https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity - evaluation of CMIP7 DECK simulations

Alejandro Bodas-Salcedo[1], John W. Rostron[1], Catherine Guiavarc'h[1], William Kay[1], Harry Mutton[1], Ed W. Blockley[1], Mark A. Ringer[1], Kuniko Yamazaki[1], Max Thomas[1, 4], Colin G. Jones[2], Timothy Andrews[1], Jane P. Mulcahy[1], Tamzin E. Palmer[1], Steven T. Rumbold[3], David M. H. Sexton[1], Yongming Tang[1], and Martin R. Willett[1]

[1]Met Office Hadley Centre, Exeter, United Kingdom
[2]National Centre for Atmospheric Science, School of Earth and the Environment, University of Leeds, Leeds, United Kingdom
[3]National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom
[4]British Antarctic Survey, Cambridge, United Kingdom

Correspondence: Alejandro Bodas-Salcedo (alejandro.bodas@metoffice.gov.uk)

Abstract. This study presents UKCM2-LL, a new configuration with constrained climate sensitivity aligned to the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) likely range. UKCM2-LL is based on the Met Office Hadley Centre Global Environmental Model Global Coupled configuration version 5.0 (HadGEM3 GC5.0). We document the development and evaluation of UKCM2-LL through CMIP7 DECK simulations, including the spin-up and equilibration of the pre-industrial control, the calculation of the Effective Climate Sensitivity (EffCS), Transient Climate Response (TCR), effective radiative forcing (ERF) and climate feedbacks from abrupt and ramped CO_2 experiments. The new UKCM2-LL configuration achieves an EffCS of 3.6 K, within the IPCC AR6 likely range of 2.5–4.0 K, primarily through more negative shortwave cloud feedbacks compared to the parent GC5.0 model (EffCS of 6.6 K). Evaluation against observations and CMIP6 models demonstrates improved simulation of the historical temperature record since the 1970s, with better agreement in interhemispheric temperature anomalies. The 1985-2014 global-mean trend reduces from 0.43 K per decade to 0.33 K per decade, closer to the observed trend of 0.18 K per decade. While present-day climatological metrics show approximately 30% higher root mean square errors (RMSE) across a wide range of metrics relative to GC5.0, climatological performance remains reasonable relative to CMIP6 models. The introduction of CMIP7 forcings degrades the simulation of the historical temperature record, especially the interhemispheric anomalies. This work demonstrates the efficacy of Perturbed Parameter Ensemble (PPE) methodologies to develop climate models with constrained emergent properties while maintaining acceptable present-day performance, providing a valuable tool for climate projection studies requiring models within the IPCC-assessed climate sensitivity range.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Copyright statement. The works published in this journal are distributed under the Creative Commons Attribution 4.0 License. This licence does not affect the Crown copyright work, which is re-usable under the Open Government Licence (OGL). The Creative Commons Attribution 4.0 License and the OGL are interoperable and do not conflict with, reduce or limit each other.

© Crown Copyright 2026

1 Introduction

The latest Global Coupled (GC) configuration of the Met Office Hadley Centre Global Environmental Model (HadGEM3) is GC5.0. The development of GC5.0 followed the standard GC development process, driven by parametrisation improvements and evaluation against the mean climate and variability of recent decades (Xavier et al., submitted). Emerging properties like the Effective Climate Sensitivity (EffCS), the Transient Climate Response (TCR) or the simulation of historical trends are not considered during the GC development process.

Equilibrium Climate Sensitivity (ECS) is the global-mean equilibrium temperature change expected from a doubling of atmospheric CO_2, and is a key metric of the climate system's response to anthropogenic forcing. Models' true ECS can only be obtained using very long runs to quasi-equilibrium, but practically the Effective Climate Sensitivity (EffCS) is used, which is calculated as an extrapolation of a linear regression of annual means of an *abrupt-4xCO2* experiment (Gregory et al., 2004). The EffCS of GC5.0 is 6.6K, which is outside the baseline range of the WCRP assessment of climate sensitivity (Sherwood et al., 2020) and the very likely range of the IPCC AR6 (Forster et al., 2021).

During the development phase of GC5.0, a consultation exercise with a range of science users was conducted. In this consultation, some users expressed a preference for a mid-range EffCS within the IPCC-assessed likely range (2.5-4.0K), in particular for carbon-cycle studies or other studies where additional feedbacks may be explored. Other users expressed a requirement for a configuration that could serve as a baseline to span a range of EffCS that includes values in the IPCC-assessed likely range using a GC5 Perturbed Parameter Ensemble (PPE). The quality of the simulated historical record was also highlighted as important by several user groups. These requirements led to the development of the UK Climate Model version 2 (UKCM2), a coupled configuration based on GC5.0 with an EffCS within the IPCC AR6 likely range. A detailed description of the methodology used to develop UKCM2 can be found in Rostron et al. (2026). The development of UKCM2 was done at a relatively low resolution (see Section 2.1), both in the atmosphere and in the ocean, and therefore the configuration is referred to as UKCM2-LL.

Here we document the outcome of that process. We present results of the UKCM2-LL simulations that will form part of the UK contribution to CMIP7. We use DECK simulations to document the UKCM2-LL response to forcing and its overall historical performance. The structure of the paper is as follows. Section 2 gives an overview of the model, the methodology used to develop the new configuration, and the simulations used in this paper. We then describe the spinup methodology and the overall quasi-equilibrium properties of the pre-industrial control simulation in Section 3. Section 4 analyses the response to CO2 forcing and compares it against GC5.0 and HadGEM3-GC3.1-LL, the low-resolution configuration submitted to CMIP6. Section 5 evaluates the simulation of the historical period, including the global mean temperature anomalies, the interhemi-

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

spheric temperature anomalies, and ocean circulation and sea-ice metrics. Section 6 evaluates the present-day climatology of UKCM2-LL in the context of CMIP6 models. Finally, Section 7 summarizes our conclusions.

2 Models, methodology, and simulations

2.1 GC5.0

UKCM2-LL is based on the Met Office's global coupled model, HadGEM3-GC5.0 (Xavier et al., submitted). It comprises Global Atmosphere and Land (Willett et al., 2026, GAL9) and Global Ocean and Sea Ice components (Guiavarc'h et al., 2025, GOSI9), with coupling provided by the OASIS3-MCT coupler (Craig et al., 2017). Compared to HadGEM3-GC3.1 (Williams et al., 2018; Kuhlbrodt et al., 2018), the UK's previous submission to CMIP6, GC5.0 has undergone significant developments in the atmosphere, land, ocean and sea-ice components, as well as in the coupling between them.

The main developments in the GAL9 component are: a new prognostic-based convective entrainment rate (Willett et al., 2026); deep convection switched off after alignment of mid- and deep-level convection; new riming parameterisation (Furtado and Field, 2017); new bimodal cloud initiation scheme (Weverberg et al., 2021b, a); and several changes targeting improvements in the large-scale circulation (Williams et al., 2020); improved parameterisation of SO_2 dry deposition (Hardacre et al., 2021) and a new representation for marine organic aerosols (Gantt et al., 2011, 2015), detailed in Mulcahy et al. (2023). For the land component, the main changes are: new calculation of surface albedo to improve agreement with observations and remove the previous scaling to climatology, and an update to the soil hydrology parameterisation that improves drainage to lower soil layers.

The ocean component of GOSI9 is based on version 4.0.4 of the NEMO ocean model (Guiavarc'h et al., 2025; Madec et al., 2019). The main differences to the previous configuration are: new equation of state (IOC et al., 2010); adaptive-implicit vertical advection that enables an increase in the model time step; a higher order tracer advection scheme to reduce numerical mixing (Guiavarc'h et al., 2025). The sea ice component of GOSI9 uses SI3 (Vancoppenolle et al., 2023, Sea Ice modelling Integrated Initiative), NEMO's new native sea ice model. Its implementation into HadGEM3-GC5 is described in Blockley et al. (2024).

The lower resolution configuration of HadGEM3-GC5, called HadGEM3-GC5.0-LL, has a horizontal resolution of $1.875°$ longitude by $1.25°$ latitude in the atmosphere and $1°$x$1°$ in the ocean. The atmosphere has 85 levels, and the ocean 75 levels.

2.2 Development of UKCM2-LL

The development of UKCM2-LL made use of a large Perturbed Parameter Ensemble (PPE) of atmosphere-only simulations of GC5.0 variants, which was used to explore the parameter space of the model and identify a set of parameters that would lead to a reduced EffCS. A detailed description of this methodology is presented in Rostron et al. (2026), here we only give an overview to provide context. The approach, based on the methodology by Peatier et al. (2022), consisted of the following steps:

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

1. A 503-member Perturbed Parameter Ensemble (PPE) of the atmospheric and land components (GAL9) was created by perturbing 73 parameters across key parameterization schemes (convection, cloud microphysics, radiation, aerosols, boundary layer, land). The ensemble sampled a wide range of atmospheric feedback strengths, including combinations capable of producing lower EffCS values than the standard configuration. For each PPE member, 5-year simulations of experiments *amip* and *amip-future-4K* were run. These provided model climatological fields for evaluating performance via root mean squared error metrics and estimates of atmospheric radiative feedbacks (λ_{amip}) as proxies for EffCS.

2. Gaussian-Process-based emulators were trained to predict (i) λ_{amip} and (ii) an aggregate AMIP performance error metric for any point in the 73-dimensional parameter space. This enabled rapid exploration of millions of parameter combinations at negligible computational cost.

3. Millions of Monte-Carlo samples were evaluated via emulators, subject to constraints on TOA energy balance, Atlantic water flux (aiming at imposing some constraints on the water cycle), and PPE convex hull coverage (see section 3.1.1 of Rostron et al. (2026)). From these, 41 candidate configurations were selected across a range of predicted λ_{amip} values and then locally optimised to minimise the aggregate error metric while keeping λ_{amip} within prescribed bounds.

4. Full *amip* simulations were run for the 41 candidates to verify emulator predictions. Thirty-four candidates showed predicted EffCS in the target range. Based on *amip* performance and λ_{amip}, 22 candidates were selected for coupled testing.

5. Twenty-two candidates were run for at least 44 years under *piControl* conditions to assess their energy balance, large-scale ocean circulation, sea-ice climatology, and drift behaviour. Six promising candidates were extended to multi-century integrations, and four underwent *abrupt-4xCO2* experiments for EffCS diagnosis.

6. One candidate was chosen as the *proto-UKCM2-LL* due to its relatively low EffCS (4.5 K) and its potential for improvement through parameter tuning.

7. The *proto-UKCM2-LL* underwent further tuning to reduce its TOA imbalance while maintaining acceptable climatological performance. This involved iterative adjustments to key parameters informed by emulator sensitivity analyses during a continuous *piControl-spinup* run. During this process, in addition to the TOA imbalance, the EffCS was further reduced to 3.6 K.

This final configuration was named UKCM2-LL and is the focus of this paper. Only two additional changes have been introduced since the end of the development process: the introduction of CMIP7 forcings, and the update to Gregorian calendar. The update to Gregorian calendar is scientifically neutral, but the introduction of CMIP7 forcings has an impact on both the model's climatology and historical performance as we show below.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

2.3 Simulations

We present results from a subset of DECK simulations from the CMIP6 (Eyring et al., 2016) and CMIP7 (Dunne et al., 2025) protocols:

- *piControl*: pre-industrial control simulation, with greenhouse gas concentrations, aerosols, land use and solar forcing representative of 1850 conditions. The results presented here use 400 years of simulation.
- *abrupt-4xCO2*: instantaneous quadrupling of CO_2 concentrations at the start of the simulation. All the other forcings are the same as in the *piControl* simulation. This simulation is run for 300 years for the UKCM2-LL CMIP7 simulations, and for 150 years for the simulations with CMIP6 forcings.
- *1pctCO2* : 1% per year CO_2 increase simulation, with an increase in CO_2 concentrations of 1% per year (compounded). The simulation is run for 150 years.
- *historical* : historical simulation, with time-evolving greenhouse gas concentrations, anthropogenic aerosol emissions, land use change, ozone, stratospheric aerosol and solar forcings. The simulation is run from 1850 until the end of 2014 for simulations run with CMIP6 forcings, and the end of 2021 for simulations run with CMIP7 forcings.
- *piClim-(x)*: 30 year atmosphere-only timeslice experiment with climatological SSTs and sea-ice derived from the *piControl* coupled atmosphere-ocean simulation. These use fixed forcings set to pre-industrial conditions for *piClim-control*, CO2 concentrations multiplied by 4 for *piClim-4xCO2*, and present day anthropogenic forcing agents for *piClim-anthro* (2014 for CMIP6 and 2021 for CMIP7; Kramer et al., 2026).

Table 1 summarises the models used in this study, and Table 2 summarises the simulations used in this study, including the ensemble size and run length for each experiment. The *piControl* simulation provides the control, unperturbed climate, that serves as baseline for the climate change experiments. The *abrupt-4xCO2* is used to calculate the EffCS and radiative feedbacks, while the *1pctCO2* is used to calculate the TCR, which is the temperature change at the time of CO_2 doubling, calculated as the temperature anomaly averaged between years 61 and 80. The *historical* simulation allows us to evaluate the model's performance against observations over a long historical period, subject to uncertainty due to ensemble size, internal variability and forcings.

Data for GC3.1 were downloaded from the Earth System Grid Federation (ESGF) data portal. The GC5.0 and UKCM2-LL simulations with CMIP6 forcings were run as part of the UKCM2 development project. The second set of UKCM2-LL simulations presented here (with CMIP7 forcings) form the set of simulations that will contribute to the CMIP7 DECK simulations (Dunne et al., 2025).

3 Spin-up and equilibration of the *piControl* simulation

In step 6 of the development phase of UKCM2-LL, one of the promising variants identified in step 5 was chosen as a *proto-UKCM2-LL* configuration. All the step 5 variants had been initialised from EN4 reanalysis (Good et al., 2013) and an atmo-

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. Combination of models and forcing datasets used in this study. LL stands for low-resolution atmosphere (N96) and low-resolution ocean (ORCA1).

Model	Description	Forcings
HadGEM3-GC31-LL	HadGEM3 Global Coupled configuration version 3.1 (Williams et al., 2018; Kuhlbrodt et al., 2018). Submitted to CMIP6.	CMIP6
HadGEM3-GC50-LL	HadGEM3 Global Coupled configuration version 5.0 (Xavier et al., submitted)	CMIP6
UKCM2-LL	UK Climate Model version 2. Same as HadGEM3-GC50-LL, but with different parameter settings (Rostron et al., 2026)	CMIP6
UKCM2-LL C7	Same as UKCM2-LL. Submitted to CMIP7.	CMIP7

spheric state from a previous model configuration. Several parameter changes were applied to the *proto-UKCM2-LL* simulation during the fine-tuning step with the aim of reaching a stable climate with a top-of-atmosphere (TOA) energy balance (Rostron et al., 2026). The fine-tuning was started in year 188 of the *proto-UKCM2-LL*, and it was implemented in 2 steps: the first was run for 86 years, and the second for 944 years. This spin-up period amounted to a total of 1218 years. This stage was performed under CMIP6 pre-industrial forcings, with fixed greenhouse gas concentrations and natural aerosols, and no land-use change.

CMIP7 forcings were introduced in several stages as they became available. First, the solar and volcanic aerosol forcings were introduced. These two forcings were introduced at the same time because the expectation was that they would provide small forcing changes in opposite directions with respect to the CMIP6 forcings, therefore minimising the change in absolute forcing. This phase of the spin-up with CMIP7 solar and volcanic aerosol forcings was run for 60 years. Then, land-use, aerosol and greenhouse forcings were updated to CMIP7 forcings and the spin-up was continued for 233 years. Finally, the ozone forcing was updated. The introduction of the ozone forcing did not require a further spin-up, as the model remained close to equilibrium. In total, the model was spun up for nearly 300 years with a combination of CMIP6 and CMIP7 piControl forcings, after the initial spin-up of more than 1200 years during the fine-tuning phase. Here we present results from the first 400 years of the *piControl* simulation of UKCM2-LL with CMIP7 forcings, i.e. after more than 1500 years of spin-up. This is the simulation that will be made public as part of the CMIP7 DECK simulations.

Table 3 summarises the global means of some key variables during the 400 years of the *piControl* simulation presented here, along with the centennial linear trends over the same period. The variables included in the table: near-surface air temperature (*tas*); net radiation at the top of the atmosphere (*rnt*, positive downwards); net shortwave radiation at the top of the atmosphere (*rsnt*); and outgoing longwave radiation at the top of the atmosphere (*rlut*). These statistics are presented for the *piControl* simulation of UKCM2-LL, and for the two other GC configurations, GC3.1 and GC5.0. The values of these variables for the *piControl* simulation of GC3.1 are included as a reference, as this was the previous low-resolution configuration of the

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. Experiments used in this study. For UKCM2-LL we use two sets of simulations, one with CMIP6 forcings that were used during the development of the model, and a second set with CMIP7 forcings that will contribute to CMIP7 DECK simulations.

Model	Experiment name	Ensemble size	Run length (years)
HadGEM3-GC31-LL	piControl	1	400
	abrupt-4xCO2	1	150
	historical	5	165
	1pctCO2	1	150
	piClim-control	1	30
	piClim-4xCO2	1	30
	piClim-anthro	1	30
HadGEM3-GC50-LL	piControl	1	400
	abrupt-4xCO2	1	150
	historical	3	165
	piClim-control	1	30
	piClim-4xCO2	1	30
UKCM2-LL	piControl	1	400
	abrupt-4xCO2	1	150
	historical	3	165
	1pctCO2	1	150
UKCM2-LL CMIP7	piControl	1	400
	abrupt-4xCO2	1	300
	historical	5	172
	1pctCO2	2	150
	piClim-control	1	30
	piClim-4xCO2	1	30
	piClim-anthro	1	30

Hadley Centre model submitted to CMIP6. This provides context for changes introduced in GC5.0 and UKCM2-LL. It is worth noting the *piControl* simulation of GC5.0 was spun up for a much shorter period than the other configurations, and therefore the values of these variables for GC5.0 may not represent the same stage of equilibration as for the other models. Also, as mentioned above, the spin up process of UKCM2-LL was constrained by IPCC timescales and several pragmatic decisions were made during the implementation of CMIP7 forcings as mentioned above. The ranges of these variables in an ensemble of 36 CMIP6 *piControl* simulations are also presented for context.

The variables show small trends over the 400 years of simulation, indicating that the model's surface state is close to steady state, and their mean values are generally close to present-day observed values. The near-surface air temperature (*tas*) has

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 3. Global mean values and trends of key variables for the *piControl* simulations of the following models: GC3.1, GC5.0, UKCM2-LL, UKCM2-LL C7 and an ensemble of 36 CMIP6 *piControl* simulations. For the CMIP6 ensemble, the standard deviations are also included in parentheses. Units are K for *tas* and W m^{-2} for the other variables. The trends are calculated as the linear trend over the 400 years of simulation and reported as W m^{-2} per century for the radiation variables and K per century for *tas*.

	GC3.1		GC5.0		UKCM2		UKCM2 C7		CMIP6	
	Mean	Trend	Mean	Trend	Mean	Trend	Mean	Trend	Mean	Trend
tas	286.9	0.03	285.9	0.06	286.1	0.07	286.2	-0.05	286.7 (0.71)	0.02 (0.07)
rnt	0.19	-0.01	0.10	0.00	0.08	0.02	0.12	0.00	0.76 (1.01)	0.00 (0.03)
rsnt	241.6	0.07	239.4	0.13	241.1	0.19	241.7	-0.10	239.6 (3.44)	0.04 (0.12)
rlut	241.4	0.08	239.3	0.13	241.0	0.16	241.6	-0.10	238.8 (3.29)	0.04 (0.14)

a mean value of 286.2 K, which is between 0.5 to 1 K colder than the estimated global-mean temperature in pre-industrial conditions (c3s, 2024). The TOA net radiation (*rnt*) has a mean value of 0.12 Wm^{-2}. This value is both within the variability in the MME and within the 0.15 Wm^{-2} target for shorter simulations during the development phase (Rostron et al., 2026). However, it is larger than the more stringent target we were aiming for longer control simulations (0.05 Wm^{-2}) but smaller than the TOA imbalance in GC3.1.

The implementation of CMIP7 forcings had a small, but not insignificant, impact on the TOA imbalance, increasing it by $0.04Wm^{-2}$. This small positive energy TOA imbalance is reflected in a linear increase of ocean heat content (OHC) with time as we shall see below. This constant accumulation of energy may have implications for the long-term stability of the control climate, as shown by Ridley et al. (2022). We did not attempt to tune away this extra energy imbalance due to the time constraints of the CMIP7 DECK simulations.

The net shortwave radiation at the top of the atmosphere (*rsnt*) has a mean value of 241.7 Wm^{-2}, and the outgoing longwave radiation at the top of the atmosphere (*rlut*) has a mean value of 241.6 Wm^{-2}. Both *rsnt* and *rlut* show small positive trends over the 400 years. Although these trends nearly cancel each other to give a near-zero trend in *rnt*, they are indicative that the model is still progressing towards equilibrium. The values of these variables and their trends are also consistent with the results from the spin-up phase of UKCM2-LL (not shown), giving confidence that upgrading to the Gregorian calendar and to a new supercomputer had a negligible impact. In general, the mean values and trends are within one standard deviation of those found in the CMIP6 MME, indicating that the global-mean climate of the model is not unusual compared to other models.

Figure 1 shows the time series of the global means of the variables presented in Table 3. All configurations show substantial decadal and multidecadal variability. All GC5-based models show larger variability than GC3.1, with GC5.0 showing a larger variability than the UKCM2 simulations. This larger variability in GC5.0 is likely due to the different physics, but we cannot rule out the possibility that it is related to the fact that the *piControl* simulation of GC5.0 was not spun up for as long as the other simulations because it is not part of the UK contribution to CMIP7. Interestingly, the larger variability in GC5.0 in *rsnt*

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

and *rlut* tend to cancel each other, leading to a similar or even smaller variability in *rnt* than in the other two configurations. The plots also show the target ranges that were used during the development of UKCM2-LL. For UKCM2-LL C7, *rsnt* and *rlut* lie outside the target ranges by ≈0.5 W m^{-2}. These targets were based on 2005-2015 CERES EBAF observations (Loeb et al., 2018), combined with early estimates to give some uncertainty bounds (L'Ecuyer et al., 2015), and UKESM1.1-simulated historical anomalies from Mulcahy et al. (2023) to obtain the pre-industrial estimates. Given the uncertainty in the pre-industrial estimates, these target ranges were not imposed as hard constraints. The full list of target ranges used during development are presented in Table 2 of Rostron et al. (2026). Also, it was difficult to tune these whilst also getting global-mean *tas* right, so a trade-off was required (Rostron et al., 2026).

All models show a growing ocean heat content anomaly (Figure 2). The OHC growth rates are consistent with the average positive energy imbalances at the top of the atmosphere shown in Table 3: GC3.1 shows the fastest growth in OHC anomaly, followed by UKCM2-LL with CMIP7 forcings, GC5.0 and UKCM2-LL with CMIP6 forcings. However, the distribution of the heat anomaly in the ocean is markedly different between the models. GC3.1 and UKCM2-LL C7 exhibit growing heat content in all ocean layers. GC5.0 exports energy from the top 700 m to the layer below (700-2000m) during the first 200 years, switching to the opposite behavior in the last 200 years. UKCM2-LL with CMIP6 forcings behaves significantly differently, with a smaller total heat content trend but notably with the largest trend below 2000m, which is largely due to energy transfer from the upper 2000 metres. This will steadily increase instability of the vertical column that could lead to sudden changes in the climatological state later in the simulation (Ridley et al., 2022). The North Atlantic is the main driver of the negative trend in the intermediate layer in UKCM2-LL. Figure 3 shows the time-averaged and zonally-averaged state of the MOC for both UKCM2 simulations for the period between years 150 and 159 of the time series shown in Figure 2. The simulation with CMIP6 forcings has a stronger upper overturning cell, which is consistent with a heat loss from the upper 2000 m to the deeper ocean in the North Atlantic. These differences are likely due to the different stages in the equilibration of the two simulations, with the CMIP6-forced simulation being affected by the changes that were made during the development of UKCM2-LL (Rostron et al., 2026).

The fact that the CMIP7-forced UKCM2-LL simulation shows a positive heat uptake anomaly in the upper 700m, while having a small global-mean *tas* negative trend, is an indication that heat is being redistributed within that upper layer. The different characteristics of vertical redistribution of heat in the different configurations highlights that, despite the relatively small trends in surface temperature, the models are still not in equilibrium and they are following different trajectories towards equilibrium.

Figure 4 shows the time series of the maximum Atlantic Meridional Overturning Circulation (AMOC) strength at 26°N and the Antarctic Circumpolar Current (ACC) through the Drake Passage. The AMOC is well below the observational time-average of 17 Sv and annual-mean standard deviation of 1.3 Sv (Moat et al., 2026; Smeed et al., 2018). We used a very conservative target range of 11-19 Sv during the development, as GC5.0 AMOC strength was already quite weak. UKCM2-LL C7 tends to show a slightly weaker AMOC than both GC5.0 and UKCM2-LL C6 forcings; however, these differences are small and may be well within the models' variability as they slowly evolve towards a steady state. In most CMIP6 models, AMOC strengthens by approximately 10% during the historical period until the end of the 20th Century, mainly due to aerosol cooling in the NH

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

(Menary et al., 2020), and then declines. Therefore, the comparison of model AMOC strength in the pre-industrial climate and with present-day observations is subject to a large uncertainty.

UKCM2 shows a clearly weaker ACC than GC5.0. Estimates of the ACC strength range between 140 and 175 Sv (Xu et al., 2020). Similar to AMOC, we used a wide target range during development: 110-190 Sv. GC5.0 shows a mean ACC strength just outside the lower end of this range, with UKCM2-LL C6 and C7 showing a substantially weaker ACC strength than GC5.0, in the range between 110 and 120 Sv. This weakening of the ACC is probably due to the effect of the cloud tunings that have weakened the SW CRE in the Southern Ocean. A weaker SW CRE leads to warmer SSTs, decreasing both the atmospheric temperature and ocean density gradients between the lower and higher latitudes. A weaker temperature gradient leads to weaker winds, which, together with the weaker meridional density gradient, weaken the ACC. The weakening of the ACC in UKCM2-LL itself may be contributing to a warmer Southern Ocean: with a weak ACC, the slumping of the isopycnals will favour the transport of warm water southward (Storkey et al., 2025), reinforcing the warm SST bias.

Figure 5 shows the time series of the global-mean sea ice fraction and volume. As discussed in Rostron et al. (2026), the target ranges for annual-mean sea-ice area were 11.5-13.0 million km^2 in the Arctic and 12.0-16.0 million km^2 in the Antarctic. These ranges were informed by present-day observations through the HadISST2.2.0.0 dataset (Titchner and Rayner, 2014), with pre-industrial corrections applied using expert judgement. In the Arctic, all GC5-based models are at the high end of that range, with UKCM2 C7 and GC5.0 being outside that range about half of the time. In the Antarctic, only GC5.0 falls within the target range, with all the other models being biased low.

CMIP7 forcings lead to larger Arctic sea ice fraction and volume, mainly due to the larger stratospheric volcanic aerosol loading in the Northern Hemisphere (NH) with some contribution from the changes in the solar forcing. In the Antarctic, the CMIP7-forced simulation has less sea-ice area and volume than the CMIP6-forced simulation that was used during development. However, this difference is probably more related to the CMIP6-forced simulation not being spun up for longer. An extension of that simulation shows a warming trend in the Southern Hemisphere extratropics, increasing its near-surface temperature by 0.5 K in 700 years. That additional warming leads to a reduction in the Antarctic sea ice fraction and thickness that explains most of the differences in Antarctic sea ice between the two UKCM2 simulations shown in Figure 5. It may have been possible to tune the model to reduce some of the impacts of the new set of forcings, but given the time constraints of the CMIP7 protocol, we have decided to keep the same parameters and accept these impacts due to the introduction of CMIP7 forcings. On the other hand, this provides an opportunity to assess the impact of the uncertainty in pre-industrial and historical forcings on UKCM2's climatology and response to forcing, which we will explore in future work.

UKCM2 shows a clear multidecadal variability in the Antarctic sea ice fraction and thickness, caused by polynyas in the Weddell Sea (not shown). Larger multidecadal variability associated with polynyas is also present in GC3.1 during an extended piControl simulation (Ridley et al., 2022). Ridley et al. (2022) hypothesise that they are a consequence of a positive, downward, top-of-atmosphere radiative balance. This continual ocean heat uptake during the model spin-up and the initial pre-industrial control simulation destabilises the ocean column, leading to these episodes of deep convection. It is not clear whether this large multidecadal variability will decrease as the model approaches equilibrium over long timescales or whether this large multidecadal variability is a property of the model's equilibrium state.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

4 Climate sensitivity, feedbacks and forcing

We calculate the EffCS using the method of Gregory et al. (2004), where a linear regression of the globally-averaged annual mean TOA net radiation anomalies against globally-averaged annual mean *tas* anomalies is performed for the first 150 years after quadrupling the CO_2 concentration in the *abrupt-4xCO2* simulation. The slope of the regression corresponds to the feedback parameter λ_{4xCO2}, while the intercept corresponds to the 4xCO2 effective radiative forcing (ERF), F_{4xCO2}. The EffCS is calculated as $-0.5F_{4xCO2}/\lambda_{4xCO2}$. The EffCS of UKCM2-LL with CMIP6 forcings is 3.6 K, which is within the likely range of the IPCC AR6 and much lower than the EffCS of GC5.0 (6.6 K). The set of forcings doesn't affect the EffCS of UKCM2-LL, which is 3.6 K with CMIP7 forcings. When the *piClim-(x)* simulations are used, the ERF is calculated as the 30 year time-mean difference in net TOA radiative flux between *piClim-4xCO2* or *piClim-anthro* and *piClim-control*.

Table 4 summarises the values of the 4xCO_2 ERF and the feedback parameters for a range of TOA radiative fluxes for GC5.0 and UKCM2-LL, along with the ranges in CMIP6 models. The reduction in EffCS in UKCM2-LL with respect to GC5.0 is mainly due to a less positive SW Cloud Radiative Effect (CRE) feedback (λ_{SWCRE}). The CRE is the difference between clear-sky and all-sky fluxes at the TOA, positive downwards (Ramanathan et al., 1989). The other feedback parameters show smaller differences between the two models, with some being slightly more positive in UKCM2-LL and others being slightly more negative. The ERF of the GC5-based models is generally lower than that of GC3.1. UKCM2-LL configurations have smaller ERF than GC5.0, however this seems to be driven by the change in the feedbacks rather than changes in the instantaneous radiative forcing given that the ERF calculated from the first 5 years after quadrupling CO_2 is slightly larger in the UKCM2-LL configurations. The values of the feedback parameters for UKCM2-LL are within the CMIP6 ranges. The ERF of UKCM2-LL is also within the range of CMIP6 models, but on the lower end of the range. The lower ERF and more negative λ_{SWCRE} in UKCM2-LL compared to GC5.0 contribute to the lower EffCS in UKCM2-LL.

In general, the values for the UKCM2-LL simulations with CMIP7 forcings show small differences with respect to those with CMIP6 forcings. The TCR estimates are subject to a non-negligible uncertainty due to both unforced variability and the effect of the subtraction of the *piControl* linear trend. For instance, the TCR of both ensemble members of UKCM2-LL C7 is 2.0K and 2.2K, and the difference in the *piControl* linear trends between the UKCM2-LL simulations with CMIP6 and CMIP7 forcings leads to a difference in the TCR of 0.1K, which is a significant fraction of the difference between the two TCR estimates. In any case, the TCR estimates for both UKCM2-LL simulations are within the likely range of the IPCC AR6 (1.4-2.2K). No *1pctCO2* experiment was performed for GC5.0, but we follow the method in Gregory et al. (2015) to provide an estimate of TCR using the *abrupt-4xCO2* experiment. The approximate TCR values using this method are shown for all models in column TCR*. The TCR* estimates are generally close to the TCR estimates from the *1pctCO2* experiments, which gives confidence that the TCR* estimate of 2.7K for GC5.0 should be close to the real value. This value is much higher than the TCR of UKCM2-LL and outside of the IPCC AR6 very likely range (1.2-2.4K), consistent with the differences in EffCS. A recent study by Gyuleva et al. (2026) proposes a revised TCR very likely range of 1.9-2.6 K, placing GC5.0 just outside of the upper end of that range.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

The UKCM2-LL *abrupt-4xCO2* experiment was run for 300 years, as requested by the CMIP7 protocol, and Table 4 includes the values of the 300-year regressions. After 300 years, the EffCS increases to 3.9K, which is due to a less negative λ_{4xCO2}, while the ERF decreases. This is an indication of a non-linearity in the Gregory plot due to a change in the relationship between the TOA net radiation and the global-mean *tas* anomalies on longer timescales. This behaviour is very common across models (Rugenstein et al., 2020), and it can be observed in Figure 6, which shows the scatter plots and regressions lines for the four model configurations analysed here. Each subplot contains the regression line over the first 150 years, with the intercept (ERF) and slope λ indicated in the legend. The change in the slope of the relationship between the radiative fluxes and temperature anomalies after 150 years is evident in UKCM2 C7 plots, which show years 151 to 300 as grey dots. If the regression is performed over years 151 to 300 ($EffCS_{151-300}$), the EffCS increases to 4.3K. This change is driven by the SW CRE (Figure 6g4), mainly due to changes in the SW cloud feedbacks on longer timescales. The ratio between $EffCS_{151-300}$ and $EffCS_{1-150}$ is 1.19. This is very close to 1.17, the median of the ratio of the best estimate of the long-term ECS and $EffCS_{1-150}$ in the multimodel ensemble used by Rugenstein et al. (2020). Therefore, 4.3K is probably a better estimate of the true model ECS.

Figure 7 shows a scatter plot of the EffCS against the global feedback parameter (λ_{4xCO2}) for a range of CMIP6 models, with the UK models analysed here highlighted in colour. As expected, the plot shows a clear relationship between the EffCS and λ_{4xCO2}, with higher EffCS corresponding to higher (less negative) λ_{4xCO2}. The EffCS of UKCM2-LL is within the range of CMIP6 models, while the EffCS of GC5.0 is much higher than any CMIP6 model. The lower EffCS of UKCM2-LL with respect to GC5.0 is associated with a more negative λ_{4xCO2}, from -0.47 to -0.79 W m^{-2}. A smaller positive SW cloud feedback (λ_{SWcre}), from 0.86 to 0.65 W m^{-2}, explains two thirds of the difference in the total feedback. Both models fall below the curve defined by the ERF multi-model mean, implying that they have a lower ERF than the multi-model mean. This plot shows that, irrespective of their EffCS, both GC5.0 and UKCM2-LL behave similarly to other CMIP6 models in terms of the relationship between *EffCS* and feedback parameter.

Figure 8 shows the differences in λ_{SWcre} feedback between UKCM2-LL and GC5.0. The regional λ_{SWcre} is calculated as the slope of the regression of anomalies in local TOA net radiation and global-mean *tas*. The smaller λ_{SWcre} is achieved by a decrease in the feedback in the midlatitudes. The Southern Hemisphere (SH) subtropics show a reduction in the positive feedback, whereas the NH subtropics show an increase in the positive SW feedback. The differences in the midlatitudes are more zonally coherent, especially in the Southern Ocean, whereas the subtropics show compensation between regions with stronger and weaker regional feedbacks than GC5.0. The pattern of differences of λ_{TOTcre} (not shown) is very similar to that in Figure 8, except for a muted difference in the deep convective clouds over the Tropical Western Pacific, where λ_{LWcre} changes in the opposite direction to λ_{SWcre}, leading to a compensation in λ_{TOTcre}. The Arctic shows large differences in λ_{SWcre} feedback between UKCM2-LL and GC5.0, but SWCRE is strongly affected by cloud-masking in those regions Soden et al. (2004, 2008), so it is likely that those changes are dominated by unforced variability of sea-ice.

It is worth noting the differences between UKCM2-LL regional λ_{SWcre} with CMIP6 and CMIP7 forcings. The main systematic differences occur in the lower midlatitudes. With CMIP7 forcings, λ_{SWcre} is weaker (less positive) in the SH lower midlatitudes, whereas a similar weakening is observed in the NH lower midlatitudes with CMIP6 forcings. These differences

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Table 4. Effective radiative forcing, feedbacks, TCR and EffCS for all the models analysed in this study. Feedbacks and ERF are calculated using a linear regression over the first 150 years after quadrupling the CO_2 concentration. ERF (5yr) is calculated using only the first 5 years after quadrupling CO_2, which is more representative of the instantaneous radiative forcing plus adjustments (Andrews et al., 2019). The TCR is calculated as the average of the GMSAT between years 60 and 80 of the 1pctCO2 experiment and the *piControl* simulation. For UKCM2-LL with CMIP7 forcings, we also include the values of these variables when the regression is performed over the first 300 years after quadrupling the CO_2 concentration. Since no *1pctCO2* experiment was performed for GC5.0, the TCR value of GC5.0 is estimated using the *abrupt-4xCO2* experiment following the method in Gregory et al. (2015). The approximate TCR values using this method are shown for all models in column TCR*. The units are: Wm^{-2} for ERFs, W m^{-2} K^{-1} for the feedback parameters, and K for TCR and EffCS.

Metric	GC3.1	GC5.0	UKCM2 C6	UKCM2 C7	UKCM2 C7 (300 yr)	CMIP6 mean (std)
F_{2xCO2}	3.44	3.10	2.90	2.85	2.51	3.54 (0.39)
F^{5y}_{2xCO2}	4.01	3.51	3.74	3.92	3.92	4.14 (0.50)
EffCS	5.50	6.61	3.58	3.58	3.90	3.80 (1.06)
$EffCS_{51-150}$	5.62	6.93	3.80	3.78	3.78	4.07 (1.22)
TCR	2.60	n/a	1.90	2.10	2.10	2.05 (0.41)
TCR*	2.60	2.70	2.10	1.90	1.90	2.09 (0.36)
λ_{4xCO2}	-0.63	-0.47	-0.81	-0.79	-0.64	-1.02 (0.34)
λ_{SW}	1.63	1.68	1.41	1.42	1.55	0.77 (0.51)
λ_{LW}	-2.26	-2.15	-2.22	-2.22	-2.20	-1.78 (0.31)
λ_{SWcs}	0.66	0.82	0.76	0.77	0.81	0.65 (0.23)
λ_{LWcs}	-1.81	-1.78	-1.85	-1.84	-1.82	-1.79 (0.12)
λ_{TOTcre}	0.52	0.49	0.28	0.28	0.37	0.12 (0.42)
λ_{SWcre}	0.97	0.86	0.65	0.65	0.74	0.11 (0.59)
λ_{LWcre}	-0.45	-0.37	-0.37	-0.37	-0.37	0.01 (0.26)

may be related to changes in the *piControl* cloud climatology due to the differences in the CMIP7 pre-industrial forcings, which lead to colder NH extratropics and warmer SH extratropics with CMIP7 forcings, as discussed above. They could also be driven by patterns of multidecadal unforced variability. The largest differences in the UKCM2 simulations with different forcings occur in high latitudes in both hemispheres (e.g. around 70°S or poleward of 75°N). However, these differences are more easily explained by the differences in the sea-ice climatology between the two pre-industrial states, which affect the SW CRE feedback through a different SW clear-sky feedback. With CMIP7 forcings, λ_{SWcs} is weaker in the Antarctic sea ice region, likely due to the smaller amount of sea ice in the pre-industrial state. Interestingly, all these regional differences compensate each other in the global mean, leading to a very similar global-mean λ_{SWcre} for both UKCM2 simulations with CMIP6 and CMIP7 forcings.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Figure 9 shows the SW CRE climatologies for the *piControl* simulations of the GC5.0 and UKCM2-LL C7 models. The large differences in SW CRE climatologies between the two configurations can probably only explain part of the differences in λ_{SWcre} (Figure 8e). While there are some regions where the differences in SW CRE climatology and feedbacks are anticorrelated (e.g. NH land, NH subtropical Eastern Pacific), suggesting a causal relationship between the changes in the climatology and the feedbacks, there are other regions where the differences in SW CRE climatology and feedbacks are positively correlated (e.g. the SH subtropical Eastern Pacific). This suggests that the differences in SW CRE feedbacks between GC5.0 and UKCM2-LL are probably due to a combination of the impact of the changes in the cloud climatology and changes in the response of clouds to warming. A more detailed analysis of the differences in cloud feedbacks between the two models will be documented elsewhere (Ringer et al., in preparation).

Alongside the *abrupt-4xCO2* experiment, the ERF_{2xCO2} can also be calculated using the *piClim-4xCO2* simulations. These simulations give the values of 4.04, 3.75, and 3.59 W m^{-2} for GC3.1, GC5.0, and UKCM2 C7, respectively. This method removes the uncertainty in ERF associated with feedbacks when using the regression method (used in Table 4) since the climatological SSTs in the piClim experiments reduce changes in TOA radiation associated with changes in global surface temperature. The ERF estimated from the *piClim-4xCO2* experiment demonstrates a reduction from GC3.1 to GC5.0 and a further reduction in UKCM2 C7. Although the ERF was not targeted in the UKCM2 model development, the reduced ERF resulting from tuning changes will have contributed to the lower climate sensitivity.

5 Simulation of the historical period

Figure 10 shows that both UKCM2-LL and GC5.0, when forced with CMIP6 forcings, reproduce the observed anomalies in near-surface air temperature more accurately than GC3.1-LL, both for the global-mean and for the interhemispheric difference. This improvement is mainly driven by changes in the aerosol forcing introduced in GC5.0 due to a better representation of SO_2 deposition (Mulcahy et al., 2023). UKCM2-LL also shows a better match to the observed interhemispheric temperature anomalies (Figure 10b). The legend in Figure 10a shows the global-mean temperature trends for the last 30 years of the CMIP6 historical period (1985-2014). The trends correlate with the EffCS of the models, with UKCM2-LL showing a 25% reduction in the trend with respect to GC5.0 , reducing the bias with respect to the observed trend. The reduction in the trend may not be entirely due to the lower EffCS, as differences in the pattern effect between the two configurations may also play a role. The pattern effect is the impact of the spatial pattern of warming on the global-mean feedback parameter, which can lead to differences in the global-mean temperature response to forcing (e.g. Senior and Mitchell, 2000; Gregory and Andrews, 2016; Armour, 2017; Andrews et al., 2018; Dong et al., 2021). The differences in the pattern effect between UKCM2-LL and GC5.0 will be explored in future studies, when larger historical ensembles become available.

The introduction of CMIP7 forcings seems to have a significant impact on the simulation of the historical period, degrading the comparison of the temperature anomalies against observations. The main differences in anomaly space happen after 1950, when the CMIP7-forced simulations show a stronger global-mean warming, and a substantially larger interhemispheric temperature anomaly. Similar anomaly plots for the NH and SH separately (not shown) show that the differences in the inter-

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

hemispheric temperature anomaly are mainly due to a stronger warming in the NH in the CMIP7-forced simulations. There is also a contribution from the Southern Hemisphere, which shows a weaker warming in the CMIP7-forced simulations. Both the larger warming in the NH and a smaller warming in the SH contribute to the larger interhemispheric temperature anomaly in the CMIP7-forced simulations. Of the ≈0.7 K difference in the interhemispheric temperature anomaly between the CMIP6- and CMIP7-forced simulations, 0.5 K is due to the NH and 0.2 K is due to the SH.

The time series of NH absolute temperature (Figure 11) shows that the main difference between the CMIP6- and CMIP7-forced runs occurs before 1900. This is partly due to the fact that the CMIP7 forcings have a larger volcanic aerosol loading in the pre-industrial period, which leads to a colder pre-industrial NH, but also because of the larger volcanic activity in the early part of the historical period in CMIP7 forcings, which leads to an initial cooling in the CMIP7-forced historical simulations. In the first decades of the historical period, the CMIP7 forcings show more active volcanism (Aubry et al., 2026b), which introduces a differential negative forcing at the beginning of the CMIP7-forced historical simulations. The imprint of the larger volcanic activity in the pre-1900 period in CMIP7-forced simulations can be clearly seen in the clear-sky outgoing SW radiation (Figure 12). In Figure 10, we use the 1850-1900 period as the baseline period for the anomalies, which makes the differences at the beginning of the historical simulations appear post-1900 in anomaly space. Also, the CMIP7-forced simulations warm more in the later part of the historical period, contributing to the larger interhemispheric anomaly post-1950. This is also likely related to differences between the aerosol forcing datasets. Some of the tropospheric aerosol species show a smaller growth in emissions over the historical period in the CMIP7 forcings dataset, with larger emissions in the pre-industrial period and smaller emissions after ≈1950. These differences are consistent with the larger warming in the NH in the CMIP7-forced simulations.

It is worth noting that the uncertainty in the forcings is significant. The CMIP7 volcanic aerosols are likely biased in terms of the latitudinal distribution of small-to-moderate magnitude eruptions pre-satellite, with insufficient tropical and southern hemisphere injections (Aubry et al., 2026a). Also, the updates of the regional patterns of emissions of tropospheric aerosols in CMIP7 may also play a role in the differences between the CMIP6- and CMIP7-forced simulations (Samset et al., 2024). The ERF_{anthro} shows a substantial increase between GC3.1 and UKCM2 C7, from 1.81 to 2.36 W m^{-2}. It is unclear whether this is caused by changes in the model, forcing dataset, or time period used (i.e. 2014 for GC3.1 and 2021 for UKCM2). Unforced variability cannot be ruled out as a potential contributor to the differences, especially given the large multidecadal variability seen in the GC5-based configurations. The explanations presented above are initial hypotheses, the drivers of the differences in the CMIP6- and CMIP7-forced historical simulations will be explored in more detail in future work.

Figure 13 shows the time series of the AMOC and ACC strength for the historical simulations. Generally, the AMOC strengthens during the historical period, with a peak around 1975, followed by a weakening. UKCM2-LL C7 seems to be an exception to this general behaviour, as it does not show a weakening of the AMOC after 1975. UKCM2-LL C7 has a weaker AMOC than the other configurations throughout most of the historical period, which is consistent with the weaker AMOC in the pre-industrial state, but by the end of the historical period, the AMOC in UKCM2-LL C7 is similar to that of the other configurations. UKCM2-LL C7 also shows a larger multidecadal variability in the AMOC than the other configurations, consistent with the larger variability in the pre-industrial state discussed above. The AMOC strength is biased low in all configurations

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

compared to the observational estimates, as shown by the grey area (17±1.3 Sv). The ACC is weak in all configurations, well below the observational range. UKCM2 shows a weaker ACC than GC3.1 and GC5.0, with the configuration with CMIP7 forcings showing the weakest ACC. The ACC in GC5.0 shows a slight weakening during the historical period whereas it remains stable in UKCM2-LL, although these differences in historical evolution may be dominated by internal centennial variability of the ACC (4).

Figure 14 shows the simulation of sea-ice area and volume. In the Arctic, UKCM2-LL C7 has a larger sea-ice area and volume than UKCM2-LL at the beginning of the historical period, but they converge to similar values after 1975. Arctic sea ice declines at a faster rate in UKCM2-LL than in GC5.0 during most of the historical period, but the opposite is true after ≈1980. The main differences between UKCM2-LL and UKCM2-LL C7 are in the Antarctic, where UKCM2-LL C7 shows substantially smaller area and volume than UKCM2-LL, as a result of the warmer Southern Ocean in the control simulation as discussed above. Compared to observations, UKCM2-LL C7 has a high sea ice area bias (±10%) in the Arctic, although that bias becomes smaller at the end of the historical simulation due to the faster area decline in the model. In the Antarctic, UKCM2-LL C7 shows a low sea ice area bias compared to OSI SAF. Antarctic sea ice variability seems to be higher in UKCM2-LL than in GC5.0, and it is certainly larger than in GC3.1. The low mean state and high variability for Antarctic sea ice will be related to the prevalence of polynyas in the Southern Ocean as discussed above. Similar behaviour is found in GC3.1 during an extended piControl simulation (Ridley et al., 2022), and Jones et al. (2024) reported unusual temperature variability over the Southern Ocean in GC3.1 related to occasional deep convection bringing heat to the surface. In particular, Jones et al. (2024) noted that it is most prominent in the model's aerosol-only simulations because the cooling from the negative forcing destabilises the water column in the Southern Ocean, exciting deep convection and bringing up relatively warmer waters.

6 UKCM2-LL in the context of the CMIP6 ensemble

The tuning process used to develop UKCM2-LL was designed to find a balance between reducing the EffCS and maintaining an acceptable climatological performance. As mentioned in Section 2, the parameter changes that led to reduction in EffCS in UKCM2-LL degraded the climatological performance as measured by the climatological root mean square error (RMSE) of the amip simulations used in the optimisation process. The relationship of this general degradation in RMSEs with structural uncertainties in the model is discussed in more detail in Rostron et al. (2026). This degradation of the climatological performance in the amip simulations carried over to the coupled simulations. Figure 15 shows a heatmap of annual-mean RMSE for key climate variables averaged over 1981–2000. The values are normalised by the RMSE of HadGEM3-GC31-MM, which shows a value of 1.0 for all variables. Values smaller than 1 indicate better agreement with observations. The heatmap shows that UKCM2-LL has a higher RMSE than GC5.0 for all the variables tested, with an average degradation of 30%. However, it is important to note that UKCM2-LL still performs reasonably well compared to the range of CMIP6 models included in the heatmap. For instance, UKCM2-LL C7 has an average normalised RMSE across all variables of 1.24, which ranks 28th out of the 67 models shown in Figure 15, is below the multi-model average of 1.35.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

The main aim of the UKCM2-LL development project was to deliver a GC5-based configuration with an EffCS within the IPCC AR6 very likely range (2.0 to 5.0K), with the IPCC AR6 likely range (2.5 to 4.0K) as a stronger constraint. The structural constraints mentioned above imply that it should be possible to build a configuration with an EffCS within the very likely range and smaller degradation in climatological performance. However, the experimental design used during the development of UKCM2-LL did not provide such a level of precision in the outcome due to the impact of the manual tuning on the final EffCS (Rostron et al., 2026).

7 Conclusions

This study documents the CMIP7 DECK simulations of UKCM2-LL, a climate model configuration based on HadGEM3 GC5.0 (Xavier et al., submitted) with constrained climate sensitivity within the IPCC AR6 likely range (2.5–4.0K). This model configuration has been developed using a perturbed parameter ensemble approach, applied to atmosphere-only simulations, where the parameter space is systematically explored, identifying configurations that reduce the EffCS from 6.6K in GC5.0 to 3.6K in UKCM2-LL, primarily through a smaller positive cloud feedback. The new configuration represents a physically plausible solution that balances the requirement for reduced climate sensitivity with acceptable present-day climatological performance. The methodology used to develop UKCM2-LL is documented in Rostron et al. (2026). We present results from the CMIP7 DECK simulations of UKCM2-LL, and assess them against observations, other model configurations (GC3.1 and GC5.0), and the CMIP6 multi-model ensemble. We focus on the spin-up and equilibration of the pre-industrial control simulation, EffCS, forcings and feedbacks, and the simulation of the historical period.

The reduction in EffCS from 6.6 K in GC5.0 to 3.6 K in UKCM2-LL is primarily driven by changes in cloud feedbacks. The shortwave cloud radiative effect feedback (λ_{SWCRE}) becomes less positive in UKCM2-LL. These feedback changes reflect the parameter adjustments made during the tuning process and represent the physical mechanism through which the perturbed parameter ensemble methodology achieved the target climate sensitivity while maintaining radiative balance in the pre-industrial control simulation. A smaller effective radiative forcing also contributes to the reduction in EffCS.

Evaluation of UKCM2-LL against observations and other models demonstrates that the lower climate sensitivity translates to improved simulation of the historical temperature record since the 1970s, with better agreement with observed interhemispheric temperature anomalies compared to GC5.0 when CMIP6 forcings are used. CMIP7 forcings lead to a stronger warming in the historical period, especially in the Northern Hemisphere, which degrades the comparison with observations. The differences between CMIP6 and CMIP7 forcings are mainly due to differences in volcanic forcing, with CMIP7 showing more active volcanism before 1900, leading to a stronger negative forcing at the beginning of the historical simulations.

While the model exhibits some degradation in climatological metrics (approximately 30% higher RMSE), the performance remains reasonable relative to CMIP6 models and reflects the inherent trade-off between constraining emergent properties and maintaining observational fidelity. The pre-industrial climate is stable with acceptable top-of-atmosphere energy balance and physically plausible ocean circulation characteristics.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

The development of UKCM2-LL demonstrates the efficacy of top-down constraint methodologies in developing climate models that satisfy requirements on emergent properties like climate sensitivity while maintaining reasonable physical consistency. This work provides a valuable tool for climate projection studies that require models within the IPCC-assessed climate sensitivity range and offers insights into the relationship between parametric uncertainty and model performance. Future applications of UKCM2-LL will benefit studies of carbon-cycle feedbacks and other processes where the choice of climate sensitivity significantly influences results.

In order to constrain emergent properties like EffCS in a bottom-up approach to development, there is a need to develop further observational constraints. Potential avenues are: cloud controlling factors (CCFs; Scott et al., 2020; Ceppi and Nowack, 2021; Furtado et al., 2023); more process-based analysis using recently-developed diagnostic tools (Takahashi et al., 2021; Shaw et al., 2025; Nakamura et al., 2026, e.g.); better use of the growing historical record as an evaluation record (Bodas-Salcedo et al., 2023; Simpson et al., 2025). Ideally, this should be available through accessible community tools like ESMValTool (Eyring et al., 2020) so that it can be routinely used in model development and tuning.

Also, PPEs can be better exploited in the development process to understand structural uncertainty, which combined with better process-based evaluation will help guide model development towards models that are more consistent with the IPCC-assessed climate sensitivity range without having to compromise on other areas of performance.

Code availability. The GOSI9 official release is available to download at https://doi.org/10.5281/zenodo.13814369 (Guiavarc'h and Storkey, 2024). The UM and/or JULES code branch(es) used in the publication have not all been submitted for review and inclusion in the UM/JULES trunk or released for general use. Due to intellectual property copyright restrictions, we cannot provide the source code for the UM or JULES, but a copy was made available to the reviewers of this work. The UM is available for use under licence. A number of research organisations and national meteorological services use the UM in collaboration with the Met Office to undertake atmospheric process research, produce forecasts, develop the UM code and build and evaluate Earth system models. To apply for a licence for the UM, go to https://www.metoffice.gov.uk/research/approach/ modelling-systems/unified-model, and for permission to use JULES, go to https://jules.jchmr.org.

Data availability. ERA Interim reanalysis data are available from the Copernicus Climate Data Store at https://cds.climate.copernicus.eu/datasets/reanalysis-era-interim, published by the European Centre for Medium Range Weather Forecasts (ECMWF) under the Creative Commons Attribution 4.0 International licence (CC BY 4.0) (Dee et al., 2011). CERES EBAF Edition 4.1 radiative flux data are publicly available from NASA ASDC (https://asdc.larc.nasa.gov/project/CERES/CERES_EBAF_Edition4.1) (Loeb et al., 2018). HadCRUT5 data were obtained from http://www.metoffice.gov.uk/hadobs/hadcrut5 and are British Crown Copyright, Met Office 2026, provided under an Open Government License, http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ (Morice et al., 2021). HadISST.2.2.0.0 sea ice concentration data are available for download from the Met Office Hadley Centre at https://www.metoffice.gov.uk/hadobs/hadisst2/data/download.html (Titchner and Rayner, 2014). EUMETSAT OSI SAF sea ice concentration data are available from https://user.eumetsat.int/catalogue/EO:EUM:DAT:0875 (EUMETSAT OSI SAF, 2020).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Code and data availability. The scripts used to process the data and reproduce the figures this paper are publicly available on Zenodo at https://doi.org/10.5281/zenodo.23162219 (Bodas-Salcedo, 2026).

Author contributions. ABS led the UKCM2-LL development, the production of the DECK simulations, and the preparation of the manuscript, including the processing code to produce the Tables and Figures. JWR, CG, WK and HM ran the DECK simulations and contributed to the analysis. JWR, EB, MAR, CGJ, JPM, TEP, STR, DMHS, YT and MRW contributed to the UKCM2-LL development. HM and TA produced the ERF calculations and their analysis. MAR contributed to the feedback analysis. WK and KY contributed to the analysis of the historical simulations. All authors contributed with comments and edits to the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. All scientific content, analysis, and interpretation were carried out by the authors. GitHub Copilot was used in Visual Studio Code to provide inline code suggestions during the editing and refactoring of data-processing and plotting scripts. All scripts were reviewed, modified, and validated by the authors, who take full responsibility for the scientific content and results. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. ABS thanks Richard Wood for his continuous support during the development of UKCM2-LL. We thank Daley Calvert and Alex West for their help implementing ocean and sea ice diagnostics in UKCM2.

Financial support. A

. Bodas-Salcedo, J. W. Rostron, C. Guiavarc'h, W. Kay, H. Mutton, E. W. Blockley, M. A. Ringer, K. Yamazaki, M. Thomas, T. Andrews, J. P. Mulcahy, T. E. Palmer, D. M. H. Sexton, Y. Tang, and M. R. Willett were supported by the Met Office Hadley Centre Climate Programme (HCCP) funded by the UK Department for Business, Innovation, Science and Trade (DBIST). C. G. Jones and S. T. Rumbold were supported by TerraFIRMA "Future Impacts, Risks and Mitigation Actions in a changing Earth system" funded by the UKRI Natural Environment Research Council (grant reference NE/W004895/1). C. G. Jones was also supported by the European Union Horizon Europe project OptimESM "Optimal High Resolution Earth System Models for Exploring Future Climate Changes" (grant agreement no. 101081193) and the UK government's Horizon Europe funding guarantee (grant numbers, 10103098, 10043072).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

References

Global Climate Highlights 2023, Tech. rep., Copernicus Climate Change Service (C3S), European Centre for Medium-Range Weather Forecasts (ECMWF), https://climate.copernicus.eu/global-climate-highlights-2023, 2024.

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., Webb, M. J., Armour, K. C., Forster, P. M., and Titchner, H.: Accounting for changing temperature patterns increases historical estimates of climate sensitivity, Geophys. Res. Lett., 45, https://doi.org/10.1029/2018GL078887, 2018.

Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., Menary, M. B., Ridley, J., Ringer, M. A., Sellar, A. A., Senior, C. A., and Tang, Y.: Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1, Journal of Advances in Modeling Earth Systems, 11, 4377–4394, https://doi.org/https://doi.org/10.1029/2019MS001866, 2019.

Armour, K. C.: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks, Nature Climate Change, 7, 331–335, https://doi.org/10.1038/nclimate3278, 2017.

Aubry, T. J., Sigl, M., Toohey, M., Chim, M. M., Verkerk, M., Schmidt, A., and Carn, S. A.: Stratospheric aerosol forcing for CMIP7 (part 2): Volcanic sulfur dioxide emissions, EGUsphere, 2026, 1–53, https://doi.org/10.5194/egusphere-2026-546, 2026a.

Aubry, T. J., Toohey, M., Khanal, S., Chim, M. M., Verkerk, M., Johnson, B., Schmidt, A., Kovilakam, M., Sigl, M., Nicholls, Z., Thomason, L., Naik, V., Rieger, L., Stiller, D., Ziegler, E., Durack, P., and Smith, I. H.: Stratospheric aerosol forcing for CMIP7 – Part 1: optical properties for pre-industrial, historical, and scenario simulations, Geoscientific Model Development, 19, 3725–3756, https://doi.org/10.5194/gmd-19-3725-2026, 2026b.

Blockley, E., Fiedler, E., Ridley, J., Roberts, L., West, A., Copsey, D., Feltham, D., Graham, T., Livings, D., Rousset, C., Schroeder, D., and Vancoppenolle, M.: The sea ice component of GC5: coupling SI^3 to HadGEM3 using conductive fluxes, Geosci. Model Devel., 17, 6799–6817, https://doi.org/10.5194/gmd-17-6799-2024, 2024.

Bodas-Salcedo, A., Sexton, J. M. G. D. M. H., and Morice, C. P.: Assessment of Large-Scale Indices of Surface Temperature during the Historical Period in the CMIP6 Ensemble, J. Climate, 36, 2055–2072, https://doi.org/10.1175/JCLI-D-22-0398.1, 2023.

Ceppi, P. and Nowack, P.: Observational evidence that cloud feedback amplifies global warming, Proceedings of the National Academy of Sciences, 118, e2026290 118, https://doi.org/10.1073/pnas.2026290118, 2021.

Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geoscientific Model Development, 10, 3297–3308, https://doi.org/10.5194/gmd-10-3297-2017, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., and et al: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., Paynter, D., Smith, C. J., and Shiogama, H.: Biased Estimates of Equilibrium Climate Sensitivity and Transient Climate Response Derived From Historical CMIP6 Simulations, Geophysical Research Letters, 48, e2021GL095 778, https://doi.org/https://doi.org/10.1029/2021GL095778, e2021GL095778 2021GL095778, 2021.

Dunne, J. P., Hewitt, H. T., Arblaster, J. M., Bonou, F., Boucher, O., Cavazos, T., Dingley, B., Durack, P. J., Hassler, B., Juckes, M., Miyakawa, T., Mizielinski, M., Naik, V., Nicholls, Z., O'Rourke, E., Pincus, R., Sanderson, B. M., Simpson, I. R., and Taylor, K. E.: An evolving Coupled Model Intercomparison Project phase 7 (CMIP7) and Fast Track in support of future climate assessment, Geoscientific Model Development, 18, 6671–6700, https://doi.org/10.5194/gmd-18-6671-2025, 2025.

EUMETSAT OSI SAF: Sea Ice Index - Multimission, https://doi.org/10.15770/EUM_SAF_OSI_0022, 2020.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation, Geosci. Model Devel., 8, 10 539–10 583, https://doi.org/10.5194/gmdd-8-10539-2015, 2016.

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E. L., Davini, P., Debeire, K., de Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T., Swaminathan, R., Torralba, V., Vegas-Regidor, J., von Hardenberg, J., Weigel, K., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geoscientific Model Development, 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D., Mauritsen, T., Palmer, M., Watanabe, M., Wild, M., and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., book section 7, pp. 923–1054, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157896.009, 2021.

Furtado, K. and Field, P.: The Role of Ice Microphysics Parametrizations in Determining the Prevalence of Supercooled Liquid Water in High-Resolution Simulations of a Southern Ocean Midlatitude Cyclone, Journal of the Atmospheric Sciences, 74, 2001 – 2021, https://doi.org/10.1175/JAS-D-16-0165.1, 2017.

Furtado, K., Tsushima, Y., Field, P. R., Rostron, J., and Sexton, D.: The Relationship Between the Present-Day Seasonal Cycles of Clouds in the Mid-Latitudes and Cloud-Radiative Feedback, Geophysical Research Letters, 50, e2023GL103 902, https://doi.org/https://doi.org/10.1029/2023GL103902, e2023GL103902 2023GL103902, 2023.

Gantt, B., Meskhidze, N., Facchini, M. C., Rinaldi, M., Ceburnis, D., and O'Dowd, C. D.: Wind speed dependent size-resolved parameterization for the organic mass fraction of sea spray aerosol, Atmospheric Chemistry and Physics, 11, 8777–8790, https://doi.org/10.5194/acp-11-8777-2011, 2011.

Gantt, B., Johnson, M. S., Crippa, M., Prévôt, A. S. H., and Meskhidze, N.: Implementing marine organic aerosols into the GEOS-Chem model, Geoscientific Model Development, 8, 619–629, https://doi.org/10.5194/gmd-8-619-2015, 2015.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res., 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Gregory, J. M. and Andrews, T.: Variation in climate sensitivity and feedback parameters during the historical period, Geophys. Res. Lett., 43, 3911–3920, https://doi.org/10.1002/2016GL068406, 2016.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, L03 205, https://doi.org/10.1029/2003gl018747, 2004.

Gregory, J. M., Andrews, T., and Good, P.: The inconstancy of the transient climate response parameter under increasing CO_2, Philos. Trans. R. Soc. London, 373, 20140 417, https://doi.org/10.1098/rsta.2014.0417, 2015.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Guiavarc'h, C., Storkey, D., Blaker, A. T., Blockley, E., Megann, A., Hewitt, H., Bell, M. J., Calvert, D., Copsey, D., Sinha, B., Moreton, S., Mathiot, P., and An, B.: GOSI9: UK Global Ocean and Sea Ice configurations, Geoscientific Model Development, 18, 377–403, https://doi.org/10.5194/gmd-18-377-2025, 2025.

Gyuleva, G., Fischer, E., Knutti, R., and Sippel, S.: Recent Temperature and Energy Imbalance Trends Point to Higher Estimates of Future Warming, Earth's Future, 14, e2026EF008 356, https://doi.org/https://doi.org/10.1029/2026EF008356, e2026EF008356 2026EF008356, 2026.

Hardacre, C., Mulcahy, J. P., Pope, R. J., Jones, C. G., Rumbold, S. T., Li, C., Johnson, C., and Turnock, S. T.: Evaluation of SO2, SO42- and an updated SO2 dry deposition parameterization in the United Kingdom Earth System Model, Atmospheric Chemistry and Physics, 21, 18 465–18 497, https://doi.org/10.5194/acp-21-18465-2021, 2021.

Jones, G. S., Andrews, M. B., Andrews, T., Blockley, E., Ciavarella, A., Christidis, N., Cotterill, D. F., Lott, F. C., Ridley, J., and Stott, P. A.: The HadGEM3-GC3.1 Contribution to the CMIP6 Detection and Attribution Model Intercomparison Project, Journal of Advances in Modeling Earth Systems, 16, e2023MS004 135, https://doi.org/https://doi.org/10.1029/2023MS004135, e2023MS004135 2023MS004135, 2024.

Kramer, R. J., Smith, C., and Andrews, T.: The Radiative Forcing Model Intercomparison Project (RFMIP2.0) for CMIP7, Geoscientific Model Development, 19, 4447–4466, https://doi.org/10.5194/gmd-19-4447-2026, 2026.

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., Hill, R., Graham, T., Ridley, J., Blaker, A., Calvert, D., Copsey, D., Ellis, R., Hewitt, H., Hyder, P., Ineson, S., Mulcahy, J., Siahaan, A., and Walton, J.: The Low-Resolution Version of HadGEM3 GC3.1: Development and Evaluation for Global Climate, Journal of Advances in Modeling Earth Systems, 10, 2865–2888, https://doi.org/https://doi.org/10.1029/2018MS001370, 2018.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Kreiner, M. B., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., and Kato, S.: Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product, Journal of Climate, 31, 895 – 918, https://doi.org/10.1175/JCLI-D-17-0208.1, 2018.

L'Ecuyer, T. S., Beaudoing, H. K., Rodell, M., Olson, W., Lin, B., Kato, S., Clayson, C. A., Wood, E., Sheffield, J., Adler, R., Huffman, G., Bosilovich, M., Gu, G., Robertson, F., Houser, P. R., Chambers, D., Famiglietti, J. S., Fetzer, E., Liu, W. T., Gao, X., Schlosser, C. A., Clark, E., Lettenmaier, D. P., and Hilburn, K.: The Observed State of the Energy Budget in the Early Twenty-First Century, Journal of Climate, 28, 8319 – 8346, https://doi.org/10.1175/JCLI-D-14-00556.1, 2015.

Madec, G., Bourdallé-Badie, R., Chanut, J., Clementi, E., Coward, A., Ethé, C., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Rousset, C., Storkey, D., Vancoppenolle, M., Müeller, S., Nurser, G., Bell, M., and Samson, G.: NEMO ocean engine, https://doi.org/10.5281/ZENODO.3878122, version Number: v4.0, 2019.

Menary, M. B., Robson, J., Allan, R. P., Booth, B. B. B., Cassou, C., Gastineau, G., Gregory, J., Hodson, D., Jones, C., Mignot, J., Ringer, M., Sutton, R., Wilcox, L., and Zhang, R.: Aerosol-Forced AMOC Changes in CMIP6 Historical Simulations, Geophysical Research Letters, 47, e2020GL088 166, https://doi.org/https://doi.org/10.1029/2020GL088166, e2020GL088166 2020GL088166, 2020.

Moat, B. I., Smeed, D., Rayner, D., Johns, W. E., Smith, R. H., Volkov, D. L., Elipot, S., Petit, T., Kajtar, J. B., Baringer, M. O., and Collins, J.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2024 (v2024.1a)., https://doi.org/10.5285/48D0BF43-0598-CEB2-E063-7086ABC062F1, 2026.

Morice, C., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/https://doi.org/10.1029/2019JD032361, e2019JD032361 2019JD032361, 2021.

Mulcahy, J. P., Jones, C. G., Rumbold, S. T., Kuhlbrodt, T., Dittus, A. J., Blockley, E. W., Yool, A., Walton, J., Hardacre, C., Andrews, T., Bodas-Salcedo, A., Stringer, M., de Mora, L., Harris, P., Hill, R., Kelley, D., Robertson, E., and Tang, Y.: UKESM1.1: development and evaluation of an updated configuration of the UK Earth System Model, Geoscientific Model Development, 16, 1569–1600, https://doi.org/10.5194/gmd-16-1569-2023, 2023.

Nakamura, Y., Suzuki, K., and Horie, H.: Incorporating W-Band Doppler Velocity Signal Simulator Into COSP2: Model Evaluation Against Ground-Based Radar Measurement, Journal of Advances in Modeling Earth Systems, 18, e2025MS004 958, https://doi.org/https://doi.org/10.1029/2025MS004958, e2025MS004958 2025MS004958, 2026.

Peatier, S., Sanderson, B. M., Terray, L., and Roehrig, R.: Investigating Parametric Dependence of Climate Feedbacks in the Atmospheric Component of CNRM-CM6-1, Geophys. Res. Lett., 49, e2021GL095 084, https://doi.org/10.1029/2021GL095084, 2022.

Ramanathan, V., Cess, R. D., Harrison, E. F., Minnis, P., Barkstrom, B. R., Ahmad, E., and Hartmann, D.: Cloud radiative forcing and climate: Results from the Earth Radiation Budget Experiment, Science, 243, 57–63, 1989.

Ridley, J. K., Blockley, E. W., and Jones, G. S.: A Change in Climate State During a Pre-Industrial Simulation of the CMIP6 Model HadGEM3 Driven by Deep Ocean Drift, Geophysical Research Letters, 49, e2021GL097 171, https://doi.org/https://doi.org/10.1029/2021GL097171, e2021GL097171 2021GL097171, 2022.

Rostron, J. W., Bodas-Salcedo, A., Sexton, D. M. H., Jones, C. G., Blockley, E. W., Kuhlbrodt, T., Mulcahy, J. P., Palmer, T. E., Peatier, S., Ringer, M. A., Rumbold, S. T., Sanderson, B. M., Tang, Y., and Willet, M. R.: UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity – methodology and development, EGUsphere, 2026, 1–46, https://doi.org/10.5194/egusphere-2026-1676, 2026.

Rugenstein, M., Bloch-Johnson, J., Gregory, J., Andrews, T., Mauritsen, T., Li, C., Frölicher, T. L., Paynter, D., Danabasoglu, G., Yang, S., Dufresne, J.-L., Cao, L., Schmidt, G. A., Abe-Ouchi, A., Geoffroy, O., and Knutti, R.: Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models, Geophysical Research Letters, 47, e2019GL083 898, https://doi.org/https://doi.org/10.1029/2019GL083898, e2019GL083898 10.1029/2019GL083898, 2020.

Samset, B. H., Wilcox, L. J., and Allen, R. J.: Broader research efforts and assessments needed to uncover the complex climate effects of regional changes in aerosol emissions, PLOS Clim, 3, e0000 508, https://doi.org/10.1371/journal.pclm.0000508, 2024.

Scott, R. C., Myers, T. A., Norris, J. R., Zelinka, M. D., Klein, S. A., Sun, M., and Doelling, D. R.: Observed Sensitivity of Low-Cloud Radiative Effects to Meteorological Perturbations over the Global Oceans, Journal of Climate, 33, 7717 – 7734, https://doi.org/10.1175/JCLI-D-19-1028.1, 2020.

Senior, C. A. and Mitchell, J. F. B.: The time dependence of climate sensitivity, Geophys. Res. Lett., 27, 2685–2688, 2000.

Shaw, J. K., Swales, D. J., DeSouza-Machado, S., Turner, D. D., Kay, J. E., and Schneider, D. P.: COSP-RTTOV-1.0: flexible radiation diagnostics to enable new science applications in model evaluation, climate change detection, and satellite mission design, Geoscientific Model Development, 18, 4935–4950, https://doi.org/10.5194/gmd-18-4935-2025, 2025.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C. S., Foster, G. L., Hausfather, Z., von der Heydt, A. S., Knutti, R., Mauritsen,

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

T., Norris, J. R., Proistosescu, C., Rugenstein, M., Schmidt, G. A., Tokarska, K. B., and Zelinka, M. D.: An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence, Rev. Geophys., 58, e2019RG000 678, https://doi.org/10.1029/2019RG000678, 2020.

Simpson, I. R., Shaw, T. A., Ceppi, P., Clement, A. C., Fischer, E., Grise, K. M., Pendergrass, A. G., Screen, J. A., Wills, R. C. J., Woollings, T., Blackport, R., Kang, J. M., and Po-Chedley, S.: Confronting Earth System Model trends with observations, Science Advances, 11, eadt8035, https://doi.org/10.1126/sciadv.adt8035, 2025.

Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L., and McCarthy, G. D.: The North Atlantic Ocean Is in a State of Reduced Overturning, Geophysical Research Letters, 45, 1527–1533, https://doi.org/https://doi.org/10.1002/2017GL076350, 2018.

Soden, B. J., Broccoli, A. J., and Hemler, R. S.: On the use of cloud forcing to estimate cloud feedback, J. Climate, 17, 3661–3665, https://doi.org/10.1175/1520-0442(2004)017, 2004.

Soden, B. J., Held, I. M., Colman, R., Shell, K. M., Kiehl, J. T., and Shields, C. A.: Quantifying climate feedbacks using radiative kernels, J. Climate, 21, 3504–3520, https://doi.org/10.1175/2007JCLI2110.1, 2008.

Storkey, D., Mathiot, P., Bell, M. J., Copsey, D., Guiavarc'h, C., Hewitt, H. T., Ridley, J., and Roberts, M. J.: Resolution dependence of interlinked Southern Ocean biases in global coupled HadGEM3 models, Geoscientific Model Development, 18, 2725–2745, https://doi.org/10.5194/gmd-18-2725-2025, 2025.

Takahashi, H., Bodas-Salcedo, A., and Stephens, G.: Warm Cloud Evolution, Precipitation, and Their Weak Linkage in HadGEM3: New Process-Level Diagnostics Using A-Train Observations, Journal of the Atmospheric Sciences, 78, 2075 – 2087, https://doi.org/10.1175/JAS-D-20-0321.1, 2021.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, Journal of Geophysical Research: Atmospheres, 119, 2864–2889, https://doi.org/https://doi.org/10.1002/2013JD020316, 2014.

Vancoppenolle, M., Rousset, C., Blockley, E., Aksenov, Y., Feltham, D., Fichefet, T., Garric, G., Guémas, V., Iovino, D., Keeley, S., Madec, G., Massonnet, F., Ridley, J., Schroeder, D., and Tietsche, S.: SI3, the NEMO Sea Ice Engine, https://doi.org/10.5281/zenodo.7534900, 2023.

Weverberg, K. V., Morcrette, C. J., and Boutle, I.: A Bimodal Diagnostic Cloud Fraction Parameterization. Part II: Evaluation and Resolution Sensitivity, Monthly Weather Review, 149, 859–878, https://doi.org/10.1175/MWR-D-20-0230.1, 2021a.

Weverberg, K. V., Morcrette, C. J., Boutle, I., Furtado, K., and Field, P. R.: A Bimodal Diagnostic Cloud Fraction Parameterization. Part I: Motivating Analysis and Scheme Description, Monthly Weather Review, 149, 841–857, https://doi.org/10.1175/MWR-D-20-0224.1, 2021b.

Willett, M. R., Brooks, M. E., Bushell, A., Earnshaw, P., Marzin, C., Smith, S., Tomassini, L., Abraham, N. L., Best, M., Edwards, J. M., Furtado, K., Hardacre, C., Hewitt, A. J., Johnson, B., Lock, A., Mulcahy, J. P., Manners, J., Sellar, A., Sheridan, P., Tennant, W., Van Weverberg, K., Varma, V., and Whitall, M.: The Met Office Unified Model and JULES Global Atmosphere and Land 9.0 configuration, EGUsphere, 2026, 1–68, https://doi.org/10.5194/egusphere-2026-2425, 2026.

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., Scaife, A. A., Schiemann, R., Storkey, D., Thorpe, L., Watterson, I. G., Walters, D. N., West, A., Wood, R. A., Woollings, T., and Xavier, P. K.: The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations, J. Adv. Model Earth Syst., 10, 357–380, https://doi.org/10.1002/2017MS001115, 2018.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.



Williams, K. D., van Niekerk, A., Best, M. J., Lock, A. P., Brooke, J. K., Carvalho, M. J., Derbyshire, S. H., Dunstan, T. D., Rumbold, H. S., Sandu, I., and Sexton, D. M. H.: Addressing the causes of large-scale circulation error in the Met Office Unified Model, Quarterly Journal of the Royal Meteorological Society, 146, 2597–2613, https://doi.org/10.1002/qj.3807, 2020.

Xavier, P., Willett, M., Graham, T., Earnshaw, P., Copsey, D., Narayan, N., Marzin, C., Sellar, A., Ackerley, D., Blaker, A., Blockley, E., Bodas-Salcedo, A., Bushell, A., Choi, N., Chua, X. R., Guiavarc'h, C., Hassim, M., Heming, J., Hudson, D., Ineson, S., Jones, A., Jones, C., Keane, R., Kim, K., Kim, J., Kuhlbrodt, T., Lee, M.-I., Levine, R., Li, C., Martin, G., Megann, A., Mccabe, A., Moise, A., Regayre, L., Ridley, J., Roberts, L., Sahany, S., Schiemann, R. K. H., Storkey, D., Tennant, W., Tomassini, L., Tsushima, Y., Weedon, G. P., West, A., Wheeler, M. C., Williams, K., Zhou, X., and Zhu, H.: Assessment of the Met Office Global Coupled model version 5 (GC5) configurations, Geoscientific Model Development, submitted.

Xu, X., Chassignet, E. P., Firing, Y. L., and Donohue, K.: Antarctic Circumpolar Current Transport Through Drake Passage: What Can We Learn From Comparing High-Resolution Model Results to Observations?, Journal of Geophysical Research: Oceans, 125, e2020JC016 365, https://doi.org/https://doi.org/10.1029/2020JC016365, e2020JC016365 2020JC016365, 2020.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

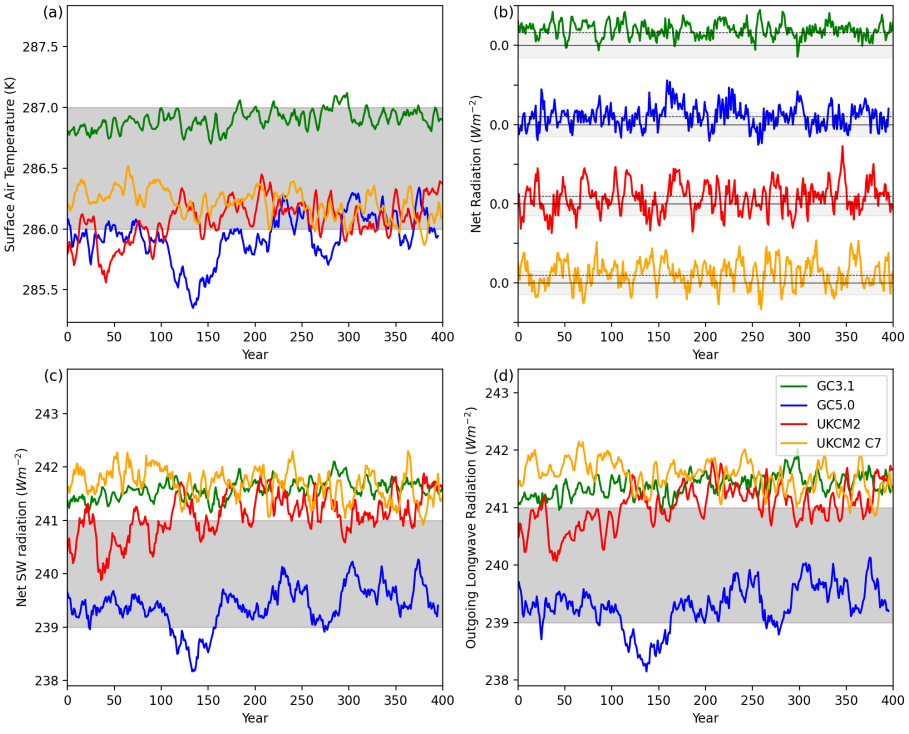


Figure 1. Time series of global mean near-surface air temperature (GMSAT) and top-of-atmosphere (TOA) radiative fluxes for the *piControl* simulation of UKCM2-LL. A moving average of 5 years has been applied to the time series of annual means. The gray shade shows the target range used during the development of UKCM2-LL (Table 2 of Rostron et al. (2026)). In panel (b), each line has been shifted to have its own zero line, with the dash line showing the time-average and the grey shade a $\pm 0.15 W m^{-2}$ region around the zero line. The distance between adjacent zero lines is 1 W m^{-2}.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

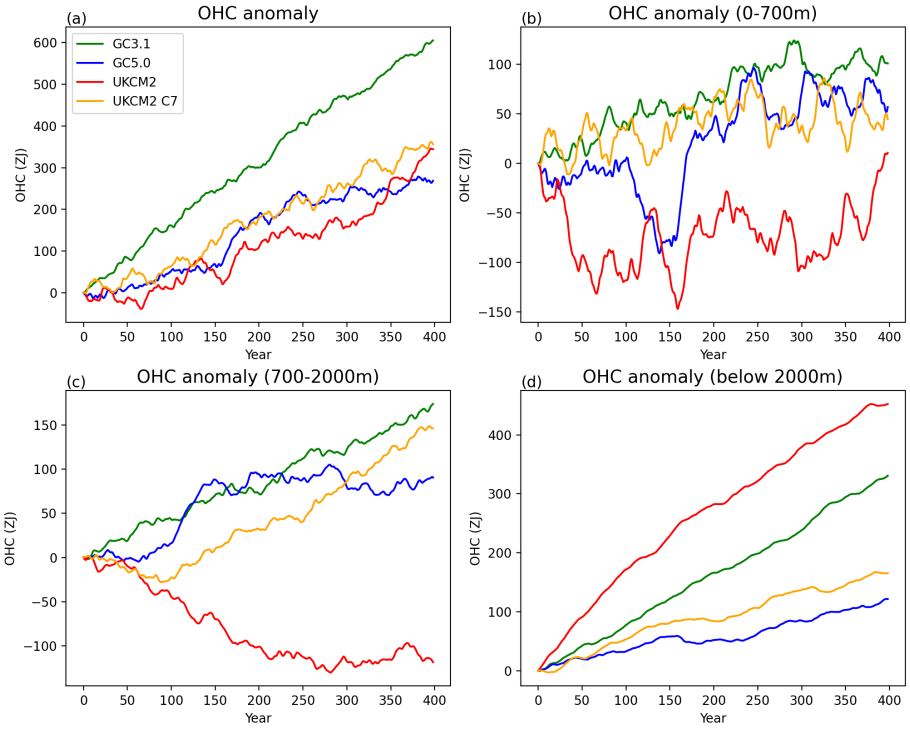


Figure 2. Time series of the Ocean Heat Content (OHC) for the *piControl* ensemble: (a) whole ocean; (b) 0-700 m; (c) 700-2000 m; (d) below 2000 m.

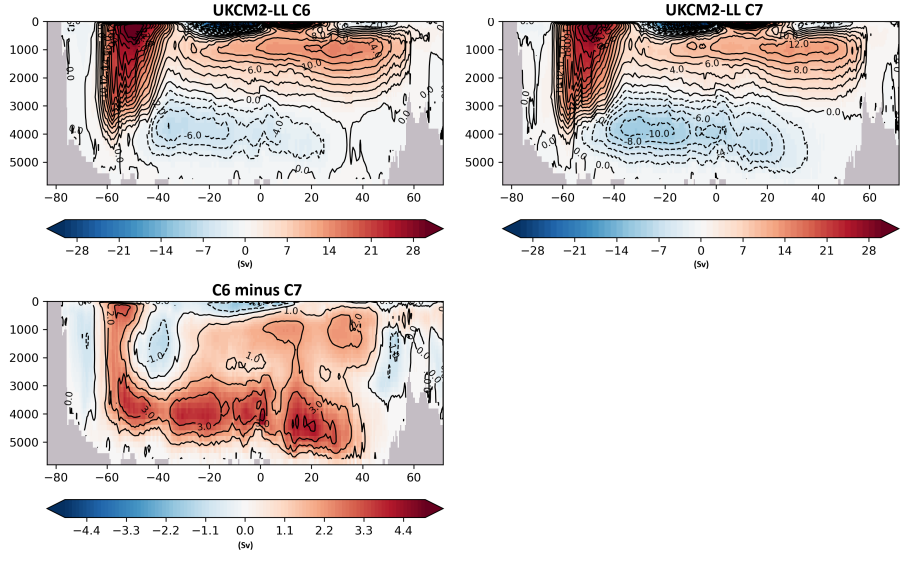


Figure 3. Time-averaged and zonally-averaged state of the MOC for both UKCM2 simulations for the period from years 150 to 159 of the piControl simulations of UKCM2: (a) UKCM2 with CMIP6 forcings; (b) UKCM2 with CMIP7 forcings; (c) difference between (a) and (b).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

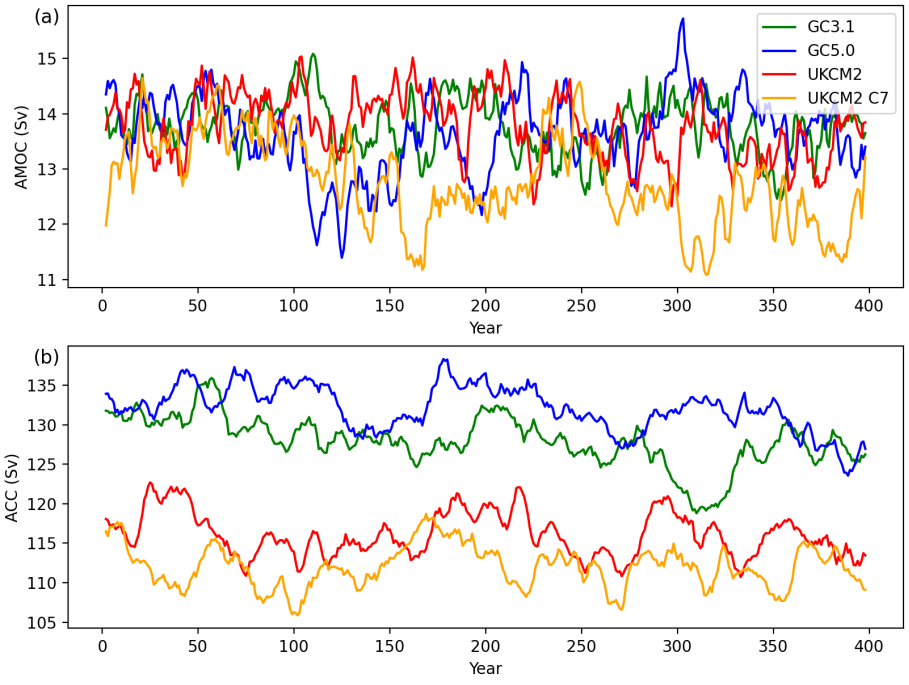


Figure 4. Time series of the Atlantic Meridional Overturning Circulation (AMOC) and the Antarctic Circumpolar Current (ACC) for the *piControl* ensemble.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

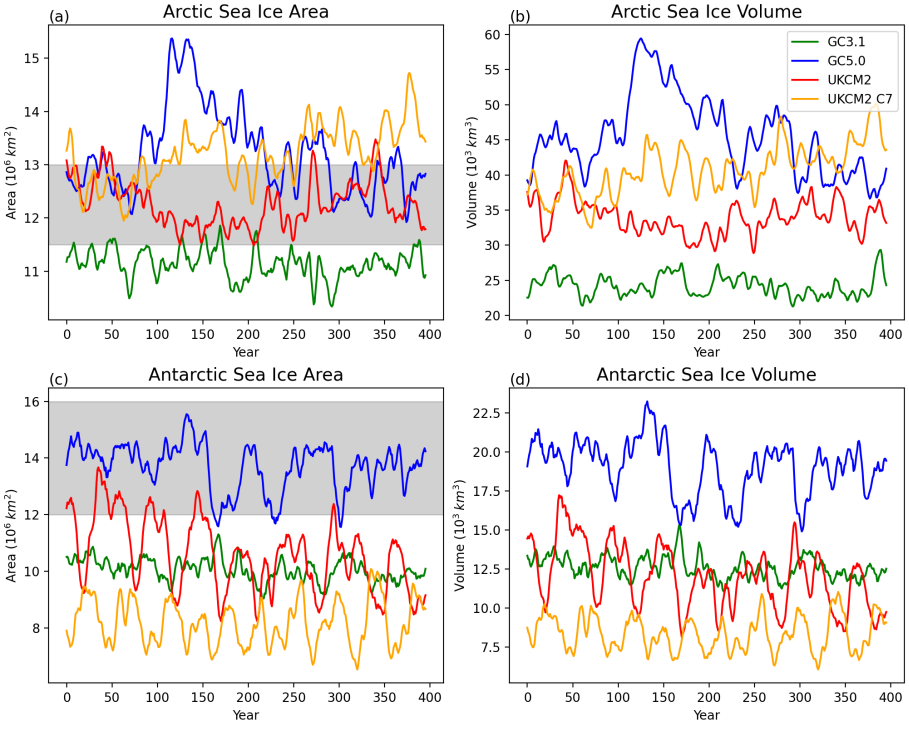


Figure 5. Time series of Arctic and Antarctic sea ice fraction and volume for the *piControl* simulations. The gray shade shows the target range used during the development of UKCM2-LL.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

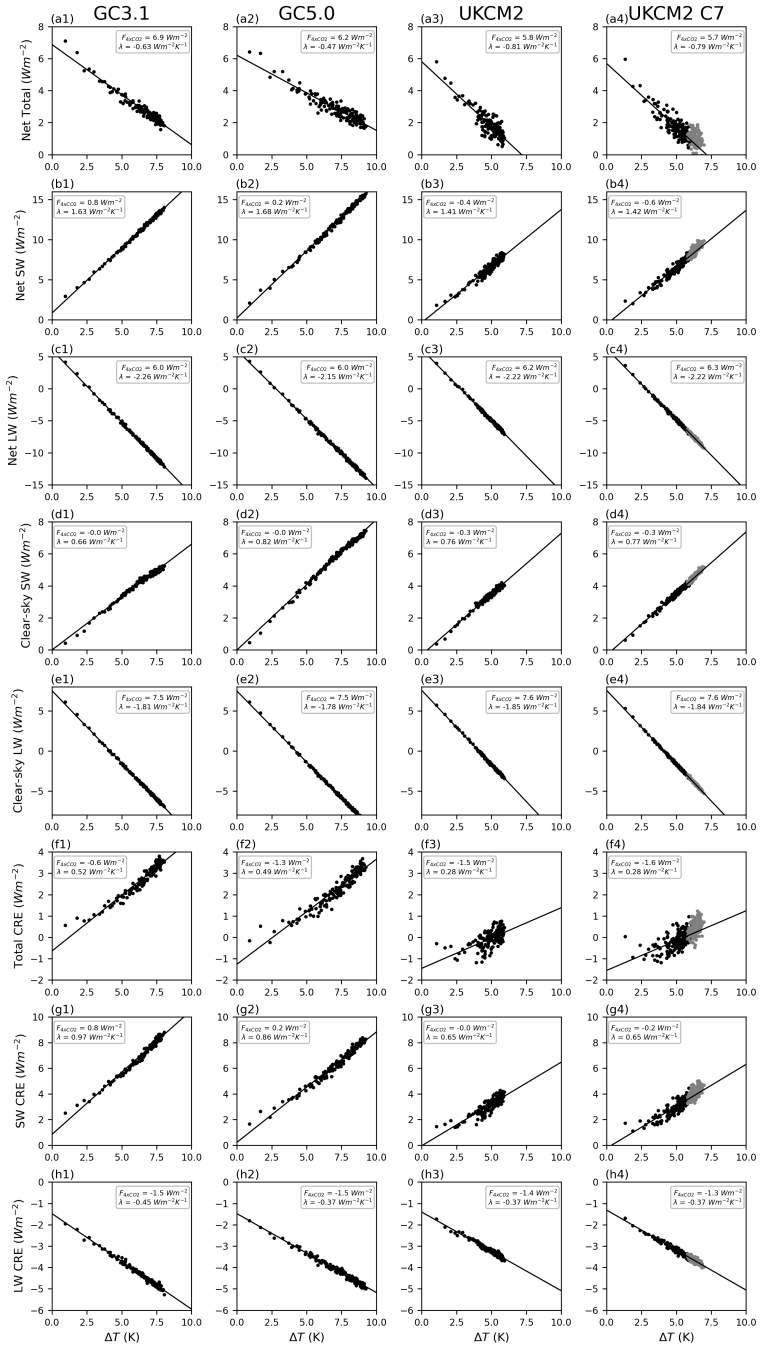


Figure 6. Gregory plots for the *abrupt-4xCO2* experiment of the small ensemble of models used in this study. Each row shows a different TOA radiative fluxe: (a) total net radiation; (b) SW net radiation; (c) LW net radiation; (d) clear-sky SW net radiation; (e) clear-sky LW net radiation; (f) total cloud radiative effect; (g) SW cloud radiative effect; (h) LW cloud radiative effect.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

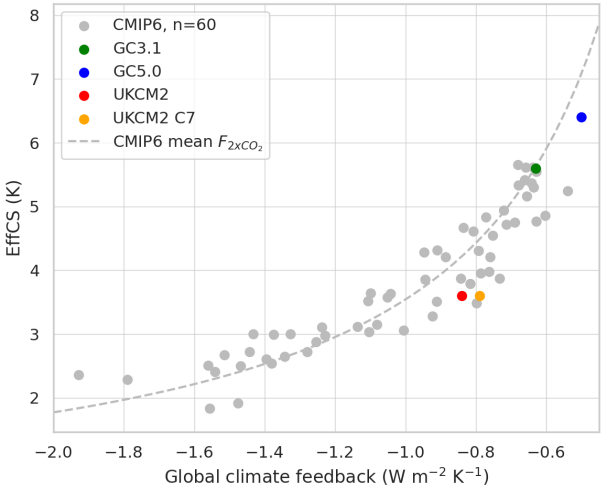


Figure 7. Scatter plot of the Effective Climate Sensitivity (EffCS) against the global feedback parameter (λ) for a range of CMIP6 models. The position of the UK models analysed in this paper are highlighted with coloured dots.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

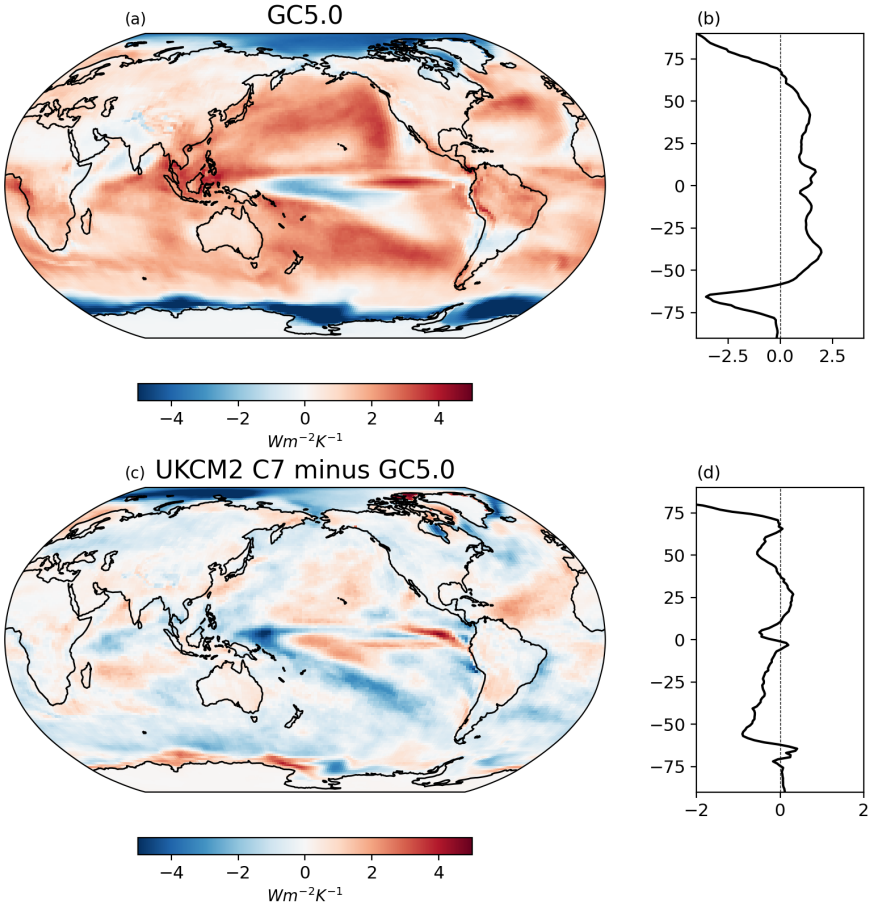


Figure 8. Shortwave CRE feedbacks for the *abrupt-4xCO2* simulations of the GC5.0 and UKCM2-LL models: (a) Regional feedback of GC5.0; (b) zonal mean of (a); (c) difference in the regional feedback between UKCM2-LL C7 and GC5.0; (d) zonal mean of the (c).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

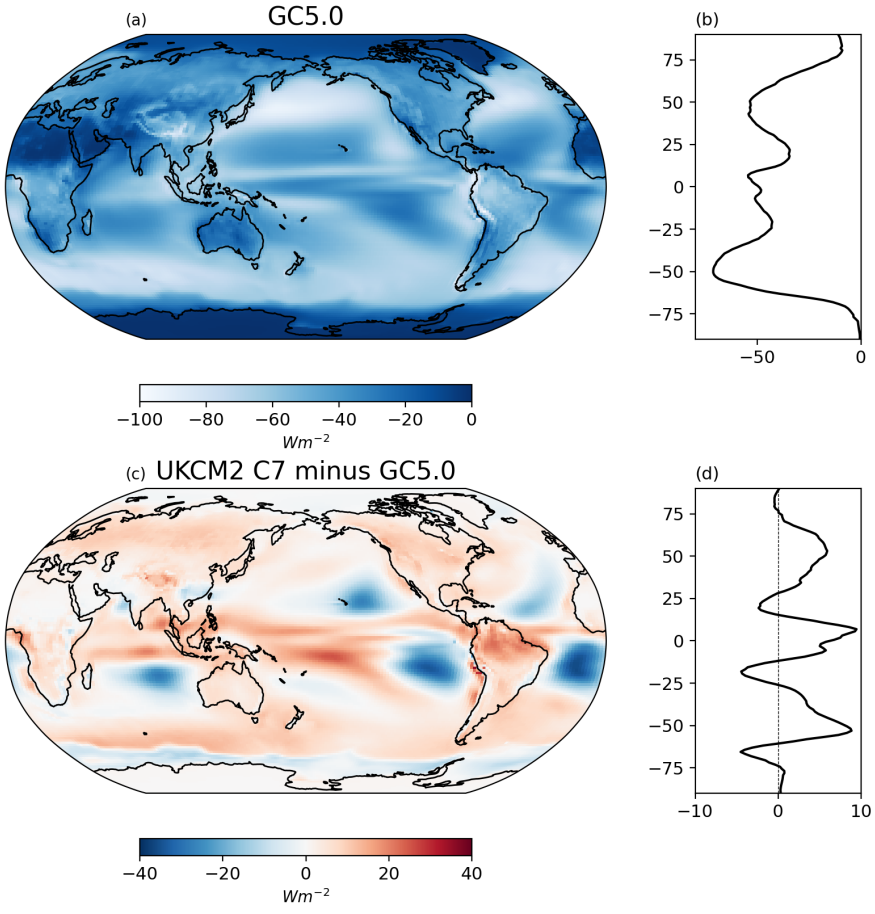


Figure 9. Shortwave CRE climatologies for the *piControl* simulations of the GC5.0 and UKCM2-LL C7 models. The top row shows the GC5.0 climatology and the bottom row shows the difference of the UKCM2-LL C7 climatology with respect to GC5.0. The right column shows the zonal means of the maps.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

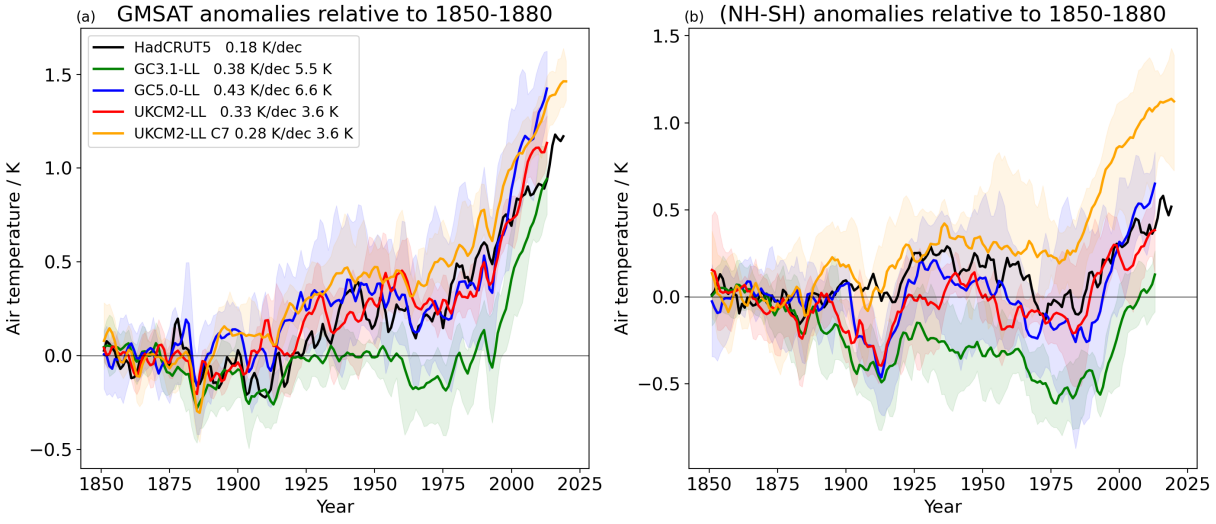


Figure 10. Evolution of the near surface air temperature anomalies relative to 1850-1900: (a) global-mean; (b) interhemispheric difference. The black line shows the observations from HadCRUT5, while the coloured lines show the model simulations. The thick lines show the ensemble mean, and the shade the 1-sigma range of the ensemble. The legend in panel (a) shows the global-mean temperature trends for the last 30 years of the CMIP6 historical period (1985-2014), as well as the EffCS of each model.

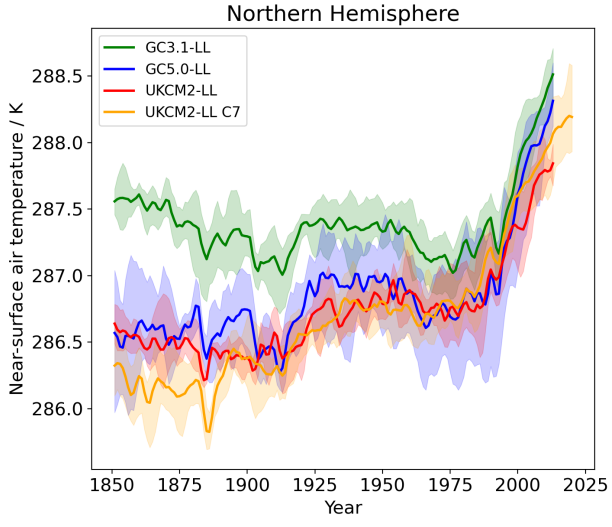


Figure 11. Time series of Northern Hemisphere near-surface air temperature for the ensembles of *historical* simulations. The thick lines show the ensemble mean, and the shade the 1-sigma range of the ensemble.

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

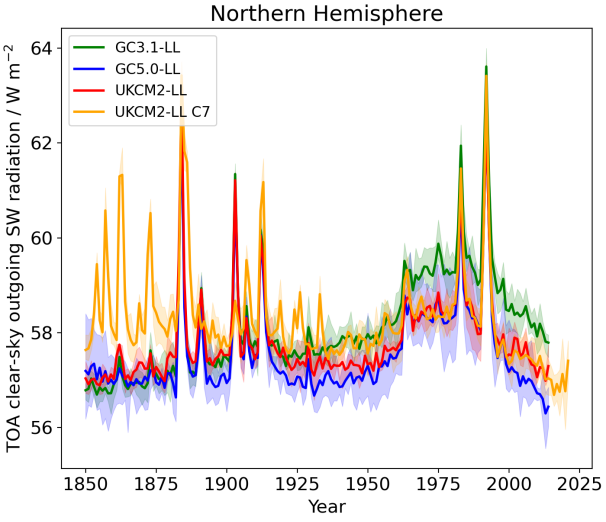


Figure 12. Time series of Northern Hemisphere clear-sky outgoing shortwave radiation for the ensembles of *historical* simulations. The thick lines show the ensemble mean, and the shade the 1-sigma range of the ensemble.

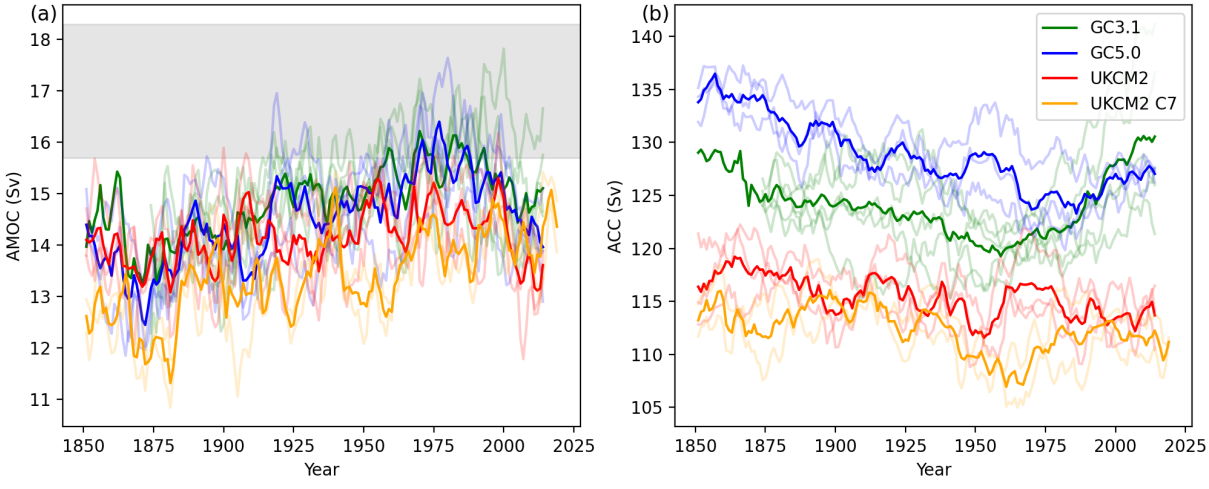


Figure 13. Strength of the ocean circulation for the model ensembles of *historical* simulations: (a) AMOC; (b) ACC. The dark colour lines show the models ensemble means, and the individual members in lighter colours. The grey shade shows the observational estimates for the AMOC (17±1.3 Sv).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

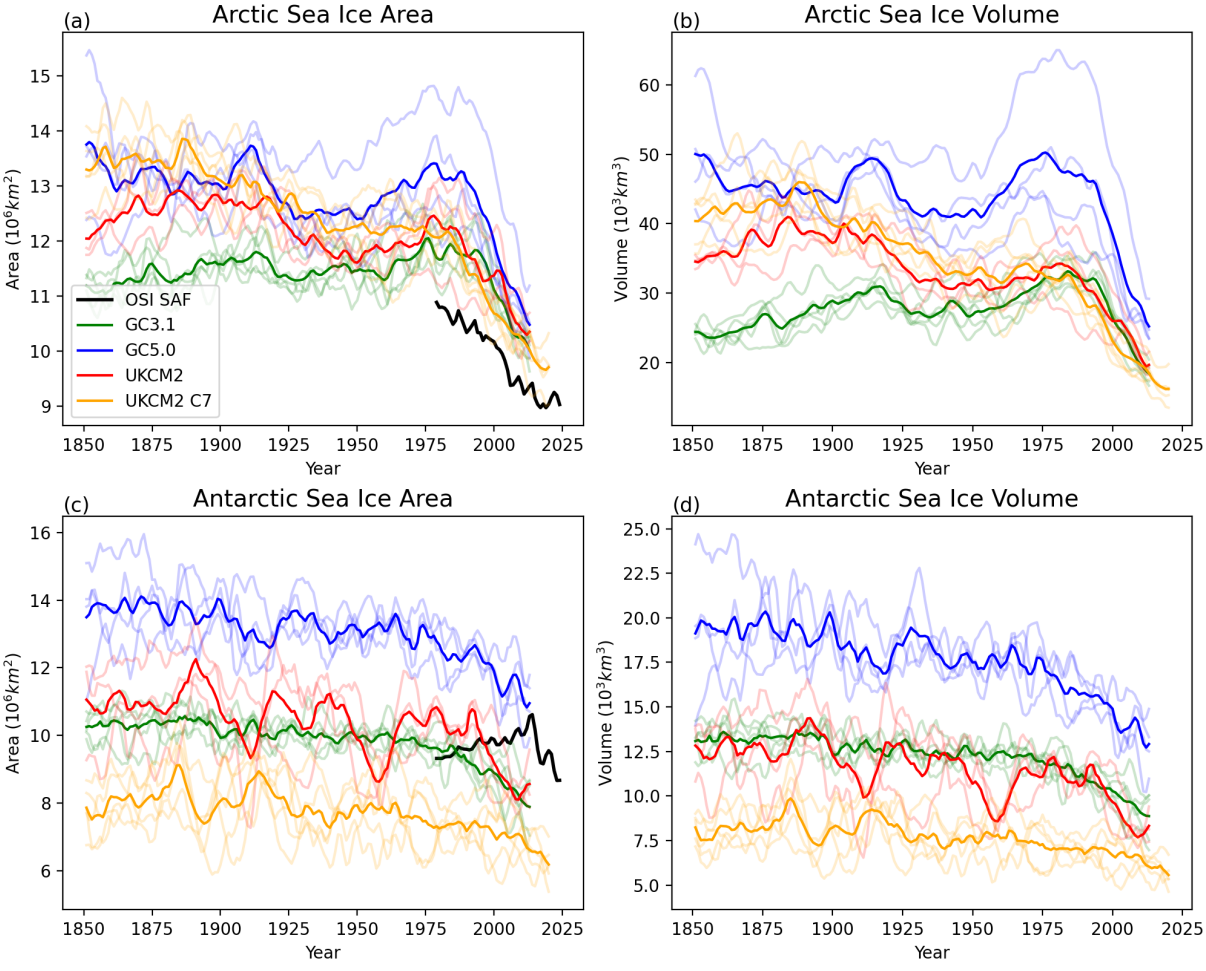


Figure 14. Time series of the sea ice properties for the model ensembles of *historical* simulations: (a) Arctic sea ice area; (b) Arctic sea ice volume; (c) Antarctic sea ice area; (d) Antarctic sea ice volume. The dark colour lines show the models ensemble means, and the individual members in lighter colours. The black lines show the observations from Version 2.3 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records (EUMETSAT OSI SAF, 2020; Lavergne et al., 2019).

https://doi.org/10.5194/egusphere-2026-6093
Preprint. Discussion started: 19 November 2026
© Author(s) 2026. CC BY 4.0 License.

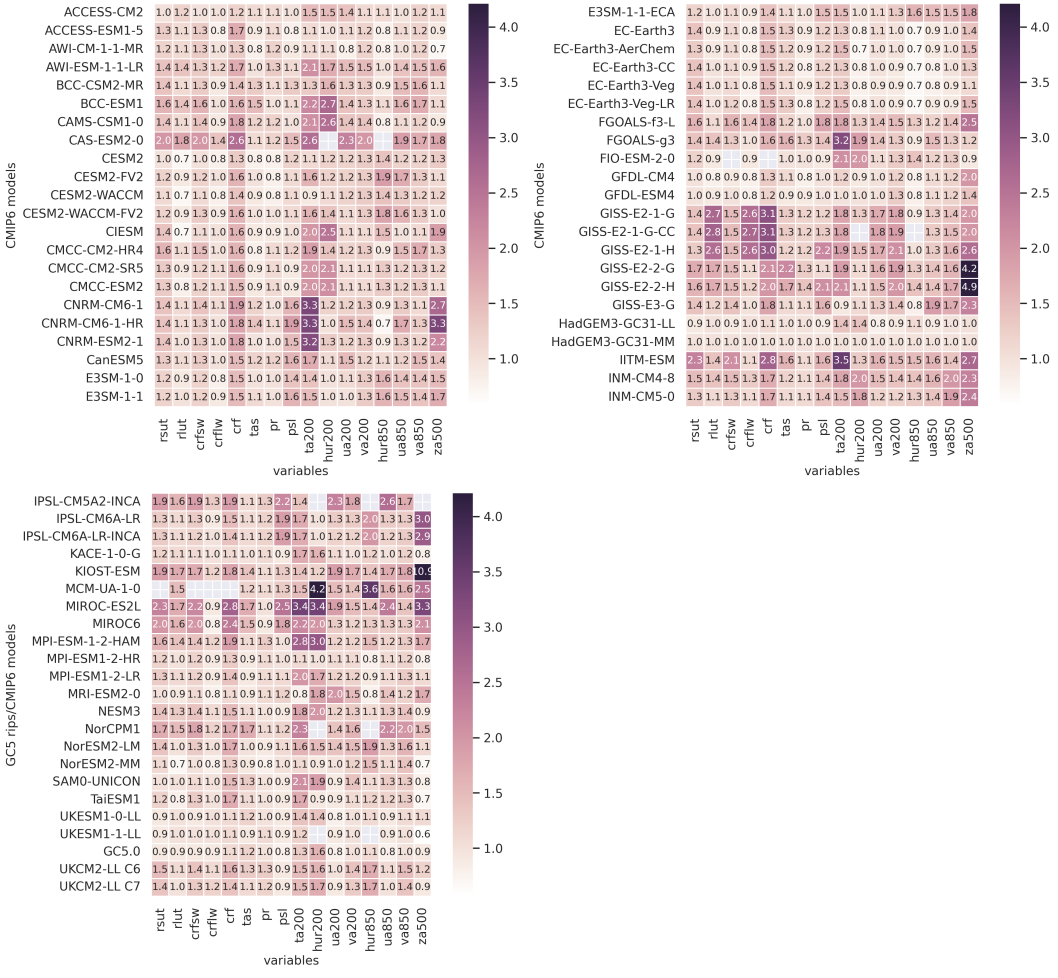


Figure 15. Heatmap of global-mean, annual-mean RMSE for key climate variables averaged over 1981-2000. Each row shows results for a single *historical* simulation. The values are normalised by the RMSE of HadGEM3-GC31-MM (row with all values equal to 1.0). Paler colours (i.e. smaller values of normalised RMSE) indicate better agreement with observations. The definitions of the acronyms are as follows: rsut: top-of-atmosphere (TOA) outgoing shortwave radiation, rlut: TOA outgoing longwave radiation, crfsw: shortwave cloud radiative forcing, crflw: longwave cloud radiative forcing, crf: total cloud radiative forcing, tas: near-surface air temperature, pr: total precipitation rate, psl: pressure at sea level, ta: temperature, hur: relative humidity, ua: zonal wind, va: meridional wind, za: geopotential height. The references for the observational dataset are: CERES (rsut, rlut), ERA-Interim (crfsw, crflw, crf, ta, ua, va, za), GPCP (pr) and HadSLP (psl). CERES data averaged over 2001-2005 are used because data for 1981-2000 do not exist.