Geophysical Research Letters

Supporting Information for

Sources of uncertainty in multi-model large ensemble projections of the winter North Atlantic Oscillation

C. M. McKenna[1] and A. C. Maycock [1]

[1] School of Earth and Environment, University of Leeds, Leeds, UK

Corresponding author: Christine McKenna (C.McKenna1@leeds.ac.uk)

Contents of this file

Introduction

This document contains additional text, figures, and tables that provide more technical detail on the methods/datasets used and investigate the sensitivity of the results to our methodological choices. Text S1 provides more detail on the uncertainty decomposition method of Maher et al. (2021). Text S2 explains how N_{min} is calculated in Figure 3. Text S3 explains how internal variability (IV) in the DJF NAO index was quantified for each model and observation-based dataset used. Figure S1 shows the historical NAO patterns used in Figures 2 and 4 to decompose an MSLP anomaly map into an NAO-congruent part and a residual. Figure S2 shows the effect of including the EA pattern in this decomposition for Figure 2. Figure S3 shows the effect of adjusting the model-based estimates of IV used in Figure 3a-d to an observation-based estimate of IV. Figures S4 and S5 are versions of Figures 2 and 4, respectively, but for the zonal wind at 850 hPa. Tables S1 and S2 respectively provide a detailed list of the MMLEA model simulations and CMIP5/6 model simulations used in the study.

Text S1. Separating uncertainty into parts due to IV and model structural differences

The total uncertainty (U) in projections of the DJF NAO index (X) across the MMLEA models is separated into a part due to IV (U_{IV}) and a part due to model structural differences (U_{MD}) using the method of Maher et al. (2021). This method is described in detail below.

The projected change in X in a single ensemble member (i) of a single MMLEA model (m) is given by:

$$\Delta X_{m,i} = \bar{X}_{m,i,\mathrm{fut}} - \bar{X}_{m,i,\mathrm{pres}}$$

where overbars indicate a time mean over a future (fut) or near-present-day (pres) 20-year epoch. The forced response in X in a single model (m) is given by the ensemble mean projected change:

$$\Delta X_{m,\mathrm{F}} = \frac{1}{N_m} \sum_{i=1}^{N_m} \Delta X_{m,i}$$

where N_m is the ensemble size for the model. The spread in ΔX across a model (m) due to IV is calculated as the inter-member standard deviation of the projected change:

$$\sigma(\Delta X_m) = \sqrt{\frac{1}{N_m - 1} \sum_{i=1}^{N_m} (\Delta X_{m,i} - \Delta X_{m,F})^2} \quad .$$

The uncertainty in ΔX due to IV (U_{IV}) is then given by the average of the IV across the models:

$$U_{\mathrm{IV}} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} \sigma^2(\Delta X_m)}$$

where M is the number of MMLEA models.

The MMM forced response in ΔX for the MMLEA models is calculated as the mean of the forced responses for each model:

$$\Delta X_{\mathrm{F}} = \frac{1}{M} \sum_{m=1}^{M} \Delta X_{m,\mathrm{F}} \quad .$$

The variance in the forced response across the models is then estimated as:

$$\sigma_{\mathrm{F}}{}^2 = D^2 - E^2$$

where D^2 is the variance in the ensemble means:

$$D^2 = \frac{1}{M-1} \sum_{m=1}^{M} (\Delta X_{m,\mathrm{F}} - \Delta X_{\mathrm{F}})^2$$

and E^2 removes the contribution of IV to the variance in the ensemble means. E^2 is equal to the average mean squared error of the models:

$$E^2 = \frac{1}{M} \sum_{m=1}^{M} \frac{\sigma^2 (\Delta X_m)}{N_m} \quad .$$

The uncertainty in ΔX due to model structural differences (U_{MD}) is then estimated as:

$$U_{\mathrm{MD}} = \sqrt{{\sigma_{\mathrm{F}}}^2} \quad .$$

We quantify the contribution of U_{MD} and U_{IV} to the total uncertainty in projections (U), by calculating the percentage variance contribution of each ($\%U_{\mathrm{MD}}$ and $\%U_{\mathrm{IV}}$) to the sum of U_{MD} and U_{IV}. To estimate the contributions of U_{MD} and U_{IV} to real-world uncertainty in the future NAO response, the model-based estimate of U_{IV} is replaced with an observation-based estimate of IV. Specifically, the IV in each MMLEA model, $\sigma(\Delta X_m)$, is replaced with an estimate of IV from Obs LE calculated as described in Text S3. Note that there are minimal differences to the results when using 20CRv3 or ERA20C.

Text S2. Calculation of N_{min}

To estimate the minimum ensemble size (N_{min}) required to detect a robust forced NAO index response of a given magnitude (X) between any two 20-year epochs, we follow the method of Screen et al. (2014).

First, we calculate the Student's t-statistic, t, for many different ensemble sizes, N, using a Student's t-test for a difference of means (von Storch & Zwiers, 1999):

$$t = \frac{X}{\sigma\sqrt{2/N}}$$

where σ is the standard deviation of 20-year epoch means due to IV (Text S3). Note it is assumed that σ is constant for all N, which Screen et al. (2014) show is a reasonable assumption.

Second, we define the difference, X, as statistically significant when $t \geq t_c$, where t_c is the cutoff value of the Student's t-distribution for a two-sided p-value of 0.025 and $2N$-2 degrees of freedom. N_{min} is the smallest value of N for which this is satisfied. Hence, N_{min} is calculated by re-arranging the Student's t-test, and replacing t with t_c and N with N_{min}:

$$N_{\mathrm{min}} = 2t_c^2 \times (\sigma/X)^2 \quad .$$

Text S3. Methods for calculating IV

All methods described below are applied to the DJF NAO index.

For each CMIP5/6 model, the IV in 20-year epoch means (Table S2) is calculated as the standard deviation in non-overlapping 20-year epoch means from the piControl simulation. This is multiplied by the square root of 2 when a difference in 20-year epoch means is of interest; this assumes the two 20-year epochs are independent and have the same variance (Collins et al., 2013). As in Collins et al. (2013), the median IV across all models is used for the MMM. Non-overlapping 20-year epochs are used to ensure each sample is independent.

The IV in 20-year epoch means for each MMLEA model (Table S1) is calculated as the inter-member standard deviation of a 20-year epoch mean, where this is pooled (i.e., averaged) for all possible 20-year epochs over 1951-2099. The same method is used for Obs LE, but over the period 1922-2014. For ERA20C and 20CRv3, we use the standard deviation of all possible 20-year epoch means over 1901-2010; given the limited temporal extent of these records, non-overlapping segments could not be used. For consistency between all datasets considered, the IV for a difference in 20-year epoch means is the IV in 20-year epoch means multiplied by the square root of 2. In all cases we assume the IV is constant in time; analysing timeseries of the inter-member standard deviation in 20-year epoch means for the MMLEA models suggests this is a reasonable assumption.

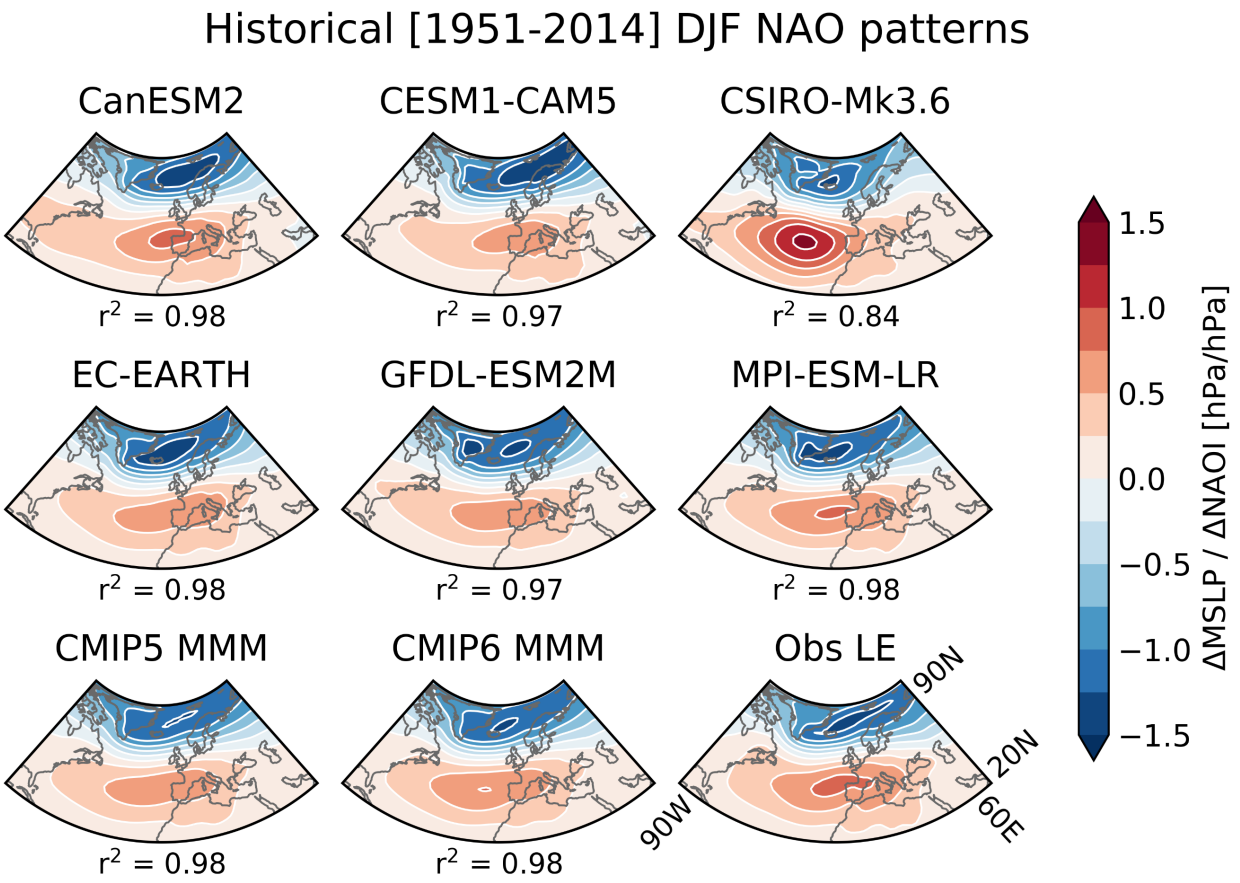

Figure S1. Historical [1951-2014] DJF NAO patterns for the MMLEA models, CMIP5/6 MMM and an observation-based dataset. Shading shows the change in MSLP (hPa) for a 1 hPa positive change in NAO index. Patterns for each CMIP5/6 model and Obs LE are calculated in the same way as for the MMLEA models (Section 2.2). Ensemble means are used to define the patterns to minimise uncertainty in the NAO pattern due to IV (e.g., see Simpson et al., 2020). Obs LE is used for the observation-based NAO pattern because it is designed to be less affected by sampling issues; note that there are minimal differences when using 20CRv3 or ERA20C, or when using a longer historical period. r^2 is the squared area-weighted pattern correlation between the modelled and observation-based NAO patterns. Largely, the modelled and observation-based patterns are highly correlated. The southern centre of action, however, is generally weaker in the models and in CSIRO-Mk3.6 the pattern is westward shifted. There is little improvement in the CMIP6 MMM compared to the CMIP5 MMM.

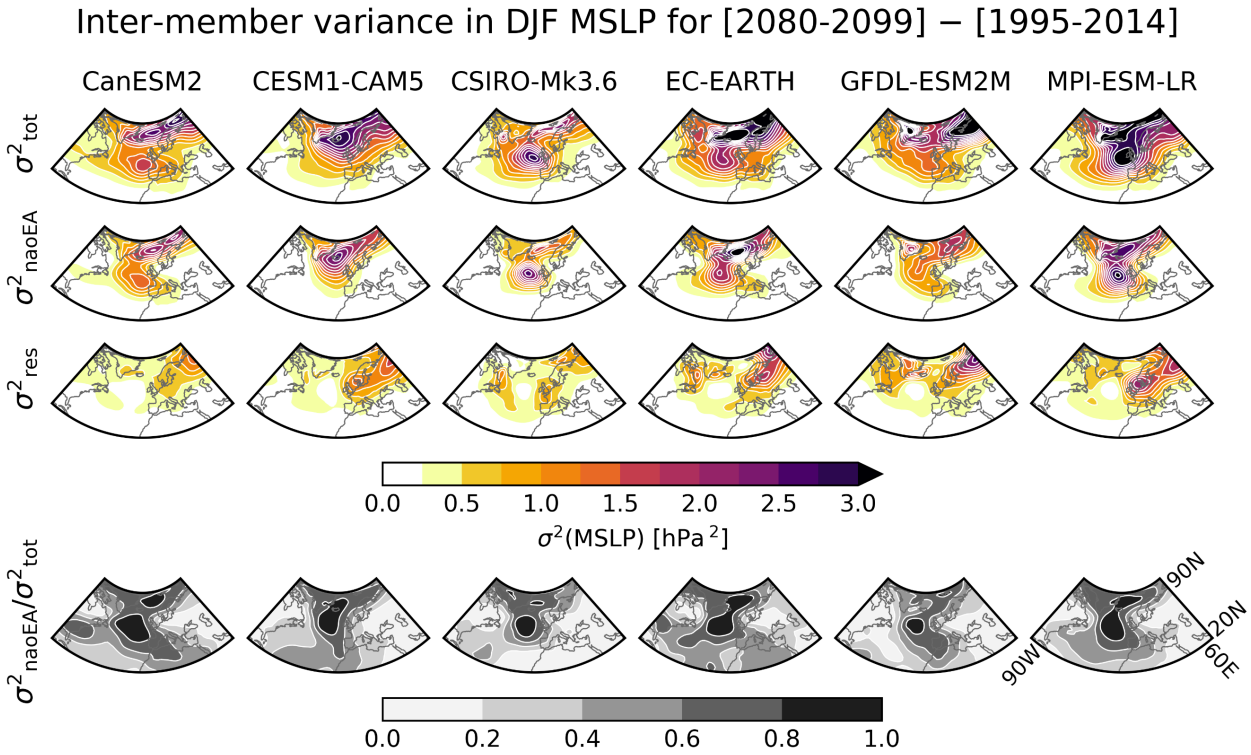

Figure S2. Same as Figure 2, but with the EA pattern included in the regression. [Top row] Total variance (σ^2_{tot}); [Second row] Variance explained by the NAO and EA (σ^2_{naoEA}); [Third row] Residual variance (σ^2_{res}); [Bottom row] Proportion of total variance explained by the NAO and EA. σ^2_{naoEA} is obtained through multivariate regression at each grid-point of the total inter-member spread in MSLP on the spread in NAO-congruent MSLP and spread in EA-congruent MSLP. σ^2_{res} is the variance in the residuals of this regression. The EA pattern is characterised by a monopole in MSLP over the mid-latitude North Atlantic ocean (Barnston & Livezey, 1987; Moore et al., 2011; Wallace & Gutzler, 1981). Following Moore et al. (2011), the EA index is calculated as the anomalous MSLP in the nearest gridbox to (52.5N, 27.5W). The EA-congruent part of the MSLP is obtained using the same procedure as for the NAO-congruent part (Section 2.2).

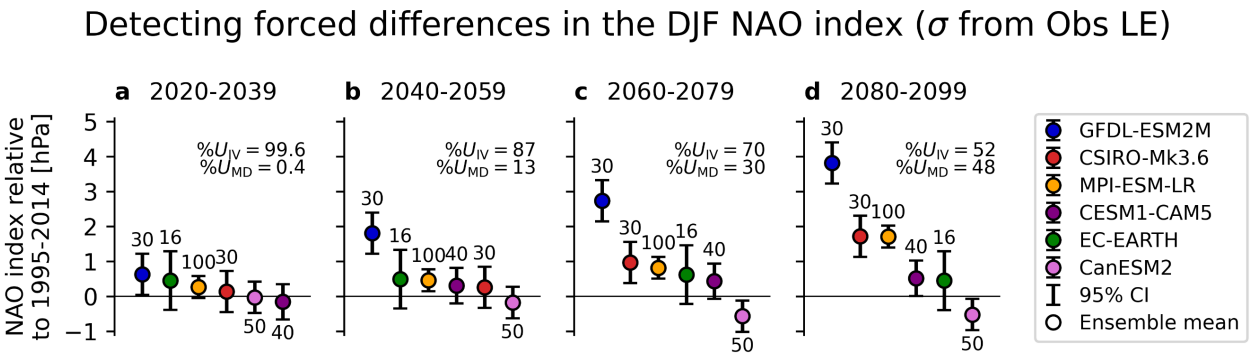

Figure S3. Same as Figure 3a-d, but with confidence intervals calculated by replacing the model-based estimates of IV with observation-based estimates of IV. Obs LE is used for the observation-based IV estimate because it is designed to be less affected by sampling issues; note that there are minimal differences when using 20CRv3 or ERA20C. IV is estimated as described in Text S3. $\%U_{IV}$ and $\%U_{MD}$ are defined as described in Text S1 using Obs LE to estimate IV.

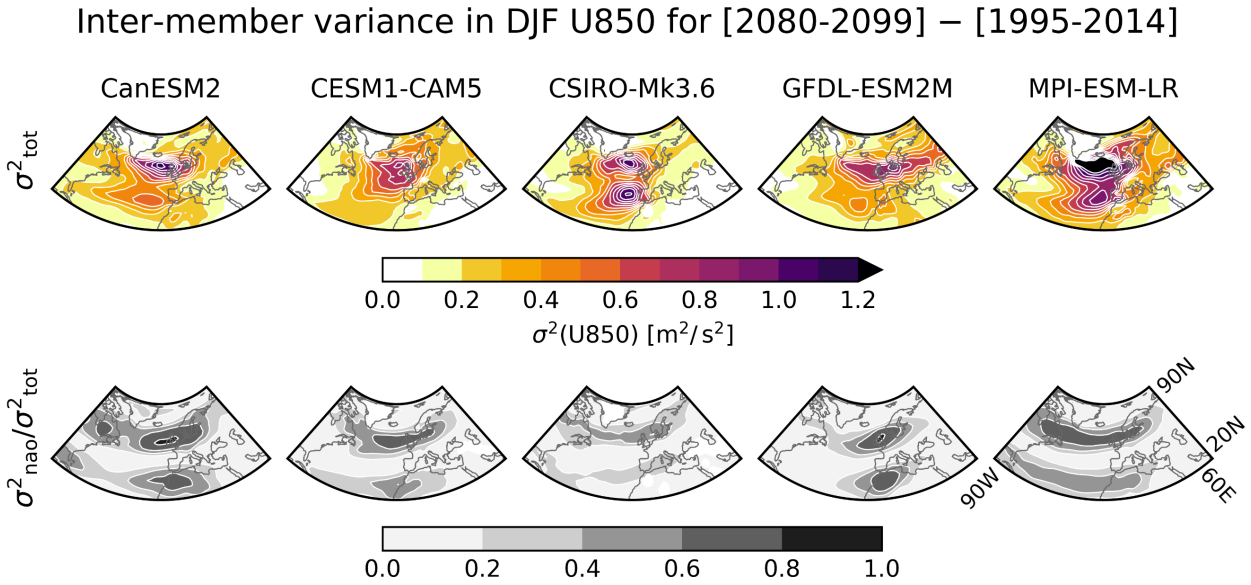

Figure S4. Same as top and bottom rows of Figure 2, but for U850. [Top row] Total variance (σ^2_{tot}); [Bottom row] Proportion of total variance explained by the NAO (σ^2_{nao} / σ^2_{tot}). σ^2_{nao} is obtained by regressing the total inter-member spread in U850 on the spread in NAO-congruent U850 at each grid-point. The NAO-congruent part of U850 is obtained using the same procedure as for MSLP (Section 2.2), but with the historical NAO pattern constructed by regressing historical timeseries of U850 at each grid-point onto the NAO index timeseries. EC-EARTH is omitted from the analysis because there was no three-dimensional zonal wind data available.

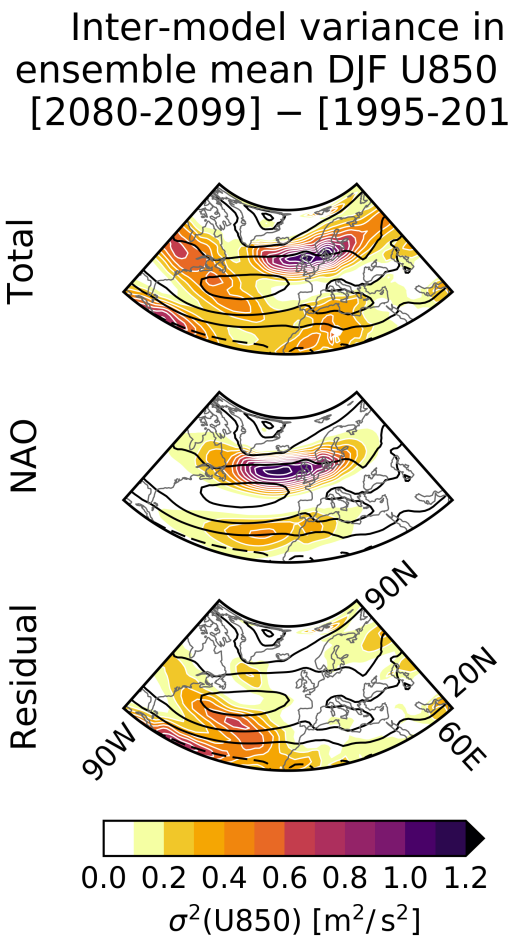

Figure S5. Same as far right column of Figure 4, but for U850. [Top] Total; [Middle] NAO-congruent part; [Bottom] Residual. The NAO-congruent part of U850 is obtained using the same procedure as for MSLP (Section 2.2), but with the historical NAO pattern constructed by regressing historical timeseries of U850 at each grid-point onto the NAO index timeseries. EC-EARTH is omitted from the analysis because there was no three-dimensional zonal wind data available. Note that similar results are obtained for the far right column of Figure 4 when EC-EARTH is removed. Black contours show the MMM near-present-day (1995-2014) U850 climatology with intervals of 10 m/s.

Table S1. List of MMLEA models with historical and RCP8.5 simulations. IV is for 20-year means of the DJF NAO index over 1951-2099 (see Text S3 for details). In all MMLEA models, this IV is underestimated compared to observation-based datasets (1.1 hPa, 1.2 hPa, and 1.2 hPa in Obs LE, 20CRv3, and ERA20C, respectively). Note that while the MMLEA does contain an ensemble for GFDL-ESM2M, there is no three-dimensional zonal wind data available for this model. We therefore use a similar 30 member ensemble from the Princeton Large Ensemble Archive (Schlunegger et al., 2019), which has three-dimensional zonal wind data available. The NAO index and MSLP results are very similar for the two ensembles.

Model	Modelling Centre	CMIP generation	Years	No. of members	IV (hPa)	Reference
CanESM2	CCCma	CMIP5	1950-2100	50	0.72	Kirchmeier-Young et al. (2017)
CESM1-CAM5	NCAR	CMIP5	1920-2100	40	0.77	Kay et al. (2015)
CSIRO-Mk3.6	CSIRO	CMIP5	1850-2100	30	0.68	Jeffrey et al. (2013)
EC-EARTH	EC-Earth Consortium	CMIP5	1860-2100	16	0.85	Hazeleger et al. (2010)
GFDL-CM3	GFDL	CMIP5	1920-2100	20	0.77	Sun et al. (2018)
GFDL-ESM2M	GFDL	CMIP5	1950-2100	30	0.93	Rodgers et al. (2015); Schlunegger et al. (2019)
MPI-ESM-LR	MPI	CMIP5	1850-2099	100	0.84	Maher et al. (2019)

Table S2. List of CMIP5/CMIP6 models with piControl, historical and RCP8.5/SSP5-8.5 simulations. Numerical labels are for bars in Figure 1. Models are ranked in order of magnitude of IV in 20-year means of the DJF NAO index from the piControl simulations (see Text S3 for details), where rank 1 has the highest IV and rank 75 has the lowest. This enables each model to be located in the grey plumes of Figure 3e-f. In most models the IV is underestimated compared to observation-based datasets (respectively 1.11 hPa, 1.18 hPa, and 1.20 hPa in Obs LE, 20CRv3, and ERA20C). Note that for CMIP5 models that are also MMLEA models, the IV magnitudes listed here do not necessarily match those in Table S1. For example, based on the piControl simulations CESM1-CAM5 has a very low IV, but based on the MMLEA simulations it has an average IV. This likely reflects that the piControl IV is calculated from a relatively short simulation (319 years) with only 15 independent samples of 20-year means, while there are 40 independent ensemble members for the MMLEA simulations. It could also be that there are differences in the magnitude of IV between the pre-industrial state and historical/RCP8.5 state, but this cannot be determined with the limited piControl simulation length.

Label	Model	Modelling Centre	CMIP generation	piControl length (years)	Number of historical/ RCP/SSP members	IV (hPa)	IV rank
1	ACCESS1.0	CSIRO-BOM	CMIP5	500	1	0.62	66
2	ACCESS1.3		CMIP5	500	1	0.84	23
3	BCC-CSM1.1	BCC	CMIP5	500	1	0.77	40
4	BCC-CSM1.1-M		CMIP5	400	1	0.86	18
5	BNU-ESM	BNU	CMIP5	559	1	1.14	2
6	CanESM2	CCCma	CMIP5	996	5	0.68	56
7	CCSM4	NCAR	CMIP5	1051	6	0.83	27
8	CESM1-BGC	NSF-DOE-NCAR	CMIP5	500	1	0.89	14
9	CESM1-CAM5		CMIP5	319	3	0.52	72
10	CESM1-WACCM		CMIP5	200	3	0.45	74
11	CMCC-CESM	CMCC	CMIP5	277	1	1.03	5

Label	Model	Modelling Centre	CMIP generation	piControl length (years)	Number of historical/ RCP/SSP members	IV (hPa)	IV rank
12	CMCC-CM	CMCC	CMIP5	330	1	0.67	59
13	CMCC-CMS		CMIP5	500	1	0.80	32
14	CNRM-CM5	CNRM-CERFACS	CMIP5	850	5	0.78	39
15	CSIRO-Mk3.6.0	CSIRO-QCCCE	CMIP5	500	10	0.67	60
16	EC-EARTH	ICHEC	CMIP5	451	8	0.80	34
17	FGOALS-g2	LASG-CESS	CMIP5	700	1	0.64	63
18	FIO-ESM	FIO	CMIP5	800	3	0.81	30
19	GFDL-CM3	NOAA-GFDL	CMIP5	500	1	0.66	61
20	GFDL-ESM2G		CMIP5	500	1	0.93	11
21	GFDL-ESM2M		CMIP5	500	1	0.74	47
22	GISS-E2-H	NASA-GISS	CMIP5	780	2	0.63	65
23	GISS-E2-H-CC		CMIP5	251	1	0.37	75
24	GISS-E2-R		CMIP5	850	2	0.80	31
25	GISS-E2-R-CC		CMIP5	251	1	0.75	44
26	HadGEM2-CC	MOHC	CMIP5	240	3	0.84	26
27	HadGEM2-ES		CMIP5	576	4	0.86	20
28	INM-CM4	INM	CMIP5	500	1	0.74	46
29	IPSL-CM5A-LR	IPSL	CMIP5	1000	4	0.82	28

Label	Model	Modelling Centre	CMIP generation	piControl length (years)	Number of historical/ RCP/SSP members	IV (hPa)	IV rank
30	IPSL-CM5A-MR	IPSL	CMIP5	300	1	0.61	67
31	IPSL-CM5B-LR		CMIP5	300	1	1.25	1
32	MIROC-ESM	MIROC	CMIP5	630	1	0.79	35
33	MIROC-ESM-CHEM		CMIP5	255	1	0.71	51
34	MIROC5		CMIP5	670	3	0.55	71
35	MPI-ESM-LR	MPI-M	CMIP5	1000	3	0.93	10
36	MPI-ESM-MR		CMIP5	1000	1	0.80	33
37	MRI-CGCM3	MRI	CMIP5	500	1	0.94	9
38	NorESM1-M	NCC	CMIP5	501	1	0.78	38
39	NorESM1-ME		CMIP5	252	1	1.03	7
40	ACCESS-CM2	CSIRO-ARCCSS	CMIP6	500	1	0.84	24
41	ACCESS-ESM1.5	CSIRO	CMIP6	900	1	0.60	69
42	AWI-CM1.1-MR	AWI	CMIP6	500	1	0.78	37
43	BCC-CSM2-MR	BCC	CMIP6	600	1	1.09	3
44	CAMS-CSM1.0	CAMS	CMIP6	500	1	0.92	12
45	CanESM5	CCCma	CMIP6	1000	25	0.86	21
46	CanESM5-CanOE			501	1	0.75	45

Label	Model	Modelling Centre	CMIP generation	piControl length (years)	Number of historical/ RCP/SSP members	IV (hPa)	IV rank
47	CESM2	NCAR	CMIP6	1200	1	0.99	8
48	CESM2-WACCM		CMIP6	499	3	0.81	29
49	CIESM	THU	CMIP6	500	1	0.69	54
50	CMCC-CM2-SR5	CMCC	CMIP6	500	1	0.76	43
51	CNRM-CM6.1	CNRM-CERFACS	CMIP6	500	6	0.85	22
52	CNRM-CM6.1-HR		CMIP6	300	1	0.63	64
53	CNRM-ESM2.1		CMIP6	500	5	1.06	4
54	EC-Earth3-Veg	EC-Earth-Consortium	CMIP6	500	1	0.76	42
55	FGOALS-f3-L	CAS	CMIP6	561	1	0.79	36
56	FGOALS-g3		CMIP6	700	1	0.68	57
57	FIO-ESM2.0	FIO-QLNM	CMIP6	575	1	0.61	68
58	GFDL-CM4	NOAA-GFDL	CMIP6	500	1	0.70	53
59	GFDL-ESM4		CMIP6	500	1	0.90	13
60	HadGEM3-GC3.1-LL	MOHC	CMIP6	500	4	0.67	58
61	HadGEM3-GC3.1-MM		CMIP6	500	4	0.86	19
62	INM-CM4.8	INM	CMIP6	531	1	0.56	70
63	INM-CM5.0		CMIP6	1201	1	0.64	62

Label	Model	Modelling Centre	CMIP generation	piControl length (years)	Number of historical/ RCP/SSP members	IV (hPa)	IV rank
64	IPSL-CM6A-LR	IPSL	CMIP6	2000	3	0.88	15
65	KACE1.0-G	NIMS-KMA	CMIP6	450	1	0.87	16
66	KIOST-ESM	KIOST	CMIP6	500	1	0.72	49
67	MIROC-ES2L	MIROC	CMIP6	500	1	0.51	73
68	MIROC6		CMIP6	800	3	0.70	52
69	MPI-ESM1.2-HR	MPI-M	CMIP6	500	1	0.86	17
70	MPI-ESM1.2-LR		CMIP6	1000	1	0.69	55
71	MRI-ESM2.0	MRI	CMIP6	701	1	0.84	25
72	NESM3	NUIST	CMIP6	500	1	0.77	41
73	NorESM2-LM	NCC	CMIP6	501	1	1.03	6
74	NorESM2-MM		CMIP6	500	1	0.71	50
75	UKESM1.0-LL	MOHC	CMIP6	1880	5	0.73	48