JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2019MS001866

Special Section:
The UK Earth System Models for CMIP6

Forcings, Feedbacks, and Climate Sensitivity in HadGEM3-GC3.1 and UKESM1

Timothy Andrews[1], Martin B. Andrews[1], Alejandro Bodas-Salcedo[1], Gareth S. Jones[1], Till Kuhlbrodt[2], James Manners[1,3], Matthew B. Menary[1,4], Jeff Ridley[1], Mark A. Ringer[1], Alistair A. Sellar[1], Catherine A. Senior[1], and Yongming Tang[1]

[1]Met Office Hadley Centre, Exeter, UK, [2]NCAS, University of Reading, Reading, UK, [3]Global Systems Institute, Exeter University, Exeter, UK, [4]Now at LOCEAN/IPSL, Sorbonne Universités (SU)-CNRS-IRD-MNHN, Paris, France

Key Points:
- HadGEM3-GC3.1 and UKESM1 have climate sensitivities of 5.5 and 5.4 K, respectively
- Our models' forcing and feedback processes are not atypical of models in general
- The relatively large climate sensitivity arises from an unusual combination of forcing and feedback

Correspondence to:
T. Andrews,
timothy.andrews@metoffice.gov.uk

Citation:
Andrews, T., Andrews, M. B., Bodas-Salcedo, A., Jones, G. S., Kuhlbrodt, T., Manners, J., et al. (2019). Forcings, feedbacks, and climate sensitivity in HadGEM3-GC3.1 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/2019MS001866

Received 15 AUG 2019
Accepted 19 NOV 2019
Accepted article online 24 NOV 2019

Abstract Climate forcing, sensitivity, and feedback metrics are evaluated in both the United Kingdom's physical climate model HadGEM3-GC3.1 at low (-LL) and medium (-MM) resolution and the United Kingdom's Earth System Model UKESM1. The effective climate sensitivity (EffCS) to a doubling of CO_2 is 5.5 K for HadGEM3.1-GC3.1-LL and 5.4 K for UKESM1. The transient climate response is 2.5 and 2.8 K, respectively. While the EffCS is larger than that seen in the previous generation of models, none of the model's forcing or feedback processes are found to be atypical of models, though the cloud feedback is at the high end. The relatively large EffCS results from an unusual combination of a typical CO_2 forcing with a relatively small feedback parameter. Compared to the previous U.K. climate model, HadGEM3-GC2.0, the EffCS has increased from 3.2 to 5.5 K due to an increase in CO_2 forcing, surface albedo feedback, and midlatitude cloud feedback. All changes are well understood and due to physical improvements in the model. At higher atmospheric and ocean resolution (HadGEM3-GC3.1-MM), there is a compensation between increased marine stratocumulus cloud feedback and reduced Antarctic sea-ice feedback. In UKESM1, a CO_2 fertilization effect induces a land surface vegetation change and albedo radiative effect. Historical aerosol forcing in HadGEM3-GC3.1-LL is -1.1 W m^{-2}. In HadGEM3-GC3.1-LL historical simulations, cloud feedback is found to be less positive than in *abrupt-4xCO_2*, in agreement with atmosphere-only experiments forced with observed historical sea surface temperature and sea-ice variations. However, variability in the coupled model's historical sea-ice trends hampers accurate diagnosis of the model's total historical feedback.

Plain Language Summary A new generation of climate models—called HadGEM3-GC3.1 and UKESM1—have been developed in the United Kingdom and will be used widely in the Coupled Model Intercomparison Project Phase 6 (CMIP6). Evaluating the models' benchmark climate sensitivity and feedback metrics is a useful first step to understanding their characteristic response to forcing. The effective climate sensitivities are found to be higher than that seen in the previous generation of models, in common with other recently developed climate models. Reasons for this are discussed.

1. Introduction

Comprehensive models of the global climate system, known as coupled atmosphere-ocean general circulation models (AOGCMs), are essential tools for understanding climate processes and projecting future climate changes (e.g., Collins et al., 2013; Flato et al., 2013). Diagnosing a model's radiative forcings, feedbacks, and climate sensitivities is a useful first step to understanding its characteristic behavior in response to forcing (e.g., Andrews, Gregory, et al., 2012). In particular, such metrics can be used to understand a climate model's simulation of historical global temperature change as well as its projection of 21st-century climate change for a given emission scenario (e.g., Forster et al., 2013). For example, a model with a relatively large climate sensitivity is expected to provide a relatively large temperature change both globally and regionally for a given 21st-century emission scenario (Grose et al., 2018).

A new generation of climate models have recently been developed by the international community and will be widely used in the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016). In the previous generation of climate models, CMIP5 (Taylor et al., 2012), the effective climate sensitivity (EffCS) to a doubling of CO_2 ranged from 2.1 to 4.7 K across the model ensemble (Andrews, Gregory, et al., 2012; Flato et al., 2013). In the CMIP6 generation of models, some modelling centers are reporting

increased sensitivities relative to this range, with EffCS values closer to ~5.0 K and greater, for example, CNRM-CM6-1 (Voldoire et al., 2019), CESM2 (Gettelman et al., 2019), and E3SMv1 (Golaz et al., 2019). Here, we will show that a new generation of U.K. physical and Earth system climate models (HadGEM3-GC3.1 and UKESM1, respectively, see below) are likewise simulating higher climate sensitivities relative to the previous generation of models.

A move to higher climate sensitivities has important consequences. For example, larger climate sensitivities will lead to greater 21st-century warmings for a given emissions scenario (e.g., Grose et al., 2018), increase the risk of state-dependent feedbacks (e.g., Bloch-Johnson et al., 2015), and impact on allowable carbon budgets for a given target (Rogelj et al., 2019). However, determining whether such sensitivities are plausible is a challenging task, since it requires evaluating and combining many lines of evidence, such as from the historical record, process understanding, and paleorecords, all of which have particular difficulties in constraining the upper bound on climate sensitivity (e.g., Stevens et al., 2016).

Here, we document benchmark climate sensitivity metrics—such as the transient climate response (TCR), EffCS, and radiative feedback parameter (see below)—as well as radiative forcings, in the new generation of U.K. climate models being used in CMIP6:

1. *HadGEM3-GC3.1* is the third Hadley Centre Global Environmental Model (HadGEM3) run under the latest coupled configuration (GC3.1; Williams et al., 2017). We make use of two atmosphere-ocean resolutions: a low atmosphere and ocean resolution (N96ORCA1; 135 km atmosphere and 1° ocean) termed HadGEM3-GC3.1-LL and a medium atmosphere and ocean resolution (N216ORCA025; 60 km atmosphere and 0.25° ocean) termed HadGEM3-GC3.1-MM (Kuhlbrodt et al., 2018). Details of the model's configuration and performance are given elsewhere within this special issue of *JAMES*. For instance, Williams et al. (2017) describe the atmosphere, ocean, sea-ice, and land surface configurations in detail; Mulcahy et al. (2018) describe the developments in aerosol processes; and Kuhlbrodt et al. (2018) and Menary et al. (2018) provide a comprehensive description of the model's climatology and variability. Hardiman et al. (2019) describe an ozone redistribution scheme that is included in the climate change simulations used here.
2. *UKESM1* is the latest state-of-the-art U.K. Earth System Model. It builds on the low-resolution "physical climate" model HadGEM3-GC3.1-LL with the inclusion of, amongst other things, terrestrial carbon and nitrogen cycles—including an interactive vegetation model—ocean biochemistry and a unified tropospheric-stratospheric chemistry scheme (Sellar et al., 2019). Details of the model components, couplings, performance, and evaluation are given in Sellar et al. (2019).

An important characteristic of our model development process is traceability and consistency in physical parameters across the model hierarchy. For example, when changing horizontal resolution between HadGEM3-GC3.1-LL and HadGEM3-GC3.1-MM, only a small number of changes—mostly in the ocean (see table 1. of Kuhlbrodt et al., 2018)—are explicitly required and a retuning of the model (e.g., to the top-of-atmosphere [TOA] radiative balance) is not needed since resolution-dependent parameterizations are avoided (Williams et al., 2017). This mean differences in forcing, feedback, and climate sensitivity between the -LL and -MM configurations can confidently be attributed to differences in horizontal resolution between the configurations rather than any change in physical parameters that might have arisen if the models had needed to be retuned. The one exception is that the -LL configuration uses a slightly lower albedo for snow on sea ice compared to the -MM resolution to compensate for a sea-ice bottom melt that is too weak in the Arctic (Kuhlbrodt et al., 2018). However, we do not believe this parameter change has a significant impact on our results (see section 4). Similarly, in the construction of UKESM1, no retuning of the HadGEM3-GC3.1-LL physical parameters was required (Sellar et al., 2019). Hence, we interpret the differences in results between HadGEM3-GC3.1-LL and UKESM1 as due to the inclusion of Earth system processes. The one exception is a difference in the model parameterization related to the partial burying of vegetation by snow, which we discuss the impacts in section 5.

Section 2 describes the experiments used and methods. Section 3 diagnoses the global climate sensitivity and feedback metrics of each model to idealized CO_2 changes. Section 4 identifies which feedback processes depend on the atmospheric and ocean resolution. Section 5 identifies Earth system feedbacks from the coupling between CO_2 and vegetation changes in UKESM1. Section 6 investigates the linearity of feedbacks to evolving patterns of temperature change. Section 7 diagnoses the radiative forcing and feedbacks in

Table 1.
Summary of Climate Change Simulations Used in This Study and Their Purpose

Experiment name	MIP	Purpose (to calculate)	Model	Ensemble size	Run length (years)
abrupt-4xCO$_2$	DECK	"Long-term" effective sensitivities: EffCS, EffF_{2x}, λ, F_{2x}	HadGEM3-GC3.1-LL	1	150
			HadGEM3-GC3.1-MM	1	150
			UKESM1	1	150
1pctCO$_2$		Transient sensitivities: TCR, T140	HadGEM3-GC3.1-LL	4	150
			HadGEM3-GC3.1-MM	1	150
			UKESM1	4	150
historical		Historical feedbacks and temperature change	HadGEM3-GC3.1-LL	4	1850–2014
abrupt-4xCO$_2$-rad	n/a	Biogeochemical feedbacks/CO$_2$ fertilization effects	UKESM1	1	150
piClim-anthro	RFMIP	1850–2014 F_{ANT} timeslice	HadGEM3-GC3.1-LL	1	30
piClim-ghg		1850–2014 F_{WMGHG} timeslice	HadGEM3-GC3.1-LL	1	30
piClim-aer		1850–2014 F_{AER} timeslice	HadGEM3-GC3.1-LL	1	30
piClim-lu		1850–2014 F_{LU} timeslice	HadGEM3-GC3.1-LL	1	30
piClim-4xCO$_2$		F_{4xCO2}	HadGEM3-GC3.1-LL	1	30
piClim-histall		Historical $F(t)$ time series	HadGEM3-GC3.1-LL	1	1850–2014
piClim-histnat		Historical $F_{NAT}(t)$ time series	HadGEM3-GC3.1-LL	1	1850–2014
piClim-histghg		Historical $F_{WMGHG}(t)$ time series	HadGEM3-GC3.1-LL	1	1850–2014
piClim-histaer		Historical $F_{AER}(t)$ time series	HadGEM3-GC3.1-LL	1	1850–2014
amip-piForcing	CFMIP	Historical $\lambda(t)$ in response to observed SST and sea-ice changes	HadGEM3-GC3.1-LL	1	1870–2014

HadGEM3-GC3.1-LL historical simulations and contrasts them to feedbacks found in *abrupt-4xCO$_2$* simulations as well as to observed sea surface temperature (SST) and sea-ice variations. Finally, section 8 presents a summary and discussion.

2. Experiments and Methods

We predominantly make use of core CMIP6 "DECK" simulations (Eyring et al., 2016) with HadGEM3-GC3.1 at two different resolutions and UKESM1 but additionally explore historical forcing and feedback in HadGEM3-GC3.1-LL with Radiative Forcing Model Intercomparison Project (RFMIP; Pincus et al., 2016) and Cloud Feedback Model Intercomparison Project (CFMIP; Webb et al., 2017) simulations. CMIP5 data used for comparison purposes are the same as that used in Forster et al. (2013) and Andrews et al. (2015).

2.1. Control Simulations

The AOGCM control of each model configuration is *piControl*—that is, a spun-up control simulation with constant year 1850 forcing conditions (Eyring et al., 2016). The control climatology and variability of HadGEM3-GC3.1 is analyzed and evaluated in Menary et al. (2018). For diagnosing in HadGEM3-GC3.1-LL (section 7), we also have an atmosphere-only GCM (AGCM) configuration of the *piControl*, called *piClim-control* (Pincus et al., 2016). This is a 30-year AGCM simulation using constant 1850 forcing conditions (e.g., year 1850 greenhouse gas concentrations and aerosol emissions) and a monthly climatology of SST and sea ice derived over the first 50 years of the *piControl* simulation as boundary conditions for the atmospheric model, following the RFMIP protocol (Pincus et al., 2016).

2.2. Climate Change Simulations and Calculations

All climate change simulations used—along with their purpose in this study—are listed in Table 1., following CMIP6 experiment names. Details of each experimental protocol are given in Eyring et al. (2016) for the DECK *abrupt-4xCO$_2$*, *1pctCO$_2$*, and *historical* simulations; Pincus et al. (2016) for the suite of *piClim* RFMIP radiative forcing experiments; and Webb et al. (2017) for the CFMIP *amip-piForcing* simulation, though each will be briefly introduced where relevant in the following analysis.

All calculations that follow are based on annual mean quantities, and differences between control and perturbation simulations are calculated in parallel (i.e., year by year) before calculating global means. If one was only interested in global mean quantities, it might be preferable to remove a linear fit of the global annual mean *piControl* time series (e.g., Andrews, Gregory, et al., 2012; Forster et al., 2013). This would remove interannual *piControl* noise and unforced model drift (assuming it to be linear and equal in the control

Table 2
Global Mean CO_2 Forcing and Sensitivity Metrics

Model	F_{2x}	TCR	T140	EffCS	EffF_{2x}	λ_{NET}	λ_{LWcs}	λ_{SWcs}	λ_{CRE}
HadGEM3-GC3.1-LL	4.05	2.48	6.59	5.54	3.49	−0.63	−1.80	0.66	0.51
HadGEM3-GC3.1-MM	4.00	2.66	6.35	5.44	3.55	−0.65	−1.82	0.57	0.60
UKESM1	4.02	2.76	6.61	5.36	3.60	−0.67	−1.88	0.71	0.50
HadGEM3-GC2.0	3.84	2.11	4.61	3.18	3.35	−1.05	−1.74	0.44	0.24
CMIP5 mean ± (5–95%)	4.08 ± 0.80[a]	1.82 ± 0.63[b]	4.58 ± 1.18[c]	3.22 ± 1.32[b]	3.44 ± 0.84[b]	−1.13 ± 0.51[b]	−1.81 ± 0.25[b]	0.71 ± 0.24[b]	−0.04 ± 0.53[b]

Note. F_{2x} is the radiative forcing from a doubling of CO_2, calculated as the ΔN-axis intercept from the regression of ΔN against ΔT in the first 5 years of the *abrupt-4xCO_2* simulation, divided by 2 (Armour, 2017). TCR is the ΔT at the point of CO_2 doubling (year 70) in the *1pctCO_2* simulation, calculated as the mean over years 61–80 (Gregory & Forster, 2008). T140 is the ΔT at the point of CO_2 quadrupling (year 140) in the *1pctCO_2* simulation, calculated as the mean over years 131–150. This is a slightly different T140 calculation to that given in Gregory et al. (2015), since they were limited to CMIP5 simulations that only ran for 140 years. TCR and T140 values for HadGEM3-GC3.1-LL and UKESM1 represent the mean values from the four ensemble members. EffCS = $-$EffF_{2x}/λ_{NET} is the effective climate sensitivity, where EffF_{2x} is the effective 2xCO_2 forcing and λ_{NET} the climate feedback parameter, calculated from the regression of ΔN against ΔT for all 150 years of the *abrupt-4xCO_2* simulation (EffF_{2x} is the ΔN-axis intercept divided by 2; λ_{NET} is the slope of the linear fit; Andrews, Gregory, et al., 2012). λ_{LWcs}, λ_{SWcs}, and λ_{CRE} are the longwave clear-sky, shortwave clear-sky, and cloud radiative effect components of the feedback parameter, respectively, calculated by regressing the change in radiative component against ΔT for all 150 years of the *abrupt-4xCO_2* simulation (Andrews, Gregory, et al., 2012). CMIP5 values come from the studies indicated; the 5–95% uncertainty ranges represent ±1.645 standard deviations across the individual model results.
[a]Armour (2017). [b]Forster et al. (2013). [c]Gregory et al. (2015).

and perturbation run). However, assuming linear drift at a grid box level might introduce errors. Hence, since we are interested in regional results as well as global means, we prefer our method as this removes unforced model drift and ensures regional results are consistent with global mean analysis. However, this is at the expense of including interannual *piControl* noise in the results.

Details of how the benchmark forcing, feedback, and sensitivity metrics are calculated are given in the caption of Table 2. All our radiative flux and sensitivity terms are defined as positive downwards, so a positive number represents a heat gain with warming (a positive, destabilizing, feedback), and a negative number represents a heat loss to space with warming (a negative, stabilizing, feedback). Thus, for example, the total climate feedback parameter (λ_{NET}, in W m^{-2} K^{-1}) is negative using our sign convention, since the climate system is overall stable.

3. Global Sensitivity and Feedbacks Metrics to Idealized CO_2 Changes

Figures 1a and 1b show the global annual mean surface air temperature change, ΔT, from the *1pctCO_2* and *abrupt-4xCO_2* experiments, respectively. Compared to the CMIP5 generation of models, both HadGEM3-GC3.1-LL and UKESM1 warm more in response to increased CO_2. Indeed, both benchmark sensitivity metrics—the TCR (2.48 to 2.76 K) and EffCS (5.54 to 5.36 K; Table 2)—are at or above the CMIP5 5–95% ranges of 1.2 to 2.5 K (for TCR) and 1.9 to 4.5 K (for EffCS; see caption of Table 2 for details of calculations and definitions). This is found to be the case in other CMIP6 models too, such as CNRM-CM6-1 (Voldoire et al., 2019), CESM2 (Gettelman et al., 2019), and E3SMv1 (Golaz et al., 2019). Indeed, both TCR and EffCS are found to be very similar to the E3SMv1 model (TCR = 2.9 K; EffCS = 5.3 K; Golaz et al., 2019).

To understand the simulated climate sensitivities of the HadGEM3-GC3.1-LL and UKESM1 configurations further, we decompose EffCS into its effective radiative forcing from a doubling of CO_2 (EffF_{2x}) and climate feedback parameter (λ_{NET}) following Andrews, Gregory, et al. (2012), noting that EffCS = $-$EffF_{2x}/λ_{NET}. As per Andrews, Gregory, et al. (2012), we further decompose the radiative feedback parameter into its radiative components: longwave clear-sky (λ_{LWcs}), shortwave clear-sky (λ_{SWcs}), and cloud radiative effect (λ_{CRE}). Cloud feedback (λ_{CRE}) in UKESM1 and the HadGEM3-GC3.1-LL configurations are found to be at the upper end of the CMIP5 5–95% range (Table 2) but below the maximum found in CMIP5 (0.7 W m^{-2} K^{-1}, Andrews, Gregory, et al., 2012). The other feedback processes (λ_{LWcs} and λ_{SWcs}) are relatively close to the CMIP5 mean, as is EffF_{2x} (Table 2).

If none of the individual forcing or feedback processes in the model are unprecedented compared to CMIP5 models, what then leads to the relatively high EffCS in these two model configurations? First, in CMIP5, there existed an anticorrelation between EffF_{2x} and λ_{NET} (Andrews, Gregory, et al., 2012; Chung & Soden,

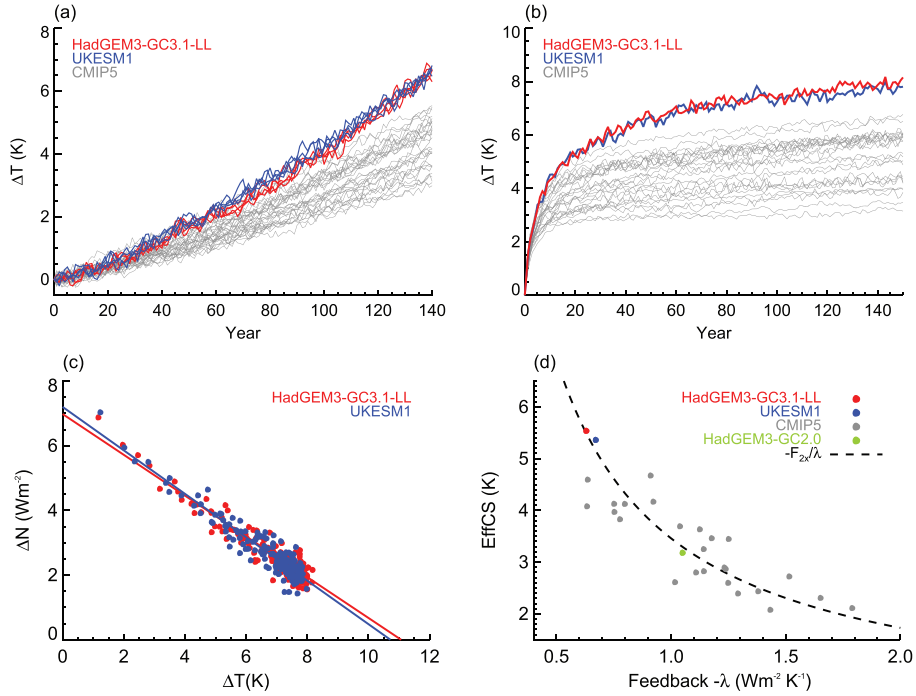

Figure 1. Global annual mean surface air temperature change (ΔT) in the (a) *1pctCO$_2$* and (b) *abrupt-4xCO$_2$* simulations. (c) Gregory plot used to estimate the EffCS. (d) Relationship between EffCS and feedback parameter (EffCS = $-F_{2x}/\lambda$); the dotted line uses the CMIP5 multimodel mean EffF_{2x}.

2017; Ringer et al., 2014). This tended to reduce the impact of a small λ_{NET} on EffCS, since models with a small λ_{NET} also had a small EffF_{2x}. This is evident in Figure 1d, which shows the relationship between EffCS and λ_{NET}, the dotted line representing the multimodel mean EffF_{2x}. All CMIP5 models with λ_{NET} < 0.9 W m^{-2} K^{-1} (six in total) have an EffF_{2x} below that of the multimodel mean (dotted line) and sometimes substantially so. In contrast, in UKESM1 and HadGEM3-GC3.1-LL, while λ_{NET} is small (but not unprecedented), a small λ_{NET} and an EffF_{2x} close to—or just above—the CMIP5 mean is unprecedented. It is this unusual combination of forcing and feedback that leads to a larger EffCS than seen in the CMIP5 generation of models. The results of Golaz et al. (2019; their figure 28) suggest this may be true for the E3SMv1 CMIP6 model too.

What model developments have led to the models relatively high EffCS? Bodas-Salcedo et al. (2019) tested the impact of all atmospheric model developments on λ_{NET} between HadGEM3-GC3.1-LL and the previous HadGEM configuration HadGEM3-GC2.0 (Williams et al., 2015), which had a "typical" EffCS of ~3 K (Senior et al., 2016). (Note that the CMIP5 configuration was HadGEM2-ES, but the vast amount of model developments since then means a traceable link between model development and changes in climate sensitivity is impractical). Bodas-Salcedo et al. (2019) found that cloud feedback in the midlatitudes substantially increased between HadGEM3-GC2.0 and HadGEM3-GC3.1-LL. This was due to (i) the inclusion of a mixed-phase cloud scheme that reduced the strength of preexisting negative cloud feedbacks in that region for physically well understood reasons, bringing the model more in line with observations, and (ii) the inclusion of a new aerosol scheme that suppressed a strong negative feedback that operated through a reduction in cloud droplet size with warming. However, Bodas-Salcedo et al. (2019) were limited to understanding the atmospheric response to prescribed SST increases only, necessarily omitting coupled atmosphere-ocean effects and sea-ice feedbacks, which we now go on to evaluate here.

We compare the feedbacks in the HadGEM3-GC2.0 and HadGEM3-GC3.1-LL coupled atmosphere-ocean *abrupt-4xCO$_2$* simulations in Table 2 and total feedback, λ_{NET}, in Figure 1d. As expected, we find that the total feedback has decreased (from −1.05 W m^{-2} K^{-1} in HadGEM3-GC2.0 to −0.63 W m^{-2} K^{-1} in HadGEM3-GC3.1-LL, Table 2) and that the results of Bodas-Salcedo et al. (2019) pull through into the coupled model's feedbacks and EffCS, that is, that cloud feedback has doubled from 0.24 to ~0.5 W m^{-2} K^{-1}

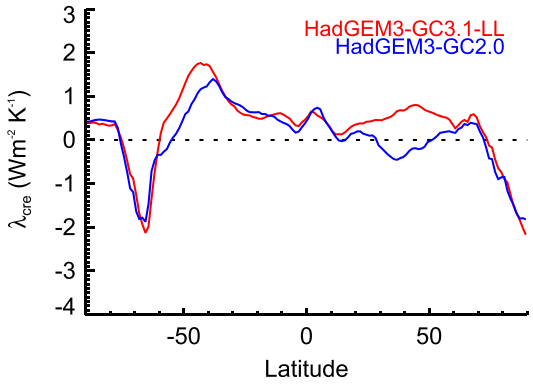

Figure 2. Zonal-mean cloud feedback parameter (λ_{CRE}) in the HadGEM3-GC3.1-LL and HadGEM3-GC2.0 *abrupt-4xCO$_2$* simulations.

(Table 2). As in Bodas-Salcedo et al. (2019), the biggest changes in cloud feedback is at midlatitudes (Figure 2), due to the reasons discussed above (i.e., the inclusion of a mixed-phase cloud scheme and improved aerosol-cloud interaction processes). However, we also identify two other reasons for the increased EffCS of the HadGEM model that Bodas-Salcedo et al. (2019) were unable to identify: (i) the SW clear-sky radiative feedback parameter, λ_{SWcs}, is substantially more positive, and (ii) the effective radiative forcing from a doubling of CO_2 (EffF_{2x}) has increased (Table 2).

We attribute the increase in λ_{SWcs} mostly to reductions in the Southern Ocean SST warm bias and consequent increases (improvements) in Antarctic sea-ice extent and volume in HadGEM3-GC3.1-LL compared to HadGEM3-GC2.0 (Williams et al., 2017; see their figure 7); the enhanced sea-ice extent in the HadGEM3-GC3.1-LL climatology leading to a greater sea-ice feedback. The Antarctic sea-ice volume and extent was too low in HadGEM3-GC2.0 (Williams et al., 2017) and resulted in an unrealistically low sea-ice feedback (Senior et al., 2016). A Southern Ocean warm bias is a common problem in AOGCMs (e.g., Sallée et al., 2013), and we have found that model developments aimed at improving the climatology of this region may either directly (e.g., through changes to mixed-phase cloud parameterizations and so feedback, e.g., Bodas-Salcedo et al., 2019) or indirectly (through changes in SST and sea-ice climatology/feedback) affect climate feedbacks and sensitivity.

The increase in EffF_{2x} can mostly be explained by improvements in the treatment of greenhouse gas absorption (see section 3.2.1 in Walters et al., 2019). Indeed, Figure 3 shows the CO_2 radiative forcing at the tropopause as determined from offline radiative transfer calculations averaged over varying background humidities following the methodology of Pincus et al. (2015) using the treatment of greenhouse gas absorption in HadGEM3-GC2.0 as well as in HadGEM3-GC3.1-LL, both compared to a reference 300 band case. The CO_2 radiative forcing in HadGEM3-GC3.1-LL is an improvement on that seen in HadGEM3-GC2.0, which had unrealistically low forcing compared to the reference case, and performs well in the context of other radiative transfer schemes and parameterizations (Pincus et al., 2015). At 2xCO_2, the offline radiative transfer calculations give an increase in forcing of ~0.3 W m^{-2} going from HadGEM3-GC2.0 and HadGEM3-GC3.1-LL. To see whether this difference in forcing due to the treatment of radiative transfer pulls through to the effective radiative forcing, we eliminate any impact of nonlinearity in the regression method of calculating EffF_{2x} by using only the first 5 years of *abrupt-4xCO$_2$* following Armour (2017), termed F_{2x} in Table 2. F_{2x} has increased by ~0.2 W m^{-2} in HadGEM3-GC3.1-LL compared to HadGEM3-GC2.0. This is slightly smaller than the difference in the offline radiative transfer calculations, suggesting that other processes—such as cloud adjustments perhaps—that are included in the effective radiative forcing but not in the offline radiative transfer calculations also differ and perhaps slightly compensate.

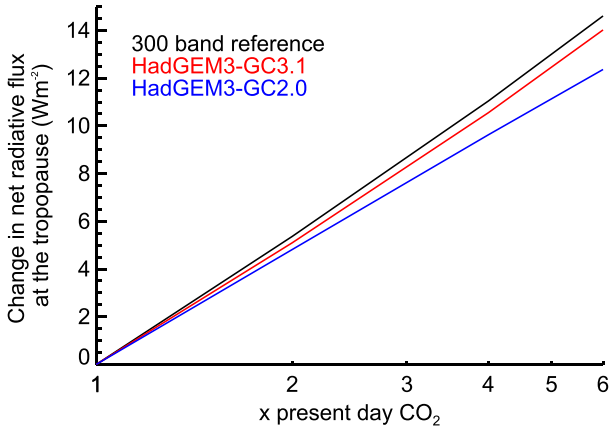

Figure 3. CO_2 radiative forcing at the tropopause from offline radiative transfer calculations using the treatment of greenhouse gas absorption in (blue) HadGEM3-GC2.0 and (red) HadGEM3-GC3.1 compared to a (black) reference 300 band case. Note that the *x* axis has a logarithmic scale. Each calculation is averaged over four different profiles with varying humidities following the method of Pincus et al. (2015).

4. Feedback Dependence on Atmospheric and Ocean Resolution

In this section, we make use of the traceability and consistency in physical parameters across our model hierarchy (discussed in section 1) to investigate whether any of these feedback and sensitivity results depend on the horizontal resolution of the model. We do this by contrasting the results between the HadGEM3-GC3.1-LL and HadGEM3-GC3.1-MM configurations.

Figure 4a and Table 2 show that the relatively high sensitivity of HadGEM3-GC3.1 is largely independent of its low (-LL) and medium (-MM) atmosphere-ocean resolution configurations. However, considering EffCS alone masks compensating processes and regional differences. Table 2 shows that λ_{SWcs} is smaller at the higher atmosphere-ocean resolution configuration (-MM), but this is mostly compensated for by a larger

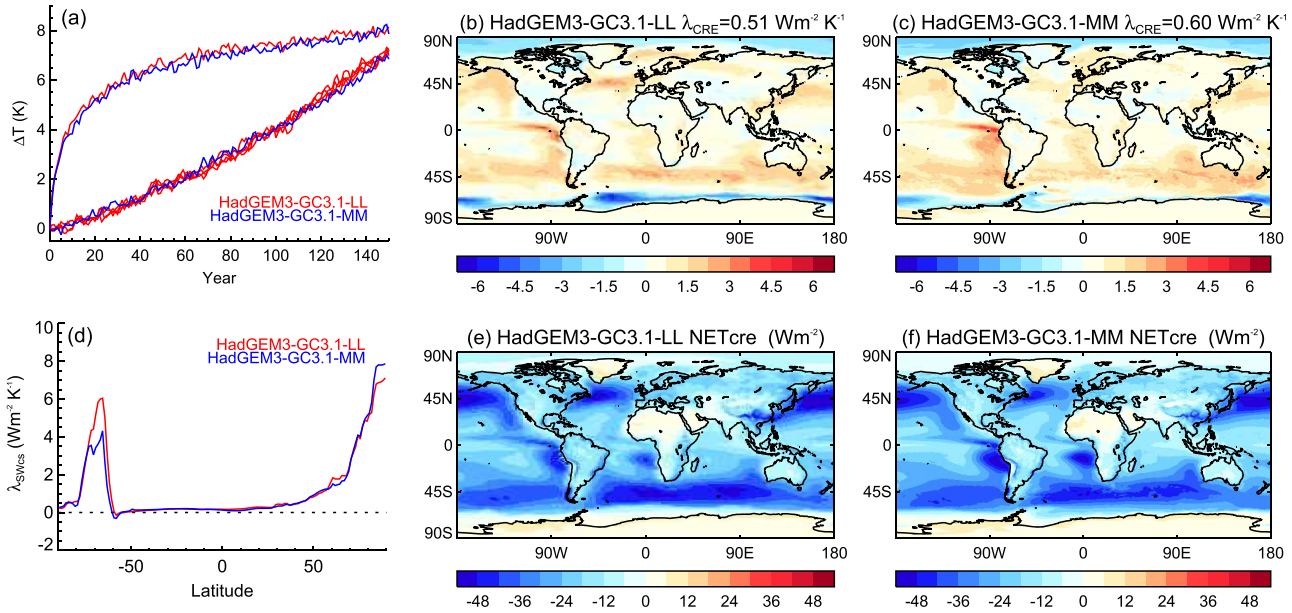

Figure 4. (a) Global annual mean surface air temperature change (ΔT) in the *1pctCO*$_2$ and *abrupt-4xCO*$_2$ simulations using a low (-LL) and medium (-MM) atmosphere-ocean resolution configuration of HadGEM3-GC3.1. Geographical distribution of the cloud feedback parameter (λ_{CRE}) in the (b) low- and (c) medium-resolution configurations, from *abrupt-4xCO*$_2$. (d) Comparison of the zonal-mean SW clear-sky feedback parameter (λ_{SWcs}) from *abrupt-4xCO*$_2$. Geographical distribution of the *piControl* cloud radiative effect climatology in the (e) low- and (f) medium-resolution configurations.

cloud feedback (λ_{CRE}). Why is this? Contrasting Figures 4b and 4c reveals a more positive and extensive cloud feedback in the regions of tropical marine stratocumulus decks, such as in the tropical southeast Pacific (off the western coast of South America) and tropical southeast Atlantic (off the western coast of Africa), in HadGEM3-GC3.1-MM. These are regions of oceanic upwelling and extensive climatological low clouds that form above the relatively cold SSTs and sit under a temperature inversion that caps the boundary layer. Kuhlbrodt et al. (2018) showed significant warm SST biases in these coastal upwelling regions in the lower-resolution configuration (-LL), which were improved at higher atmosphere-ocean resolution due to the better representation of upwelling. Gent et al. (2010) provide a possible mechanism for this, in which the better resolved orography at higher atmospheric resolution allows stronger winds to develop closer to the coasts in the upwelling regions, increasing the upwelling and reducing the SST warm bias. A consequence of the colder SSTs is to increase the cloudiness in these regions (Gent et al., 2010). Indeed, Figures 4e and 4f show that the elimination of the warm bias in these upwelling regions has increased the climatological cloudiness (measured by the climatological cloud radiative effect, which becomes more negative and extensive in the -MM configuration). The consequence of increased cloudiness in these regions is a more positive and extensive low cloud feedback.

In contrast, the higher atmospheric-ocean resolution configuration has a smaller cloud feedback in North Atlantic (Figures 4b and 4c). This is because of a poorer representation of the North Atlantic Current—which is too zonal—at the lower-resolution model, leading to a significant SST cold bias (up to 6 K) in the northwest Atlantic (see Kulhbrodt et al., 2019). This is a common problem for 1° ocean models (Danabasoglu et al., 2014). The improved representation of the northward and eastward flow of the North Atlantic Current in the higher ocean resolution configuration reduces this cold bias, reducing cloudiness (Figures 4e and 4f) and cloud feedback (Figures 4b and 4c). This is dependent on the ocean resolution rather than atmospheric resolution, since Storkey et al. (2018) showed significant improvements in the North Atlantic Current with ocean-only simulations at different resolutions, unlike the coastal upwelling regions discussed previously that likely depend on the atmospheric resolution too.

Finally, while these differences are all expected to be improvements at higher resolution, Kulhbrudt et al. (2019) also showed that the higher-resolution configuration had a worse Southern Ocean warm SST bias than the lower-resolution model. Consequently, the -MM configuration has much less Antarctic sea-ice extent and volume than the -LL configuration (see Kuhlbrodt et al., 2018; their figure 15), which likely

impacts on the sea-ice feedback in this region. Indeed, Figure 4d shows the zonal-mean λ_{SWcs} feedback parameter at the two resolutions; it is clear that the reduced Antarctic sea-ice due to the SST warm bias in the -MM configuration results in a much smaller albedo feedback in this region. As noted in section 1, one physical parameter that is different between the -LL and -MM configurations is a slightly lower albedo for snow on ice (Kulhbrodt et al., 2018; their table 1.). However, this ought to reduce (albeit only very slightly) the radiative effect of a given sea-ice fraction change in the -LL configuration, since the contrast in albedo between snow-covered ice and open water is reduced, which is the opposite of what we find. Hence, we do not think this can explain the different sea-ice feedback results between the -LL and -MM configurations.

In this section, we have shown that the high EffCS of the HadGEM3-GC3.1 model is largely independent of horizontal resolution, but this results from a fortuitous cancellation of changes to individual feedback processes. There exists a compensation between an increased marine stratocumulus cloud feedback and a reduced Antarctic sea-ice feedback at the higher-resolution configuration, both linked to changes in the underlying climatology resulting from the change in resolution. While the more positive cloud feedback is linked to an improved SST climatology in the regions of low clouds, the compensating sea-ice feedback change results from a Southern Ocean SST warm bias that is worse at the higher-resolution configuration.

5. Earth System Feedbacks

Our model hierarchy also allows us to investigate the impact of including Earth system processes in our model configuration on forcing, feedback, and sensitivity. We do this by comparing the results of HadGEM3-GC3.1-LL and UKESM1 (see section 1). Similarity between the global forcing and feedback metrics in HadGEM3-GC3.1-LL and UKESM1 (Table 2) suggests that—at the global scale—the inclusion of Earth system processes in the physical climate model does not substantially alter the top-level sensitivity metrics. Andrews, Ringer, et al. (2012) found this to be the case with the previous Hadley Centre Earth System Model HadGEM2-ES also but due to compensating Earth system processes.

Here, we follow the methodology of Andrews, Ringer, et al. (2012) of performing an additional *abrupt-4xCO_2* experiment with UKESM1, but this time, only the radiation scheme sees the increased CO_2 level (the vegetation continues to see control CO_2 levels)—termed *abrupt-4xCO_2-rad*. Differencing *abrupt-4xCO_2* and *abrupt-4xCO_2-rad* allows the impact of CO_2 stomatal and fertilization effects to be quantified, which has been identified as an important Earth system response in many models (e.g., Bala et al., 2006; Matthews, 2007; O'Ishi et al., 2009). Assuming linearity, this difference is equivalent to the biogeochemically (BGC) coupled CO_2 experimental designs of C4MIP (the Coupled Climate-Carbon Cycle Model Intercomparison Project; Jones et al., 2016) where only the carbon cycle sees the increased CO_2, except here, we have applied it to the *abrupt-4xCO_2* simulation rather than the *1pctCO_2* simulation used in C4MIP. Hence, we refer to this difference as BGC. The assumption of linearity is supported by Gregory et al. (2009); their figure 3a), who showed in *1pctCO_2* simulations that the global temperature evolution, ΔT, seen in "BGC coupled" and "rad coupled" simulations combined linearly to give ΔT evaluated in "fully coupled" simulations.

When only the radiation scheme of UKESM1 is forced with 4xCO_2 (*abrupt-4xCO_2-rad*), the global ΔT increase is smaller (Figure 5a), with the difference (BGC, averaged over the last 50 years of the simulations) large (1 to 2 K and more) over most continental regions (Figure 5b). In other words, including the BGC coupling with the CO_2 change increases the warming in the model. Applying the same sensitivity analysis of section 3 reveals that the EffCS reduces from 5.4 to 5.1 K *in abrupt-4xCO_2-rad*, due to a reduction in SW clear-sky feedback (λ_{SWcs}; from 0.71 to 0.67 W m^{-2} K^{-1}). This difference (the BGC feedback) is shown in Figure 5c, showing more positive feedbacks over many continental regions and especially over the Northern Hemisphere midlatitudes to high latitudes.

The increased EffCS and radiative feedback due to the BGC coupling come about due to CO_2 fertilization effects whereby the extra CO_2 encourages growth of (darker) trees at expense of (brighter) grasses, reducing the surface albedo and hence increasing the amount of solar radiation absorbed (e.g., Betts, 2000; Bala et al., 2006; Matthews, 2007; O'Ishi et al., 2009). This is clear in Figures 5c–5f, which show substantial increases in trees (Figure 5e) at the expense of grasses (Figure 5f; the area-weighted spatial correlation $r = -0.78$),

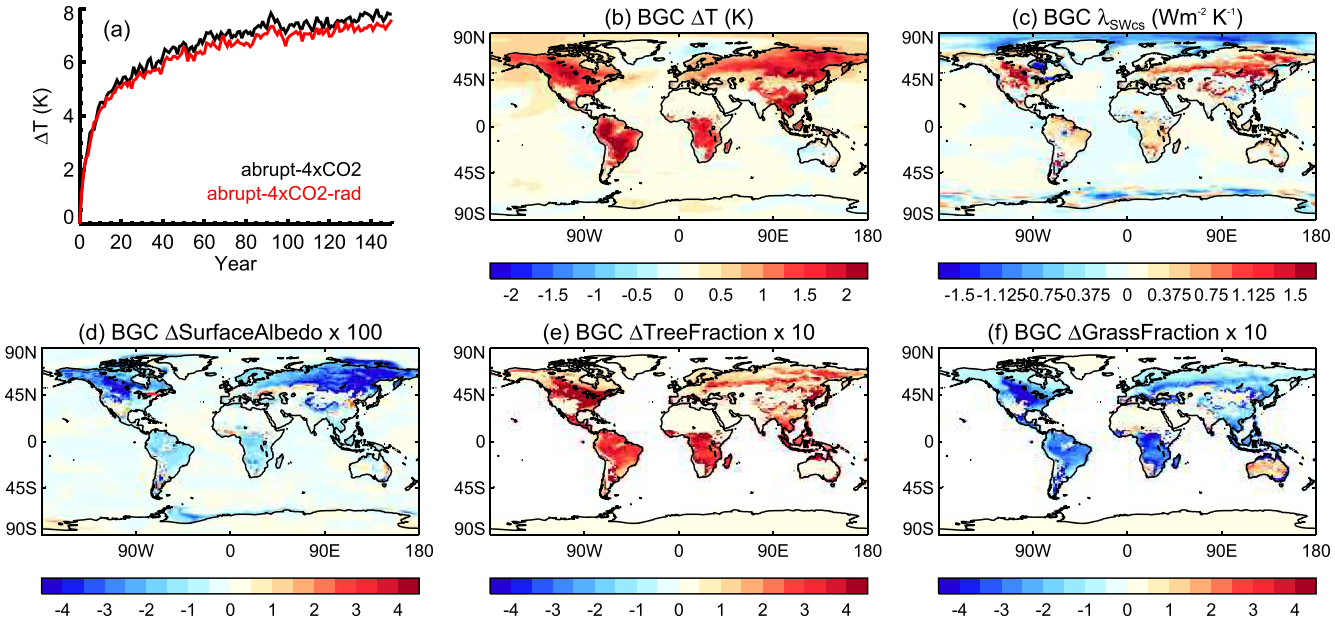

Figure 5. (a) Global annual mean surface air temperature change (ΔT) in the UKESM1 *abrupt-4xCO$_2$* and *abrupt-4xCO$_2$-rad* (only the radiation scheme sees the increased CO$_2$). (b) Difference in ΔT between *abrupt-4xCO$_2$* and *abrupt-4xCO$_2$-rad*, termed the biogeochemical (BGC) response, averaged over the last 50 years of the simulation. (c) The difference in SW clear-sky feedback parameter (λ_{SWcs}), for example, the BGC component of the radiative feedback. The BGC change averaged over the last 50 years of the simulations in (d) surface albedo, (e) tree fraction, and (f) grass fraction.

reducing the surface albedo (Figure 5d) and hence increasing the SW radiative feedback parameter (Figure 5c). A particularly important region is the expansion of the boreal forest that co-locates with seasonal snow cover, reducing the surface albedo even more than when trees simply replace grasses. For example, despite large vegetation changes in the tropics (Figures 5e and 5f), the albedo and radiative effect is much smaller than found at midlatitudes to high latitudes (Figures 5c and 5d).

A potential issue with this vegetation-snow BGC response in UKESM1 is that it co-locates with a region of model bias. Comparison of UKESM1 against satellite observations shows that this region is too bright in the present day and is likely to be too bright in the *piControl* also (Sellar et al., 2019). This bias is related to model parametrization of the partial burying of vegetation by snow, and Sellar et al. (2019) conclude that this effect may be too strong in the model, for grasses in particular. If regions of partially snow-covered vegetation are too bright climatologically, then when the forests expand due to the CO$_2$ fertilization effect in these regions, the respective surface albedo change (and so λ_{SWcs}) maybe overstated. Unfortunately, there are difficulties in the parameterization of surface albedo from snow-vegetation interactions (e.g., Bright et al., 2015), and Earth system processes are not as well understood or constrained as many other drivers of climate change (Collins et al., 2011), so the fidelity of the BGC response is difficult to quantify. However, here, we have highlighted an Earth system response to CO$_2$ change in UKESM1 that would be useful to quantify in other models, perhaps revealing an important source of uncertainty in Earth System Model feedbacks and land temperature trends.

Here, we have highlighted a small but positive feedback associated with the inclusion of Earth system processes, yet the top-level sensitivity metrics (EffCS) are similar or—if anything—slightly smaller (Table 2). Hence, there must be compensating negative feedback processes in the Earth System Model that are not included in HadGEM3-GC3.1-LL that we are currently unable to identify. Table 2 shows that LW clear-sky feedback processes (λ_{LWcs}) are slightly more negative in UKESM1 compared to HadGEM3-GC3.1-LL (λ_{LWcs} is equal to -1.88 and 1.80 W m^{-2} K^{-1}, respectively). This term is typically associated with the temperature (Plank and lapse rate) and water vapor feedbacks, yet there is no reason to suspect that the inclusion of Earth system processes has altered these. Perhaps more likely are new feedback processes included in UKESM1. In particular, UKESM1 includes a unified tropospheric-stratospheric chemistry scheme (Sellar et al., 2019) that would allow for changes in atmospheric composition with ΔT, from ozone, for example, and will be the focus of future work.

6. *abrupt-4xCO$_2$* Pattern Effects

Recent work has shown feedbacks to vary within a given model due to an evolving pattern of surface warming in *abrupt-4xCO$_2$* simulations (e.g., Andrews et al., 2015; Andrews, Gregory, et al., 2012; Geoffroy et al., 2013; Rugenstein et al., 2016), referred to as "pattern effects" (Stevens et al., 2016). Andrews et al. (2015) quantified this effect in CMIP5 models by comparing λ diagnosed from the first 20 years and remaining 130 years of the *abrupt-4xCO$_2$* simulation. We repeat their analysis here with HadGEM3-GC3.1-LL and UKESM1 and compare against CMIP5 AOGCMs in Figure 6, to see if these models are unusual in this behavior given their high sensitivities.

Figure 6a shows that both models are rather linear, unlike most CMIP5 models: λ_{NET} from years 1–20 and years 21–150 are similar (Figure 6b, i.e., these models lie close to the 1:1 line), while most CMIP5 models show a decrease in λ_{NET} (i.e., fall below the 1:1 line; larger EffCS) in the later period. We therefore rule out any large change in feedback strength as a reason for the large EffCS in HadGEM3-GC3.1-LL and UKESM1. However, we see that the EffCS from the first 20 years of *abrupt-4xCO$_2$* is larger than any CMIP5 model, but the EffCS from years 21 to 150 is not (Figure 6c). It is therefore unusual (relative to CMIP5) to have such a large EffCS early in the *abrupt-4xCO$_2$* simulation but not unusual if using an EffCS definition based on years 21–150 of *abrupt-4xCO$_2$*.

The reason for the increased EffCS sensitivity in the later (years 21–150) compared to the early (years 1–20) part of the *abrupt-4xCO$_2$* simulation (albeit only a small increase in HadGEM3-GC3.1-LL and UKESM1 models) follows that of other CMIP5 models. That is, SW cloud feedback increases during the simulation, while other feedback processes mostly remain unchanged (Figure 6d). Note that Figure 6d also shows that the LW and SW components of the model's cloud feedback are outside the CMIP5 range, while the net cloud feedback is not (section 3). This has been a feature of the previous Hadley Centre model HadGEM3-GC2.0 too and relates to reductions in high cloud with surface warming that have opposing LW and SW CRE changes (Senior et al., 2016), though this cancellation need not be exact (e.g., Mauritsen & Stevens, 2015).

7. Historical Forcing and Feedback

In this final section, we make use of RFMIP experiments to diagnose the historical radiative forcings in HadGEM3-GC3.1-LL and then use them in conjunction with the model's coupled atmosphere-ocean *historical* simulation to determine the model's historical feedbacks and EffCS. We then compare these feedbacks to that found in *abrupt-4xCO$_2$*, as well as in response to observed historical SST and sea-ice variations. Note that evaluation of the model's historical simulation against observations—including a rigorous evaluation of its temperature trends—will be given elsewhere in this special issue.

7.1. Historical Forcing

Table 3 documents the 1850–2014 effective radiative forcings (ERF) in HadGEM3-GC3.1-LL diagnosed from the change in radiative balance in 30-year fixed-SST timeslice experiments following the RFMIP tier 1 experiments (Forster et al., 2016; Pincus et al., 2016). The ERFs represent the change in radiative balance (averaged over the 30-year experiments) between a fixed-SST control experiment (*piClim-control*, see section 2.1), based on 1850 forcing conditions and a perturbation experiment that is identical except the relevant forcing agent (e.g., aerosol emissions) changed to year 2014 levels. The present-day aerosol ERF (equal to the mean change in net TOA radiative balance between *piClim-control* and a perturbation experiment—*piClim-aer*—with anthropogenic aerosol emissions of sulfur dioxide (SO_2), black carbon, and organic carbon changed from year 1850 to 2014) is -1.10 W m^{-2}. This aerosol forcing is close to the CMIP5 multimodel mean (-1.2 ± 0.5 W m^{-2} [5–95%]) reported in Zelinka et al. (2014) and counter to the paradigm that models with large climate sensitivities have large aerosol forcings (e.g., Kiehl, 2007), consistent with Forster et al. (2013) who found no relationship between historical forcing and climate sensitivity in the CMIP5 ensemble.

The radiative components (LW/SW clear-sky and CRE) of the ERF are also shown in Table 3 and are consistent with physical expectations of how each climate driver perturbs the Earth's energy budget. For example, the forcing from well-mixed greenhouse gases (F_{WMGHG}) acts predominantly in the LW clear-sky (since greenhouse gases principally reduce outgoing LW radiation), while aerosol forcing (F_{AER}) has a large negative component in SW clear-sky (due to aerosol-radiation interactions, e.g., the direct scattering and

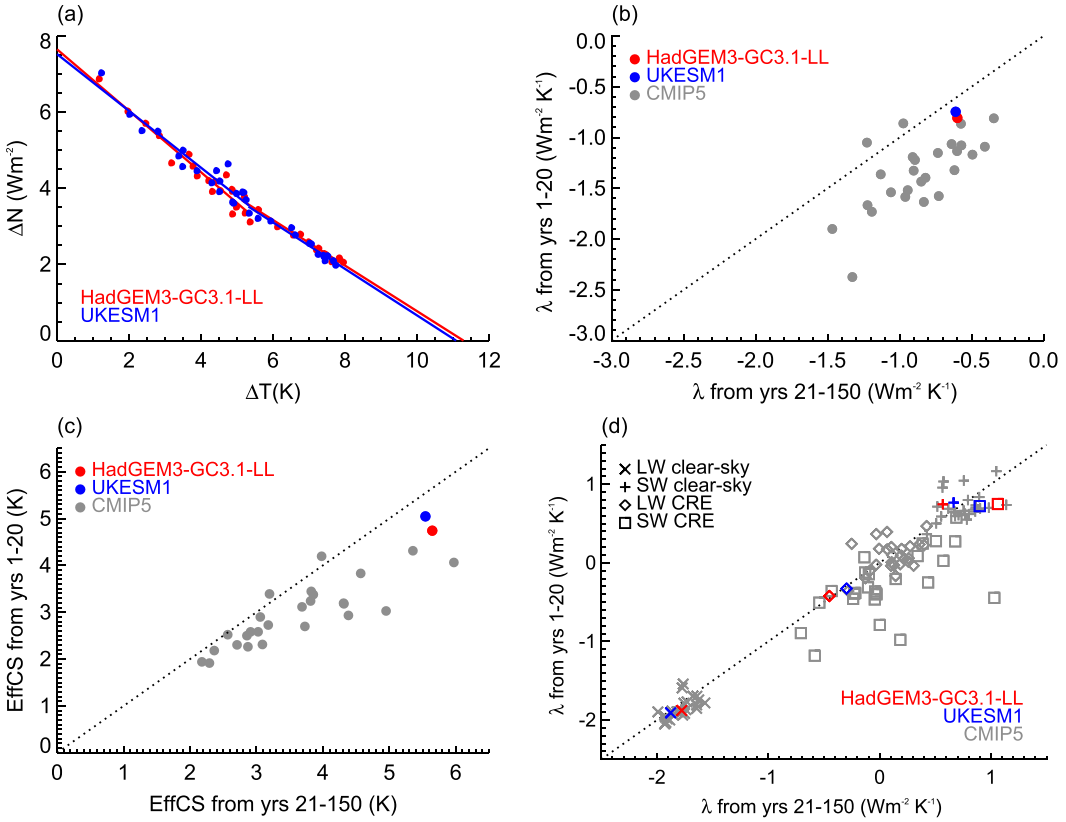

Figure 6. (a) Gregory plot separated at year 20, that is, a linear regression fit for years 1–20 and 21–150, for HadGEM3-GC3.1-LL and UKESM1 *abrupt-4xCO$_2$* simulations. (b) Comparison of the feedback parameter derived the two time periods across models. (c) Comparison of the EffCS derived from the two time periods. (d) Comparison of the feedback parameter components derived the two time periods across models.

absorption of SW radiation) and clouds (due to aerosol-cloud interactions). Land use forcing (F_{LU}; predominantly from deforestation) increases the surface albedo, so its largest component is in SW clear-sky (Table 3).

Figure 7a shows the annual mean historical time series of the total (anthropogenic plus natural), WMGHG, aerosol, and natural only (F_{NAT}) forcings (natural forcings include time-varying volcanic and solar forcings, including their effect on stratospheric O_3). These are derived from analogous fixed-SST experiments but with time-varying forcing constituents following the RFMIP tier 2 transient experiments (Pincus et al., 2016) and are found to agree with the 2014 timeslice experiments (Table 3) at the present day (filled circles, Figure 7a). The model's negative aerosol forcing time series largely balances the positive forcing from increases in WMGHGs up to the point in which aerosol forcing levels off around the 1970s (Figure 7a), after which

Table 3
1850 to 2014 Effective Radiative Forcings Simulated by HadGEM3-GC3.1-LL

Forcing (W m^{-2})	Includes	F_{NET}	F_{LWcs}	F_{SWcs}	F_{CRE}
F_{ANT} (total anthropogenic)	WMGHGs, O_3, aerosols, and land use	1.81	3.36	−1.08	−0.47
F_{WMGHG} (well-mixed GHGs)	WMGHGs only	3.09	2.94	0.16	−0.01
F_{AER} (anthropogenic aerosols)	Anthropogenic SO_2, OC, and BC	−1.10	0.20	−0.82	−0.48
F_{LU} (land use)	Land use only	−0.11	0.07	−0.27	0.09
F_{4x} (4xCO_2)	4xCO_2 only	8.08	7.27	0.28	0.53

Note. Numbers are diagnosed from the change in radiative balance (averaged over the length of the experiment) between a 30-year fixed-SST control experiment based on year 1850 conditions (*piClim-control*) and a perturbation experiment that is identical except the relevant forcing agent changed to year 2014 levels, following RFMIP protocols (Forster et al., 2016; Pincus et al., 2016). The aerosol perturbation includes changing anthropogenic aerosol emissions of sulfur dioxide (SO_2), black carbon (BC), and organic carbon (OC) from year 1850 to 2014. Also included is the 4xCO_2 forcing as well as the radiative components (LW/SW clear-sky/cloud).

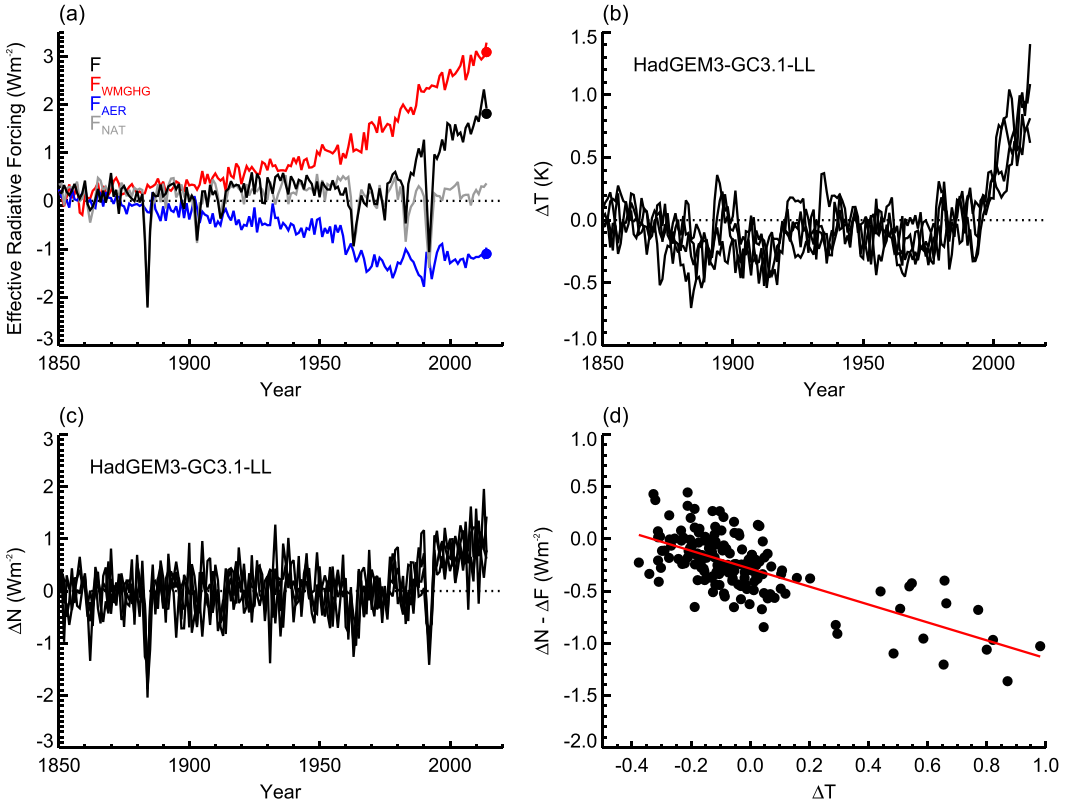

Figure 7. (a) Time series of the global annual mean historical effective radiative forcing, *F*, in HadGEM3-GC3.1-LL, including its separation into WMGHGs (red), aerosols (blue), and natural forcings (grey). Circles at 2014 represent the 1850–2014 timeslice ERF experiments (Table 3). Note that black circle represents the total anthropogenic forcing, whereas the black line (time series) includes anthropogenic and natural forcing (RFMIP protocol does not include a present-day natural ERF timeslice experiment). Global annual mean (b) surface air temperature change, ΔT, and (c) net TOA radiation change, ΔN, in the HadGEM3-GC3.1-LL four-member *historical* simulation. (d) The relationship between historical $\Delta N - \Delta F$ and ΔT for the ensemble mean.

the increasing forcing from WMGHGs increasingly dominates and historical ΔT in the coupled atmosphere-ocean simulation follows accordingly (Figure 7b). The large negative spikes in total and natural forcing are consistent with volcanic eruptions. The total and natural forcing time series does not go to zero at 1850, since *piControl* conditions are spun-up with the historical time-mean volcanic forcing to avoid long-term ocean drifts in response to historical volcanic forcing (Eyring et al., 2016). In practice, this means a lack of volcanic activity in the historical period gives a small but positive (~0.2 W m^{-2}) volcanic forcing, compared to the historical time-mean volcanic forcing.

7.2. Historical Feedbacks and EffCS

With the model's historical forcing time series known (Figure 7a and section 7.1), we can use it in conjunction with the change in net TOA radiation (Figure 7c)—and its LW/SW clear-sky/CRE components—and ΔT (Figure 7b) from coupled atmosphere-ocean *historical* simulations to estimate the model's climate feedbacks during its *historical* simulation, via $\lambda = (\Delta N - \Delta F)\Delta T$, fitted via ordinary least square regression of $\Delta N - \Delta F$ against ΔT (Figure 7d). Across the four-member *historical* ensemble, $\lambda_{NET} = -0.86 \pm 0.4$ W m^{-2} K^{-1} (5–95%), which is larger (smaller EffCS) in the ensemble mean response than that found in the *abrupt-4xCO₂* simulation but not distinct when considering the variability across the ensemble (Table 4). However, λ_{LWcs} is more negative, and λ_{CRE} less positive (both tending to reduce EffCS in the *historical* simulation) compared to *abrupt-4xCO₂* (Table 4). This is consistent with recent studies (e.g.,

Table 4
HadGEM3-GC3.1-LL Climate Feedback Parameters (W m^{-2} K^{-1}) in the abrupt-4xCO₂, amip-piForcing, and historical Simulations

Experiment	λ_{NET}	λ_{LWcs}	λ_{SWcs}	λ_{CRE}
abrupt-4xCO₂	−0.63	−1.80	0.66	0.51
amip-piForcing	−1.32	−2.27	0.74	0.20
Historical mean ± (5–95%)	−0.86 ± 0.40	−1.94 ± 0.06	0.88 ± 0.19	0.19 ± 0.15

Note. The 5–95% uncertainty ranges in the historical feedbacks represent ±1.645 standard deviations across the four-member ensemble results.

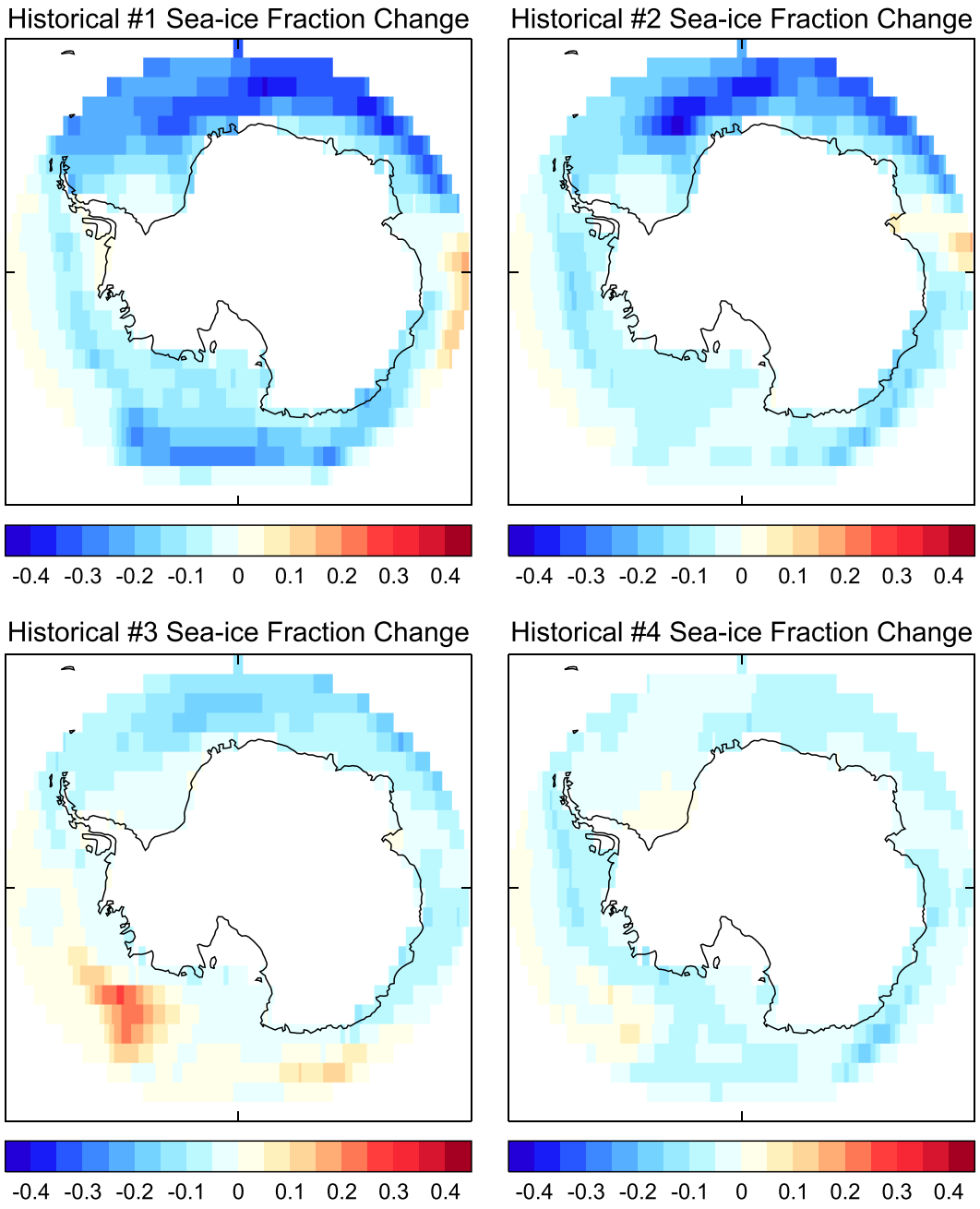

Figure 8. Historical (2000–2014 minus 1850) Southern Ocean sea-ice fraction change in the four-member ensemble historical HadGEM3-GC3.1-LL simulations.

Andrews et al., 2018; Ceppi & Gregory, 2017; Gregory & Andrews, 2016; Zhou et al., 2016) that have argued that historical LW clear-sky and cloud feedbacks may be more stabilizing (smaller EffCS) during historical climate change than long-term EffCS experiments due to pattern effects. While accurately estimating historical total feedback and EffCS in a single model is difficult due to large internal variability in the simulated historical high-latitude feedbacks (Adams & Dessler, 2019; Dessler et al., 2018), there is still of course value in evaluating the relationship between historical and long-term EffCS across a multimodel ensemble, where the statistics will improve, such as CMIP6 when available.

Indeed, we find that determining the total pattern effect between the *historical* and *abrupt-4xCO$_2$* simulations (i.e., the difference in feedbacks and EffCS) is primarily hindered by significant variability in

the models historical sea-ice feedback (since $\lambda_{SWcs} = 0.88 \pm 0.19$ W m^{-2} K^{-1} [5–95%]) across the ensemble (Table 4). This is due to significant decadal variability in the historical sea-ice trends in this model, particularly in the Antarctic. For example, Figure 8 shows the sea-ice fraction change in each ensemble member averaged over the final 15 years (2000 to 2014) of the *historical* simulation relative to the *piControl*. Two ensemble members show significant sea-ice loss, while two others show relatively little change. This is consistent with Dessler et al. (2018) and Adams and Dessler (2019) who found large uncertainties in historical feedbacks from high latitudes in the Max Planck Institute Earth System Model (MPI-ESM 1.1) 100-member *historical* ensembles. Large decadal variability in historical sea-ice trends is a common feature of AOGCMs (e.g., Brown et al., 2016; Kay et al., 2011; Rosenblum & Eisenman, 2017) and thus hinders our ability to accurately diagnose the total feedback and EffCS to forced climate change in *historical* simulations within a single model.

Finally, we calculate the feedbacks in HadGEM3-GC3.1-LL to real-world historical SST and sea-ice variations following Andrews et al. (2018) using the *amip-piForcing* simulation in CFMIP (Webb et al., 2017). *amip-piForcing* uses the atmospheric component of the model and forces it with observed monthly varying SST and sea ice from 1870 to 2014 using the AMIP II data set (Hurrell et al., 2008), while keeping all forcing agents (GHGs, aerosols etc.) constant at *piControl* conditions. Thus, λ_{NET} (and its components, calculated analogously) can be diagnosed simply from the linear fit of ΔN and ΔT (Andrews et al., 2018). This is useful quantity because Andrews et al. (2018) showed that AGCMs forced with real-world historical SST and sea-ice variations simulate more stabilizing cloud and lapse-rate feedbacks than those found in the same models forced by long-term CO_2 changes, with substantially lower EffCS in *amip-piForcing* than *abrupt-4xCO$_2$*. This "pattern effect" between feedbacks in response to observed historical SST patterns and those simulated under long-term CO_2 changes suggests observed EffCS constraints based on the historical record maybe biased low compared to long-term climate change (Andrews et al., 2018). With a move to higher EffCS in *abrupt-4xCO$_2$*, it would be informative to calculate this pattern effect in HadGEM3-GC3.1-LL, to test whether the historical pattern effect is larger in a higher EffCS model.

In common with other models, HadGEM3-GC3.1-LL simulates strongly stabilizing feedbacks in *amip-piForcing* compared to *abrupt-4xCO$_2$*, mostly due to differences in λ_{LWcs} and λ_{CRE} (Table 4). This is also the case in the model's coupled atmosphere-ocean historical simulation (Table 4). Using $F_{2x} = 4.05$ W m^{-2} (Table 2) and $\lambda = 1.32$ W m^{-2} K^{-1} (Table 4), the historical EffCS in *amip-piForcing* is found to be ~3.1 K, in contrast to the 5.5 K estimated from *abrupt-4xCO$_2$*. However, the change in feedback (equal to λ_{NET} in *abrupt-4xCO$_2$* minus λ_{NET} in *amip-piForcing*) ~0.7 W m^{-2} K^{-1} is close to the multimodel mean reported by Andrews et al. (2018; they gave a mean and uncertainty of 0.6 ± 0.4 [5–95%] W m^{-2} K^{-1} across a six-model ensemble). Thus, there is little indication, from this single high EffCS model at least, that the range of historical pattern effect determined by Andrews et al. (2018) needs significantly adjusting for higher-sensitivity models.

8. Summary and Discussion

A new generation of U.K. climate models has been developed and is being widely used in CMIP6. Here, we have evaluated benchmark forcing, feedback, and climate sensitivity metrics from these models. The EffCS to a doubling of CO_2 is found to be 5.5 K for HadGEM3.1-GC3.1-LL and 5.4 K for UKESM1 using the benchmark method of Andrews, Gregory, et al. (2012) that was adopted in IPCC AR5 (Flato et al., 2013). The TCR is 2.5 and 2.8 K, respectively. While the EffCS is larger than that seen in the previous generation of models, a move to higher sensitivity is in common with some other recently published modelling centers (Gettelman et al., 2019; Golaz et al., 2019; Voldoire et al., 2019).

The reasons for the increased sensitivity across the modelling ensemble will become clearer once CMIP6 data are widely available for analysis. For HadGEM3-GC3.1-LL and UKESM1, none of the individual forcing or feedback processes are found to be atypical of that found in the previous generation of models (CMIP5), though the net cloud feedback is towards the higher end and there exists a large high cloud response (which tends to have opposing LW and SW effects and so cancels in the net, as in HadGEM3-GC2.0, see Senior et al., 2016). The relatively large EffCS results from an unusual (relative to CMIP5) combination of a typical effective CO_2 forcing (close to the CMIP5 multimodel mean) with a relatively small feedback parameter (but within the CMIP5 5–95% range). This also appears to be the case for the E3SMv1 model (Golaz et al.,

2019; their figure 28). In CMIP5, an anticorrelation between CO_2 forcing and feedback existed that tended to minimize the impact of small feedback parameters on EffCS (i.e., models with small feedback parameters [higher EffCS]—as in the case here—tended to have small effective CO_2 forcings [lower EffCS]; Andrews, Gregory, et al., 2012; Ringer et al., 2014; Chung & Soden, 2017). Given that HadGEM3-GC3.1-LL, UKESM1, and E3SMv1 tend to go against this relationship, it would be useful to reinvestigate this anticorrelation between CO_2 forcing and feedback across the CMIP6 ensemble, when data are available, as a potential reason for increased EffCS relative to CMIP5.

Compared to the previous U.K. physical climate model, HadGEM3-GC2.0, the EffCS has increased from 3.2 to 5.5 K due to an increase in (i) CO_2 radiative forcing, (ii) surface albedo radiative feedback, and (iii) midlatitude cloud feedbacks. All changes are well understood—especially in the atmosphere (see Bodas-Salcedo et al., 2019)—and due to physical improvements in the model. For example, an improved treatment of greenhouse gas absorption (Pincus et al., 2015; Walters et al., 2019) has led to an increase in CO_2 radiative forcing; a reduction in the southern ocean SST bias has increased Antarctic sea-ice extent (Williams et al., 2017) and radiative feedback; and the inclusion of a mixed-phase cloud scheme and improved aerosol-cloud interaction processes (Mulcahy et al., 2018) has increased midlatitude cloud feedback (Bodas-Salcedo et al., 2019).

At higher atmospheric and ocean resolution (HadGEM3-GC3.1-MM; 60 km atmosphere and 0.25° ocean), the EffCS is largely unchanged compared to the lower-resolution configuration (HadGEM3-GC3.1-LL; 135 km atmosphere and 1° ocean), but there exists a compensation between an increased marine stratocumulus cloud feedback and a reduced Antarctic sea-ice feedback at the higher-resolution configuration (-MM). The increased cloud feedback arises from a better representation of coastal upwelling with higher resolution, reducing a warm SST bias that existed at the lower-resolution configuration (-LL) and so increasing climatological cloudiness and cloud feedback. In contrast, the reduced Antarctic sea-ice feedback at higher resolution maybe due to a Southern Ocean warm bias that is worse at higher resolution, reducing the climatological sea-ice amounts and feedback.

In UKESM1, we identified a CO_2 fertilization effect that induces a land surface vegetation and albedo change, which enhances the effective sensitivity of the model: The enhanced CO_2 favors growth of (darker) trees at the expense of (brighter) grasses, thus reducing the surface albedo and increasing the amount of solar radiation absorbed (e.g., Betts, 2000; Bala et al., 2006; Matthews, 2007; O'Ishi et al., 2009). This effect is a particularly strong effect when co-located with regions of seasonal snow cover, such as an expansion of the boreal forest, since the contrast in the albedo of snow-covered land to forest is even greater (e.g., Heinze et al., 2019). While we have highlighted a small but positive feedback associated with the inclusion of Earth system processes, the total feedback and EffCS between UKESM1 and HadGEM3-GC3.1-LL are similar, indicating a compensating negative feedback process (or processes) in the Earth System Model that is not included in HadGEM3-GC3.1-LL that we are currently unable to identify.

Historical aerosol forcing in HadGEM3-GC3.1-LL is -1.1 W m^{-2}, close to the CMIP5 multimodel mean. In HadGEM3-GC3.1 *historical* simulations, cloud feedback is found to be less positive than in *abrupt-4xCO_2*, in agreement with atmosphere-only experiments forced with observed historical SST and sea-ice variations. However, variability in the coupled model's historical sea-ice trends hampers accurate evaluation of the model's total historical feedback and EffCS, as found in other models (e.g., Adams & Dessler, 2019; Dessler et al., 2018). This hinders calculation of the "pattern effect" between historical climate change and long-term CO_2 changes (e.g., Andrews et al., 2018) within a coupled atmosphere-ocean single-model framework.

We highlight one final important characteristic of high-sensitivity models. That is, the EffCS to a doubling of CO_2 is itself sensitive to small changes in feedbacks in the model, λ_{NET}, due to the inverse relationship between EffCS and λ_{NET} (Figure 1d; Roe & Baker, 2007). This is important when quantifying the relative importance of changes in feedback on EffCS. For example, imagine a model development that altered the cloud feedback and climate sensitivity of that model, perhaps from the inclusion of a better representation of mixed-phase clouds (e.g., Tan et al., 2016) or tunings to aerosol-cloud interactions (e.g., Gettelman et al., 2019). Now imagine this development altered the cloud feedback of that model by +0.1 W m^{-2} K^{-1}. For a typical CMIP5 model, with say $\lambda_{\mathrm{NET}} = -1.1$ W m^{-2} K^{-1}, this would alter the model's EffCS

by only a small amount, from 3.1 K originally to 3.4 K (assuming a model mean EffF_{2x} = 3.4 W m^{-2}, Table 2). In contrast, for higher-sensitivity models, like UKESM1 and HadGEM3-GC3.1, with say λ_{NET} = −0.6 W m^{-2} K^{-1}, the same small change in feedback process would lead to big swing in EffCS from 5.7 to 6.8 K. Hence, the impact of a model development on EffCS clearly depends on the baseline sensitivity of that model. We therefore recommend reporting changes in model sensitivity in λ rather than—or as well as—EffCS space.

Acknowledgments
This work was supported by the Met Office Hadley Centre Climate Programme funded by Department for Business, Energy and Industrial Strategy (BEIS) and Department for Environment, Food and Rural Affairs (Defra). T. K. was funded by the Natural Environment Research Council (NERC) national capability grant for the U.K. Earth System Modelling project, Grant NE/N017951/1, and by the EU Horizon 2020 Framework Programme (H2020) CRESCENDO project, Grant Agreement 641816. We thank Peter Cadwell for useful discussions on the impact of ocean/atmosphere resolutions on climate sensitivity. We are grateful to Thorsten Mauritsen and two anonymous reviewers for positive and constructive comments that improved the clarity of the manuscript. The simulation data used in this study will be made available in the CMIP6 distributed data archive operated by the Earth System Grid Federation (ESGF; https://esgf-node.llnl.gov/projects/cmip6/). In the meantime, simulation data are archived at the Met Office and are available for research purposes through the JASMIN platform (www.jasmin.ac.uk) maintained by the Centre for Environmental Data Analysis (CEDA); for details, please contact um_collaboration@metoffice.gov.uk referencing this paper.

References

Adams, B. K., & Dessler, A. E. (2019). Estimating transient climate response in a large-ensemble global climate model simulation. *Geophysical Research Letters*, *46*(1), 311–317. https://doi.org/10.1029/2018GL080714

Andrews, T., Gregory, J. M., Paynter, D., Silvers, L. G., Zhou, C., Mauritsen, T., et al. (2018). Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophysical Research Letters*, *45*(16), 8490–8499. https://doi.org/10.1029/2018GL078887

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, *28*(4), 1630–1648. https://doi.org/10.1175/JCLI-D-14-00545.1

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*, L09712. https://doi.org/10.1029/2012GL051607

Andrews, T., Ringer, M. A., Doutriaux-Boucher, M., & Webb, M. J. (2012). Sensitivity of an Earth system climate model to idealized radiative forcing. *Geophysical Research Letters*, *39*, L10702. https://doi.org/10.1029/2012GL051942

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

Bala, G., Caldeira, K., Mirin, A., Wickett, M., Delire, C., & Phillips, T. J. (2006). Biogeophysical effects of CO_2 fertilization on global climate. *Tellus B*, *58*, 620–627. https://doi.org/10.1111/j.1600-0889.2006.00210.x

Betts, R. A. (2000). Offset of the potential carbon sink from boreal forestation by decreases in surface albedo. *Nature*, *408*(6809), 187–190. https://doi.org/10.1038/35041545

Bloch-Johnson, J., Pierrehumbert, R. T., & Abbot, D. S. (2015). Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, *42*, 4973–4980. https://doi.org/10.1002/2015GL064240

Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019). Strong dependence of atmospheric feedbacks on mixed-phase microphysics and aerosol-cloud interactions in HadGEM3. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1735–1758. https://doi.org/10.1029/2019MS001688

Bright, R. M., Myhre, G., Astrup, R., Antón-Fernández, C., & Strømman, A. H. (2015). Radiative forcing bias of simulated surface albedo modifications linked to forest cover changes at northern latitudes. *Biogeosciences*, *12*, 2195–2205. https://doi.org/10.5194/bg-12-2195-2015

Brown, P. T., Li, W., Jiang, J. H., & Su, H. (2016). Spread in the magnitude of climate model interdecadal global temperature variability traced to disagreements over high-latitude oceans. *Geophysical Research Letters*, *43*, 12,543–12,549. https://doi.org/10.1002/2016GL071442

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(50), 13,126–13,131. https://doi.org/10.1073/pnas.1714308114

Chung, E.-S., & Soden, B. J. (2017). On the compensation between cloud feedback and cloud adjustment in climate models. *Climate Dynamics*, *50*(3-4), 1267–1276. https://doi.org/10.1007/s00382-017-3682-1

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., et al. (2013). Long-term climate change: Projections, commitments and irreversibility. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1029–1136). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781107415324.024

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., et al. (2011). Development and evaluation of an Earth-System model—HadGEM2. *Geoscientific Model Development*, *4*, 1051–1075. https://doi.org/10.5194/gmd-4-1051-2011

Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., et al. (2014). North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modelling*, *73*, 76–107. https://doi.org/10.1016/j.ocemod.2013.10.005

Dessler, A. E., Mauritsen, T., & Stevens, B. (2018). The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. *Atmospheric Chemistry and Physics*, *18*(7), 5147–5155. https://doi.org/10.5194/acp-18-5147-2018

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 741–882). Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9781107415324.020

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research*, *118*, 1139–1150. https://doi.org/10.1002/jgrd.50174

Forster, P. M., Richardson, T., Maycock, A. C., Smith, C. J., Samset, B. H., Myhre, G., et al. (2016). Recommendations for diagnosing effective radiative forcing from climate models for CMIP6. *Journal of Geophysical Research: Atmospheres*, *121*, 460–12,475. https://doi.org/10.1002/2016JD025320

Gent, P. R., Yeager, S. G., Neale, R. B., Levis, S., & Bailey, D. A. (2010). Improvements in a half degree atmosphere/land version of the CCSM. *Climate Dynamics*, *34*, 819–833. https://doi.org/10.1007/s00382-009-0614-8

Geoffroy, O., Saint-Martin, D., Bellon, G., Voldoire, A., Olivié, D. J., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *Journal of Climate*, *26*, 1859–1876. https://doi.org/10.1175/JCLI-D-12-00196.1

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., et al. (2019). High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters*, *46*, 8329–8337. https://doi.org/10.1029/2019GL083978

Golaz, J., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *11*, 2089–2129. https://doi.org/10.1029/2018MS001603

Gregory, J. M., & Andrews, T. (2016). Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, *43*, 3911–3920. https://doi.org/10.1002/2016GL068406

Gregory, J. M., Andrews, T., & Good, P. (2015). The inconstancy of the transient climate response parameter under increasing CO_2. *Philosophical Transactions of the Royal Society A*, *373*, 20140417. https://doi.org/10.1098/rsta.2014.0417

Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, *113*, D23105. https://doi.org/10.1029/2008JD010405

Gregory, J. M., Jones, C. D., Cadule, P., & Friedlingstein, P. (2009). Quantifying Carbon Cycle Feedbacks. *Journal of Climate*, *22*, 5232–5250. https://doi.org/10.1175/2009JCLI2949.1

Grose, M. R., Gregory, J. M., Colman, R., & Andrews, T. (2018). What climate sensitivity index is most useful for projections? *Geophysical Research Letters*, *45*, 1559–1566. https://doi.org/10.1002/2017GL075742

Hardiman, S. C., Andrews, M. B., Andrews, T., Bushell, A. C., Dunstone, N. J., Dyson, H., et al. (2019). The impact of prescribed ozone in climate projections run with HadGEM3-GC3.1. *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/2019MS001714

Heinze, C., Eyring, V., Friedlingstein, P., Jones, C., Balkanski, Y., Collins, W., et al. (2019). Climate feedbacks in the Earth system and prospects for their evaluation. *Earth System Dynamics*, *10*, 379–452. https://doi.org/10.5194/esd-10-379-2019

Hurrell, J., Hack, J., Shea, D., Caron, J., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *Journal of Climate*, *21*(19), 5145–5153. https://doi.org/10.1175/2008JCLI2292.1

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., et al. (2016). C4MIP—The Coupled Climate–Carbon Cycle Model Intercomparison Project: Experimental protocol for CMIP6. *Geoscientific Model Development*, *9*, 2853–2880. https://doi.org/10.5194/gmd-9-2853-2016

Kay, J. E., Holland, M. M., & Jahn, A. (2011). Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophysical Research Letters*, *38*, L15708. https://doi.org/10.1029/2011GL048008

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, *34*, L22710. https://doi.org/10.1029/2007GL031383

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al. (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2865–2888. https://doi.org/10.1029/2018MS001370

Matthews, H. D. (2007). Implications of CO_2 fertilization for future climate change in a coupled climate–carbon model. *Global Change Biology*, *13*, 1068–1078. https://doi.org/10.1111/j.1365-2486.2007.01343.x

Mauritsen, T., & Stevens, B. (2015). Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geoscience*, *8*, 346–351. https://doi.org/10.1038/ngeo2414

Menary, M. B., Kuhlbrodt, T., Ridley, J., Andrews, M. B., Dimdore-Miles, O. B., Deshayes, J., et al. (2018). 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

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., et al. (2018). Improved aerosol processes and effective radiative forcing in HadGEM3 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *10*, 2786–2805. https://doi.org/10.1029/2018MS001464

O'Ishi, R., Abe-Ouchi, A., Prentice, I. C., & Sitch, S. (2009). Vegetation dynamics and plant CO_2 responses as positive feedbacks in a greenhouse world. *Geophysical Research Letters*, *36*, L11706. https://doi.org/10.1029/2009GL038217

Pincus, R., Forster, P. M., & Stevens, B. (2016). 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

Pincus, R., Mlawer, E. J., Oreopoulos, L., Ackerman, A. S., Baek, S., Brath, M., et al. (2015). Radiative flux and forcing parameterization error in aerosol-free clear skies. *Geophysical Research Letters*, *42*(13), 5485–5492. https://doi.org/10.1002/2015GL064291

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

Roe, G. H., & Baker, M. B. (2007). Why is climate sensitivity so unpredictable? *Science*, *318*(5850), 629–632. https://doi.org/10.1126/science.1144735

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., & Seferian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature*, *571*(7765), 335–342. https://doi.org/10.1038/s41586-019-1368-z

Rosenblum, E., & Eisenman, I. (2017). Sea ice trends in climate models only accurate in runs with biased global warming. *Journal of Climate*, *30*, 6265–6278. https://doi.org/10.1175/JCLI-D-16-0455.1

Rugenstein, M. A., Gregory, J. M., Schaller, N., Sedláček, J., & Knutti, R. (2016). Multiannual ocean–atmosphere adjustments to radiative forcing. *Journal of Climate*, *29*, 5643–5659. https://doi.org/10.1175/JCLI-D-16-0312.1

Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013). Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response. *Journal of Geophysical Research, Oceans*, *118*, 1830–1844. https://doi.org/10.1002/jgrc.20135

Sellar, A., Jones, C. G., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*. https://doi.org/10.1029/2019MS001739

Senior, C. A., Andrews, T., Burton, C., Chadwick, R., Copsey, D., Graham, T., et al. (2016). Idealized climate change simulations with a high-resolution physical model: HadGEM3-GC2. *Journal of Advances in Modeling Earth Systems*, *8*, 813–830. https://doi.org/10.1002/2015MS000614

Stevens, B., Sherwood, S. C., Bony, S., & Webb, M. J. (2016). Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, *4*(11), 512–522. https://doi.org/10.1002/2016EF000376

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., et al. (2018). UK Global Ocean GO6 and GO7: A traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*, 3187–3213. https://doi.org/10.5194/gmd-11-3187-2018

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, *352*(6282), 224–227. https://doi.org/10.1126/science.aad5300

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, *11*, 2177–2213. https://doi.org/10.1029/2019MS001683

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., et al. (2019). 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

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). 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

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2017). 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

Williams, K. D., Harris, C. M., Bodas-Salcedo, A., Camp, J., Comer, R. E., Copsey, D., et al. (2015). The Met Office Global Coupled model 2.0 (GC2) configuration. *Geoscientific Model Development*, *8*, 1509–1524. https://doi.org/10.5194/gmd-8-1509-2015

Zelinka, M. D., Andrews, T., Forster, P. M., & Taylor, K. E. (2014). Quantifying components of aerosol-cloud-radiation interactions in climate models. *Journal of Geophysical Research*, *119*, 7599–7615. https://doi.org/10.1002/2014JD021710

Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, *9*, 871–874. https://doi.org/10.1038/ngeo2828