UKESM1: Description and evaluation of the UK Earth System Model

Alistair A. Sellar[1], Colin G. Jones[2,3], Jane Mulcahy[1], Yongming Tang[1], Andrew Yool[6], Andy Wiltshire[1], Fiona M. O'Connor[1], Marc Stringer[2,4], Richard Hill[1], Julien Palmieri[6], Stephanie Woodward[1], Lee de Mora[7], Till Kuhlbrodt[2,4], Steve Rumbold[2,4], Douglas I. Kelley[5], Rich Ellis[5], Colin E. Johnson[2,3], Jeremy Walton[1], Nathan Luke Abraham[2,11], Martin B. Andrews[1], Timothy Andrews[1], Alex T. Archibald[2,11], Ségolène Berthou[1], Eleanor Burke[1], Ed Blockley[1], Ken Carslaw[2,3], Mohit Dalvi[1], John Edwards[1], Gerd A. Folberth[1], Nicola Gedney[1], Paul T. Griffiths[2,11], Anna B. Harper[12], Maggie A. Hendry[1], Alan J. Hewitt[1], Ben Johnson[1], Andy Jones[1], Chris D. Jones[1], James Keeble[2,11], Spencer Liddicoat[1], Olaf Morgenstern[13], Robert J. Parker[8,9], Valeriu Predoi[2,4], Eddy Robertson[1], Antony Siahaan[10], Robin S. Smith[2,4], Ranjini Swaminathan[4,8], Matthew T. Woodhouse[14], Guang Zeng[13], Mohamed Zerroukat[1]

[1]Met Office Hadley Centre, Exeter, UK
[2]National Centre for Atmospheric Science, UK
[3]University of Leeds, Leeds, UK
[4]University of Reading, Reading, UK
[5]Centre for Ecology and Hydrology, Wallingford, UK
[6]National Oceanography Centre, Southampton, UK
[7]Plymouth Marine Laboratory, Plymouth, UK
[8]National Centre for Earth Observation, UK
[9]University of Leicester, Leicester, UK
[10]British Antarctic Survey, Cambridge, UK
[11]University of Cambridge, Cambridge, UK
[12]University of Exeter, Exeter, UK
[13]NIWA, Wellington, New Zealand
[14]CSIRO, Melbourne, Australia

Corresponding author: Alistair Sellar, `alistair.sellar@metoffice.gov.uk`

Key Points:

- UKESM1 represents a major advance over its predecessor HadGEM2-ES, both in the complexity of its components and its internal coupling

- The complex coupling presents challenges to the model development; we document the tuning process employed to obtain acceptable performance

- UKESM1 performs well, having a stable pre-industrial state and showing good agreement with observations in a wide variety of contexts

Abstract

We document the development of the first version of the United Kingdom Earth System Model UKESM1. The model represents a major advance on its predecessor HadGEM2-ES, with enhancements to all component models and new feedback mechanisms. These include: a new core physical model with a well-resolved stratosphere; terrestrial biogeochemistry with coupled carbon and nitrogen cycles and enhanced land management; tropospheric-stratospheric chemistry allowing the holistic simulation of radiative forcing from ozone, methane and nitrous oxide; two-moment, five-species, modal aerosol; and ocean biogeochemistry with two-way coupling to the carbon cycle and atmospheric aerosols. The complexity of coupling between the ocean, land and atmosphere physical climate and biogeochemical cycles in UKESM1 is unprecedented for an Earth system model. We describe in detail the process by which the coupled model was developed and tuned to achieve acceptable performance in key physical and Earth system quantities, and discuss the challenges involved in mitigating biases in a model with complex connections between its components. Overall the model performs well, with a stable pre-industrial state, and good agreement with observations in the latter period of its historical simulations. However, global mean surface temperature exhibits stronger-than-observed cooling from 1950 to 1970, followed by rapid warming from 1980 to 2014. Metrics from idealised simulations show a high climate sensitivity relative to previous generations of models: equilibrium climate sensitivity (ECS) is 5.4 K, transient climate response (TCR) ranges from 2.68 K to 2.85 K, and transient climate response to cumulative emissions (TCRE) is 2.49 K/TtC to 2.66 K/TtC.

Plain Language Summary

We describe the development and behaviour of UKESM1, a novel climate model that includes improved representations of processes in the atmosphere, ocean and on land. These processes are inter-related: for example, dust is produced on the land and blown up into the atmosphere where it affects the amount of sunlight falling on the Earth. Dust can also be dissolved in the ocean, where it affects marine life. This in turn changes both the amount of carbon dioxide absorbed by the ocean and the material emitted from the surface into the atmosphere, which has an affect on the formation of clouds. UKESM1 includes many processes and interactions such as these, giving it a high level of complexity. Ensuring realistic process behaviour is a major challenge in the development of our

model, and we have carefully tested this. UKESM1 performs well, correctly exhibiting stable results from a continuous pre-industrial simulation (used to provide a reference for future experiments) and showing good agreement with observations towards the end of its historical simulations. Results for some properties—including the degree to which average surface temperature changes with increased amounts of carbon dioxide in the atmosphere—are examined in detail.

1 Introduction

Government and society require reliable guidance on how the Earth's climate will evolve in the coming century and how this can be mitigated through changes in human activity. This requires detailed understanding of how the climate will respond to future emissions of pollutants, as well as to anthropogenic land use. In addition to the physical climate response, there is a growing need to understand the response of the global biosphere and associated biogeochemical cycles, both to a changing climate and to increasing concentrations of atmospheric CO_2. Changes in the efficiency of the Earth's natural carbon cycle may provide a strong feedback onto future climate warming, reducing available carbon emissions to realise key policy targets, such as limiting global warming to less than 2C above pre-industrial levels (Friedlingstein, 2015). Warming and increased absorption of CO2 by the global oceans is expected to have negative impacts on seawater chemistry and ocean ecosystems (Gehlen et al., 2014), while warming on land may trigger vegetation change (Walther, 2010) as well as permafrost thaw and release of ancient carbon stores (Burke, Chadburn, Huntingford, & Jones, 2018; Chadburn et al., 2017). Changes in atmospheric composition and temperature have the potential to alter the rate and balance of chemical reactions with implications for stratospheric ozone recovery (Chipperfield et al., 2017) and future air quality (Silva et al., 2017). To address these challenges, the UK's Met Office and Natural Environment Research Council (NERC) have jointly developed the UK Earth System model (UKESM1) as a successor to the HadGEM2-ES model (Collins et al., 2011). UKESM1 will deliver simulations to the Coupled Model Intercomparison Project phase 6 (CMIP6; Eyring et al., 2016) and will be a community research tool for studying past and future climate.

UKESM1 uses the coupled climate model HadGEM3-GC3.1 (GC3.1 hereafter, Kuhlbrodt et al., 2018; Williams et al., 2018) as its physical core, with the following components interactively coupled to GC3.1: terrestrial carbon and nitrogen cycles, including dynamic

vegetation and representation of agricultural land use change; ocean biogeochemistry (BGC) with prognostic diatom and non-diatom concentrations and a unified troposphere-stratosphere chemistry model, tightly coupled to a multi-species modal aerosol scheme. These additional Earth system components, as well as couplings across components and coupling with the physical climate model, allow a number of important policy and science questions to be addressed that would not be possible with GC3.1. Primary amongst these are estimates of available carbon emission budgets commensurate with limiting global warming below specific targets (Matthews et al., 2017). Introducing prognostic atmospheric chemistry into GC3.1 allows an assessment of the mitigation potential from reducing short lived climate forcers such as methane, black carbon and tropospheric ozone (Collins et al., 2018; Stohl et al., 2015), while coupling atmospheric chemistry to terrestrial biogeochemistry opens up the potential to study the risk of permafrost thaw, methane release and their impacts on Arctic warming. A full atmosphere treatment of ozone chemistry supports a joined-up study of stratospheric ozone recovery and climate change as well as analysis of interactions between climate warming, tropospheric ozone and vegetation (Sadiq, Tai, Lombardozzi, & Val Martin, 2017). In the marine domain, interactive treatment of ocean heat and carbon uptake, carbonate chemistry and ocean biology enables future risks around ocean acidification and warming to be assessed (Bopp et al., 2013). Finally, feedbacks from changes in Earth system components of the model on future climate warming can also be investigated.

To provide a model capable of addressing the above questions, in developing UKESM1 we emphasised Earth system process-completeness, particularly with respect to couplings between prognostic model components, across different model domains (e.g. atmosphere, ocean, land, ice) and between different process parametrisations. We aimed to make UKESM1 internally consistent, using the model to predict processes as much as possible, rather than prescribe them from an external data source. This treatment increases the model degrees of freedom when predicting future changes in the coupled Earth system. As one example we highlight some of the interactions involved in the UKESM1 representation of dust; similar interactions exist for numerous other model processes and variables. Dust is produced from the land surface, depending on model predicted vegetation type, soil moisture and surface winds. The dust produced is then advected by the model atmosphere and interacts with the radiation parametrisation. Atmospheric dust can be deposited into the model ocean, thousands of kilometres from where it was produced, and

act as a source of soluble iron influencing ocean biological activity and thus marine up-take of CO_2. Surface chlorophyll resulting from this marine biological activity, feeds into surface emission parametrisations of dimethyl sulphide (DMS) and primary marine organic aerosol, both of which increase cloud condensation nuclei in the model atmosphere, influencing the model's radiation and cloud microphysical schemes and thus the physical climate response. Retaining such interactions in a fully prognostic sense means UKESM1 is one of the most process-complete Earth system models available today.

Developing an Earth system model that retains this degree of process realism, process couplings and prognostic treatment comes with risks. Primary among these is constraining the performance of the new Earth system components that often do not have the same degree of observational cover as more traditional physical climate processes. In addition, the increased degree of coupling between processes, as well as the preferential use of model-predicted fields as input to dependent parametrisations (e.g. predicted vegetation type, seawater DMS, atmospheric chemical oxidants, etc.) rather than prescribed, temporally constant fields, increases the risk that errors in one model component will negatively impact dependent variables or parametrisations in another part of the model, with the danger of a cascade of biased interactions driving the overall simulated climate to be significantly in error. In developing UKESM1, when a new Earth system process was included or made prognostic when it had previously been constrained by prescribed input, we carefully assessed the performance of the full model as well as the dependent model components. Our aim was to ensure new processes performed as realistically as possible, and that dependent processes as well as the full model retained an acceptable level of skill when assessed against a basket of performance indicators. If the introduction of a new Earth system process or prognostic coupling was neutral or positive for model performance then it was included. If the impact was negative, the degree of performance degradation was weighed against the potential benefit of more realistic future projections. Wherever possible we leaned towards inclusion of increased process realism or coupling.

In section 2 we describe the component models of UKESM1, and how they are coupled to one another. Section 3 discusses the methodology for tuning the fully coupled Earth system model to achieve acceptable performance. In section 4 we evaluate the model against observations and examine a few of its emergent properties, and finally in section 5 we summarise and discuss some of the challenges arising from the development of UKESM1.

Appendix C describes some non-standard configurations of the model: firstly configurations of reduced complexity intended for carbon cycle research, and secondly the configuration which runs with prescribed emissions (rather than concentrations) of CO_2. Appendix D lists the code versions and simulation identifiers for experiments used in this work.

2 Component models and coupling

UKESM1 is built upon component models which have been developed to study climate processes for specific science domains such as ocean biogeochemistry or atmospheric composition. Each of these has been evaluated and tuned in uncoupled configurations (atmosphere-only, ocean-only or offline-land), before their inclusion in the coupled Earth system model. Its physical core GC3.1 has been tested in coupled atmosphere-land-ocean-seaice mode, but without feedbacks from any of the additional Earth system components in UKESM1. The components themselves have been developed as models in their own right, with corresponding descriptions in the literature. Below we list the component models and their primary references; in appendix A we give more detailed descriptions.

- Physical core atmosphere-land-ocean-seaice model: HadGEM3-GC3.1 (Kuhlbrodt et al., 2018; Williams et al., 2018)

- Terrestrial biogeochemistry: JULES (Joint UK Land Environment Simulator; Clark et al., 2011), with developments to plant physiology and functional types (Harper et al., 2016, 2018), land use and the introduction of nitrogen cycling.

- Ocean biogeochemistry: MEDUSA (Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification; Yool, Popova, & Anderson, 2013), with the addition of a diagnostic submodel for DMS surface seawater concentration (Anderson et al., 2001) to support interactive DMS emissions to the atmosphere.

- Atmospheric composition: UKCA (United Kingdom Chemistry and Aerosol model; Archibald et al., 2019; Mulcahy et al., 2018) with unified stratospheric-tropospheric chemistry and close coupling between chemistry and aerosols.

UKESM1 is a successor to HadGEM2-ES (Collins et al., 2011), and many of its components, namely the land surface, atmospheric chemistry and the core physical model, have been developed from that model's components. The most significant developments for century-scale climate projections are (a) the inclusion of interactive nitrogen cycling

Table 1.   Coupled interactions between Earth system components

Components	Coupling	Reason for inclusion
Land-ocean-atmosphere	Carbon pools coupled to atmospheric CO_2 concentration and climate	Simulate changes in natural carbon sinks and sources
Land	Carbon cycle coupled to nitrogen availability	Improve carbon uptake process representation and enable future coupling to atmospheric nitrogen chemistry
Chemistry-radiation	Radiation uses 3D concentrations of N_2O^*, CH_4 and O_3 from chemistry	Improve representation of GHG radiative forcing and simulate radiative response to changes in chemistry
Chemistry-water vapour	Two-way coupling between water vapour and chemistry*	Improve representation of H_2O sources and simulate radiative response to changes in chemistry
Aerosol-radiation-cloud	Prognostic aerosol size distribution is used by radiation and cloud microphysics*	Improve representation of aerosol radiative forcing
Chemistry-aerosol	Oxidation processes for secondary organic* and sulphate aerosol coupled to chemistry	Improve representation of aerosol radiative forcing
Clouds-chemistry	Photolytic reaction rates depend on 3D radiative intensity*	Simulate chemical response to cloud and ozone changes
Vegetation-chemistry-aerosol	Interactive emissions of biogenic volatile organic compounds*	Simulate chemical and aerosol response to changes in vegetation productivity and distribution
Vegetation-aerosol-ocean	Dust emissions are coupled to simulated bare soil fraction	Include dependence of airborne dust on vegetation distribution
Aerosol-ocean	Deposition of iron in mineral dust	Simulate biological response to dust production and transport
Ocean-aerosol	Emissions of PMOA* & DMS depend on ocean BGC	Changes in ocean biology can impact aerosol-cloud feedbacks

*Indicates an enhancement in capability relative to HadGEM2-ES.

in the terrestrial carbon cycle, which addresses a common bias across CMIP5 models to simulate excessive land carbon uptake, and (b) interactive stratospheric chemistry, which enables us to predict the radiative forcing of ozone, methane and nitrous oxide in the full model domain, and allows stratospheric ozone recovery to evolve with climate change. Additionally, aerosol is represented by prognostic mass and number, allowing better representation of aerosol size distribution, and hence cloud-aerosol forcing.

Of prime importance for an Earth system model (ESM) is the coupling between its components, since these couplings determine which feedbacks the model is capable of simulating, and hence the potential applications of the model. Table 1 lists the couplings which are active in the default configuration of UKESM1, with a summary of the rationale for their inclusion. UKESM1 uses the OASIS3-MCT coupler (Craig, Valcke, & Coquart, 2017) as the interface between marine and atmospheric components. Coupling information is exchanged between the ocean and atmosphere every three hours. Between the land and atmosphere the coupling frequency is every dynamical model timestep (20 minutes). Couplings involving aerosols, chemistry or radiation are performed once per hour. Ocean physics, sea-ice and BGC are coupled every ocean model timestep (45 minutes).

3 Performance tuning of final coupled model

In the final stages of development of any coupled climate model or ESM, there is a phase of fine-tuning model parameters or settings in order to minimise model biases and satisfy high-level performance metrics such as global TOA radiation balance, water and energy conservation. For discussion of these issues and a summary of common practices, see e.g. Hourdin et al. (2016); Mauritsen et al. (2012); Schmidt et al. (2017). This process is generally more complex for an ESM because of the large number of interactions between component models, and the concomitant potential for biases in one component to degrade the performance of other parts of the coupled system. In this section we summarise all of the parameters tuned in the development of the fully coupled UKESM1, and the rationale for that tuning. Note that prior to coupled testing, the component models were developed and tuned in un-coupled mode and this tuning is described in the documentation papers for those components (Harper et al., 2016, 2018; Morgenstern et al., 2009; Mulcahy et al., 2018; O'Connor et al., 2014; Walters et al., 2019; Williams et al., 2018; Yool et al., 2013). The parameter values from that component model tun-

–9–

230 ing acted as the starting point for the tuning of UKESM1; in the case of physical model

231 parameters the starting point was the values used in GC3.1 (the physical core of UKESM1).

232   The goals of the UKESM1 tuning were twofold: firstly to keep long-term global-

233 mean carbon fluxes and net TOA radiation close to zero under pre-industrial forcing, to

234 avoid significant drifts in either the climate or carbon pools; and secondly to avoid large

235 biases between observations and the model state. Despite the majority of reference ob-

236 servations being taken in the last 40 years, most of the UKESM1 coupled model tuning

237 was performed using pre-industrial tests with perpetual 1850 forcing, during the model's

238 spin-up phase. Of course, present-day tests would allow a cleaner comparison to obser-

239 vations but such tests are complicated by the fact that present-day climate and the car-

240 bon cycle are not in equilibrium with their forcing, and by the difficulty of choosing ini-

241 tial conditions. The only truly reliable way to test the present-day impact of changes to

242 ESM components is to allow the change to spin up in a pre-industrial configuration and

243 then run transient historical tests through to the present day. However, this is prohibitively

244 slow to perform iteratively and instead we made pragmatic judgements to account for

245 expected differences between the pre-industrial and present-day model state.

246   A wide range of model outputs were used to inform the tuning process. These in-

247 cluded metrics (single numbers such as global mean TOA radiation or global total plant

248 productivity) as well as maps of key model variables. In appendix B, we list the main

249 metrics and variables which were used during the tuning process to identify priority bi-

250 ases and monitor the stability of the pre-industrial state. While these summarise the pri-

251 mary tuning targets, other variables were examined from time to time, particularly when

252 a particular bias or process needed checked in more detail.

253   Now that we have completed pre-industrial spin-up and historical simulations with

254 the tuned model, it is possible to assess the accuracy of these judgements with hindsight.

255 To this end we have performed two "untuned" experiments parallel to the pre-industrial

256 and historical simulations described in section 4. The untuned configuration is identi-

257 cal to the final model except that the parameters described below have been reset to their

258 original pre-tuned values. The pre-industrial untuned run branched from the beginning

259 of the UKESM1 pre-industrial control (i.e. from the end of the spinup described in sec-

260 tion 4), while the untuned historical run branched from the year 1990 of a single mem-

261 ber of the UKESM1 historical ensemble. The pre-industrial control itself Table 2 presents

a set of metrics summarising the biases which were tackled by this tuning exercise. The first five years have been discarded to exclude initial spin-up effects, and there is minimal drift in these metrics during the 20-year period over which they are calculated. In practice the tuning was guided by a wide array of information including the pattern of biases across multiple quantities as described above; the metrics in Table 2 serve only to illustrate the *impact* of the tuning at a headline level.

3.1 Snow-vegetation interactions

The most significant tuning of the coupled ESM concerned the interaction between shortwave (SW) radiation, snow and vegetation. We found it necessary to tune the parameters governing this interaction in order to reduce a substantial SW radiation bias. The extent to which vegetation is covered by snow makes an enormous difference to surface albedo, and there is a large diversity in how models parametrise this process (Boisier et al., 2012; Bright, Myhre, Astrup, Antón-Fernández, & Strømman, 2015; Ménard, Essery, Pomeroy, Marsh, & Clark, 2014). JULES represents the bending and partial burying of vegetation by snow (Ménard et al., 2014) on a per-PFT basis, by reducing the LAI used in the albedo calculation as snow depth increases (Walters et al., 2019). When the depth of snow (d) is less than the canopy height h_i of a given PFT, the partial burying is described by the equation

$$L_i' = L_i \left(1 - \frac{d}{h_i} \right)^{n_i} \tag{1}$$

where i denotes the PFT, L_i is the true LAI for each PFT and L_i' is the "exposed" LAI used in the albedo calculation. The parameter n_i thus governs the rate at which vegetation is covered by snow as snow depth increases, but it is not well constrained by observations and must be tuned for a given LAI distribution. The LAI in GC3.1 is prescribed using a seasonal climatology derived from MODIS satellite observations, while the phenology in UKESM1 is simulated interactively using the scheme of Clark et al. (2011). Figure 1 shows LAI for C3 grass in the two models in JJA and DJF. GC3.1 shows an almost complete loss of LAI in winter in central Asia and N America. This seasonal reduction is more extreme than seen in other observational products (Garrigues et al., 2008), and may in part result from the obscuring of vegetation by snow as seen by the satellite data which is used in constructing the LAI climatology. Even the needleleaf tree PFT in this climatology sees LAI reductions of 80-95% in regions dominated by evergreen needleleaf trees (not shown). Therefore, in GC3.1 there is little need for the parametrisation

figs/figure1-lai_drop_c3grass.pdf

Figure 1. Seasonal changes in leaf area index. Left: prescribed LAI in GC3.1, right: simulated LAI in UKESM1 1995-2014 (mean of 9 historical simulations), top row: June-August, middle row: December-February, bottom row: JJA minus DJF expressed as a percentage of the JJA value.

in equation 1, since LAI is already very small, and n_i is accordingly tuned to very low values.

In contrast, the LAI in UKESM1 is more similar between summer and winter for grasses and evergreen PFTs, with reductions of 20-35 %. This wintertime LAI may be too high because UKESM1 does not fully represent the phenology of grasses, but the large difference from the LAI climatology used in GC3.1 requires a different tuning for the snow-vegetation parametrisation. Accordingly n_i is given higher values in UKESM1, particularly for grasses, so that the radiative effect of snow partially covering vegetation is taken into account. The specific values of n_i were chosen to maximise the similarity in grid-box mean wintertime surface albedo between UKESM1 and GC3.1, to avoid large differences in regional mean radiative fluxes. Using the GC3.1 values of n_i in UKESM1 produces a widespread negative bias in DJF clear-sky outgoing SW across northern mid-latitude land masses (Figure 2c). The tuning of n_i increases the wintertime albedo in seasonally snow-covered regions, increasing the TOA outgoing clear-sky SW (Figure 2b), which reduces the negative bias but does introduce positive bias in some regions (Figure 2d). The net impact of this tuning can be summarised by the second metric in Table 2, defined as TOA outgoing clear-sky SW mean over land from 30°N to 60°N for DJF, representing the clear-sky radiative impact of snow on TOA radiation. A crude observational range of 50.5 - 53.6 $\mathrm{W\,m^{-2}}$ has been derived by calculating the same metric for two satellite datasets: CERES-EBAF Ed4.0 (Loeb et al., 2018) and ISCCP-FD (Y. Zhang, Rossow, Lacis, Oinas, & Mishchenko, 2004), using the 2000-2010 period common to both datasets. Table 2 shows that the impact of the tuning was to increase this metric from well below the observational range to just above it.

figs/figure2-0388_Clear-sky_outgoing_SW_TOA_djf_v_ebaf.png

Figure 2. Clear-sky outgoing SW $(\mathrm{W\,m^{-2}})$ during DJF from UKESM1 historical simulations (1995-2015). a) Tuned UKESM1 model climatology, b) UKESM1 minus parallel simulation with original GC3.1 values of n_i, c) bias of untuned simulation relative to CERES-EBAF, d) bias of tuned UKESM1 model relative to CERES-EBAF.

3.2 Bare soil and dust emissions

In UKESM1, emissions of mineral dust are tightly coupled to the fraction of bare soil in a grid box, and hence are very sensitive to biases in vegetation fraction in arid and semi-arid areas. In particular the model, prior to tuning, was prone to excessive emissions of dust in Australia, India and central Asia. We tuned bare soil fraction in these regions by increasing the vegetation dynamics parameter lai_{min} (Clark et al., 2011) for grass PFTs. When LAI for a given PFT exceeds this lai_{min}, some excess productivity is allocated to horizontal spread of the PFT within the grid box. For the same productivity, smaller values of lai_{min} allow greater vegetation fractions and hence smaller bare soil area. Two satellite land cover datasets were used as a reference for this tuning: European Space Agency (ESA) Climate Change Initiative Land Cover data (CCI-LC; Poulter et al., 2015), and that from the International Geosphere-Biosphere Programme Land Use and Cover Change project (IGBP-LUCC; Loveland et al., 2000). The impact of increasing lai_{min} is illustrated by considering the total area of Australian bare soil in the model, which is reduced by nearly a factor of two to $2.8 \times 10^6 \, \mathrm{km}^2$, covering around one third of the continent. The resulting bare soil area lies between the equivalent values for IGBP-LUCC and CCI-LC of $2.5 \times 10^6 \, \mathrm{km}^2$ and $4.0 \times 10^6 \, \mathrm{km}^2$ respectively. Bare soil areas in central Asia were similarly reduced but the same is not true for India where the model suffers from a deficit in monsoon rainfall (Williams et al., 2018), which leads to such low plant productivity that the lai_{min} parameter has little or no impact.

Even with reduced biases in vegetation cover, there was also a need to tune the parameters of the dust emission scheme itself. These have been described by Woodward (2011), and in summary they control the emission response to soil moisture and friction velocity, and include an overall scaling factor for the total emission. These were tuned to maximise agreement with in-situ surface concentration measurements from the University of Miami network (18 sites), and Aerosol Robotic Network (Holben et al., 2001, AERONET) total aerosol optical depth (AOD) measurements at eight dust-dominated sites. While regional metrics are used in the tuning process (see section 4.6), overall performance is summarised by a single global dust error metric, defined as the RMS difference between $\sqrt{model}$ and $\sqrt{obs}$ for all of these University of Miami and AERONET observations. The square root is used as a weighting function to increase the contribution of lower concentrations at remote sites to the global metric. The calculation uses seasonal means of the model and observations; each site receives equal weight, thus the con-

centration observations contribute more strongly than AOD: 18 versus 8 sites. The final UKESM1 configuration has a global dust error of 0.47 (unit-less), which is just over half the error of the untuned configuration; the reduction comes from a combination of the changes to dust parameters, and the reduction in bare soil area described above. For context, the value of this metric is 0.51 for GC3.1 and was 0.98 for HadGEM2-ES.

3.3 Terrestrial productivity

Carbon fluxes and stores are key performance indicators for an Earth system model, due to their importance in regulating carbon feedbacks. The global total GPP of plants in UKESM1 was tuned downward to better agree with observations by reducing the quantum efficiency of photosynthesis. Quantum efficiency is specified separately for each PFT, and the UKESM1 values were adjusted to the lower end of the ranges found by the literature review of Skillman (2008). To inform this tuning we combined GPP from our pre-industrial tests with an estimate of the likely pre-industrial-to-present-day change based on previous models, and compared this against the present-day GPP estimated by the FLUXNET model tree ensembles (MTE) data product of Jung et al. (2011). The goal was to improve both the annual mean and seasonal cycle of the global total GPP. The global total GPP for the last 20 years of the resulting UKESM1 historical simulations is 130 GtC/yr, a little higher than the Jung et al. (2011) 5-95 % range of 119±6 GtC/yr.

3.4 Surface chlorophyll

The ocean-atmosphere coupling of PMOA emissions described in section A.5 makes aerosol emission dependent upon the interactively simulated surface chlorophyll concentrations, which allows the model to capture potential feedbacks between ocean biogeochemistry and climate. However, ocean models typically have larger biases in ocean chlorophyll than other marine biogeochemistry properties (Kwiatkowski et al., 2014; Yool et al., 2013), posing the risk of introducing a large bias in aerosol load. We mitigate this risk by multiplying the chlorophyll concentration by a global scaling factor. A value of 0.5 is used for this factor, chosen to produce a similar global mean outgoing SW radiation flux at TOA to that in the atmosphere-only implementation of the PMOA emission scheme, which was driven by the observation-based GlobColour dataset (Maritorena, D'Andon, Mangin, & Siegel, 2010). This scaling factor reflects the fact that MEDUSA

–15–

chlorophyll concentrations are higher than those of GlobColour. Note that this scaling of surface chlorophyll is performed only within the PMOA emission parametrisation: there is no impact on the prognostic chlorophyll simulated by the ocean model.

3.5 Energy balance

The physical core of UKESM1 was already well-tuned before adding ESM components (Kuhlbrodt et al., 2018; Williams et al., 2018), and with the exception of the snow-vegetation interaction above it was not deemed necessary to re-tune the physical model parameters. The TOA radiation balance was however altered by the addition of ESM components, through a combination of changes in cloud, surface albedo and radiatively active gases. Without further tuning the net downward radiation at TOA in 1850 was $-0.81\,\mathrm{W\,m^{-2}}$, which would have resulted in a significant downward drift in model temperatures. We brought the TOA radiation into balance by tuning parameters in the Anderson et al. (2001) parametrisation of DMS sea-water concentration to permit lower minimum values of DMS. The standard configuration of this parametrisation has a prescribed minimum value of 2.29 nM for seawater DMS concentration, while gridded observational DMS datasets, such as those of Kettle et al. (1999) or Lana et al. (2011), contain significantly smaller values than this over large regions of the ocean. Anderson et al. (2001) themselves point out that the data from which their parametrisation is derived is likely to contain a sampling bias towards higher values, so a lower minimum is reasonable. We reduced the minimum from 2.29 nM to 1.00 nM, extending the Anderson et al. (2001) linear relationship (between DMS concentration and log_{10} of chlorophyll concentration, surface SW and nutrient availability) to lower values of DMS. The ensuing reduction in DMS gave widespread decreases in cloud droplet number (and thus cloud albedo) across the Southern Ocean and stratocumulus regions, resulting in a drop in reflected SW at TOA of 2 to 5 $\mathrm{Wm^{-2}}$ over large areas of the ocean.

Using the ocean DMS parametrisation to balance the TOA, rather than for example cloud parameters, allows greater traceability between UKESM1 and its physical core GC3.1, by minimising the number of physical model parameter changes between the two models. While the goal of this tuning was to balance the pre-industrial TOA, we note that the mean TOA over the last 20 years of the tuned historical run is within the Johnson, Lyman, and Loeb (2016) present-day observational estimate of 0.61-0.81 $\mathrm{W\,m^{-2}}$ (5-95 % confidence).

–16–

Table 2. Metrics illustrating biases targeted by the coupled model tuning, showing the impact of this tuning on pre-industrial and present-day performance.

	Pre-industrial		Present-day		Present-day
	untuned	tuned	untuned	tuned	observations
Net TOA down ($\mathrm{W\,m^{-2}}$)	-0.81	-0.09	0.23	0.81	0.61 - 0.81
Snow SW metric[a] ($\mathrm{W\,m^{-2}}$)	44.4	52.5	45.1	54.3	50.5 - 53.6
Bare soil Aus. (10^6 km^2)	5.2	3.0	4.7	2.8	2.5 - 4.0
Dust error[b] (unit-less)	0.89	0.44	0.82	0.47	0
GPP (GtC/yr)	122	117	139	130	113 - 125

Note: Pre-industrial refers to years 5-25 of the pre-industrial control. Present-day refers to years 1995-2014 of the 5th ensemble member of the historical experiment. See text for observations details and section 4 for model experiment definitions. [a]TOA outgoing clear-sky SW over land, 30°N-60°N mean. [b]$\mathrm{RMS}(\sqrt{m_i} - \sqrt{o_i})$ for dust concentration and AOD observations (o_i) and their model counterparts (m_i). RMS: root-mean-square.

3.6 Effectiveness of model tuning

Despite the difficulties noted at the start of this section, the tuning exercise was successful in reducing the biases it sought to address. The metrics in table 2 which summarise these biases are all improved relative to the untuned equivalents; that is the pre-industrial net TOA is closer to zero and present day values of the other metrics are closer to their observational values. Recall that the tuning decisions were based mainly on pre-industrial tests, so the present-day performance could not be assured until historical simulations were completed. In the case of the snow-vegetation interactions affecting the outgoing SW metric, the tuning possibly went too far since the present-day value of the "snow SW" metric changed from low-biased to high-biased, though still closer to the observations. This illustrates the imprecise nature of the tuning exercise.

Finally we note that, while the parallel runs presented in Table 2 give a good indication of the the magnitude of the direct impact of tuning, they are likely to underestimate the full impact. This is because these tests are designed to be relatively short (25 years), in order to minimise confounding internal variability. While there is minimal drift after the first five years, the increased GPP and dust emissions in particular would

produce slow responses in terrestrial carbon stores and ocean biogeochemistry respectively which could take more than a century to equilibrate. Assessing the full impact of these changes would require the untuned pre-industrial configuration to be run to equilibrium, before initialising a new ensemble of historical simulations; such an exercise is too costly to perform at this time.

We have not yet assessed the impact of this tuning on emergent properties such as radiative forcing or climate sensitivity (as done for example by Mauritsen et al., 2012); this will be considered in future work.

4 Evaluation and characterisation of model behaviour

In this section we characterise the behaviour of the model, using coupled simulations that form the core set of experiments for CMIP6 (Eyring et al., 2016):

1. piControl: a pre-industrial control with 1850 forcing; over 1100 years of this simulation have been completed.

2. abrupt4xCO2: as piControl but with CO_2 concentration instantaneously quadrupled; 150 years duration

3. 1pctCO2: as piControl but with CO_2 concentration increasing at 1% per year; 150 years duration

4. historical: simulating the evolving climate since the pre-industrial era, with transient forcing from 1850 to 2014.

The pre-industrial control is initialised from the end of a 500-year coupled spin-up. The coupled spin-up was itself preceded by 5000 years of ocean-only simulation which used an asynchronous coupling procedure, and by 1000 years of land-only spinup. The spin-up methodology and evolution will be documented in a companion paper in this special issue. Historical results in the present paper make use of nine members of an initial condition ensemble, with each member initialised from a different date in the pre-industrial control. Unless otherwise stated, results from the historical simulation use a mean of these nine ensemble members.

In section 4.1 we explore the drift and variability of the pre-industrial control. Sections 4.2-4.6 evaluate the latter period of the historical simulations with reference to ob-

servations. Finally in section 4.7 we examine emergent properties of the model diagnosed from the full historical period and from the idealised CO_2 experiments.

4.1 Pre-industrial trends and variability

The pre-industrial control simulation is important for quantifying the model's natural variability and any long-term drifts in the model state. Significant drifts in climate or the carbon cycle would compromise the model's utility for many applications, particularly in assessing future Earth system responses to anthropogenic emissions. Figure 3 shows how five key properties of the model vary over the pre-industrial control. Generally drift is very small relative to internal variability.

There is a relatively constant outgassing of CO_2 from the ocean at a rate of ≈ 0.02 GtC/yr, slowing slightly as the run progresses, and a release of soil carbon of a similar magnitude. Both of these, and the total carbon uptake, are well below the ± 0.1 GtC/yr threshold set by C4MIP (C. D. Jones et al., 2016) as a guide for acceptable drift in carbon pools, shown as dotted lines in Figure 3.

Net TOA radiation is very close to zero, which ensures that there are no discernible drifts in surface temperature or sea-ice extent. TOA radiation does drift very slowly over the course of the run with a mean of 0.009 W m^{-2} over the first 500 years, and 0.038 W m^{-2} over years 500 to 1000.

Surface air temperature shows strong multi-century variability, most notably during years 400 to 900, with swings of up to 0.3 K over a 100-year period (Figure 3). This is linked to episodes of oceanic deep convection in the Southern Ocean which bring warm saline deep Atlantic water to the surface, leading to a significant reduction in sea-ice and loss of accumulated ocean heat to the model atmosphere. The warm, convecting, phases of these cycles coincide with reductions in Antarctic sea-ice cover.

Aggregated vegetation types show variability on a range of timescales. Grasses expand and retract into regions of bare soil on interannual and decadal timescales in response to local variations in rainfall and surface climate, resulting in a clear anti-correlation between grass and bare soil in Figure 3. Tree coverage evolves on a much slower timescale, with a small decrease of around 1% during the first 600 years of the run, followed by a slight recovery, mostly at the expense of shrubs.

~~figs/figure3~~timeseries_pistability_u-aw310.pdf

Figure 3. Time-series of key model properties for the first 1125 years of the pre-industrial control. From top to bottom: global mean net TOA radiation; global mean 1.5m air temperature; hemispheric sea-ice extent (area of model grid boxes with >15 % ice cover); global total area of aggregated plant functional types (note reduced scale for shrubs); cumulative carbon uptake (dashed lines are ±0.1 GtC/yr). In the upper three panels, thin lines represent annual means and thick lines a centred 21-year rolling mean. In the lower two panels, all lines represent annual means or accumulations.

The UKESM1 historical simulations are initialised using atmosphere-sea-ice-land-ocean conditions taken at different points in the pre-industrial control simulation. Ideally, these initial conditions will sample the model's multi-decadal internal variability so that each historical run is initialised using well-separated states of the pre-industrial coupled model. To achieve this, we consider two large-scale modes of sea-surface temperature (SST) multi-decadal variability, the Inter-decadal Pacific Oscillation (IPO; Power, Casey, Folland, Colman, & Mehta, 1999; Y. Zhang et al., 1997) and the Atlantic Multi-decadal Oscillation (AMO; Kerr, 2000).

The IPO pattern is the largest multi-decadal quasi-periodic SST variability over the Pacific basin. It is represented by the first area-weighted Empirical Orthogonal Function (EOF) of deseasonalised, detrended and low-pass filtered monthly SSTs. Figure 4 shows the IPO pattern for HadISST1.1 (1870-2017; Rayner et al., 2003), and for the UKESM1 pre-industrial control simulation. The IPO index is the principal component of this EOF. The AMO pattern is a quasi-periodic oscillation of North Atlantic SSTs. The AMO index is calculated as the area-weighted mean timeseries of the deseasonalised, detrended and low-pass filtered North Atlantic SSTs. Low-pass filtering the data minimises the influence of the El Nino/Southern Oscillation (ENSO). Observed and modelled AMO pattern are shown in Figure 5.

The observed and simulated patterns are similar, for both modes of variability. The most notable difference between the two IPO patterns is the westward extent of the equatorial warm region, which extends further towards the maritime continent in the simulation. The AMO patterns differ in terms of the northerly position of the warmest region. Nevertheless, the similarity of these climate mode patterns provides confidence that

–20–

Figure 4. First area-weighted EOF of Pacific SST, representing the spatial pattern of the observed IPO, scaled by the standard deviation of the IPO index. a) HadISST1.1 b) Pre-industrial control. Monthly SST and sea-ice data are used in both cases. Grid points containing sea-ice during any point in the dataset are masked prior to calculating the mode patterns.

510 the simulation is, to first order, replicating the observed multi-decadal variability in the

511 Pacific and North Atlantic basins.

512 The phase space spanned by the observed and modelled IPO and AMO indices is

513 shown in Figure 6. The observations show no clear relationship between the IPO and

514 AMO, and populate all four sectors of phase space. The model has a much longer time-

515 series, and hence explores the phase space more fully. Consistent with the observations,

516 the IPO and AMO appear to be independent of each other in the model simulation. We

517 use the simulation phase-space to choose well-separated initial conditions for each his-

518 torical run (marked in Figure 6), thus ensuring that the beginning of the historical en-

519 semble samples the multi-decadal variability of the ocean. For the same reason, we also

–21–
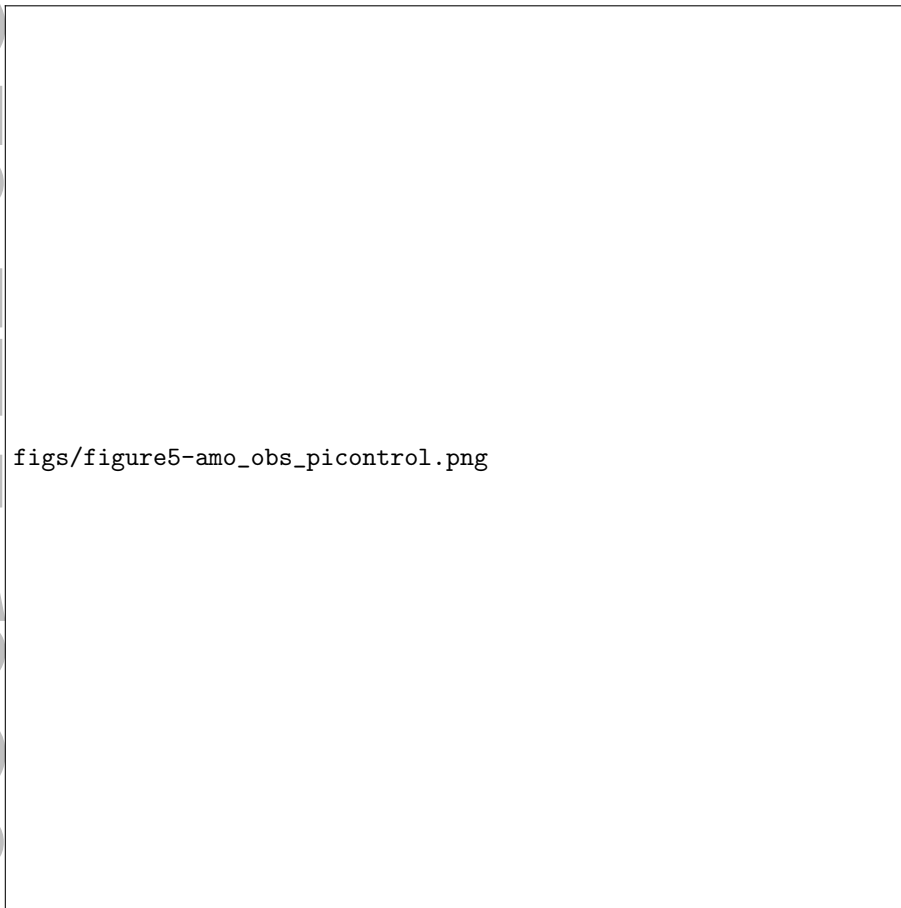
Figure 5.   AMO pattern, calculated as the North Atlantic SST correlation with the area-weighted mean North Atlantic SST time-series, scaled by the standard deviation of the AMO index. a) HadISST1.1 b) Pre-industrial control. Monthly SST and sea-ice data are used in both cases. Grid points containing sea-ice during any point in the dataset are masked prior to calculating the mode patterns.

Figure 6. The phase space of IPO versus AMO monthly indices, showing timeseries of these indices expressed as deviations from their mean value, normalised by their respective standard deviation. Left: observations (148 years); right: the first 833 years of the pre-industrial control simulation. The initial conditions chosen for each historical run are marked with a red star.

520 prescribed that initial conditions for historical members were separated by at least 30

521 years in the pre-industrial control.

522   The model also exhibits a realistic simulation of the El Nino/Southern Oscillation

523 (ENSO). The addition of ESM components has made no significant difference to the ENSO

524 behaviour of the physical model GC3.1 documented in Kuhlbrodt et al. (2018). We have

525 repeated the ENSO analysis of Kuhlbrodt et al. (2018), and the results are statistically

526 indistinguishable (not shown).

figs/figure7-sst_UKESM1r1-r9_minus_HadISST_6up.pdf

Figure 7. Seasonal mean sea surface temperature (1995-2014) for HadISST1.1 observations (Rayner et al., 2003, top), UKESM1 historical ensemble mean (middle), and UKESM1 minus HadISST (bottom). Left column: DJF; right column: JJA.

figs/figure8-tas_UKESM1r1-r9_minus_CRUTEM_6up.pdf

Figure 8. Seasonal mean air temperature at 1.5m (1995-2014) for CRUTEM4.6 observations (P. Jones et al., 2012, top), UKESM1 historical ensemble mean (middle), and UKESM1 minus CRUTEM (bottom). Left column: DJF; right column: JJA.

figs/figure9-precip_UKESM1r1-r9_minus_GPCP_6up.pdf

Figure 9. Seasonal mean precipitation (1995-2014) for GPCP2.3 (Adler et al., 2003, top), UKESM1 historical ensemble mean (middle), and UKESM1 minus GPCP (bottom). Left column: DJF; right column: JJA.
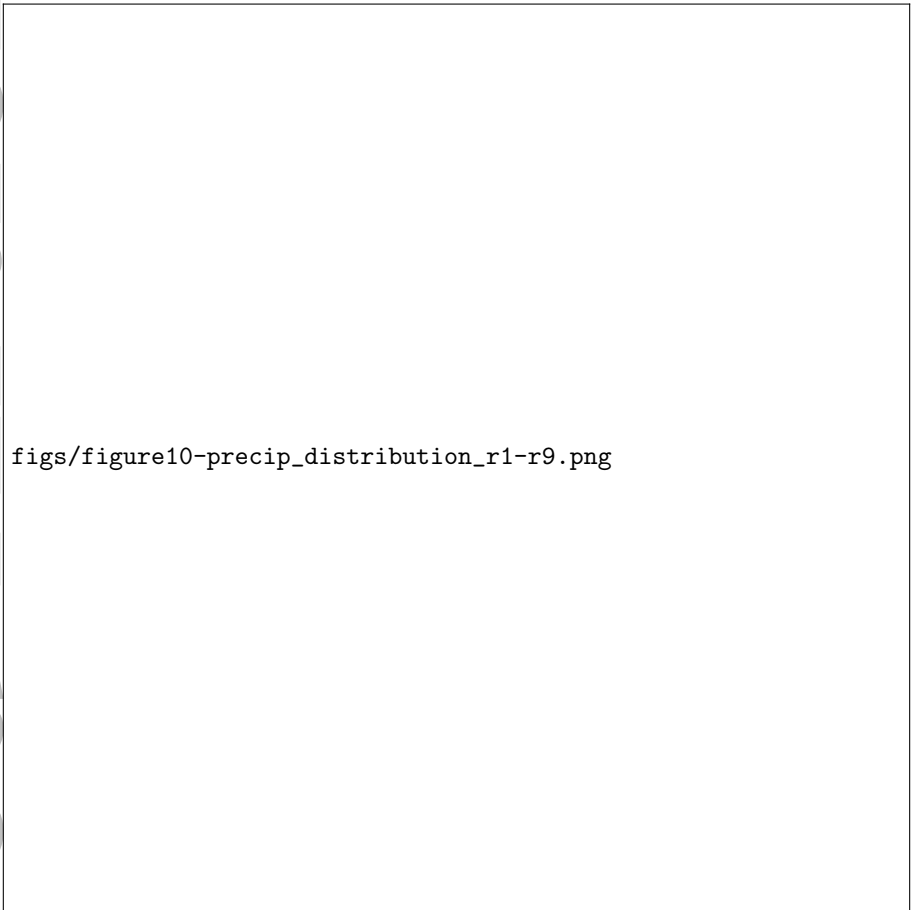
4.2 Present-day surface climate

In Figure 7 we evaluate climatological seasonal mean sea surface temperature against the HadISST1.1 observation dataset (Rayner et al., 2003). As is common in ocean models of this resolution (Flato et al., 2013), UKESM1 exhibits a strong cold bias in the NW Atlantic, related to the too-zonal representation of the Gulf Stream and N Atlantic current. Again, similar to other low-resolution models, we see a year-round cold bias in the N Pacific, and a too-westward extension of the equatorial Pacific cold tongue during JJA. Overall the pattern of SST bias is similar to that of GC3.1 (Kuhlbrodt et al., 2018), including a warm bias of $0.5\,\mathrm{K}$ to $2\,\mathrm{K}$ across much of the Southern Ocean.

Figures 8 and 9 make similar seaonal comparisons for climatological 1.5m air temperature over land and precipitation, against the observation datasets of CRUTEM4.6 (P. Jones et al., 2012) and GPCP2.3 (Adler et al., 2003), respectively. UKESM1 shows a wintertime cold bias over N America, Europe and NW Russia. One driver of this cold bias relates to the SST biases noted above, influencing land temperatures through atmospheric advection. These continental cold biases are larger than the mean upwind SST bias, suggesting a potential feedback involving snow cover, snow albedo and surface temperatures whereby a cold surface temperature bias is maintained in the model through excess snow cover, both spatially and temporally through the winter season.

In common with the CMIP5 generation of models (Flato et al., 2013), UKESM1 exhibits a double-inter tropical convergence zone in the tropical Pacific, linked to the Equatorial cold tongue bias. Tropical Atlantic precipitation is too far south, and in DJF this pattern extends westward over S America. In JJA UKESM1 exhibits the same dry Indian monsoon bias described by Williams et al. (2018) for GC3.1. This dry bias leads to a vegetation deficit (described in section 4.3) and some of the Earth system feedbacks discussed by Martin and Levine (2012).

While our emphasis in this paper is on evaluating UKESM1 at larger spatial scales (i.e. continental or ocean basin scales) and longer time scales (multi-decadal to centennial), UKESM1 will be used to make a set of future projections following the CMIP6 scenarioMIP protocol (O'Neill et al., 2016). Output from these projections will be used to drive impacts models (e.g. to assess future risks to food security or water resources). Such models require high temporal and spatial resolution surface climate data as forcing. Furthermore, assessing a model's higher time frequency statistics increases confidence in the

Figure 10. Daily precipitation contributions to mean precipitation rate for the period 1980-2014, for UKESM1 members and observational datasets, for four regions and different seasons (see labels in each panel). The extent of each region is shown in grey in the top panel: AMZ, Amazon; EUR, Europe; SEA, South-East Asia; EAS, East Asia. Precipitation rates are binned with pseudo-exponential bin sizes (see Klingaman et al., 2017) to account for more frequent rain events at low rain rates. Each bin frequency is then multiplied by the mean bin rate, which gives the precipitation contribution from each bin. The x axis is plotted with a log scale so the area under the curve is approximately the mean precipitation.

figs/figure11-historical9_uncorrected_ensemble_T1p5_seasons_24April2019.png

Figure 11. Frequency of occurrence of monthly mean near surface air temperature over the period 1961-2014 binned into $1\,°C$ wide classes for the same four regions as in Figure 10. Observations are from the CRUTS4.02 data set (thick black line) and 9 UKESM1 historical ensemble members are shown (thin coloured lines). Only grid points that include a fraction of land greater than 0.995 are included in the spatial means.

559 simulated long-term means, variability and future changes, all of which are based on the

560 underlying high time frequency behaviour. As an initial evaluation of UKESM1, Figure

561 10 shows simulated and observed daily timescale precipitation for four distinct climatic

562 regions and Figure 11 shows a similar analysis for monthly mean near-surface air tem-

563 perature.

564 For each of the regions, Figure 10 plots the fractional contribution from different

565 daily precipitation rates to the total seasonal accumulation, calculated for the seasons

566 specified in the figure labels. Observations include the GPCC/HOAPS v1 global daily

567 precipitation (Andersson et al., 2010; Schamm et al., 2014), based on both gauge and

568 satellite observations, the Global Precipitation Climatology Project (GPCPv1.1, Adler

569 et al., 2003) and the Tropical Rainfall Measuring Mission 3B42v6 product (TRMM, Huff-

570 man et al., 2007). All observational data were first re-gridded onto the UKESM1 grid.

571 TRMM data is only used over the SEA and AMZ regions, whereas GPCC is not used

572 over the predominantly ocean area of SEA.

573 Over EUR we show winter (DJF, full lines) and summer (JJA, dashed lines) pre-

574 cipitation, with the nine UKESM1 members in black and the observations in different

575 shades of blue. In both seasons the distributions are skewed towards a long tail of light

576 precipitation ($<1\,\mathrm{mm\,day^{-1}}$) that contributes to the seasonal accumulation. UKESM1

577 captures this skewness quite well. In DJF the model slightly overestimates precipitation

578 contributions from moderate to heavy daily rates (3 to $10\,\mathrm{mm\,day^{-1}}$), but does capture

579 the heavy rain tail of the distribution. Conversely, in JJA moderate to heavy rainfall is

580 underestimated by the model while very heavy events ($>20\,\mathrm{mm\,day^{-1}}$) contribute too

581 much to the seasonal accumulation. As the moderate to heavy daily rates contribute more

582 to the seasonal accumulation (i.e. they occur more frequently) than the very heavy rates,

583 the underestimate of the former results in a modest dry bias over EUR in JJA, primar-

584 ily in central and south-east Europe (Figure 9).

585 Turning to the EAS region, where precipitation is dominated by the summer mon-

586 soon, we plot distributions for the summer (wet) season (JJAS) and the winter (dry) sea-

587 son (NDJFM). In JJAS UKESM1 significantly overestimates the contribution of rain-

588 fall events $>30\,\mathrm{mm\,day^{-1}}$, with some unrealistic daily rates in excess of $100\,\mathrm{mm\,day^{-1}}$.

589 Figure 9 indicates that this results in a significant wet bias, primarily over the ocean part

590 of EAS (i.e. the west Pacific warm pool region). In NDJFM seasonal precipitation is sig-

–28–

nificantly lower than JJAS in both observations and the model, although the model systematically overestimates precipitation from most of the bins. Again Figure 9 suggests this overestimate is mainly along the coastal ocean in EAS. To the south of the EAS region, over the maritime continent and west Pacific warm pool (SEA) the rainfall distribution in the main wet months (NDJFM) is very well captured, with only a modest overestimate in the model across most bins. Finally, the AMZ region shows precipitation over the Amazon rainforest. While the overall spread of the observed distribution is captured by the model, there is a significant underestimate of moderate to heavy rainfall days (approx. 8 to 30 mm day^{-1}) that leads to the seasonal mean dry bias seen over the central Amazon basin in Figure 9, which is surrounded to the south by a wet bias of equivalent magnitude.

In Figure 11 for the same four regions we compare spatially averaged 1.5m monthly mean air temperatures between UKESM1 and CRUTS4.02 observations (Harris, Jones, Osborn, & Lister, 2014). From this we wish to see if UKESM1 captures the spread of cold/warm months around its climatological mean value, providing an evaluation of the inter-annual variability of near surface climate in the model. For all regions we first interpolate the 0.5° CRUTS data onto the UKESM1 grid and then mask out ocean and coastal points, using the UKESM1 land-sea mask and a criterion that a grid point has to have fractional land amount > 0.995 to be included in the spatial average.

Considering EUR, in winter (DJF) a systematic bias is evident in the model distribution, with mode values shifted relative to CRUTS by 3-4C. Nevertheless, the spread of cold/warm months around the mode of the distribution, as well as the shape (skewness), are well captured in the model. Figure 8 indicates the bulk of the cold bias in EUR is over Scandinavia and European Russia, as noted above. In the summer (JJA) a smaller cold bias remains in the model distribution, which continues to be located in the northern part of EUR. A similar cold bias is seen over EAS during boreal winter, while the summer distribution is well simulated, albeit with a slight tendency for too frequent cold months (surface temperatures <18 °C) not seen in CRUTS. Over AMZ, the mode of the UKESM1 distribution is quite accurate, although the model now has a bias towards extreme warm months (mean temperatures >27 °C) not seen in CRUTS. These are likely associated with the dry precipitation bias over the region (Figure 9), shifting the surface energy balance from predominantly evaporative towards a larger sensible heat component. Finally, the SEA region shows a cold bias in the model distribution. It should

–29–

Figure 12. Area covered by snow greater than a threshold depth of snow, expressed as snow water equivalent (SWE). Thresholds are 2 mm, 10 mm, 40 mm and 100 mm. Model: UKESM1 historical ensemble mean 1990-2009 (red line) and ensemble spread (red shading). Observations: GlobSnow (Takala et al., 2011) monthly mean SWE (black) and monthly maximum SWE (grey). Model fields are masked where observations have missing data.

be noted for SEA that due to the small number of land points used in the spatial means and the extreme topography of islands in this region, neither the model nor CRUTS data are likely a robust representation of true land temperatures in this region. We plan to extend our analysis of UKESM1 near surface temperatures in a future study, considering daily temperature variability, as well as daily maximum and minimum values to better isolate model processes controlling biases in regional surface temperature.

Figure 12 compares the areal extent of snow with the GlobSnow observation product (Takala et al., 2011) for four different snow depths. In addition to the area for observed monthly mean snow depths, the equivalent for monthly maximum depth is shown

Figure 13. Comparison of permafrost extent in the model against the dataset of Brown et al. (2002). Purple regions show permafrost diagnosed from the model: defined as grid points for which the maximum monthly-mean soil temperature at 2m depth is below 0 °C for the period 1990-2009. Grey regions are where modelled annual mean air temperature is below 0 °C. The green and orange lines are, respectively, the limits of continuous and discontinuous permafrost from Brown et al. (2002).

633   as an indicator of inter-annual variability. The model has too much very thin snow compared with the observations, but a comparable amount of deeper snow. The deeper snow tends to thaw slightly later than the observations in the spring, but increase at a similar rate to the observations in the autumn.

637   Figure 13 compares permafrost extent diagnosed from the model against the dataset of Brown et al. (2002). The model has permafrost in the majority of the regions where the Brown et al. (2002) dataset suggest there is continuous permafrost. There is permafrost

–31–

Figure 14. UKESM1 historical ensemble mean 1990-2009 Arctic (top) and Antarctic (bottom) sea ice concentration (shading) for September (left) and March (right). Areas with sea ice less than 15% are masked out. Also shown (orange lines) are corresponding 15% sea ice concentration contours from the HadISST.2.2.0.0 dataset (Titchner & Rayner, 2014). Inlaid boxes show the observed and modelled sea ice and the corresponding Integrated Ice Edge Error in units of 10^6 km^2 (IIEE; Goessling et al., 2016).

640 for approximately 47% of the land surface where the air temperature is less than zero.

641 This falls within the range of observations which is 40 % to 72 %.

642 Figure 14 compares September and March sea ice concentration/extent, in both

643 hemispheres, against the HadISST.2.2.0.0 dataset of Titchner and Rayner (2014) for the

644 period 1990-2009. Also shown are the modelled and observed sea ice extent (the total

645 area of ocean with at least 15% ice cover) and the corresponding Integrated Ice Edge Er-

646 ror (IIEE). The IIEE metric, introduced by Goessling et al. (2016), is a measure of the

647 error in ice edge placement, and is essentially the area integral of all grid cells where the

model and observations disagree about whether sea ice concentration is above 15% (see Goessling et al., 2016, for further details).

In general there is very good agreement between the modelled (shading) and the observed (orange lines) location of the ice edge as depicted by the 15% concentration contour. In the Arctic the model sea ice is, overall, a little more extensive than the observations, particularly so in the north Pacific (Bering Sea and Sea of Okhotsk) in winter (March). The model also has too much winter sea ice north of Iceland and does not reproduce the Oden ice tongue, which is captured in the HadISST2.2 dataset. In the summer the Arctic sea ice is slightly over extensive within the Arctic Ocean basin, except for in the Kara Sea where the sea ice extent is much smaller than observations. In the Antarctic the sea ice is slightly less extensive than observed, particularly in the Ross Sea and east of the Weddell Sea sector south of Africa.

In response to the cool northern hemisphere surface temperature reported above and in section 4.7, the present-day sea ice is much thicker than observational estimates. Mean September sea ice thickness for 1990-2009 is over 4.5m throughout the central Arctic (not shown). However, despite the high modelled thicknesses, the spatial extent of sea ice in UKESM1 agrees well with the HadISST observational dataset.

4.3 Terrestrial biogeochemistry

To have confidence in the processes responsible for terrestrial carbon uptake, one needs to evaluate model simulated vegetation structure, surface carbon fluxes and carbon stores. If a model performed well for one or two of these properties, but not all three, it would suggest that the model was getting the right answer for the wrong reasons. In this section we evaluate and characterise the behaviour of each of these three aspects.

UKESM1 represents vegetation structure with nine natural PFTs and four crop/pasture PFTs, and the fractional coverage of each PFT within a gridbox is determined by a competition hierarchy. We evaluate the vegetation distribution using two complementary approaches: firstly with spatial maps of aggregated PFTs (Figure 15), and secondly by aggregating fractions within biomes (Figure 16). The biomes are regions of the world defined by Olson et al. (2006) based on their common ecosystem characteristics. Within each of these regions Figure 16 shows the fractional coverage of aggregated PFTs for UKESM1 and the IGBP-LUCC and CCI-LC land cover datasets; the same region definition is used

–33–

figs/figure15-veg_fractions_map.png

Figure 15.   UKESM1 aggregated PFT fractions compared to the IGBP and CCI land cover observation-based datasets. Model data is for the year 2005 to coincide with the validity time of the CCI-LC data, using the mean vegetation fraction of the historical ensemble.

figs/figure16-veg_fractions_biome.png

Figure 16. Fractional coverage of aggregated PFTs within seven "biome" regions defined by Olson et al. (2006). Each biome shows three columns: UKESM1 (historical ensemble mean for the year 2005), IGBP land cover and CCI land cover.

for all three datasets. Broadly the model does a very good job in representing vegetation structure of the biomes. Notable biases includes an excess of C3 grass in tundra regions which the observations indicate should contain more bare ground; this can be seen in the northern Siberian C3 grass fraction in Figure 15. Some of this bias is a side effect of the lai_{min} tuning described in section 3, which improved the bare soil fraction in dust-producing regions but has allowed grass to spread too extensively in northern Russia and Canada. In contrast, Saharan bare soil extends too far south, with a deficit of grassland in the Sahel. Similarly there is excess bare soil in eastern India. Both of these biases result from precipitation deficits in these regions, associated with errors in the position and intensity of monsoon rains (see Williams et al., 2018). There is a small overestimation of tree fraction in savanna biomes, most notably on the southern edge of the Amazon forest. This is due to the lack of fire disturbance, the inclusion of which would be expected to improve vegetation structure in these regions (Burton et al., 2019).

Figure 17 shows the evolution of terrestrial carbon fluxes in the ensemble mean of the historical simulations. From 1850 to 1980, decadal-mean GPP slowly increases from 118 Pg/yr to 122 Pg/yr, after which it increases rapidly to 134 Pg/yr in the last decade of the experiment. GPP in UKESM1 is slightly higher than observational estimates, which are around 120 Pg/yr in the last few decades (Jung et al., 2011). Net ecosystem productivity (NEP), the balance of GPP and respiration, also increases more rapidly after 1980. NEP is balanced by a small land-use change emission flux and a much larger crop harvest flux; the sum of these three fluxes is the net biome productivity (NBP), the total carbon flux from the atmosphere to the terrestrial biosphere. The land is a net source of carbon prior to 1980: cumulative NBP is -40 Pg from 1850 to 1979. After 1980 the land becomes a sink of atmospheric CO_2: cumulative NBP between 1980-2015 is 22 Pg.

Generally the model has an excellent zonal distribution of vegetation carbon (Figure 18), lying between the two observational estimates. The exception is a high-bias between 25°S and 10°S, reflecting the excess of trees in savanna biomes noted above. Global total vegetation carbon is 551 Pg in year 2000 and 583 Pg in 1850 which is consistent with observation-based estimates (450-650 Pg in 1750; Ciais, Willem, Friedlingstein, & Munhoven, 2013). 72% of vegetation carbon is stored in tropical forests and 26% in extra-tropical forests, with the remaining 2% being stored in shrubs and grasses.

figs/figure17-carbon_fluxes_r1-r9.png

Figure 17. Global total carbon fluxes. Gross primary productivity (GPP), net primary productivity (NPP), heterotrophic respiration (Rh), net ecosystem productivity (NEP), landuse emissions (fLuc), crop harvest carbon flux (fHarvest) and net biome productivity (NBP). A 10-year running mean has been applied to all data.

figs/figure18-carbon_soil_veg_zontot.pdf

Figure 18. Zonal mean carbon stores for model (black, 1995-2014) and observation datasets (colours). Left: Vegetation carbon, observations: GEOCARBON (Avitabile et al., 2016); Saatchi et al. (2011). Right: Soil carbon, observations: WISE30sec (Batjes, 2016); IGBP-DIS (Global Soil Data Task Group, 2002); Carvalhais et al. (2014, thin green lines show ensemble inter-quartile range).

710   At 1780 Gt, the UKESM1 global total soil carbon is less than that estimated by

711 the Carvalhais et al. (2014) data set for the whole soil profile (2450 GtC) and slightly

712 less than that estimated by Batjes (2016) for the top 2 m (1970 GtC). The model rep-

713 resents the top 3 m of soil. There is not enough soil carbon simulated for the northern

714 high latitudes (Figure 18); this is partly because turnover times are too fast (not shown)

715 but also because there is not yet a representation of permafrost carbon included within

716 UKESM1. The simulated peak of soil carbon in the tropics is slightly further south than

717 in any of the observational data sets.

4.4 Ocean biogeochemistry

719   Ocean biochemistry in UKESM1 plays a fundamental role in the carbon cycle, act-

720 ing as a source or sink of atmospheric CO_2, and in the aerosol simulation via natural emis-

721 sions of DMS and PMOA. In this section we focus on these two roles, evaluating the present-

722 day surface flux of CO_2 and the productivity which underpins the aerosol emissions.

723   Figure 19 shows simulated net primary production, averaged over the period 2000-

724 2009, compared with three satellite-derived observational estimates for the same period

725 (Behrenfeld & Falkowski, 1997; Carr et al., 2006; Westberry, Behrenfeld, Siegel, & Boss,

726 2008, these three estimates are averaged here to a single field, as per Yool et al., 2013).

727 In general, the model simulates the main geographical patterns of high productivity in

728 upwelling regions, low production in subtropical oligotrophic areas, and seasonally-high

729 productivity in high latitude areas. However, modelled productivity is more tightly fo-

730 cused on the equator, especially in the Pacific Ocean. Oligotrophic areas exhibit lower

731 production in the model, in part due to unrepresented mesoscale processes, and high lat-

732 itude productivity is elevated above that observed, especially in the Southern Ocean.

733   Figure 20 shows simulated net air-sea CO_2 flux, averaged for the period 2000-2009,

734 and compared with the observational estimate of Rödenbeck et al. (2013). The model

735 captures the main characteristics of the broad pattern of CO_2, with outgassing (blue)

736 in the equatorial and tropical band, especially in the Pacific Ocean, and ingassing (red)

737 at higher latitudes, in particular the North Atlantic. The model does show some discrep-

738 ancies, for instance excessive outgassing in the upwelling areas associated with the Hum-

739 boldt and Benguela Currents, and excess ingassing close to Antarctica linked to a deficit

740 in sea-ice cover in spring and summer.

Figure 19. Mean ocean net primary production for 2000-2009. Left: an average of three satellite-derived observational estimates (see text). Right: UKESM1 historical ensemble mean.

figs/figure20-co2_airsea_flux.png
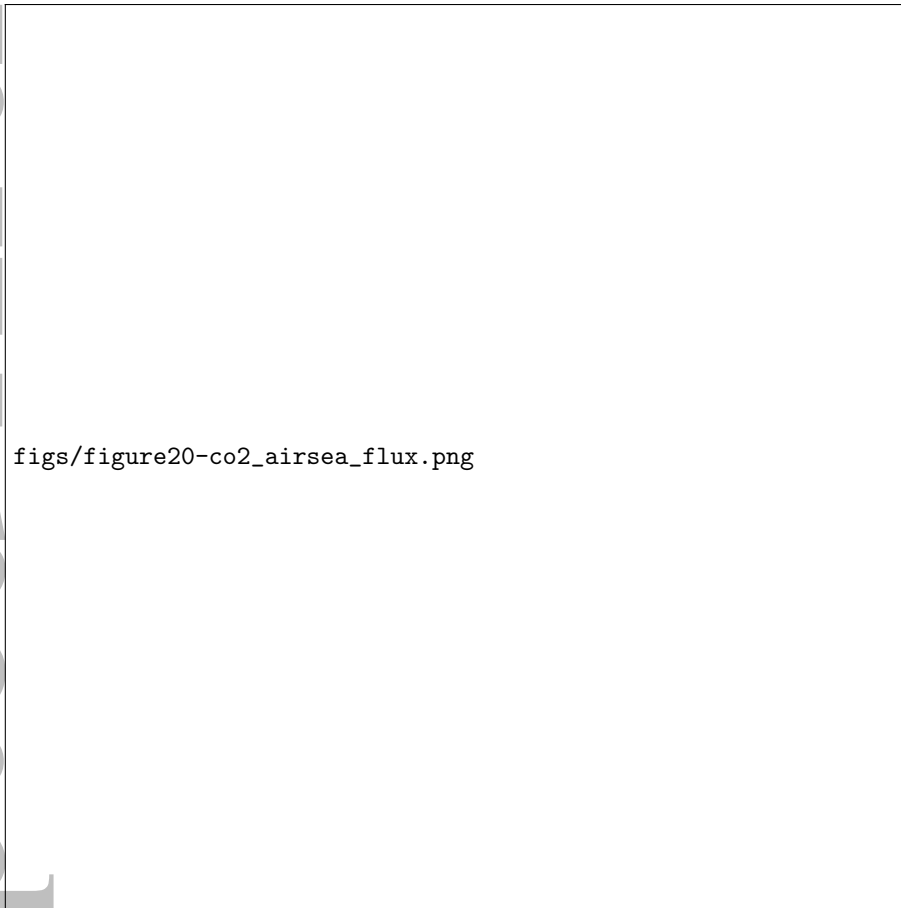
Figure 20. Net air-to-sea flux of CO_2, 2000-2009 mean. Left: Rödenbeck et al. (2013) observation-based estimate. Right: UKESM1 historical ensemble mean.

figs/figure21-ozone_col.png

Figure 21. Left: Zonal- and multi-annual mean total column ozone (Dobson Units, DU), averaged over two historical simulations, for 1995-2014. Right: Corresponding model bias versus the NIWA-Bodeker Scientific TCO database version 3.4 (updated from Bodeker et al., 2005) for the same years.

figs/figure22-ozone_spole_timeseries.jpg

Figure 22. Monthly-mean total column ozone (DU) at the South Pole, for the spring and summer months. Colours: Two UKESM1 historical simulations. (thin lines: full variability; thick lines: smoothed with a 7-year boxcar filter). Symbols: Monthly-means derived from the South Pole Dobson measurements starting in 1964 (WOUDC, 2014). All monthly means are shown that are derived from measurements taken both before and after mid-month (hence March and September are excluded).

4.5 Atmospheric chemistry

A thorough evaluation of the chemistry performance of UKESM1 can be found in Archibald et al. (2019); here we focus on those aspects which represent major enhancements relative to HadGEM2-ES, namely stratospheric chemistry, interactive photolysis and biogenic emissions of VOCs.

First we assess the quality of the simulation of total-column ozone (TCO), which is dominated by the stratospheric ozone layer, versus the NIWA-Bodeker Scientific TCO database (Bodeker, Shiona, & Eskes, 2005, version3.4) covering 1979-2016. We also use the TCO record measured at South Pole using a Dobson photospectrometer (Rosen, Kjome, & Oltmans, 1993; WOUDC, 2014).
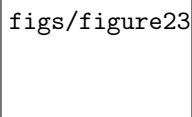
figs/figure23-ozone_col_trend.png

Figure 23. (left) Trends in zonal-mean total column ozone (DU/year) averaged over two historical UKESM1 simulations, for 1979-1997 (19 years). Stippling indicates non-significance at 95% confidence. (right) Bias in the trend versus the NIWA-Bodeker Scientific climatology for the same years.

751   Figure 21 shows the zonal-mean TCO averaged over 1995-2014, averaged over two
752 historical UKESM1 simulations, and the difference versus the climatology. In the extra-
753 polar regions, there is a general overestimation of ozone of around 20 to 40 Dobson Units
754 (DU), whereas in the Antarctic, the model underestimates ozone relative the climatol-
755 ogy by up to -60 DU in November and December. Since satellite-based observation of
756 TCO is particularly difficult at high latitudes, we compare this result directly to the Dob-
757 son long-term measurements at South Pole (Figure 22). The comparison shows a good
758 agreement with the Dobson record in October, but in November and December the model
759 only tracks the most stable, deepest ozone holes and does not reproduce the observed
760 interannual variability. Linear trends during the period of increasing ozone-depleting sub-
761 stances (1979-1997) mirror this finding (Figure 23). In the tropics the model does not
762 produce significant trends, in agreement with the climatology, but in both polar regions
763 the model produces too large negative trends during this period during spring and sum-
764 mer in both hemispheres. A separate analysis of trends during the ozone recovery pe-
765 riod (1998-2014) shows generally non-significant trends, in agreement with the observa-
766 tions.

767   An analysis of the vertical distribution of ozone and its trends (not shown) reveals
768 a generally good correspondence with observations but too strong a downdraught of meso-
769 sphere air during polar winter which contributes to low TCO during spring, and also the
770 general overestimation of ozone in the extra-polar ozone layer discussed above.

771   The good correspondence in October suggests that the sensitivity of ozone to an-
772 thropogenic chemical ozone depletion is generally correct, broadly in agreement with sci-
773 entific consensus on the underlying gas-phase chemistry (World Meteorological Organ-

⁷⁷⁴ ization, 2018), but the underestimation of variability in November and December sug-

⁷⁷⁵ gest insufficient, and insufficiently variable, meridional heat transport towards both poles

⁷⁷⁶ which would serve to inject variability into TCO and would shorten, on average, the life-

⁷⁷⁷ time of the polar vortex (Garfinkel et al., 2013). This is consistent with a stratospheric

⁷⁷⁸ cold bias of 5-8 K over the Antarctic and a polar night jet that is too strong by more

⁷⁷⁹ than 10 m s^{-1} (not shown).

⁷⁸⁰ The general overestimation of ozone in the extra-polar regions would be partly ad-

⁷⁸¹ dressed by a more comprehensive inclusion of heterogeneous reactions on sulphate aerosols

⁷⁸² involving bromine compounds (Dennison et al., 2019), which is not part of UKESM1.

⁷⁸³ The nine CMIP5 models that did simulate stratospheric ozone interactively exhibited

⁷⁸⁴ substantial biases (Eyring et al., 2013). Relative to this previous generation of models,

⁷⁸⁵ UKESM1 compares well to the reference climatology.

⁷⁸⁶ Turning now to the troposphere, Figure 24 shows that for the historical simulation

⁷⁸⁷ of UKESM1, the burden of ozone in the troposphere (defined as in Tilmes et al. (2016))

⁷⁸⁸ has increased steadily over the period 1850-2014. The burden of ozone has recently been

⁷⁸⁹ evaluated for the present day (2014-2016) by Gaudel et al. (2018) as 300 ± 12 Tg from

⁷⁹⁰ 60°S-60°N. The UKESM1 tropospheric ozone burden, subset over the same latitude range

⁷⁹¹ for the year 2014 is 314 Tg, in excellent agreement with the observations. UKESM1 sim-

⁷⁹² ulates an increase over the historical period (1850-2014) of more than 60 Tg of ozone in

⁷⁹³ the troposphere, an increase of approximately 23 % from a pre-industrial baseline of 260 ± 4 Tg

⁷⁹⁴ (1850-1875) to 320 ± 7 Tg (1995-2010).

⁷⁹⁵ Methane lifetime is a core metric for atmospheric chemistry models due to the role

⁷⁹⁶ methane plays in multiple chemical cycles, with impacts on ozone, aerosols (via oxidant

⁷⁹⁷ concentrations), as well as exerting a strong direct radiative effect in its own right. The

⁷⁹⁸ methane lifetime in UKESM1 is 8.4 years (calculated as total atmospheric burden di-

⁷⁹⁹ vided by chemical loss and deposition for the period 1995-2010). This lies within the 9.1 ± 0.9 year

⁸⁰⁰ range estimated by Prather, Holmes, and Hsu (2012) based on constraints from methyl

⁸⁰¹ chloroform, and the 7.6 to 14 year range of SPARC (2013). Table 3 summarises methane

⁸⁰² lifetime along with five other metrics of the global chemical state of the model.

⁸⁰³ Figure 25 shows the global distribution of isoprene and monoterpene emissions sim-

⁸⁰⁴ ulated by the interactive BVOC scheme in UKESM1. The emissions in this figure rep-

⁸⁰⁵ resent the mean over the last ten years (2005-2014) of the historical run. Annual global

–43–

Figure 24. Time series of the evolution of the tropospheric ozone burden in 6 members of the historical ensemble of the UKESM1 simulations (coloured lines). The black line shows the ensemble mean. Tropospheric ozone burden is calculated as the total mass of ozone between the surface and the tropopause, using a tropopause based on monthly-mean ozone mixing ratio of 150 ppb.

Table 3. Metrics summarising the state of atmospheric chemistry in the latter part of the historical period (1995-2010).

Metric	UKESM1
CH_4 lifetime (whole atm)	8.4 years
OH mean concentration	1.25×10^6 molecules/cm^3
OH ratio NH:SH	1.3
CO burden	304 Tg
CH_4 burden (trop.)	4134 Tg
O_3 burden (trop.)	320 Tg

figs/figure25-TotGlobAnnMeanTerpIsopEms.png

Figure 25. Annual mean emissions (TgC/yr) of monoterpene (top) and isoprene (bottom) from the UKESM1 historical simulations (2005-2014).

Figure 26. Annual mean aerosol optical depth at 440 nm from UKESM1 ensemble mean 1992-2008. Left: Model field (contours) overlaid with circles depicting the observed AOD at 67 ground-based AERONET stations. Right: Scatterplot of AERONET versus UKESM1 at the AERONET station locations. AERONET = Aerosol Robotic Network; AOD = Aerosol optical depth; RMSE = Root-mean-square error.

total isoprene and monoterpene emissions amount to approximately 520 TgC/yr and 120 TgC/yr, respectively. The spatial pattern of emissions is broadly consistent with state-of-the-art BVOC emission models (e.g. Arneth, Monson, Schurgers, Niinemets, & Palmer, 2008; A. B. Guenther et al., 2012; Messina et al., 2016; Müller et al., 2008; Sindelarova et al., 2014; Stavrakou et al., 2009; Wiedinmyer, Tie, Guenther, Neilson, & Granier, 2006; Young, Arneth, Schurgers, Zeng, & Pyle, 2009).
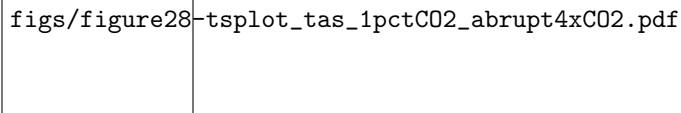
figs/figure27-dust__scatter_ensmean_r1-9_1995-2014.png

Figure 27. Scatterplot of 1995-2014 mean seasonal mean near-surface dust concentration and AODs at dust-dominated stations from the UKESM1 historical simulations versus data from the University of Miami network and AERONET stations. Colours indicate season: black DJF; blue MAM; green JJA and red SON. Symbols indicate areas for concentrations: asterisks Atlantic; squares N Pacific; triangles S Pacific; diamonds Southern Ocean; and pluses AOD. The model:obs ratio meaned over stations and seasons for each area are: Atlantic 1.10; N Pacific 2.07; S Pacific 2.03; Southern Ocean 0.48; AOD 0.61.

4.6 Aerosols

Figure 26 shows the UKESM1 ensemble annual mean AOD at 440nm for the period 1992 to 2008. Annual mean observations of AOD covering the same period from the ground-based, global AERONET (Holben et al., 2001) are overlaid on the simulated global AOD map. Overall, the model captures the spatial distribution of AOD well. This is reflected in the good correlation found in the point-wise comparison shown in Figure 26 ($r^2 = 0.63$). AOD over continental polluted regions and tropical biomass regions is generally well-represented in the model, including the east-west contrast in continental North America. There are fewer observations in remote marine regions but where observations exist the model is able to capture the low AOD values in the Pacific Ocean and downwind of anthropogenic pollution source regions in the Atlantic Ocean.

–47–

figs/figure28-tsplot_tas_1pctCO2_abrupt4xCO2.pdf

Figure 28. Change in global mean surface air temperature in the abrupt4xCO2 (red) and 1pctCO2 (blue) experiments. The parallel portion of the pre-industrial control has been subtracted from each experiment to remove any background temperature drift.

The model AOD in Figure 26 is biased low at sites close to dust sources in West Africa and the tropical Atlantic. This is examined in more detail in Figure 27 which compares the model against seasonal mean AERONET AOD at dust-dominated sites and near-surface dust concentrations from the University of Miami observing network. The AODs in this figure are consistent with the annual mean result, with seasonal mean AOD lower than the observations by 39% on average. We note that these stations are all at similar latitudes in a region strongly affected by the West African Monsoon, which is challenging to simulate, and small biases in wind-speed or precipitation in this area would strongly affect local dust production. Thus the bias in AODs may not be representative of the wider dust simulation. Indeed, dust concentrations over the Atlantic are generally well-simulated (Figure 27). Further afield, there is some high bias in concentration over the Pacific and low bias over the Southern Ocean.

4.7 Emergent behaviour of the coupled model

We consider some emergent properties of UKESM1 using metrics of global climate sensitivity and carbon uptake. Firstly, equilibrium climate sensitivity (ECS) estimates the equilibrium response of global mean surface air temperature to a doubling of CO_2. While this cannot be directly evaluated against observations, it is a useful metric for comparing the response of models to CO_2 perturbations across multiple generations of models. Figure 28 shows the temporal evolution of surface air temperature in response to an abrupt CO_2 quadrupling (the CMIP6 abrupt4xCO2 experiment; Eyring et al., 2016). We calculate effective climate sensitivity as in Gregory et al. (2004) by regressing the change in TOA net radiation against surface temperature over the first 150 years of the run. The

845 projected equilibrium temperature change from the abrupt4xCO2 experiment is divided

846 by 2 to correspond to a CO_2 doubling. The resulting ECS for UKESM1 is 5.4 K.

847 Transient climate response (TCR) is calculated from the CMIP6 1pctCO2 exper-

848 iment in which CO_2 is increased at 1 % per year. TCR is defined as the temperature change

849 from the pre-industrial control after CO_2 has doubled, averaged over years 61-80. A four-

850 member initial condition ensemble for UKESM1 (Figure 28) have TCR in the range 2.68 K

851 to 2.85 K.

852 Transient climate response to cumulative emission (TCRE; Gillett, Arora, Matthews,

853 & Allen, 2013) extends the TCR concept to use carbon feedbacks from an ESM in or-

854 der to project the warming due to a cumulative CO_2 emission of 1 TtC. Following Gillett

855 et al. (2013) we calculate TCRE as TCR (as defined above) divided by the cumulative

856 compatible anthropogenic emission up to year 70 of the same 1pctCO2 experiment. CO_2

857 concentration is prescribed in these experiments, but the anthropogenic emission com-

858 patible with the model's carbon uptake can be calculated as the change in the total mass

859 of carbon in the ocean, land and atmosphere. To remove the (very small) background

860 drift in ocean and land carbon stores (Figure 3) we calculate the compatible emission

861 as the difference between the 1pctCO2 store at year 70 and that of pre-industrial con-

862 trol in the same year, yielding a range of compatible emissions from 1.07 TtC to 1.08 TtC

863 and TCRE from 2.49 K/TtC to 2.66 K/TtC.

864 The UKESM1 values of ECS, TCR and TCRE are all higher than those of CMIP5

865 models (respectively, 2.1 K to 4.7 K, 1.0 K to 2.6 K, and 0.8 K/TtC to 2.4 K/TtC An-

866 drews, Gregory, Webb, & Taylor, 2012; Gillett et al., 2013). Elsewhere in this special is-

867 sue, Bodas-Salcedo et al. (2019) analyse the increase in atmospheric climate feedbacks

868 in HadGEM3-GC3.1 relative to the previous version of HadGEM3, whose ECS (3.2 K;

869 Senior et al., 2016) is within the range of CMIP5 models. They find that the feedbacks

870 have become more positive as a result of improvements to cloud microphysics and cloud-

871 aerosol interactions. Both of these developments reduce errors in the present-day sim-

872 ulation of clouds relative to observations. These stronger feedbacks contribute to increases

873 in all three of these metrics; TCRE is additionally increased relative to CMIP5 models

874 by the inclusion of nitrogen limitation which reduces the terrestrial carbon sink for a given

875 CO_2 atmospheric concentration and hence increases the CO_2 concentration and corre-

876 sponding warming compatible with a 1 TtC cumulative emission. More detailed anal-

Table 4. Cumulative 1850-2014 carbon uptake and land use emissions for the UKESM1 historical ensemble mean and estimated by GCB.

Cumulative flux (GtC)	UKESM1	GCB
Ocean sink	135	150 ± 20
Land sink	158	185 ± 50
Land use emissions	152	195 ± 75

Note: UKESM1 land use emissions in this table include only the CO_2 flux from above-ground biomass; GCB land use emission estimate includes both above- and below-ground biomass.

ysis of the climate sensitivity and carbon feedbacks in UKESM1 will be reported in subsequent papers.

Carbon uptake can also be diagnosed from historical, rather than idealised, simulations and thus compared against estimates rooted in historical observations. In Table 4 we compare cumulative 1850-2014 carbon uptake and land-use emissions against observation-based estimates from the Global Carbon Budget (GCB, Le Quéré et al., 2018). The ocean and land sinks are both within the GCB range. UKESM1 land use emissions in Table 4 represent the carbon flux due to clearance of above-ground biomass, including crop harvest and deforestation. The model's total emission due to land use includes an additional contribution from decomposition of below-ground biomass which we have not diagnosed. The total land-use emissions will hence will be closer to, or above, the GCB central estimate.

Surface temperature is one of the few variables for which reliable observations cover the full period of the 1850-2014 historical simulation, which allows us to evaluate the model's first-order climate response to the evolving forcing over this period. The UKESM1 global mean surface temperature anomaly in the historical ensemble shown in Figure 29, alongside the HadCRUT4 observation dataset (Morice, Kennedy, Rayner, & Jones, 2012). The observations represent only a single realisation of the internal variability of the climate system, so one should not expect the model ensemble to be centred on the observations, but rather that the range of observational uncertainty overlaps the ensemble range (under the assumption that the model ensemble is large enough to sample the relevant in-

figs/figure29-tsmean_global_nhem_shem_ukesm1.pdf
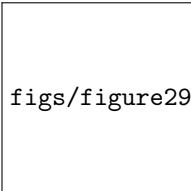
Figure 29. Global and hemispheric mean surface temperature anomaly, relative to 1851-1880 mean. Thin blue lines are ensemble members of the historical experiment; the thick blue line is their ensemble mean. The HadCRUT4.6.0.0 median and 5-95% uncertainty range are shown in black and grey, respectively. For consistency with the observations, model fields are constructed using a combination of 1.5 m air temperature over land points and SST at ocean points, and masked to match the temporally evolving coverage of HadCRUT. Model timeseries are de-trended using the linear trend of the pre-industrial control over the equivalent period.

ternal variability). Most ensemble members begin to warm in the early 20th century, then cool strongly between 1950 and 1970 before warming rapidly through to the end of the simulation. The observations show a limited cooling of 0.1 K to 0.2 K during 1940 to 1970, but the model ensemble mean cools by nearly 0.4 K over the same period. All ensemble members also show a stronger cooling response to the large volcanic eruptions of 1883, 1963 and 1991 than is seen in the observations.

Figure 29 also separates the mean surface temperature into northern and southern hemisphere timeseries. The stronger-than-observed cooling is restricted to the northern hemisphere which, together with its temporal evolution, points to either aerosol or land use forcing as the prime driver. Further investigation into this discrepancy will be the subject of future work. In the southern hemisphere the model ensemble overlaps with the observational uncertainty for the entire duration of the experiment, with the exception of the dip following the Mt. Pinatubo eruption in 1991 noted above.

5 Discussion and conclusions

In this paper we describe and evaluate the new Earth system model UKESM1, which has been jointly developed by NERC and the Met Office. UKESM1 represents a major advance on its predecessor model HadGEM2-ES, both in the sophistication of its component models and in the complexity of the couplings between these components. The most important advances are:

–51–

1. inclusion of nitrogen cycling within the terrestrial carbon cycle, and increased ocean biological complexity, in order to better predict allowable anthropogenic emissions;

2. major advances in the treatment of land management, with the separation of crop and pasture areas, and the introduction of a harvest carbon flux;

3. a well-resolved stratosphere with interactive stratospheric chemistry to assess the impact of changes in ozone depleting substances;

4. a state-of-the-art aerosol scheme with prognostic size distribution for a better understanding of aerosol radiative forcing;

5. coupling between the ocean biogeochemistry and atmospheric composition via interactive emissions of primary marine organic aerosol;

6. coupling between the terrestrial biosphere and atmospheric composition in the form of interactive BVOC emissions from vegetation;

7. coupling between atmospheric chemistry and physics through water vapour changes arising from chemical reactions.

Achieving acceptable performance in the presence of this additional model complexity posed major challenges during the development and tuning of UKESM1. The greatest difficulties arose where one component model was sensitive to biases in the variables it received from another component. One example of this is the dependence of dust emissions on bare soil fraction, and the sensitivity of bare soil to the model's climate in semi-arid regions. Through careful tuning of vegetation and dust parameters we were able to significantly reduce both bare soil and dust errors significantly. However, in some regions (particularly western India) the climate biases were such that we could not obtain a good simulation of vegetation. In this respect we still have not answered some of the difficult questions posed by Martin and Levine (2012) regarding coupling and component model biases in Earth system models, though we have made significant steps forward.

Generally, the model performs very well. The pre-industrial control is stable, with acceptably small drifts in carbon pools and virtually no drift in surface climate over 1100 years. The Southern Ocean exhibits large multi-century oscillations in surface temperature and sea-ice extent. Further work is in progress to understand the mechanisms behind this variability, but there is a growing body of evidence which indicates that such oscillations may be a feature of the natural state of the Earth system (Kurtakoti et al.,

2018; Reintges, Martin, Latif, & Park, 2017; L. Zhang, Delworth, Cooke, & Yang, 2018, e.g.).

Evaluation of the latter portion of the historical experiments against observations is generally positive. Particular highlights are: vegetation structure, which has reduced or eliminated biases in the Boreal forests and Australia that were seen in the predecessor model HadGEM2-ES (Collins et al., 2011); and stratospheric ozone, which exhibits smaller biases than the previous generation of coupled chemistry-climate models with interactive stratospheric chemistry (Eyring et al., 2013).

The most significant deficiency in the model's performance is its stronger-than-observed global mean surface cooling in the mid-20th century, followed by stronger-than-observed warming after 1990. The cooling is most likely related to aerosols and/or land use change, suggesting that the combined radiative forcing of these processes is excessive during this period. While these processes will not have a first-order impact on century-scale projections at a global scale (which are dominated by greenhouse gas forcing), understanding and improving this bias will be a priority for future configurations of UKESM1. The strong warming towards the end of the century may be due to the reduction of aerosol emissions in Europe and N America, potentially adding weight to the hypothesis that aerosol forcing is too strong in this model prior to 1990. Or it could be a result of the model's high climate sensitivity (discussed below), or a combination of both. There are large uncertainties in the magnitude of the true historical forcing due to aerosols and land-use change, and these results present an opportunity to better understand and constrain the relevant processes, with the potential to improve ESMs more generally.

In the introduction to this special issue, Senior et al. (2019) outline the development strategy for UKESM1 and its physical core HadGEM3-GC3.1, which included a step (before model finalisation) of estimating effective radiative forcing (ERF) due to key anthropogenic forcing agents (GHGs, aerosols, land use change), in order to avoid unrealistic historical performance at a global mean scale. See Mulcahy et al. (2018) in this special issue for action which was taken on aerosol forcing as a result of this strategy. These ERF experiments assessed only the 1850-2000 forcing, but as Figure 29 demonstrates, it is possible to have good agreement in global mean surface temperature towards the end of the historical period (implying that the integrated historical forcing is accurate, perhaps through a compensation of errors) and still have significant deviations prior

–53–

to this. This brings interesting new evidence to bear on the questions about model development strategy which have been discussed by e.g. Hourdin et al. (2016) and Schmidt et al. (2017), and will prompt an evolution of the strategy summarised in Senior et al. (2019).

The high sensitivity of the model to CO_2 concentrations and emissions, as evidenced by the high values of ECS, TCR and TCRE will also be a focus of further study. Bodas-Salcedo et al. (2019) show that some of the changes responsible for the high sensitivity represent significant improvements in process evaluation, but better observations and more in-depth analysis are required to assess all the processes involved. Note that these climate sensitivity and carbon feedback metrics are emergent properties of the model, and were not subject to tuning during the development process.

Near- to medium-term plans for model development in the UKESM project include new model configurations with additional capabilities. First will come a model with interactive ice sheets on Greenland and Antarctica, with two-way coupling between the ice sheets and both the atmosphere and ocean. Following that, we are developing a hybrid-resolution functionality to allow the most expensive model components (e.g. atmospheric composition) to run at a reduced horizontal resolution relative to the model physics. These ambitious developments will be reported in future papers.

Other priorities for further work include evaluation of the interactions between component models. The parametrisations used in coupling schemes (such as those for emissions or deposition of gases and aerosols) are generally based on good physical understanding and empirical relationships derived from in-situ observation, but it is hard to evaluate these relationships at the scale of a model gridbox, let alone regional or global scales. For example, is the response of the model's bare soil fraction to its climate more or less sensitive than that of the real world? Such questions have an important bearing not only for obtaining a good present-day simulation, but also for interpreting its response in model projections.

The complexity of UKESM1, both in terms of its component models and its internal coupling, is unprecedented for an Earth system model. The sophisticated interactions were included to facilitate the exploration of future feedbacks, but increase the difficulty of obtaining skilful simulation of observed period. In the face of this challenge, the generally successful performance of the model is a major accomplishment.

A Component model details

In this appendix we summarise the component models of UKESM1. We do not include a detailed description of each component; rather we refer the reader to previously published descriptions of each component and focus instead on developments which have been made since these publications, and on summarising the key model elements from an Earth system perspective.

A.1 Underlying physical model

The components of an ESM are dependent on the core physical model for accurate simulation of winds and ocean currents (which advect the biogeochemical tracers), atmospheric moist and turbulent processes (which control chemical reaction rates and aerosol-cloud interactions) and temperature, precipitation and soil moisture (upon which aerosol and gas removal processes, vegetation and CO_2 uptake depend). Biases in the physical model which are tolerable in their own right can impact disproportionately on Earth system components (e.g. Hardiman et al., 2015; Martin & Levine, 2012). Simulated ocean-land surface temperature and surface and top-of-atmosphere (TOA) radiative fluxes retain their primary importance, even for the ESM components, because of their role in feedbacks.

The Earth system components of UKESM1 are built upon the state-of-the-art physical climate model HadGEM3-GC3.1 (Kuhlbrodt et al., 2018; Williams et al., 2018). The ocean component of GC3.1 (Storkey et al., 2018) uses the NEMO dynamical ocean on a nominally 1° tripolar grid with 75 vertical levels and an explicit nonlinear free surface. The sea-ice (Ridley et al., 2018) is modelled by CICE on the same horizontal grid as NEMO, with 5 ice thickness categories, multi-layer thermodynamics with four layers of ice and one of snow, and elastic-viscous-plastic rheology. The atmosphere (Walters et al., 2019) uses the Met Office Unified Model (UM) with 85 vertical levels on terrain-following hybrid height coordinate, with an 85 km model top. The land surface is modelled using JULES (Best et al., 2011). Land and atmosphere share the same horizontal grid: a regular latitude-longitude grid with 1.25° x 1.875° resolution. Surface and sub-surface runoff is transported to the ocean using the TRIP river routing scheme (Total Runoff Integrating Pathways; Oki & Sud, 2006). Comparing UKESM1 to its predecessor HadGEM2-ES, the ocean and sea-ice models have been completely replaced, and the atmosphere

has a new dynamical core (Wood et al., 2014), a well resolved stratosphere, major advances in cloud physics including prognostic cloud, condensate and rain (Wilson et al., 2008), and a new modal aerosol scheme (Mann et al., 2010). The coupled physical model as a whole is evaluated in Kuhlbrodt et al. (2018) and has been shown to represent a substantial improvement upon its predecessors.

We believe that we have achieved our goal of keeping the physical core of UKESM1 as close to GC3.1 as possible. The differences are restricted to the following components:

1. Enhancements to aerosol emissions to improve their process realism (adding couplings between the biosphere and atmospheric composition which were previously missing), described in section A.5;

2. Tuning parameters which control dust emissions and snow-vegetation interactions, the motivation for which is explained in section 3;

3. A change to the soil hydrology scheme to address a problem with soil moisture. GC3.1, as in recent versions of HadGEM3, uses an adaptation of the van Genuchten (1980) empirical equations for soil hydraulic conductivity properties (Wiltshire et al., 2019). During early tests of UKESM1, it was discovered that the implementation of these equations in JULES produced unrealistic soil moisture profiles in the Earth system model. In some regions the soil moisture profile was inverted relative to observed profiles, with less moisture in the deepest soil layer (1-3 m) than the surface layer (0-10 cm). Such a lack of moisture at depth would limit the availability of moisture to vegetation during dry periods, and had the potential to distort the dynamic simulation of vegetation distribution. To remedy this, UKESM1 instead uses equations based on the work of Brooks and Corey (1964) with consistent ancillaries files for soil properties, as employed in HadGEM2. This results in much higher soil hydraulic conductivity and thus more moisture in the deeper soil layers. The change of scheme makes little difference to the surface energy budget. Work is in progress to refine the van Genuchten implementation for future versions of HadGEM3 in order to remedy this issue;

4. The land surface characteristics of the Greenland and Antarctic ice sheets are represented using the sub-gridscale elevation class tiling scheme and albedo and snow pack physics described in Shannon et al. (2019); we do not however use the glacier-specific tunings or precipitation and surface wind downscaling from that work. In
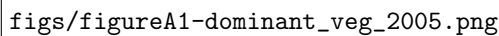
UKESM1 we use 10 irregularly-spaced elevation classes, and allow for up to 10 discrete layers in the snow pack. These schemes provide significant improvements to the surface mass balance of the ice sheets simulated by UKESM1 compared to GC3.1.

Additionally, the physical simulation in UKESM1 is affected by its Earth system components through the radiative impact of changes in atmospheric composition and through the impact of vegetation structure on surface energy and momentum fluxes. These couplings are described alongside the Earth system component models in the rest of this appendix.

A.2 Terrestrial biogeochemistry

The terrestrial biogeochemistry in UKESM1 is based largely on that of HadGEM2-ES (Collins et al., 2011), though the underlying code has been migrated from MOSES (Met Office Surface Exchange Scheme) to JULES (Best et al., 2011; Clark et al., 2011) and major enhancements have been developed for UKESM1. As in HadGEM2-ES, fractional cover and canopy height of plant functional types (PFTs) are simulated by the TRIFFID vegetation dynamics (Cox, 2001), and the resulting carbon fluxes drive the RothC soil carbon model (Coleman & Jenkinson, 1999). The model simulates the fractional extent of wetland within each grid box, which is used to derive an emission of CH_4; in the standard configuration of the model this emission is diagnostic, but the model has the capability to run with interactive CH_4 emissions.

Parameters related to photosynthesis, respiration and leaf turnover have been updated based on the TRY database (Kattge et al., 2011). These include nitrogen content of leaves, roots, and stems, leaf mass per unit area, leaf growth rate, and the relationship between leaf nitrogen and the maximum rate of carboxylation of the enzyme Rubisco at $25\,°C$ (Harper et al., 2016, 2018). The number of natural PFTs was increased from five to nine to represent the distinction between evergreen and deciduous plants and between tropical and temperate evergreen trees. The canopy scheme was updated so that the exponential decay of both photosynthetically available radiation and leaf nitrogen through the canopy now depend on the total leaf area index. In offline land-only tests, these developments improved model simulation of gross and net primary productivity (GPP and NPP, respectively) at the site level and globally (Harper et al., 2016). In particular, the NPP of C4 grasses was reduced and GPP and NPP of trees was increased.

Figure A.1. Dominant PFT within each model grid box for the year 2005, using the mean vegetation fraction of the historical ensemble. BLD: broadleaf deciduous tree; BLETr: tropical broadleaf evergreen tree; BLETe: temperate broadleaf evergreen tree; NLD: needleleaf deciduous tree; NLE: needleleaf evergreen tree; C3G: natural C3 grass; C4G: natural C4 grass; Sh: Shrub; Cr: Crop (C3 + C4 grass); Pa: Pasture (C3 + C4 grass); BS: bare soil.

¹¹⁰⁶ Plants still compete for space based on their NPP and phenology as in Cox (2001), ¹¹⁰⁷ but the competition is now based on a pure height ranking, where the tallest trees get ¹¹⁰⁸ access to space in a gridbox first. In uncoupled component model tests, the new dynamic ¹¹⁰⁹ vegetation and PFTs yield a closer match to observed vegetation distribution, with par- ¹¹¹⁰ ticular improvements to tropical and boreal forests and the high latitudes, where pre- ¹¹¹¹ viously the model overestimated the distribution of shrubs (Harper et al., 2018).

¹¹¹² Figure A.1 illustrates the overall distribution of these PFTs with the dominant veg- ¹¹¹³ etation type, i.e. the PFT having the greatest fractional coverage in each grid box. The ¹¹¹⁴ additional PFT sub-divisions added by Harper et al. (2016) can be seen most clearly in

the Boreal forests, with needleleaf deciduous trees generally dominating over needleleaf evergreen trees in more northerly continental climates. Tropical broadleaf evergreen trees dominate most of S America, sub-Saharan Africa, SE Asia and the maritime continent. In contrast, there are very few areas in which broadleaf deciduous trees or temperate broadleaf evergreen dominate, as these tend to grow sparsely in regions dominated by grasses or shrubs.

Terrestrial carbon uptake is limited by the availability of nitrogen and other nutrients. An ecosystem becomes limited when insufficient nitrogen is available for plants to allocate net photosynthate to growth. In such cases photosynthesis is down-regulated to match the available inorganic nitrogen. In UKESM1, nitrogen does not directly affect photosynthetic capacity through leaf nitrogen concentrations but acts indirectly by controlling the biomass and leaf area index within the TRIFFID vegetation dynamics. A second mechanism acts through the soil by limiting the decomposition of litter into soil carbon in the RothC model. This occurs when insufficient nitrogen is available to decompose high C:N litter into low C:N soil organic matter. The vegetation model includes retranslocation of nitrogen during senescence of leaves and roots into a labile pool to supply nutrients for the following season's leaf growth. The soil model simulates mineralisation and immobilisation, with mineralised nitrogen becoming available for plant uptake and ecosystem loss. Inorganic nitrogen is represented by a single gridbox pool from which all PFTs have equal access. As PFTs have differing nitrogen demands and turnover, the inclusion of nitrogen processes influences vegetation competition and therefore distribution.

To simulate the impact of agriculture and anthropogenic land-use on the climate (via biophysical effects) and carbon cycle (via biogeochemical effects), a representation of land-use and land-management is implemented. Going beyond the HadGEM2-ES treatment of all agriculture as "disturbed" in a common way (C. D. Jones et al., 2011), UKESM1 distinguishes crop and pasture land. This is done by reserving a fraction of each grid box for crops and a fraction for pasture; these land-use fractions are prescribed as an external forcing and may vary in time. Within the crop (pasture) fraction only the C3 crop (pasture) and C4 crop (pasture) PFTs are allowed to grow, with the area of each crop PFT determined by TRIFFID. Only the 9 natural PFTs (including natural C3 and C4 grass) may grow the remainder of the gridbox. When the prescribed crop or pasture fraction increases, natural vegetation is removed from the portion of the grid box into which

agriculture has expanded, representing anthropogenic land clearance. The vegetation carbon thus removed is distributed as follows: root biomass enters the soil carbon pools and above-ground biomass is distributed to three product pools which release CO_2 to the atmosphere at turnover rates of 1, 10 and 100 years to represent a range of fast-, medium- and slow-decay wood products. Conversely, when crop and pasture areas are reduced, the natural PFTs are allowed to re-colonise the vacated gridbox fraction.

The crop and pasture PFTs are physiologically identical to the natural grasses, but simple representations of fertiliser application and harvesting are applied to the crop PFTs. The crop PFTs are prevented from being nitrogen-limited by a fertiliser flux, which is calculated to exactly meet the nitrogen demand which cannot be met by available soil nitrogen. A crop harvest carbon flux is included by preventing 30% of crop PFT litter from entering the soil; instead this carbon flux enters the atmosphere, representing the rapid turnover of crop products. The harvest flux prevents unrealistic accumulation of soil carbon in croplands.

Finally, JULES now simulates biogenic emissions of VOCs, described in section A.4.

A.3 Ocean biogeochemistry

The Model of Ecosystem Dynamics, nutrient Utilisation, Sequestration and Acidification (MEDUSA: Yool et al., 2013) is an intermediate-complexity plankton ecosystem model designed to incorporate sufficient complexity to address key feedbacks between anthropogenically-driven changes (climate, acidification) and oceanic biogeochemistry. MEDUSA-2 resolves a dual size-structured ecosystem of small (nanophytoplankton and microzooplankton) and large (microphytoplankton and mesozooplankton) components. This explicitly includes the biogeochemical cycles of nitrogen, silicon and iron nutrients as well as the cycles of carbon, alkalinity and dissolved oxygen. Large phytoplankton are treated as diatoms and utilise silicic acid in addition to nitrogen, iron and carbon.

Similar to its living components, MEDUSA-2's detrital components are split into two size classes. Small, slow-sinking detritus is represented explicitly with separate nitrogen and carbon tracers; a fixed Fe:N ratio is assumed. Large, fast-sinking detritus is represented implicitly, and created and remineralised within a single timestep. Fast-sinking detritus consists of organic nitrogen, carbon and iron, together with silicon and calcium carbonate biominerals that are involved in a ballast parameterisation (Armstrong, Lee,

Hedges, Honjo, & Wakeham, 2001). At the seafloor, MEDUSA-2 resolves five reservoirs to store sinking organic material reaching the sediment.

The model's nitrogen, silicon and alkalinity cycles are closed and conservative (e.g. no riverine inputs), while the other three cycles (carbon, iron, oxygen) are open. The iron cycle includes aeolian and benthic sources, and is depleted by scavenging. The ocean's carbon cycle exchanges CO_2 with the atmosphere. The oxygen cycle exchanges with the atmosphere, and dissolved oxygen is additionally created by primary production and depleted by remineralisation. The various elemental cycles include both fixed and variable stoichiometry: iron has a fixed ratio with nitrogen throughout; nitrogen and carbon have fixed (but different) ratios in phytoplankton and zooplankton, and variable ratios in detritus; diatom silicon has a variable ratio with nitrogen; calcium carbonate (cf. alkalinity) is produced at a variable rate relative to organic carbon; oxygen production and consumption reflects the C:N ratio of organic matter produced and consumed.

During its development for UKESM1, several changes have occurred to the science and code of MEDUSA-2 compared to that originally described in Yool et al. (2013). The principal science changes are:

1. The carbonate chemistry sub-model in MEDUSA-2 has been upgraded to MOCSY-2.0 (Orr & Epitalon, 2015), increasing the calculation accuracy of pCO_2 and thus CO_2 air-sea flux.

2. A diagnostic submodel for DMS surface seawater concentration (Anderson et al., 2001) was added to support interactive DMS emissions to the atmosphere (see section 4.6).

3. For consistency with earlier work, the diel light cycle within UKESM1 is time-averaged for MEDUSA-2.

A.4 Atmospheric chemistry

Atmospheric composition in UKESM1 is simulated by the United Kingdom Chemistry and Aerosols model (UKCA). UKCA has several chemistry configurations available; the specific configuration used in UKESM1 is described in Archibald et al. (2019), and comprises a combination of the stratospheric scheme of Morgenstern et al. (2009) (with minor updates to reaction rates (Morgenstern et al., 2017)) and the "TropIsop" tropospheric chemistry of O'Connor et al. (2014), which includes extensions to the volatile or-

ganic compound (VOC) representation relative to the tropospheric chemistry in HadGEM2-ES. The merged scheme thus simulates interactive chemistry from the surface to the top of the model (85 km), enabling a holistic treatment of ozone, the third most important radiatively active gas in the atmosphere and the primary source of the hydroxyl radical (OH) in the troposphere. In total, the chemistry solves 291 thermal and photolytic reactions which describe the chemistry of 81 species. This includes the oxidation reactions which are a source of sulphate and secondary organic aerosol, and hence the chemistry is tightly coupled to aerosol production.
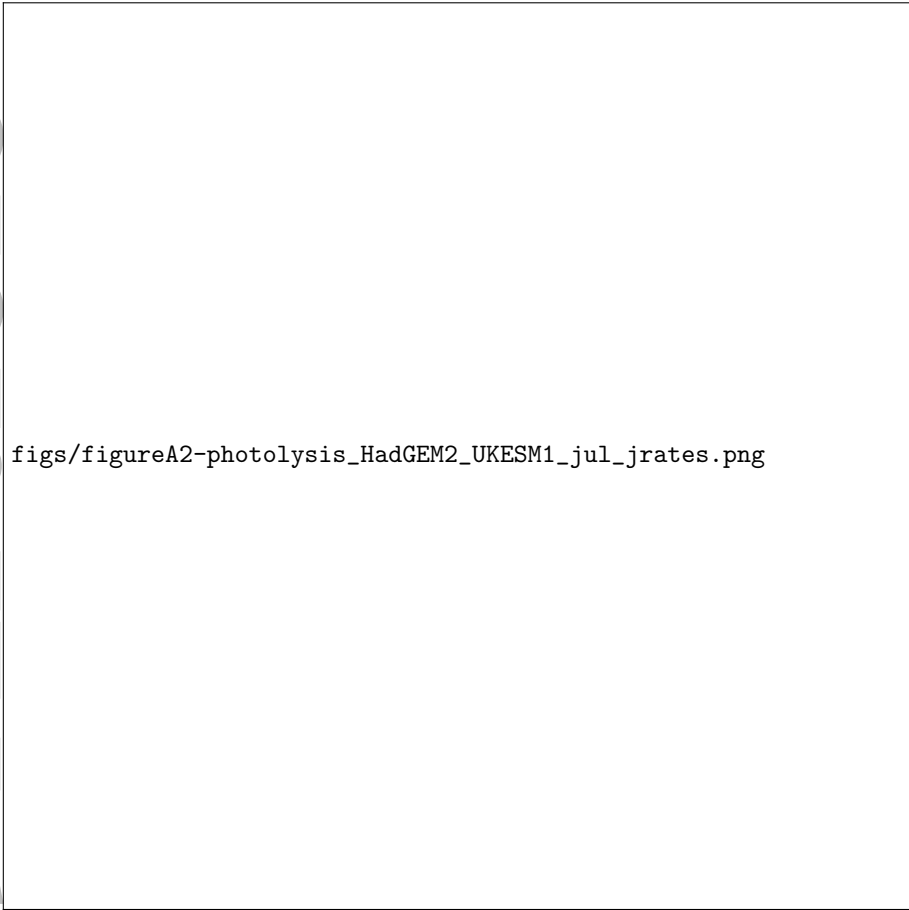
In UKESM1, prognostic chemistry influences radiation through the three-dimensional concentrations of four gases: O_3, CH_4, N_2O and H_2O. CH_4 and N_2O are forced with globally constant surface concentrations, and above the surface are treated as interactive chemical tracers. This allows the model to represent the spatiotemporal variability of these species (in particular their decay in the upper troposphere) while constraining their global-scale evolution to match past observations or a particular forcing scenario. The model has the capability to run with interactive CH_4 emissions, but in the standard configuration surface concentrations are prescribed as above.

In contrast, O_3 is fully interactive and is forced by surface emissions of ozone precursors such as VOCs, three-dimensional emissions of NOx (including emissions from aircraft and interactive production by lightning), and concentrations of ozone depleting substances. Chemical reactions which produce or destroy water vapour increment the humidity tracer of the dynamical core. This allows interactive simulation of chemical water vapour feedbacks, the most significant of which is due to stratospheric methane oxidation (see e.g. Hegglin et al., 2014; Hurst et al., 2011; Rohs et al., 2006). Concentrations of ozone depleting substances (CFCs, HCFCs and Halons) are treated as in Morgenstern et al. (2009), and thus are not coupled between UKCA and radiation.

As an inert gas, CO_2 is not modified by UKCA chemistry, and is specified as a global constant in the default configuration of UKESM1. See Appendix C.2 for a description of the CO_2 emission-driven configuration.

The photolysis scheme that is used to drive the photochemistry in UKCA is that of Fast-JX (Neu, Prather, & Penner, 2007; Telford et al., 2013) v6.4, augmented by the offline scheme after Lary and Pyle (1991) above 60 km for wavelengths between 113 and 177 nm. Fast-JX includes scattering calculations over a broad wavelength range (177-

Figure A.2. Comparison of 2005-2014 July mean photolysis rates from a single historical run, for O^1D and NO_2 below 20 km between the offline photolysis scheme used in the CMIP5 configuration of HadGEM2-ES and the Fast-JX scheme in UKESM1.

850 nm) to cover the stratosphere and lower mesosphere. This allows photolysis rates to vary with the optical depth associated with cloud liquid and frozen water as well as both the surface albedo calculated by the model's radiation scheme (Edwards & Slingo, 1996) and UKCA's own prognostic ozone distribution. However, there is no coupling to the aerosol scheme, and so the interactive photolysis rates are independent of aerosol loading.

The main purpose of this interactive calculation is to allow photolysis and hence methane lifetime to respond to stratospheric ozone recovery. This limitation made the future methane lifetime projections of HadGEM2-ES (which used fixed photolysis rates) anomalous relative to other models in the Atmospheric Chemistry and Climate Model Intercomparison Project (Voulgarakis et al., 2013). Figure A.2 compares interactive rates

–63–

in UKESM1 against the offline tabulated rates used by HadGEM2-ES, for July 2005-2014. The morphology and magnitude of the photolysis rates match reasonably well and there is a strong correlation between the offline and online rates ($r^2 > 0.96$). The Fast-JX scheme gives rise to photolysis rates that are more variable, and higher than the offline rates by a factor of 1.2-1.3. This difference persists throughout the annual cycle (not shown) and leads to the methane lifetime being approximately 25% shorter with interactive photolysis than that with offline photolysis (O'Connor et al., 2014; Telford et al., 2013). For UKESM1, such a reduction in methane lifetime would push it outside of the observational range, highlighting the importance of the improved, interactive, scheme.

Biogenic volatile organic compounds (BVOCs) play an important role in atmospheric composition and consequently in climate. BVOCs act as ozone precursors, impact strongly on the chemical lifetime of methane, and their oxidised derivatives contribute to the formation of secondary organic aerosols (SOA). The most important BVOCs for climate and air quality are isoprene and terpenes, which typically contribute 50-70% of the total BVOC emission flux in contemporary chemistry-climate models. In UKESM1, isoprene feeds into the chemical mechanism for isoprene photooxidation, whilst a monoterpene tracer (which oxidises and condenses to form SOA) is used to represent terpenes.

In UKESM1, biogenic emissions of isoprene and monoterpene are calculated interactively by means of the iBVOC model (e.g. Pacifico, Folberth, Jones, Harrison, & Collins, 2012; Pacifico et al., 2015), which applies the isoprene emission scheme of Pacifico et al. (2011) and the terpene emission parametrisation from A. Guenther et al. (1995). These parametrisations take as inputs the modelled temperature and gross primary productivity (GPP), and hence add an important coupling between climate, atmospheric composition and vegetation dynamics. In order to minimise differences in organic carbon aerosol load between UKESM1 and GC3.1, the per-PFT monoterpene emission capacity parameters were adjusted within observed ranges to achieve a present-day global total monoterpene emission of approximately 125 TgC/yr, to match the emission dataset of A. Guenther et al. (1995) used by GC3.1. Similarly, isoprene emission factors were adjusted in line with published values to achieve a present-day global total emission of approximately 500 TgC/yr (Arneth et al., 2008; A. B. Guenther et al., 2012; Lathière et al., 2006; Messina et al., 2016).

A.5 Aerosols

The aerosol configuration of UKESM1 is the same as that of GC3.1, and is described by Mulcahy et al. (2018). It uses the UKCA-GLOMAP modal scheme (Mann et al., 2010) for sulphate, black carbon, organic carbon and sea salt, and a variant of the Woodward (2011) bin scheme for mineral dust. UKESM1 does however differ from GC3.1 in its treatment of natural emissions of aerosols and aerosol precursors, as follows:

1. Biogenic emissions of monoterpene, a precursor of SOA, are interactively calculated in the land surface scheme as described in section A.4, whereas GC3.1 prescribes a climatological emission which is representative of the present day.

2. DMS emissions are coupled to ocean biogeochemistry, using the Anderson et al. (2001) parametrisation of sea water DMS concentration rather than prescribed using a present-day observational climatology.

3. UKESM1 introduces an emission of primary marine organic aerosol (PMOA), using the parametrisation of Gantt et al. (2012, 2011), acting as a source of Aitken mode organic aerosol. This parametrisation is coupled to the interactive surface chlorophyll concentration of MEDUSA. GC3.1 has no explicit PMOA representation, but instead applies a scale factor of 1.7 to its DMS emissions as a proxy for PMOA (Mulcahy et al., 2018). This DMS scale factor is not applied in UKESM1 since PMOA is explicitly simulated.

These three couplings allow the model aerosol and cloud-radiative behaviour to respond directly to changes in both the marine and terrestrial biospheres, as well as indirectly through the impact of climate on the biosphere. A detailed assessment of the impact of these emissions developments on the aerosol and radiation performance of the model is in progress.

Another difference in the treatment of aerosol between UKESM1 and GC3.1 is that the latter prescribes time-invariant concentrations of the chemical oxidants involved in the formation of sulphate and secondary organic aerosol, namely O_3, OH, NO_3, HO_2 and H_2O_2. In UKESM1 these oxidants are interactively simulated by the chemistry component, and are themselves depleted by the oxidation reactions. This oxidant coupling means that changes in methane and other ozone precursors can affect the model's aerosol loading, and aerosol radiative forcing.

1315　　Emissions of mineral dust in UKESM1 are dependent upon the modelled bare soil

1316　fraction as well as on soil moisture and near-surface wind. Deposition of dust to the ocean

1317　in turn acts as a source of iron nutrient to MEDUSA, fertilising phytoplankton growth

1318　wherever iron is limited. Dust therefore acts as a unique link between the terrestrial and

1319　marine biospheres. The dust scheme in UKESM1 is similar to that in GC3.1, and is de-

1320　scribed in Woodward (2011), with the following changes. Firstly, in UKESM1 the bare

1321　soil fraction (which dictates the area available for dust emissions) is interactively sim-

1322　ulated by the TRIFFID vegetation dynamics, rather than prescribed. This significant

1323　difference in driving data necessitated tuning of the dust emission parameters, as described

1324　in section 3. Secondly, dust emission from seasonally vegetated sources was disabled in

1325　UKESM1 in order to reduce the complexity of the coupling to TRIFFID, so that dust

1326　emissions depend on bare soil fraction, but not on simulated leaf area index. Note that

1327　following a careful tuning of the vegetation simulation, also described in section 3, it was

1328　not found necessary to impose a map of preferential dust source regions such as those

1329　used in earlier models (e.g. Collins et al., 2011; Ginoux et al., 2001; Tegen et al., 2002;

1330　Zender, Newman, & Torres, 2003).

B　Metrics and variables used in model tuning

1332　　Table B.1 lists the primary metrics which were monitored during the tuning and

1333　spinup of UKESM1. These were calculated based on annual mean model outputs (in some

1334　cases with longer-term meaning applied to smooth interannual noise) and plotted as time-

1335　series. These timeseries plots were updated routinely as test runs progressed, to rapidly

1336　understand the model's response to parameter changes. Priority 1 metrics were checked

1337　more frequently than the priority 2 metrics.

1338　　In addition to these timeseries, we also periodically examined seasonal mean, long-

1339　term averaged maps of these and other quantities to ensure there wasn't error cancel-

1340　lation in the spatial means, and to understand better the processes which were likely re-

1341　sponsible for driving differences between test runs.

Table B.1.   Priority metrics monitored during the tuning of UKESM1

Model variable	Location or averaging domain
Priority 1	
TOA radiation: SW, LW and net flux	Global and latitude bands
Cloud fraction	Global and latitude bands
Surface air temperature	Global and latitude bands
AMOC strength	26N
ACC strength	Drake passage
Sea surface temperature	Global and ocean basins
Depth-mean water temperature	Global and ocean basins
CO2 air-sea flux	Global
Priority 2	
Sea ice extent and thickness	Arctic and Antarctic
Surface emissions of DMS, PMOA, BVOCs, dust	Global and latitude bands
Soil and vegetation carbon	Global
Ocean mixed-layer depth	Global and ocean basins
Ocean surface DIC, iron, nitrate	Global
Primary productivity	Global land and ocean

Notes: Latitude bands are: 90-30S, 30S-30N, 30N-90N. Priority ocean regions

monitored were: N Atlantic, Southern Ocean, Weddell Sea, Ross Sea.

C Special configurations of the model

C.1 Faster configurations: UKESM1-CN and UKESM1-C

Because the chemistry scheme in UKESM1 is particularly expensive, we have implemented a cheaper configuration of the model, known as UKESM1-CN (a carbon-nitrogen cycle model) which reduces computational expense by utilising the simpler "offline oxidant" chemistry mechanism of GC3.1 (Walters et al., 2019). This scheme simulates sulphur chemistry for the production of sulphate aerosol using prescribed oxidants, but excludes other interactive chemical processes. In addition, UKESM1-CN does not have interactive emissions of BVOCs. Apart from these two changes, UKESM1-CN is identical to the full UKESM1 configuration. It runs approximately twice as fast as UKESM1. UKESM1-CN will not be used for submissions to CMIP6, but is intended for use in carbon-nitrogen cycle research and other applications which do not depend upon interactive chemistry.

Because it is lacking fully interactive chemistry, the UKESM1-CN configuration needs additional input files of the following variables:

1. O_3, for radiation and as an oxidant for aerosol chemistry,

2. other oxidants for aerosol chemistry (OH, NO_3, HO_2, H_2O_2),

3. and biogenic emissions of monoterpene.

For consistency with the full model, each of these inputs are derived from the interactive model fields produced by a prior run with UKESM1. Experience thus far indicates that headline metrics such as ECS are consistent between UKESM1 and UKESM1-CN. Future work will study the traceability between these configurations of differing complexity.

A further simplification incorporates only the carbon cycle (UKESM1-C), primarily for understanding the impact of the nitrogen processes. The nitrogen processes do not add significant cost, so UKESM1-C runs at the same speed as UKESM1-CN.

C.2 CO_2 emission-driven configuration

All of the results in section 4, and most of the simulations which will be submitted to CMIP6, are driven by CO_2 concentrations rather than CO_2 emissions. When UKESM1

Figure C.1. Climatological seasonal cycle of near-surface atmospheric CO_2 concentration at Mauna Loa, Hawaii (upper) and Point Barrow, Alaska (lower) for the period 1960-2008. Observations are from CDIAC. Model values are from the UKESM1 emission-driven historical experiment.

runs in CO_2 concentration-driven mode, atmospheric CO_2 concentrations are externally prescribed as a global constant and the biogeochemical fluxes of CO_2 are purely diagnostic. From these natural fluxes, compatible anthropogenic emissions can be derived, as per Friedlingstein et al. (2006).

When UKESM1 runs in CO_2 emission-driven mode, atmospheric CO_2 is held as an advected 3D tracer, updated by the natural surface fluxes and anthropogenic emissions. In this mode CO_2 concentrations are free to evolve, and thus drive a radiative response the model's carbon cycle feedbacks. The 3D nature of the free tracer means that (unlike in concentration-driven mode) the model resolves the spatial structure of CO_2, including hemispheric and stratosphere-troposphere differences, and its seasonal cycle.

Two emission-driven simulations have been performed: a pre-industrial control with zero anthropogenic emissions, and a historical run from 1850 with evolving CMIP6 an-

figs/figureC2-CO2_MAUNA_LOA_TS_MONTHLY_UKESM1_0m_05092019.pdf
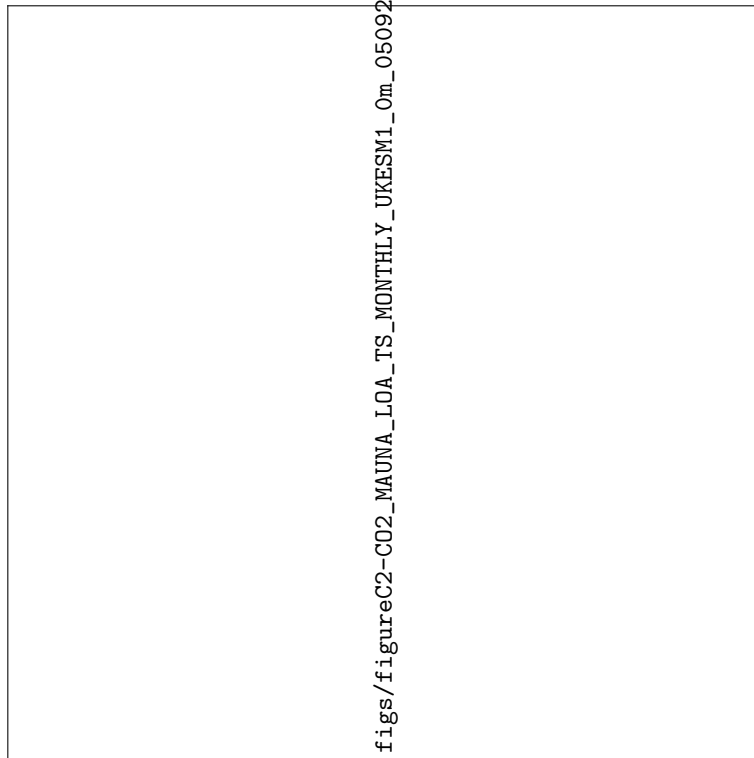
Figure C.2. Monthly timeseries of near-surface CO_2 at Mauna Loa, Hawaii, from UKESM1 emission-driven piControl and historical experiments, and CDIAC observations.

thropogenic emissions. In Figure C.1 we compare the seasonal cycle of the historical run against observations from the Carbon Dioxide Information Analysis Centre (Keeling, Piper, Bollenbacher, & Walker, 1994, CDIAC;). These show that the timing of the model's seasonal cycle compares well against the observations, though the summer draw-down begins too early. The seasonal range is weaker than observations at Mauna Loa by approximately 40%, and stronger at Barrow by approximately 20%.

Figure C.2 shows the temporal evolution of CO_2 for these experiments at Mauna Loa, again compared to historical observations from CDIAC. The pre-industrial control concentrations vary on decadal timescales within the range 280-288 ppm, which encompasses the observed value of 284 ppm prescribed in concentration-driven experiments (Meinshausen et al., 2017). The historical experiment drifts slightly higher than the observations, reaching a bias of about 10 ppm by 2000. Such close agreement with the observations is a good result given the number of degrees of freedom in the emission-driven configuration, and compares well with the CMIP5 generation of models.

It is critical that the fully coupled model conserves carbon to a high accuracy: a significant loss of conservation would confound the analysis of carbon feedbacks. The UKESM1 atmospheric convection and boundary layer schemes only conserve tracer mass to within order 0.1 Gt of CO2 per year, which is not sufficiently accurate for emission-driven simulations. An atmospheric mass correction has therefore been implemented, as was required in HadGEM2-ES (C. D. Jones et al., 2011). In UKESM1 this correction uses an implementation of the Priestley (2002) conservation algorithm to distribute the correction increments globally. The ocean and land components did not require any such correction scheme to achieve this level of conservation.

The finalised model conserves carbon to high accuracy. Over the first 64 years of the emission-driven pre-industrial control and historical simulations, the model gained an average of -1.3×10^{-3} GtC/yr and 1.2×10^{-4} GtC/yr respectively. These rates are two or more orders of magnitude smaller than the C4MIP-recommended $0.1\,\mathrm{Gt\,yr^{-1}}$ threshold for acceptable drift in carbon pools in CO_2 concentration-driven experiments (C. D. Jones et al., 2016), so we consider this an acceptable loss of conservation for CO_2 emission-driven experiments.

D Simulation identifiers

All simulations used in this work were performed using version 10.9 of the UM, version 5.0 of JULES, NEMO version 3.6, CICE version 5.1.2 and OASIS3-MCT version 3.0. Simulation identifiers are as follows:

1. piControl: u-aw310
2. 4xCO2: u-bb446
3. 1pctCO2 u-bb448, u-bd334, u-bd335, u-bd336
4. historical: u-bc179, u-bc292, u-bc370, u-bb075, u-az513, u-az515, u-az524, u-bb277, u-bc370, u-bc470
5. un-tuned pre-industrial test: u-bd881 (the control is u-aw310)
6. un-tuned present-day test: u-bd880 (the control is u-az513)

Acknowledgments

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., . . .
Arkin, P.   (2003).   The Version 2 Global Precipitation Climatology Project

({GPCP}) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeorol.*, *4*, 1147–1167.

Anderson, T., Spall, S., Yool, A., Cipollini, P., Challenor, P., & Fasham, M. (2001, 8). Global fields of sea surface dimethylsulfide predicted from chlorophyll, nutrients and light. *Journal of Marine Systems*, *30*(1-2), 1–20. doi: 10.1016/S0924-7963(01)00028-8

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., & Schulz, J. (2010). The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data HOAPS-3. *Earth System Science Data*, *2*(2), 215–234. doi: 10.5194/essd-2-215-2010

Andrews, T., Gregory, J. M., Webb, M. J., & Taylor, K. E. (2012). Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters*, *39*(9). doi: 10.1029/2012GL051607

Archibald, A. T., O'Connor, F. M., Abraham, N. L., S., A.-N., P., C. M., M., D., . . . Zeng, G. (2019). Description and evaluation of the UKCA stratosphere-troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM1. *Geoscientific Model Development Discussions*. doi: 10.5194/gmd-2019-246

Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., & Wakeham, S. G. (2001). A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep Sea Research Part II: Topical Studies in Oceanography*, *49*(1-3), 219–236. doi: 10.1016/S0967-0645(01)00101-1

Arneth, A., Monson, R. K., Schurgers, G., Niinemets, , & Palmer, P. I. (2008, 8). Why are estimates of global terrestrial isoprene emissions so similar (and why is this not so for monoterpenes)? *Atmospheric Chemistry and Physics*, *8*(16), 4605–4620. doi: 10.5194/acp-8-4605-2008

Avitabile, V., Herold, M., Heuvelink, G. B., Lewis, S. L., Phillips, O. L., Asner, G. P., . . . Willcock, S. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*, *22*(4), 1406–1420. doi: 10.1111/gcb.13139

Batjes, N. H. (2016). Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma*, *269*, 61–68. doi: 10.1016/j.geoderma.2016.01.034

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*. doi: 10 .4319/lo.1997.42.1.0001

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., . . . Harding, R. J. (2011, 9). The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. doi: 10.5194/gmd-4-677-2011

Bodas-Salcedo, A., Mulcahy, J. P., Andrews, T., Williams, K. D., Ringer, M. A., Field, P. R., & Elsaesser, G. S. (2019, 6). Strong Dependence of Atmospheric Feedbacks on Mixed-Phase Microphysics and Aerosol-Cloud Interactions in HadGEM3. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1735–1758. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001688` doi: 10.1029/2019MS001688

Bodeker, G. E., Shiona, H., & Eskes, H. (2005, 9). Indicators of Antarctic ozone depletion. *Atmospheric Chemistry and Physics*, *5*(10), 2603–2615. Retrieved from `http://www.bodekerscientific.com/data/total-column-ozone` doi: 10.5194/acp-5-2603-2005

Boisier, J. P., De Noblet-Ducoudré, N., Pitman, A. J., Cruz, F. T., Delire, C., Van Den Hurk, B. J., . . . Voldoire, A. (2012). Attributing the impacts of land-cover changes in temperate regions on surface temperature and heat fluxes to specific causes: Results from the first LUCID set of simulations. *Journal of Geophysical Research Atmospheres*. doi: 10.1029/2011JD017106

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., . . . Vichi, M. (2013). Multiple stressors of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*, *10*(10), 6225–6245. doi: 10.5194/bg-10-6225-2013

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

Brooks, R. H., & Corey, A. T. (1964). *Hydraulic properties of porous media* (Vol. 3).

Brown, O., J., Ferrians, J. A., Heginbottom, & Melnikov, E. (2002). *Circum-Arctic*

Map of Permafrost and Ground-Ice Conditions, Version 2. Boulder, Colorado USA: NSIDC: National Snow and Ice Data Center.

Burke, E. J., Chadburn, S. E., Huntingford, C., & Jones, C. D. (2018). CO2 loss by permafrost thawing implies additional emissions reductions to limit warming to 1.5 or 2 C. *Environmental Research Letters*, *13*(2). doi: 10.1088/1748-9326/aaa138

Burton, C., Betts, R., Cardoso, M., Feldpausch, R. T., Harper, A., Jones, C. D., ... Wiltshire, A. (2019). Representation of fire, land-use change and vegetation dynamics in the Joint UK Land Environment Simulator vn4.9 (JULES). *Geoscientific Model Development*, *12*(1), 179–193. doi: 10.5194/gmd-12-179-2019

Carr, M.-E., Friedrichs, M. A., Schmeltz, M., Noguchi Aita, M., Antoine, D., Arrigo, K. R., ... Yamanaka, Y. (2006, 3). A comparison of global estimates of marine primary production from ocean color. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5-7), 741–770. doi: 10.1016/J.DSR2.2006.01.028

Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., ... Reichstein, M. (2014). Global covariation of carbon turnover times with climate in terrestrial ecosystems. *Nature*, *514*(7521), 213–217. doi: 10.1038/nature13731

Chadburn, S. E., Burke, E. J., Cox, P. M., Friedlingstein, P., Hugelius, G., & Westermann, S. (2017). An observation-based constraint on permafrost loss as a function of global warming. *Nature Climate Change*, *7*(5), 340–344. doi: 10.1038/nclimate3262

Chipperfield, M. P., Bekki, S., Dhomse, S., Harris, N. R., Hassler, B., Hossaini, R., ... Weber, M. (2017). Detecting recovery of the stratospheric ozone layer. *Nature*, *549*(7671), 211–218. doi: 10.1038/nature23681

Ciais, P., Willem, J., Friedlingstein, P., & Munhoven, G. (2013). *Carbon and Other Biogeochemical Cycles: In Climate Change 2013* (Vol. 6; Tech. Rep.). doi: 10.1017/CBO9781107415324.015

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., ... Cox, P. M. (2011, 9). The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. doi: 10.5194/gmd-4-701-2011

Coleman, K., & Jenkinson, D. S. (1999). *RothC-26.3: a model for the turnover of*

carbon in soil. Model description and Users' guide (Tech. Rep.). Harpenden, UK: Lawes Agricultural Trust.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., ... Woodward, S. (2011). Development and evaluation of an Earth-System model-HadGEM2. *Geosci. Model Dev*. doi: 10.5194/gmd-4-1051-2011

Collins, W. J., Webber, C. P., Cox, P. M., Huntingford, C., Lowe, J., Sitch, S., ... Powell, T. (2018). Increased importance of methane reduction for a 1.5 degree target. *Environmental Research Letters*, *13*(5). doi: 10.1088/1748-9326/aab89c

Cox, P. M. (2001). Description of the "TRIFFID" Dynamic Global Vegetation Model. *Hadley Centre technical note*.

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT-3.0. *Geoscientific Model Development*, *10*(9), 3297–3308. doi: 10.5194/gmd-10-3297-2017

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

Edwards, J. M., & Slingo, A. (1996). Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1256/smsqj.53106

Eyring, V., Arblaster, J. M., Cionni, I., Sedláček, J., Perlwitz, J., Young, P. J., ... Watanabe, S. (2013). Long-term ozone changes and associated climate impacts in CMIP5 simulations. *Journal of Geophysical Research Atmospheres*. doi: 10.1002/jgrd.50316

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

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., ... Rummukainen, M. (2013). Evaluation of Climate Models. In *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 741–

866). doi: 10.1017/CBO9781107415324

Friedlingstein, P. (2015). Carbon cycle feedbacks and future climate change. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2054). doi: 10.1098/rsta.2014.0421

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., ... Zeng, N. (2006, 7). Climate–Carbon Cycle Feedback Analysis: Results from the C 4 MIP Model Intercomparison. *Journal of Climate*, *19*(14), 3337–3353. doi: 10.1175/jcli3800.1

Gantt, B., Johnson, M. S., Meskhidze, N., Sciare, J., Ovadnevaite, J., Ceburnis, D., & O&apos;Dowd, C. D. (2012, 9). Model evaluation of marine primary organic aerosol emission schemes. *Atmospheric Chemistry and Physics*, *12*(18), 8553–8566. doi: 10.5194/acp-12-8553-2012

Gantt, B., Meskhidze, N., Facchini, M. C., Rinaldi, M., Ceburnis, D., & O&apos;Dowd, C. D. (2011, 8). Wind speed dependent size-resolved parameterization for the organic mass fraction of sea spray aerosol. *Atmospheric Chemistry and Physics*, *11*(16), 8777–8790. doi: 10.5194/acp-11-8777-2011

Garfinkel, C. I., Oman, L. D., Barnes, E. A., Waugh, D. W., Hurwitz, M. H., & Molod, A. M. (2013). Connections between the Spring Breakup of the Southern Hemisphere Polar Vortex, Stationary Waves, and Air–Sea Roughness. *Journal of the Atmospheric Sciences*, *70*(7), 2137–2151. doi: 10.1175/jas-d-12-0242.1

Garrigues, S., Lacaze, R., Baret, F., Morisette, J. T., Weiss, M., Nickeson, J. E., ... Yang, W. (2008, 6). Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. *Journal of Geophysical Research: Biogeosciences*, *113*(G2), n/a-n/a. doi: 10.1029/2007JG000635

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., ... Ziemke, J. (2018). Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation. *Elem Sci Anth*. doi: 10.1525/elementa.291

Gehlen, M., Séférian, R., Jones, D. O., Roy, T., Roth, R., Barry, J., ... Tjiputra, J. (2014). Projected pH reductions by 2100 might put deep North Atlantic biodiversity at risk. *Biogeosciences*, *11*(23), 6955–6967. doi:

10.5194/bg-11-6955-2014

Gillett, N. P., Arora, V. K., Matthews, D., & Allen, M. R. (2013). Constraining the ratio of global warming to cumulative CO2 emissions using CMIP5 simulations. *Journal of Climate*, *26*(18), 6844–6858. doi: 10.1175/JCLI-D-12-00476.1

Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., & Lin, S. J. (2001). Sources and distributions of dust aerosols simulated with the GOCART model. *Journal of Geophysical Research Atmospheres*. doi: 10.1029/2000JD000053

Global Soil Data Task Group. (2002). *Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS)* (Tech. Rep.). doi: 10.3334/ORNLDAAC/569

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., & Jung, T. (2016). Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, *43*(4), 1642–1650. doi: 10.1002/2015GL067232

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., . . . Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). doi: 10.1029/2003GL018747

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., . . . Zimmermann, P. (1995, 5). A global model of natural volatile organic compound emissions. *Journal of Geophysical Research*, *100*(D5), 8873. doi: 10.1029/94JD02950

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., & Wang, X. (2012, 11). 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. Retrieved from `https://www.geosci-model-dev.net/5/1471/2012/` doi: 10.5194/gmd-5-1471-2012

Hardiman, S. C., Boutle, I. A., Bushell, A. C., Butchart, N., Cullen, M. J., Field, P. R., . . . Woodhouse, M. T. (2015). Processes controlling tropical tropopause temperature and stratospheric water vapor in climate models. *Journal of Climate*, *28*(16), 6516–6535. doi: 10.1175/JCLI-D-15-0075.1

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch,

S., . . . Bodegom, P. v. (2016, 7). 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. doi: 10.5194/gmd-9-2415-2016

Harper, A. B., Wiltshire, A. J., Cox, P. M., Friedlingstein, P., Jones, C. D., Mercado, L. M., . . . Duran-Rojas, C. (2018, 7). Vegetation distribution and terrestrial carbon cycle in a carbon cycle configuration of JULES4.6 with new plant functional types. *Geoscientific Model Development*, *11*(7), 2857–2873. doi: 10.5194/gmd-11-2857-2018

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, *34*(3), 623–642. doi: 10.1002/joc.3711

Hegglin, M. I., Plummer, D. A., Shepherd, T. G., Scinocca, J. F., Anderson, J., Froidevaux, L., . . . Weigel, K. (2014, 10). Vertical structure of stratospheric water vapour trends derived from merged satellite data. *Nature Geoscience*, *7*(10), 768–776. doi: 10.1038/ngeo2236

Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., . . . Zibordi, G. (2001, 6). An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. *Journal of Geophysical Research: Atmospheres*, *106*(D11), 12067–12097. doi: 10.1029/2001JD900014

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., . . . Williamson, D. (2016, 7). The Art and Science of Climate Model Tuning. *Bull. Amer. Meteor. Soc.*, *98*(3), 589–602. doi: 10.1175/bams-d-15-00135.1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., . . . Stocker, E. F. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.*, *8*, 38–55. doi: 10.1175/JHM560.1

Hurst, D. F., Oltmans, S. J., Vömel, H., Rosenlof, K. H., Davis, S. M., Ray, E. A., . . . Jordan, A. F. (2011). Stratospheric water vapor trends over Boulder, Colorado: Analysis of the 30 year Boulder record. *Journal of Geophysical Research Atmospheres*, *116*(2). doi: 10.1029/2010JD015065

Johnson, G. C., Lyman, J. M., & Loeb, N. G. (2016, 7). Improving estimates of Earth's energy imbalance. *Nature Climate Change*, *6*(7), 639–640. doi: 10

1686 .1038/nclimate3043

1687 Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., . . .

1688 Zaehle, S. (2016). C4MIP-The Coupled Climate-Carbon Cycle Model Inter-

1689 comparison Project: Experimental protocol for CMIP6. *Geoscientific Model*

1690 *Development*, *9*(8), 2853–2880. doi: 10.5194/gmd-9-2853-2016

1691 Jones, C. D., Hughes, J., Bellouin, N., Hardiman, S., Jones, G., Knight, J., . . .

1692 Zerroukat, M. (2011). The HadGEM2-ES implementation of CMIP5

1693 centennial simulations. *Geoscientific Model Development*, *4*(3). doi:

1694 10.5194/gmd-4-543-2011

1695 Jones, P., Lister, D., Osborn, T., Harpham, C., Salmon, M., & Morice, C. (2012).

1696 Hemispheric and large-scale land-surface air temperature variations: An ex-

1697 tensive revision and an update to 2010. *Journal of Geophysical Research*

1698 *Atmospheres*, *117*(5). doi: 10.1029/2011JD017139

1699 Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain,

1700 M. A., . . . Williams, C. (2011, 9). Global patterns of land-atmosphere fluxes

1701 of carbon dioxide, latent heat, and sensible heat derived from eddy covariance,

1702 satellite, and meteorological observations. *Journal of Geophysical Research*,

1703 *116*(G3), G00J07. doi: 10.1029/2010JG001566

1704 Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., . . . Wirth,

1705 C. (2011). TRY - a global database of plant traits. *Global Change Biology*.

1706 doi: 10.1111/j.1365-2486.2011.02451.x

1707 Keeling, R. F., Piper, S. C., Bollenbacher, A., & Walker, J. (1994). Atmo-

1708 spheric {CO}_2 records from sites in the {SIO} air sampling network.

1709 In *Trends: A compendium of data on global change* (pp. 16–26). doi:

1710 10.3334/CDIAC/atg.035

1711 Kerr, R. A. (2000, 6). A north atlantic climate pacemaker for the centuries. *Science*,

1712 *288*(5473), 1984–5. doi: 10.1126/science.288.5473.1984

1713 Kettle, A. J., Andreae, M. O., Amouroux, D., Andreae, T. W., Bates, T. S.,

1714 Berresheim, H., . . . Uher, G. (1999). A global database of sea surface

1715 dimethylsulfide (DMS) measurements and a procedure to predict sea surface

1716 DMS as a function of latitude, longitude, and month. *Global Biogeochemical*

1717 *Cycles*, *13*(2), 399–444. doi: 10.1029/1999GB900004

1718 Klingaman, N. P., Martin, G. M., & Moise, A. (2017). ASoP (v1.0): a set of meth-

ods for analyzing scales of precipitation in general circulation models. *Geoscientific Model Development*, *10*(1), 57–83. doi: 10.5194/gmd-10-57-2017

Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., ... Walton, J. (2018, 11). The Low-Resolution Version of HadGEM3 GC3.1: Development and Evaluation for Global Climate. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2865–2888. doi: 10.1029/2018MS001370

Kurtakoti, P., Veneziani, M., Stössel, A., Weijer, W., Kurtakoti, P., Veneziani, M., ... Weijer, W. (2018, 12). Preconditioning and Formation of Maud Rise Polynyas in a High-Resolution Earth System Model. *Journal of Climate*, *31*(23), 9659–9678. doi: 10.1175/JCLI-D-18-0392.1

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., ... Cox, P. M. (2014). IMarNet: An ocean biogeochemistry model intercomparison project within a common physical ocean modelling framework. *Biogeosciences*, *11*(24), 7291–7304. doi: 10.5194/bg-11-7291-2014

Lana, A., Bell, T. G., Simó, R., Vallina, S. M., Ballabrera-Poy, J., Kettle, A. J., ... Liss, P. S. (2011). An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean. *Global Biogeochemical Cycles*, *25*(1). doi: 10.1029/2010GB003850

Lary, D. J., & Pyle, J. A. (1991). Diffuse radiation, twilight, and photochemistry - II. *Journal of Atmospheric Chemistry*. doi: 10.1007/BF00057754

Lathière, J., Hauglustaine, D. A., Friend, A. D., De Noblet-Ducoudré, N., Viovy, N., & Folberth, G. A. (2006, 6). Impact of climate variability and land use changes on global biogenic volatile organic compound emissions. *Atmospheric Chemistry and Physics*, *6*(8), 2129–2146. doi: 10.5194/acp-6-2129-2006

Le Quéré, C., Andrew, R., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., ... Zheng, B. (2018). Global Carbon Budget 2018. *Earth System Science Data*, *10*(4), 2141–2194. doi: 10.5194/essd-10-2141-2018

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., ... Kato, S. (2018, 1). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate*, *31*(2), 895–918. doi: 10.1175/JCLI-D-17-0208.1

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*. doi: 10.1080/014311600210191

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., ... Johnson, C. E. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development*. doi: 10.5194/gmd-3-519-2010

Maritorena, S., D'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2010.04.002

Martin, G. M., & Levine, R. C. (2012, 11). The influence of dynamic vegetation on the present-day simulation and future projections of the South Asian summer monsoon in the HadGEM2 family. *Earth System Dynamics*, *3*(2), 245–261. doi: 10.5194/esd-3-245-2012

Matthews, H. D., Landry, J. S., Partanen, A. I., Allen, M., Eby, M., Forster, P. M., ... Zickfeld, K. (2017). Estimating Carbon Budgets for Ambitious Climate Targets. *Current Climate Change Reports*, *3*(1), 69–77. doi: 10.1007/s40641-017-0055-0

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... Tomassini, L. (2012, 3). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*(3), n/a-n/a. doi: 10.1029/2012MS000154

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

Ménard, C. B., Essery, R., Pomeroy, J., Marsh, P., & Clark, D. B. (2014, 1). A shrub bending model to calculate the albedo of shrub-tundra. *Hydrological Processes*, *28*(2), 341–351. doi: 10.1002/hyp.9582

Messina, P., Lathière, J., Sindelarova, K., Vuichard, N., Granier, C., Ghattas, J., ... Hauglustaine, D. A. (2016, 11). Global biogenic volatile organic compound

emissions in the ORCHIDEE and MEGAN models and sensitivity to key parameters. *Atmospheric Chemistry and Physics*, *16*(22), 14169–14202. doi: 10.5194/acp-16-14169-2016

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*. doi: 10.5194/gmd-2-43-2009

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

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012, 4). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, *117*(D8), n/a-n/a. doi: 10.1029/2011JD017187

Mulcahy, J. P., Jones, C., Sellar, A., Johnson, B., Boutle, I. A., Jones, A., ... McCoy, D. T. (2018, 11). Improved Aerosol Processes and Effective Radiative Forcing in HadGEM3 and UKESM1. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2786–2805. doi: 10.1029/2018MS001464

Müller, J.-F. F., Stavrakou, T., Wallens, S., De Smedt, I., Van Roozendael, M., Potosnak, M. J., ... Guenther, A. B. (2008, 3). Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environment model. *Atmospheric Chemistry and Physics*, *8*(5), 1329–1341. doi: 10.5194/acp-8-1329-2008

Neu, J. L., Prather, M. J., & Penner, J. E. (2007, 6). Global atmospheric chemistry: Integrating over fractional cloud cover. *Journal of Geophysical Research*, *112*(D11), D11306. doi: 10.1029/2006JD008007

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

Oki, T., & Sud, Y. C. (2006). Design of Total Runoff Integrating Pathways (TRIP)—A Global River Channel Network. *Earth Interactions*, *2*(1), 1–37.

doi: 10.1175/1087-3562(1998)002⟨0001:dotrip⟩2.3.co;2

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., . . . Kassem, K. R. (2006). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*. doi: 10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2

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

Orr, J. C., & Epitalon, J.-M. (2015, 3). Improved routines to model the ocean carbonate system: mocsy 2.0. *Geoscientific Model Development*, *8*(3), 485–499. doi: 10.5194/gmd-8-485-2015

Pacifico, F., Folberth, G. A., Jones, C. D., Harrison, S. P., & Collins, W. J. (2012, 11). Sensitivity of biogenic isoprene emissions to past, present, and future environmental conditions and implications for atmospheric chemistry. *Journal of Geophysical Research: Atmospheres*, *117*(D22). doi: 10.1029/2012JD018276

Pacifico, F., Folberth, G. A., Sitch, S., Haywood, J. M., Rizzo, L. V., Malavelle, F. F., & Artaxo, P. (2015). Biomass burning related ozone damage on vegetation over the Amazon forest: A model sensitivity study. *Atmospheric Chemistry and Physics*. doi: 10.5194/acp-15-2791-2015

Pacifico, F., Harrison, S. P., Jones, C. D., Arneth, A., Sitch, S., Weedon, G. P., . . . Schurgers, G. (2011, 5). Evaluation of a photosynthesis-based biogenic isoprene emission scheme in JULES and simulation of isoprene emissions under present-day climate conditions. *Atmospheric Chemistry and Physics*, *11*(9), 4371–4389. doi: 10.5194/acp-11-4371-2011

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., . . . Peylin, P. (2015). Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative. *Geoscientific Model Development*. doi: 10.5194/gmd-8-2315-2015

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

Prather, M. J., Holmes, C. D., & Hsu, J. (2012, 5). Reactive greenhouse gas sce-

narios: Systematic exploration of uncertainties and the role of atmospheric chemistry. *Geophysical Research Letters*, *39*(9). doi: 10.1029/2012GL051440

Priestley, A. (2002). A Quasi-Conservative Version of the Semi-Lagrangian Advection Scheme. *Monthly Weather Review*, *121*(2), 621–629. doi: 10.1175/1520 -0493(1993)121⟨0621:aqcvot⟩2.0.co;2

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., ... Kaplan, A. (2003, 7). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*(D14), 4407. doi: 10.1029/2002JD002670

Reintges, A., Martin, T., Latif, M., & Park, W. (2017). Physical controls of Southern Ocean deep-convection variability in CMIP5 models and the Kiel Climate Model. *Geophysical Research Letters*, *44*(13), 6951–6958. doi: 10.1002/2017GL074087

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., & Schroeder, D. (2018, 2). The sea ice model component of HadGEM3-GC3.1. *Geoscientific Model Development*, *11*(2), 713–723. doi: 10.5194/gmd-11-713-2018

Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean pCO2 and sea-Air CO2 flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Science*. doi: 10.5194/os-9-193-2013

Rohs, S., Schiller, C., Riese, M., Engel, A., Schmidt, U., Wetter, T., ... Aoki, S. (2006). Long-term changes of methane and hydrogen in the stratosphere in the period 1978-2003 and their impact on the abundance of stratospheric water vapor. *Journal of Geophysical Research Atmospheres*, *111*(14). doi: 10.1029/2005JD006877

Rosen, J. M., Kjome, N. T., & Oltmans, S. J. (1993, 7). Simultaneous ozone and polar stratospheric cloud observations at South Pole station during winter and spring 1991. *Journal of Geophysical Research*, *98*(D7), 12741. doi: 10.1029/93JD00880

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., ... Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sci-*

ences, *108*(24), 9899–9904. doi: 10.1073/pnas.1019576108

Sadiq, M., Tai, A. P., Lombardozzi, D., & Val Martin, M. (2017). Effects of ozone-vegetation coupling on surface ozone air quality via biogeochemical and meteorological feedbacks. *Atmospheric Chemistry and Physics*, *17*(4), 3055–3066. doi: 10.5194/acp-17-3055-2017

Schamm, K., Ziese, M., Becker, A., Finger, P., Meyer-Christoffer, A., Schneider, U., ... Stender, P. (2014). Global gridded precipitation over land: a description of the new GPCC First Guess Daily product. *Earth System Science Data*, *6*(1), 49–60. doi: 10.5194/essd-6-49-2014

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., ... Saha, S. (2017, 9). Practice and philosophy of climate model tuning across six US modeling centers. *Geoscientific Model Development*, *10*(9), 3207–3223. doi: 10.5194/gmd-10-3207-2017

Senior, C. A., Andrews, T., Burton, C., Chadwick, R., Copsey, D., Graham, T., ... Tsushima, Y. (2016). Idealized climate change simulations with a high-resolution physical model: HadGEM3-GC2. *Journal of Advances in Modeling Earth Systems*, *8*(2), 813–830. doi: 10.1002/2015MS000614

Senior, C. A., Jones, C. G., Wood, R. A., Sellar, A., Belcher, S., Gregory, J., ... Andrews, T. (2019). UK Community Earth System Modelling. To be submitted to UK-CMIP6 special issue of J. Adv. Model. Earth Systems. *Journal of Advances in Modeling Earth Systems*.

Shannon, S., Smith, R., Wiltshire, A., Payne, T., Huss, M., Betts, R., ... Harrison, S. (2019). Global glacier volume projections under high-end climate change scenarios. *Cryosphere*, *13*(1), 325–350. doi: 10.5194/tc-13-325-2019

Silva, R. A., West, J. J., Lamarque, J. F., Shindell, D. T., Collins, W. J., Faluvegi, G., ... Zeng, G. (2017). Future global mortality from changes in air pollution attributable to climate change. *Nature Climate Change*, *7*(9), 647–651. doi: 10.1038/nclimate3354

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

Skillman, J. B. (2008). Quantum yield variation across the three pathways of photo-

synthesis: Not yet out of the dark. In *Journal of experimental botany.* doi: 10 .1093/jxb/ern029

SPARC. (2013). SPARC Report on the Lifetimes of Stratospheric Ozone-Depleting Substances, Their Replacements, and Related Species. *SPARC Rep. 6, WCRP-15/2013*, edited by M. Ko et al.

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., van der Werf, G. R., Giglio, L., & Guenther, A. (2009, 2). Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns. *Atmospheric Chemistry and Physics*, *9*(3), 1037–1060. doi: 10.5194/acp-9-1037-2009

Stohl, A., Aamaas, B., Amann, M., Baker, L. H., Bellouin, N., Berntsen, T. K., . . . Zhu, T. (2015). Evaluating the climate and air quality impacts of short-lived pollutants. *Atmospheric Chemistry and Physics*, *15*(18), 10529–10566. doi: 10.5194/acp-15-10529-2015

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., . . . Sinha, B. (2018, 8). UK Global Ocean GO6 and GO7: a traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*(8), 3187–3213. doi: 10.5194/gmd-11-3187-2018

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J. P., . . . Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, *115*(12), 3517–3529. doi: 10.1016/j.rse.2011.08.014

Tegen, I., Harrison, S. P., Kohfeld, K., Prentice, I. C., Coe, M., & Heimann, M. (2002). Impact of vegetation and preferential source areas on global dust aerosol: Results from a model study. *Journal of Geophysical Research Atmospheres*. doi: 10.1029/2001JD000963

Telford, P. J., Abraham, N. L., Archibald, A. T., Braesicke, P., Dalvi, M., Morgenstern, O., . . . Pyle, J. A. (2013, 2). Implementation of the Fast-JX Photolysis scheme (v6.4) into the UKCA component of the MetUM chemistry-climate model (v7.3). *Geoscientific Model Development*, *6*(1), 161–177. doi: 10.5194/gmd-6-161-2013

Tilmes, S., Lamarque, J. F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia,

R. R., ... Blake, N. (2016). Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI). *Geoscientific Model Development*. doi: 10.5194/gmd-9-1853-2016

Titchner, H. A., & Rayner, N. A. (2014). The met office hadley centre sea ice and sea surface temperature data set, version 2: 1. sea ice concentrations. *Journal of Geophysical Research*, *119*(6), 2864–2889. doi: 10.1002/2013JD020316

van Genuchten, M. T. (1980). *A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils* (Vol. 44; Tech. Rep. No. 5). doi: 10.2136/sssaj1980.03615995004400050002x

Voulgarakis, A., Naik, V., Lamarque, J.-F., Shindell, D. T., Young, P. J., Prather, M. J., ... Zeng, G. (2013, 3). Analysis of present day and future OH and methane lifetime in the ACCMIP simulations. *Atmospheric Chemistry and Physics*, *13*(5), 2563–2587. doi: 10.5194/acp-13-2563-2013

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., ... Zerroukat, M. (2019, 5). 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. doi: 10.5194/gmd-12-1909-2019

Walther, G. R. (2010). Community and ecosystem responses to recent climate change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1549), 2019–2024. doi: 10.1098/rstb.2010.0021

Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008, 6). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, *22*(2), n/a-n/a. doi: 10.1029/2007GB003078

Wiedinmyer, C., Tie, X., Guenther, A., Neilson, R., & Granier, C. (2006). Future changes in biogenic isoprene emissions: How might they affect regional and global atmospheric chemistry? *Earth Interactions*. doi: 10.1175/EI174.1

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., ... Xavier, P. K. (2018, 2). 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*(2), 357–380. doi: 10.1002/2017MS001115

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., Morcrette, C., & Bodas-Salcedo, A. (2008). PC2: A prognostic cloud fraction and condensation scheme. Part II: Results of climate model simulations. *Submitted to Q. J. R.*

Meteorol. Soc..

Wiltshire, A. J., Duran Rojas, C., Edwards, J., Gedney, N., Harper, A. B., Hartley, A., ... Smout-Day, K. (2019, 7). JULES-GL7: The Global Land Configuration of the Joint UK Land Environment Simulation version 7.0. *Geoscientific Model Development Discussions*, 1–34. doi: 10.5194/gmd-2019-152

Wood, N., Staniforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., ... Thuburn, J. (2014). An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Quarterly Journal of the Royal Meteorological Society*, *140*(682), 1505–1520. doi: 10.1002/qj.2235

Woodward, S. (2011). *Mineral dust in HadGEM2, Tech Report 87* (Tech. Rep.). Exeter: Hadley Centre, Met Office.

World Meteorological Organization. (2018). Executive Summary: Scientific Assessment of Ozone Depletion: 2018. In *Global ozone research and monitoring project - report no. 58* (p. 67pp).

WOUDC. (2014). *WOUDC Ozone Monitoring Community, World Meteorological Organization-Global Atmosphere Watch Program (WMO-GAW)/World Ozone and Ultraviolet Radiation Data Centre (WOUDC) [Data]*. Retrieved from `http://www.woudc.org` doi: 10.14287/10000001

Yool, A., Popova, E. E., & Anderson, T. R. (2013, 10). 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. doi: 10.5194/gmd-6-1767-2013

Young, P. J., Arneth, A., Schurgers, G., Zeng, G., & Pyle, J. A. (2009). The CO2 inhibition of terrestrial isoprene emission significantly affects future ozone projections. *Atmospheric Chemistry and Physics*. doi: 10.5194/acp-9-2793-2009

Zender, C. S., Newman, D., & Torres, O. (2003, 9). Spatial heterogeneity in aeolian erodibility: Uniform, topographic, geomorphic, and hydrologic hypotheses. *J. Geophys. Res.*, *108*(D17), 4543+. doi: 10.1029/2002jd003039

Zhang, L., Delworth, T. L., Cooke, W., & Yang, X. (2018, 12). Natural variability of Southern Ocean convection as a driver of observed climate trends. *Nature Climate Change*, 1. doi: 10.1038/s41558-018-0350-3

Zhang, Y., Rossow, W. B., Lacis, A. A., Oinas, V., & Mishchenko, M. I. (2004).

Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research D: Atmospheres*, *109*(19). doi: 10.1029/2003JD004457

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

Figure 1.

GC3.1 C3 grass LAI: JJA
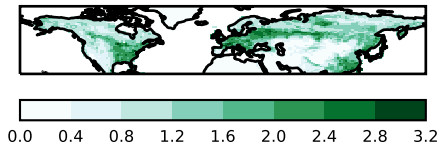
UKESM1 C3 grass LAI: JJA
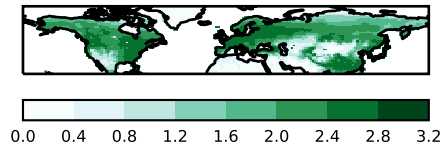
GC3.1 C3 grass LAI: DJF
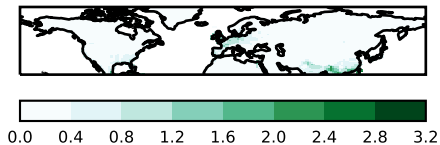
UKESM1 C3 grass LAI: DJF
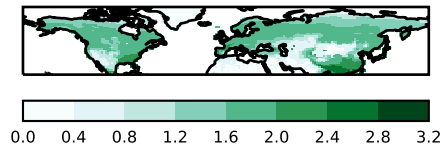
GC3.1 C3 grass: JJA minus DJF (%)
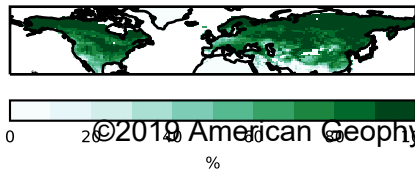
UKESM1 C3 grass: JJA minus DJF (%)
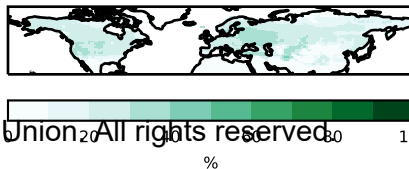
Figure 2.

Clear-sky outgoing shortwave (TOA) for djf
a) UKESM1
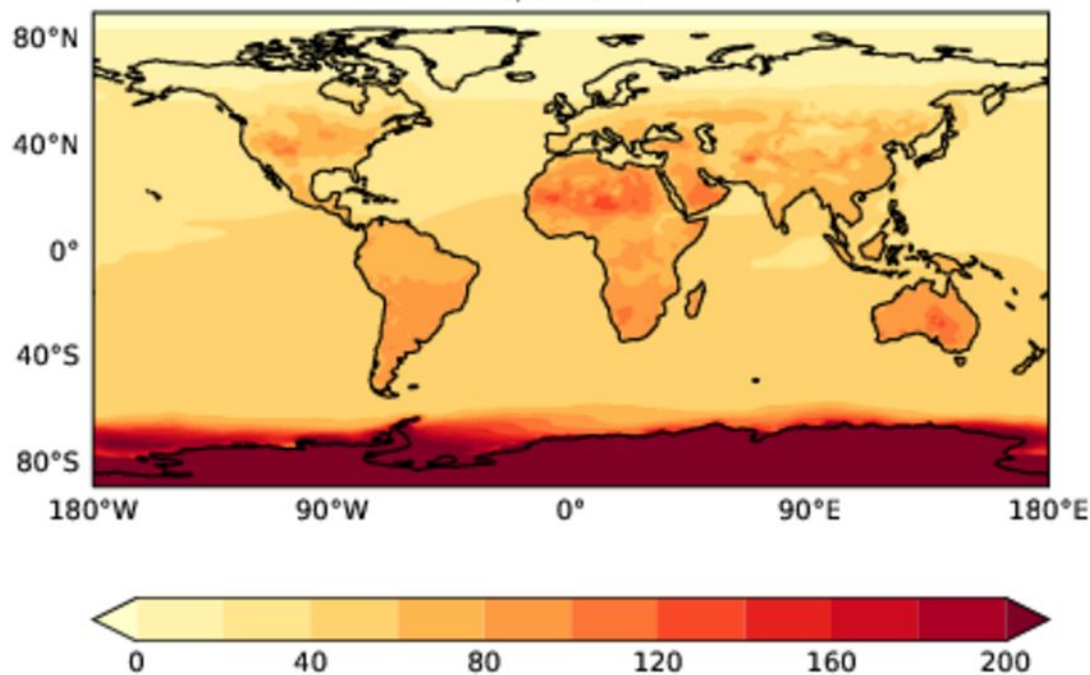
Clear-sky outgoing shortwave (TOA) for djf
b) UKESM1 minus original n_i

Area-weighted rms diff = 4.57
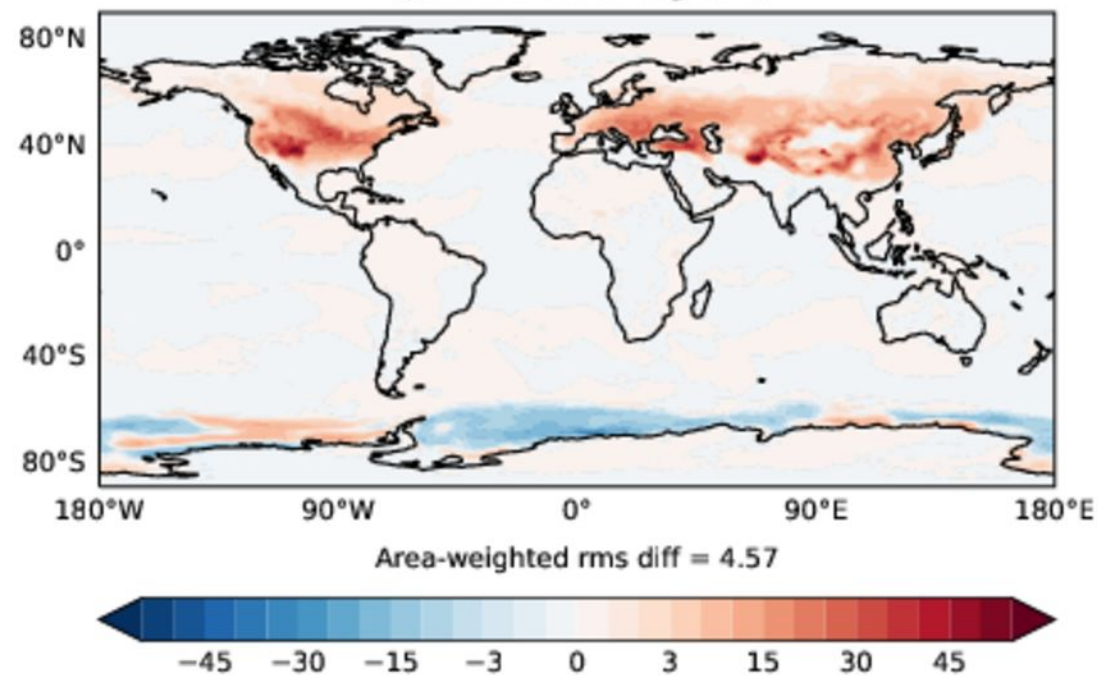
Clear-sky outgoing shortwave (TOA) for djf
c) original n_i minus CERES-EBAF (2000-2010)

Area-weighted rms diff = 8.58
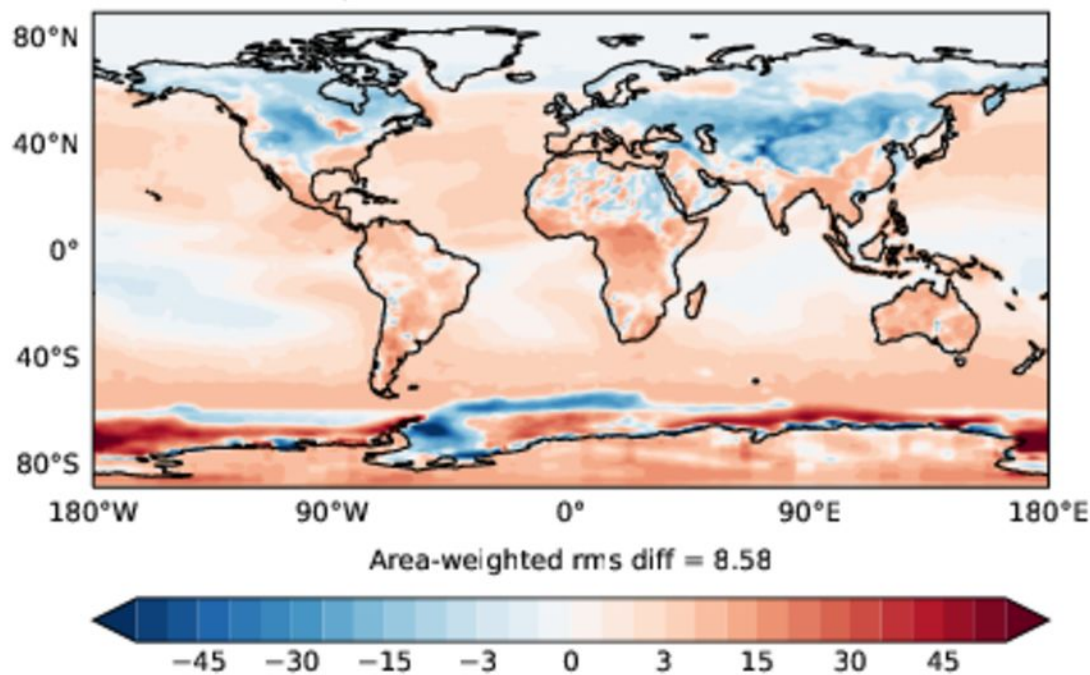
Clear-sky outgoing shortwave (TOA) for djf
d) UKESM1 minus CERES-EBAF (2000-2010)
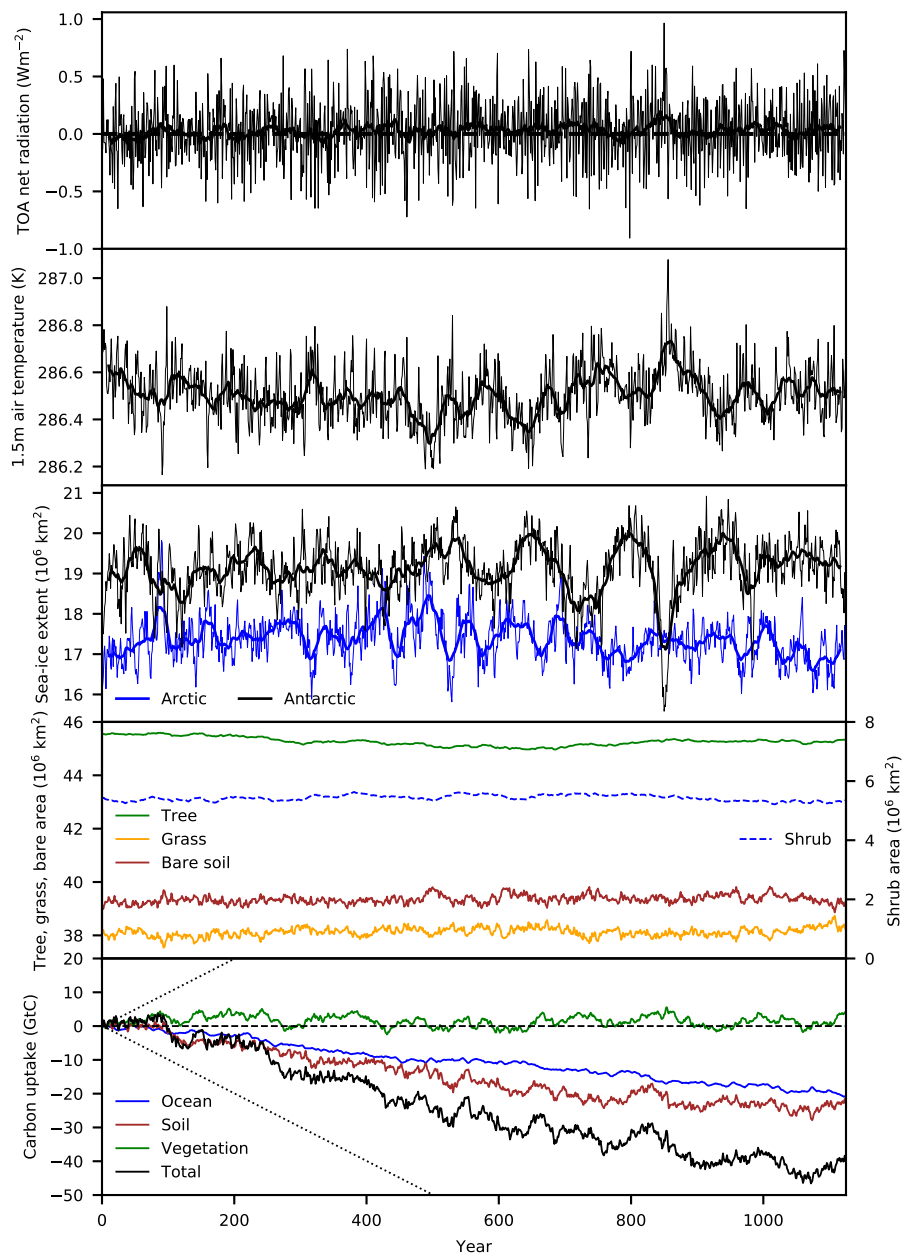
Area-weighted rms diff = 8.49

Figure 3.

Figure 4.

HadISST1.1   UKESM1 piControl

a

b

−0.45   −0.30   −0.15   0.00   0.15   0.30   0.45

K

−0.45   −0.30   −0.15   0.00   0.15   0.30   0.45

K

Figure 5.
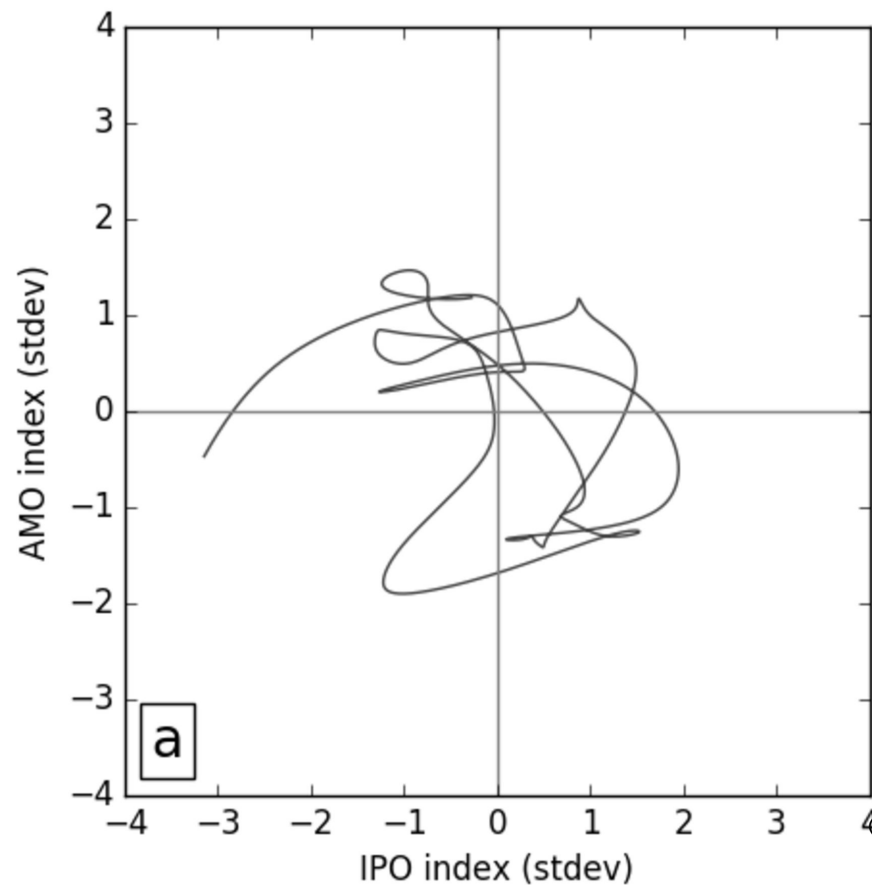
HadISST1.1   UKESM1 piControl

Figure 6.

HadISST1.1

UKESM1 piControl

a

b

IPO index (stdev)

AMO index (stdev)
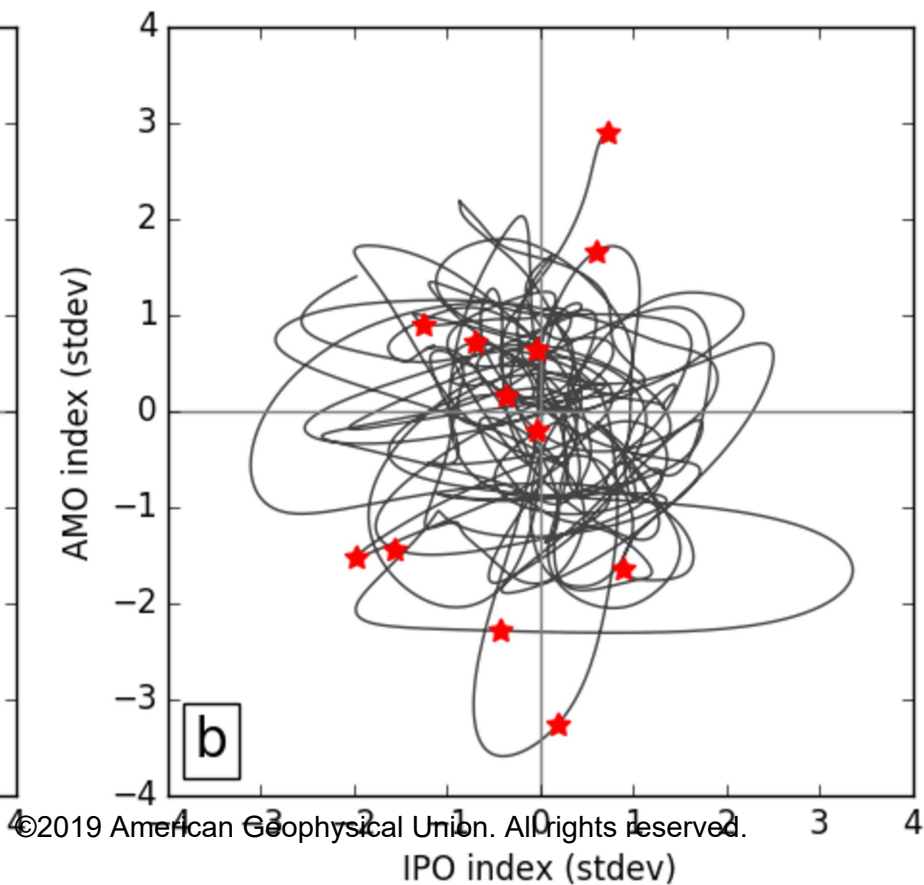
Figure 7.
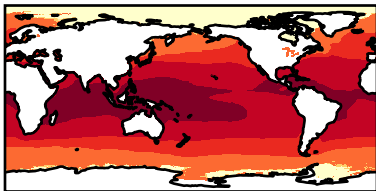
HadISST, DJF HadISST, JJA
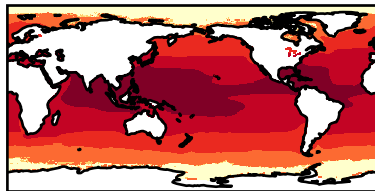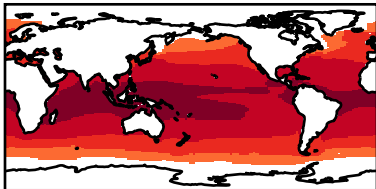
UKESM1, DJF UKESM1, JJA
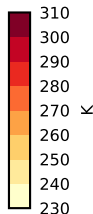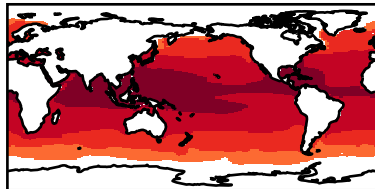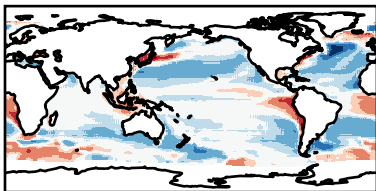
UKESM1 - HadISST, DJF UKESM1 - HadISST, JJA

K

K

RMSE = 0.98 K RMSE = 1.01 K
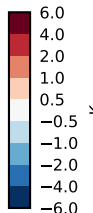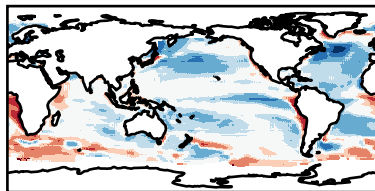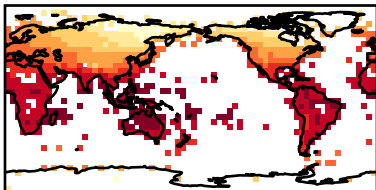
Figure 8.

CRUTEM, DJF   CRUTEM, JJA
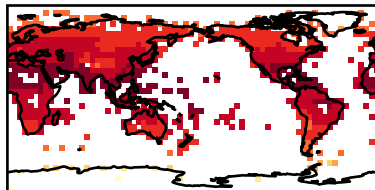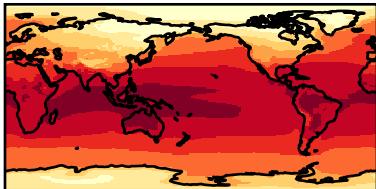
UKESM1, DJF   UKESM1, JJA
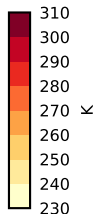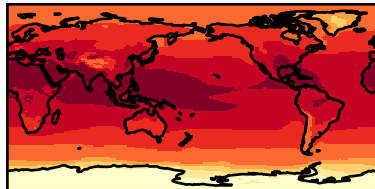
UKESM1 - CRUTEM, DJF   UKESM1 - CRUTEM, JJA

RMSE = 3.42 K   RMSE = 1.81 K
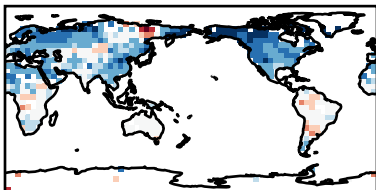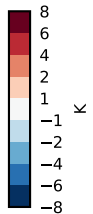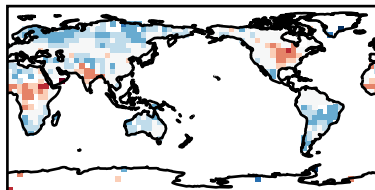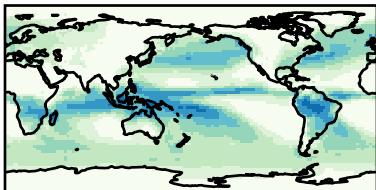
Figure 9.
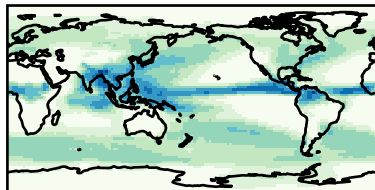
GPCP, DJF GPCP, JJA
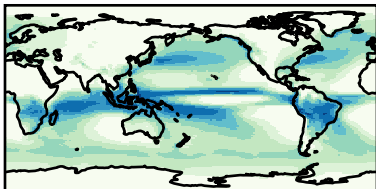
UKESM1, DJF UKESM1, JJA
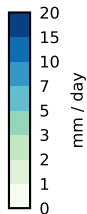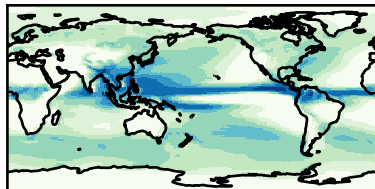
UKESM1 - GPCP, DJF UKESM1 - GPCP, JJA

mm / day

mm / day

RMSE = 1.67 mm/day RMSE = 1.67 mm/day

Figure 10.

Figure 11.

Figure 12.

Area where SWE is greater than 2 mm

Area where SWE is greater than 10 mm

Area where SWE is greater than 40 mm

Area where SWE is greater than 100 mm

Legend:
- GlobSnow-mean
- GlobSnow-max
- UKESM1
- Range GlobSnow-mean
- Range GlobSnow-max
- Range UKESM1

Figure 13.

Figure 14.

Arctic - September

OBS =	7.82
MOD =	8.76
IIEE =	1.59

HadISST.2.2.0.0

Arctic - March

OBS =	17.05
MOD =	17.84
IIEE =	2.02

HadISST.2.2.0.0

Antarctic - September

OBS =	20.68
MOD =	19.66
IIEE =	2.56

HadISST.2.2.0.0

Antarctic - March
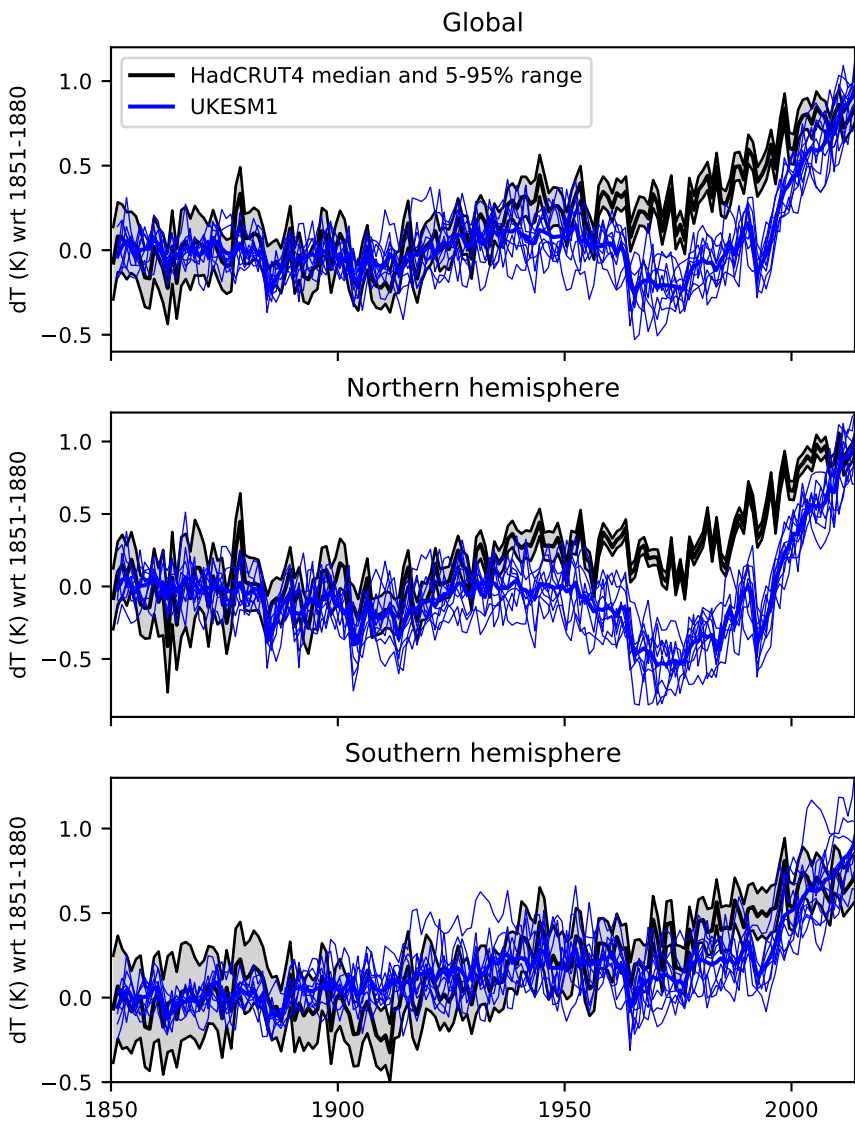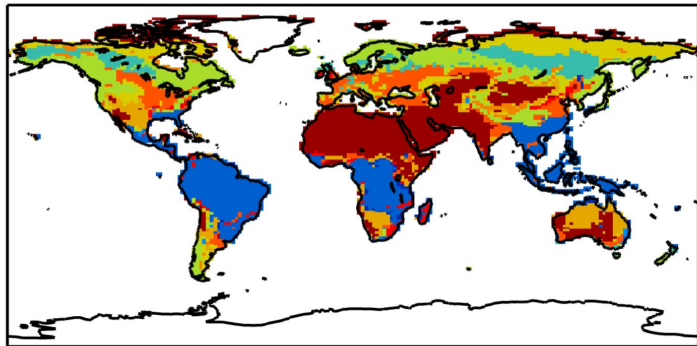
OBS =	6.45
MOD =	6.25
IIEE =	1.32

HadISST.2.2.0.0

Figure 15.

Figure 16.

Accepted Article

Figure 17.
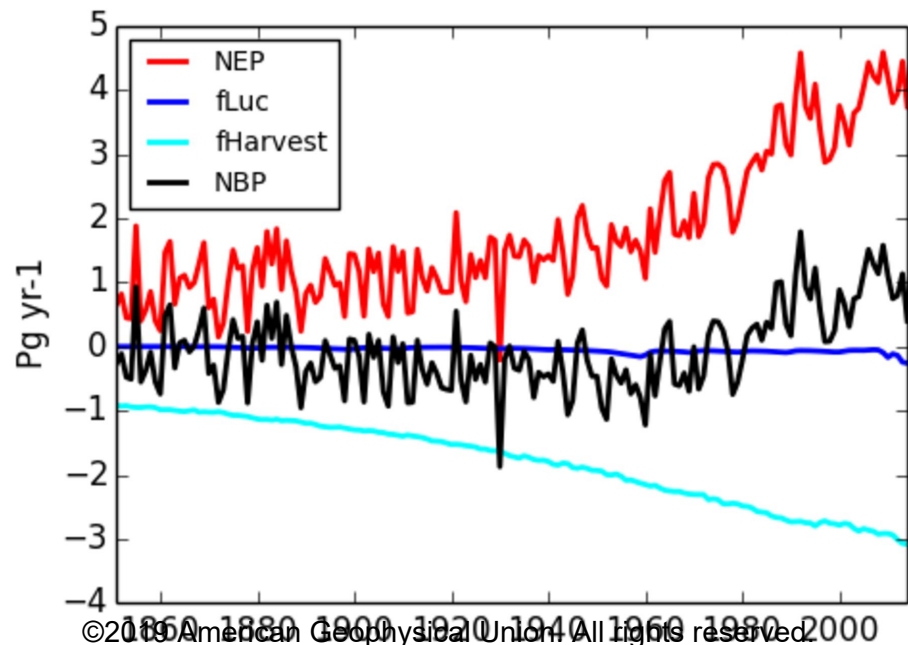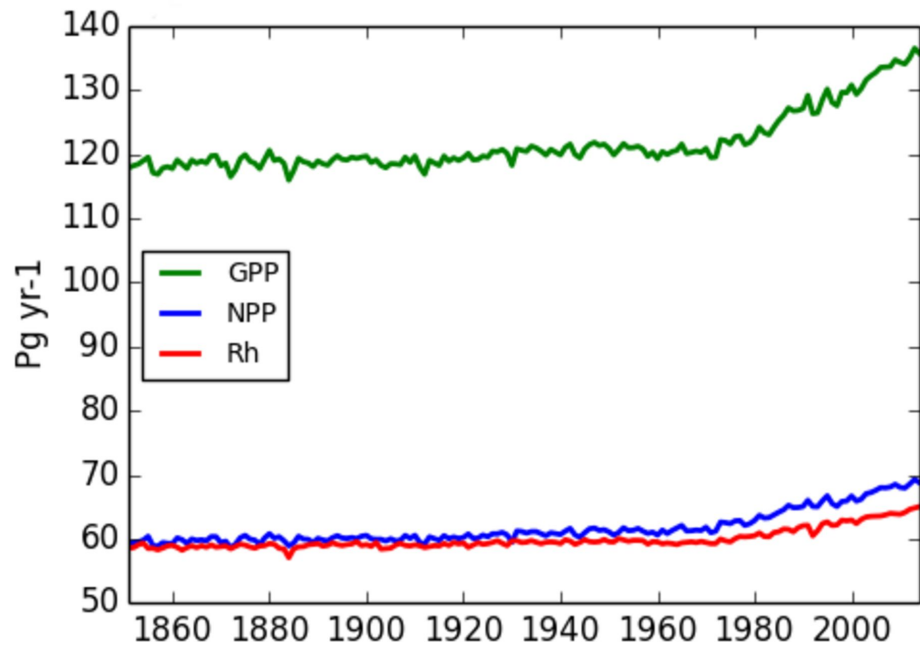
Figure 18.
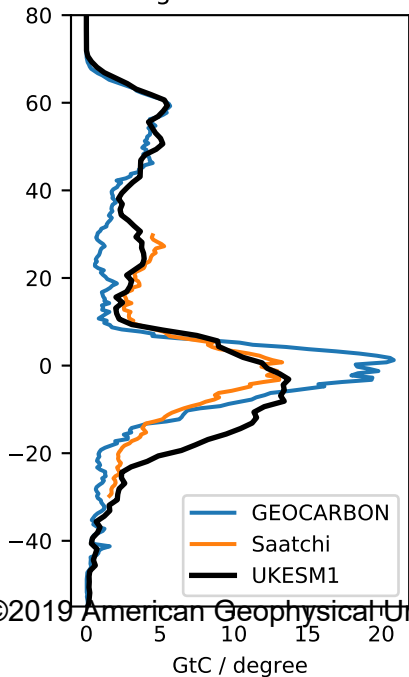
Vegetation carbon — Soil carbon
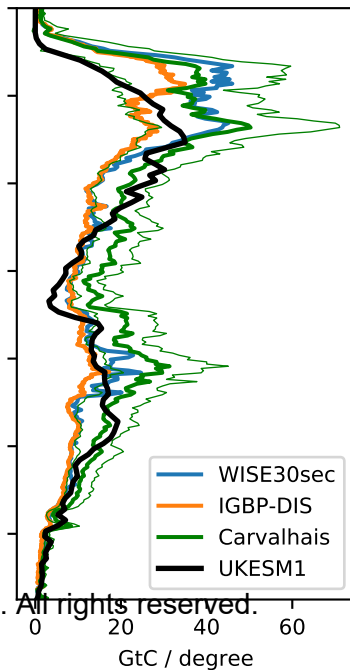
Figure 19.
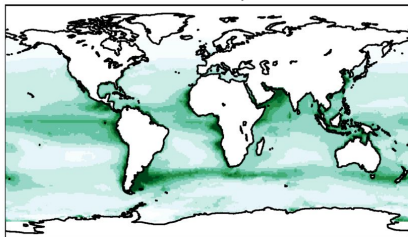
| Observed, DJF | Simulated, DJF |
| Observed, JJA | Simulated, JJA |
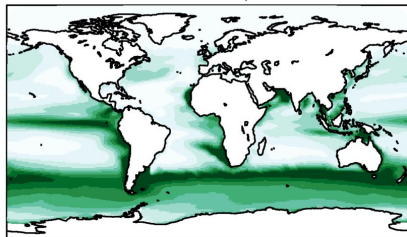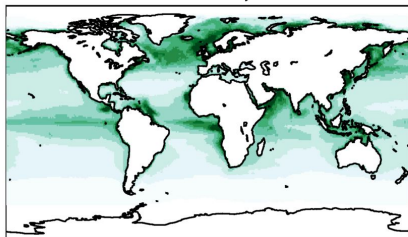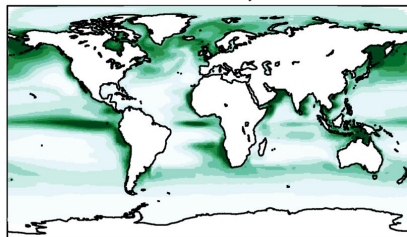
Net primary production [g C m^{-2} d^{-1}]

Figure 20.
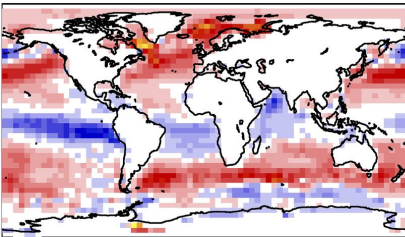
Observed   Simulated
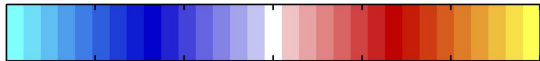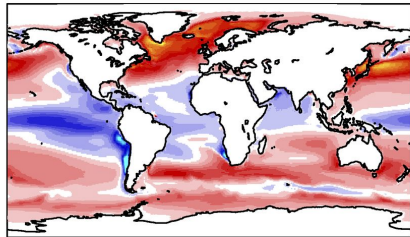
Air−sea CO_2 flux [mol C m^{-2} month^{-1}]

Figure 21.
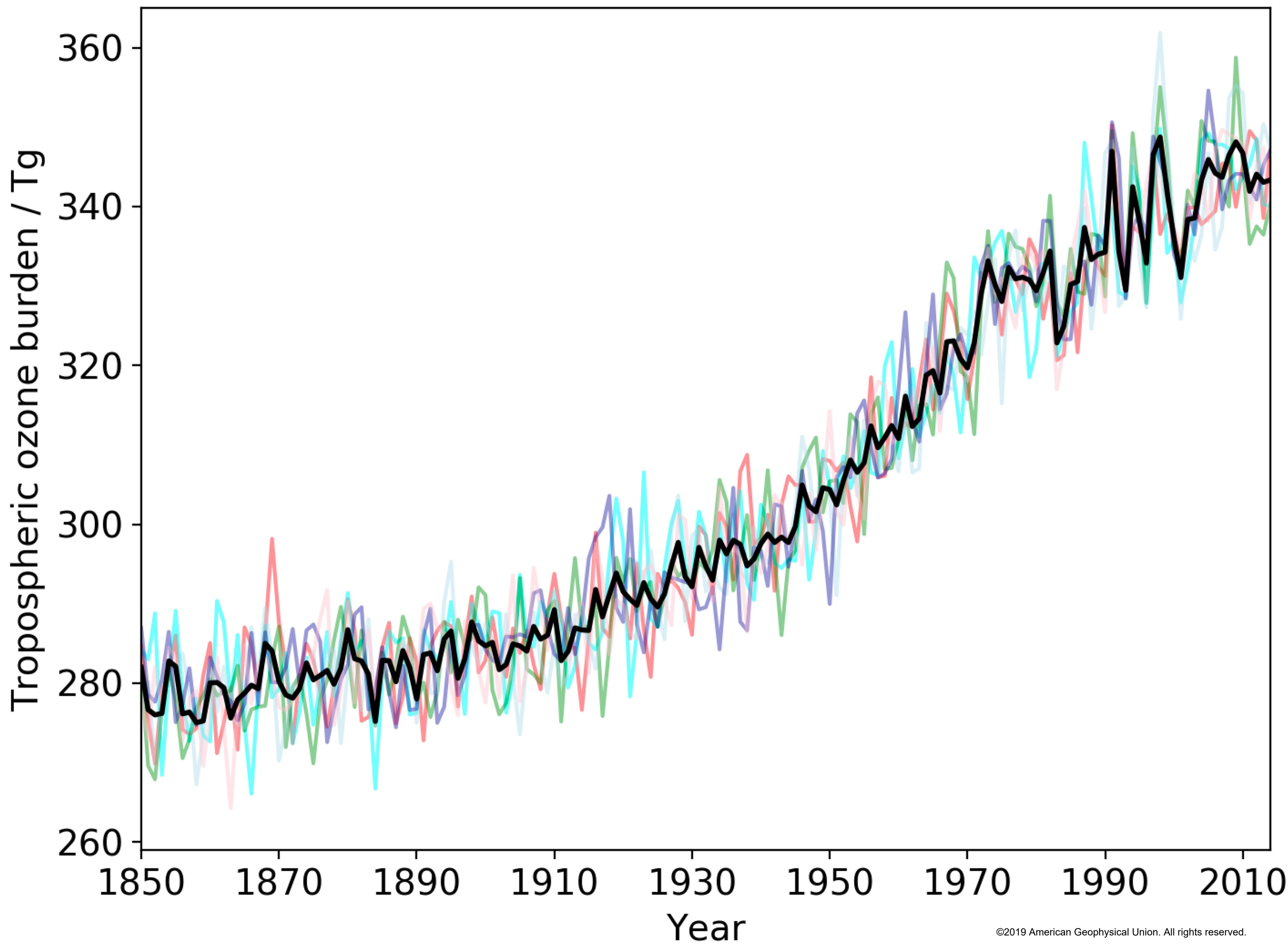
Figure 22.

South Pole TCO October

South Pole TCO November

South Pole TCO December

South Pole TCO January

South Pole TCO February

Figure 23.

Figure 24.

Figure 25.

Monoterpene emissions

Mean = 0.00608009

Isoprene emissions

Mean = 0.0274958
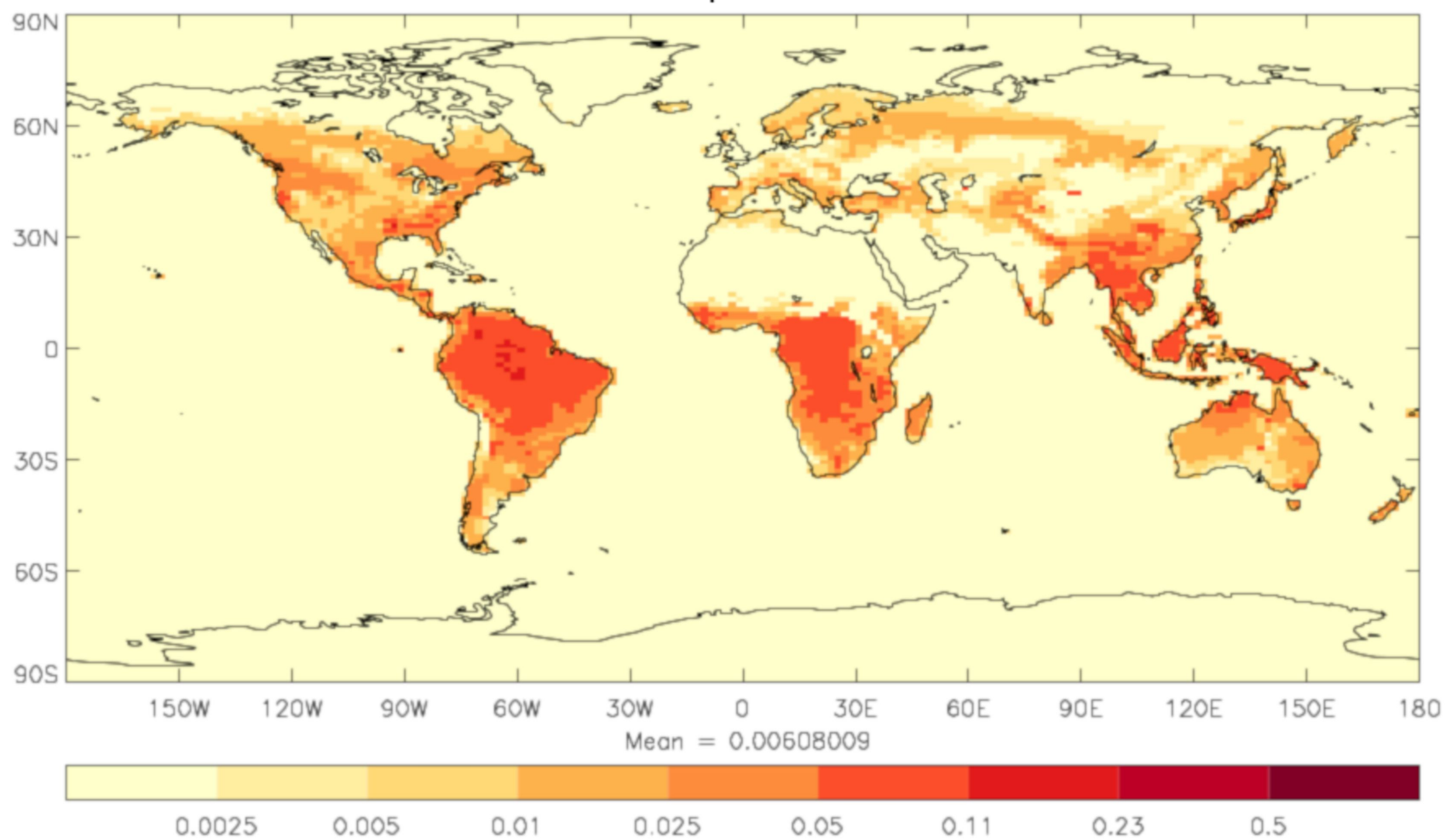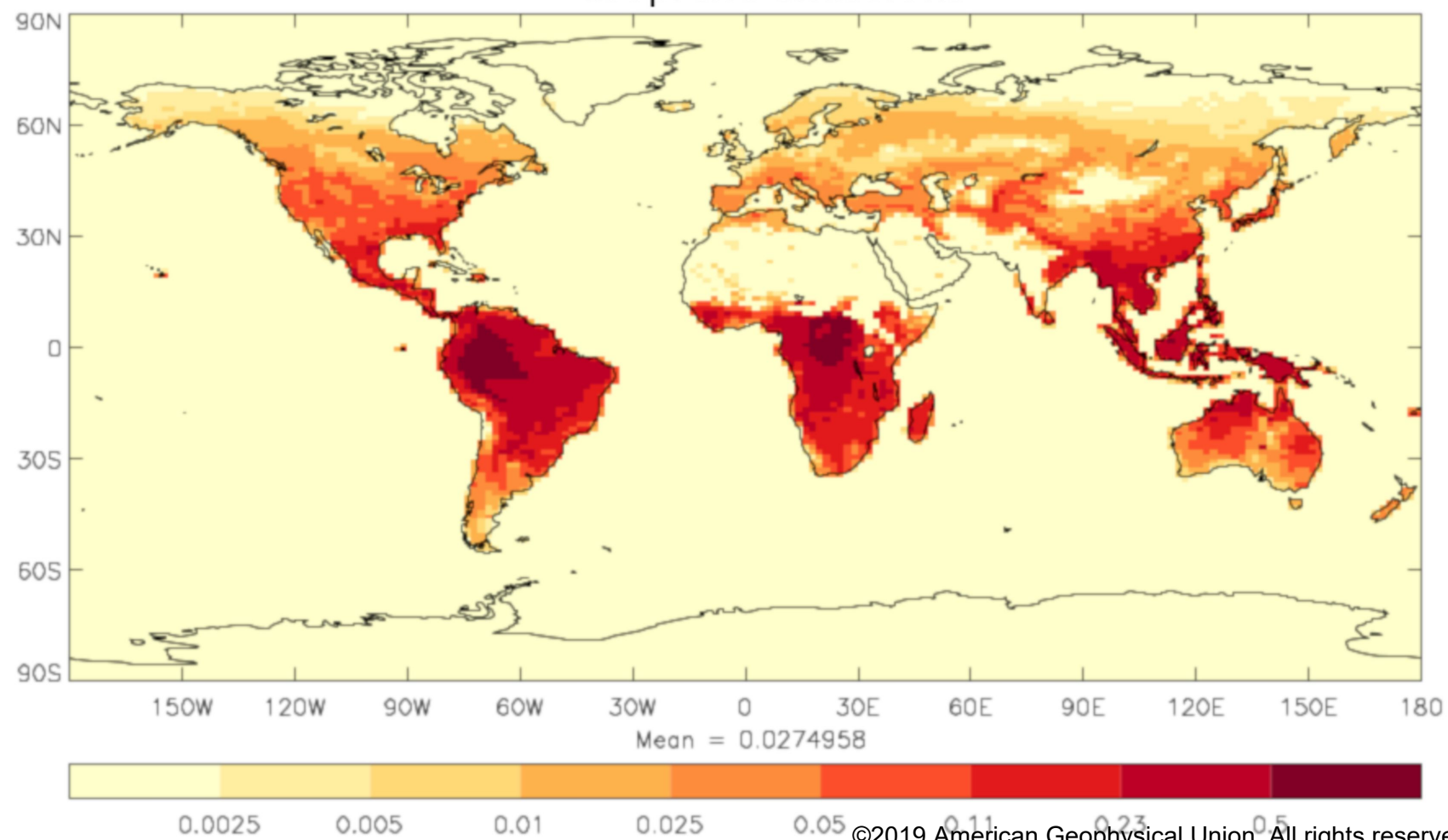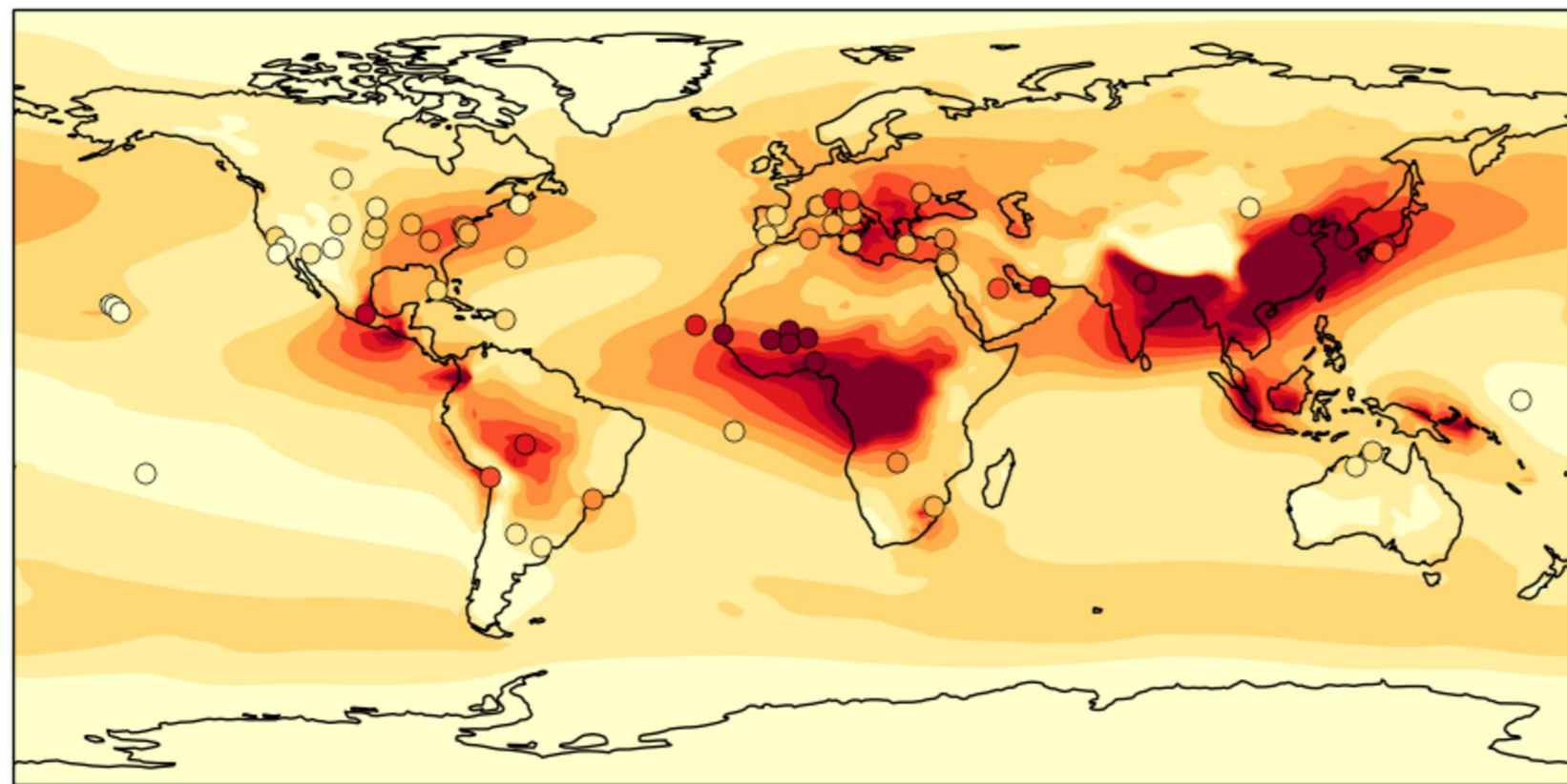
Figure 26.

MEAN: 0.175. RMSE v Aeronet: 0.100

Figure 27.

Figure 28.

Figure 29.

Global

Northern hemisphere

Southern hemisphere

dT (K) wrt 1851-1880

HadCRUT4 median and 5-95% range
UKESM1

Figure A1.

BLD BLETr BLETe NLD NLE C3G C4G Sh Cr Pa BS

Figure A2.

a) HadGEM2-ES: j(O^{1}D) / 1.0e6s-1

b) HadGEM2-ES: j(NO$_2$) / 1.0e4s-1

c) UKESM1: j(O^{1}D) / 1.0e6s-1

d) UKESM1: j(NO$_2$) / 1.0e4s-1

e) UKESM1 vs HadGEM2 j(O^{1}D)

Linear fit

1:1

$y = 1.3x + 1.3$
$r^2 = 0.97$

f) UKESM1 vs HadGEM2 j(NO$_2$)

Linear fit

1:1

$y = 1.18x + 3.79$
$r^2 = 0.96$

Seasonal cycle of CO_2 at Mauna Loa

Seasonal cycle of CO_2 at Point Barrow

Figure C2.
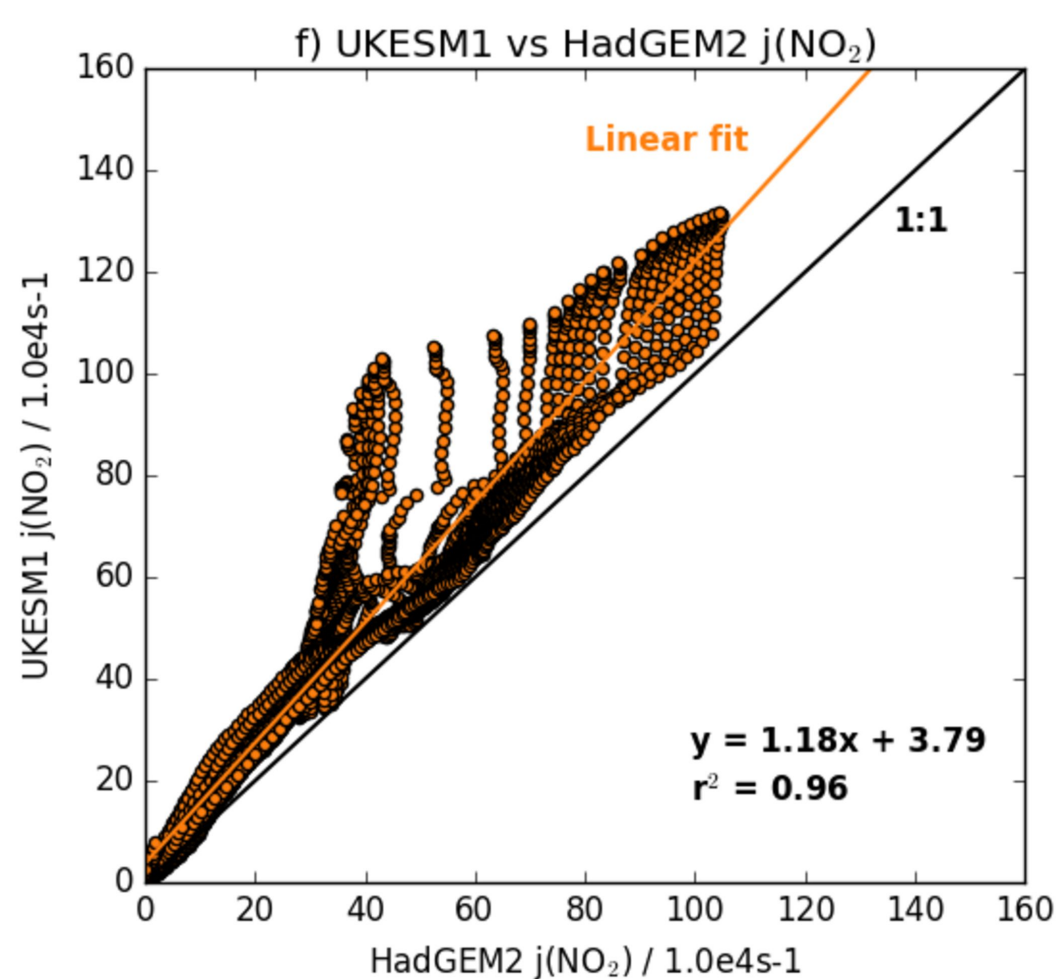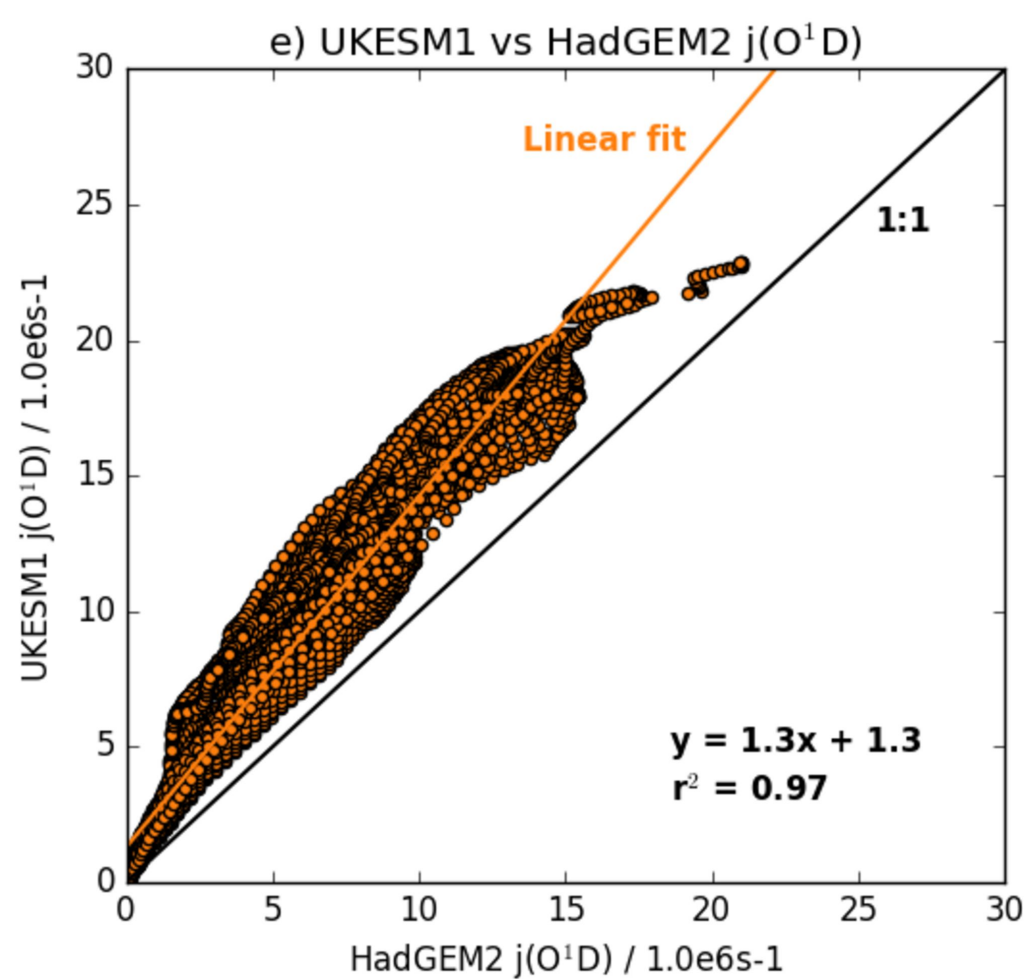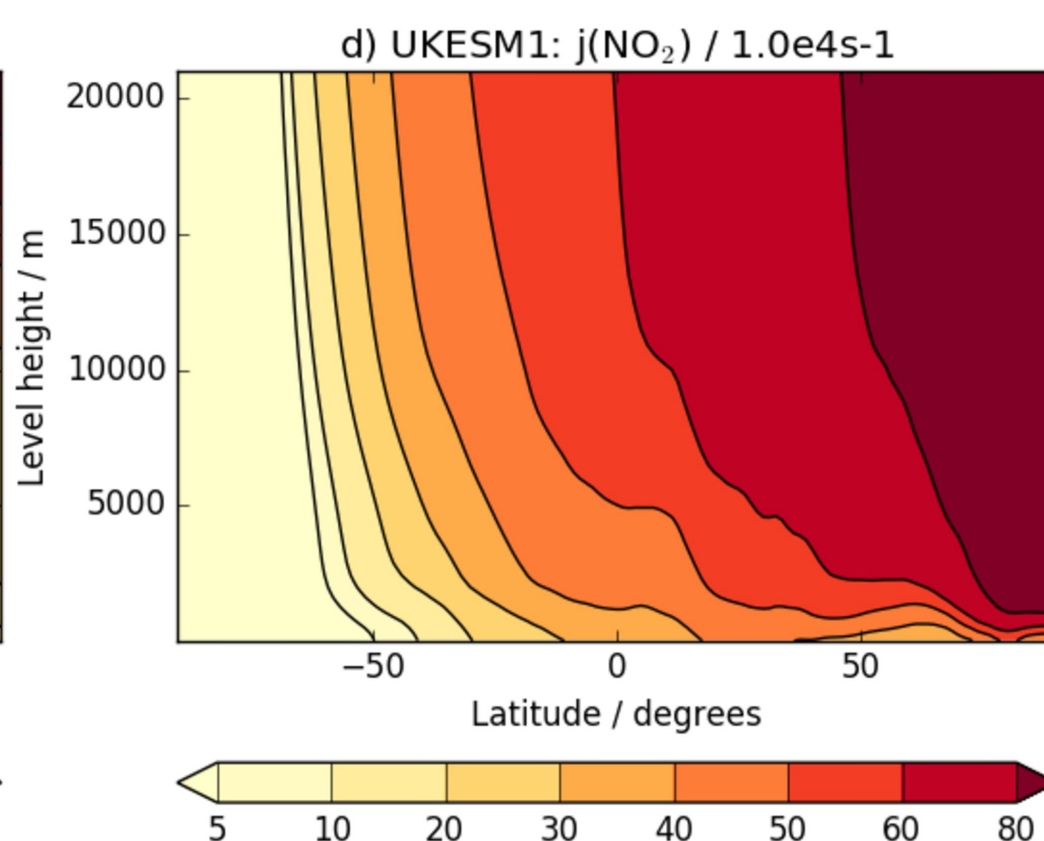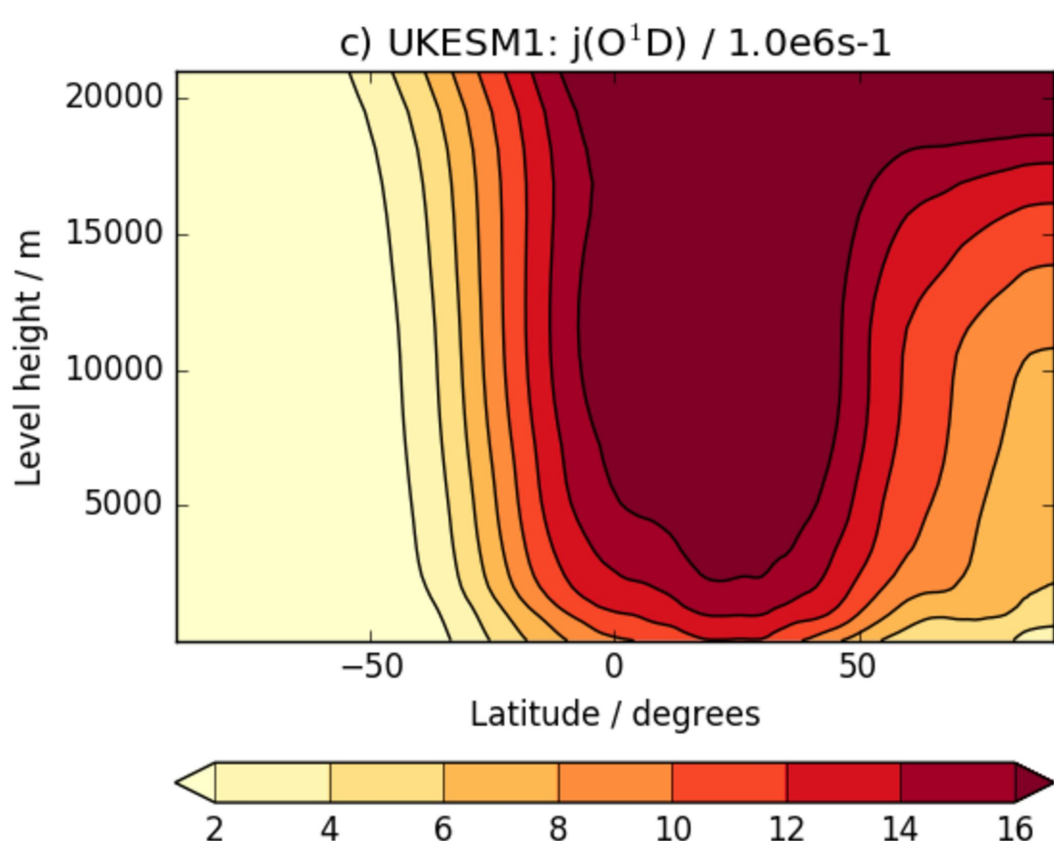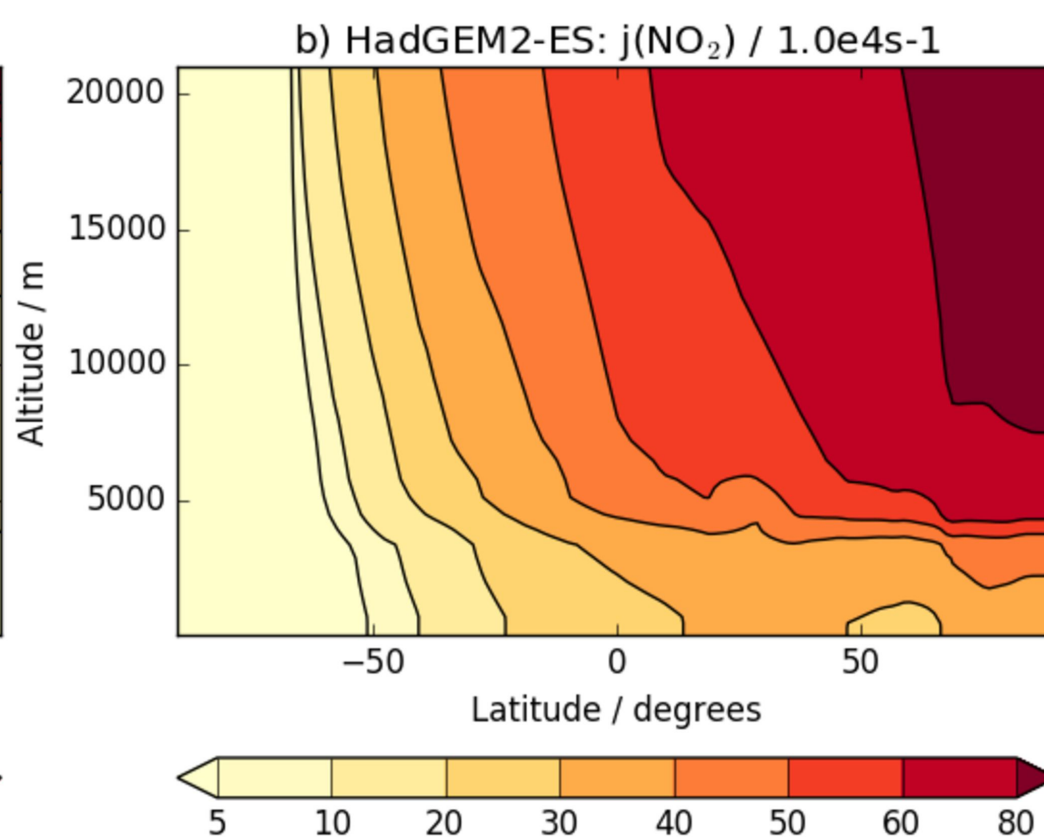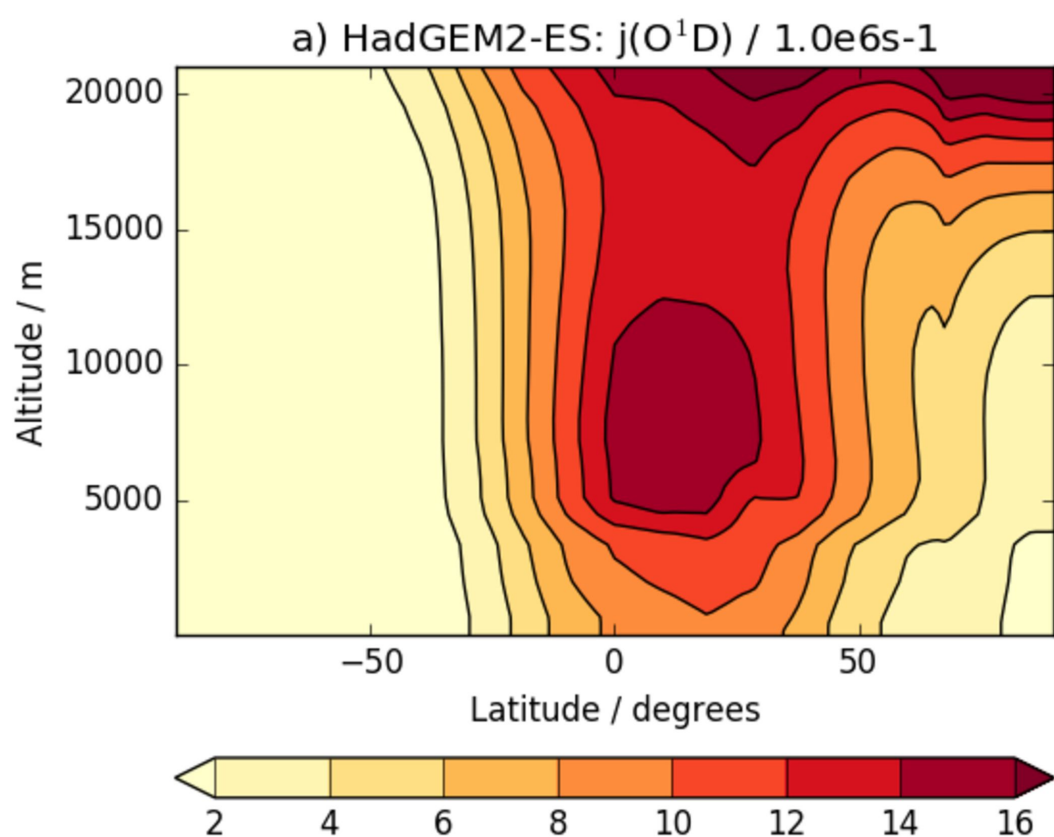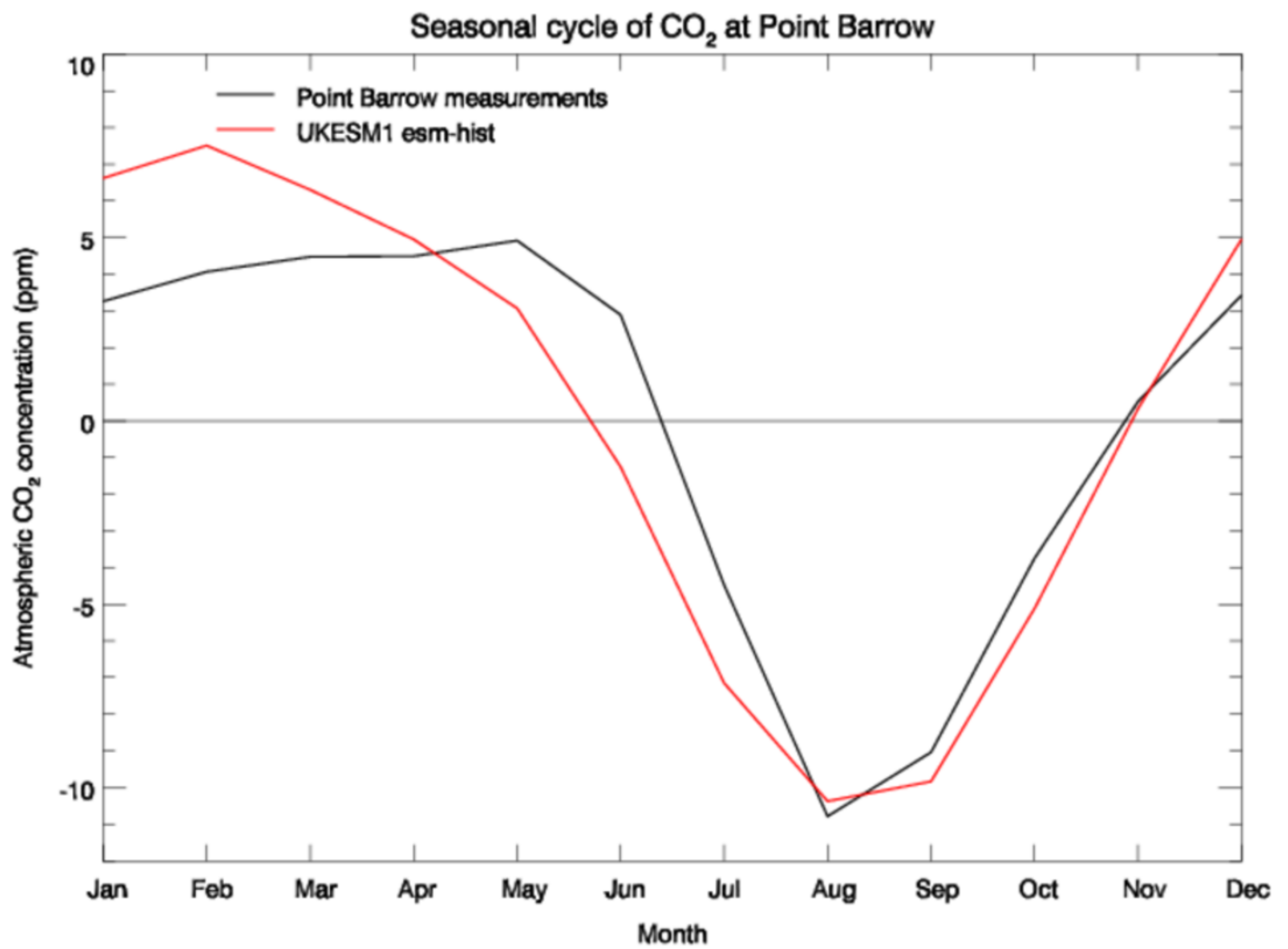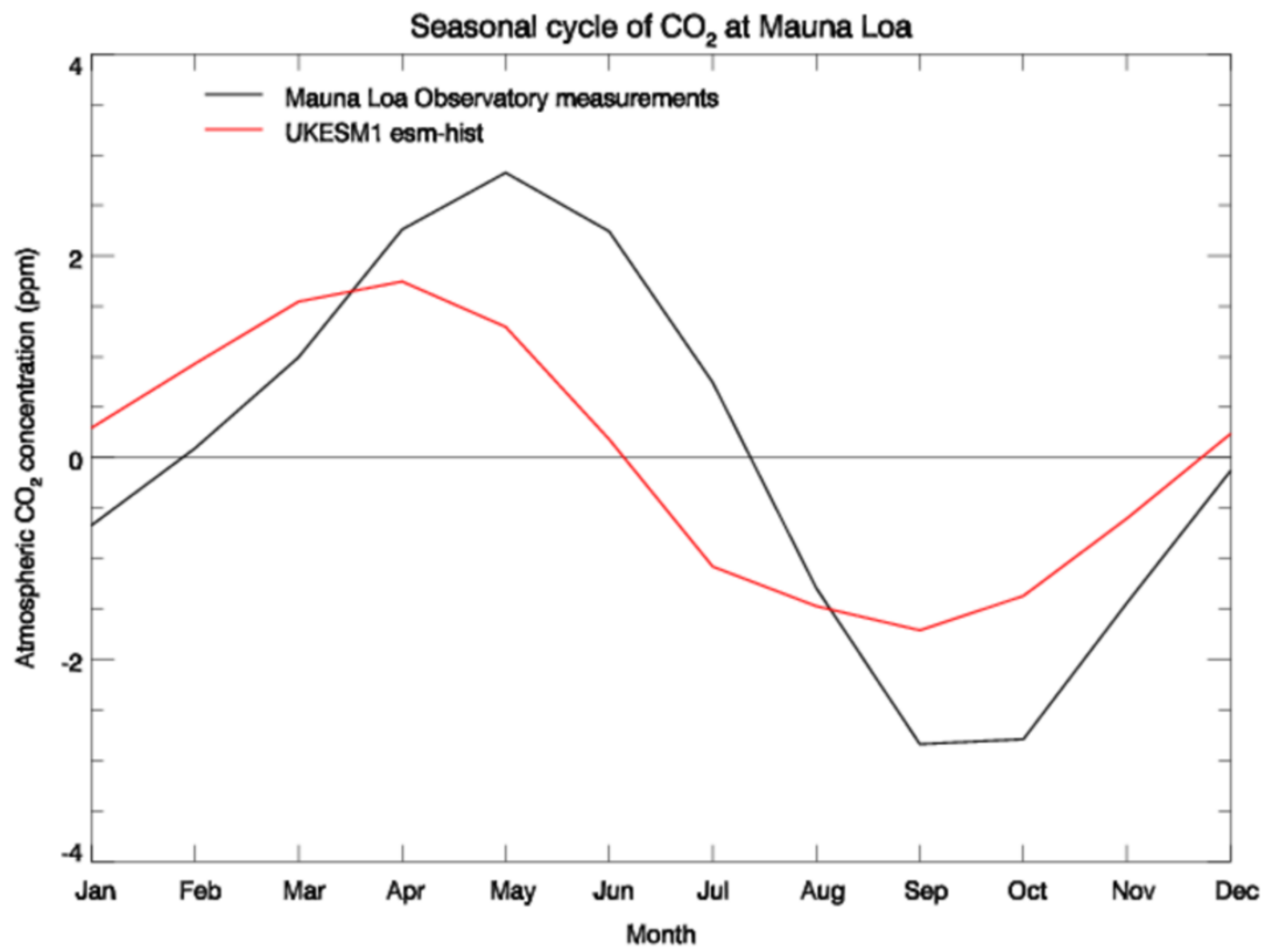
Accepted Article

UKESM1: CO$_2$ concentration at Mauna Loa

UKESM1 esm-piControl and esm-hist

Observations from Mauna Loa Observatory

ppm