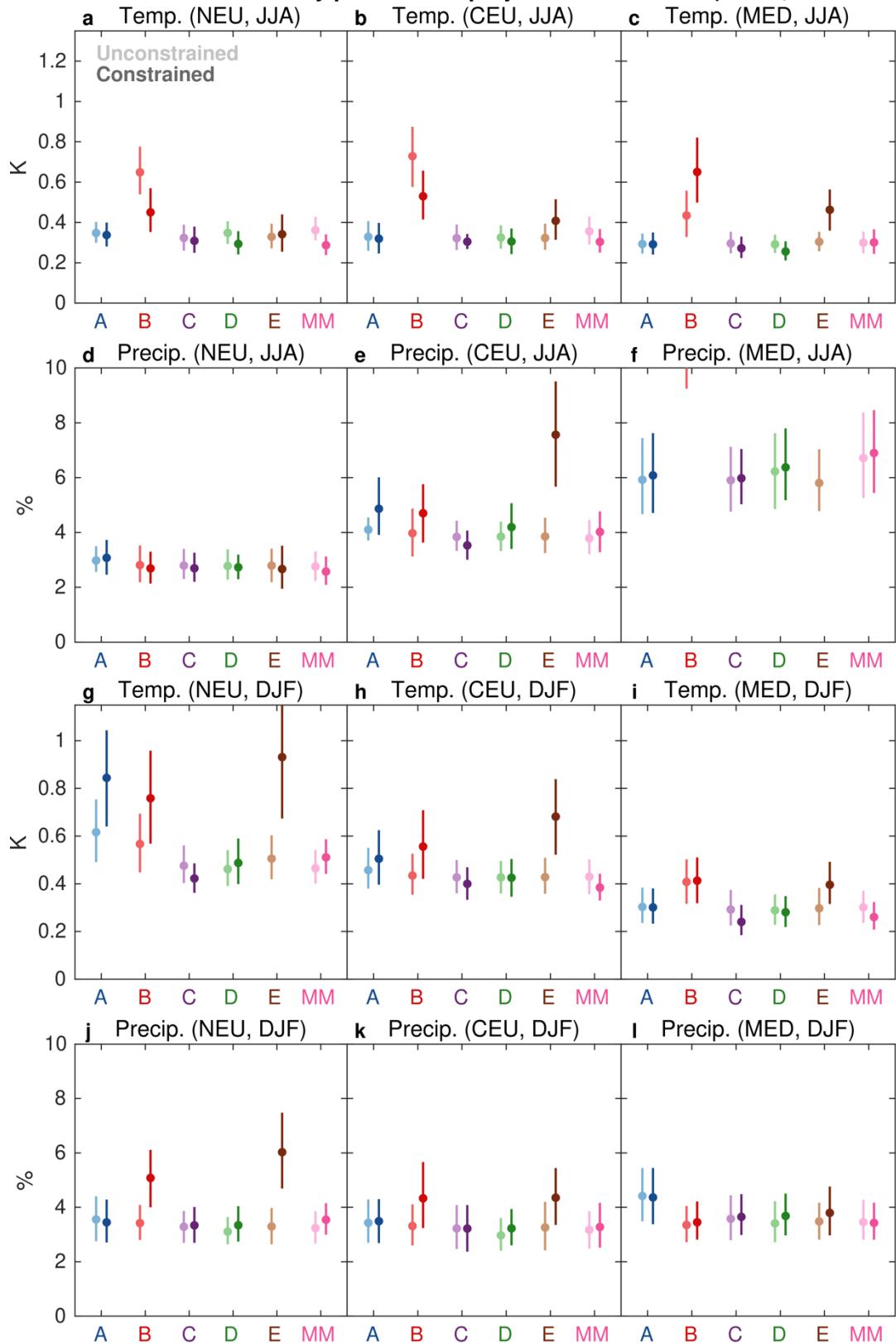


Supplementary Figure 17. Distribution of the 2041-2060 changes in the pseudo-observational datasets for the summer season, the high climate sensitivity (high-CS) subset, the low climate sensitivity (low-CS) subset and the CMIP5 ensemble (the exact set used by Method C is shown here).



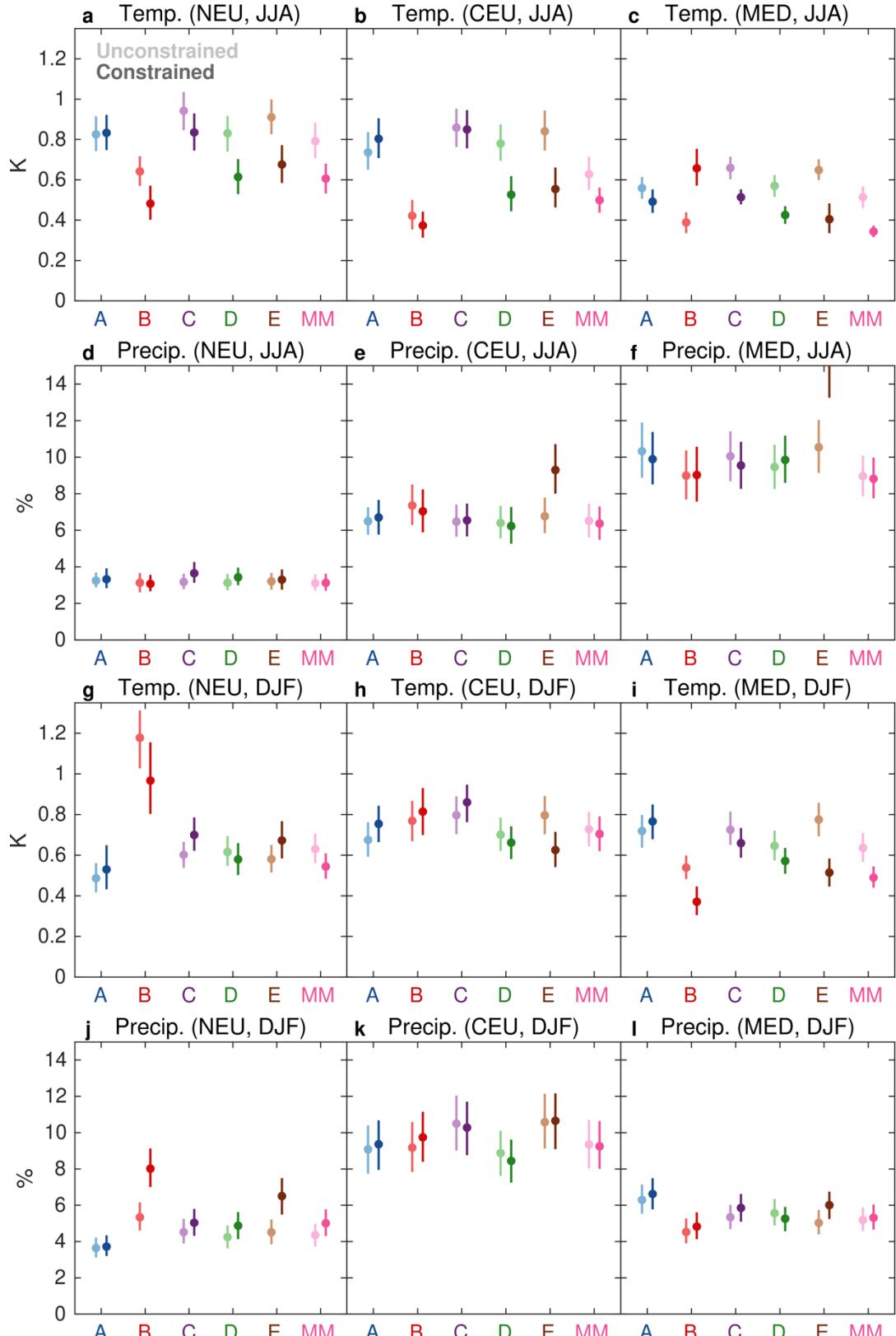
Supplementary Figure 18. Distribution of the 2041-2060 changes in the pseudo-observational datasets for the winter season, the high climate sensitivity (high-CS) subset, the low climate sensitivity (low-CS) subset and the CMIP5 ensemble (the exact set used by Method C is shown here).

Lower climate sensitivity pseudo-obs projection verification (CRPS, 2041-2060)

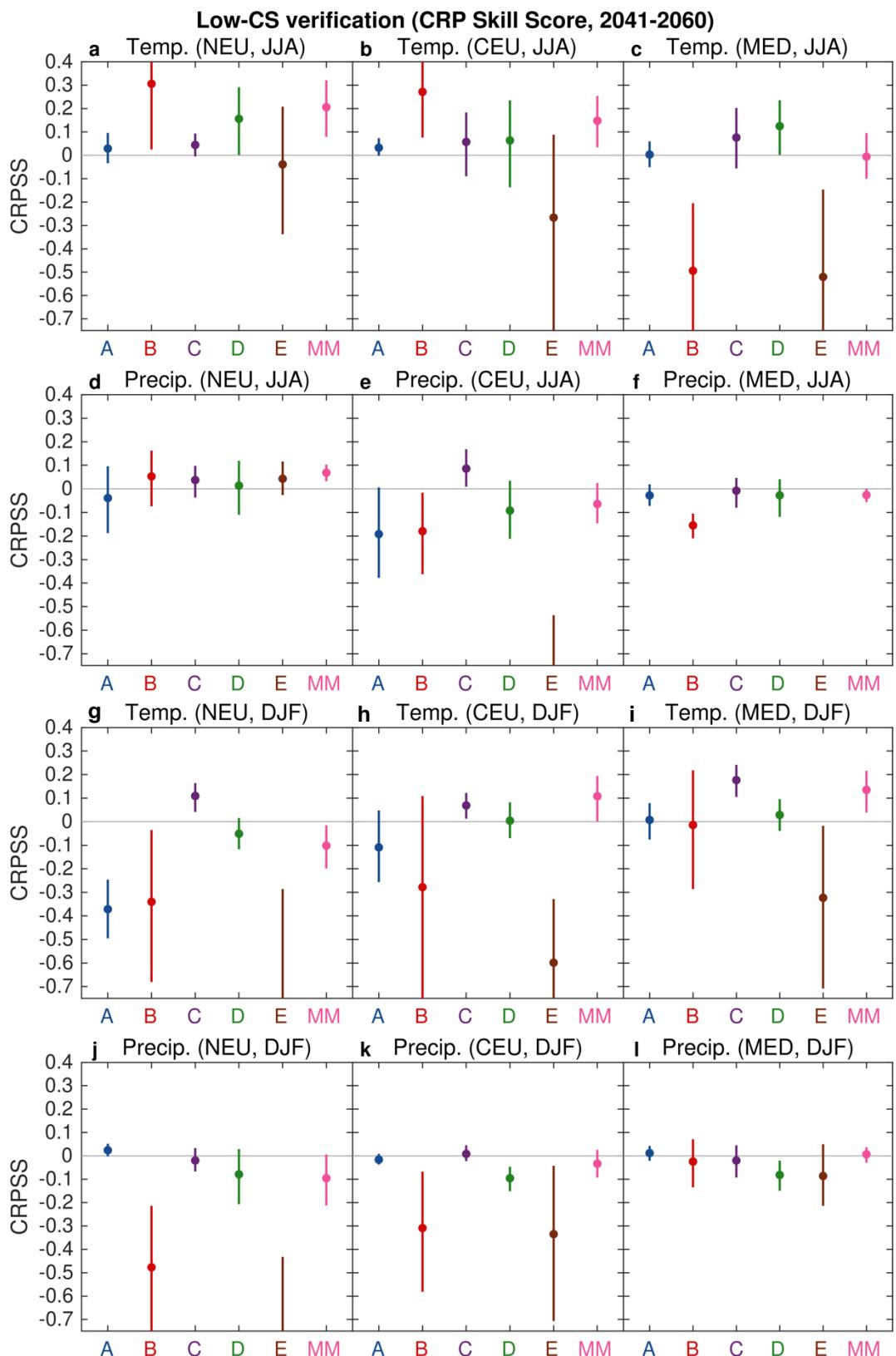


Supplementary Figure 19. CRPS for all methods and seasons for the subset of 38 pseudo-observations from CMIP6 models with lower Equilibrium Climate Sensitivities that are within the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

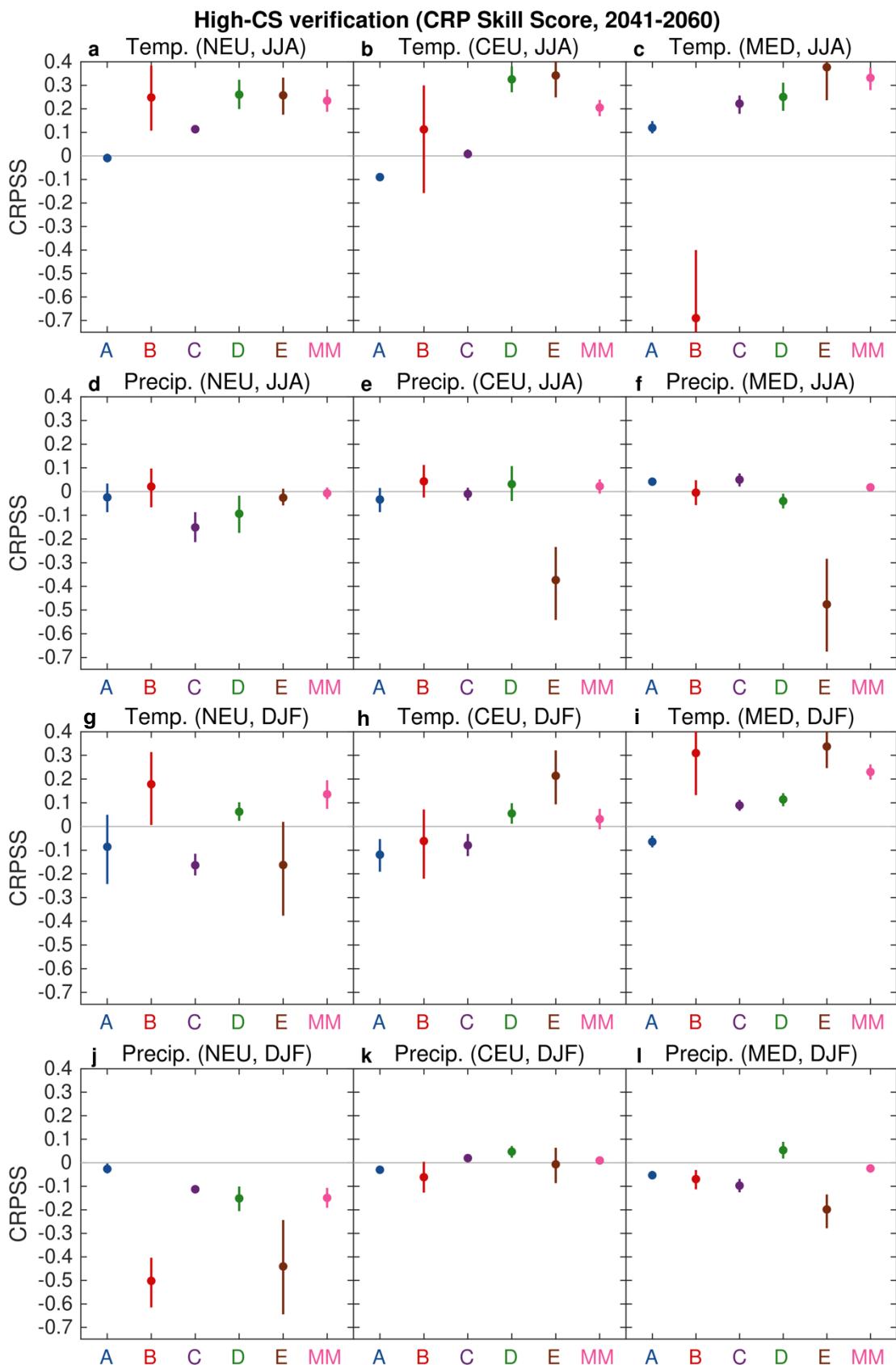
Higher climate sensitivity pseudo-obs projection verification (CRPS, 2041-2060)



Supplementary Figure 20. CRPS for all methods and seasons for the subset of 87 pseudo-observations from CMIP6 models with higher Equilibrium Climate Sensitivities that are outside the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

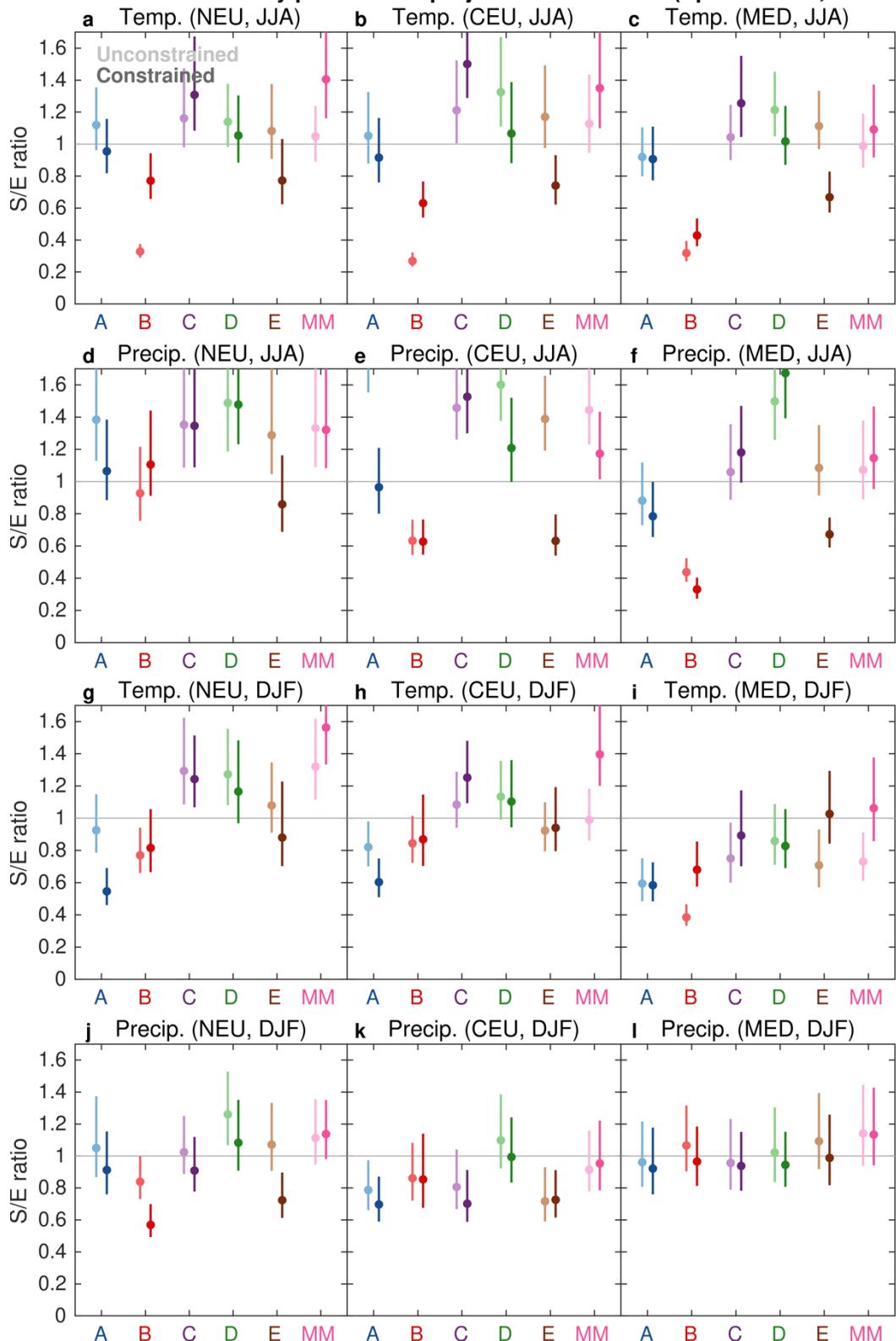


Supplementary Figure 21. CRPSS for all methods and seasons for the subset of 38 pseudo-observations from CMIP6 models with lower Equilibrium Climate Sensitivities that are within the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



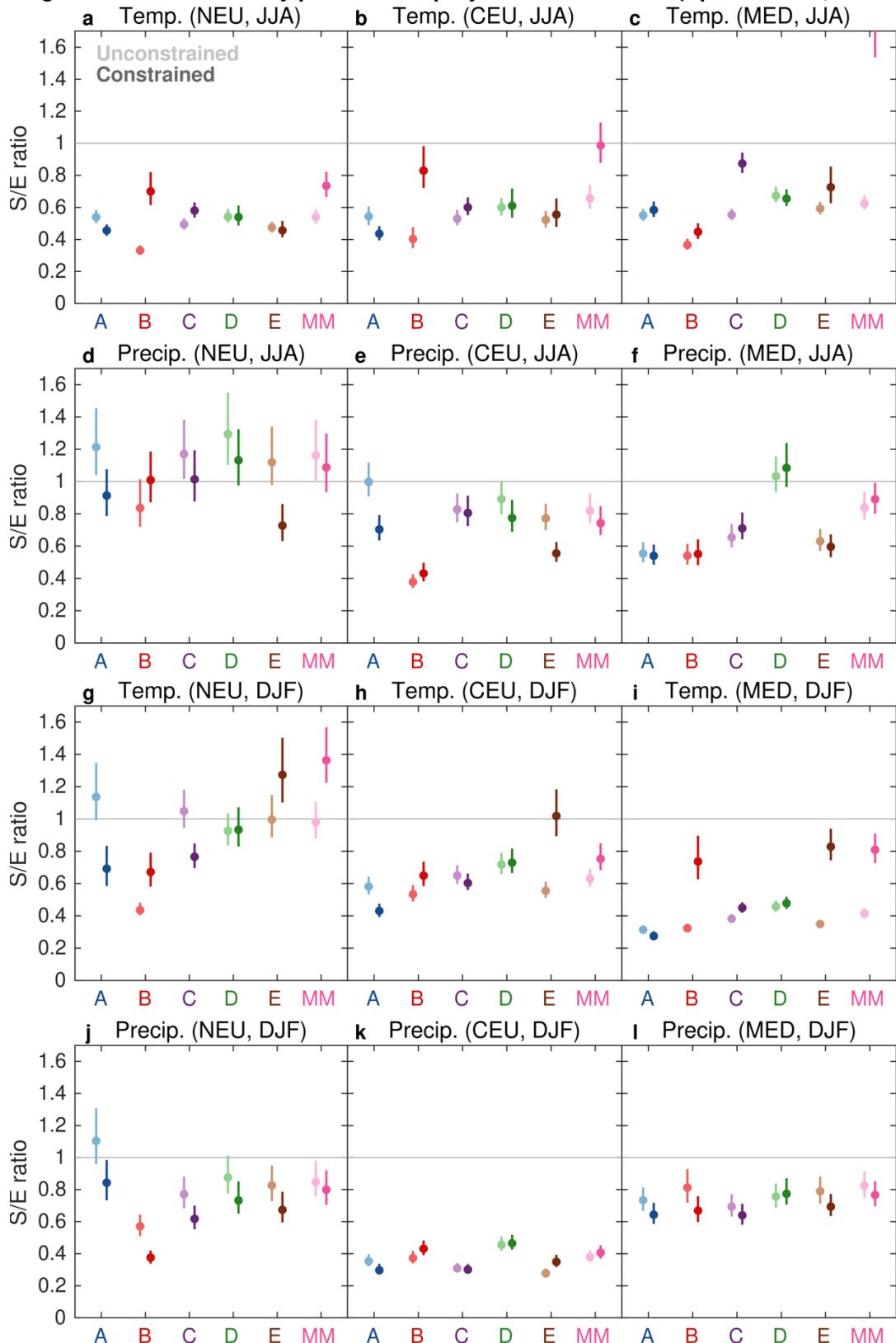
Supplementary Figure 22. CRPSS for all methods and seasons for the subset of 87 pseudo-observations from CMIP6 models with higher Equilibrium Climate Sensitivities that are outside the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

Lower climate sensitivity pseudo-obs projection verification (Spread/Error, 2041-2060)

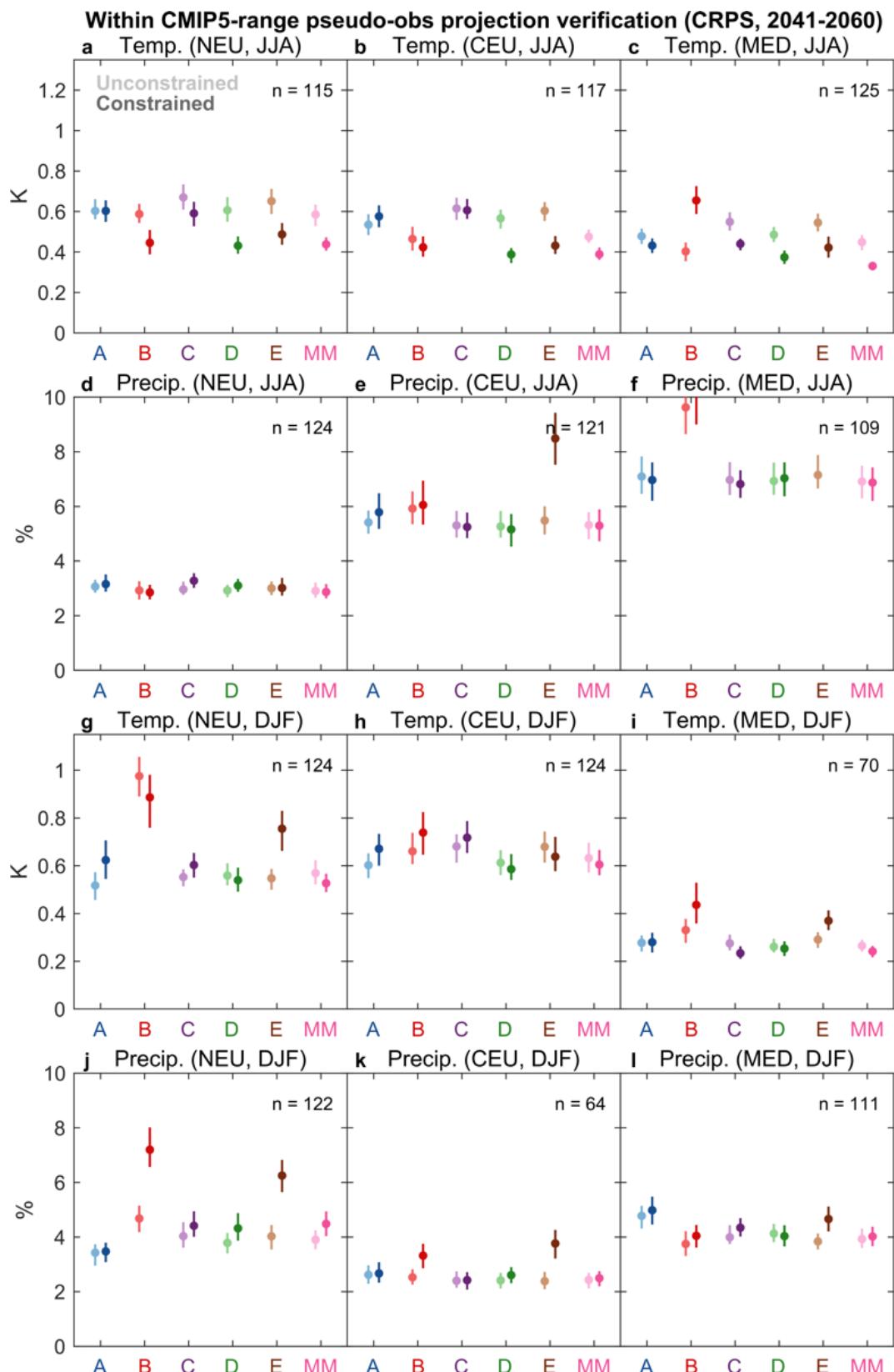


Supplementary Figure 23. Spread/Error ratio for all methods and seasons for the subset of 38 pseudo-observations from CMIP6 models with lower Equilibrium Climate Sensitivities that are within the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

Higher climate sensitivity pseudo-obs projection verification (Spread/Error, 2041-2060)



Supplementary Figure 24. Spread/Error ratio for all methods and seasons for the subset of 87 pseudo-observations from CMIP6 models with higher Equilibrium Climate Sensitivities that are within the minimum-maximum range of CMIP5 models (taken from Zelinka et al., 2020). The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).



Supplementary Figure 25. CRPS for all methods and seasons for the subset of pseudo-observations in which the future change falls within the minimum-maximum range of the CMIP5 models (shown in Figures S16 & S17). The numbers in each panel show the number of pseudo-observations that fall within the CMIP5 range. The dots show the measured values and the lines indicate the 95% confidence intervals based on a bootstrap resampling (see Methods).

	A	B	C	D	E
	REA	CALL	ClimWIP	KCC	ASK
Assumes truth centered	✓				
Constrained range can lie beyond unconstrained range		✓			✓
Spatial scale at which constraint or performance weighting is calculated	Local	Target region	Target region	Global + local	Europe
Multiple variables used to weight each target variables			✓		
Samples structural model uncertainty	✓		✓	✓	✓
Estimate of method error			✓		

Supplementary Table 1. An overview of some key characteristics of the different constraining methods used in this study. This table was adapted from Brunner et al (2020).

METHOD	T-JJA-NEU	T-JJA-CEU	T-JJA-MED	T-JJA-MEAN	RMSE rank										
	5	5	4	4.67	T-DIF-NEU	T-DIF-CEU	T-DIF-MED	T-DIF-MEAN	P-JJA-NEU	P-JJA-CEU	P-JJA-MED	P-JJA-MEAN			
A	5	5	4	4.67	4	1	6	3.67	4	3	3.33	1	2	6	3
B	1	1	6	2.67	6	6	2	4.67	1	5	3.67	5	5	1	3.67
C	6	6	3	5	3	5	5	4.33	6	2	3.33	3	4	4	3.67
D	3	2	2	2.33	1	3	3	2.33	5	4	4.33	4	1	3	2.67
E	4	4	5	4.33	5	4	4	4.33	3	6	6	6	5	5	5.67
NN	2	3	2	1.67	2	2	1	1.67	1	2	1	1.33	2	2	2.33

METHOD	T-DIF-NEU	T-DIF-CEU	T-DIF-MED	T-DIF-MEAN	CRPS rank							
	P-JJA-NEU	P-JJA-CEU	P-JJA-MED	P-JJA-MEAN	P-DIF-NEU	P-DIF-CEU	P-DIF-MED	P-DIF-MEAN	P-DIF-NEU	P-DIF-CEU	P-DIF-MED	P-DIF-MEAN
A	6	5	4	4.67	6	1	5	4	3	3	6	3.33
B	1	1	6	2.67	6	6	1	4.33	2	5	4	4
C	5	6	5	5.33	3	5	5	4.33	6	3	2	3.67
D	3	2	2	2.67	2	1	4	2.33	4	1	2	3
E	4	3	3	3.67	5	3	3	3.67	6	5	5	5.33
NN	2	1	2	1.67	1	2	1	1.67	4	2	2	2.67

Supplementary Table 2. Comparison of the accuracy rank (in terms of RMSE and CRPS) of the constrained projections for each of the regions/variables analysed in this study. Also shown is the average rank across the three European regions for each variable. The numbers indicate the accuracy ranking of the constrained projections (from 1-6) based on the lowest RMSE/CRPS across, where 1=best (i.e. lowest error).