https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity – methodology and development

John W. Rostron[1], Alejandro Bodas-Salcedo[1], David M. H. Sexton[1], Colin G. Jones[2], Edward W. Blockley[1], Till Kuhlbrodt[3], Jane P. Mulcahy[1], Tamzin E. Palmer[1], Saloua Peatier[3], Mark A. Ringer[1], Steven T. Rumbold[3], Benjamin M. Sanderson[4], Yongming Tang[1], Martin R. Willett[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK
[2]National Centre for Atmospheric Science, School of Earth and the Environment, University of Leeds, Leeds, UK
[3]National Centre for Atmospheric Science, University of Reading, Reading, UK
[4]CICERO Center for International Climate Research, Oslo, Norway

Correspondence to: John W. Rostron (john.rostron@metoffice.gov.uk)

Abstract. The Global Coupled model version 5 (GC5) of the Met Office Unified Model incorporates substantial developments across its model components and shows improved performance for a range of applications. However, it exhibits a very high effective climate sensitivity (EffCS) of 6.7 K and an excessive rise in recent global-mean surface air temperatures (GMSAT), limiting its suitability for some climate applications. This motivated the development of an alternative GC5-based configuration with EffCS constrained to lie within the IPCC Sixth Assessment Report "very likely" range, improved simulation of historical temperatures, and acceptable climatological performance. This new configuration, called UKCM2-LL, will form part of the UK's submission to CMIP7.

We describe a two-stage methodology used to develop UKCM2-LL. First, a 503-member atmosphere-only perturbed parameter ensemble (PPE) of GC5 variants was used to train statistical emulators that predict climatological performance metrics and atmosphere-only feedbacks. These emulators were then used to generate 41 candidate configurations predicted to have substantially reduced EffCS relative to GC5. Second, coupled preindustrial control and abrupt $4\times CO_2$ experiments were used to evaluate the candidates against large-scale climatological metrics and to diagnose EffCS, progressively narrowing the set of viable configurations through an expert-led evaluation process. Final fine-tuning of a single candidate was performed manually using the coupled experiments and was informed by parameter sensitivity information derived from the PPE. The resulting UKCM2-LL configuration has an EffCS of 3.6 K and exhibits an improved simulation of historical temperatures relative to GC5. However, achieving this lower EffCS required a degradation in climatological performance, reflecting a structural constraint of the GC5 model.

This work demonstrates the value of PPE-based approaches for systematically exploring such structural constraints, and for parameter tuning during model development. We discuss potential improvements to the methodology and consider the implications of explicitly constraining climate sensitivity for future model development and multi-model ensemble diversity.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

1. Introduction

The Global Coupled model version 5 (GC5) of the Met Office Unified Model (UM) was developed using an established approach to model development (Willett et al., in prep.; Xavier et al., in prep.). GC5 builds on the GC3/3.1 (Williams et al., 2018) and GC4 (Xavier et al., 2024) configurations with substantial improvements to convection, cloud, and land-surface processes, alongside an updated ocean model component (Guiavarc'h et al., 2025) and a new sea ice model, SI^3 (Blockley et al., 2024). These developments have led to improvements in the large-scale circulation, a better representation of the diurnal cycle of convection, more realistic monsoon behaviour, reduced Southern Ocean SST and sea-ice biases, and improved NWP forecast skill.

However, GC5 also exhibits a very strong emergent response to CO_2 forcing. The lower resolution version of GC5 has an effective climate sensitivity (EffCS) of 6.7 K, which substantially exceeds the 5 K upper bound of the "very likely" range for equilibrium climate sensitivity (ECS) assessed in the IPCC AR6 WG1 report (Intergovernmental Panel on Climate Change (IPCC), 2023). GC5 also simulates a rapid increase in global-mean surface air temperature ("GMSAT") over 1985-2015 of 0.44 K $decade^{-1}$, which is considerably higher than the observed trend of 0.19 K $decade^{-1}$, based on HadCRUT5 (Morice et al., 2021).

Such a strong CO_2 response affects the usability of GC5 for longer-timescale climate applications. For example, a similar (but reduced) discrepancy in the late-20^{th}/early-21^{st} century GMSAT trend is present in HadGEM3-GC3.1, which underpins the UK's national climate projections (Murphy et al., 2018). This was recognised both within that study and by others (Hausfather et al., 2020), who noted that such trends pose challenges for applying these projections in climate adaptation work. A poor simulation of historical temperature can also be a critical limitation for decadal prediction, where systematic model biases that develop after initialisation (model "drift") can significantly degrade the accuracy of predicted climate anomalies. And the extremely high EffCS for GC5 also limits its use as the physical component of Earth-system models (e.g. UKESM1.1; Mulcahy et al., 2023). For example, carbon-cycle studies with Earth-system models provide a means to explore feedbacks that are not represented in physical-only models. But coupling to a physical model with such a high EffCS means that even small additional uncertainties can easily generate implausible simulations, limiting the range of feedbacks that can be explored.

Given these potential limitations, we initiated a project to explore alternative configurations of GC5, with more moderate EffCS value that could be better suited to these longer-timescale climate applications. This ultimately led to the UKCM2-LL configuration, which has an EffCS of 3.6 K. This paper describes the methodology by which UKCM2-LL was developed, through the tuning of GC5 model parameters. A companion paper by Bodas-Salcedo et al. (in prep.) provides a more detailed assessment of the climatological performance of UKCM2-LL. Both studies focus only on lower resolution simulations and so

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

we drop the term "-LL" and refer to the configuration as "UKCM2". Together, these two papers provide the documentation for UKCM2-LL, which will form part of the UK's submission to the CMIP7 project.

The established UM development process aims to reduce model biases in present-day climate and NWP simulations by addressing known deficiencies in the model's physical parameterisations. However, the model's emergent response to external forcing is not explicitly considered during this process. In contrast, the development of UKCM2 followed a different approach: perturbed parameter ensembles (PPEs) were used to tune GC5 model parameters to ensure that its emergent response to forcings satisfied the following criteria:

1. an EffCS within the IPCC AR6 very likely range of 2 – 5 K; and
2. an improved simulation of historical surface temperatures relative to GC5.

In addition, the adoption of the final UKCM2 configuration required that an acceptable level of climatological performance was maintained. These criteria were defined following an internal consultation exercise involving potential users of alternative GC5 configurations across multiple applications (see Bodas-Salcedo et al., in prep. for details of this consultation exercise).

In practice, only the EffCS and climatological performance were actively considered during the development of UKCM2. Because of the high computational cost, only a very small number of historical simulations were conducted, and metrics of historical surface temperature (e.g. Bodas-Salcedo et al., 2023) were therefore not used as development targets. Instead, it was assumed that lowering the EffCS would improve the simulation of historical temperature, while acknowledging that any such improvement could arise from compensating errors associated with incorrectly simulated SST patterns (Andrews et al., 2022; Armour et al., 2024). An evaluation of UKCM2's historical temperature simulation *was* conducted prior to its formal adoption and is described in Bodas-Salcedo et al. (in prep.).

Although there have been examples of similar approaches, whereby models are tuned or selected based on their emergent properties (Boucher et al., 2020; Mauritsen et al., 2019), this is not common practice in climate model development (Hourdin et al., 2017; Schmidt et al., 2017) and it is notably different to the established UM development process. We discuss the potential implications of this approach, e.g. for the diversity in multi-model ensembles, in Sect. 5.2.

The use of PPEs was another novel aspect of the development of UKCM2. PPEs of Hadley Centre models have been used in a wide range of studies, including uncertainty estimation (Murphy et al., 2004, 2007; Sexton et al., 2012), parameter sensitivities and constraints (Johnson et al., 2018, 2020; Tsushima et al., 2020), causal-based process understanding (Sexton et al., 2024; Yamazaki et al., 2021), investigations of structural model biases (Furtado et al., 2023; McNeall et al., 2016; Regayre et al., 2023; Rostron et al., 2025) and the development of national climate projections (Murphy et al., 2018). More

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

broadly, PPEs are increasingly being used in model development for model calibration and tuning, supported by advances in computationally efficient climate model emulators and parameter-optimisation techniques (Bonnet et al., 2025; Elsaesser et al., 2025; Lee et al., 2011; Watson-Parris et al., 2021; Williamson et al., 2013; Yarger et al., 2024).

To date, however, PPEs have not been used routinely in the development of the UM. For UKCM2, we adopted a PPE-based method that comprised two stages. The first stage used atmosphere-only experiments to explore alternative GC5 parameter settings and to optimise a set of promising "candidates". The second stage used coupled-model experiments to test the large-scale climatological performance and climate sensitivities of a smaller subset of these candidates. This two-stage process was required because GC5 is a relatively slow model (its speed is similar to HadGEM3-GC3.1-LL, which is compared to other CMIP6 models in Acosta et al., 2024). Running a large PPE of coupled simulations would therefore have been prohibitively computationally expensive.

The atmosphere-only based optimisation followed the method of Peatier et al. (2022), who used a PPE of the atmospheric component of CNRM-CM6-1, to produce alternative model configurations spanning a range of net feedback parameter values, while minimising a climatological error metric. We adopted a very similar approach, using a PPE of the atmospheric component of GC5 to optimise 41 UKCM2 candidates which were expected to have low EffCS. These candidates were subsequently evaluated using coupled pre-industrial control and abrupt-4xCO2 experiments. In the later stages of development, parameter sensitivity information derived from the PPE was used, in part, to guide the manual fine-tuning required to produce the final UKCM2 configuration.

The remainder of this paper is organised as follows: Sect. 2 describes the models and methods used in the development of UKCM2, including the GC5 PPE and its emulation, the atmosphere-only and coupled experiments conducted, and the metrics used to evaluate climatological performance and climate sensitivity. Sect. 3 outlines the methodology for developing UKCM2, including the atmosphere-only-based parameter optimisation (Sect. 3.1), the subsequent evaluation of UKCM2 candidates in coupled experiments (Sect. 3.2), and the fine-tuning used to finalise UKCM2 (Sect. 3.3). In Sect. 4 we analyse the key differences in parameter values between GC5 and UKCM2, and in Sect. 5 we discuss potential improvements to the methodology, how lessons learned from the development of UKCM2 could be incorporated into future UM model development, and the broader implications of employing similar approaches in the development of climate models.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

2. Models and methods

2.1 GC5

The starting point for the development of UKCM2 was the Global Coupled model version 5 (GC5) of the Met Office Unified Model (UM), which was developed using the established UM development process in collaboration with national and international partners. GC5 comprises the Global Atmosphere and Land ("GAL9" Willett et al., in prep.) and Global Ocean and Sea Ice components ("GOSI9"; Blockley et al., 2024; Guiavarc'h et al., 2025), with coupling provided by the OASIS3-MCT coupler (Craig et al., 2017). Here, we describe the main developments for GC5, relative to GC3.1 which formed the basis for the UK's submission to CMIP6 (Williams et al., 2018). Further details can be found in the aforementioned documentation papers, and an evaluation of GC5 is given in Xavier et al. (in prep.).

GAL9 incorporates developments in most areas of the model physics relative to GA7.1 (used in GC3.1; Walters et al., 2019). These include: the introduction of a prognostic-based convective entrainment rate into the mid-level convection scheme (Willett et al., 2025, in prep.; Willett and Whitall, 2017); alignment of mid- and deep-level convection (with deep convection subsequently switched off); a new parameterisation for riming (Furtado and Field, 2017); bimodal cloud initiation (Weverberg et al., 2021a, b); and several changes to improve the large-scale circulation (Williams et al., 2020). Updates to the aerosol scheme include an improved parameterisation of SO_2 dry deposition (Hardacre et al., 2021) and a new representation for marine organic aerosols (Gantt et al., 2011, 2015). For the JULES land component (Best et al., 2011), changes were made to the calculation of surface albedo to improve agreement with observations and to remove the previous scaling to climatology. The soil hydrology parameterisation was also updated, improving drainage to lower soil layers.

The ocean component of GC5 is based on version 4.0.4 of the NEMO ocean model (Guiavarc'h et al., 2025; Madec et al., 2019). It includes a new equation of state (IOC et al., 2010); uses adaptive-implicit vertical advection to enable significant increases in the model time step; and employs a higher-order tracer advection scheme to reduce numerical mixing (Guiavarc'h et al., 2025). GC5 includes a new sea-ice component, based on NEMO's native SI^3 model (Sea Ice modelling Integrated Initiative; Vancoppenolle et al., 2023). The GC5 sea-ice component and its coupling processes are documented in Blockley et al. (2024).

The development of UKCM2 was carried out using lower-resolution climate simulations, with a horizontal resolution of approximately 135 km in the atmosphere (midlatitudes) and 1° in the ocean (known as "N96ORCA1"). UKCM2 will form part of the UK's submission to the CMIP7 project, and there the naming convention "-LL" is used to reflect the lower resolutions used for the atmosphere and ocean model components. For the remainder of this paper, however, we drop these naming conventions and simply use the names "GC5" and "UKCM2" for these configurations.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

2.2 The GC5 PPE

The development of UKCM2 from GC5 was based primarily on changes to parameters in GAL9. Prior to this, the parameter space of GC5/GAL9 was explored using a 503-member PPE, and this PPE was used extensively in the subsequent development of UKCM2. The PPE was created slightly before the final GC5 configuration was defined, meaning that it did not include a fix to address unrealistic surface temperatures from very occasionally developing for a single timestep under highly specific conditions (see GMED ticket #604 in Willett et al., in prep.). This change was found to have a negligible effect on lower-resolution AMIP simulations, and so we refer to the PPE as the "GC5 PPE".

The GC5 PPE was built by making simultaneous perturbations to 73 independent model parameters across seven parameterisation schemes (the boundary layer, gravity wave drag, convection, cloud microphysics, cloud/cloud radiation, aerosols and land). The majority of the parameters had already been included in earlier UM-based PPEs (Rostron et al., 2025; Sexton et al., 2019, 2021). Additional parameters, and their prior distributions were elicited from model experts to target key uncertainties in GC5's simulation of present-day and future climate, with a focus on generating a spread of outcomes in radiative feedbacks. A summary of the key parameters used in the development of UKCM2 is given in Table 1.

500 of the members were defined using Latin hypercube sampling of the prior distributions, which provides an efficient way to explore the effects of perturbing the parameters across their ranges, and the interactions between them. Three more members were added to this: one used the default GC5 parameter settings (i.e. "GC5"), and two were defined using the modal and median parameter values from the elicited parameter distributions as these have been shown to represent PPE mean behaviour reasonably well (Rostron et al., 2025). (In practice, many of the distributions have flat tops so the mode is not uniquely defined. In these cases, we use the central value from the flat part of the distribution to define the mode.)

2.3 Experiments

2.3.1 Atmosphere-only experiments

The 503 GC5 PPE members, together with a subsequent set of 41 UKCM2 candidate configurations (Sect. 3.1), were tested using two five-year atmosphere-only experiments. These experiments were designed to assess the present-day climatological performance and radiative feedbacks. Their design closely follows the CMIP6 *amip* and *amip-future4K* experiments (Eyring et al., 2016; Webb et al., 2017), but with modifications as described below. In previous PPE studies using UM-based models, these experiments were referred to as "ATMOS" and "SSTFuture" (e.g. Sexton et al., 2021). However, despite the departures from the standard CMIP protocol, we refer to them here as "*amip*" and "*amip-future4K*" for simplicity. We also use the term "*amip* experiments" when referring to both experiments collectively.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Table 1 Key parameter perturbations used in the development of UKCM2, from the *amip*-based optimisation and/or the coupled-model based fine-tuning.

Scheme	Parameter name	Description	GC5 value	UKCM2 value	Range tested
Convection	ent_fac_sh	Scaling factor for entrainment rate for shallow convection	1.000	2.253	0.33 - 3.00
	thpixs_mid	Size of the initial buoyancy perturbation for mid-level convection prior to level pressure thickness scaling.	0.500 K	0.762 K	0 – 1 K
	c_mass_sh	Scaling factor for the cloud-base mass-flux for shallow convection	0.030	0.058	0.01 – 0.09
Cloud and cloud radiation	dbsdtbs_turb_0	Parameter controlling the rate at which unresolved sub-grid motions mix clear and cloudy air and hence removes liquid condensate and evaporates liquid cloud fraction.	1.50×10^{-4} s^{-1}	6.25×10^{-4} s^{-1}	$0 – 10 \times 10^{-4}$ s^{-1}
	two_d_fsd_factor	Scaling factor applied to empirically derived one-dimensional cloud condensate variability to make it represent two-dimensional variability.	1.65	1.27	1 – 2
	dp_corr_strat	Parameter controlling the amount of vertical overlap between clouds in the sub-column in the cloud generator used to calculate the radiative impact of clouds.	10000 Pa	35851 Pa	2500 – 40000 Pa
Cloud microphysics	m_ci_rp	Scaling factor for the cloud ice fall speed	1.000	1.393	0.3 – 3.0
Aerosols	sea_salt_ems_scaling	Scaling factor for emissions of sea salt	1.0	1.4	0.12 – 4.00
	ps_natl_dms_emiss	Scaling factor for emissions of natural DMS	1.0	1.5	0.5 – 2.0
	marine_pom_ems_scaling	Scaling factor for emission of primary marine organic aerosol	1.0	1.4	0.25 – 4.00

Compared with the CMIP6 *amip* protocol, the experiment conducted here was shorter (5 years instead of 36) and used prescribed daily sea surface temperatures (SSTs) and sea ice concentrations from the HadISST2 observational dataset (Titchner and Rayner, 2014), rather than the standard CMIP6 boundary conditions (Durack et al., 2022). The simulations spanned 1

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

September 2004 to 1 December 2009, with the first three months used as a spin-up period which was excluded from all subsequent analysis.

The *amip-future-4K* experiment follows the CMIP6 protocol by perturbing the *amip* boundary conditions with a prescribed 4 K global-mean SST warming pattern derived from CMIP3 coupled model simulations (Webb et al., 2017). Sea ice concentrations were held fixed relative to the *amip* experiment.

For both the *amip* and *amip-future-4K* experiments, we also ran 32 realisations of the unperturbed GC5 configuration, which differed only in the random seed used to initialise the stochastic schemes used in GC5. We refer to these as the "stochastic" experiments. These simulations are used to characterise the atmospheric internal variability in GC5, under the implicit assumption that this variability is independent of the model parameter values.

The results from these experiments were used primarily to train emulators (Sect. 2.4) for the *amip* feedback parameter, and for a set of variables used to assess the present-day *amip* performance (Sect. 2.5.1 and 3.1). These emulators were subsequently used throughout the development of UKCM2.

2.3.2 Coupled experiments

The most promising candidates for UKCM2, identified using the *amip* experiments, were subsequently evaluated using coupled pre-industrial control and abrupt-4xCO2 experiments. These coupled simulations were conducted to assess the climatological performance and the EffCS of each candidate.

The pre-industrial control experiment followed the standard CMIP6 protocol, with all relevant forcings - well-mixed greenhouse gases, ozone, solar irradiance, tropospheric and stratospheric aerosols, and land surface properties - set to their 1850 values (Eyring et al., 2016; Menary et al., 2018). The simulations for each candidate were initialised using the EN4 ocean reanalysis, averaged over the years 1950-1954 (Good et al., 2013).

Twenty-two candidates were tested using the pre-industrial control experiment, and each was integrated for a minimum of 44 years to allow the atmosphere, land surface, ocean mixed layer, and sea ice to adjust before an initial assessment of their climatological performance. The simulations for a subset of the best-performing candidates were extended, with at least of 174 additional years of integration (Sect. 3.2). A single "proto-UKCM2" candidate was then fine-tuned, during which parameter adjustments were applied at later points in the pre-industrial control simulation (i.e. the simulation was not re-initialised from the EN4 ocean reanalysis each time a parameter was adjusted; Sect. 3.3).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

We refer to the initial 44-year phase as the '*picontrol-spinup*' experiments and the subsequent phases as the '*picontrol*' experiments, following CMIP6 terminology.

The duration of these experiments was considerably shorter than is typical for CMIP *picontrol* simulations, and the candidates were therefore not expected to reach equilibrium during the development. This was a pragmatic decision, reflecting the computational expense of running coupled simulations for multiple candidates. Although the ultimate objective was to develop a configuration capable of producing a stable long-term simulation, equilibrium was not required during the development itself. Instead, the aim was to identify the most promising parameter settings by iteratively discarding candidates that were least likely to result in an acceptable final configuration. Running each candidate and fine-tuning test to near equilibrium would have been inefficient and was therefore not attempted. As a result, particular care was required to account for model drifts when evaluating the performance of the candidates.

A small number of *abrupt-4×CO2* experiments were conducted to assess the EffCS of the most promising candidates (Sects. 2.5.2, 3.2 and 3.3). These experiments also followed the CMIP6 protocol, whereby a simulation branches from the *picontrol* and its CO_2 concentrations are instantaneously quadrupled relative to their 1850 value, while all other pre-industrial forcings remain unchanged. Each of the *abrupt-4×CO2* experiments was integrated for 150 years.

2.4 Emulators

To make systematic use of the large number of *amip* simulations (Sect. 2.3.1), we employed statistical emulators. These emulators are statistical models, trained on climate model output, that map climate model parameter values to simulated quantities. They allow us to estimate the expected model output at any point within the explored parameter space, while reducing the influence of internal variability.

We trained emulators using only the *amip* and *amip-future4K* experiments because the limited number of candidates evaluated with the coupled experiments was not sufficient to train reliable emulators. The *amip* emulators were used extensively in the development of UKCM2 - to optimise an initial set of candidates (Sect. 3.1) and to quantify *amip* parameter sensitivities (Sects. 3.3 and 4). Further details of their construction and application are provided in those sections.

Our emulators implement the Gaussian Process methodology described in Sexton et al. (2019). In summary, a smooth, continuous model across the parameter space is created by simultaneously fitting a linear model, based on a subset of the most sensitive parameters, along with a Gaussian Process model (fitted to the residuals of the linear model), which relies on optimised decorrelation length scales for each parameter. To account for internal variability in the data, the stochastic experiments were used to specify a nugget term, which reduces the chance of the emulator fitting to noise.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

2.5 Performance assessment, feedbacks and EffCS

2.5.1 *amip* performance and feedbacks

As discussed in the Introduction, we applied a method adapted from Peatier et al. (2022) to optimise and select 41 UKCM2 candidates. The optimisation was based on the climatological performance and the feedback parameter derived from the *amip* experiments. We define these here, whilst the details of the optimisation are described in Sect. 3.1.

The *amip* climatological performance was quantified using root-mean-squared errors (RMSEs) of the 5-year climatologies for 18 variables (which are listed in Table A1). This is a larger set of variables than was used by Peatier et al. (2022), chosen to capture the performance across a range of atmospheric processes and to reduce the chance of inadvertently rewarding compensating errors. One of the variables considered was an implied ocean heat transport metric ("fmassefp"), which is the energy flux potential of the net downwelling energy flux into the ocean (Pearce and Bodas-Salcedo, 2023). This variable was introduced to try and reduce adjustments when the candidates are run in coupled mode. A smaller error in fmassefp implies a pattern of surface energy flux that is closer to the observed pattern, and therefore it is expected to be more compatible with the present-day divergent ocean heat transport.

Following Peatier et al. (2022), the RMSEs were calculated using spatial fields truncated to the leading five empirical orthogonal functions (EOFs). These EOFs were defined using the 503-member GC5 PPE. Truncation at five EOFs was chosen because this captured much of the variance across the PPE while retaining good signal-to-noise in emulated predictions (Peatier et al., 2022). The observational reference fields were truncated in a consistent manner, by projecting them onto the EOF bases defined by the PPE. Peatier et al. (2022) argue that this focuses the optimisation on aspects of the real climate system that are captured by the model across parameter space. For emulated candidates, the truncated spatial field for each variable was reconstructed as the linear combination of the five EOF basis fields, weighted by the corresponding emulated EOF amplitudes.

The overall performance for each candidate was summarised using an "*amip* aggregate error" metric, defined as the mean of the 18 RMSE values, where each had first been normalised to the RMSE value for GC5. For the remainder of this paper, the *amip* aggregate error is denoted as "E_{amip}".

The observation and reanalysis datasets used to calculate the errors are given in the final column of Table A1. For the radiative fluxes we used the CERES Energy Balanced and Filled (EBAF) Edition4.1 dataset (Loeb et al., 2018; NASA/LARC/SD/ASDC, 2019), while ERA-Interim reanalysis was used for temperature, relative humidity and wind variables (Dee et al., 2011). For precipitation, we used the Global Precipitation Climatology Project (GPCP) monthly analysis version 2.3 (Adler et al., 2018); and for surface pressure we used the HadSLP2 dataset (Allan and Ansell, 2006). For fmassefp we used DEEP-C v5 (Liu et al., 2017).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Global-mean radiative feedbacks were diagnosed from the *amip-future-4K* experiment as the ratio of the change in five-year global-mean top-of-atmosphere (TOA) flux to the corresponding change in global-mean surface air temperature (relative to *amip*). Feedbacks were calculated for several TOA flux components; however, we focused primarily on the all-sky net TOA feedback, hereafter denoted λ_{amip} (the *amip* feedback parameter).

Atmospheric feedbacks are a key component of the equilibrium climate response to CO_2 forcing in coupled experiments (Qin et al., 2022; Ringer et al., 2014), and λ_{amip} was used to provide information on the likely EffCS of the candidates during the *amip*-based stage of the development of UKCM2 (Sect. 3.1). However, the *amip-future-4K* experiment is insensitive to several coupled feedbacks (e.g. the sea-ice-albedo feedback) and does not quantify the CO_2 forcing. Consequently, accurate assessments of the EffCS of candidates was only possible using the coupled *abrupt-4×CO2* experiments.

2.5.2 Coupled-assessment metrics and EffCS

The suitability of UKCM2 candidates during the coupled-model-based development was evaluated using target ranges for 11 large-scale metrics assessed from the *picontrol-spinup* and *picontrol* climatologies, along with a target range of 2 – 5 K for the EffCS. The 11 metrics were: global-mean net, net shortwave and outgoing longwave TOA fluxes ("net TOA", "netSW" and "OLR", respectively); global-mean surface air temperature ("GMSAT"); the Atlantic meridional overturning circulation ("AMOC") transport at 26.5°N; the Antarctic Circumpolar Current ("ACC") transport through Drake Passage; Arctic and Antarctic sea ice area means and annual ranges; and Arctic winter sea ice volume. These metrics are summarised in Table 2.

We used these "coupled-assessment metrics" during the development to assess the stability and large-scale climatological performance of the candidates, across the different model components. The target ranges were defined to ensure that the final configuration provided a plausible large-scale climate and would be expected to perform acceptably in the subsequent, more detailed assessment described in Bodas-Salcedo et al. (in prep.).

The use of target ranges, rather than the RMSE-based metrics used in the *amip* experiments, also reflected a shift in our development approach during the coupled testing. While the *amip*-based optimisation relied on a single, well-defined cost-function (E_{amip}) within an optimisation algorithm (Sect. 3.1), the coupled development followed a more subjective, expert-led process typical in model development (Sects. 3.2 and 3.3). The use of target ranges provided the flexibility required to explore the performance trade-offs that were expected in defining a final UKCM2 configuration.

A key aim in the development of UKCM2 was to achieve a net TOA flux close to zero, in order to minimise systematic accumulation or loss of heat by the ocean. Accordingly, a target range of ±0.15 W m^{-2} was adopted for the net TOA flux.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

During earlier stages of the coupled development (Sect. 3.2), a wider target range of ±1.0 W m^{-2} was used to avoid rejecting candidates before fine-tuning had been attempted.

For the remaining variables, the target ranges were chosen to reflect observational or reanalysis estimates, and their associated uncertainties. In many cases these estimates were based on present-day observations, and adjustments were therefore required to convert them to values appropriate for the *picontrol* climatologies. For netSW and OLR, we based the target ranges on observed estimates from Loeb et al. (2018) and L'Ecuyer (2015), with present-day to pre-industrial adjustments taken from the UKESM1.1 model under CMIP6 forcings (Mulcahy et al., 2023). This resulted in a target range of 239 – 241 W m^{-2} for both netSW and OLR. For the GMSAT we adopted a target range of 286 – 287 K, based on an estimate for the 1850 – 1900 pre-industrial average reported in the Global Climate Highlights 2023 report (Copernicus Climate Change Service (C3S), 2024). A range of 1 K was chosen to reflect the challenges in establishing an absolute temperature for the pre-industrial period, given the uncertainties associated with poor spatial coverage and the lack of standardised measurement techniques. As with the net TOA target range, a wider range of 285 – 288 K was adopted during earlier phases of the coupled development (Sect. 3.2).

For AMOC, we adopted a conservative target range of 11 – 19 Sv. This incorporated estimates from the Estimating the Circulation and Climate of the Ocean project (ECCO) and the RAPID array (Bonan et al., 2025), with an adjustment for present-day to pre-industrial conditions based on typical anomalies from CMIP6 models (noting that these anomalies are thought to be overestimated; Menary et al., 2020; Robson et al., 2022). For the ACC, we adopted a similarly conservative target range of 110 – 190 Sv, spanning both the canonical estimate of 134 Sv (with an uncertainty up to 27 Sv; Cunningham et al., 2003; Whitworth and Peterson, 1985) and a more recent higher estimate of 173.3 Sv (Donohue et al., 2016). No present-day to pre-industrial adjustment was applied to the ACC target range.

For the sea ice variables, the target ranges for sea ice areas 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. For the annual means, the target ranges were 11.5 – 13.0 x 10^6 km^2 for the Arctic and 12.0 – 16.0 x 10^6 km^2 for the Antarctic, while for the annual ranges the targets were 7.5 – 9.5 x 10^6 km^2 for the Arctic (March – September) and 7.5 – 9.5 x 10^6 km^2 for the Antarctic (September – February). The Arctic winter (March) sea-ice volume target was informed by the Pan-Arctic Ice-Ocean Modeling and Assimilation System (PIOMAS) sea ice volume reanalysis (Schweiger et al., 2011). This was also adjusted to a pre-industrial value using expert judgment, resulting in a target range of 30 – 45 x 10^3 km^3.

EffCS values were diagnosed following the method of Gregory et al. (2004). This regresses global-mean net TOA flux anomalies (ΔN) against corresponding GMSAT anomalies (ΔT) from the 150-year *abrupt-4×CO2* experiment. Anomalies are used to remove model drifts, which are estimated by fitting linear trends to N and T over the coincident period in the

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

corresponding *picontrol* simulation. The coupled feedback parameter, λ_{4xCO2}, is estimated using the slope of this regression, while the effective radiative forcing for a doubling of CO_2 is estimated as half of the regression intercept ($EffF_{4xCO2}/2$). The EffCS is then computed as $EffCS = -EffF_{4xCO2}/(2 \times \lambda_{4xCO2})$ following Andrews et al. (2019).

Following an internal consultation exercise with a range of potential users of a lower-EffCS configuration of GC5, a target range of 2 – 5 K was adopted, based on the "very likely" range for equilibrium climate sensitivity (ECS) assessed in the IPCC AR6 WG1 report (Intergovernmental Panel on Climate Change (IPCC), 2023). This was felt to be an achievable aim that would satisfy the requirements for multiple applications, as discussed in the Introduction.

3. Development of UKCM2-LL

Initial analyses of the GC5 PPE showed that typical values of λ_{amip} were between -1.43 and -0.74 W m^{-2} K^{-1} (2.5 - 97.5 percentiles; see Fig. 1), with a minimum value of -1.55 W m^{-2} K^{-1}. For context, the λ_{amip} value for GC5 (with an EffCS of 6.7 K) was -1.22 W m^{-2} K^{-1}. We also estimated that the upper limit of the target range for EffCS (5 K) corresponded to a λ_{amip} value of -1.37 W m^{-2} K^{-1}, under the simplifying assumption that, for GC5, the missing coupled feedbacks and the CO_2 forcing do not vary with the PPE parameter perturbations.

These results were a key motivation for the development of UKCM2. They indicated that, although most GC5 PPE members would exhibit high EffCS values, alternate parameter combinations could potentially produce configurations with "acceptable" EffCS values i.e., within the target range of 2 – 5 K.

At this stage, however, it was not clear whether such alternate configurations would produce plausible *amip* simulations, or how they would behave when tested in the coupled *picontrol* and *abrupt-4×CO2* experiments. The following sections describe how these questions were addressed through the development of UKCM2, including:

- an automated "*amip*-based optimisation" stage, where the GC5 PPE was used to generate 41 UKCM2 candidates that were predicted to have an acceptable EffCS, while maintaining E_{amip} as far as possible (Sect. 3.1);
- an expert-led "coupled-model evaluation" stage, where the candidates were evaluated using the large-scale coupled-assessment metrics and EffCS (defined in Sect. 2.5.2), ultimately resulting in a single "proto-UKCM2" configuration (Sect. 3.2); and,
- an expert-led "fine-tuning" stage, where model parameter values were adjusted to achieve acceptable performance in the large-scale coupled-assessment metrics (Sect. 3.3).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

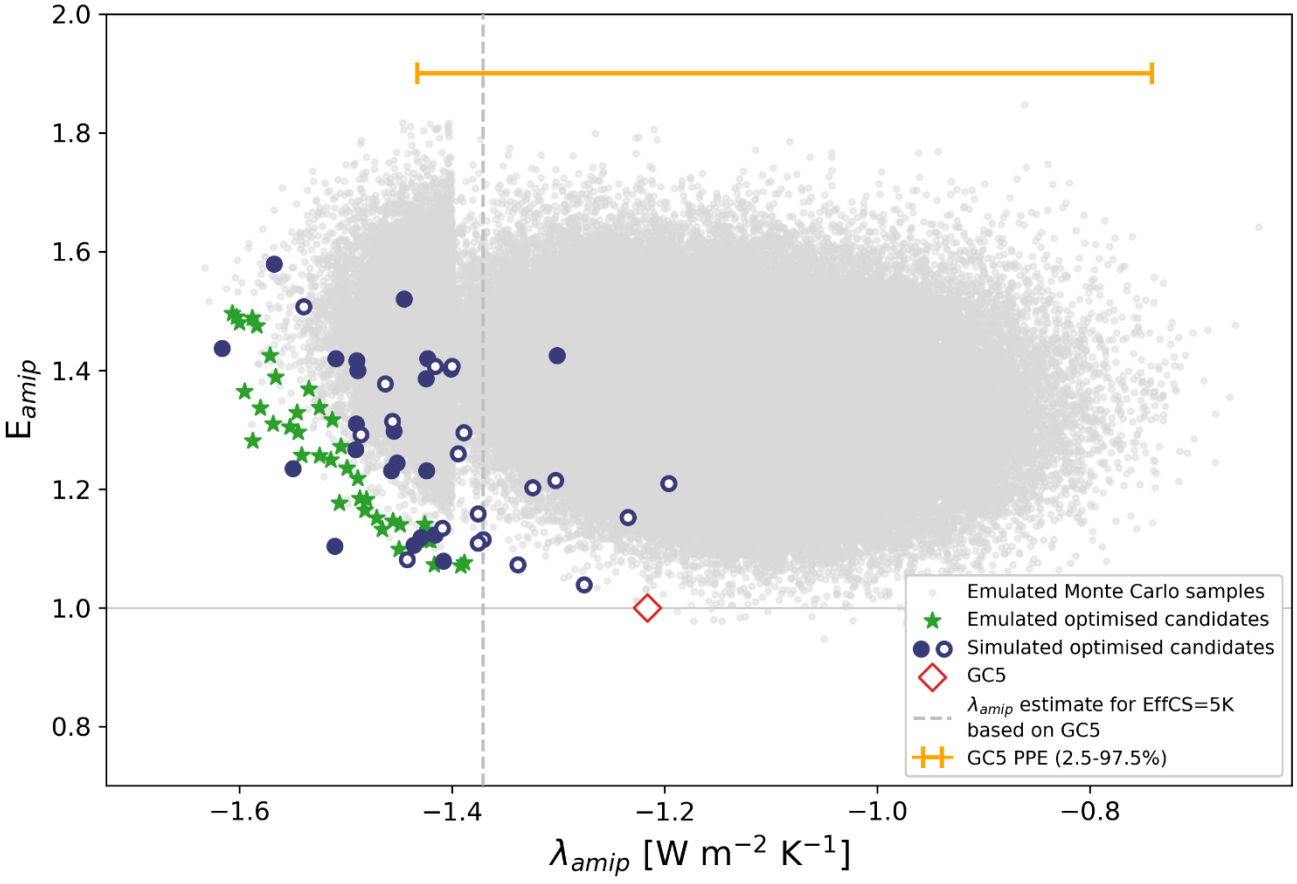

Figure 1: E_{amip} vs λ_{amip}. Grey points denote the emulated Monte Carlo samples, and the green stars are the emulated predictions for the 41 optimised UKCM2 candidates. Blue points show the corresponding simulated values for these 41 candidates from the *amip* experiments. The 22 filled blue points indicate the candidates that were subsequently tested using the *picontrol-spinup* experiment; the 19 open blue points correspond to candidates which were not tested any further. The red diamond marks the GC5 simulation (i.e. the unperturbed GC5 PPE member) which has an E_{amip} of 1 by construction. The 2.5 - 97.5 percentile range of λ_{amip} values from the GC5 PPE is shown in orange. The dashed vertical line is an estimate of λ_{amip} for EffCS=5 K.

3.1 *amip*-based optimisation

We used information from the *amip* simulations of the 503 GC5 PPE members to generate a set of 41 UKCM2 candidates which could subsequently be tested using coupled experiments. Our approach closely followed that of Peatier et al. (2022), and consisted of the following steps:

1. Emulators were trained to predict λ_{amip} and the *amip* aggregate error (E_{amip}) for arbitrary GC5 parameter combinations.
2. Large Monte Carlo samples of GC5 parameter combinations were generated, and predictions of λ_{amip} and E_{amip} were made using the emulators.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

3. Forty-one promising UKCM2 candidates were selected from these samples and subsequently optimised using a Nelder-Mead algorithm.
4. The 41 optimised candidates were tested using *amip* experiments.

3.1.1 Emulated predictions for λ_{amip} and E_{amip}

Emulators for λ_{amip} and E_{amip} were trained using the 503-member GC5 PPE, enabling predictions across the GC5 parameter space. For E_{amip}, separate emulators were trained for the amplitudes of the leading five EOFs for each of the 18 variables listed in Table A1. Emulated predictions of the 18 RMSEs, and E_{amip}, were then constructed as described in Sect. 2.5.1. Predictions of λ_{amip} and E_{amip} across the GC5 parameter space were made by evaluating two large Monte Carlo samples, each containing several million parameter sets. To reduce computational costs, only the 37 most influential parameters (from the original 73) were perturbed. A parameter was perturbed only if it explained at least 3% of the variance across the PPE for at least three of the 18 RMSEs contributing to E_{amip}. Sample values for these 37 parameters were drawn from their prior distributions (Sect. 2.2), while the remaining 36 parameters were fixed at their default GC5 values.

Several constraints were applied to both Monte Carlo samples. First, emulated predictions of global-mean net TOA flux were required to be between -2 and +3 W m^{-2}, to exclude parameter combinations likely to produce coupled *picontrol* simulations that were substantially out of radiative equilibrium. Second, a constraint was applied to emulated predictions of Atlantic water flux, which was calculated as the 5-year mean of precipitation plus run-off minus evapotranspiration, summed over the Atlantic basin. This was constrained to be between -1.0 and -0.2 Sv (Ganachaud and Wunsch, 2003); again, this broad range was chosen to exclude samples that were clearly implausible.

A further constraint was applied to remove samples whose emulated predictions lay outside the convex hull of the 503 PPE simulations, for the 18 emulated variables. This was intended to mitigate the impact of extrapolation, by restricting predictions to regions of parameter space sampled by the PPE simulations. For the second Monte Carlo sample, an additional constraint required λ_{amip} to be less than -1.4 W m^{-2} K^{-1}. This was done to provide denser sampling in the most promising regions of parameter space, which were expected to produce EffCS values less than 5 K. After applying all of the constraints, the two samples contained a combined total of 124884 parameter sets. The resulting emulated predictions of λ_{amip} and E_{amip} are shown in grey in Fig. 1.

For λ_{amip} values below -1.4 W m^{-2} K^{-1}, the lower envelope of these predictions shows that the minimum achievable E_{amip} increases as λ_{amip} decreases. This highlights a key structural constraint of the GC5 model: within the sampled parameter space, achieving lower λ_{amip} (and hence lower EffCS) requires degradations in present-day climate performance. This trade-off between EffCS and climatological performance was a key consideration in the subsequent development of UKCM2.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

3.1.2 Optimising 41 UKCM2 candidates

Using the constrained Monte Carlo samples, we generated 41 UKCM2 candidates using a two-stage optimisation process. First, parameter sets with predicted λ_{amip} between -1.64 (the minimum) and -1.41 W m^{-2} K^{-1} were split into 22 equally spaced λ_{amip} bins. Within each bin, the two parameter sets with the lowest predicted E_{amip} were selected. The lowest three bins contained only one sample each, giving a total of 41 selected parameter sets.

In the second stage, these 41 parameter sets were optimised using a Nelder-Mead algorithm, with E_{amip} acting as the cost function. The constraints applied to the Monte Carlo samples – for global-mean net TOA, Atlantic water flux and the convex hull – were also used in this optimisation. In addition, the optimised λ_{amip} for each candidate was required to remain within the bounds of its original bin. The resulting 41 optimised UKCM2 candidates are shown as green stars in Fig. 1.

Hereafter, these candidates are uniquely identified using the labels “p01” to “p41”.

After UKCM2 was finalised, we identified three errors in this optimisation process. First, we found that the normalised RMSE values for fmassefp were not as large as intended (possibly due to an issue with the values used for the observations), meaning that this metric effectively did not constrain the resulting parameter values. We tested the impact of correcting this error and found that for most candidates it would only have affected the Nelder–Mead optimisation, where its influence was negligible. For a small number of candidates, however, it would have affected the initial selection and led to different parameter sets being chosen in these cases.

Second, we found that although the parameter ps_natl_dms_emiss was not perturbed in the Monte Carlo samples, it *was* included in the Nelder-Mead optimisation. This had a negligible impact however, and the optimised values remained very close to the GC5 value of 1.0 (typically within 0.99 – 1.01; note this parameter was later perturbed during the coupled fine-tuning – see Table 1 and Sect. 3.3). Finally, the parameters mp_dz_scal and param_mp_tau_lim *were* perturbed in the Monte Carlo samples but were mistakenly reset to their unperturbed GC5 values at the start of the Nelder-Mead optimisation and remained fixed thereafter.

The combination of these three issues did not adversely affect the optimisation, and so the 41 candidates remained very good candidates. This is evident from the fact that they track the lower envelope of E_{amip} vs λ_{amip} values from the Monte Carlo samples in Fig. 1 (green stars compared with grey points); and ultimately these errors did not prevent a suitable final UKCM2 configuration being produced.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

3.1.3 *amip* simulations for the 41 UKCM2 candidates

The 41 optimised candidates were assessed using the *amip* and *amip-future4K* experiments. The resulting E_{amip} and λ_{amip} values are shown in Fig. 1 (open and closed blue points). While E_{amip} values were reasonably well predicted by the emulators, λ_{amip} values were less accurately predicted. The differences between the simulated and emulated λ_{amip} values had a mean and standard deviation of 0.082 ± 0.078 W m^{-2} K^{-1}, indicating the emulated values were generally biased low, and highly variable. (We note that the comparison here is not influenced by the methodological errors noted in Sect. 3.1.2).

Despite this, the results from the *amip* simulations showed the same broad structural relationship for E_{amip} and λ_{amip} that was noted in Sect. 3.1.1, i.e., that the candidates with lower λ_{amip} typically have higher E_{amip}. Notably, although GC5 has a high λ_{amip}, it performs better than all of the 41 candidates, highlighting the effectiveness of the expert-led tuning to present-day performance metrics (but not emergent responses) in the conventional UM model development process.

Importantly, 34 of the 41 candidates had λ_{amip} values lower than -1.37 W m^{-2} K^{-1}, corresponding to EffCS estimates lower than the upper bound of the target range (5 K). Based on these results, a subset of the 41 candidates was selected for further evaluation using *picontrol-spinup* simulations.

3.2 Coupled-model evaluation

The *amip* experiments were well suited to a broad exploration of the GC5 parameter space and enabled the automated optimisation of the 41 UKCM2 candidates. However, coupled experiments were required to diagnose their EffCS and to assess their performance more robustly, both for the large-scale coupled-assessment metrics considered here (defined in Sect. 2.5.2), and for the more detailed assessment described in Bodas-Salcedo et al. (in prep.). In this section, we describe how we progressively reduced the initial set of 41 candidates by evaluating them against the coupled-assessment target ranges defined in Sect. 2.5.2. This process ultimately resulted in the selection of a single "proto-UKCM2" candidate, which was subsequently fine-tuned.

In contrast to the *amip*-based optimisation, which used an automated, cost-function-driven approach, the coupled-model based development involved a greater degree of subjectivity, requiring consideration of several factors. In this "coupled-model evaluation" stage, we used the target ranges as reference points for judging performance and identifying candidates that were clearly unsuitable, with the understanding that more promising candidates could be fine-tuned later to better match the targets (Sect. 3.3). However, because the *picontrol* experiments were initialised from an observed ocean state and were unlikely to have reached radiative balance, substantial model drifts were expected. These drifts, together with the fact that the target ranges were defined for an equilibrated *picontrol* climate, had to be considered when evaluating the performance of each candidate.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

As discussed in Sect. 3.1, the structural behaviour of GC5 meant that no candidate was expected to exhibit both strong overall performance and an acceptable EffCS (Fig. 1 and Fig. 2a). Therefore, trade-offs between these characteristics were likely to be necessary. Additionally, given that climate models are often affected by structural biases (Furtado et al., 2023; McNeall et al., 2016; Regayre et al., 2023; Rostron et al., 2025; Tian and Dong, 2020), trade-offs *between* the coupled-assessment metrics were also possible, both in this phase and during the fine-tuning. This would require subjective choices about which metrics, if any, should be prioritised.

Finally, careful choices were needed to make effective use of the limited HPC and human resources available, for example determining how many candidates to test and which experiments to prioritise.

Table 2 The coupled-assessment metrics and target ranges used in the coupled-model phases of the development of UKCM2. All metrics are multi-year annual means. The GC5 and UKCM2 values are means over a 200-year period from the *picontrol* experiment: years 251 - 450 for GC5 and years 144 - 343 for UKCM2.

Assessment metric	Unit	Region	Target range	GC5 value	UKCM2 value
Net TOA flux [Net TOA]	W m^{-2}	Global	±0.15 [±1.0].	0.09	0.06
Net SW flux [netSW]	W m^{-2}	Global	239 – 241	239.6	241.2
Outgoing longwave flux [OLR]	W m^{-2}	Global	239 – 241	239.5	241.2
Surface air temperature [GMSAT]	K	Global	286 – 287 [285 – 288]	286.0	286.1
Atlantic meridional overturning circulation transport [AMOC]	Sv	26.5°N	11 – 19	14.1	14.1
Antarctic circumpolar current transport [ACC]	Sv	Drake Passage	110 - 190	131	116
Sea ice area annual mean	10^6 km^2	Arctic	11.5 – 13.0	12.9	12.0
Sea ice area annual range (March – September)	10^6 km^2	Arctic	7.5 – 9.5	8.6	8.5
Winter (March) sea ice volume	10^3 km^3	Arctic	30 – 45	50.6	40.3
Sea ice area annual mean	10^6 km^2	Antarctic	12.0 – 16.0	14.9	10.4
Sea ice area annual range (September – February)	10^6 km^2	Antarctic	12.0 – 14.0	13.6	13.8
Effective Climate Sensitivity [EffCS]	K	Global	2 – 5	6.7	3.6

To begin, the 44-year *picontrol-spinup* experiment was performed for 22 of the 41 UKCM2 candidates. The choice of these 22 candidates was based on the *amip* experiments. Fourteen were chosen because they had $\lambda_{amip} < -1.40$ W m^{-2} K^{-1} or exhibited low E_{amip} values (Fig. 1). The remaining eight were selected using less stringent criteria, although all but one still had λ_{amip} values below -1.40 W m^{-2} K^{-1}.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

The 30-year climatologies of the coupled-assessment metrics (using years 15-44 of the *picontrol-spinup* experiment) are shown in Fig. 2. Although many candidates achieved net TOA fluxes within the more relaxed target range of ±1 W m^{-2}, they typically exhibited cold biases (GMSAT < 285 K), with the component TOA fluxes lying well outside the target range of 239–241 W m^{-2} for both netSW and OLR. The sea ice metrics also showed substantial positive biases, especially for the Arctic winter volume, the Arctic annual-mean area and Antarctic annual area range. Systematic behaviour in the sea-ice area biases was also apparent for the Antarctic (Fig. 2g), indicating that compromises in the performance of these variables may be required. However, at this stage, no changes to parameters in the sea ice model (SI^3) had been explored.

Only six candidates were selected for further testing using *picontrol* simulations. A diversity of options was maintained with these six candidates (highlighted in colour in Fig. 2): p01, p05 and p22 had lower λ_{amip} values but poorer performance against the assessment target ranges (particularly for net TOA flux); whereas p08, p10 and p31 had higher λ_{amip}, but better performance across the assessment metrics.

The *picontrol* simulations for these six candidates were extended to between 218 and 557 simulated years. The climatologies for the assessment metrics are shown in Fig. 3 and are based on means over the final 50 simulated years in each case.

For some of the candidates, model drifts led to substantial changes in these assessment metrics relative to the end of the *picontrol-spinup* phase. This behaviour was expected given the initial imbalances in net TOA flux, particularly for candidates p01, p05 and p22 (Fig. 2a). In addition, shortly after the *picontrol-spinup* phase was completed, adjustments to a small number of SI^3 parameters were applied for p08, p10 and p31 in order to reduce the sea ice in these candidates. These modifications are discussed further in Sect. 3.3, where they motivated a change to the conductivity of snow on sea ice in the final UKCM2 configuration.

Preliminary fine-tuning tests for candidate p01 indicated that reducing the net TOA flux through parameter changes could only be achieved by significantly lowering both netSW and OLR, which risked adversely affecting the climatology and the EffCS. However, as this candidate adjusted towards a smaller positive bias in net TOA and larger positive biases in netSW and OLR, fine-tuning became more feasible (see the black point in Fig. 3c compared with Fig. 2c). The associated increase in GMSAT resulted in a substantial warm bias and drove a reduction in sea ice; these biases would also need to be addressed during fine-tuning.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

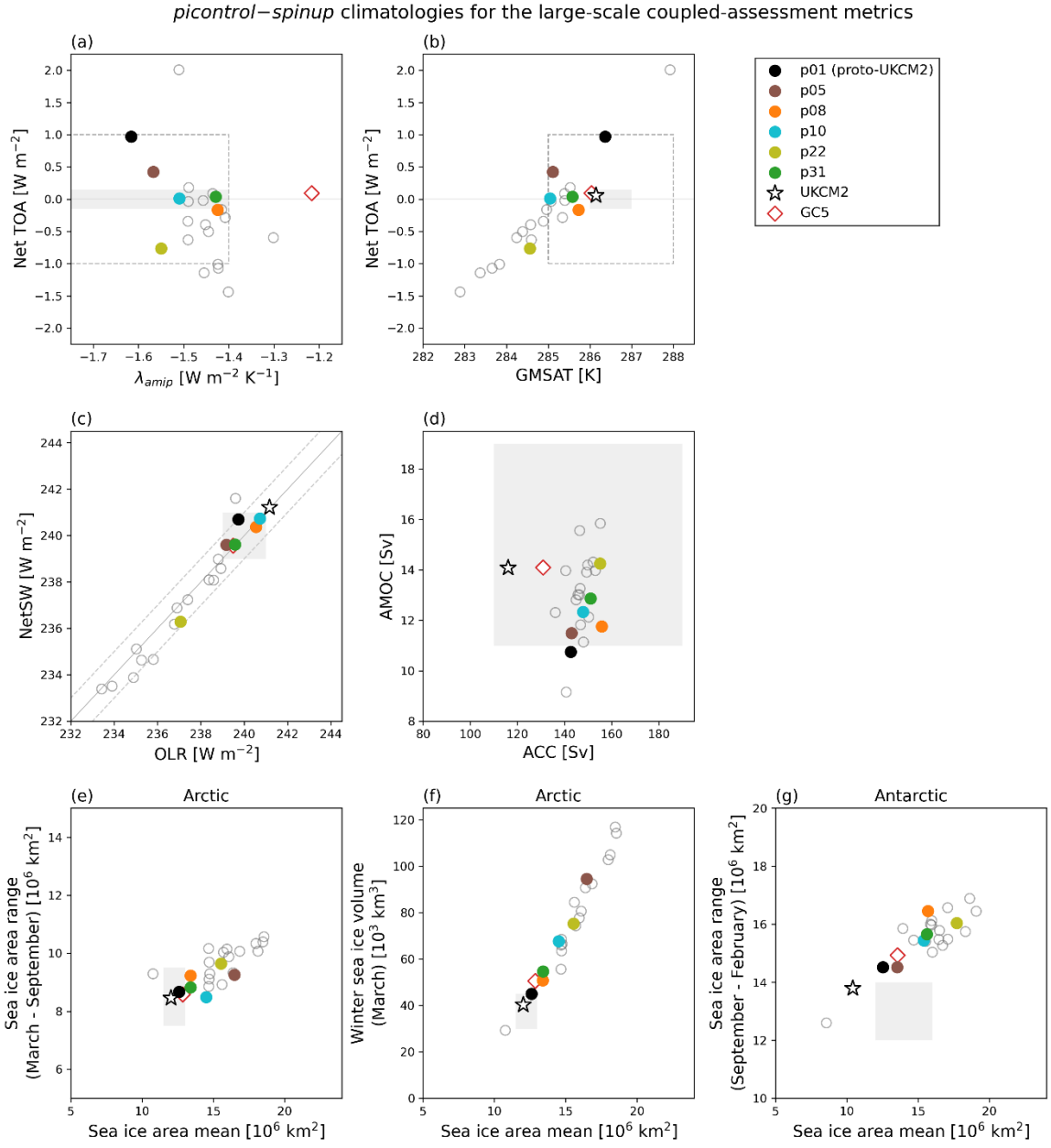

Figure 2: *picontrol-spinup* climatologies of the coupled-assessment metrics for the UKCM2 candidates (circles), with the GC5 (red diamond) and UKCM2 (black star) *picontrol* climatologies shown for reference. The mean values were calculated using simulation years 15-44 for the candidates, years 144-343 for UKCM2 and years 251-450 for GC5. The plots show (a) net TOA flux vs λ_{amip} (the latter being diagnosed from the *amip* experiments); (b) net TOA flux vs GMSAT; (c) netSW vs OLR; (d) AMOC vs ACC; (e), (g) Annual range vs annual mean for sea-ice area in the Arctic and Antarctic, respectively; and (f) the winter sea-ice volume vs annual mean sea-ice area in the Arctic. The candidates here are the same as those shown in filled blue points in Fig. 1. The circles shown in colour are the 6 candidates which were selected for extended *picontrol* experiments. The coupled-assessment target ranges are shown as grey boxes, and the wider ranges for net TOA and GMSAT are indicated as dashed boxes. In (c) lines are also shown for net TOA values of -1, 0 and +1 W m^{-2}.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

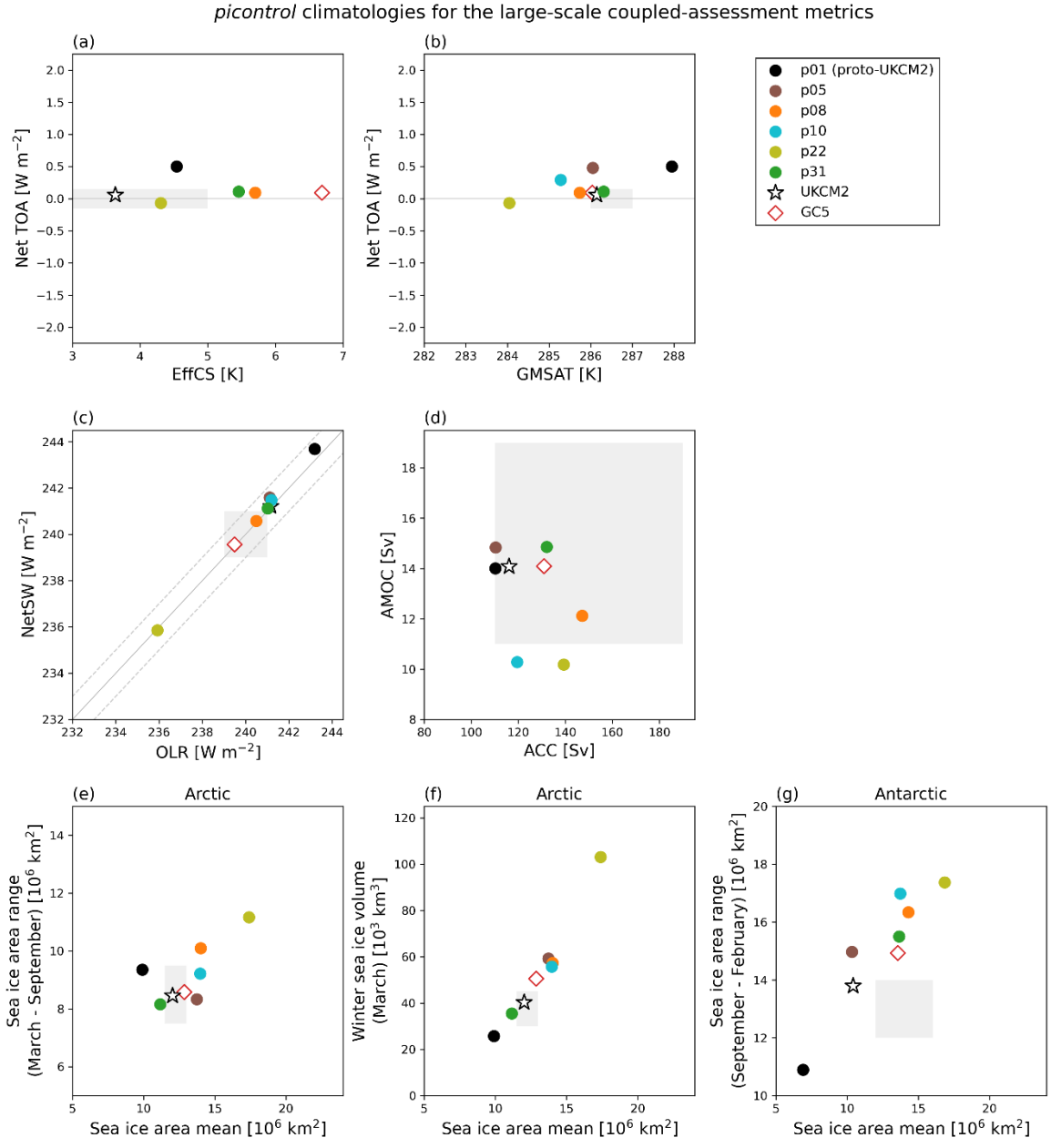

Figure 3 *picontrol* climatologies of the coupled-assessment metrics for 6 UKCM2 candidates (circles), GC5 (red diamond) and UKCM2 (black star). The mean values were calculated using the final 50 simulation years for each of the candidates, and using years 144-343 for UKCM2 and years 251-450 for GC5. The variables in each plot are the same as in Fig. 2, except for (a) where EffCS values are shown instead of λ_{amip} (where available). The target ranges are shown as grey boxes. In (c) lines are also shown for net TOA values of -1, 0 and +1 W m^{-2}.

560

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

EffCS values were diagnosed using the *abrupt-4×CO2* experiment for four of the remaining candidates: p01, p08, p22 and p31. p01 (black points in Fig. 3) was selected because it had the lowest λ_{amip} and, as discussed above, we judged that its biases could be reduced through tuning to achieve acceptable performance against the target ranges. p08 (orange) and p31 (green) were selected because they performed reasonably well across the assessment metrics, despite having somewhat higher λ_{amip} values. p05 (brown) was not tested with an *abrupt-4×CO2* experiment because its λ_{amip} was not as low as that of p01 and it only matched 2 of the 11 target ranges, while p10 (cyan) was not chosen because its performance was worse than that of p08 and p31 for almost all of the assessment metrics. p22 (olive-green) was considered an interesting candidate and was found to have the lowest EffCS; however, its climatology was clearly inconsistent with the assessment target ranges, and it was therefore not considered further.

The EffCS values for the three remaining candidates were: 4.5 K for p01, 5.7 K for p08 and 5.5 K for p31 (in line with the relative expectation from the λ_{amip} values). These results indicated that further fine-tuning was required, as none of the candidates performed satisfactorily against all of the target ranges (including the EffCS). For p01, fine-tuning was needed to improve the performance issues noted above while maintaining its acceptable EffCS. In contrast, p08 and p31 required reductions in EffCS to below 5 K without degrading their already reasonable performance.

We subsequently found that p01 showed the most promise in initial fine-tuning tests, and thereafter the development focused solely on this candidate. The fine-tuning of this candidate, hereafter referred to as "proto-UKCM2", is described in the following section.

3.3 Fine-tuning UKCM2

Proto-UKCM2 was selected on the basis of its acceptable EffCS of 4.5 K and the expectation that its *picontrol* climatology could be fine-tuned to better align with the coupled-assessment target ranges defined in Sect. 2.5.2 (and Table 2). Here we describe this fine-tuning process, which led to the final UKCM2 configuration.

As discussed in Sect. 3.2, the main issues with proto-UKCM2 were the positively biased global-mean TOA fluxes, a warm bias in its GMSAT, and too little sea ice in both the Arctic and Antarctic (black points in Fig. 3). These metrics were therefore the primary focus during the fine-tuning, although it was also necessary to preserve the good performance in other aspects of the climatology (e.g. the AMOC) and, crucially, the acceptable EffCS.

Figure 4 summarises the fine-tuning applied to proto-UKCM2, illustrating how adjustments to GC5 parameters progressively reduced the positive net TOA flux imbalance, ultimately leading to the UKCM2 configuration which performed acceptably against the coupled-assessment target ranges. Corresponding figures for netSW, OLR and GMSAT are shown in Figs. B1-B3.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

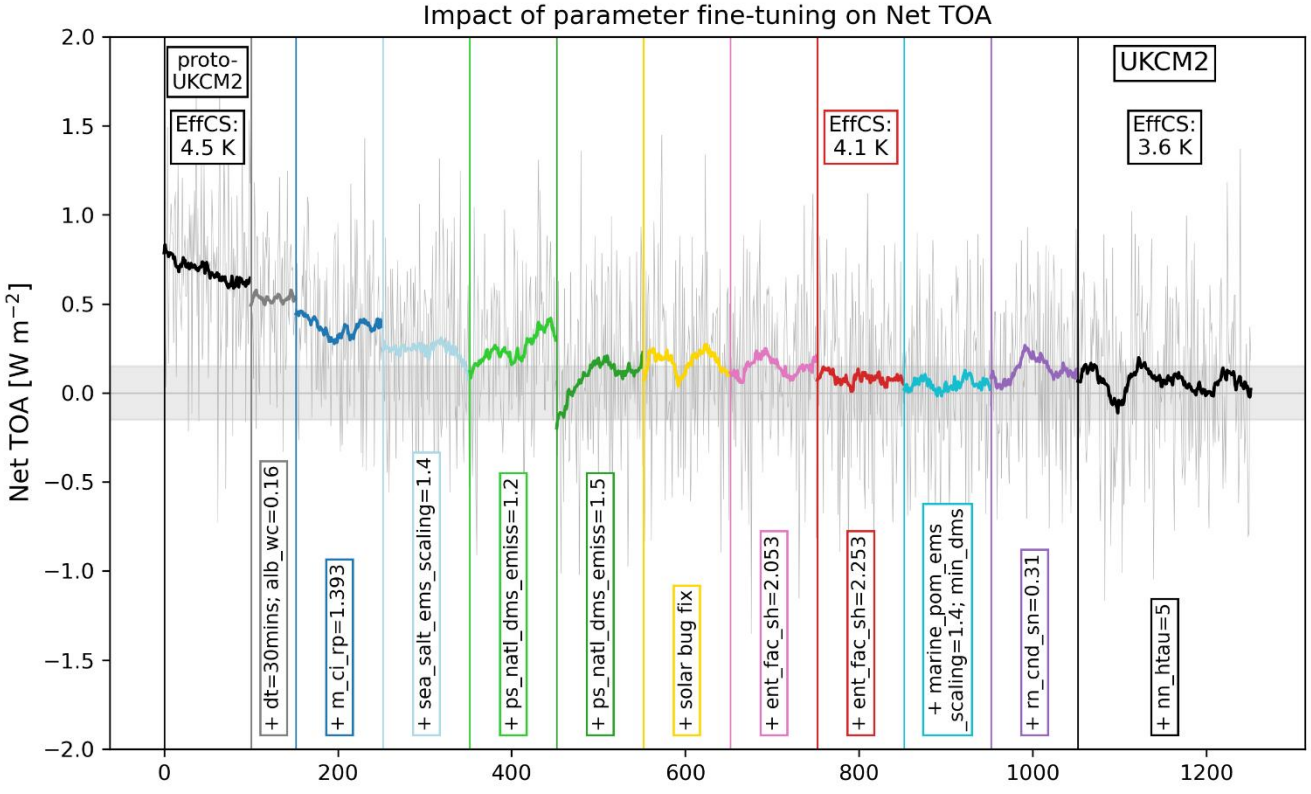

Figure 4 Evolution of Net TOA flux during the fine-tuning of proto-UKCM2, ultimately resulting in the final UKCM2 configuration. For each change, only the final 100 years of that simulation is shown (or less for shorter tests), except for UKCM2 where the assessed period of years 144-343 is shown. For proto-UKCM2, the years immediately preceding the first fine-tuning change are shown (years 90-189). This means the displayed timeseries is not continuous but rather highlights the impact of each change, after the simulation has had an opportunity to equilibrate. The coloured lines show the timeseries' smoothed using 31-year rolling means. The fine-tuning changes are indicated along the bottom, and EffCS values are shown where available.

595 The reduction in net TOA was achieved primarily by adjusting parameters to increase outgoing shortwave radiation, thereby decreasing the netSW flux. Although this involved some trial and error, the specific parameter adjustments were guided by (i) the existing knowledge of model experts and (ii) quantitative parameter sensitivity information derived from emulators of the *amip* experiments (here using the original 503 GC5 PPE variants plus the 41 UKCM2 candidates; Fig. 5).

600 Several of the parameter adjustments involved scaling natural aerosol emissions, including sea-salt (sea_salt_ems_scaling), primary marine organic aerosols (PMOA; marine_pom_ems_scaling), and dimethyl sulphide (DMS; ps_natl_dms_emiss). We also imposed a minimum DMS emission flux to address the known difficulty of representing point DMS measurements in gridded simulations (see e.g. Anderson et al., 2001). The resulting increases in aerosol optical depths (AOD) and cloud droplet number concentrations (CDNC) reduced the netSW flux more than they reduced the OLR (Figs. B1 and B2), leading to

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

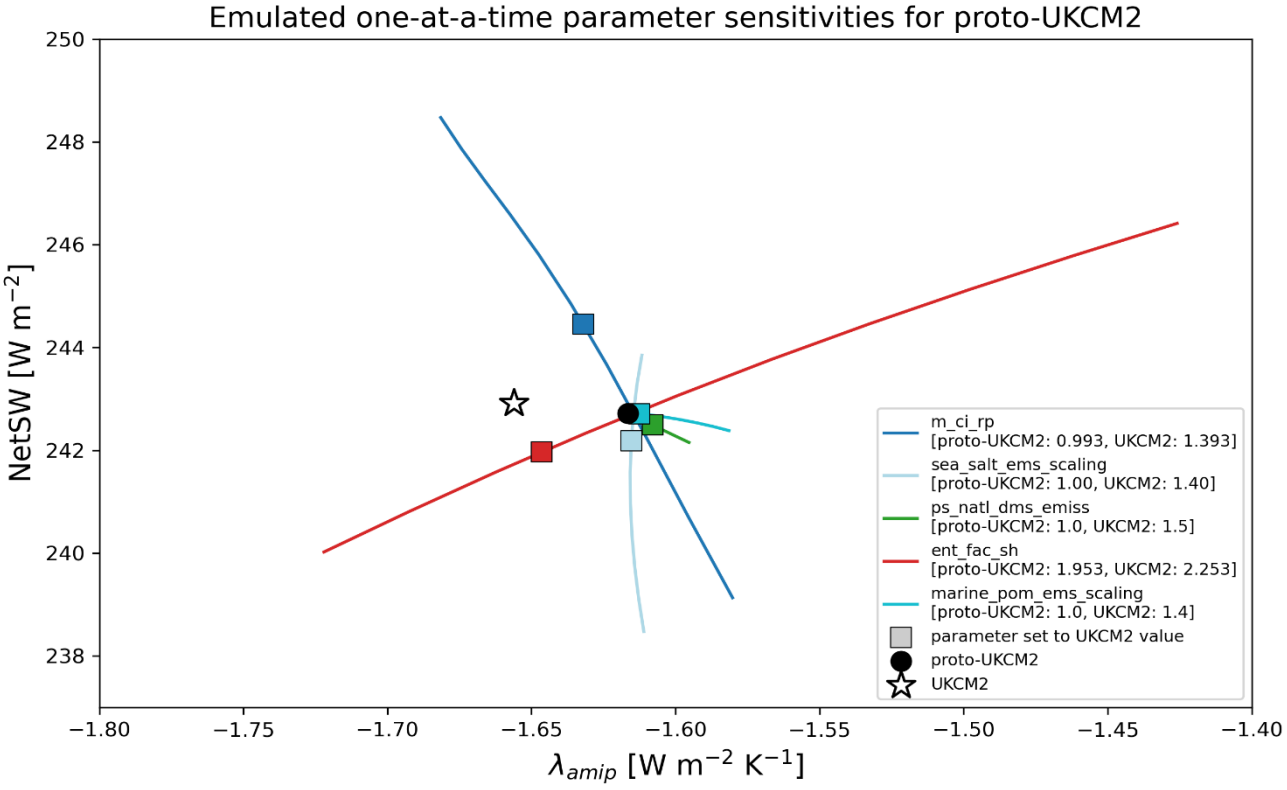

Figure 5 Parameter sensitivities derived from emulated predictions of netSW and λ_{amip} from the *amip* experiments. Five of the parameters used during the coupled-model fine-tuning are shown. The lines show how netSW and λ_{amip} change as each parameter is moved between the minimum and maximum of their elicited range, with all other values held at the proto-UKCM2 values. The black circle and black star show the emulated predictions for proto-UKCM2 and UKCM2, respectively. The squares show the emulated predictions where that parameter is set to the UKCM2 value, with all other parameters set to the proto-UKCM2 values. These values are indicated in the legend. The GC5 values for these parameters are all equal to 1.0 (Table 1).

605

moderate reductions in net TOA (Fig. 4). Additionally, the *amip* parameter sensitivities indicated that these perturbations would have only a minimal impact on λ_{amip} and therefore were not expected to substantially alter the EffCS (Fig. 5).

The final values of these scalings (given in Table 1) were chosen so that the global-mean emissions, AOD, and CDNC remained
610 within the upper bounds of observational estimates (adjusted to the pre-industrial period). These estimates carry substantial uncertainty and provided flexibility to explore a wide range of scaling values.

Modifications to cloud-related parameters were also tested, including adjustments within the microphysics, convection, and cloud/cloud-radiation schemes. Two of these modifications were adopted. The first was an increase in the ice fall-speed

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

parameter (m_ci_rp), which primarily reduces cloud ice water content, leading to lower high-cloud fractions. This change increased the biases in both netSW and OLR but improved (i.e., reduced) the net TOA because the longwave impact of the reduced high cloud was larger than the corresponding shortwave impact. The second adopted change was an increase in the entrainment factor for shallow convection (ent_fac_sh), which reduced both netSW and OLR, and also produced a modest improvement in the net TOA.

Again, these adjustments were partly motivated by the *amip* sensitivities, which indicated that both changes would reduce the EffCS (Fig. 5). This was confirmed with an *abrupt-4×CO2* experiment, performed after applying the adjustments to m_ci_rp and ent_fac_sh, which yielded an EffCS of 4.1 K (compared to 4.5 K for proto-UKCM2; Fig. 5 also shows the direction of the netSW response was accurately predicted).

Further changes were applied to the sea ice and ocean model components. GC5-based configurations at N96ORCA1 resolution were typically found to have too much sea ice, including the UKCM2 candidates as discussed in Sect. 3.2. This bias was partly driven by an overly weak northward ocean heat transport through Fram Strait in the 1° ocean model, as was found for HadGEM3-GC3.1-LL (Kuhlbrodt et al., 2018). The low global-mean temperatures of these configurations are also likely to have contributed to the excessive sea ice. Although proto-UKCM2 was comparatively warm and actually exhibited a sea ice deficit (Fig. 3), the fine-tuning adjustments led to a cooler climate (Fig. B3) which will have increased sea ice area and volume.

We tested two changes to SI^3 model parameters to address this issue. First, we reduced the snow on sea ice albedo values for visible and infrared wavelengths by 0.02, following the approach of Kuhlbrodt et al. (2018) for HadGEM3-GC3.1-LL (this reduction remained within the observational uncertainty). Second, we tested a reduction in the conductivity of snow on sea ice (rn_cnd_s), which limits winter sea-ice growth and, unlike the albedo changes, does not directly affect TOA fluxes. During the development of GC5, rn_cnd_s was increased from its previous default of 0.31 to 0.50 W m^{-1} K^{-1} (Blockley et al., 2024); here we assessed the impact of reverting it to 0.31 W m^{-1} K^{-1}. In tests of candidates p08, p10, and p31 (Sect. 3.2), both of these changes reduced sea ice areas and volumes, although the impact of rn_cnd_s on Antarctic sea ice was minimal. For UKCM2, we chose to adopt the reduction in rn_cnd_s.

A change to the GOSI9 configuration was introduced to reduce wind-driven mixing depths in the Southern Hemisphere, between 13°S and 48.5°S. This change produced a shallower summer mixed layer in this region, which led to slightly warmer summer SSTs, thereby amplified an existing positive bias. This caused increases in OLR and netSW (Figs. B1 and B2) but, crucially, helped reduce the net TOA flux so that it fell within the target range of ±0.15 W m^{-2} (Fig. 4).

Two corrections were also applied to the model. First, the prescribed ocean whitecap albedo parameter (alb_wc) in the open-sea albedo scheme (Jin et al., 2011; Willett et al., 2025) was reduced from its default value of 0.55 to 0.16. The default

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

corresponds to fresh, dense foam patches and is therefore almost certainly an overestimate. The reduced value of 0.16 was chosen as a more realistic estimate that accounts for the ageing process of foam patches and streaks (Koepke, 1984). Second, a scaling error in the solar constant was corrected. The previous implementation inadvertently applied a value corresponding to the orbital time-average rather than the value at 1 astronomical unit. Both corrections had been tested previously and were found to have minimal impacts on simulation results.

An *abrupt-4×CO2* experiment for UKCM2 yielded an EffCS value of 3.6 K, fulfilling one of the key aims of the project. Values for the remaining coupled-assessment metrics are listed in the final column of Table 2, and are shown as the black star in Fig. 3 (the equivalent values for GC5 are shown as the red diamond). The UKCM2 values were calculated for a 200-year period covering years 144 to 343 into the UKCM2 simulation, noting that this was started in a well-spun-up state as it was preceded by 274 years of integration from proto-UKCM2 and the fine-tuning tests. We used this period because it is the same as that used in the assessment exercise on which the formal adoption of UKCM2 was based.

UKCM2 performs well across the coupled-assessment metrics, with 9 of the 12 target ranges being met. It demonstrates a comparable level of performance to GC5 for these large-scale metrics, while achieving a substantially reduced EffCS. Importantly, it is very close to radiative balance, with a mean value of +0.06 W m^{-2} for the assessed period.

However, some compromises were required in the large-scale climatological performance of UKCM2. The targets that were not met were the component TOA fluxes (netSW and OLR) and the Antarctic annual-mean sea ice area. The biases for netSW and OLR were small - only 0.2 Wm^{-2} above the upper limit of their respective target ranges. This was driven by a trade-off between these fluxes and the GMSAT, which was at the lower end of its target range (Fig. 3b, c). The fine-tuning primarily targeted reductions in netSW, which reduced the surface temperature and OLR, and so it would have been difficult to correct the component fluxes without adversely affecting the GMSAT (Figs. B1-B3). For the Antarctic sea ice area, as noted in Sect. 3.2, none of the candidates (including GC5) fell within the target range of both the annual mean and annual range of sea ice area (Fig. 3g). For UKCM2 this compromise resulted in a negative bias in the Antarctic annual-mean sea ice area. Further testing and tuning of the SI^3 model parameters may have reduced this bias; however, we did not pursue this due to resource constraints.

As noted in Sect. 2.5.2, the coupled-assessment metrics were designed to test the large-scale climatological performance and stability of the models. The final adoption of the UKCM2 configuration was subject to a more comprehensive assessment of the performance. That assessment showed that, while the overall performance of UKCM2 is degraded compared to GC5 (as expected from our finding in the *amip* experiments), it does have an improved simulation of historical temperatures, and its performance is competitive with CMIP6 models across a wide range of variables. Therefore, the main aims of the project were achieved.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

4. Key parameter changes for UKCM2

In Sect. 3.2, we showed how *amip* parameter sensitivities were used to inform the fine-tuning of proto-UKCM2 such that the TOA fluxes could be adjusted without increasing the EffCS. Here, we use the *amip* parameter sensitivities in a different way: to identify the key parameter changes between GC5 and UKCM2 and to assess their impact. We use emulators of the *amip* experiments because they can be trained reliably, due to the large size of the PPE. In contrast, the coupled experiments were performed for a relatively small number of candidates, and each involved simultaneous changes to many parameters. This prevented us from training reliable coupled emulators.

Fig. 6 shows the impact of key parameter changes in UKCM2 on the emulated predictions of λ_{amip} and the global-mean SW cloud-radiative effect (SW CRE). For each parameter, the changes in λ_{amip} and SW CRE resulting from perturbing its value from the GC5 setting to the UKCM2 setting are indicated by squares. These perturbations include contributions from both the *amip* optimisation (Sect. 3.1) and the fine-tuning (Sect. 3.3), but in practice the *amip* optimisation accounts for most of the changes. This is evident from the differences between the emulated predictions for GC5, proto-UKCM2 and UKCM2 in Fig. 6 (red diamond, black point and black star, respectively).

Fig. 6 (and Fig. 5) shows that the reduction in λ_{amip} (and hence EffCS) between GC5 and UKCM2 resulted from modifications to many parameters. This is a reflection of the fact that, in addition to reducing λ_{amip}, the parameter changes were optimising performance across a broad range of physical processes (represented by the 18 RMSE variables used during the *amip* optimisation, and the 11 assessment metrics used during the coupled development). It would have been unlikely for a change to a single parameter, or a small number of parameters, to reduce the EffCS from the GC5 value of 6.7 K to within the target range of 2–5 K while maintaining acceptable performance across all these variables.

Additionally, different parameter changes produced compensating effects on individual performance metrics. This compensating behaviour is shown in Fig. 6, where the substantial individual changes to the SW CRE largely cancel out, so that the UKCM2 value remains reasonably close to observations. Some perturbations even increased λ_{amip} (e.g. two_d_fsd_factor), presumably because, within the automated *amip* optimisation, their impacts on the RMSE variables were used to offset the effects of the parameters that *were* driving lower values of λ_{amip}. This is demonstrated in Fig. 6, where the positive SW CRE bias for UKCM2 would have been much larger were it not for the negative SW CRE change from the

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

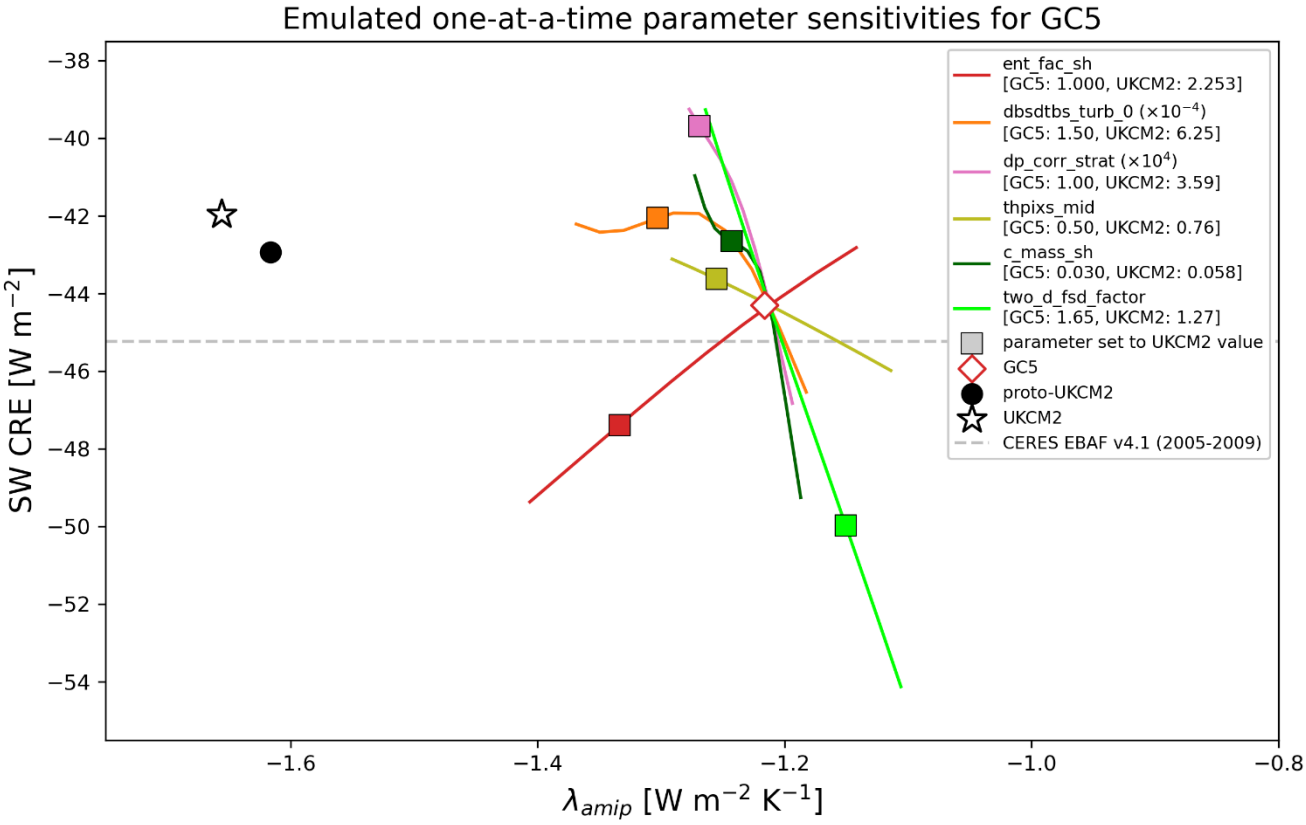

Figure 6 Parameter sensitivities derived from emulated predictions of SW CRE and λ_{amip} from the *amip* experiments. Six of the parameters with the biggest impact on λ_{amip} are shown. The lines show how netSW and λ_{amip} change as each parameter is moved between the minimum and maximum of their elicited range, with all other values held at the GC5 values. The red diamond, black circle and black star show the emulated predictions for GC5, proto-UKCM2 and UKCM2, respectively. The squares show the emulated predictions where that parameter is set to the UKCM2 value, with all other parameters set to the GC5 values. These values are indicated in the legend.

two_d_fsd_factor perturbation. However, SW CRE is only one of the 18 variables used in the optimisation, and the role of the parameter changes in balancing the overall climatological performance, while also reducing λ_{amip}, is likely to be more complex than suggested by this single variable.

This sensitivity analysis indicates that the optimisation was dominated by cloud-related processes, with the key changes applied to parameters in the convection, cloud/cloud radiation and microphysics schemes. These parameters influence the distribution of cloud (e.g. cloud amount and height) and their impact on radiation. Additionally, we found that the largest contribution to the change in λ_{amip} (for proto-UKCM2 vs GC5) came from cloud feedbacks, and primarily the SW CRE feedbacks. The total change in the simulated λ_{amip} was -0.400 W m^{-2} K^{-1}, of which -0.318 W m^{-2} K^{-1} was due to the global-mean SW CRE feedback.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

In contrast, the changes in the global-mean LW CRE and clear-sky feedbacks were much smaller, at -0.004 W m^{-2} K^{-1} and -0.078 W m^{-2} K^{-1}, respectively.

The details of how these parameters affect λ_{amip} are likely to be complex. Parameter sensitivities vary between regions and between flux components because different physical processes dominate at different locations (Tsushima et al., 2020). The sensitivities shown for λ_{amip} in Fig. 6 may be masking these variations, for example, where opposing regional responses to a given parameter cancel in the global mean. A detailed analysis of how parameter perturbations affect feedbacks in UKCM2 – such as isolating specific cloud regimes or flux components – is beyond the scope of this study, but is planned as follow-up work.

5. Summary and Discussion

UKCM2 was developed for use in applications such as the production of national climate projections, decadal prediction, and Earth-system modelling, where the GC5 configuration was not suitable due to its extremely high EffCS and rapid late-20th/early-21st century temperature rise. The key aims for UKCM2 were to have an EffCS within the IPCC AR6 very likely range of 2–5 K and to improve the simulation of surface temperatures during the historical period, while maintaining acceptable climatological performance.

Together with Bodas-Salcedo et al. (in prep.), we have shown that these aims were achieved: UKCM2 exhibits an EffCS of 3.6 K, has an improved simulation of historical surface temperatures compared to GC5, and shows a climatological performance that, while degraded with respect to GC5, *is* competitive with CMIP6 models.

In the following sections we discuss some potential improvements to the methodology we used to develop UKCM2 (Sect. 5.1), and how lessons learned from this project may feed back into model development, along with some of the wider implications of this work (Sect. 5.2).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

5.1 Improvements to the methodology

Our method to develop UKCM2 followed a two-stage approach based on PPEs, comprising an automated *amip*-based optimisation of parameters using cheap experiments, followed by an expert-led "manual" evaluation and fine-tuning stage, based on more expensive coupled-model experiments.

In both stages the aim was to tune parameter values against a broad set of performance metrics, along with EffCS or λ_{amip}. The number and diversity of performance metrics was an important choice, to reduce the risk of over-fitting to a small number of targets. However, balancing performance across many variables was challenging, particularly during the coupled fine-tuning when the turnaround for testing was much slower. The use of PPEs was invaluable here – both in the automated *amip*-based optimisation and as a tool to guide the manual fine-tuning.

While this approach was successful, several potential improvements could make the process more efficient and increase the likelihood of developing better-performing configurations with acceptable climate sensitivities. In general, the method could be improved by reducing the burden on the resource-intensive coupled stages of the development (particularly the manual fine-tuning) by expanding the role of automated methods. This could be achieved by (i) improving predictions of the coupled performance and EffCS values of the candidates and (ii) improving the speed of the model.

One area for improvement would be more accurate emulation of the *amip* simulations. This would provide more reliable predictions of model performance and λ_{amip} across the GC5 parameter space, helping to reduce discrepancies between emulated and simulated outcomes. This would be particularly important for λ_{amip}, where errors in the emulated predictions typically ranged from 0.1 to 0.3 W m^{-2} K^{-1} (Fig. 1). For example, multiple 'waves' of the PPE could be conducted, with observational and/or λ_{amip} constraints being applied after each iteration (e.g. Elsaesser et al., 2025; Sexton et al., 2021). By progressively narrowing the parameter space, this approach could allow promising parameter combinations to be identified earlier (Elsaesser et al., 2025), increasing the likelihood that selected candidates would perform well in subsequent coupled testing. As discussed in Sect. 4, the parameter sensitivities for λ_{amip} are complex, combining multiple processes that vary spatially. New AI/ML techniques which are well suited to provide emulation of spatial sensitivities (e.g. CNNs; Watson-Parris et al., 2021) could provide improvements here.

Another major improvement would be to include coupled spin-up performance metrics directly in the automated optimisation stage. To emulate these metrics, we would need a PPE of coupled spin-ups that had the same atmospheric and land parameter perturbations as a subset of the AMIP PPE. This would allow us to use statistical methods to identify AMIP indices that predict coupled spin-up outputs. For any untried parameter combination, the full AMIP PPE could first be used to emulate the indices, and those emulated indices could then be used to predict coupled metrics. Because coupled spin-ups are computationally

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

expensive, the coupled PPE size must balance the cost against the need for robust index estimates and a reliable mapping from AMIP indices to coupled metrics. Although atmospheric parameters would be inherited from the AMIP PPE, a sufficiently large coupled PPE could also sample a small number of ocean and sea-ice parameters (ideally at least six times as many simulations as oceanic parameters), potentially improving performance in these components (e.g. addressing the systematic behaviour in Antarctic sea ice highlighted in Sects. 3.2 and 3.3).

Some relatively simple improvements could also be made through more effective use of metrics. For example, including aerosol variables such as global-mean AOD and CDNC in the *amip* aggregate error metric would have helped identify candidates with unrealistic aerosol emissions earlier in the process. This would have addressed the case of candidate p08 which was found – at a relatively late stage of testing - to have unrealistically large sea-salt emissions. Correcting the processing error for the implied ocean heat transport metric (fmassefp; Sect. 3.1) may have increased the likelihood of selecting candidates with smaller initial coupling "shocks" and more stable coupled simulations, thereby simplifying the coupled-model evaluation. And early insights into historical global-mean temperature simulations could have been obtained by estimating aerosol effective radiative forcings (ERFs), for example using the RFMIP piClim-histaerO3 experiment (Pincus et al., 2016).

As noted in the Introduction, GC5 is a relatively slow model, and this was part of the motivation for adopting the two-stage development process. While this was successful, it meant that the process relied heavily on the *amip* predictions of coupled behaviour. Having a faster model would improve all stages of the development, as it would allow a wider range of candidates to be tested using more effective experiments (e.g. it could enable a PPE of coupled simulations, as discussed above).

A simple change to increase model speed, which was adopted here during the fine-tuning, was to increase the atmospheric model timestep from 20 min to 30 min. This resulted in a speed-up of more than 10%. More complex techniques to improve model speed are being explored, including using reduced floating-point precision, running slower model components at lower resolution (e.g. the UKCA component), and improving process parallelism. Finally, one active area of research is the acceleration of model spin-ups, which would enable more reliable assessments of coupled performance because they would be conducted with models that were closer to equilibrium.

5.2 Implications for model development and the wider modelling community

In developing UKCM2 we have demonstrated that PPEs, along with techniques such as emulation, are a valuable tool for model development because they provide a way to efficiently explore behaviour across a model's parameter space. If they were incorporated more routinely into model development processes they could, for example, be used to quickly generate test candidates in the development of major releases of a model, or help developers understand the sensitivities of model parameters. PPEs can also give developers a better understanding of the structural behaviour of the model. They are particularly

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

well suited to this because they readily expose behaviours which persist across the model's parameter space (Furtado et al., 2023; McNeall et al., 2016; Regayre et al., 2023; Rostron et al., 2025).

The most important structural feature of the GC5 model for this project is shown in Fig. 1 i.e. that lower expected values of EffCS require degradations in the model performance. We showed this explicitly in the atmosphere-only experiments in Sect. 3.1, and the results from the coupled experiments are consistent with this: compared to GC5, UKCM2 has a substantially lower EffCS (3.6 K vs 6.7 K) but higher root-mean-squared errors, driven by degradations in spatial patterns which were not captured in the large-scale coupled-assessment metrics used during the development (Bodas-Salcedo et al., in prep.). This structural feature of GC5 is not expected to be seen in all models – indeed, Peatier et al. (2022) found a much flatter relationship between their error metric and λ_{amip} for the CNRM-CM PPE (see Figure 2e in that paper). A more direct comparison of this relationship across different model structures, using a standardised error metric, would be an interesting topic for future work.

The structural relationship between model performance and EffCS for GC5 has forced us to confront the tensions that arise when using our models in multiple contexts. GC5 is a state-of-the-art UM configuration that performs well for a range of applications, most notably NWP; but, as we discussed in the Introduction, it is not suitable for applications that require a more plausible EffCS and/or a better simulation of the historical temperature record. This is ultimately what motivated the development of UKCM2.

Future UM configurations may have different structural relationships of this kind, and may not require trade-offs between performance and EffCS in the same way. The extent to which we will actively monitor, or modify, these relationships in future UM model development will need to be considered. However, it is very possible that these trade-offs will continue to be required for future UM configurations, and indeed for models produced by other modelling centres, so it is important to formulate a coherent, physically credible strategy for dealing with this.

This highlights another tension: when these models are used in the wider community, in projects such as CMIP. Here, there can be a tension between the useability of a model for applications specific to each modelling centre (e.g. producing national climate projections) and the value it provides to the wider community. If all modelling centres contributing to CMIP did as we have done and tuned their models' EffCS to be within the IPCC very likely range (perhaps to meet their own user requirements), then this would reduce the diversity in the CMIP ensemble. Clearly, this is not desirable – very high and low sensitivity models are valuable for certain applications (e.g., studying high-impact low-likelihood events) and as research tools e.g., to increase our understanding of the robustness of the upper limits of assessed EffCS ranges. Related to this, there are questions about whether the upper bound of the IPCC very likely range was too conservative, due to misrepresentation of pattern effect biases (Armour et al., 2024; Myhre et al., 2025).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Our experience suggests that a more pragmatic approach to model development may be necessary, in which alternative model configurations are produced according to the requirements of specific users (or groups of users as we have done for UKCM2). However, this would need to be balanced against the high resource costs required to develop and maintain new configurations of a climate model.

These issues are clearly worthy of wider discussion within the community: in particular, how to balance specific user needs with the desire to maintain multi-model diversity. The need to address this could become increasingly important as more reliable observational or emergent constraints are developed.

Appendix A: Table of variables used in the *amip*-based optimisation of 41 UKCM2 candidates

Table A1 Variables used to assess present-day climatological performance in the *amip* simulations, for the generation of 41 UKCM2 candidates. The observational comparison data used are given in for each variable, along with their references.

Variable	Season/Annual	Observational dataset	Reference
Precipitation	DJF, JJA	GPCP version 2.3	(Adler et al., 2018)
Surface pressure	DJF, JJA	HadSLP2	(Allan and Ansell, 2006)
1.5m temperature	Annual	ERA-Interim	(Dee et al., 2011)
Zonal winds at 850hPa	Annual	ERA-Interim	
Meridional winds at 850hPa	Annual	ERA-Interim	
Relative humidity at 850hPa	Annual	ERA-Interim	
Temperature at 200hPa	Annual	ERA-Interim	
Zonal winds at 200hPa	Annual	ERA-Interim	
Relative humidity at 200hPa	Annual	ERA-Interim	
LW clear-sky outgoing flux at TOA	Annual	CERES-EBAF Ed4.1	(Loeb et al., 2018; NASA/LARC/SD/ASDC, 2019)
LW cloud radiative effect	Annual	CERES-EBAF Ed4.1	
SW cloud radiative effect	DJF, JJA	CERES-EBAF Ed4.1	
LW clear-sky downwelling surface flux	Annual	CERES-EBAF Ed4.1	
LW downwelling surface flux	Annual	CERES-EBAF Ed4.1	

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Implied ocean heat transport metric (fmassefp)	Annual	DEEP-C v5	(Liu et al., 2017)

860

Appendix B: Impact of parameter fine-tuning on netSW, OLR and GMSAT

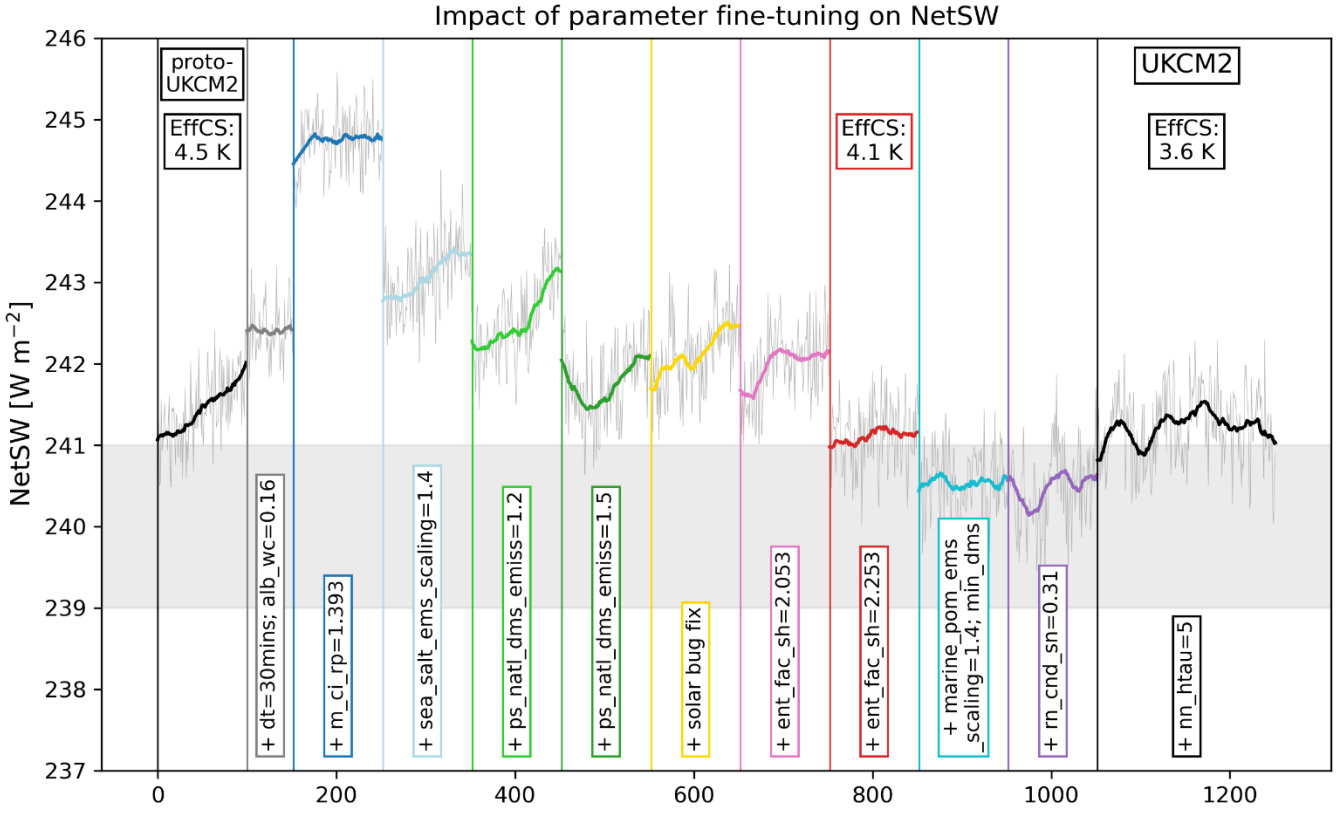

Figure B1 As Fig. 4 but for netSW.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

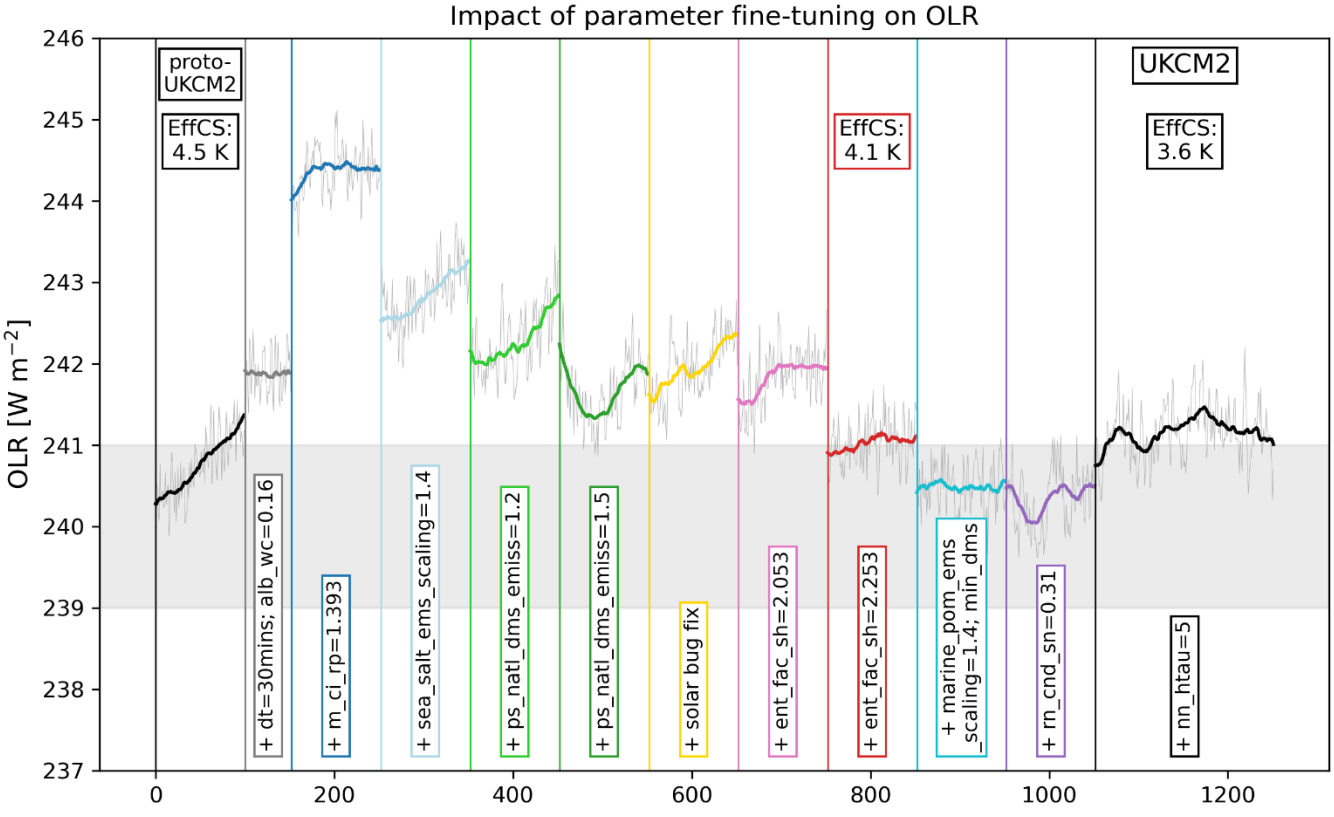

Figure B2 As Fig. 4 but for OLR.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

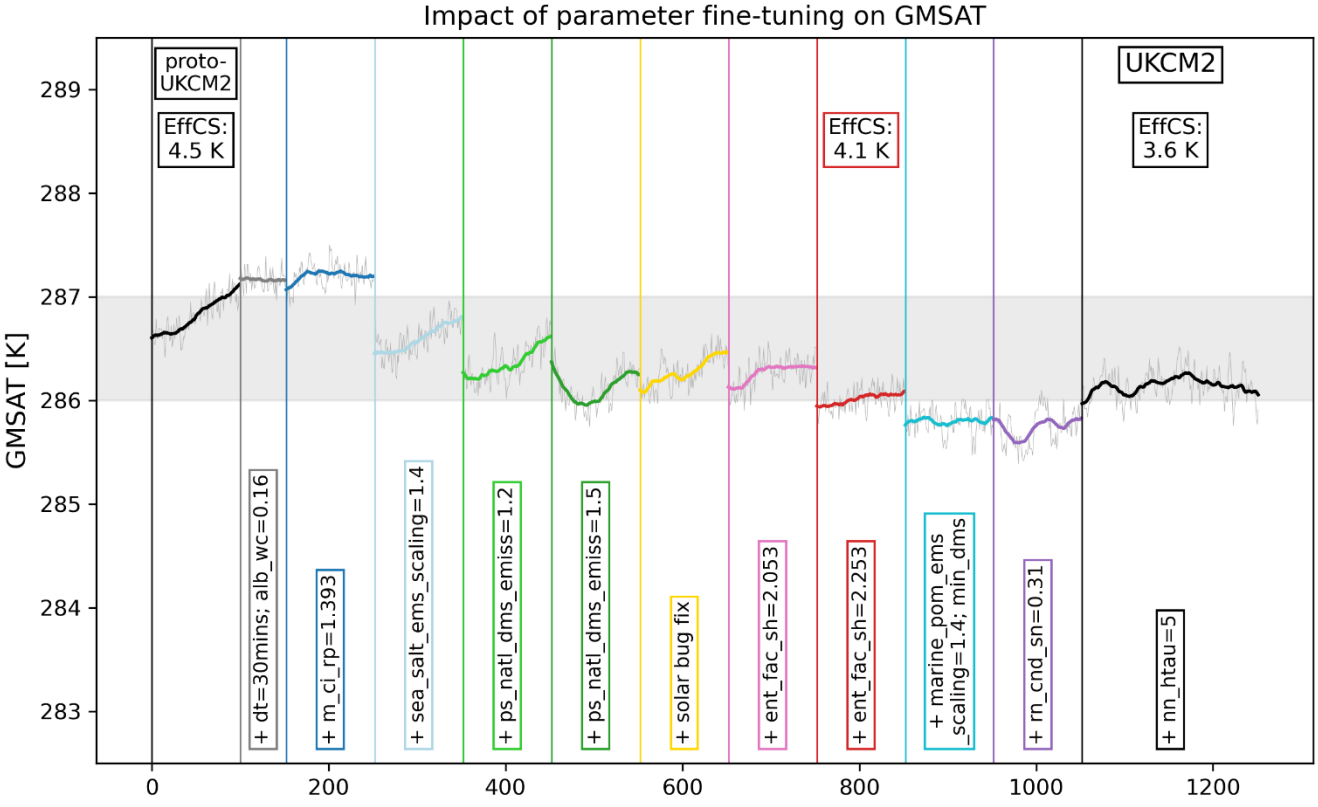

Figure B3 As Fig. 4 but for GMSAT.

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 (last access: March 2026), and for permission to use JULES, go to https://jules.jchmr.org.

The analysis and plotting scripts used to process the data and reproduce the figures and tables in this paper are publicly available on Zenodo at https://doi.org/10.5281/zenodo.19205160 (Rostron, 2026).

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Data availability

Data used in the development and evaluation of the UKCM2-LL climate model configuration, including the data used to produce the figures and tables in this paper, are publicly available on Zenodo at https://doi.org/10.5281/zenodo.19267353 (Rostron and Sexton, 2026).

GPCP v2.3 monthly precipitation data are publicly available from NOAA NCEI (https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00979) (Adler et al., 2018). 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). The DEEP-C v5 dataset is available from the University of Reading Research Data Archive (https://researchdata.reading.ac.uk/347/) (Liu et al., 2017). HadSLP2 data were obtained from https://www.metoffice.gov.uk/hadobs/hadslp2 (Allan and Ansell, 2006). 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). PIOMAS reanalysis data are available from the Polar Science Center web page at http://psc.apl.uw.edu/research/ projects/arctic-sea-ice-volume-anomaly (Schweiger et al., 2011).

Author contributions

JWR led the preparation of the manuscript, including the processing code to produce the Tables and Figures, and developed the GC5 PPE with DMHS. ABS led the project to develop UKCM2, from its inception to its final adoption. DMHS led the *amip* analysis to select the 41 UKCM2 candidates, along with SP and BMS. ABS, DMHS, CGJ, EWB, TK, JPM, TEP, MAR and MRW contributed their expertise and analyses to the coupled-model-based stages of the development. CGJ led the fine-tuning, along with SR and YT who ran the simulations, EWB who led on the sea-ice model tuning and TEP who led on the changes to nn_htau. All co-authors made substantial contributions to the manuscript, through comments and edits to the text.

Competing interest

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Acknowledgements

Microsoft Copilot was used to assist with improving the clarity of the manuscript text. 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.

Financial support

John W. Rostron, Alejandro Bodas-Salcedo, David M. H. Sexton, Edward W. Blockley, Jane P. Mulcahy, Tamzin E. Palmer, Mark A. Ringer, Yongming Tang and Martin R. Willett were supported by the Met Office Hadley Centre Climate Programme funded by DSIT. Colin G. Jones, Till Kuhlbrodt and Steven 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). Colin 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). Saloua Peatier was supported by the Horizon Europe projects OptimESM (grant agreement no. 101081193) and TipESM (grant agreement no. 101137673), and by the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number 10090271]. Benjamin M. Sanderson acknowledges funding from Norwegian Funding Council project "NorESM4CMIP7 - Norwegian Earth System Modeling for CMIP7", Project Number: 2732835.s

References

Acosta, M. C., Palomas, S., Paronuzzi Ticco, S. V., Utrera, G., Biercamp, J., Bretonniere, P.-A., Budich, R., Castrillo, M., Caubel, A., Doblas-Reyes, F., Epicoco, I., Fladrich, U., Joussaume, S., Kumar Gupta, A., Lawrence, B., Le Sager, P., Lister, G., Moine, M.-P., Rioual, J.-C., Valcke, S., Zadeh, N., and Balaji, V.: The computational and energy cost of simulation and storage for climate science: lessons from CMIP6, Geoscientific Model Development, 17, 3081–3098, https://doi.org/10.5194/gmd-17-3081-2024, 2024.

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., and Shin, D.-B.: The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation, Atmosphere, 9, 138, https://doi.org/10.3390/atmos9040138, 2018.

Allan, R. and Ansell, T.: A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004, Journal of Climate, 19, 5816–5842, https://doi.org/10.1175/JCLI3937.1, 2006.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Anderson, T. R., Spall, S. A., Yool, A., Cipollini, P., Challenor, P. G., and Fasham, M. J. R.: Global fields of sea surface dimethylsulfide predicted from chlorophyll, nutrients and light, Journal of Marine Systems, 30, 1–20, https://doi.org/10.1016/S0924-7963(01)00028-8, 2001.

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/10.1029/2019MS001866, 2019.

Armour, K. C., Proistosescu, C., Dong, Y., Hahn, L. C., Blanchard-Wrigglesworth, E., Pauling, A. G., Jnglin Wills, R. C., Andrews, T., Stuecker, M. F., Po-Chedley, S., Mitevski, I., Forster, P. M., and Gregory, J. M.: Sea-surface temperature pattern effects have slowed global warming and biased warming-based constraints on climate sensitivity, Proceedings of the National Academy of Sciences, 121, e2312093121, https://doi.org/10.1073/pnas.2312093121, 2024.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

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, Geoscientific Model Development, 17, 6799–6817, https://doi.org/10.5194/gmd-17-6799-2024, 2024.

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

Bodas-Salcedo, A., Rostron, J. W., and Sexton, D. M. H.: UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity – assessment of CMIP7 DECK simulations, in prep.

Bonan, D. B., Thompson, A. F., Schneider, T., Zanna, L., Armour, K. C., and Sun, S.: Observational constraints imply limited future Atlantic meridional overturning circulation weakening, Nat. Geosci., 18, 479–487, https://doi.org/10.1038/s41561-025-01709-0, 2025.

Bonnet, P., Pastori, L., Schwabe, M., Giorgetta, M., Iglesias-Suarez, F., and Eyring, V.: Tuning the ICON-A 2.6.4 climate model with machine-learning-based emulators and history matching, Geoscientific Model Development, 18, 3681–3706, https://doi.org/10.5194/gmd-18-3681-2025, 2025.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., Lionel, Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, Journal of Advances in Modeling Earth Systems, 12, e2019MS002010, https://doi.org/10.1029/2019MS002010, 2020.

Copernicus Climate Change Service (C3S): Global Climate Highlights 2023, European Centre for Medium-Range Weather Forecasts (ECMWF), 2024.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

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.

Cunningham, S. A., Alderson, S. G., King, B. A., and Brandon, M. A.: Transport and variability of the Antarctic Circumpolar Current in Drake Passage, Journal of Geophysical Research: Oceans, 108, https://doi.org/10.1029/2001JC001147, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Donohue, K. A., Tracey, K. L., Watts, D. R., Chidichimo, M. P., and Chereskin, T. K.: Mean Antarctic Circumpolar Current transport measured in Drake Passage, Geophysical Research Letters, 43, 11,760-11,767, https://doi.org/10.1002/2016GL070319, 2016.

Durack, P. J., Taylor, K. E., Ames, S., Po-Chedley, S., and Mauzey, C.: PCMDI AMIP SST and sea-ice boundary conditions version 1.1.8, https://doi.org/10.22033/ESGF/input4MIPs.16921, 2022.

Elsaesser, G. S., van Lier-Walqui, M., Yang, Q., Kelley, M., Ackerman, A. S., Fridlind, A. M., Cesana, G. V., Schmidt, G. A., Wu, J., Behrangi, A., Camargo, S. J., De, B., Inoue, K., Leitmann-Niimi, N. M., and Strong, J. D. O.: Using Machine Learning to Generate a GISS ModelE Calibrated Physics Ensemble (CPE), Journal of Advances in Modeling Earth Systems, 17, e2024MS004713, https://doi.org/10.1029/2024MS004713, 2025.

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 organization, Geoscientific Model Development, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

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, e2023GL103902, https://doi.org/10.1029/2023GL103902, 2023.

Ganachaud, A. and Wunsch, C.: Large-Scale Ocean Heat and Freshwater Transports during the World Ocean Circulation Experiment, Journal of Climate, 16, 696–705, https://doi.org/10.1175/1520-0442(2003)016%3C0696:LSOHAF%3E2.0.CO;2, 2003.

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.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

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, Journal of Geophysical Research: Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Gregory, J. M.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, L03205, https://doi.org/10.1029/2003GL018747, 2004.

Guiavarc'h, C. and Storkey, D.: JMMP-Group/GO_RELEASES: GOSI9 release, , https://doi.org/10.5281/zenodo.13814369, 2024.

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.

Hardacre, C., Mulcahy, J. P., Pope, R. J., Jones, C. G., Rumbold, S. T., Li, C., Johnson, C., and Turnock, S. T.: Evaluation of SO_2, SO_4^{2-} and an updated SO_2 dry deposition parameterization in the United Kingdom Earth System Model, Atmospheric Chemistry and Physics, 21, 18465–18497, https://doi.org/10.5194/acp-21-18465-2021, 2021.

Hausfather, Z., Drake, H. F., Abbott, T., and Schmidt, G. A.: Evaluating the Performance of Past Climate Model Projections, Geophysical Research Letters, 47, e2019GL085378, https://doi.org/10.1029/2019GL085378, 2020.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning, https://doi.org/10.1175/BAMS-D-15-00135.1, 2017.

Intergovernmental Panel on Climate Change (IPCC) (Ed.): The Earth's Energy Budget, Climate Feedbacks and Climate Sensitivity, in: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, 923–1054, https://doi.org/10.1017/9781009157896.009, 2023.

IOC, SCOR, and IAPSO: The international thermodynamic equation of seawater – 2010: Calculation and use of thermodynamic properties, UNESCO, Paris, 2010.

Jin, Z., Qiao, Y., Wang, Y., Fang, Y., and Yi, W.: A new parameterization of spectral and broadband ocean surface albedo, Opt. Express, OE, 19, 26429–26443, https://doi.org/10.1364/OE.19.026429, 2011.

Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M. H., Rostron, J. W., Booth, B. B. B., and Carslaw, K. S.: The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing, Atmospheric Chemistry and Physics, 18, 13031–13053, https://doi.org/10.5194/acp-18-13031-2018, 2018.

Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Turnock, S. T., Browse, J., Sexton, D. M. H., Rostron, J. W., Schutgens, N. A. J., Partridge, D. G., Liu, D., Allan, J. D., Coe, H., Ding, A., Cohen, D. D., Atanacio, A., Vakkari, V., Asmi, E., and Carslaw, K. S.: Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing, Atmospheric Chemistry and Physics, 20, 9491–9524, https://doi.org/10.5194/acp-20-9491-2020, 2020.

Koepke, P.: Effective reflectance of oceanic whitecaps, Appl. Opt., AO, 23, 1816–1824, https://doi.org/10.1364/AO.23.001816, 1984.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 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/10.1029/2018MS001370, 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.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, Atmospheric Chemistry and Physics, 11, 12253–12273, https://doi.org/10.5194/acp-11-12253-2011, 2011.

Liu, C., Allan, R. P., Mayer, M., Hyder, P., Loeb, N. G., Roberts, C. D., Valdivieso, M., Edwards, J. M., and Vidale, P.-L.: Evaluation of satellite and reanalysis-based global net surface energy flux and uncertainty estimates, Journal of Geophysical Research: Atmospheres, 122, 6250–6272, https://doi.org/10.1002/2017JD026616, 2017.

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.

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, 2019.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenéz-de-la-Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2, Journal of Advances in Modeling Earth Systems, 11, 998–1038, https://doi.org/10.1029/2018MS001400, 2019.

McNeall, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., and Sexton, D.: The impact of structural error on parameter constraint in a climate model, Earth System Dynamics, 7, 917–935, https://doi.org/10.5194/esd-7-917-2016, 2016.

Menary, M. B., Kuhlbrodt, T., Ridley, J., Andrews, M. B., Dimdore-Miles, O. B., Deshayes, J., Eade, R., Gray, L., Ineson, S., Mignot, J., Roberts, C. D., Robson, J., Wood, R. A., and Xavier, P.: Preindustrial Control Simulations With HadGEM3-GC3.1 for CMIP6, Journal of Advances in Modeling Earth Systems, 10, 3049–3075, https://doi.org/10.1029/2018MS001495, 2018.

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, e2020GL088166, https://doi.org/10.1029/2020GL088166, 2020.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Morice, C. P., 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, e2019JD032361, https://doi.org/10.1029/2019JD032361, 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.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations., Nature, 430, 768–72, https://doi.org/10.1038/nature02771, 2004.

Murphy, J. M., Booth, B. B. B., Collins, M., Harris, G. R., Sexton, D. M. H., and Webb, M. J.: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles, Philos Trans A Math Phys Eng Sci, 365, 1993–2028, https://doi.org/10.1098/rsta.2007.2077, 2007.

Murphy, J. M., Harris, G. R., Sexton, D. M. H., Kendon, E. J., Bett, P. E., Clark, R. T., Eagle, K. E., Fosser, G., Fung, F., Lowe, J. A., McDonald, R. E., McInnes, R. N., McSweeney, C. F., Mitchell, J. F. B., Rostron, J. W., Thornton, H. E., Tucker, S., and Yamazaki, K.: UKCP18 Land Projections: Science Report, 2018.

Myhre, G., Hodnebrog, Ø., Loeb, N., and Forster, P. M.: Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models, Science, 388, 1210–1213, https://doi.org/10.1126/science.adt0647, 2025.

NASA/LARC/SD/ASDC: CERES Energy Balanced and Filled (EBAF) TOA Monthly means data in netCDF Edition4.1, 2019.

Pearce, F. A. and Bodas-Salcedo, A.: Implied Heat Transport from CERES Data: Direct Radiative Effect of Clouds on Regional Patterns and Hemispheric Symmetry, Journal of Climate, 36, 4019–4030, https://doi.org/10.1175/JCLI-D-22-0149.1, 2023.

Peatier, S., Sanderson, B. M., Terray, L., and Roehrig, R.: Investigating Parametric Dependence of Climate Feedbacks in the Atmospheric Component of CNRM-CM6-1, Geophysical Research Letters, 49, 1–9, https://doi.org/10.1029/2021GL095084, 2022.

Pincus, R., Forster, P. M., and Stevens, B.: The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6, Geoscientific Model Development, 9, 3447–3460, https://doi.org/10.5194/gmd-9-3447-2016, 2016.

Qin, Y., Zelinka, M. D., and Klein, S. A.: On the Correspondence Between Atmosphere-Only and Coupled Simulations for Radiative Feedbacks and Forcing From CO2, Journal of Geophysical Research: Atmospheres, 127, e2021JD035460, https://doi.org/10.1029/2021JD035460, 2022.

Regayre, L. A., Deaconu, L., Grosvenor, D. P., Sexton, D. M. H., Symonds, C., Langton, T., Watson-Paris, D., Mulcahy, J. P., Pringle, K. J., Richardson, M., Johnson, J. S., Rostron, J. W., Gordon, H., Lister, G., Stier, P., and Carslaw, K. S.: Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing, Atmospheric Chemistry and Physics, 23, 8749–8768, https://doi.org/10.5194/acp-23-8749-2023, 2023.

Ringer, M. A., Andrews, T., and Webb, M. J.: Global-mean radiative feedbacks and forcing in atmosphere-only and coupled atmosphere-ocean climate change experiments, Geophysical Research Letters, 41, 4035–4042, https://doi.org/10.1002/2014GL060347, 2014.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Robson, J., Menary, M. B., Sutton, R. T., Mecking, J., Gregory, J. M., Jones, C., Sinha, B., Stevens, D. P., and Wilcox, L. J.: The Role of Anthropogenic Aerosol Forcing in the 1850–1985 Strengthening of the AMOC in CMIP6 Historical Simulations, Journal of Climate, 35, 6843–6863, https://doi.org/10.1175/JCLI-D-22-0124.1, 2022.

Rostron, J.: Supplementary material for manuscript "UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity – methodology and development": Scripts., , https://doi.org/10.5281/zenodo.19205161, 2026.

Rostron, J. and Sexton, D.: Supplementary material for manuscript "UKCM2-LL: a new low-resolution GC5 configuration with constrained climate sensitivity – methodology and development": Data., https://doi.org/10.5281/zenodo.19267354, 2026.

Rostron, J. W., Sexton, D. M. H., Furtado, K., and Tsushima, Y.: A clearer view of systematic errors in model development: two practical approaches using perturbed parameter ensembles, Clim Dyn, 63, 354, https://doi.org/10.1007/s00382-025-07717-5, 2025.

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., and Saha, S.: Practice and philosophy of climate model tuning across six US modeling centers, Geoscientific Model Development, 10, 3207–3223, https://doi.org/10.5194/gmd-10-3207-2017, 2017.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, Journal of Geophysical Research: Oceans, 116, https://doi.org/10.1029/2011JC007084, 2011.

Sexton, D. M. H., Murphy, J. M., Collins, M., and Webb, M. J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology, Climate Dynamics, 38, 2513–2542, https://doi.org/10.1007/s00382-011-1208-9, 2012.

Sexton, D. M. H., Karmalkar, A. V., Murphy, J. M., Williams, K. D., Boutle, I. A., Morcrette, C. J., Stirling, A. J., and Vosper, S. B.: Finding plausible and diverse variants of a climate model. Part 1: establishing the relationship between errors at weather and climate time scales, Clim Dyn, 53, 989–1022, https://doi.org/10.1007/s00382-019-04625-3, 2019.

Sexton, D. M. H., McSweeney, C. F., Rostron, J. W., Yamazaki, K., Booth, B. B. B., Murphy, J. M., Regayre, L., Johnson, J. S., and Karmalkar, A. V.: A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations, Climate Dynamics, 56, 3395–3436, https://doi.org/10.1007/s00382-021-05709-9, 2021.

Sexton, D. M. H., Yamazaki, K., Rostron, J. W., Dunstone, N. J., Fereday, D. R., Hardiman, S. C., Ineson, S., and Knight, J. R.: Effect of resolution on simulated teleconnections to winter North Atlantic circulation inferred from a causal network derived from expert elicitation, Clim Dyn, 63, 32, https://doi.org/10.1007/s00382-024-07497-4, 2024.

Tian, B. and Dong, X.: The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 Models Based on Annual Mean Precipitation, Geophysical Research Letters, 47, 1–11, https://doi.org/10.1029/2020GL087232, 2020.

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/10.1002/2013JD020316, 2014.

Tsushima, Y., Ringer, M. A., Martin, G. M., Rostron, J. W., and Sexton, D. M. H.: Investigating physical constraints on climate feedbacks using a perturbed parameter ensemble, Climate Dynamics, 55, 1159–1185, https://doi.org/10.1007/s00382-020-05318-y, 2020.

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.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., Willett, M., Browse, J., Bushell, A., Carslaw, K., Dalvi, M., Essery, R., Gedney, N., Hardiman, S., Johnson, B., Johnson, C., Jones, A., Jones, C., Mann, G., Milton, S., Rumbold, H., Sellar, A., Ujiie, M., Whitall, M., Williams, K., and Zerroukat, M.: The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations, Geoscientific Model Development, 12, 1909–1963, https://doi.org/10.5194/gmd-12-1909-2019, 2019.

Watson-Parris, D., Williams, A., Deaconu, L., and Stier, P.: Model calibration using ESEm v1.1.0 – an open, scalable Earth system emulator, Geoscientific Model Development, 14, 7659–7672, https://doi.org/10.5194/gmd-14-7659-2021, 2021.

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., Klein, S. A., Marchand, R., Medeiros, B., Siebesma, A. P., Skinner, C. B., Stevens, B., Tselioudis, G., Tsushima, Y., and Watanabe, M.: The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6, Geoscientific Model Development, 10, 359–384, https://doi.org/10.5194/gmd-10-359-2017, 2017.

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, 2021a.

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, 2021b.

Whitworth, T. and Peterson, R. G.: Volume Transport of the Antarctic Circumpolar Current from Bottom Pressure Measurements, Journal of Physical Oceanography, 15, 810–816, https://doi.org/10.1175/1520-0485(1985)015%3C0810:VTOTAC%3E2.0.CO;2, 1985.

Willett, M. and Whitall, M.: A Simple Prognostic based Convective Entrainment Rate for the Unified Model: Description and Tests, Met Office Forecasting Research Technical Report, no. 617., 2017.

Willett, M. R., Brooks, M., Bushell, A., Earnshaw, P., Smith, S., Tomassini, L., Best, M., Boutle, I., Brooke, J., Edwards, J. M., Furtado, K., Hardacre, C., Hartley, A. J., Hewitt, A., Johnson, B., Lock, A., Malcolm, A., Mulcahy, J., Müller, E., Rumbold, H., Rooney, G. G., Sellar, A., Ujiie, M., van Niekerk, A., Wiltshire, A., and Whitall, M.: The Met Office Unified Model Global Atmosphere 8.0 and JULES Global Land 9.0 configurations, https://doi.org/10.5194/egusphere-2025-1829, 2025.

Willett, M. R., Brooks, M., Bushell, A., Earnshaw, P., Smith, S., Tomassini, L., Abraham, N. L., Best, M., Edwards, J. M., Furtado, K., Hardacre, C., Hewitt, A., Johnson, B., Lock, A., Mulcahy, J., Manners, J., Sellar, A., Sheridan, P., Tennant, W., Van Weverberg, K., Varma, V., and Whitall, M.: The Met Office Unified Model Global Atmosphere and Land 9.0 configuration, in prep.

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. B., 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, Journal of Advances in Modeling Earth Systems, 10, 357–380, https://doi.org/10.1002/2017MS001115, 2018.

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.

https://doi.org/10.5194/egusphere-2026-1676
Preprint. Discussion started: 14 April 2026

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, Climate Dynamics, 41, 1703–1729, https://doi.org/10.1007/s00382-013-1896-4, 2013.

Xavier, P., Willett, M., Graham, T., Earnshaw, P., Copsey, D., Marzin, C., Sellar, A., Ackerley, D., Alves, O., Blockley, E., Bodas-Salcedo, A., Bushell, A., Butchart, N., Calvert, D., Cho, J.-A., Copsey, D., de Burgh-Day, C., Edwards, J., Earnshaw, P., Furtado, K., Field, P., Guiavarch, C., Hardy, S., Harris, C., Heywood, K., Heming, J., Hendon, H., Hewitt, H., Hyder, P., Hyun, Y.-K., Hyun, S.-H., Ineson, S., Jones, R., Kim, J., Kim, K.-Y., Klingaman, N., Levine, R., Lee, S.-M., Lekakou, K., Lock, A., Martin, G., Mathiot, P., Megann, A., Meijers, A., Moon, J.-Y., Morgenstern, O., North, R., Nurser, G., Park, Y.-H., Regayre, L., Roberts, M., Rodriguez, J., Ridley, J., Rawlins, R., Sinha, B., Shin, B., Semple, A., Storkey, D., Stephens, D., Shim, T., Tomassini, L., Tsushima, Y., Titley, H., Tennant, W., Varma, V., Vellinga, M., Weedon, G., Williams, K., Yang, Y.-M., Zhao, M., Zhou, X., and Zhu, H.: Assessment of the Met Office Global Coupled model version 4 (GC4) configurations, Met Office Forecasting Research Technical Report, no. 661. https://doi.org/10.62998/uzui3766, 2024.

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

Yamazaki, K., Sexton, D. M. H., Rostron, J. W., McSweeney, C. F., Murphy, J. M., and Harris, G. R.: A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 2: global performance and future changes, Climate Dynamics, https://doi.org/10.1007/s00382-020-05608-5, 2021.

Yarger, D., Wagman, B. M., Chowdhary, K., and Shand, L.: Autocalibration of the E3SM Version 2 Atmosphere Model Using a PCA-Based Surrogate for Spatial Fields, Journal of Advances in Modeling Earth Systems, 16, e2023MS003961, https://doi.org/10.1029/2023MS003961, 2024.