Implementation of U.K. Earth System Models for CMIP6

Alistair A. Sellar[1], Jeremy Walton[1], Colin G. Jones[2,3], Richard Wood[1], Nathan Luke Abraham[2,4], Miroslaw Andrejczuk[1], Martin B. Andrews[1], Timothy Andrews[1], Alex T. Archibald[2,4], Lee de Mora[5], Harold Dyson[1], Mark Elkington[1], Richard Ellis[6], Piotr Florek[1], Peter Good[1], Laila Gohar[1], Stephen Haddad[1], Steven C. Hardiman[1], Emma Hogan[1], Alan Iwi[7], Christopher D. Jones[1], Ben Johnson[1], Douglas I. Kelley[6], Jamie Kettleborough[1], Jeff R. Knight[1], Marcus O. Köhler[4,8,9], Till Kuhlbrodt[1,2,10], Spencer Liddicoat[1], Irina Linova-Pavlova[1], Matthew S. Mizielinski[1], Olaf Morgenstern[1,11], Jane Mulcahy[1], Erica Neininger[1], Fiona M. O'Connor[1], Ruth Petrie[7], Jeff Ridley[1], Jean-Christophe Rioual[1], Malcolm Roberts[1], Eddy Robertson[1], Steven Rumbold[1,2,10], Jon Seddon[1], Harry Shepherd[1], Sungbo Shim[1,2,12], Ag Stephens[7], Joao C. Teixiera[1], Yongming Tang[1], Jonny Williams[1,11], Andy Wiltshire[1], and Paul T. Griffiths[2,4]

[1]Met Office Hadley Centre, Exeter, UK, [2]National Centre for Atmospheric Science, UK, [3]School of Earth and Environment, University of Leeds, Leeds, UK, [4]Department of Chemistry, University of Cambridge, Cambridge, UK, [5]Plymouth Marine Laboratory, Plymouth, UK, [6]UK Centre for Ecology and Hydrology, Wallingford, UK, [7]Science and Technology Facilities Council, UK Research and Innovation, Didcot, UK, [8]Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, UK, [9]Now at European Centre for Medium-Range Weather Forecasts, Reading, UK, [10]Department of Meteorology, University of Reading, Reading, UK, [11]National Institute of Water and Atmospheric Research, Wellington, New Zealand, [12]National Institute of Meteorological Sciences, Seogwipo-si, South Korea

Abstract We describe the scientific and technical implementation of two models for a core set of experiments contributing to the sixth phase of the Coupled Model Intercomparison Project (CMIP6). The models used are the physical atmosphere-land-ocean-sea ice model HadGEM3-GC3.1 and the Earth system model UKESM1 which adds a carbon-nitrogen cycle and atmospheric chemistry to HadGEM3-GC3.1. The model results are constrained by the external boundary conditions (forcing data) and initial conditions. We outline the scientific rationale and assumptions made in specifying these. Notable details of the implementation include an ozone redistribution scheme for prescribed ozone simulations (HadGEM3-GC3.1) to avoid inconsistencies with the model's thermal tropopause, and land use change in dynamic vegetation simulations (UKESM1) whose influence will be subject to potential biases in the simulation of background natural vegetation. We discuss the implications of these decisions for interpretation of the simulation results. These simulations are expensive in terms of human and CPU resources and will underpin many further experiments; we describe some of the technical steps taken to ensure their scientific robustness and reproducibility.

Plain Language Summary Complex models of the Earth system are valuable tools for understanding the processes responsible for our changing climate. The Coupled Model Intercomparison Project (CMIP) is a well-established activity of the World Climate Research Programme that brings together results from these models to better understand their process representation and to pool their projections for robust understanding of future climate pathways. The latest phase of CMIP (CMIP6) is larger and more ambitious than previous phases. We detail the setup of two U.K. models (HadGEM3-GC3.1 and UKESM1) for a core set of experiments contributing to CMIP6, including simulations of historical and future periods covering 1850 to 2300. We highlight assumptions made in applying the prescribed CMIP6 input data to these models. We outline the technical steps to ensure the reproducibility of these simulations.

1. Introduction

Complex models of the Earth system are valuable tools for understanding the processes responsible for our changing climate. The Coupled Model Intercomparison Project (CMIP) is a well-established activity

of the World Climate Research Programme that brings together results from these models to better understand their process representation and to pool their projections for robust understanding of future climate pathways. CMIP facilitates fundamental climate and carbon cycle research (e.g., Friedlingstein et al., 2014; Sherwood et al., 2014) and supports assessments of climate science literature (e.g., IPCC, 2014). The United Kingdom has a long history of contributing to past phases of CMIP (e.g., Gordon et al., 2000; Johns et al., 1997, 2006; Jones et al., 2011) and has made a strong commitment to the current, sixth, phase (CMIP6; Eyring et al., 2016). The U.K. contribution to CMIP6 is a collaborative endeavor, with model development and simulation shared between the Met Office Hadley Centre and a number of research centers under the auspices of the Natural Environment Research Council (NERC) and the Science and Technology Facilities Council. Indeed, contributions with U.K. models are not limited to the United Kingdom: Key simulations are being performed by the Korean Meteorological Administration and New Zealand's National Institute of Water and Atmospheric Research.

CMIP6 specifies nearly 300 experiments, organized by about 20 CMIP6-endorsed Model Intercomparison Projects (MIPs) around a central set of experiments (Eyring et al., 2016). The protocols for these experiments dictate how the models should be configured and what forcing data should be used as input. Nevertheless, models differ in their capabilities and assumptions, and some aspects of the protocols need to be interpreted in light of these assumptions. Jones et al. (2011) detailed the implementation of the previous generation UK Earth system model HadGEM2-ES (Collins et al., 2011) for CMIP5 experiments. In this paper we do the same for the latest generation of U.K. models, HadGEM3-GC3.1 (Kuhlbrodt et al., 2018; Williams et al., 2018), and UKESM1 (Sellar et al., 2019), as applied to a core set of experiments in CMIP6, specifically the DECK and historical experiments (Eyring et al., 2016) and a set of projections for ScenarioMIP (O'Neill et al., 2016). We focus on these experiments as they have high policy relevance and underpin the majority of experiments in the other MIPs.

The purpose of this paper is threefold: to aid reproducibility by documenting the setup of U.K. models for core CMIP6 experiments; to highlight where scientific decisions have been taken in defining this setup, particularly where these may impact the interpretation of the results; and to outline some of the technical methodology used to ensure robustness, traceability, and reproducibility of the model experiments. We do not document the models themselves since they are presented in existing literature nor do we analyze the results of these experiments as that will be the focus of future work. The paper is structured as follows: section 2 introduces the model configurations used in the core simulations, section 3 summarizes the scientific choices and assumptions made in implementing the CMIP6 forcing for these models, section 4 explains the strategy used in initializing historical simulations to maximally span the models' modes of internal variability, and section 5 outlines the technical steps we took to ensure the reproducibility of these simulations. Finally, section 6 discusses the implications of some of the scientific choices for the interpretation of the results of these simulations and briefly summarizes the amount of resource used and the size of data generated during our participation in CMIP6.

2. Model Configurations Used for CMIP6

The U.K. contribution to CMIP6 is based on two models: the physical climate model HadGEM3-GC3.1 (Kuhlbrodt et al., 2018; Williams et al., 2018) and the Earth system model UKESM1 (Sellar et al., 2019). The two models are closely related, with UKESM1 consisting of GC3.1 as its physical core, plus the addition of component models for atmospheric chemistry and for marine and terrestrial biogeochemistry. Two horizontal resolutions of GC3.1 have been used to make DECK, historical, and ScenarioMIP simulations: the first employs an N96 atmospheric grid (192×144 grid points) and a tripolar ocean grid with a resolution of nominally 1° (eORCA1, 360×330 grid points; Madec & Imbard, 1996), while the second has an N216 atmospheric grid (432×325 grid points) and a 0.25° tripolar ocean grid (eORCA025, $1,440 \times 1,205$ grid points). The full names of these configurations are HadGEM3-GC3.1-N96ORCA1 and HadGEM3-GC3.1-N216ORCA025, respectively. Hereafter, HadGEM3-GC3.1 is referred to as GC3.1. For UKESM1, only the N96ORCA1 resolution has been used for CMIP6 simulations. For all these model configurations the atmosphere model uses 85 vertical levels from the surface to 85 km, with a hybrid sigma-height coordinate that follows terrain near the surface and evolves toward geopotentials at 18 km and above (Walters et al., 2019); the ocean uses 75 vertical levels with a z^* coordinate in which the reference levels are geopotentials but for which cell thicknesses vary in time as the nonlinear free surface evolves (Storkey et al., 201).

Table 1
CMIP6 MIP Submissions Using U.K. Model Configurations

Model	HadGEM3-GC3.1		UKESM1.0
Resolution	N96ORCA1	N216ORCA025	N96ORCA1
CMIP6 `source_id`	HadGEM3-GC31-LL	HadGEM3-GC31-MM	UKESM1-0-LL
DECK & historical	x	x	x
AerChemMIP			x
C4MIP			x
CDRMIP			x
CFMIP	x		
DAMIP	x		
DCPP		x	
PAMIP		x	
FAFMIP	x		
GeoMIP			x
GMMIP		x	
HighResMIP[a]	x	x	
ISMIP6			x
LS3MIP	x		
LUMIP			x
OMIP		x	x
PMIP	x		x
RFMIP	x		x[b]
ScenarioMIP	x[c]	x[c]	x
VolMIP			x
ZECMIP			x
Diagnostic-only MIPs			
CORDEX			x
DynVarMIP	x	x	x
SIMIP	x	x	x
VIACS AB	x	x	x

Note. See text and Table 2 for an explanation of the trailing letters in `source_id`.
[a]HighResMIP uses a matrix of ocean and atmosphere resolutions of GC3.1, which are tied to the DECK submissions of these two configurations. [b]The Tier 1 fixed-SST experiments (referred to as RFMIP-lite by Pincus et al., 2016) have been performed with UKESM1.0. [c]A subset of Tiers 1 and 2 scenarios will be performed with GC3.1.

Within CMIP6, each model configuration is uniquely identified by a short string, `source_id`, whose value for each configuration is shown in Table 1. The pair of characters at the end of each `source_id` indicates the resolution (**L**ow, **M**edium, or **H**igh) of first the atmosphere and second the ocean. The correspondence between these characters and the resolution of the atmosphere and ocean is documented in Table 2. These three configurations will also be used for a wide range of CMIP6-endorsed MIPs. The choice of models for each MIP is documented in Table 1.

3. Forcing Implementation

External forcing data sets are provided by various expert groups under the coordination of CMIP. These provide many of the inputs required by our models but in many cases need to be augmented with additional input data. For example, the CMIP6-mandated aerosol and gas emissions include anthropogenic and biomass burning sources but not natural sources. This section documents scientific choices and assumptions made in implementing the CMIP6 forcing data, as well as those data sets used to augment them. For many of these data sets the implementation differs between GC3.1 and UKESM1 because their differing levels of complexity imply different requirements for driving data. In general, UKESM1 requires fewer input data

Table 2
Resolutions Corresponding to the Final Pair of Characters in CMIP6 `source_id`

Resolution indicator	Atmosphere resolution	Ocean resolution
L	N96	eORCA1
M	N216	eORCA025
H[a]	N512	eORCA12

[a]H is used only in configurations employed by HighResMIP. N512 has $1,024 \times 768$ grid points. ORCA12 is a nominally $(1/12)°$ tripolar grid with $4,320 \times 3,604$ points.

sets because it interactively simulates more components of the Earth system, such as atmospheric chemistry and vegetation cover.

Within this section we cite the scientific papers which document the methodology and evaluation of these data sets, where available. In Appendix A we document the specific versions of the data sets employed, with reference to the data set citations (including digital object identifiers) which uniquely identify them.

Where specific processing has been performed for the historical or future period, we describe this below. For the period 2100–2300 we follow the ScenarioMIP protocol (O'Neill et al., 2016) in using transient forcing data for solar irradiance, greenhouse gas concentrations, and land use while keeping all other forcing agents fixed at 2100 levels.

3.1. Solar Variability

The solar radiation forcing used by GC3.1 and UKESM1 is derived from the recommended solar data sets for CMIP6 (Matthes et al., 2017). Two CMIP6 data sets are used: one for preindustrial control simulations and the other for historical-future scenario simulations. The preindustrial control data set, consisting of total solar irradiance (TSI) and solar spectral irradiance (SSI) data, is constructed of time-averaged historical data corresponding to 1850–1873 mean conditions. The historical-future data set consists of TSI and SSI data at monthly resolution, using historical reconstructions for the period 1850–2014 and a future projection for the period 2015–2299 (Matthes et al., 2017). The solar forcing data provided by Matthes et al. (2017) include information on particle forcing, but the CMIP6 protocol provides no recommendation on whether it should be used; for consistency with previous rounds of CMIP we do not use solar particle forcing.

The CMIP6 SSI data span the wavelength interval 10–100,000 nm. The interval accepted by GC3.1 and UKESM1 is 200–10,000 nm across 12 bands. Given that the TSI is consistent with the SSI in the CMIP6 data set, it is important to include the radiation flux below 200 nm and above 10,000 nm in the model. The spectral irradiance below and above the model frequency range is added to the lowest and highest model bands, respectively, that is, the 10- to 220-nm irradiance is included in the 200- to 220-nm model band, and the 2,380- to 100,000-nm irradiance is included in the 2,380- to 10,000-nm model band. This approach ensures that the TSI and SSI applied to the simulations are consistent with one another.

UKESM1 includes photolytic reactions in its interactive simulation of ozone, but the photolysis rates use fixed solar inputs and therefore have no dependence on the solar variability.

3.2. Stratospheric Volcanic Aerosol
3.2.1. Radiative Properties

Explosive volcanic eruptions inject SO_2 into the stratosphere leading to the formation of sulfate aerosols that scatter solar radiation back to space. This leads to a negative radiative forcing (Myhre et al., 2013), although this can be offset to some degree by the absorption of outgoing longwave radiation and other rapid adjustments (Schmidt et al., 2018). Stratospheric aerosol loading varies considerably with time, depending on the magnitude, altitude, and location of eruptions. Accurate simulations of past climate therefore rely on a reasonable representation of stratospheric aerosol radiative properties (Shindell et al., 2003).

For CMIP6, stratospheric aerosol properties have been made available via a zonal mean data set specifying aerosol extinction, absorption, and scattering asymmetry as a function of altitude (5 to 40 km in 0.5-km steps), latitude (5° resolution), and time (monthly resolution through the historical period; 1850–2014) (Arfeuille et al., 2014; Thomason et al., 2018). This climatology was supplied by the Swiss Federal Institute of Technology (ETH) and conveniently provided as averages over the spectral bands of the Met Office Unified Model's radiation scheme.
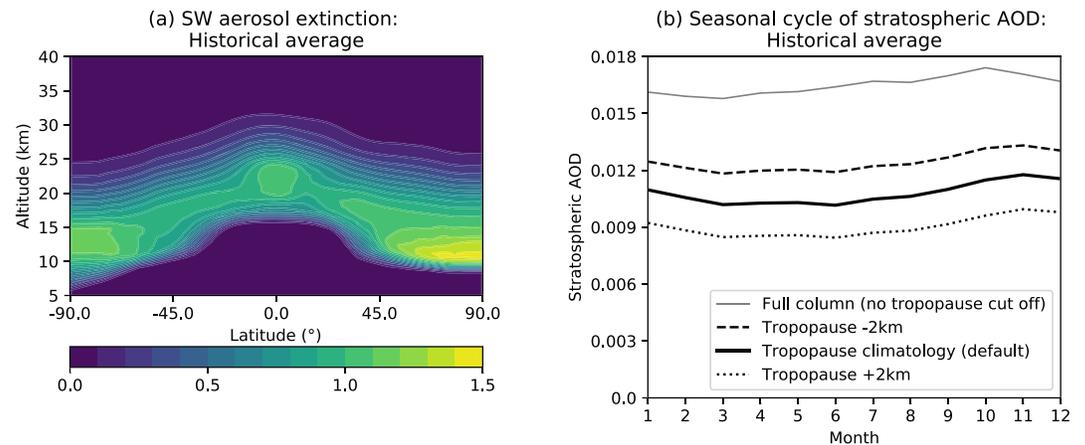
Figure 1. (a) Shortwave aerosol extinction as a function of altitude and latitude based on the annual mean from the historical average (averaged climatology used in the preindustrial control simulation and for future scenarios from 2025 onward), (b) stratospheric aerosol optical depth for the averaged climatology as function of month and depending on the implementation of tropopause height in the offline processing. Both plots show averages across the wavelength interval 0.32–0.69 μm).

The aerosol properties in this climatology were set to zero below the modeled tropopause to screen out upper-tropospheric aerosol and prevent unintended impacts on tropospheric meteorology. This screening was applied offline using a monthly and zonal mean tropopause height climatology derived from a 10-year simulation using HadGEM3-GA7.0 (Walters et al., 2019) with year 2000 forcings. Tropopause heights were diagnosed by thermal stratification following the standard World Meteorological Organization definition (WMO, 1992). To provide a smooth transition with height, a 2-km buffer zone was included straddling the tropopause such that aerosol was zero 1 km below the tropopause and unmodified 1 km above the tropopause, with a linear ramp in between. Figure 1a illustrates the resulting smooth variation of aerosol shortwave extinction with height and latitude in the resulting data set.

The magnitude of the volcanic forcing was somewhat sensitive to the assumed tropopause height, as indicated in Figure 1b. Setting the tropopause 2 km higher decreased the stratospheric aerosol optical depth (AOD) by 16% and setting it 2 km lower increased AOD by 15%. The AOD shown here is the average AOD across the wavelength interval 0.32–0.69 μm and is approximately 6% higher than the 0.55 μm AOD.

In line with the CMIP6 experimental protocol (Eyring et al., 2016), for preindustrial control simulations, a historical mean climatology was used where aerosol properties for each month are the average from all corresponding months during the period (1850–2014). The historical mean climatology has a small and fairly smooth seasonal cycle, as shown in Figure 1b, in part due to the seasonal variation of tropopause height.

The time series of stratospheric 0.32- to 0.69-μm AOD for the historical and future periods 1850–2100 are shown in Figure 2. From the year 2025 onward the historical mean climatology is employed (identical to that used in preindustrial simulations). For the period from January 2015 to December 2024 aerosol values are assumed to gradually return to this climatology with a linear ramp to create a smooth transition between the historical and future periods. During the ramp period the aerosol values in a given month are a linear combination of the aerosol values in December 2014 and the historical mean values for the given month so that the seasonal cycle of the averaged climatology gradually returns during the ramp period.

3.2.2. Surface Area Density for Heterogeneous Chemistry

As well as having an important climatic effect through their impacts on the transmission of radiation, large volcanic eruptions are known to have significant impacts on atmospheric chemistry. The stratospheric aerosol resulting from large eruptions acts as a medium on which heterogeneous reactions occur. Stratospheric heterogeneous reactions are known to play an important role in depletion of ozone as they enable the activation of reservoirs of ozone-depleting substances (Schmidt et al., 2018).

In UKESM1 this chemistry is simulated using the U.K. Chemistry and Aerosol model (UKCA; Archibald et al., 2019), following the approach outlined in Morgenstern et al. (2009), with prescribed aerosolf surface area
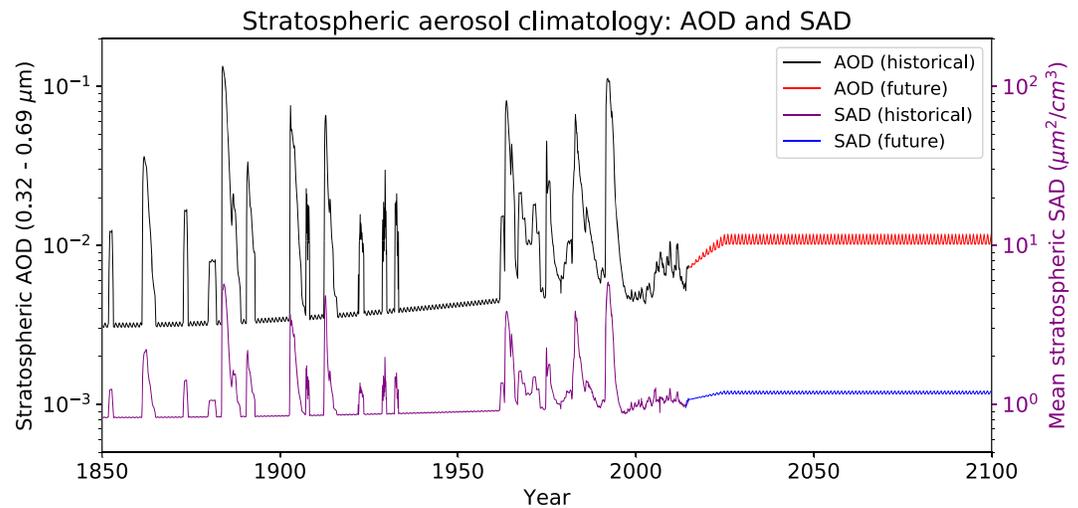
Figure 2. Time series of aerosol optical depth (AOD) and aerosol surface area density (SAD) for the stratospheric aerosol climatologies, including the historical (1850–2014) and future (2015–2100) periods.

density (SAD) from the same source as the radiative properties described above (Beiping, 2017). The temporal evolution of this SAD is shown in Figure 2. These inputs are provided with the same height, latitude, and time resolution as the radiative properties, and the same time processing is applied to the transition from the historical time series to the climatology used post-2025.

Three (out of 19) UKESM1 historical simulations inadvertently use periodic 1850 SAD data, instead of 1850–2014 time series, thus breaking with the CMIP6 experimental protocol. This will reduce the variability of stratospheric ozone in those ensemble members, most notably following volcanic eruptions when time series forcing results in a decline in ozone. Outside of the stratosphere, we do not expect this difference in forcing to have a significant impact on model behavior. Although the periodic 1850 SAD data were applied in error, we have published these experiments as part of our CMIP6 data because we feel that, first, it will allow users to analyze the impact of time-varying SAD on stratospheric chemistry by providing a control with periodic forcing, and second, for components of the model unaffected by the change, these simulations increase the size of the experimental ensemble. Table A8 shows how users of the published data can use the CMIP6 file metadata to differentiate these three historical experiments from the rest of the ensemble.

We have made an initial examination of the impact of this erroneous inclusion of 1850 SAD in three ensemble members by subtracting their zonal mean total column ozone from that of the other 13 members currently published on the Earth system grid. The temporal and meridional variability of this difference is shown in Figure 3; these differences reflect both forced variability in heterogeneous chemistry and unforced meteorological variability. From theory we would expect that in times of high chlorine loading (such as during Pinatubo) an injection of sulfur into the stratosphere should promote ozone depletion by enhancing production of active chlorine through heterogeneous reactions. By contrast under low chlorine conditions (during Krakatoa and Agung), these reactions are not so important, and instead, nitrogen deactivation is enhanced by the volcanic aerosol, which leads to ozone increases. Indeed, Figure 3 shows some sign of weak positive anomalies of ozone after the eruptions of Krakatoa and Agung, whereas after Pinatubo 1 or 2 years of quite low ozone ensued. However, these possible signals are small relative to the background variability and we have not assessed their significance. A systematic analysis of the impact of SAD variability in these simulations is planned for future work. The overriding conclusion of this initial look is that missing of SAD variability in these three ensemble members does not introduce an obvious bias or error to the modeled ozone concentrations.

In addition to this external input of SAD for sulfate aerosol, UKESM1 interactively calculates the aerosol surface area produced from the formation of nitric acid trihydrate and mixed ice/nitric acid trihydrate polar stratospheric clouds (Keeble et al., 2014).
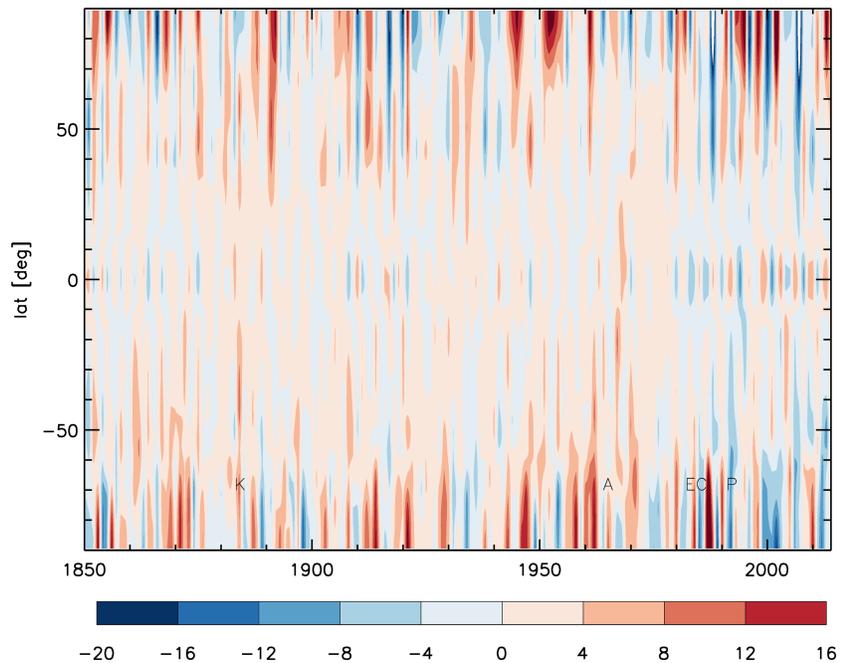
Figure 3. Annual mean zonal mean total column ozone (Dobson units) for the mean of 13 historical ensemble members with variable SAD minus the same for the mean of three ensemble members with 1850 SAD. Labels at 70S denote the four large volcanic eruptions occurring in the historical period: Krakatoa, 1883; Agung, 1963; El Chichon, 1982; Pinatubo 1991.

3.3. Well-Mixed Greenhouse Gases

We use the global mean annual mean concentrations of well-mixed greenhouse gases (WMGHG) provided by Meinshausen et al. (2017) for the historical period and Meinshausen et al. (2019) for future scenarios. For 1pctCO2 experiments (one of the DECK simulations; Eyring et al., 2016) the model calculates the annual increase of the CO_2 concentration directly, with CO_2 incremented by 1% at the beginning of each year.

3.3.1. GC3.1

The source data provide concentrations of 43 WMGHG species, many more than the models' radiation scheme can cater for. Hence, as in Jones et al. (2011), we use the equivalent concentrations HFC-134a-eq and CFC-12-eq that summarize the radiatively less important species. This means that we specify five WMGHG concentrations in total: CO_2, CH_4, N_2O, HFC-134a-eq, and CFC-12-eq. These concentrations are updated once per year and applied in a spatially uniform manner, with no horizontal or vertical variation.

3.3.2. UKESM1

In the standard configuration of UKESM1, the radiative treatment of CO_2, CFCs, and HFCs is identical to that described above for GC3.1. In contrast, the CH_4 and N_2O concentrations interacting with radiation in UKESM1 are represented by interactive three-dimensional tracers in the UKCA chemistry and only their surface concentrations are prescribed. Above the surface, these tracers are modified by chemical depletion and advection, resulting a vertical distribution which is more realistic than prescribing a single value at all heights. These CH_4 and N_2O surface concentrations are taken from the data of Meinshausen et al. (2017), as in GC3.1.

3.3.3. CO_2 Emissions for UKESM1

The default configuration of UKESM1 uses prescribed CO_2 concentrations as described above. However, the model can also run with interactive CO_2 concentrations under prescribed CO_2 emissions, as required for the CMIP6 experiments esm-piControl and esm-historical (Eyring et al., 2016). In this CO_2 emission-driven configuration, CO_2 is a three-dimensional tracer which is subject to atmospheric advection and surface exchange with the marine and terrestrial biosphere. For further details see Sellar et al. (2019).

We use the CO_2 emissions data of Hoesly et al. (2018). We combine all emission sources (including aircraft) and release the emission at the surface. After horizontal interpolation to the model grid and conversion to the 360-day calendar used by these models, emission data were scaled to ensure that annual global-total

Table 3
Model Emission Type for Anthropogenic Primary Carbonaceous Aerosol

Model emission	Source sector
Fossil fuels	Energy, industrial, shipping, transportation, solvents, waste
Biofuels	Agriculture, residential/other

emissions agree with the totals provided by the CEDS project (http://www.globalchange.umd.edu/ceds/ceds-cmip6-data/). This ensures that the model's cumulative global-total CO_2 emissions are exactly as provided by Hoesly et al. (2018). For a long-lived species such as CO_2 this conservation of cumulative emission is the key consideration, contrast this with the treatment of emissions of shorter-lived species below.

3.4. Emissions of Tropospheric Aerosols and Reactive Gases

Both UKESM1 and GC3.1 use emissions of primary carbonaceous aerosol (the model has separate tracers for black carbon and organic carbon) and gas phase sulfur dioxide (CO_2) and dimethyl suphide (($CH_3)_2S$, DMS), which act as a precursors to sulfate aerosol. Additionally, UKESM1 uses emissions of ethane (C_2H_6), propane (C_3H_8), methanol (CH_3OH), formaldehyde (CHO), acetone (($CH_3)_2CO$), acetaldehyde (CH_3CHO), carbon monoxide (CO), and nitric oxide (NO). The CMIP6 protocol provides for these emissions from anthropogenic and biomass burning sources, and we augment these with emissions of reactive gases from biogenic and other natural sources.

In contrast to the treatment of CO_2 emissions described above, there is no scaling of these emissions to compensate for the length of year in the models' 360-day calendar. As a result, the annual total of the emission received by the model is lower than the annual total of the source data by 1.4%, but daily and weekly totals match those of the source data. We make this choice because these species have lifetimes of order days to weeks, and therefore, the short-term emission rate is considered more important than the long-term accumulation.

3.4.1. Anthropogenic Sources

Anthropogenic emissions of carbonaceous aerosol and reactive gases are provided by the CEDS project (Community Emissions Data System; Hoesly et al., 2018). Carbonaceous aerosols are treated as arising from the burning of either fossil fuels or biofuels according to the source sector (Table 3). Aerosols from biofuel burning are emitted with larger sizes than those from fossil fuels: the geometric mean diameter of emitted aerosols is 150 and 60 nm, respectively.

UKESM1 and GC3.1 differ for SO_2 emission as a result of their differing levels of chemical complexity: Both models simulate the oxidation of SO_2 and terpenes in the production of secondary aerosol, but UKESM1 interactively simulates the chemistry of gases responsible for this oxidation (OH, O_3, NO_3, HO_2, and H_2O_2), while in GC3.1 these oxidants are prescribed from a monthly climatology representing present-day conditions (Walters et al., 2019). For GC3.1, SO_2 emission heights are dependent on sector, and emissions are split between the surface level emission and a "high-level" emission at 0.5 km, representing emissions thermally lofted from chimney level (Table 4). This follows the implementation used for the previous generation model (HadGEM2; Jones et al., 2011). In contrast, UKESM1 emits all anthropogenic SO_2 into the lowest model level for consistency with emissions of other chemically active gases.

UKESM1 uses emissions from aircraft for CO_2 (see section 3.3.3) and NO_x only. NO_x aircraft emissions are prescribed as three-dimensional fluxes using the altitude information provided in the source data. Following previous implementations of UKCA (e.g., Morgenstern et al., 2017;; O'Connor et al., 2014), aircraft emissions

Table 4
Split of SO_2 Emissions Between Surface and High Level for GC3.1

Emission sector	Low level %	High level %
Energy	0	100
Industrial	50	50
Shipping, transportation, solvents,		
Waste, agriculture, residential/other	100	0

Table 5
Annual Total Prescribed Natural Emissions Used by UKESM1

NO	Terrestrial	12.0 Tg
C_2H_6[a]	Terrestrial	31.08 Tg
	Oceanic	2.5 Tg
C_3H_8[b]	Terrestrial	15.61 Tg
	Oceanic	2.9 Tg
CO	Terrestrial	88.59 Tg
	Oceanic	20.0 Tg
HCHO	Terrestrial	4.58 Tg
$(CH_3)_2CO$	Terrestrial	37.43 Tg
CH_3CHO	Terrestrial	21.51 Tg
CH_3OH	Terrestrial	129.26 Tg
SO_2[c]	Volcanic	28.77 Tg

[a]Includes contributions due to C_2H_4 and C_2H_2. [b]Includes contributions due to C_3H_6. [c]Volcanic SO_2 is used also by GC3.1.

for other species are ignored because their contributions have negligible effect on atmospheric chemistry or aerosol loading.

3.4.2. Biomass Burning

Both UKESM1 and GC3.1 use primary emissions of black carbon and organic carbon from biomass burning provided by van Marle et al. (2017). Emissions from forest burning sectors are spread evenly over the lowest 3 km, while other sectors treated as a surface emission. Aerosols emitted from biomass burning sources are treated with a geometric mean diameter of 150 nm.

3.4.3. Natural Sources

GC3.1 and UKESM1 both interactively simulate emissions of sea salt, dust, and DMS as described in Walters et al. (2019). In GC3.1 the DMS emission parametrization takes as input the DMS seawater concentration data set of Lana et al. (2011), while in UKESM1 it uses interactively simulated seawater DMS (Sellar et al., 2019). Additionally, UKESM1 has interactive emissions of primary marine organic aerosol (PMOA) and biogenic emissions of the volatile organic compounds isoprene and monoterpene, as described in Sellar et al. (2019). In GC3.1, DMS emissions are scaled by a factor of 1.7 to account for missing PMOA emissions (Mulcahy et al., 2018); because UKESM1 includes PMOA emissions, the DMS scaling is reset to 1.0.

In addition to these interactive emissions, both models receive prescribed climatological natural emissions of various species. These have no secular or interannual variation. Common to both GC3.1 and UKESM1 are emissions of SO$_2$ from continuously degassing volcanoes. These are represented by the present-day three-dimensional climatology of Dentener et al. (2006), a temporally constant data set with no seasonal variation. UKESM1 also makes use of biogenic emissions of C_2H_6, C_3H_8, CH_3OH, HCHO, $(CH_3)_2CO$, CH_3CHO, CO, NO, and DMS, which are included through a climatological seasonal cycle. Emissions of ethene (C_2H_4) and ethyne (C_2H_2), species not represented by the model, were combined with those for C_2H_6, and similarly, propene (C_3H_6) emissions were added to C$_3$H$_8$. Land-based emissions were taken from Model of Emissions of Gases and Aerosols from Nature monthly emissions fluxes compiled for the Monitoring Atmospheric Composition and Climate project (MEGAN-MACC, Guenther et al., 2012; Sindelarova et al., 2014) averaged over the time period 2001–2010, except for NO where an annual flux of 12 Tg without seasonality was assumed (Yienger & Levy, 1995). Oceanic emissions were taken from the Precursors of Ozone and their Effects in the Troposphere project (POET, Granier et al., 2005; Olivier et al., 2003) consisting of 12 monthly emission fluxes for the year 1990. Annual total emissions fluxes for these prescribed natural emissions are listed in Table 5.

3.5. Ozone
3.5.1. GC3.1

In GC3.1, ozone is derived from the CMIP6 ozone data sets which are weighted means of historical and projection Chemistry-Climate Model Initiative simulations by the CESM1-WACCM (Solomon et al., 2015) and the CMAM (Jonsson, 2004) models (ChecGarcia et al., 2018; Morgenstern et al., 2017). The data sets are provided on pressure levels. For usage in GC3.1 we linearly interpolate them to the models' terrain-following

hybrid-height coordinate. The interpolation requires an estimate of the geopotential height, z, of the pressure levels of the source data, which we derive from the provided zonal mean temperature using the "hypsometric equation"

$$z(\lambda, p, t) = \frac{R}{g} \int_{p}^{p_0} T(\lambda, p', t) \quad \mathrm{d} \ln p' \tag{1}$$

under the assumptions of hydrostatic balance and a constant surface pressure $p_0 = 1013$ hPa. In equation (1), $R = 287.058$ J·kg^{-1}·K^{-1} is the specific gas constant, $g = 9.81$ m s^{-2} is the Earth's gravitational acceleration at the surface, λ is latitude, p is pressure, and t is time.

Prescribing ozone concentrations in climate simulations leads to a mismatch between the internally generated thermal tropopause height and prescribed ozone tropopause height. With the relatively high vertical resolution used in these models, this mismatch becomes greater than in previous generations of models, and it poses the greatest problem in abrupt-4xCO2 simulations, in which preindustrial ozone concentrations are prescribed. In this example, where the thermal tropopause is higher than the ozone tropopause, erroneously high ozone concentrations are prescribed in the upper troposphere, leading to an increase in cold point temperature, excessive stratospheric water vapor, and hence increased radiative heating of the troposphere. Without correction this effect drives a nonphysical positive feedback under a warming climate.

Thus, in the majority of GC3.1 simulations for CMIP6, the prescribed ozone concentrations are redistributed following the method of Hardiman et al. (2019) to ensure that they are consistent with the model thermal tropopause height. For details and discussion of the method, see Hardiman et al. (2019), but there follows a brief summary. For each simulation, at the end of each model year, the monthly mean, zonal mean, and thermal tropopause height is calculated at each latitude from data for the previous two model years. Then, the ozone tropopause is defined at 1 km below the thermal tropopause by setting ozone concentrations there to 80 ppbv and smoothing across the tropopause. The mass of ozone removed from the troposphere is added to the stratosphere by multiplying stratospheric ozone concentrations everywhere by a constant to conserve the total global mass of ozone. This scheme avoids the nonphysical positive feedback in strongly warming simulations, and Hardiman et al. (2019) find that it reduces the apparent effective climate sensitivity derived from abrupt4xCO2 experiments by approximately 10%.

This redistribution was not included in the preindustrial control (piControl) simulations, the N216 historical simulations prior to 1951, or decadal hindcasts and forecasts for the Decadal Climate Prediction Project (DCPP; Boer et al., 2016), all of which were completed prior to the implementation of this scheme and could not be rerun due to time constraints and limited computational resources. All other DECK and CMIP6 MIP contributions with GC3.1 include the ozone redistribution scheme. Sensitivity tests indicate no significant impact of the remapping on global mean quantities (e.g., long-term mean top-of-atmosphere radiative flux, surface temperature, and equatorial tropopause temperature) or zonal mean latitude-height profiles of temperature and specific humidity in piControl simulations (Hardiman et al., 2019). Based on the analysis performed thus far, the lack of ozone remapping prior to 1950 in the N216 historical runs has negligible impact on results, largely because there is little tropospheric warming in this model before 1950. Further analysis is underway to confirm this and will be reported elsewhere in this special collection. Table A8 shows how users of the published data can use the CMIP6 file metadata to determine whether or not the remapping is used in a given simulation.

3.5.2. UKESM1

In UKESM1, ozone is fully interactive and calculated using a medium complexity, coupled stratosphere-troposphere chemistry scheme similar to the one described by Morgenstern et al. (2017, their section 2.11 regarding "MetUM-based participants"). This interactive ozone field feeds both into radiation, affecting shortwave and longwave radiative transfer, as well as into the aerosol scheme, contributing to oxidation of aerosol precursors (Archibald et al., 2019).

Surface conditions are prescribed for the ozone-depleting substances CFC-11, CFC-12, and CH_3Br, as well as for H_2 and carbonyl sulfide (COS) in the same manner as described in section 3.3.2 for CH_4 and N_2O (Morgenstern et al., 2009). The values of CFC-11, CFC-12, and CH_3Br are prescribed using the global mean surface concentrations of Meinshausen et al. (2017). CFC-11, CFC-12, and CH_3Br also contain contributions from other Cl- and Br-containing species to ensure the correct stratospheric chlorine and bromine loading (see Table 6). The surface mixing ratios of H_2 and COS are fixed at 500 ppbv and 482.8 pptv, respectively.

Table 6
Species Contributing to the Surface Specification of CFC-11, CFC-12, and CH₃Br

CFC-11	CFC-12	CH_3Br
CCl_4	CFC-113	H-1211[a]
CH_3CCl_3	CFC-114	H-1202
HCFC-141b	CFC-115	H-1301
HCFC-142b	HCFC-22	H-2402
H-1211[a]		
CH_3Cl		

Note. Contributions are included by moles of Cl or Br.
[a]H-1211 contributes to both CFC-11 and CH_3Br as it contains both Cl and Br.

3.6. Land Use

The implementation of land use change differs between UKESM1 and GC3.1 because they handle vegetation in different ways. UKESM1 simulates vegetation cover interactively, and the land use areas must be specified in a manner which constrains the dynamic vegetation scheme. On the other hand, vegetation fractions are prescribed in GC3.1, and the time-varying forcing data have to be projected onto the model's input data set.

3.6.1. GC3.1

To represent land cover in GC3.1 model, a historical database of nine land surface types is required, including urban, bare soil, lake, ice, and five plant functional types (PFTs), namely, C3 and C4 grasses, needleleaf and broadleaf trees, and shrubs (Essery et al., 2003). The CMIP6 Land Use Harmonization project v2h data set (LUHv2h) data set does not contain land cover which maps to all of these surface types. We therefore follow a similar approach to that used by Baek et al. (2013) for HadGEM2-AO and by Betts et al. (2006) for earlier Hadley Centre models, in which a present-day land cover map is adjusted to follow the time-varying agricultural areas in the forcing data set.

Our starting point is a near present-day vegetation climatology derived from the International Geosphere and Biosphere Programme DISCover land cover data set (IGBP-DIS, Loveland et al., 2000) that has routinely been used in previous Met Office climate models (Baek et al., 2013; Martin et al., 2011; Walters et al., 2019). The IGBP-DIS land cover data set land cover classifications are mapped onto the model's land surface-type fractions (defined in Walters et al., 2019). To construct time-varying land cover maps including the impact of historical changes in anthropogenic land use, we combine this observed specification of land cover with mappings to the LUHv2h historical reconstruction of land cover. The approach is to represent changes in crops and pasture from the LUH2v2 data set as a combination of changes in C3 and C4 grasses at the expense/addition of clearing/planting a corresponding fraction of trees (needleleaf and broadleaf) and shrubs. Changes are implemented in such a way as to preserve the observed climatological proportions of C3 to C4 grasses and of needleleaf trees to broadleaf trees and to shrubs. We do this in anomaly space, relative to the specified present-day observed land cover maps, to ensure that absolute fractions of all surface types are consistent with our observed specification at the present day. As in UKESM1, we do not include rangeland in our definition of pasture (see below). The fractions of bare soil, inland water, urban and ice remain unchanged.

3.6.2. UKESM1

The land use scheme within UKESM1 designates a portion of each gridbox as cropland and a portion as pasture land, where only crops and pasture grasses can grow, respectively, to the exclusion of trees and shrubs (Sellar et al., 2019). In the remainder of the gridbox, nine natural PFTs compete for space, which determines the distribution of forests, grasslands, shrublands, and bare soil (Harper et al., 2016). UKESM1 therefore requires prescribed time series of the fractional cover of each of crop and pasture within each gridbox. We define cropland as the sum of five crop types (C3/C4 annual/perennial and C3 nitrogen) from the time-dependent historical reconstruction of land cover classification change in the LUHv2h, updated from Hurtt et al. (2011) for use in CMIP6. We define pasture land using the managed pasture land use class in the LUHv2h data set. The LUHv2h data set splits grazing land into managed pasture and rangeland, with drier and less populated land more likely to be classified as rangeland. In implementing this forcing data set, we therefore face a choice in whether to include the rangeland areas within the pasture fraction given to the model. We regard rangeland as regions where animals graze on and around natural vegetation,
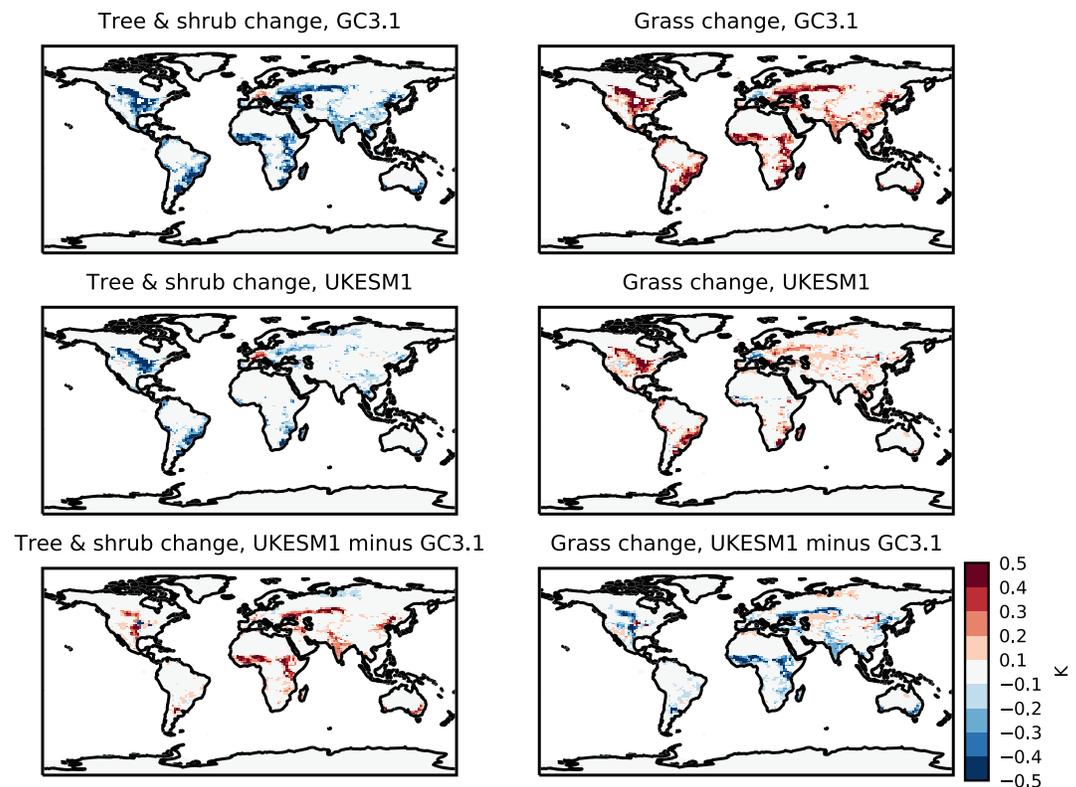
Figure 4. Changes from preindustrial to present-day of tree and shrub PFTs (left column) and grass PFTs (right column), calculated as the 1995–2014 mean of the CMIP6 historical experiment ensemble mean, minus the first 50 years of the piControl (representing 1850 conditions). Top row: GC3.1; middle row: UKESM1; bottom row: UKESM1 minus GC3.1.

and hence, no clearing or deforestation occurs. Because only grass can grow in the designated pasture area within UKESM1, we therefore do not include rangeland in the prescribed pasture forcing data.

3.6.3. Land Cover Changes in GC3.1 and UKESM1

While the two methods of implementing land use change are different, they affect biophysical forcing (i.e., surface albedo and surface fluxes of heat, water, and momentum) via the same mechanism: changes in the fractional cover of vegetation types. This mechanism can be compared between the two models to give an indication of the potential impact of the differing implementation method. Figure 4 contrasts the preindustrial to present-day change in aggregated grass and woody PFTs between the two models. In GC3.1 these changes are fully prescribed, while in UKESM1 they are partially prescribed and partially simulated. In general, the magnitude of the net transition from trees to grasses is larger in GC3.1. Because of this difference, one would expect UKESM1 to exhibit a weaker response to land use change, all else being equal. Work is underway to determine how much of this difference in land cover change is due to biases in UKESM1's background simulation of natural vegetation and how much is due to differences in the land use implementation. de Noblet-Ducoudré et al. (2012) found that the primary driver of differences between model responses to land use change was the translation of land use into changes in land cover. Dedicated experiments within the CMIP6 Land Use MIP (LUMIP; Lawrence et al., 2016) aim to directly compare different models response to the same change in land cover.

3.7. Nitrogen Deposition

The JULES land surface component (Best et al., 2011) within UKESM1 includes a new scheme to represent limitation of carbon uptake by vegetation when available nitrogen is scarce. It is important to represent the natural deposition to the land surface of nitrogen from the atmosphere, which provides a source term to the land carbon-nitrogen system, offsetting losses due to leaching and gaseous emission. We use nitrogen deposition from the same climate-chemistry models as provided the ozone data sets (ChecGarcia et al., 2018). The deposited nitrogen is added directly to the inorganic nitrogen pool, which is available for plants to

take up via their roots; the scheme makes no distinction between different nitrogenous species, and the total amount of nitrogen deposited is assumed to be entirely accessible by plants. The four nitrogen-containing species provided for nitrogen deposition forcing by input4MIPs, wet and dry fluxes of oxidized (NO_y), and reduced (NH_x) forms are therefore combined into one deposition source term.

The MEDUSA2 ocean biogeochemistry component in UKESM1 has a closed nitrogen budget (i.e., no lateral or surface flux) as described in Yool et al. (2013). We therefore make no use of nitrogen deposition forcing at ocean points.

### 3.8. Surface Boundary Forcing for AMIP Experiment

The AMIP experiment, in which the atmosphere and land surface components run uncoupled from the ocean (Eyring et al., 2016), requires surface forcing data that are not needed in coupled experiments. As per the AMIP experiment protocol, both GC3.1 and UKESM1 use prescribed sea surface temperatures (SST) and sea ice concentration. UKESM1 requires additional surface forcing, described below.

The SST and sea ice are taken from the unmodified data set of Durack and Taylor (2017a) and horizontally interpolated to the grids of the two model resolutions (N96 and N216) using an area-weighted regridding. Following the method of Taylor et al. (2000), modifications to SST and sea ice concentration values were then made to ensure that monthly means of the daily values derived by the model during its simulation are the same as in the original interpolated data. This is needed because these daily values are obtained by linear interpolation between midmonth values, which can lead to damping of seasonal and interannual variability. The resulting data are applied at the lower boundary of the atmospheric model. The model applies a limit to small sea ice concentrations, setting values of less than 0.3 to 0.0. This cutoff in sea ice concentration was originally required in an earlier version of the model due to limitations in the radiation and boundary layer schemes used at that time. It is not known whether such an approach is still required in the latest version of the model, and we will revisit this before the next round of AMIP simulations. This means that the effective marine boundary conditions in the AMIP simulations differ slightly from the CMIP6 AMIP specification.

The atmosphere component of UKESM1 calculates fluxes of DMS and PMOA at the ocean surface, driven by the ocean biogeochemistry component's interactively simulated seawater concentrations of DMS and chlorophyll-a, respectively. The chlorophyll-a is also used in the calculation of ocean surface albedo using the parametrization of Jin et al. (2011). For the AMIP experiment, we replace these interactive inputs with DMS and chlorophyll-a monthly climatologies diagnosed from a coupled historical experiment in order to maintain traceability with the coupled model. We chose to use climatologies rather than transient time series as the interannual variability in the latter is related to the model's temperature variability, which would be out of phase with the observed SSTs imposed in the AMIP experiment. The climatologies were compiled from the 1979–2014 period of a single UKESM1 historical ensemble member (r5i1p1f3; see Appendix A). As in the coupled model, the ocean's chlorophyll-a values are scaled by 0.5 before use in the atmosphere model to reduce the influence on PMOA emissions and surface albedo of a mean state bias in the ocean model (Sellar et al., 2019).

GC3.1 also requires inputs of DMS seawater concentration to drive DMS emissions and chlorophyll-a concentration to drive the surface albedo calculation only (GC3.1 does not simulate PMOA emissions). As in its coupled counterpart, the GC3.1 AMIP experiment uses present-day observation-based monthly climatologies for these fields: specifically, the Lana et al. (2011) DMS data set and a climatology of surface chlorophyll derived from the GlobColour merged satellite product (Ford et al., 2012; Walters et al., 2019).

Given the multicentury timescales required to spin-up the terrestrial carbon and nitrogen cycle in UKESM1, it is not computationally affordable to produce an initial state for an AMIP simulation in which the carbon-nitrogen cycle is in balance with the climate driven by the imposed SSTs. Thus, in order to maintain consistent forcing due to land use change between the UKESM1 coupled and AMIP experiments, dynamic vegetation is deactivated in the latter and replaced by prescribed vegetation properties from a coupled historical situation. The fields prescribed are vegetation fraction, leaf area index (LAI), and canopy height, all taken from the same member of the historical ensemble as the ocean DMS and chlorophyll-a climatologies. As with the ocean fields, we use 1979–2014 climatologies for LAI and canopy height to avoid any imprint of the coupled model's interannual SST variability (monthly and annual mean climatology for LAI and canopy height, respectively). In contrast in the case of vegetation fraction, we retain the transient land use signal by applying a 1979–2014 monthly mean time series. The vegetation fractions respond to climate variability
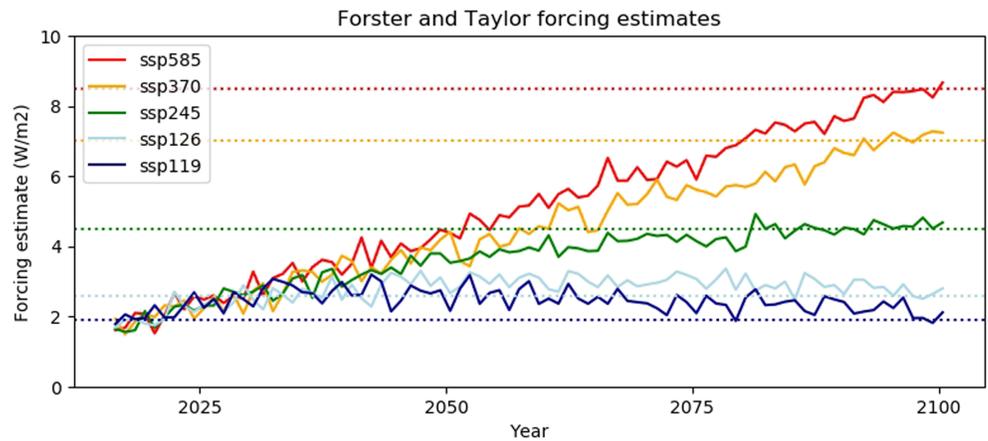
Figure 5. Time series of Forster and Taylor forcing for the four ScenarioMIP Tier 1 scenarios and SSP1-1.9. Horizontal dotted lines mark the intended forcing at 2100 for each scenario.

slowly (with a timescale of several years) and thus contain little imprint of the coupled model's interannual variability.

3.9. Freshwater Forcing From Ice Sheets

With no accounting of the ice sheet mass balance, the ice sheets will accumulate snow indefinitely and the ocean will lose mass. Thus, outflow of ice sheets to the ocean is simulated by three components in GC3.1 and UKESM1: surface runoff, icebergs, and ice shelf basal melt. Surface melt runoff occurs in the JULES land surface model with outflow to the ocean through through the TRIP river routing scheme (Oki & Sud, 2006).

The remaining flux is calibrated based on the mean surface mass balance over a 100-year period of the model spin-up, prior to the start of the piControl. The surface mass balance is calculated separately for each hemisphere as the mean rate of increase in snow mass over the calibration period. In the Southern Hemisphere, 45% of this flux is applied as an iceberg calving source term for NEMO's Lagrangian iceberg scheme; the calving flux is distributed spatially by scaling the climatology of Marsh et al. (2015) to achieve the desired total flux. The other 55% is applied as a basal shelf melt, spatially distributed according to the three-dimensional melt pattern of Mathiot et al. (2017). The Northern Hemisphere flux is implemented solely as an iceberg calving flux, again by scaling the Marsh et al. (2015) data set. Note that this approach differs from the interactive calculation described in Williams et al. (2018), in order to allow the ocean volume to evolve in historical and future simulations.

3.10. Estimated Radiative Forcing

To check that the ScenarioMIP forcings are performing as expected, the net top-of-atmosphere radiative forcing is estimated using the method of Forster and Taylor (2006). This is a global mean energy balance approach, based on the equation:

$$N = F - \lambda T \tag{2}$$

where N is the net top-of-atmosphere radiative flux, F the forcing, λ the feedback parameter, and T the global mean temperature. λ may be estimated from the abrupt-4xCO2 experiment using linear regression (Gregory, 2004). Assuming that λ is approximately constant, F may then be estimated for each year of a given scenario using annual model values of N and T.

Results using UKESM1 are shown in Figure 5 for the five ScenarioMIP scenarios designated as highest priority for the IPCC sixth assessment report: SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. This shows that the estimated forcings match closely the intended forcing at 2100 and that the time series are reasonable (c.f. O'Neill et al., 2016).

4. Initialization Strategy

An important requirement when using climate models to simulate the historical past and investigate possible future Earth system change is that an ensemble of simulations is used that samples the full range of
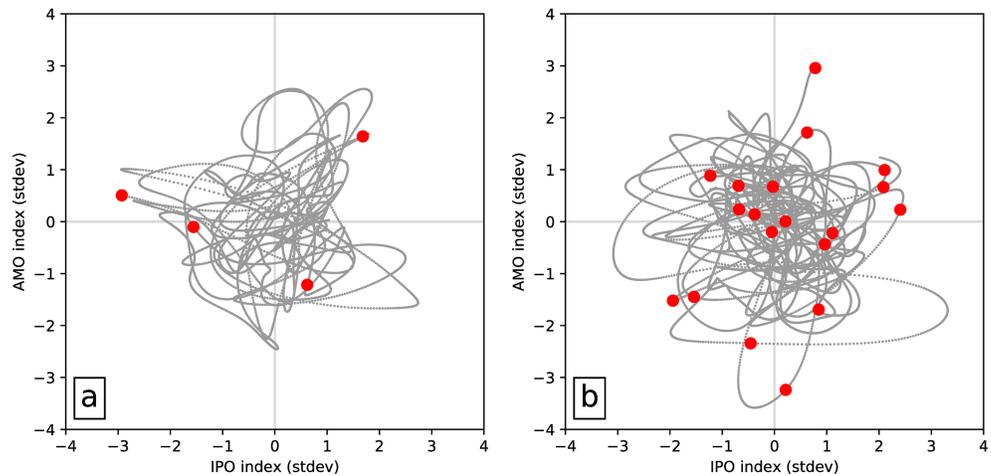
Figure 6. IPO:AMO phase space for (a) N96ORCA1 GC3.1 and (b) N96ORCA1 UKESM1 preindustrial control experiments revealing the monthly evolution of the climate modes over 500 and 858 years, respectively. The initial conditions chosen for 4 GC3.1 historical experiments and 19 UKESM1 historical experiments are indicated with red dots. The same approach (not shown) is used to select N216ORCA025 historical initial conditions from the N216ORCA025 preindustrial control experiment with GC3.1.

unforced variability simulated by the model and, ideally, observed in the real world. Such an ensemble supports a number of important analyses: (i) evaluation of the simulated modes of natural variability against equivalent observed modes, (ii) assessment of possible future changes in major modes of variability (e.g., changes in the frequency of occurrence, intensity, or geographical location of such modes) under different warming scenarios, and (iii) through ensemble averaging, provide a clearer identification of the forced climate change signal, separated from the confounding influence of trends due to natural variability. The most common method for generating such an ensemble is to ensure that the initial conditions used to start each historical simulation (in CMIP6 defined as starting 1 January 1850) sample the major modes of variability in the model's preindustrial control, from which each historical member is initialized.

For both the GC3.1 and UKESM1 historical ensembles, we aimed to ensure that each set of initial conditions fully sampled the model's preindustrial simulation of both the Interdecadal Pacific Oscillation (IPO; Power et al., 1999; Zhang et al., 1997) and the Atlantic Multidecadal Oscillation (AMO; Kerr, 2000), which manifest themselves by significant basin-scale variability in SSTs. To realize this, we calculated the first Empirical Orthogonal Function (EOF) of SST variability for each basin and plot the principal components of these EOFs in a phase space diagram (Figure 6). Using these diagrams, we identify years from the two piControl simulations that sample each model's internal variability in the joint EOF time series. Model states for these years act as initial conditions for subsequent historical simulations made with each model. In addition to sampling the joint IPO/AMO variability, we further require that selected piControl model states are a minimum of 30 years apart; increasing the likelihood that the ocean states in each historical initial condition are distinctly different. Table 7 shows the dates chosen for each member of the historical ensemble, for each of the model configurations.

The UKESM1 piControl exhibits extended periods of multidecadal to centennial variability in SST across the Southern Ocean of sufficient magnitude to influence global mean SST. This variability is linked to periodic deep ocean overturning in regions close to the Antarctic coast, such as the Weddell and Ross Seas. Overturning brings accumulated warm, saline deep water to the surface inducing large-scale reductions in Antarctic sea ice (i.e., polynyas; Campbell et al., 2019), warm SSTs, and ventilation of ocean heat to the atmosphere. Such variability has been seen in other coupled models (de Lavergne et al., 2014; Martin et al., 2013) and may help explain recent observed climate trends in the Antarctic region (Zhang et al., 2019). The timescale of this variability in UKESM1 is controlled by accumulation of sufficient warm water in the deep Southern Ocean to erode the near-surface density gradient between cold, fresh (lighter) water overlying warm, saline (denser) water. Once this density barrier is overwhelmed, convective mixing from below occurs and the accumulated heat is lost to the atmosphere. As this variability influences global mean SST on centennial timescales, we decided that it was important to sample in the UKESM1 historical initial conditions in

Table 7
Dates at Which Historical Ensemble Members Branched From the Respective piControl

Realization index	piControl branch date		
	UKESM1.0	GC3.1-N96ORCA1	GC3.1-N216ORCA025
r1	2250-01-01	1850-01-01	1850-01-01
r2	2165-01-01	1885-01-01	1885-01-01
r3	2120-01-01	1930-01-01	1930-01-01
r4	1960-01-01	1970-01-01	1960-01-01
r5	2020-01-01		
r6	2050-01-01		
r7	1995-01-01		
r8	2395-01-01		
r9	2285-01-01		
r10	2340-01-01		
r11	2460-01-01		
r12	2210-01-01		
r13	2565-01-01		
r14	2685-01-01		
r15	2745-01-01		
r16	2629-01-01		
r17	2716-01-01		
r18	2760-01-01		
r19	2815-01-01		

Note. As a perpetual-1850 simulation, the choice of start year for piControl is an arbitrary label. The UKESM1.0 piControl begins in 1960, the GC3.1 piControls begin in 1850. Dates are formatted as YYYY-MM-DD.
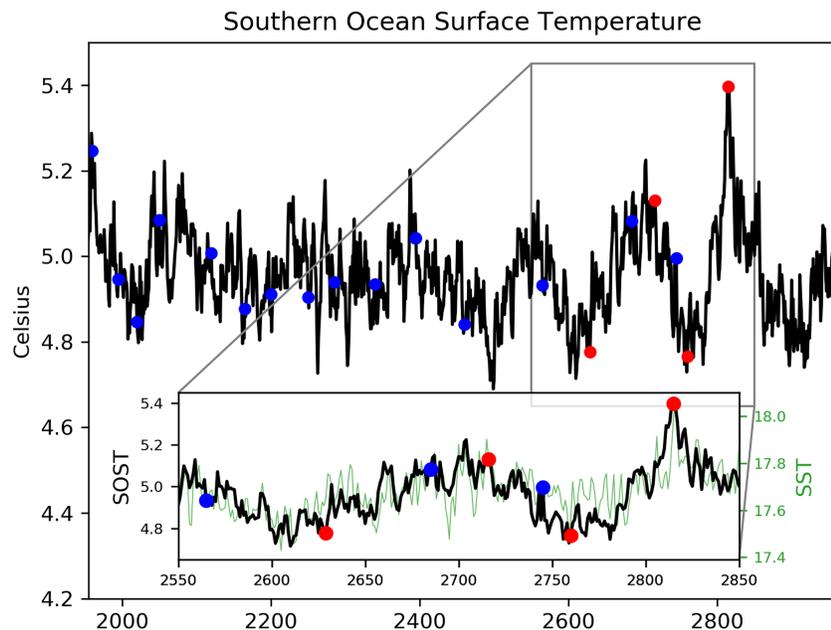
addition to the IPO/AMO variability. Figure 7 shows annual mean SST, averaged over the Southern Ocean (south of 40°), for the 1,000 years of the UKESM1 piControl from which the historical initial conditions are drawn. Blue dots on the figure show the first 15 initial conditions, selected solely by sampling the UKESM1 IPO/AMO phase space diagram. The four red dots indicate selected initial condition dates for four UKESM1 historical members, where we instead chose piControl states that capture two maxima and two minima of the centennial timescale Southern Ocean SST variability. The inset in Figure 7 focuses on the 300 years of the piControl when these four model states occur. The full black line shows the mean Southern Ocean SST, while the light green line plots simulated global mean SST over the same period. Clearly, the Southern Ocean SST variability has a significant imprint on the global mean SST.

Prior to their respective piControl simulations, the models were spun up as follows.

- GC3.1-N96ORCA1 was initialized from 1950 ocean conditions taken from the EN4 analysis data set (Good et al., 2013) and run for 615 years under CMIP6 preindustrial forcing.
- GC3.1-N216ORCA025 was initialized from EN4 2000 ocean conditions and run for 232 years under CMIP6 preindustrial forcing.
- The UKESM1.0 spin-up was more complex and will be documented in a companion paper in this special issue, but in summary it involved a 500-year coupled simulation preceded by separate ocean and land spin-up simulations of duration 5,000 and 1,000 years respectively.

5. Experiment Reproducibility

The experiments that are performed for CMIP6 require large amounts of compute resource and development effort, and their results will be used over a period of several years. In view of this expense and protracted lifetime, we have invested a good deal of effort in ensuring that our results are reliable. Our approach has been informed by lessons learned during our participation in CMIP5, and we hope that documenting the

Figure 7. Full figure: Time series of annual mean sea surface temperature (SST) averaged over the Southern ocean (south of 40°) for 1,000 years of the UKESM1 piControl simulation. Inset figure: Time series of annual mean Southern ocean mean SST (black line) and annual mean global mean SST (green line) centered on 300 years of the UKESM1 piControl from which initial conditions the UKESM1 historical members 16–19 were selected. Blue dots indicate the piControl dates selected as initial conditions for UKESM1 historical simulations using the IPO/AMO phase space diagram. Red Dots indicate UKESM1 initial condition dates, selected based on sampling variability in Southern ocean SST.

infrastructure we have used here will be helpful to those in future phases of CMIP and other large intercomparisons. It includes using a controlled environment for the development of model code, designing models which are restartable, documenting model parameters for a given experiment, checking results while the model is running, and testing the reproducibility of model behavior on different platforms. This effort also extends to the preparation and curation of the climate forcing data for the different experiments. We describe each of these aspects in more detail in the following subsections.

We note in passing that our models produce results in their native format; our procedure for converting this to the standardized format required by CMIP will be described elsewhere. In case it is useful to other centers, we mention here that version 01.00.29 of the CMIP6 data request was used in publishing data on the Earth System Grid.

5.1. Code Development
The source code for the various components of the model (atmosphere, ocean, land, etc.) is maintained in repositories in order to manage changes in a controlled fashion and to facilitate collaborative development. We note that each component is under more or less continuous development by communities of engineers and scientists that are distributed across multiple specialized institutions, which makes the use of a controlled development environment essential. We also follow software engineering good practice in the development of the model components, including the use of code standards and documentation, regression testing, and code review.

Besides being used for component development, some of these practices are also employed elsewhere—for example, we use repositories for the encapsulation and maintenance of model parameters for different experiments—see section 5.3—and two stages of code review are used for the applications which generate the forcings—see section 5.6, below.

5.2. Model Restarts
Owing to their complexity, climate simulations typically execute for long periods of CPU time—typically, of the order of several months. This period exceeds the maximum execution time usually offered by HPC environments (typically no longer than a day). It is therefore critical to have an efficient checkpoint-restart

Table 8
Resource Usage and Output Size for CMIP6 Experiments Run at the Met Office

CMIP6 source_id	Computer usage (CPU core years)	Model output size[a] (PB)
HadGEM3-GC31-LL	17,123	3.38
HadGEM3-GC31-MM	8,219	2.59
UKESM1-0-LL	27,283	7.99
Total	52,625	13.96

[a]Data volume of native model output, prior to conversion to CMIP6 data format (see section 5).

strategy, in order to allow simulations to start from saved states. In addition, model runs may subsequently need to be restarted at any point of the simulation in order to study or reproduce a subset of the model output. This can happen at any point after the initial simulation has been completed. This means that the model must be restartable in a bit-for-bit manner from checkpoints held in data archives; that is, a run of n steps should generate identical results to a run of m steps that had been started from the endpoint of a run of n-m steps. Because of the multicomponent nature of the model, restart states are usually distributed across multiple files and the restartability process can be fragile. We have taken great care when designing and testing the model to ensure that bit-reproducible restarting is maintained throughout the development; in particular, we have built regression tests for restartability into the workflows for our experiments.

5.3. Experiment Configuration

The configuration of each experiment, incorporating the setting of model parameters, specification of input data (including forcing data; see section 5.6), and definition of the dependencies and scheduling of component tasks, is encapsulated as a Rose suite (The Rose toolkit, 2012). Rose is a framework for developing and running application configurations and uses the Cylc (Oliver et al., 2018) workflow engine.

The experiments' Rose suites are maintained in a repository in the same fashion as the source code for the model components (see section 5.1, above) which facilitates the maintenance of experiment configuration, allowing changes in, for example, parameter settings to be logged and documented by scientists.

5.4. Quality Control

As noted above, these simulation runs typically require months of CPU time; in addition, they generate large amounts of output—usually of the order of hundreds of GB per model year. Both factors increase the chance of the occurrence of failures in hardware or other infrastructure such as tape drives (used for archiving output files) during the course of the run. Experience has shown that it is illuminating to perform checks on the results as they are being produced, instead of after the simulation has come to an end. More specifically, we have incorporated into our workflow tests which check for the presence of expected output files in the tape archive while the run is in progress. This has enabled early detection of archive or disk problems and avoided the expense of repeating part of a simulation after it has completed.

5.5. Reproducibility Across Platforms

Although we require that two runs of the same model generate bit-wise identical results from the same starting point (see section 5.2, above), this is not generally possible if the same model is run on different machines. This is because of the unpredictable way in which minute differences propagate through the simulation on different platforms, leading to differing results. This applies both to machines in different physical locations and to the same machine before and after an upgrade which does not preserve bit-level compatibility (e.g., a major update of the operating system or compiler).

Instead of asking for identical model behavior on two different machines, we seek to verify that each model configuration is scientifically consistent with the other; that is, could each have been sampled from the same ensemble of results generated on either machine? To check this, we create an ensemble of short (24 hr) runs on each machine by perturbing selected variables in their initial conditions using a perturbation whose numerical value is comparable with the machine's precision. The spread of results (at each point in time and space) on each platform can then be used to determine whether they could have come from a common ensemble. The statistical methodology used in this comparison will be reported in future work.

We have used this method, together with other analyses, such as comparison of the mean state from multidecade simulations, to verify the consistency of ports of UKESM1. This model was developed and tested on the internal U.K. Met Office HPC before being ported to ARCHER, a U.K. national supercomputing platform, and to machines run by the National Institute of Meteorological Science/Korean Meteorological Administration and New Zealand's National Institute of Water and Atmospheric Research.

5.6. Forcing Data

The forcing data for a given experiment define the boundary conditions for the model and represent an important aspect of experiment reproducibility. We obtained forcing data from the Input4MIPs data archive (Durack et al., 2018) to take advantage of common metadata standards and versioning policies. In Appendix A we tabulate the specific data sets from Input4MIPs which were used in these core simulations. There were two cases in which, by necessity, we used data from sources other than Input4MIPs:

1. Stratospheric aerosol radiative properties (see section 3.2.1) were calculated specific to our model wavelength bands and obtained directly from ETH (ftp://iacftp.ethz.ch/pub_read/luo/CMIP6).
2. CO_2 emission global annual totals used in scaling gridded emissions (see section 3.3.3) were downloaded online (http://www.globalchange.umd.edu/ceds/ceds-cmip6-data/).

To ensure that the forcing files are reproducible, metadata to identify the processing code and the source files were stored as attributes for netCDF output files or as an accompanying JSON file. Specifically, the metadata were as follows:

- The command and arguments used to generate the output file, which was stored in the `history` attribute as per version 1.6 of the conventions for climate and forecast metadata (CF conventions and metadata, 2019).
- The repository URL and revision of the source code used.
- A checksum for the source files used.

For a multistage data processing pipeline, the metadata was appended at each stage. Hence, the metadata for the final file contains the complete processing chain back to the original source files from input4MIPs (Durack et al., 2018).

As mentioned in section 5.1, two stages of code review were performed for the forcing processing: a scientific review to verify that the code was an accurate representation of the science, followed by a technical review to ensure that the code used an interface suitable for integration into a Rose suite.

Processing of the forcing data was performed on JASMIN, the NERC/Science and Technology Facilities Council data cluster which incorporates the U.K. node of the Earth System Grid. JASMIN also mirrors the input4MIPs repository; additional required source files were held in a single directory alongside the input4MIPs data, and this was archived to tape to preserve reproducibility. After review, the final forcing files were stored in a directory with restricted permissions; this was then mirrored to the sites where the model runs were being performed and also archived to tape. The tape archive of the forcings would have enabled us to recover from a catastrophic failure more quickly than reproducing all files from original sources and will ensure that identical model input files are available even after system upgrades prevent bit-identical runs of the processing code.

6. Summary and Discussion

We have documented the implementation of U.K. models for a central set of CMIP6 experiments, that is, the DECK and historical simulations and future scenarios. We have outlined the main technical processes used to ensure reproducibility of the simulations, to ensure the scientific integrity of the results, and to minimize costs associated with technical failures. And we have described the technical and scientific implementation of forcing data sets and the model-specific assumptions made in using these data sets.

In some cases these assumptions will have significant effects on the effective forcing arising from these forcing agents. For example, the UKESM1 implementation of land use, which prevents the model's dynamic vegetation scheme from growing natural PFTs in areas of prescribed crop and pasture, makes the biogeochemical and biophysical impacts of land use change subject to background model biases in the simulation of natural vegetation cover. This is a necessary consequence of including a level of process complexity which enables us to simulate interactions between climate change, the carbon-nitrogen cycle, and land use change,

but it needs to be borne in mind when interpreting the results from this model. Andrews et al. (2017) discuss the implications of these choices in more detail, in the context of the predecessors to GC3.1 and UKESM1.

We do not include rangeland within the agricultural land use states imposed in either UKESM1 or GC3.1, on the assumption that in rangeland animals graze on or around natural vegetation, without land clearance. Neither of these models are able to represent the effect of grazing on vegetation (indeed, very few models can, Pongratz et al., 2018), and so the only choice available within the framework of these models is whether or not to include rangeland within the prescribed grass PFTs used to represent pasture. Including rangeland in the UKESM1 pasture forcing, or the grass PFTs imposed in GC3.1, would remove all shrubs and trees from these areas, implying complete clearance of natural vegetation, in contradiction with our understanding of the definition of rangeland. By neglecting rangeland, we may underestimate land cover change in some regions. Grazing can have significant effects on the biophysical and biogechemical properties of vegetation (Erb et al., 2017), although it is not clear that it is important in rangelands, where the intensity of grazing is low. Conversely, including rangeland within the prescribed grass areas would overestimate land cover change in other regions. For example, a particular region of interest is Australia in which the inclusion of rangeland would lead to excessive natural vegetation being removed. Finally, we note that the decision to exclude rangeland was consistent with the recommended use of LUH2 at the time of implementation.

Similarly, the choices made for the prescription of marine biogeochemical fluxes and terrestrial vegetation in the UKESM1 AMIP experiment will impact the simulation results. Our choices were driven by the desire to understand the direct impact of coupled model temperature biases on the atmosphere model, while keeping other Earth system properties traceable between the two. One could equivalently attempt to derive observation-based data sets for these some of inputs in order to analyze the atmosphere in the absence of coupled model biases, but this would have severely reduced traceability to the coupled model by altering the pattern of aerosol radiative effect and vegetation-climate interactions.

As noted in section 3.1, UKESM1 includes photolytic reactions in its interactive simulation of ozone, but the photolysis rates use fixed solar inputs and therefore have no dependence on the solar variability. Dennison et al. (2019) show that including solar variability in photolysis calculations can have a noticeable effect on ozone production, resulting in variations of order 1% in extratropical total column ozone. This mode of ozone variability will therefore be absent from the UKESM1 CMIP6 simulations but will be a priority for inclusion in future versions of UKESM.

Particular care has been taken in the GC3.1 model configurations, which do not simulate interactive chemistry, to avoid inconsistencies between the model thermodynamics and prescribed ozone. Such inconsistencies have the potential to lead to nonphysical feedbacks under high-end climate change scenarios, and to prevent this, an interactive redistribution of the ozone field is performed. This remapping alters the spatial distribution of ozone, particularly near the tropopause, and while it is done in such a way as to minimize the impact on global mean radiative forcing, this redistribution should be considered when analyzing results of GC3.1 simulations. The potential for these inconsistencies exists in all models which do not simulate ozone interactively, particularly those which have a high climate sensitivity, and in GC3.1 affected estimates of ECS by around 10% (Hardiman et al., 2019). We encourage other modeling groups to describe how they mitigate this risk and suggest that future phases of CMIP include recommendations for handling this issue which would minimize unwanted model divergence due to nonphysical feedbacks and aid understanding of how widespread such issues are in CMIP simulations.

Finally, we note that—as has been widely recognized—CMIP6 is larger and more complicated than previous phases of CMIP, prescribing hundreds of experiments and thousands of model output variables. A summary of the computing resources used by the experiments that have been run on the Met Office supercomputer (a Cray XC40) is presented in Table 8. Nearly 14 PB of native model output has been produced thus far, and we expect to publish around 5 PB of CMIP6 data from all experiments performed in the United Kingdom. The size of this published data will be equivalent to 4.5 trillion pages of text and more than three times greater than the *entire* CMIP5 data archive. A similar comparison for computer usage is complicated by differences in models, machine specification, and architecture, but the Met Office's CMIP5 experiments used around 920 CPU core years on an IBM Power7 machine. Thus, the U.K.'s CPU resource for CMIP6 simulations is approximately 2 orders of magnitude larger than that for CMIP5, reflecting the increase in the scope of CMIP, as well as higher model resolution and process complexity. We are confident that the new understanding of the Earth system that will be derived from these model results will similarly be greatly enhanced relative to previous projects.

The simulations described in this paper form the core of the U.K.'s contribution to CMIP6 and will underpin many further experiments for CMIP6-endorsed MIPs and other science. We hope that the careful documentation of our experimental configuration will assist others in analysis of the simulations and in setting up new experiments based on these runs.

Appendix A: Forcing Data Set Versions

Tables A1 to A7 document the specific versions of the input data sets used in forcing the majority of DECK, historical, and scenarioMIP experiments, with some exceptions detailed below. For the DECK and historical experiments, these versions correspond to release v6.2.0 of the CMIP6 forcing data. For the purposes of our models, v6.2.0 is identical to the subsequent and final CMIP6 release v6.2.1, as these releases differ only in SO_2 emissions from aircraft, which we do not use.

Table A1
Input4MIPs Forcing Data Sets Used for Greenhouse Gases

Experiment	Data set ID	Data set citation
PI, historical	UoM-CMIP-1-2-0	Meinshausen and Vogel (2016)
SSP1-1.9	UoM-IMAGE-ssp119-1-2-0	Meinshausen and Nicholls (2018c)
SSP1-2.6	UoM-IMAGE-ssp126-1-2-0	Meinshausen and Nicholls (2018d)
SSP2-4.5	UoM-MESSAGE-GLOBIOM-ssp245-1-2-0	Meinshausen and Nicholls (2018e)
SSP3-7.0	UoM-AIM-ssp370-1-2-0	Meinshausen and Nicholls (2018a)
SSP4-3.4	UoM-GCAM4-ssp434-1-2-0	Meinshausen and Nicholls (2018b)
SSP5-3.4-over	UoM-REMIND-MAGPIE-ssp534-over-1-2-0	Meinshausen and Nicholls (2018f)
SSP5-8.5	UoM-REMIND-MAGPIE-ssp585-1-2-0	Meinshausen and Nicholls (2018g)

Table A2
Input4MIPs Forcing Data Sets Used for Land Use

Experiment	Data set ID	Data set citation
PI, historical	UofMD-landState-high-2-1-h	Hurtt et al. (2019c)
SSP1-1.9	UofMD-landState-IMAGE-ssp119-2-1-f	Hurtt et al. (2019a)
SSP1-2.6	UofMD-landState-IMAGE-ssp126-2-1-f	Hurtt et al. (2017a)
SSP2-4.5	UofMD-landState-MESSAGE-ssp245-2-1-f	Hurtt et al. (2018)
SSP3-7.0	UofMD-landState-AIM-ssp370-2-1-f	Hurtt et al. (2017b)
SSP4-3.4	UofMD-landState-GCAM-ssp434-2-1-f	Hurtt et al. (2017c)
SSP5-3.4-over	UofMD-landState-MAGPIE-ssp534-2-1-f	Hurtt et al. (2019b)
SSP5-8.5	UofMD-landState-MAGPIE-ssp585-2-1-f	Hurtt et al. (2017d)

Table A3
Input4MIPs Forcing Data Sets Used for Emissions

Experiment	Data set ID	Data set citation
PI, historical	VUA-CMIP-BB4CMIP6-1-2	van Marle et al. (2016)
	CEDS-2017-08-30	Hoesly et al. (2017a)
	CEDS-2017-08-30-supplemental-data	Hoesly et al. (2017b)
SSP1-1.9	IAMC-IMAGE-ssp119-1-1	Gidden et al. (2018c)
SSP1-2.6	IAMC-IMAGE-ssp126-1-1	Gidden et al. (2018d)
SSP2-4.5	IAMC-MESSAGE-GLOBIOM-ssp245-1-1	Gidden et al. (2018e)
SSP3-7.0	IAMC-AIM-ssp370-1-1	Gidden et al. (2018a)
SSP4-3.4	IAMC-GCAM4-ssp434-1-1	Gidden et al. (2018b)
SSP5-3.4-over	IAMC-REMIND-MAGPIE-ssp534-over-1-1	Gidden et al. (2018f)
SSP5-8.5	IAMC-REMIND-MAGPIE-ssp585-1-1	Gidden et al. (2018g)

Table A4
Input4MIPs Forcing Data Sets Used for Ozone

Experiment	Data set ID	Data set citation
PI, historical	UReading-CCMI-1-0	Hegglin et al. (2016)
SSP1-1.9	UReading-CCMI-ssp119-1-1	Hegglin et al. (2019a)
SSP1-2.6	UReading-CCMI-ssp126-1-0	Hegglin et al. (2018a)
SSP2-4.5	UReading-CCMI-ssp245-1-0	Hegglin et al. (2018b)
SSP3-7.0	UReading-CCMI-ssp370-1-0	Hegglin et al. (2018c)
SSP4-3.4	UReading-CCMI-ssp434-1-1	Hegglin et al. (2019b)
SSP5-3.4-over	UReading-CCMI-ssp534os-1-1	Hegglin et al. (2019c)
SSP5-8.5	UReading-CCMI-ssp585-1-0	Hegglin et al. (2018d)

Table A5
Input4MIPs Forcing Data Sets Used for Nitrogen Deposition

Experiment	Data set ID	Data set citation
PI, historical	NCAR-CCMI-2-0	Hegglin et al. (2016)
SSP1-1.9	NCAR-CCMI-ssp119-1-0	Hegglin et al. (2019a)
SSP1-2.6	NCAR-CCMI-ssp126-2-0	Hegglin et al. (2018a)
SSP2-4.5	NCAR-CCMI-ssp245-2-0	Hegglin et al. (2018b)
SSP3-7.0	NCAR-CCMI-ssp370-2-0	Hegglin et al. (2018c)
SSP4-3.4	NCAR-CCMI-ssp434-1-0	Hegglin et al. (2019b)
SSP5-3.4-over	NCAR-CCMI-ssp534os-1-0	Hegglin et al. (2019c)
SSP5-8.5	NCAR-CCMI-ssp585-2-0	Hegglin et al. (2018d)

Table A6
Input4MIPs Forcing Data Sets Used for Natural Forcing (PI, Historical, and Future)

Forcing	Data set ID	Data set citation
Volcanic	IACETH-SAGE3lambda-3-0-0	
Solar	SOLARIS-HEPPA-3-2	Matthes et al. (2017)

Table A7
Input4MIPs Forcing Data Sets Used for SST and Sea Ice

Experiment	Data set ID	Data set citation
AMIP	PCMDI-AMIP-1-1-3	Durack and Taylor (2017b)

Table A8
Forcing Index Definition for CMIP6 Metadata

Forcing index	GC3.1	UKESM1
f1	v6.1.1; no ozone remapping	Not used
f2	v6.2.0; no ozone remapping	v6.2.0
f3	v6.2.0; with ozone remapping	v6.2.0; historical simulations use 1850 forcing for stratospheric SAD

The exceptions to the use of v6.2.0 are the piControl, 1pctCO2, and abrupt-4xCO2 simulations with GC3.1-N216ORCA025, which were started with v6.1.1 before the updates in v6.2.0 were released. There are two differences between v6.1.1 and v6.2.0:

- Aircraft emissions of all species are corrected. GC3.1 makes no use of aircraft emissions so this difference has no impact.

- Historical stratospheric aerosol properties are updated to remove errors in some years. The update was applied in such a way as to preserve the global mean radiative forcing of the historical 1850–2014 average, which acts as the forcing data set for preindustrial simulations. Therefore, for the purposes of the piControl, 1pctCO2, and abrupt-4xCO2 experiments, v6.1.1 is consistent with v6.2.0.

The CMIP6 metadata conventions enable the encoding of forcing configuration via the "f" component of the `variant-id` file attribute. We have used this to record the use of v6.1.1 versus v6.2.0 in the published simulation data, as indicated in Table A8. Table A8 also shows how this index is used to record the use of ozone remapping for GC3.1 (see section 3.5.1) and the SAD configuration in UKESM1 historical simulations (see section 3.2.2).

References

Andrews, T., Betts, R. A., Booth, BenB. B., Jones, C. D., & Jones, G. S. (2017). Effective radiative forcing from historical land use change. *Climate Dynamics*, *48*(11-12), 3489–3505. https://doi.org/10.1007/s00382-016-3280-7

CF conventions and metadata (2019). http://cfconventions.org

Archibald, A. T., O'Connor, F. M., Abraham, N. L., Archer-Nicholls, S., Chipperfield, M., Dalvi, A., et al. (2019). Description and evaluation of the UKCA stratosphere-troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM1. Geoscientific Model Development Discussions. https://doi.org/10.5194/gmd-2019-246

Arfeuille, F., Weisenstein, D., Mack, H., Rozanov, E., Peter, T., & Brönnimann, S. (2014). Volcanic forcing for climate modeling: A new microphysics-based data set covering years 1600â present. *Climate of the Past*, *10*(1), 359–375. https://doi.org/10.5194/cp-10-359-2014

Baek, H.-J., Lee, J., Lee, H.-S., Hyun, Y.-K., Cho, C., Kwon, W.-T., et al. (2013). Climate change in the 21st century simulated by HadGEM2-AO under representative concentration pathways. *Asia-Pacific Journal of Atmospheric Sciences*, *49*(5), 603–618. https://doi.org/10.1007/s13143-013-0053-7

Beiping, L. (2017). CMIP6 stratospheric aerosol climatology. ftp://iacftp.ethz.ch/pub_read/luo/CMIP6/UMUKCA

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description. Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Betts, R. A., Goldewijk, K. K., & Ramankutty, N. (2006). Hadley Centre Technical Note 70: Radiative forcing by anthropogenic surface albedo change before and since 1750. Exeter, UK: Met Office Hadley Centre.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*(10), 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Campbell, E. C., Wilson, E. A., Moore, G. W. K., Riser, S. C., Brayton, C. E., Mazlof, M. R., & Talley, L. D. (2019). Antarctic offshore polynyas linked to Southern Hemisphere climate anomalies. *Nature*, *570*, 319–325. https://doi.org/10.1038/s41586-019-1294-0

ChecGarcia, R., Hegglin, M. I., Kinnison, D., Plummer, D. A., & Shine, K. P. (2018). Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophysical Research Letters*, *45*, 3264–3273. https://doi.org/10.1002/2017GL076770

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*, *1*, 997–1062. https://doi.org/10.5194/gmd-4-1051-2011

de Lavergne, C., Palter, J. B., Galbraith, E. B., Bernadello, R., & Marinov, I. (2014). Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. *Nature Climate Change*, *4*, 278–282. https://doi.org/10.1038/nclimate2132

de Noblet-Ducoudré, N., Boisier, J.-P., Pitman, A., Bonan, G. B., Brovkin, V., Cruz, F., et al. (2012). Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: Results from the first set of LUCID experiments. *Journal of Climate*, *25*(9), 3261–3281. https://doi.org/10.1175/JCLI-D-11-00338.1

Dennison, F., Keeble, J., Morgenstern, O., Zeng, G., Abraham, N. L., & Yang, X. (2019). Improvements to stratospheric chemistry scheme in the UM-UKCA (v10.7) model: Solar cycle and heterogeneous reactions. *Geoscientific Model Development*, *12*(3), 1227–1239. https://doi.org/10.5194/gmd-12-1227-2019

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., et al. (2006). Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. *Atmospheric Chemistry and Physics*, *6*(12), 4321–4344. https://doi.org/10.5194/acp-6-4321-2006

Durack, P., & Taylor, K. E. (2017a). PCMDI AMIP SST and sea-ice boundary conditions version 1.1.3. https://doi.org/10.22033/ESGF/input4MIPs.1735

Durack, P. J., & Taylor, K. E. (2017b). PCMDI AMIP SST and sea-ice boundary conditions version 1.1.3. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1735

Durack, P., Taylor, K., Eyring, V., Ames, S., Hoang, T., Nadeau, D., et al. (2018). Toward standardized data sets for climate model experimentation. *Eos*, *99*. https://doi.org/10.1029/2018EO101751

Erb, K.-H., Luyssaert, S., Meyfroidt, P., Pongratz, J., Don, A., Kloster, S., et al. (2017). Land management: Data availability and process understanding for global change studies. *Global Change Biology*, *23*(2), 512–533. https://doi.org/10.1111/gcb.13443

Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M., & Taylor, C. M. (2003). Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology*, *4*(3), 530–543. https://doi.org/10.1175/1525-7541(2003)004<0530:EROSHI>2.0.CO;2

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

Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., & Demaria, J. (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*, *8*(5), 751–771. https://doi.org/10.5194/os-8-751-2012

Forster, P. M. F., & Taylor, K. E. (2006). Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *Journal of Climate*, *19*(23), 6181–6194. https://doi.org/10.1175/JCLI3974.1

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, *27*(2), 511–526. https://doi.org/10.1175/JCLI-D-12-00579.1

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018a). input4mips.cmip6.scenariomip.iamc.iamc-aim-ssp370-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2482

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018b). input4mips.cmip6.scenariomip.iamc.iamc-gcam4 -ssp434-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2483

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018c). input4mips.cmip6.scenariomip.iamc.iamc-image -ssp119-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2485

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018d). input4mips.cmip6.scenariomip.iamc.iamc-image -ssp126-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2486

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018e). input4mips.cmip6.scenariomip.iamc.iamc-message -globiom-ssp245-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2487

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018f). input4mips.cmip6.scenariomip.iamc.iamc-remind -magpie-ssp534-over-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2488

Gidden, M., Riahi, K., Smith, S., Fujimori, S., Luderer, G., Kriegler, E., et al. (2018g). input4mips.cmip6.scenariomip.iamc.iamc -remind-magpie-ssp585-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2489

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). 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

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., et al. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, 16(2-3), 147–168. https://doi.org/10.1007/s003820050010

Granier, C., Lamarque, J. F., Mieville, A., Muller, J. F., Olivier, J., Orlando, J., et al. (2005). POET, a database of surface emissions of ozone precursors. http://www.pole-ether.fr/eccad

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

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): An extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, 5(6), 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012

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 HadGEM-GC3.1. *Journal of Advances in Modeling Earth Systems*, 11, 3443–3453. https://doi.org/10.1029/2019MS001714

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., et al. (2016). Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information. *Geoscientific Model Development*, 9(7), 2415–2440. https://doi.org/10.5194/gmd-9-2415-2016

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2016). CCMI nitrogen surface fluxes in support of CMIP6-version 2.0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1125

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2018a). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp126-2-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9661

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2018b). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp245-2-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9662

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2018c). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp370-2-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9663

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2018d). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp585-2-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9664

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2019a). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp119-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9944

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2019b). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp434-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9945

Hegglin, M., Kinnison, D., & Lamarque, J.-F. (2019c). input4mips.cmip6.scenariomip.ncar.ncar-ccmi-ssp534os-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9947

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2016). CCMI ozone in support of CMIP6-version 1.0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1115

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2018a). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp126-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9541

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2018b). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp245-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9542

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2018c). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp370-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9543

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2018d). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp585-1-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9544

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2019a). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp119-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.10381

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2019b). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp434-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.10382

Hegglin, M., Kinnison, D., Lamarque, J.-F., & Plummer, D. (2019c). input4mips.cmip6.scenariomip.ureading.ureading-ccmi-ssp534os-1-1. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.10384

Hoesly, R., Smith, S., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2017a). input4mips.pnnl-jgcri.emissions.cmip. ceds-2017-08-30. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1604

Hoesly, R., Smith, S., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2017b). input4mips.pnnl-jgcri.emissions.cmip. ceds-2017-08-30-supplemental-dat a. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1605

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, 11(1), 369–408. https://doi.org/10.5194/gmd-11-369-2018

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, 109(1), 117. https://doi.org/10.1007/s10584-011-0153-2

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2017a). Harmonization of global land use scenarios (LUH2): SSP126 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1661

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2017b). Harmonization of global land use scenarios (LUH2): SSP370 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1663

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2017c). Harmonization of global land use scenarios (LUH2): SSP434 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1841

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2017d). Harmonization of global land use scenarios (LUH2): SSP585 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1662

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2018). Harmonization of global land use scenarios (LUH2): SSP245 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1886

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2019a). Harmonization of global land use scenarios (LUH2): SSP119 v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9948

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2019b). Harmonization of global land use scenarios (LUH2): SSP534os v2.1f 2015-2100. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.9949

Hurtt, G., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., et al. (2019c). input4mips.cmip6.cmip.uofmd.uofmd-landstate-high-2-1-h. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.11261

IPCC (2014). *Climate change 2013—The physical science basis*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

Jin, Z., Qiao, Y., Wang, Y., Fang, Y., & Yi, W. (2011). A new parameterization of spectral and broadband ocean surface albedo. *Optics Express*, *19*(27), 26429. https://doi.org/10.1364/OE.19.026429

Johns, T. C., Carnell, R. E., Crossley, J. F., Gregory, J. M., Mitchell, J. F. B., Senior, C. A., et al. (1997). The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. *Climate Dynamics*, *13*(2), 103–134. https://doi.org/10.1007/s003820050155

Johns, T. C., Durman, C. F., Banks, H. T., Roberts, M. J., McLaren, A. J., Ridley, J. K., et al. (2006). The New Hadley Centre Climate Model (HadGEM1): Evaluation of coupled simulations. *Journal of Climate*, *19*(7), 1327–1353. https://doi.org/10.1175/JCLI3712.1

Jones, C. D., Hughes, J. K., Bellouin, N., Hardiman, S. C., Jones, G. S., Knight, J., et al. (2011). The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geoscientific Model Development*, *4*(3), 543–570. https://doi.org/10.5194/gmd-4-543-2011

Jonsson, A. I. (2004). Doubled CO_2-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback. *Journal of Geophysical Research*, *109*, D24103. https://doi.org/10.1029/2004JD005093

Keeble, J., Braesicke, P., Abraham, N. L., Roscoe, H. K., & Pyle, J. A. (2014). The impact of polar stratospheric ozone loss on Southern Hemisphere stratospheric circulation and climate. *Atmospheric Chemistry and Physics*, *14*(24), 13,705–13,717.

Kerr, R. A. (2000). A North Atlantic climate pacemaker for the centuries. *Science*, *288*(5473), 1984–5. https://doi.org/10.1126/science.288.5473.1984

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*, 2865–2888. https://doi.org/10.1029/2018MS001370

Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., et al. (2011). An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean. *Global Biogeochemical Cycles*, *25*, GB1004. https://doi.org/10.1029/2010GB003850

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., et al. (2016). The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geoscientific Model Development*, *9*(9), 2973–2998. https://doi.org/10.5194/gmd-9-2973-2016

Loveland, T. R., Reed, B. C., Ohlen, D. O., Brown, J. F., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, *21*, 1303–1330. https://doi.org/10.1080/014311600210191

Madec, G., & Imbard, M. (1996). A global ocean mesh to overcome the North Pole singularity. *Climate Dynamics*, *12*(6), 381–388. https://doi.org/10.1007/BF00211684

Marsh, R., Ivchenko, V. O., Skliris, N., Alderson, S., Bigg, G. R., Madec, G., et al. (2015). NEMOâĂŞICB (v1.0): Interactive icebergs in the NEMO ocean model globally configured at eddy-permitting resolution. *Geoscientific Model Development*, *8*(5), 1547–1562. https://doi.org/10.5194/gmd-8-1547-2015

Martin, G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., et al. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, *4*(3), 723–757. https://doi.org/10.5194/gmd-4-723-2011

Martin, T., Park, W., & Latif, M. (2013). Multi-centennial variability controlled by Southern Ocean convection in the Kiel Climate Model. *Climate Dynamics*, *40*, 2005–2020. https://doi.org/10.1007/s00382-012-1586-7

Mathiot, P., Jenkins, A., Harris, C., & Madec, G. (2017). Explicit representation and parametrised impacts of under ice shelf seas in the coordinate ocean model NEMO 3.6. *Geoscientific Model Development*, *10*(7), 2849–2874. https://doi.org/10.5194/gmd-10-2849-2017

Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., et al. (2017). Solar forcing for CMIP6 (v3.2). *Geoscientific Model Development*, *10*(6), 2247–2302. https://doi.org/10.5194/gmd-10-2247-2017

Matthes, K., Funke, B., Kruschke, T., & Wahl, S. (2017). input4mips.solaris-heppa.solar.cmip.solaris-heppa-3-2. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1122

Meinshausen, M., & Nicholls, Z. R. J. (2018a). Uom-aim-ssp370-1-2-0: Aim-ssp370 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2342

Meinshausen, M., & Nicholls, Z. R. J. (2018b). Uom-gcam4-ssp434-1-2-0: Gcam4-ssp434 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2343

Meinshausen, M., & Nicholls, ZebedeeR. J. (2018c). Uom-image-ssp119-1-2-0: Image-ssp119 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2345

Meinshausen, M., & Nicholls, ZebedeeR. J. (2018d). Uom-image-ssp126-1-2-0: Image-ssp126 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2346

Meinshausen, M., & Nicholls, ZebedeeR. J. (2018e). Uom-message-globiom-ssp245-1-2-0: Message-globiom-ssp245 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2347

Meinshausen, M., & Nicholls, ZebedeeR. J. (2018f). Uom-remind-magpie-ssp534-over-1-2-0: Remind-magpie-ssp534-over ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2348

Meinshausen, M., & Nicholls, ZebedeeR. J. (2018g). Uom-remind-magpie-ssp585-1-2-0: Remind-magpie-ssp585 ghg concentrations. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.2349

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., et al. (2019). The SSP greenhouse gas concentrations and their extensions to 2500. Geoscientific Model Development Discussions. https://doi.org/10.5194/gmd-2019-222

Meinshausen, M., & Vogel, E. (2016). input4mips.uom.ghgconcentrations.cmip.uom-cmip-1-2-0. Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1118

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., et al. (2017). Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geoscientific Model Development*, *10*(5), 2057–2116. https://doi.org/10.5194/gmd-10-2057-2017

Morgenstern, O., Braesicke, P., O'Connor, F. M., Bushell, a. C., Johnson, C. E., Osprey, S. M., & Pyle, j. a. (2009). Evaluation of the new UKCA climate-composition model Part 1: The stratosphere. *Geoscientific Model Development*, *2*, 43–57. https://doi.org/10.5194/gmd-2-43-2009

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., et al. (2017). Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI). *Geoscientific Model Development*, *10*(2), 639–671. https://doi.org/10.5194/gmd-10-639-2017

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

Myhre, G., Shindell, D., Breon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, 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. 659–740). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.018

O'Connor, F. M., Johnson, C. E., Morgenstern, O., Abraham, N. L., Braesicke, P., Dalvi, M., et al. (2014). Evaluation of the new UKCA climate-composition model. Part 2: The troposphere. *Geoscientific Model Development*, *7*(1), 41–91. https://doi.org/10.5194/gmd-7-41-2014

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Oki, T., & Sud, Y. C. (2006). Design of Total Runoff Integrating Pathways (TRIP) A global river channel network. *Earth Interactions*, *2*(1), 1–37. https://doi.org/10.1175/1087-3562(1998)002<0001:dotrip>2.3.co;2

Oliver, H. J., Shin, M., & Sanders, O. (2018). Cylc: A workflow engine for cycling systems. *Journal of Open Source Software*, *3*(27), 737. https://doi.org/10.21105/joss.00737

Olivier, J., Peters, J., Granier, C., Petron, G., Muller, J. F., & Wallens, S. (2003). Present and future surface emissions of atmospheric compounds: POET report #2. EU project EVK2-1999-00011.

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

Pongratz, J., Dolman, H., Don, A., Erb, K.-H., Fuchs, R., Herold, M., et al. (2018). Models meet data: Challenges and opportunities in implementing land management in Earth system models. *Global Change Biology*, *24*(4), 1470–1487. https://doi.org/10.1111/gcb.13988

Power, S., Casey, T., Folland, C., Colman, A., & Mehta, V. (1999). Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, *15*(5), 319–324. https://doi.org/10.1007/s003820050284

Schmidt, A., Mills, M. J., Ghan, S., Gregory, J. M., Allan, R. P., Andrews, T., et al. (2018). Volcanic radiative forcing from 1979 to 2015. *Journal of Geophysical Research: Atmospheres*, *123*, 412–491. https://doi.org/10.1029/2018JD028776

Sellar, A. 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*, 4513–4558. https://doi.org/10.1029/2019MS001739

Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, *505*(7481), 37–42. https://doi.org/10.1038/nature12829

Shindell, D. T., Schmidt, G. A., Miller, R. L., & Mann, M. E. (2003). Volcanic and solar forcing of climate change during the preindustrial era. *Journal of Climate*, *16*(24), 4094–4107. https://doi.org/10.1175/1520-0442(2003)016%3C4094:VASFOC%3E2.0.CO;2

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., et al. (2014). Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years. *Atmospheric Chemistry and Physics*, *14*, 9317–934. https://doi.org/10.5194/acp-14-9317-2014

Solomon, S., Kinnison, D., Bandoro, J., & Garcia, R. (2015). Simulation of polar ozone depletion: An update. *Journal of Geophysical Research: Atmospheres*, *120*, 7958–7974. https://doi.org/10.1002/2015JD023365

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*(8), 3187–3213. https://doi.org/10.5194/gmd-11-3187-2018

Taylor, K. E., Williamson, D. L., & Zwiers, F. W. (2000). The sea surface temperature and sea-ice concentration boundary conditions for AMIP II simulations (*60*). Livermore, CA, USA: Program for Climate Model Diagnosis and Intercomparison.

The Rose toolkit (2012). https://metomi.github.io/rose/doc/html/index.html

Thomason, L. W., Ernest, N., Millán, L, Rieger, L., Bourassa, A., Vernier, J.-P., et al. (2018). A global space-based stratospheric aerosol climatology: 1979–2016. *Earth System Science Data*, *10*(1), 469–492. https://doi.org/10.5194/essd-10-469-2018

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2016). Biomass burning emissions for CMIP6 (v1.2). Earth System Grid Federation. https://doi.org/10.22033/ESGF/input4MIPs.1117

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., et al. (2017). Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015). *Geoscientific Model Development*, *10*(9), 3329–3357. https://doi.org/10.5194/gmd-10-3329-2017

WMO (1992). International meteorological vocabulary (2nd ed.). Geneva: Secretariat of the World Meteorological Organization pp. 636-637.

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*(5), 1909–1963. https://doi.org/10.5194/gmd-12-1909-2019

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). 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

Yienger, J. J., & Levy, H. (1995). Empirical model of global soil-biogenic NOx emissions. *Journal of Geophysical Research*, *100*(D6), 11447. https://doi.org/10.1029/95JD00370

Yool, A., Popova, E. E., & Anderson, T. R. (2013). MEDUSA-2.0: An intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies. *Geoscientific Model Development*, *6*(5), 1767–1811. https://doi.org/10.5194/gmd-6-1767-2013

Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2019). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, *9*(1), 59–65. https://doi.org/10.1038/s41558-018-0350-3

Zhang, Y., Wallace, J. M., Battisti, D. S., Zhang, Y., Wallace, J. M., & Battisti, D. S. (1997). ENSO-like interdecadal variability: 1900-93. *Journal of Climate*, *10*(5), 1004–1020. https://doi.org/10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2