AGU PUBLICATIONS

Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1002/2017MS001115

The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations

K. D. Williams[1], D. Copsey[1], E. W. Blockley[1], A. Bodas-Salcedo[1], D. Calvert[1], R. Comer[1], P. Davis[1], T. Graham[1], H. T. Hewitt[1], R. Hill[1], P. Hyder[1], S. Ineson[1], T. C. Johns[1], A. B. Keen[1], R. W. Lee[2], A. Megann[3], S. F. Milton[1], J. G. L. Rae[1], M. J. Roberts[1], A. A. Scaife[1], R. Schiemann[2], D. Storkey[1], L. Thorpe[1], I. G. Watterson[4], D. N. Walters[1], A. West[1], R. A. Wood[1], T. Woollings[5], and P. K. Xavier[1]

[1]Met Office, Exeter, UK, [2]National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK, [3]National Oceanography Centre, Southampton, UK, [4]CSIRO, Aspendale, Australia, [5]Atmospheric Oceanic and Planetary Physics, Oxford, UK

Key Points:
- Description of the Global Coupled 3 (GC3) configuration of the Met Office Unified Model
- A cross-time-scale evaluation of the GC3 configuration is presented
- Overall, GC3 is an improvement on previous configurations

Correspondence to:
K. Williams,
keith.williams@metoffice.gov.uk

Citation:
Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., . . . Xavier, P. K. (2017). The Met Office Global Coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, *9*. https://doi.org/10.1002/2017MS001115

Received 13 JUL 2017
Accepted 5 DEC 2017
Accepted article online 19 DEC 2017

Abstract The Global Coupled 3 (GC3) configuration of the Met Office Unified Model is presented. Among other applications, GC3 is the basis of the United Kingdom's submission to the Coupled Model Intercomparison Project 6 (CMIP6). This paper documents the model components that make up the configuration (although the scientific descriptions of these components are in companion papers) and details the coupling between them. The performance of GC3 is assessed in terms of mean biases and variability in long climate simulations using present-day forcing. The suitability of the configuration for predictability on shorter time scales (weather and seasonal forecasting) is also briefly discussed. The performance of GC3 is compared against GC2, the previous Met Office coupled model configuration, and against an older configuration (HadGEM2-AO) which was the submission to CMIP5. In many respects, the performance of GC3 is comparable with GC2, however, there is a notable improvement in the Southern Ocean warm sea surface temperature bias which has been reduced by 75%, and there are improvements in cloud amount and some aspects of tropical variability. Relative to HadGEM2-AO, many aspects of the present-day climate are improved in GC3 including tropospheric and stratospheric temperature structure, most aspects of tropical and extratropical variability and top-of-atmosphere and surface fluxes. A number of outstanding errors are identified including a residual asymmetric sea surface temperature bias (cool northern hemisphere, warm Southern Ocean), an overly strong global hydrological cycle and insufficient European blocking.

1. Introduction

The Met Office produces forecasts across a range of time scales from numerical weather predictions (NWPs) for days ahead or less, through monthly seasonal-decadal forecasts, to climate change projections. Using the unified coupled model framework described by Hewitt et al. (2011), model development now progresses on a quasi-annual cycle with a new configuration of the coupled atmosphere-land-ocean-sea ice model (and components, e.g., atmosphere-land for short-range NWP) being released every 1–2 years for use across prediction time scales by the Met Office and its collaborators (Walters et al., 2011). This seamless approach has yielded numerous benefits including improved ability to investigate model systematic errors, greater scientific robustness of the model, and a more efficient use of resources (scientists, high performance computing, etc.; Brown et al., 2012).

The latest configuration of the coupled model, released March 2016, is known as Global Coupled configuration 3.0 (GC3.0). This comprises component configurations Global Atmosphere 7.0 (GA7.0), Global Land 7.0 (GL7.0), Global Ocean 6.0 (GO6.0), and Global Sea Ice 8.0 (GSI8.0). GA7.0 and GL7.0 are fully documented by Walters et al. (2017b), while GO6.0 is described by Storkey et al. (2017) and GSI8.0 by Ridley et al. (2017). Upon freezing GC3.0, it was found that the anthropogenic aerosol forcing (as a difference in the top-of-atmosphere radiation balance between an atmosphere-only simulation with 1,860 forcings and a simulation with year 2000 aerosol forcing) was excessively strong to the point where the total anthropogenic effective radiative forcing would almost certainly be negative, failing an acceptance criteria, and a project was initiated to understand and address this through a limited set of physical improvements to the model. The work carried out is fully described in companion papers in this special issue with the main change being

the inclusion of a representation of spectral dispersion in the calculation of the cloud droplet effective radius following Liu et al. (2008). This work resulted in a GA7.1 atmosphere configuration (Walters et al., 2017b). Following on from this, further tuning of the sea ice was undertaken within the coupled model (leading to GSI8.1) and together these changes form GC3.1. The sea ice tuning involved increasing the ice-ocean drag coefficient and increasing the ice thickness for which the bare-ice albedo is reduced to account for the darker ocean below being visible through it. However, the majority of the impact on the sea ice volume and extent between GC3.0 and GC3.1 was caused by the changes to the clouds and aerosols in the development of GA7.1. It should be noted that most of the differences between GC3.0 and GC3.1, both in terms of model changes and resulting systematic errors, are relatively small and localised. Hereafter GC3 will be used to mean collectively GC3.0 and GC3.1 and we only show results from the two where we have found notable differences. It is expected that GC3.1 will be the basis for the UK community's submission to the Coupled Model Intercomparison Project 6 (CMIP6).

In this paper, we provide a technical description of the coupling between the components to complete the model description and then present the coupled model performance in terms of mean bias, variability, and skill. We do not discuss climate projections from GC3 as these will be presented in companion papers in this special issue. The "physical" model presented here does not include earth system components such as interactive vegetation, ice sheets or land, and ocean biogeochemistry. GC3.1 is being developed into an earth system model (UKESM1) and this will be described in companion papers in this special issue.

The present paper documents the impact of changes relative to the last coupled configuration GC2 (Williams et al., 2015), which is currently used operationally at the Met Office for seasonal forecasting and has been used for some climate studies. While GC3 has been developed for use across time scales, and results from climate, seasonal and NWP simulations are presented, we focus on the longer time scale climate simulations (referred to as HadGEM3-GC3) since this will be the physical model configuration submitted to CMIP6. For the same reason, we also compare back to the much older configuration, HadGEM2-AO (Martin et al., 2011), results from which were submitted to CMIP5. Note that under current Met Office nomenclature, GCx refers to the configuration (precisely defining the science settings being run), whereas HadGEMx (Hadley Centre Global Environmental Model) is the climate system name (the way of running the model, its initialization, etc.). This nomenclature was developed after HadGEM2-AO, so HadGEM2-AO defines both the science and system.

In the next section, we provide details of the coupling and experiments subsequently presented. In section 3, the climatological biases and overall skill of the model are discussed, while systematic errors in variability are presented in section 4. We summarize in section 5.

2. Coupled Model Details

The configurations used in this paper are listed in Table 1. The GC3.0 configuration is defined by the combination of the component model scientific configurations (GA7.0, GL7.0, GO6.0, and GSI8.0) and associated

Table 1
List of Coupled Atmosphere-Land-Ocean-Sea Ice Configurations From Which Results Are Presented, Their Component Configurations, the Resolution Used in This Study and References

Coupled config.	Components	Resolution used here atmos/ocean	Component references
GC3	GA7, GL7, GO6, GSI8	N216 (60 km) L85/ORCA025 (0.25°) L75	Walters et al. (2017b), Storkey et al. (2017), Ridley et al. (2017)
GC2	GA6, GL6, GO5, GSI6	N216 (60 km) L85/ORCA025 (0.25°) L75	Walters et al. (2017a), Megann et al. (2014), Rae et al. (2015)
HadGEM2-AO	N/A	N96 (135 km) L38/1° (tropic enhanced) L40	Martin et al. (2011)

GC: Global Coupled
GA: Global Atmosphere
GL: Global Land
GO: Global Ocean
GSI: Global Sea Ice

HadGEM: climate simulation system
GloSea5: seasonal forecast system
NWP: numerical weather prediction

choices about the way these model components are coupled together. The component models are fully documented in the model description sections of Walters et al. (2017b), Storkey et al. (2017), and Ridley et al. (2017), while the technical details of the coupling are described in Appendix A. GC3.1 uses the component models GA7.1 and GSI8.1 in place of GA7.0 and GSI8.0, however, details of the coupling are identical to GC3.0.

Relative to GC2, GC3 has a new modal aerosol scheme, new multilayer snow scheme, introduction of multilayer sea ice and a number of parametrization changes in all the model components, including a set relating cloud and radiation and revision to the numerics of convection (these are all described in the component papers). HadGEM2-AO predates GC2, so relative to HadGEM2-AO, GC3 has numerous additional changes including a new ocean model, new sea ice model, major revision to the atmosphere dynamical core, new cloud scheme, and considerable revisions to all of the existing parametrization schemes.

The vertical resolution is set by the component model definitions, being 85 levels in the atmosphere (with a top at 85 km), 4 soil levels, 75 levels in the ocean (with a 1 m top level), 4 layers in the sea ice with 1 layer of snow on the top and a multilayer terrestrial snow scheme. The GA7 science is run over horizontal resolutions from N96 (135 km in midlatitudes) to N1280 (10 km in midlatitudes) on a regular latitude-longitude grid with no explicit changes to model parametrizations. The ocean uses a tripolar grid and can also be run at multiple resolutions (1°, 1/4°, 1/2°) but requires a small number of explicit changes only related to horizontal resolution (Storkey et al., 2017). Results presented here all use a horizontal resolution of N216 (60 km in midlatitudes) in the atmosphere and ORCA025 (1/4°) in the ocean. There will be a submission of the physical model to CMIP6 at this N216-ORCA025 resolution, however, UKESM1 will initially use the N96-ORCA1 (135 km atmosphere; 1° ocean) resolution (described elsewhere in this special issue). Results shown here from GC2 are at an identical horizontal and vertical resolution to GC3 (the ocean grid in GC3 has been extended from GC2 to encompass ice cavities, but these are not yet implemented, so the grids used are effectively identical). HadGEM2-AO uses an N96 (135 km) L38 atmosphere (with a top at 40 km) and 1° (plus tropically enhanced) L40 ocean.

A full description of the coupling is included in Appendix A for completeness, although many aspects are unchanged from GC2 (Williams et al., 2015) and the original description in Hewitt et al. (2011).

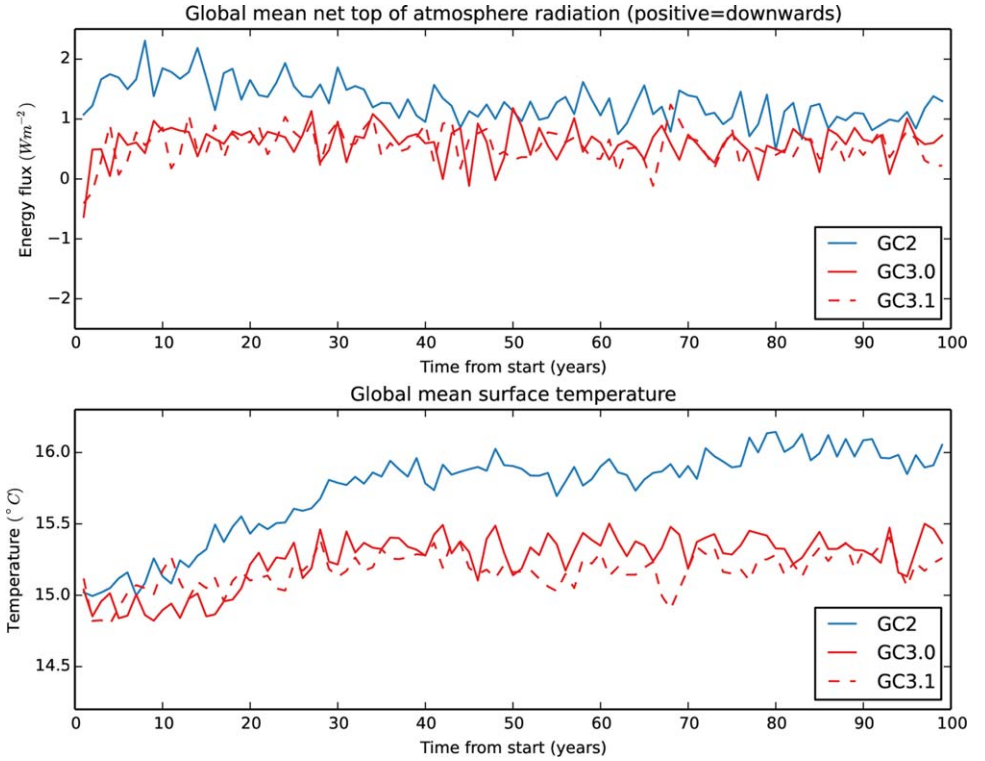

Figure 1. Time series of global-mean net top-of-atmosphere radiative flux and surface temperature from the HadGEM3-GC2, HadGEM3-GC3.0, and HadGEM3-GC3.1 simulations.

2.1. Experimental Design

Results from three types of coupled model experiment are presented in the following sections: (1) a long present-day climate simulation, (2) seasonal hindcasts, and (3) NWP hindcasts.

The majority of results are from the present-day climate experiment, results from which use the "HadGEM" system label (e.g., HadGEM3-GC3). This is a 100 year free-running simulation with forcings set to use values from the year 2000 (this is the same as experiment 2 in CMIP3). Where appropriate (e.g., for aerosol emissions), these forcings vary through the annual cycle. The ocean is initialized from EN3 climatology (Ingleby & Huddleston, 2007). Unless otherwise stated average results over the period 50–100 years after initialization from EN3 are shown. By this time, the sea surface temperatures (SSTs) are in quasi-equilibrium (Figure 1).

Results presented from seasonal hindcasts of GC2 and GC3 use the GloSea5 (Global Seasonal forecasting) system (MacLachlan et al., 2015). The hindcasts are for the years 1993–2015, each of 210 days in length. Within each year, there are three start dates for the DJF hindcasts (25 October, 1 November, and 9 November) and three for JJA (25 April, 1 May, and 9 May). Each start date has an eight-member initial condition ensemble, resulting in 552 hindcasts for each of DJF and JJA. The ocean and sea ice are initialized from Met Office Forecast Ocean Assimilation Model (FOAM; Blockley et al., 2014), the atmosphere from ECMWF (European Centre for Medium Range Weather Forecasts) Interim Reanalysis (ERA-I; Dee et al., 2011), and soil moisture from a climatology of the JULES (Joint UK Land Environment Simulator) land surface model forced with ERA-I. More details on the initialization can be found in MacLachlan et al. (2015).

NWP hindcast results are from cases over the annual period 1 September 2011 to 31 August 2012. Cases are initialized at 00Z every day (which is when the ocean reanalyses are available) and each is 15 days in length. Ocean and sea ice are initialized from reanalyses produced with research versions of the Met Office ocean analysis system (Waters et al., 2015) consistent with GO5 and GSI6 physics (for GC2 runs) or GO6 and GSI8 physics (for GC3 runs) and the corresponding ORCA025 grid. Atmosphere and land are initialized from archived Met Office operational NWP 00Z analyses of the day. Somewhat older (GA3.1 and GL3.1; Walters et al., 2011) global atmosphere and land physics versions were used operationally during 2011–2012; the operational NWP resolution used then was also different from the resolution used here, so the atmosphere analyses are interpolated to the NWP hindcast resolution (N216L85). The GC3 NWP hindcasts, as for seasonal hindcasts, initialize soil moisture and snow from GC3-compatible climatologies, but GC2 hindcasts initialize soil moisture and snow by interpolation from the analyses of the day. Consequently, the ocean and sea ice initial states can broadly be considered "native" to the NWP hindcasts with respect to resolution and

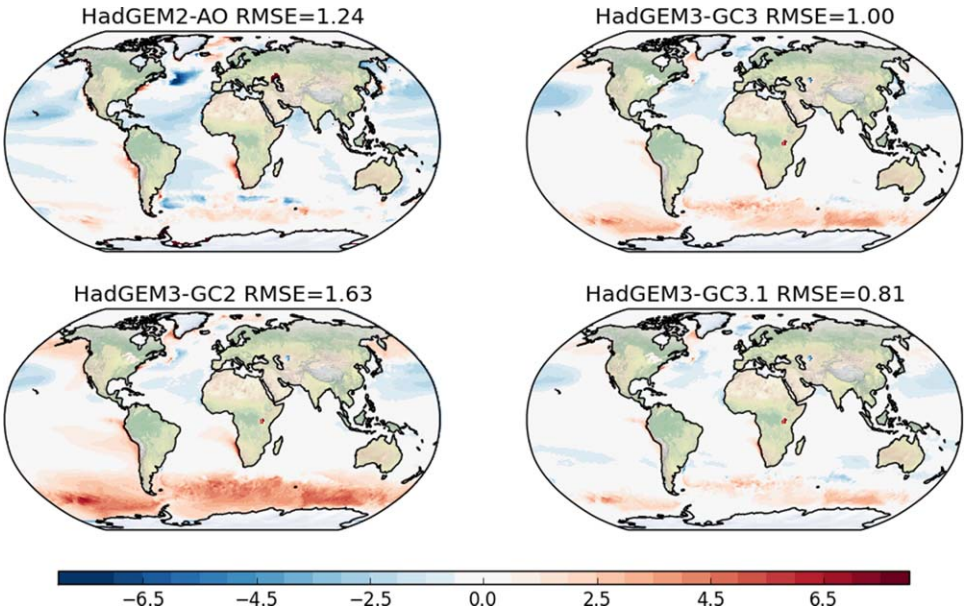

Figure 2. Mean SST bias (K) against ESA CCI (European Space Agency Climate Change Initiative) SST observations (Merchant et al., 2014) for HadGEM2-AO, HadGEM3-GC2, HadGEM3-GC3.0, and HadGEM3-GC3.1.

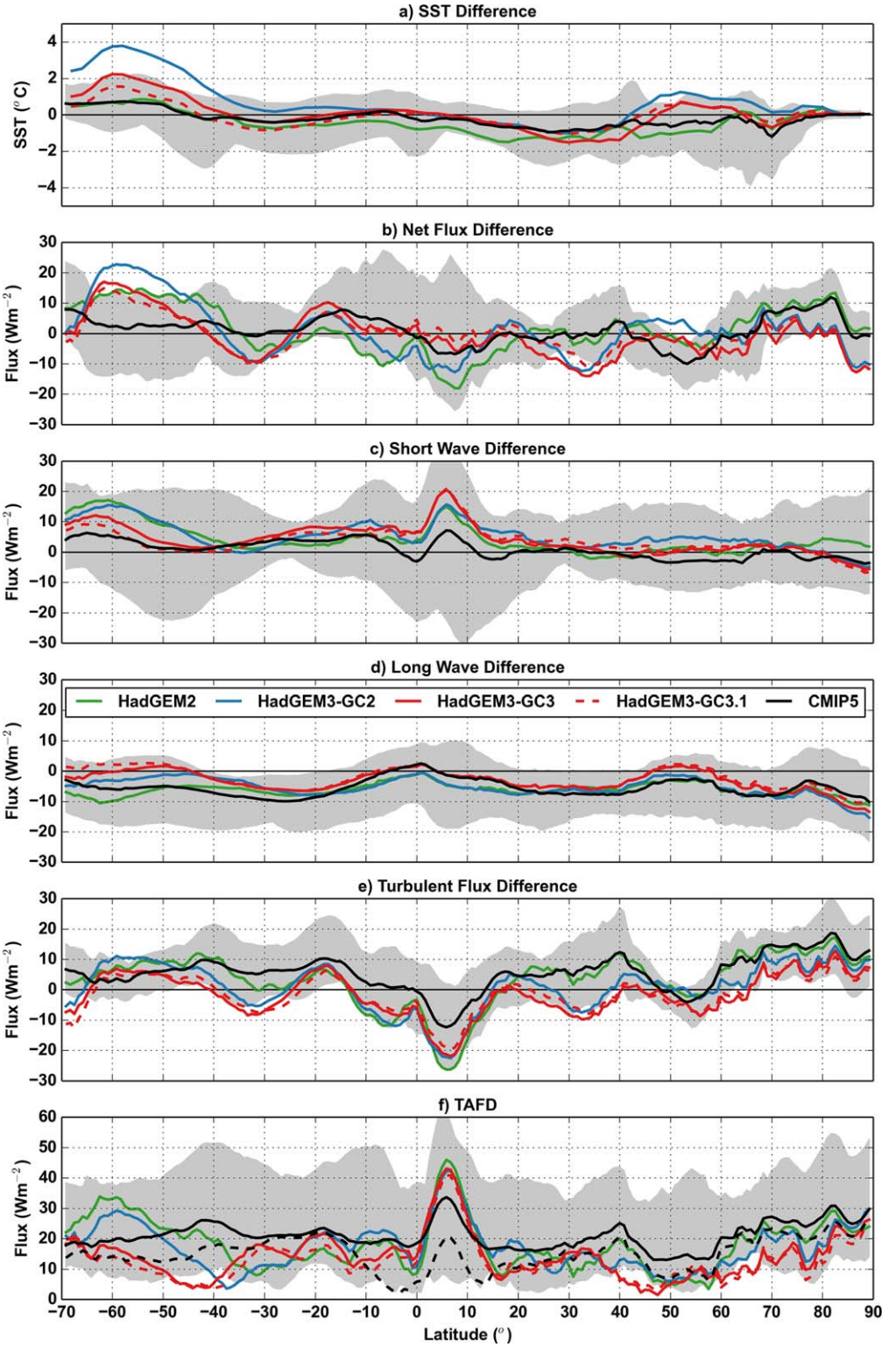

Figure 3. (a) Zonal mean SST bias in the coupled models against ESA CCI SST. (b–e) Zonal mean net downward surface flux bias in AMIP simulations and its decomposition into SW, LW, and turbulent flux errors. The grey shading shows the spread of the CMIP5 ensemble and black is the ensemble mean. The observed surface flux used to calculate the error is Liu et al. (2015). (f) Total absolute flux error, i.e., the sum of the magnitude of the error in each of the components shown in Figures 2c–2e. The solid black line is the mean of the total absolute flux error across the CMIP5 ensemble while the dashed black line is the total absolute flux error of the ensemble mean of the heat fluxes (i.e., using the solid black lines in Figures 2c–2e), which is lower due to the cancellation of errors across the ensemble.

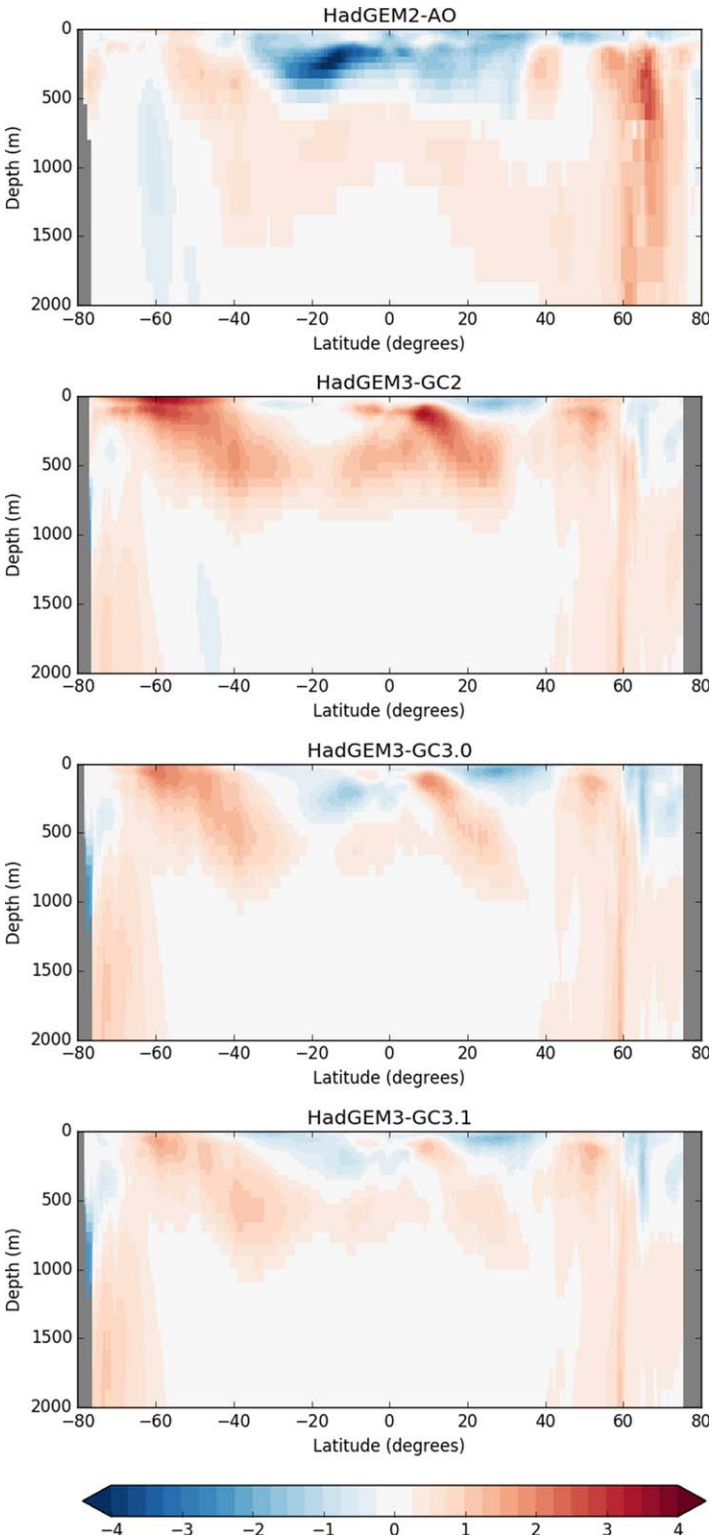

Figure 4. Ocean zonal annual-mean temperature error (K) compared with EN4 (Good et al., 2013) for HadGEM2-AO, HadGEM3-GC2, HadGEM3-GC3.0, and HadGEM3-GC3.1.

GC physics versions, but the same is not true for the atmosphere and land states. Some initialization shock might be expected as a direct consequence of this, but initialization shock can also arise as a result of the initial states themselves being products of separate uncoupled (forced) systems (Mulholland et al., 2015).

Both the NWP and seasonal hindcasts use prescribed aerosol concentrations. For GC2 this is from a HadGEM2-AO AMIP (Atmosphere Model Intercomparison Project) simulation and for GC3 it is from a GA6 AMIP simulation, but in both cases the direct and indirect aerosol effects are calculated interactively as for the climate experiment (Walters et al., 2017a).

3. GC3 Mean Biases and Summary Metrics

HadGEM2-AO develops a cold mean SST bias over much of the world's oceans, with a particularly intense cold anomaly in the central North Atlantic associated with the northward turn of the Gulf Stream (Figure 2). Exceptions are the Southern Ocean and Southern Hemisphere subtropical stratocumulus regions where a small warm bias is present. The increased ocean resolution in HadGEM3-GC2 compared to HadGEM2-AO leads to an improved representation of the path of the Gulf Stream (Scaife et al., 2011) which largely eliminates the cold North Atlantic anomaly. However, the Southern Ocean has a large warm bias in HadGEM3-GC2 (locally >6 K). This is reduced in HadGEM3-GC3.0 and the magnitude of the remaining Southern Ocean warm bias is comparable with the magnitude of cold SST biases in the northern hemisphere. The hemispheric asymmetry in the SST bias is improved in HadGEM3-GC3.1 such that the magnitudes of local SST errors are mostly less than 2 K. As the simulation uses constant present-day forcing throughout, we would expect the global-mean SST for the final 50 years assessed here to be slightly warmer than observed due to the effect of "committed warming," although Figure 1 shows this drift to be slow.

The Southern Ocean warm bias received considerable attention in the development of HadGEM3-GC3 and remains the subject of ongoing research. An AMIP simulation parallel to HadGEM3-GC2 has a particularly large net downward surface heat flux error 40°S–60°S, which spatially corresponds well with the warm bias in the coupled model, and is in excess of that from other CMIP models (Figure 3b). This suggests that at least part of the SST error over the region (and elsewhere) is likely to come from the atmosphere. Large uncertainties exist in estimates of surface heat fluxes. Here we follow the method of Liu et al. (2015) which combines satellite-based radiative fluxes (Allan et al., 2014) with ERA-I estimates of atmospheric column energy storage and horizontal divergences. We opt to do this since the spread in observational estimates of turbulent fluxes is considerably larger than the spread between observational estimates of radiative fluxes and this approach results in a consistent set of surface flux components which, by design, add up to the estimate of the net flux. The turbulent flux component is arguably the least certain with this methodology since it is a residual in the calculation. Further discussion of the uncertainties with this method are in Liu et al. (2015).

The positive net flux error in Figure 3b over the Southern Ocean arises from a combination of errors in the shortwave (SW) and turbulent fluxes. These are comparable to HadGEM2-AO, however, unlike HadGEM2-

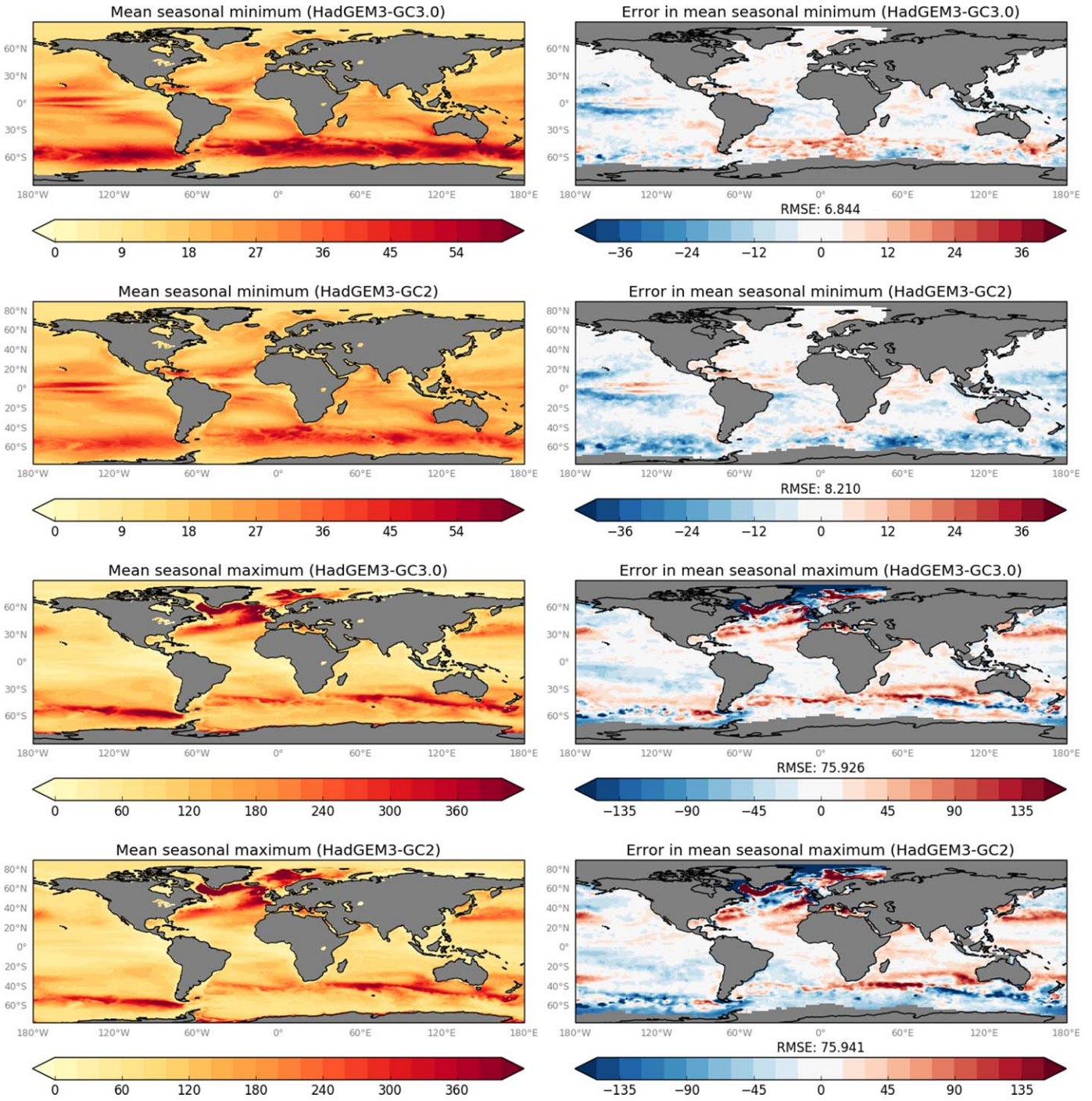

Figure 5. (left) HadGEM3-GC3.0 and HadGEM3-GC2 seasonal maximum and minimum mixed-layer depth and (right) error against the de Boyer Montégut et al. (2004) observational data set. The multiannual mean of each month is calculated first and then maximum and minimum over the seasonal cycle.

AO, HadGEM3-GC2 does not have a compensating negative flux error in the longwave (LW). This improvement in the LW surface flux is due to work having been done to increase low cloud amounts, reducing a bias over the region (Bodas-Salcedo et al., 2012). Detailed analysis has shown the outstanding SW error to be primarily related to the microphysical cloud processes (especially the lack of supercooled liquid water in the simulated cloud), and the introduction of turbulent product on of mixed-phase cloud (Furtado et al., 2016) and the modal aerosol scheme in HadGEM3-GC3 have improved the Southern Ocean cloud microphysics. As a result of these model changes, the net flux error in HadGEM3-GC3 has been reduced relative

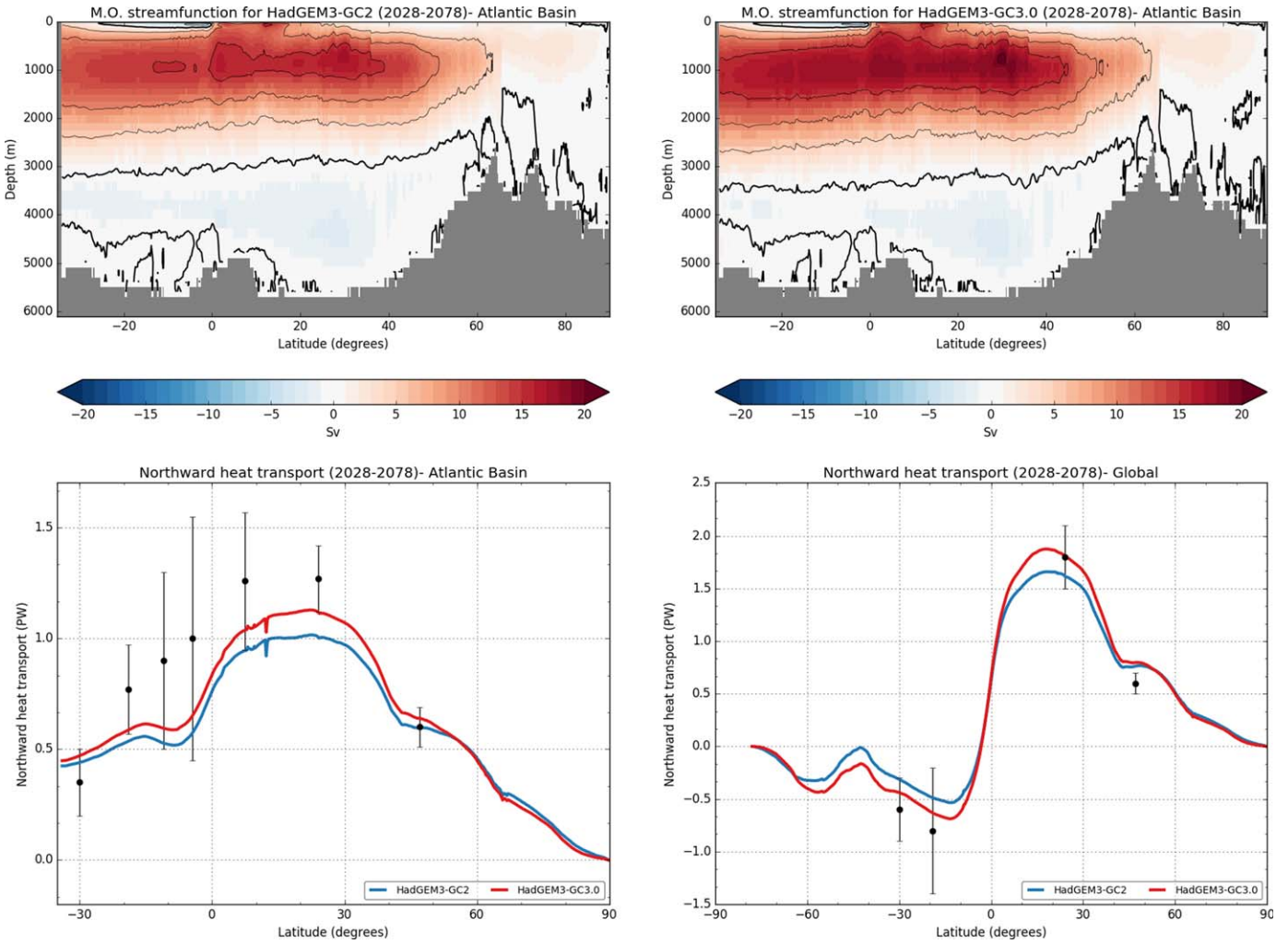

Figure 6. (top) Atlantic meridional overturning streamfunction in HadGEM3-GC2 and HadGEM3-GC3.0 (thicker black line marks the zero contour). (bottom left) Atlantic and (bottom right) Global northward heat transport (PW). Black dot and whiskers show observations from Ganachaud and Wunsch (2003).

to HadGEM3-GC2 and is now comparable with HadGEM2-AO. However, the total absolute error (sum of the magnitude of shortwave, longwave, and turbulent component errors) is much smaller and among the lowest of the CMIP5 models indicating that significant removal of compensating errors has taken place (Figure 3f). More generally, away from the tropics, the total absolute error is comparable with, or smaller than, the total absolute error of the CMIP5 ensemble mean flux errors (black dashed line in Figure 3f) which is notable since the ensemble mean can be hard to beat due to cancelling errors between different models. The high total absolute error just north of the equator is due to SW errors associated with cloud errors along the ITCZ and turbulent flux errors. It is likely that a contribution to the latter is excess evaporation associated with an overly strong hydrological cycle (see below), however, the uncertainty in the observationally derived estimate of the turbulent flux is also likely to be large in this region.

The negative flux error around 30°N in Figure 3b, which leads to the cold bias in the northern hemisphere, became worse in HadGEM3-GC3.0 compared with HadGEM3-GC2, but was subsequently improved in HadGEM3-GC3.1. In both cases, sensitivity studies have shown that the changes responsible were related to cloud-aerosol interaction, although the impact on the surface flux is through an indirect effect on the turbulent flux rather than directly through the SW as might be expected. The initial reduction in the downward surface flux was due to the revised activation scheme within the new modal aerosol scheme, while the subsequent improvement in HadGEM3-GC3.1 was due to the inclusion of a representation of the spectral dispersion in the calculation of the effective radius following (Liu et al., 2008).

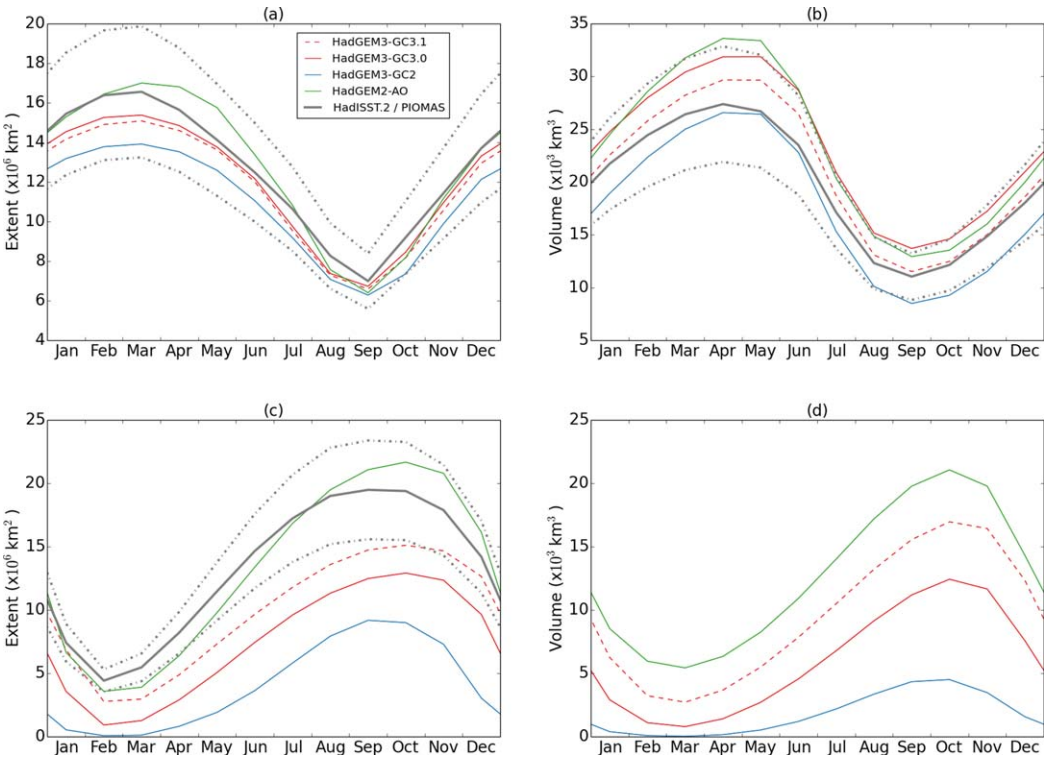

Figure 7. Annual cycle of (a, b) Arctic and (c, d) Antarctic sea (a, c) ice extent and (b, d) volume. Shown in black are HadISST observations (Rayner et al., 2003) of sea ice extent 1990–2009 for Arctic and Antarctic, and PIOMAS analyses (Schweiger et al., 2011) of sea ice volume 1990–2009 for Arctic only, with dashed lines indicating 20% intervals.

The improvements to the surface fluxes along with changes within the ocean model, primarily a retuning of the vertical and along-isopycnal mixing (as described by Storkey et al., 2017), have all contributed to the improved SSTs in HadGEM3-GC3, and generally to the zonal mean temperature profile through the ocean (Figure 4). A large tropical subsurface cold bias in HadGEM2-AO and warm bias in HadGEM3-GC2 has been considerably reduced in HadGEM3-GC3 and the Southern Ocean warm bias improved throughout the top 750 m of the ocean. There is an indication that the winter ocean mixed layer might now be slightly too deep in the mode water formation regions (Figure 5) and a future evaluation of mode water properties should be considered. There has been a strengthening of the Atlantic Meridional Overturning Circulation (MOC) by 2–3 Sv, bringing the heat transport within the uncertainty estimates of current observations (Figure 6). However, the strengthening of the MOC may, in part, be due to deepening of convection in the Labrador Sea, which increases an existing bias in HadGEM3-GC2 and can be seen as a large positive error in the mixed-layer depths for the region (Figure 5).

The accurate simulation of sea ice extent and thickness in the present-day climate is likely to be important for the estimation of climate sensitivity, projections of when the Arctic will be ice free in summer, and seasonal forecasts of sea ice extent (e.g., to inform the use of Arctic routes by shipping). HadGEM3-GC3 reproduces the seasonal cycle of Arctic ice extent well (Figure 7a). The Arctic ice volume tends to have a slightly larger amplitude of the seasonal cycle in all of the models compared with the Pan-Arctic Ice-Ocean Modelling and Assimilation System (PIOMAS; Schweiger et al., 2011). In HadGEM3-GC3.1, this results in the summer Arctic mean being close to the PIOMAS value and the winter volume being slightly higher (Figure 7b), although still within the estimated observational uncertainty. As noted by Laxon et al. (2013), PIOMAS may underestimate winter ice thickness. The large Southern Ocean warm bias in the HadGEM3 family of models results in the Antarctic ice extent and volume being considerably below what is observed, providing a positive feedback on the warm bias. The improvements to the SST error from HadGEM3-GC2 to HadGEM3-GC3.1 are reflected in the Antarctic ice extent and volume (Figures 7c and 7d) such that the Antarctic ice extent in HadGEM3-GC3.1 is now around 80% of what is observed, although still below HadGEM2-AO.

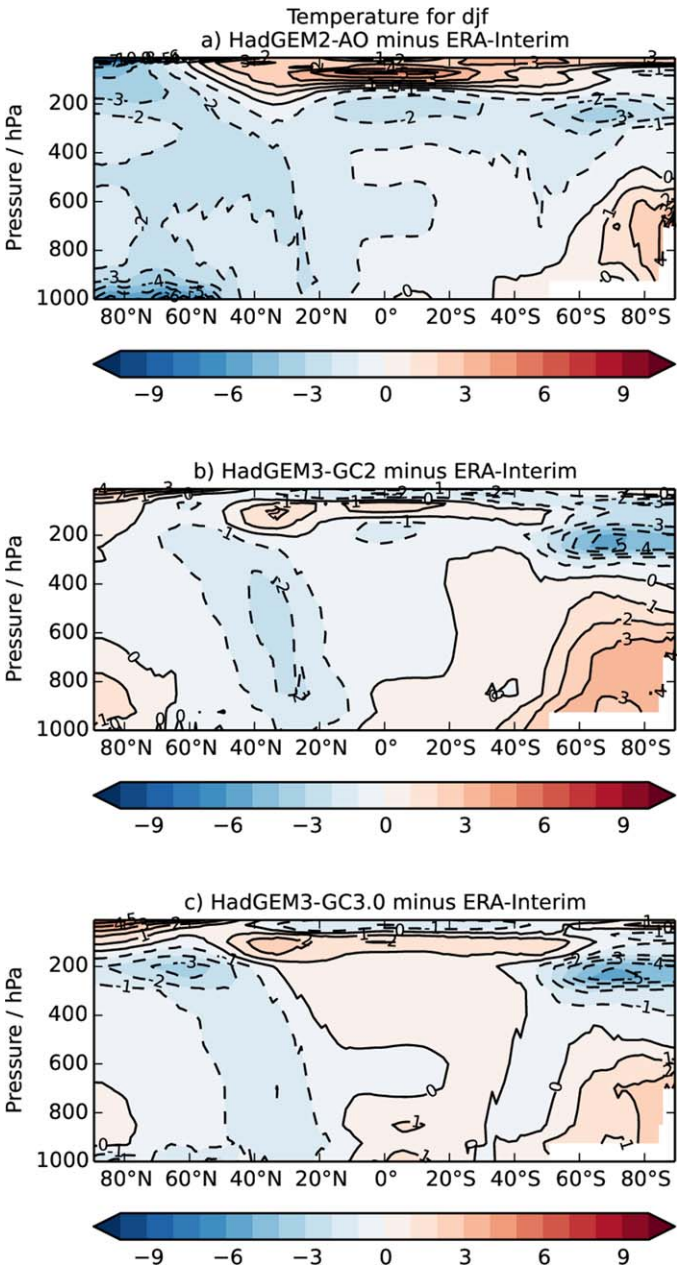

Figure 8. DJF zonal mean temperature error (K) compared with ERA-I (Dee et al., 2011) for (a) HadGEM2-AO, (b) HadGEM3-GC2, and (c) HadGEM3-GC3.0.

The improved SSTs globally are associated with improvements to lower tropospheric temperatures (Figure 8) with a large northern hemisphere cold bias being reduced between HadGEM2-AO and HadGEM3-GC2, and a Southern Hemisphere warm bias increasing in HadGEM3-GC2 but then reduced again in HadGEM3-GC3. Hardiman et al. (2015) describe the sensitivity of the tropical tropopause layer (TTL) temperature to various physical processes. Biases in this region are important since the temperature here controls the amount of water vapour entering the stratosphere (Brewer, 1949), upon which the interactive chemistry in an ESM is highly dependent. In HadGEM2-AO there was a large warm bias (>5 K) in the TTL region (Figure 8a) which has been considerably reduced, primarily through a change to the stencil used by the semi-Lagrangian dynamics in the calculation of the departure point for the vertical advection of potential temperature to make it more accurate in regions of large gradients (the so called Hermite cubic interpolation; Walters et al., 2017a). However, a cold bias has now developed near the polar tropopause (particularly in summer in each hemisphere) which requires investigation. Improvements to the height of convection, most recently through the revised numerics of the convection scheme (the so called 6A convection scheme; Walters et al., 2017b), have eliminated a cold bias in the upper tropical troposphere.

A detailed, cross-time-scale, evaluation of cloud against a range of observational data for the atmosphere components of HadGEM3-GC2 and HadGEM3-GC3 has been undertaken by Williams and Bodas-Salcedo (2017). They illustrate how the many cloud, radiation and convection changes in the most recent model configuration have improved the simulation, reducing or eliminating a number of biases (e.g., excess cirrus; excess light drizzle). Figure 9 is a reproduction of Williams and Bodas-Salcedo (2017, Figure 2b) but with the HadGEM2-AO AMIP simulation added. This shows the mean vertical profile of cloud in the tropics from CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations; Chepfer et al., 2010;Winker et al., 2010) and from the models using the CFMIP (Cloud Feedback Model Intercomparison Project) Observational Simulator Package (COSP; Bodas-Salcedo et al., 2011; Chepfer et al., 2008). Despite the accurate simulation of cloud being a particular strength of HadGEM2-AO (e.g., Jiang et al., 2012; Klein et al., 2013; Nam et al., 2012), Figure 9 reveals some of the issues in terms of detrainment from deep convection not being at a sufficiently high altitude, too little thin cirrus, too little midtop convective cloud (e.g., cumulus congestus) and too little boundary layer cloud (especially in trade cumulus regions). The amount of boundary layer cloud, and height of convection are somewhat improved in HadGEM3-GC2. The amount of subvisual cirrus also increased considerably and became excessive. This is addressed in HadGEM3-GC3, along with the height of condensate detrainment from convection being further improved, consistent with the reduced upper tropospheric cold bias in Figure 8.

The improvements in cloud follow through to improved top-of-atmosphere outgoing shortwave radiation (OSR) and outgoing longwave radiation (OLR) (Figure 10). The spatial root-mean-square-error (RMSE) of the multiannual-mean bias is reduced by over 20% for OSR and just under 10% for OLR between HadGEM2-AO and HadGEM2-GC3.1. HadGEM2-AO has excess reflected shortwave in the subtropical transition regions which has been addressed through the introduction of the new prognostic cloud and condensate scheme (PC2; Wilson et al., 2008). As discussed above, excess shortwave radiation reaches the Southern Ocean in HadGEM3-GC2 and this can also be seen in the OSR bias. Much of the error is believed to be due to a lack of supercooled liquid water at cloud top which has been improved, although not completely fixed, in

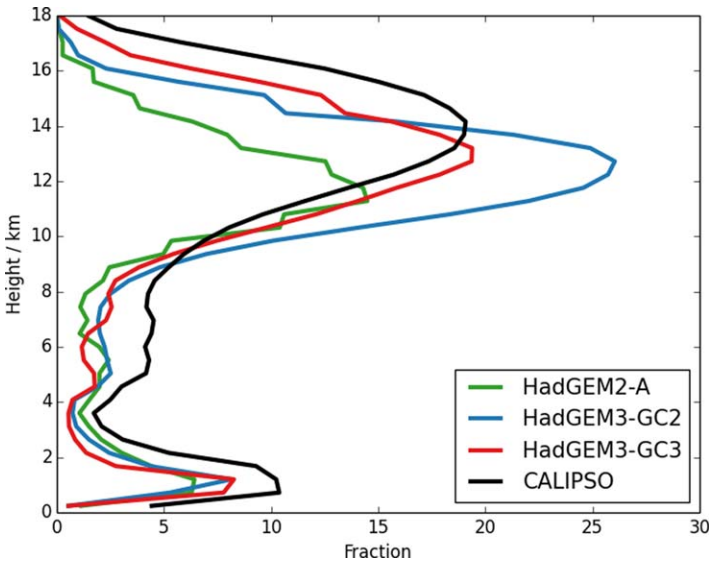

Figure 9. Tropical-mean (20°N–20°S) vertical profile of cloud fraction in atmosphere-only simulations of HadGEM2-AO, HadGEM3-GC2, and HadGEM3-GC3.0 using the CALIPSO simulator from COSP. The observed profile from CALIPSO is shown black.

HadGEM3-GC3. The OLR bias points to lack of detrained cloud from deep convection over the equatorial western Pacific in HadGEM2-AO and this has been improved. However, there remains a lack of convective cloud over tropical land (northern South America, central Africa, India, and the Maritime Continent). Throughout the extra tropics, there is too little OLR in all three model configurations. This bias is from the clear sky rather than cloud (not shown) and may relate to the upper tropospheric cold bias at higher latitudes noted above and/or excess tropospheric water vapour.

Mean precipitation errors are dominated by systematic errors along the intertropical convergence zone (ITCZ). The leading order spatial pattern of the error has changed little through the configurations (Figure 11), although the DJF dry bias over the equatorial Pacific in HadGEM2 has largely been replaced by a wet bias in more recent configurations. Consistent with the OLR bias, there is a lack of precipitation over the Sahel, India, and the Maritime Continent during JJA and excess precipitation over the oceans. In most places, these biases have been slightly reduced in HadGEM3-GC3 compared with HadGEM3-GC2, most notably over India where the dry bias has been reduced by a third. There is a southward displacement of the Atlantic ITCZ which, as well as affecting the Sahel, may be important for teleconnections to the midlatitudes (Scaife et al., 2017). During MAM, the southward displaced ITCZ also extends across the Pacific (not shown). This southward displacement is likely to be partially a result of the asymmetric SST bias, however, there is the indication of this displacement (albeit not as large) in parallel AMIP simulations which use observed SSTs suggesting that the asymmetric SST bias in the coupled configuration is not the sole cause. The annual-mean global-mean precipitation is 3.15 mm d^{-1} in HadGEM3-GC3 which is higher than observational estimates which generally have a value below 3.0 mm d^{-1} (e.g., Behrangi et al., 2014). The figure in the parallel AMIP simulation is almost identical indicating that it is not a result of SST biases in the coupled simulation. This excess precipitation is a common model error and here reflects an overly strong hydrological cycle which has been a long-term issue in the UM. Associated with this strong hydrological cycle is excess oceanic evaporation which accounts for the large tropical turbulent flux error evident in Figure 3.

Watterson et al. (2014) propose a basic overall metric for climate models based on their ability to simulate screen-level temperature, precipitation, and pressure at mean sea level (PMSL) over the continents. Table 2 shows the score and its components for the models considered in this paper. Development from HadGEM2-AO to HadGEM3-GC2 saw small improvements to the temperature and precipitation metrics, however, there was significant degradation in PMSL (resulting in a drop in the overall metric) which arose through a combination of the introduction of the ENDGame (Even Newer Dynamics for Global Atmospheric Modelling of the Environment; Wood et al., 2014) dynamical core and revised gravity wave drag scheme (Vosper et al., 2009). These model developments, while improving many aspects of variability, resulted in a large high pressure bias over the polar oceans and low pressure bias over the midlatitude continents. This transfer of mass toward the pole can be seen early in atmosphere-only NWP forecasts and investigations to date suggest excess surface drag in the midlatitudes may be responsible. It is also possible that these circulation errors may contribute to the local cloud and surface flux errors discussed above. All three metric components have improved through HadGEM3-GC3.0 to HadGEM3-GC3.1. In the case of PMSL, it should be noted that much of the improvement in the score between HadGEM3-GC2 and HadGEM3-GC3.0 was due to starting the simulation from a different atmospheric state. The UM conserves atmospheric mass well, so the global-mean PMSL is largely determined by the mass in the initial dump, which in turn resulted from an operational data assimilation on a particular day. For HadGEM3-GC3, the atmospheric start dump was changed to be consistent across AMIP and coupled simulations and have the soil moisture already spun-up, hence the new start dump had been produced by a previous AMIP simulation and so a different original analysis. By chance, the global PMSL in the new HadGEM3-GC3 initialization file is very close to 1,013.25 hPa, the value specified by a number of international model intercomparison projects (e.g., CFMIP; Webb et al., 2017; the Aqua-Planet

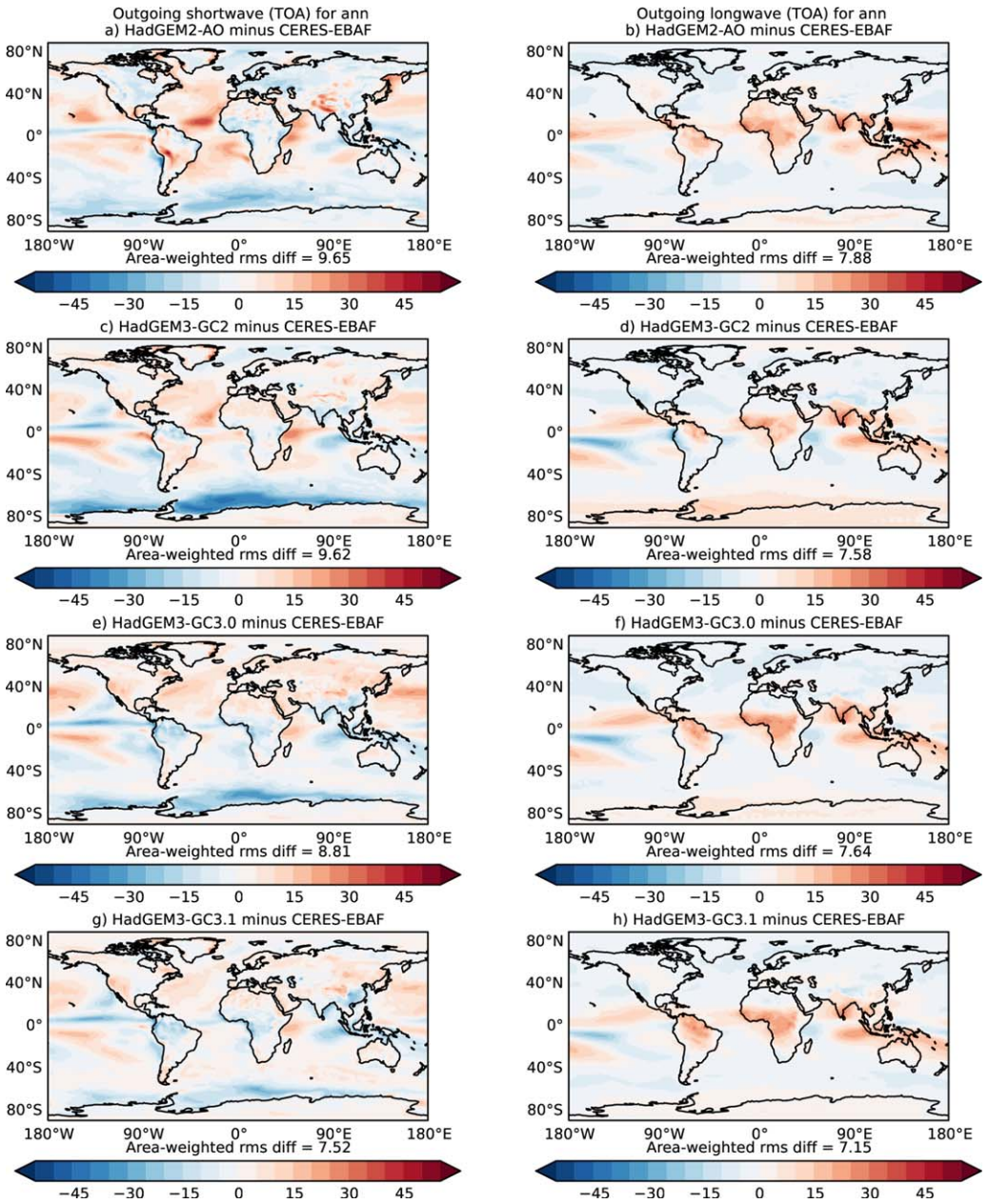

Figure 10. Annual-mean error in (a, c, e, and g) OSR and (b, d, f, and h) OLR (W m^{-2}) compared with CERES-EBAF (Loeb et al., 2009) for (a, b) HadGEM2-AO, (c, d) HadGEM3-GC2, (e, f) HadGEM3-GC3.0, and (g, h) HadGEM3-GC3.1.

Experiment, APE; Williamson et al., 2012). This is around 1.2 hPa higher than in the initialization of HadGEM3-GC2 and reduces the low PMSL bias over low latitude land. There have also been genuine improvements in the local PMSL error resulting in the score for HadGEM3-GC3.1 being higher again, although the polar high pressure bias remains and its effect may be being underrepresented in Table 2 due to the metric only considering land areas.

Overall GC3 may be considered an improvement on GC2 and this is reflected in its NWP performance. Currently the Met Office and its partners use atmosphere-only configurations operationally for week 1 forecasts, however, the use of coupled models for this forecast range is being tested for possible future implementation. Here the overall performance of GC3 is compared with GC2 in terms of forecast RMSE for a number of key variables (Figure 12). Compared against observations, the forecasts are at least overall neutral and many are significantly improved at most forecast lead times. Against ECMWF analyses,

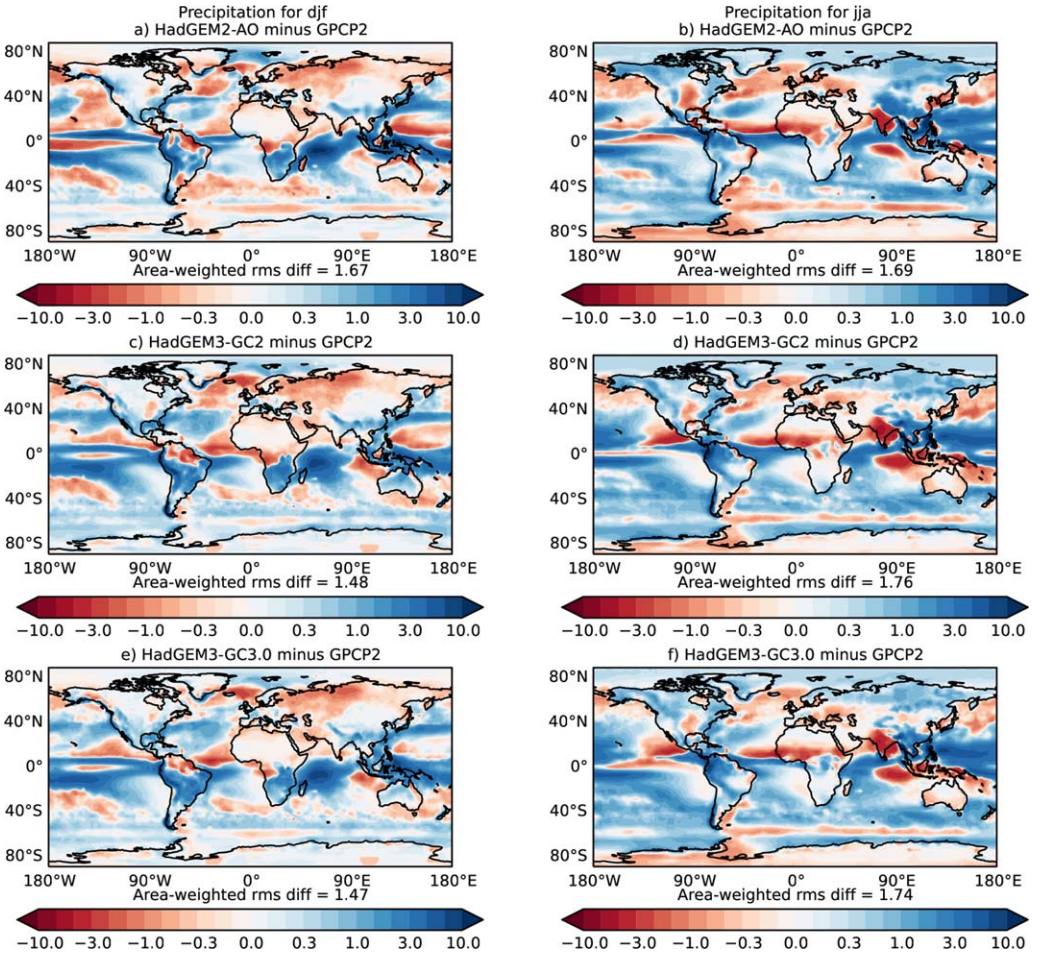

Figure 11. (a, c, and e) DJF and (b, d, and f) JJA mean precipitation rate (mm/d) against GPCP observations (Adler et al., 2003) for (a, b) HadGEM2-AO, (c, d) HadGEM3-GC2, and (e, f) HadGEM3-GC3.0.

moving from GC2 to GC3 is neutral at longer lead times, but there is degradation at short lead times. This short-range degradation is likely to be due to differences between the Met Office operational analysis of the day which have been used to initialize the model, ECMWF analysis used for verification, and that neither are necessarily consistent with an analysis which would be produced from an atmosphere data assimilation system using the GA7 science. Investigations around implementation of the atmosphere-only configuration for NWP have shown that the performance of GA7 at short forecast lead times can be quite different (more so than was seen with previous configurations) with full cycling data assimilation compared with starting from existing analyses. These investigations have also shown that a recalculation of the covariance statistics used in the data assimilation is likely to be required, hence the longer forecast lead times in Figure 12 are likely to be more representative of a system using consistent data assimilation.

Table 2
Global Climate Scores Following Watterson et al. (2014*)*

	Overall	Screen temperature	Precipitation	PMSL
HadGEM2-AO	690	804	588	680
HadGEM3-GC2	686	838	600	621
HadGEM3-GC3.0	711	848	623	662
HadGEM3-GC3.1	727	864	636	681

Note. The overall score for all continents and individual scores for contributing elements are shown. A higher number indicates an improvement. In an idealized case of a model field being identical to an observational field, the metric is 1,000.

For the use of the configuration for seasonal forecasting at the Met Office, the predictability of the North Atlantic Oscillation (NAO) is a key metric (Scaife et al., 2014). Figure 13 shows the standardized NAO from GloSea5-GC2 and GloSea5-GC3 winter hindcasts. The correlation with observations from the two configurations are statistically indistinguishable, suggesting that GloSea5-GC3 forecast skill is likely to be comparable with the currently operational GloSea5-GC2. Despite there being a statistically significant correlation with the observations, Figure 13 is a standardized plot and Scaife et al. (2014) and Eade et al. (2014) highlight the anomalously low "signal to noise" in the GloSea5 ensemble. The strength of predictable signal is still

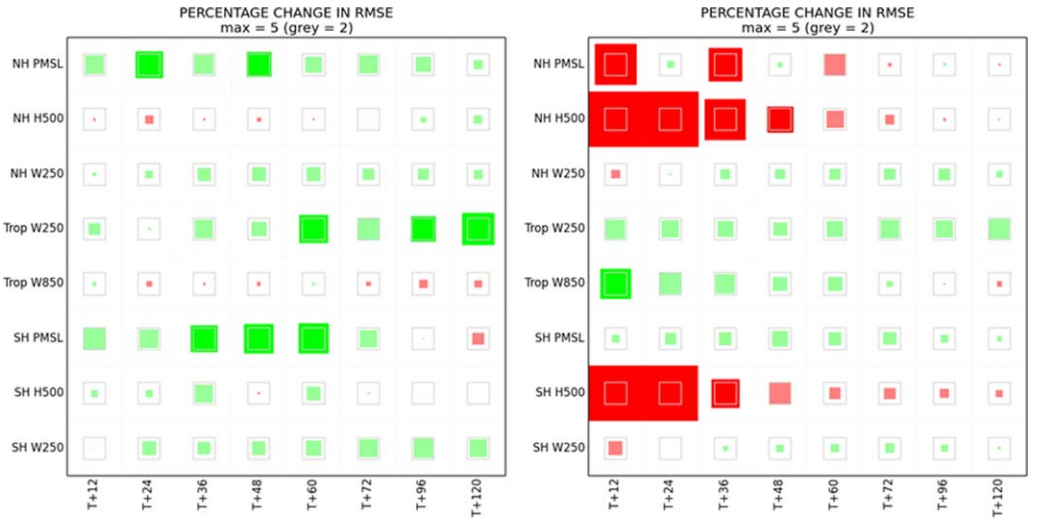

Figure 12. Relative difference in RMSE between GC2 and GC3.0 against (a) synop observations and (b) operational ECMWF analyses of the day for the NWP experiment. Reduced RMSE (i.e., improvements) between GC2 and GC3.0 are green and increased RMSE are red. The size of the coloured box represents the fractional change with a 2% change filling the outline grey box and a 5% change filling the largest square.

weak as illustrated in Figure 14 which shows the PMSL difference between a winter with a strong positive NAO (2011/2012) and a strong negative NAO (2009/2010). Both configurations have the correct sign and pattern of anomaly, however, the signal in both configurations remains small.

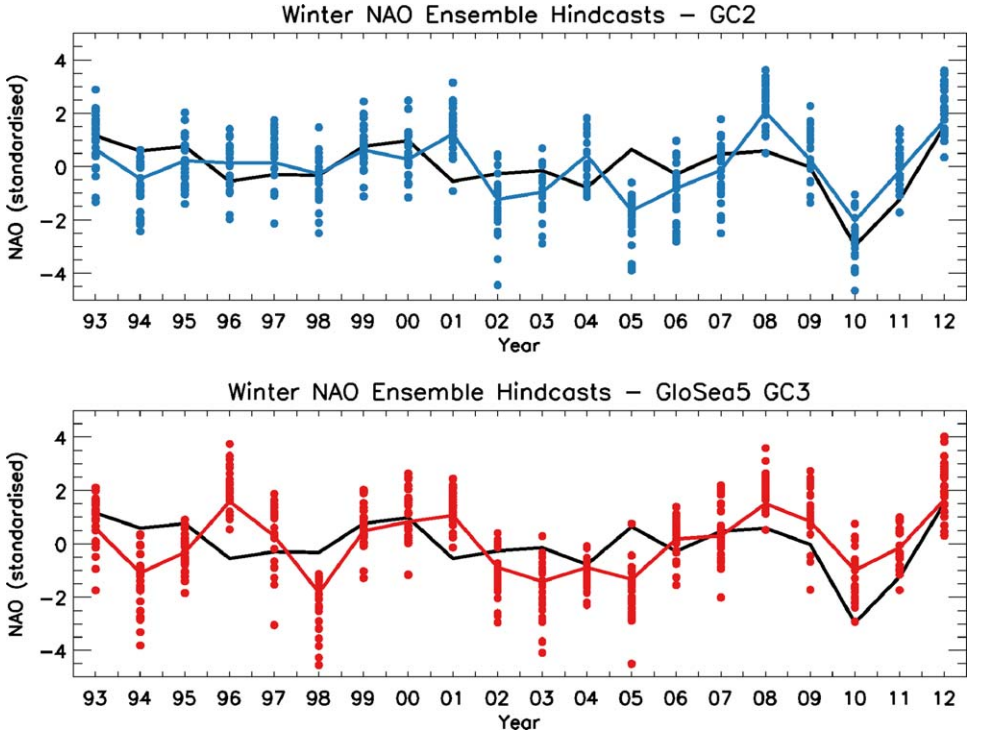

Figure 13. Winter normalized NAO index from the seasonal hindcasts for GloSea5-GC2 and GloSea5-GC3.0 following Scaife et al. (2014). Each hindcast member is shown with a dot and the ensemble mean with the coloured line. Black is HadSLP2 (Allan & Ansell, 2006) observations.

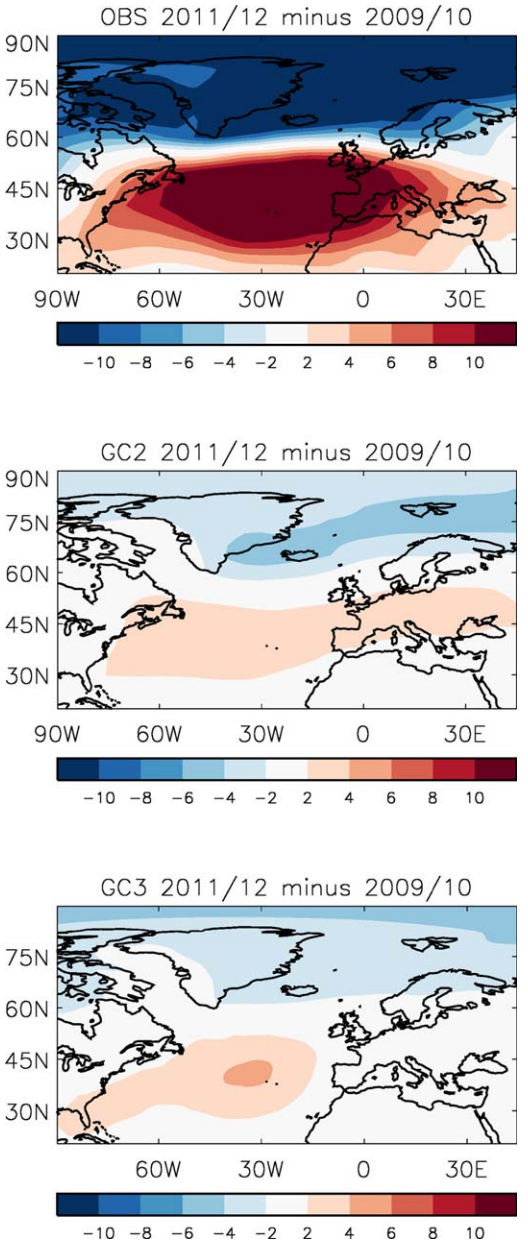

Figure 14. Mean PMSL difference (hPa) over the N. Atlantic between the winter 2009/2010 and 2011/2012. HadSLP2 observations are shown along with the mean of the hindcasts for these years from GloSea5-GC2 and GloSea5-GC3.0.

4. GC3 Variability

Williams et al. (2015) show the improvements in tracked extratropical cyclones in both hemispheres in HadGEM3-GC2 resulting from the ENDGame revision to the dynamical core which generally increases synoptic variability. HadGEM3-GC3 is very similar to HadGEM3-GC2 in this regard.

The jet over the North Atlantic displays particularly complicated variability (compared with other ocean basins), with a trimodal latitudinal frequency distribution being present in the winter and a skewed structure in the summer (Figure 15; Woollings et al., 2010). The winter jet positions have been shown to correspond to three of the principal synoptic regimes over the region (Ferranti et al., 2015). Generally, CMIP5 models have struggled to capture this latitudinal variability, either not representing the multimodal structure, or too readily favouring a particular peak (e.g., the southern jet position in the winter; Anstey et al., 2013). Both HadGEM3-GC2 and HadGEM3-GC3 capture the winter jet well, both in terms of the latitudinal position and temporal frequency of each preferred location. In summer, the reanalysis indicates a main central peak with a less frequent secondary maximum to the north. Both HadGEM3-GC2 and HadGEM3-GC3 have an indication of a double maxima, however, the two peaks have a similar frequency (i.e., the main southern position is too infrequent and the northern position too frequent in the models). Probably connected with this, there also appears to be a general northward displacement in the summer. Based on understanding gained from atmosphere-only NWP simulations, we speculate that these summer jet errors may be associated with biases in low cloud amount affecting the surface fluxes over the North Atlantic and North America, and hence baroclinic development. Sensitivity experiments have also suggested that the representation of drag (orographic and boundary layer) may play a role.

Blocking of transient eddies is an important regime which can lead to some of the most extreme weather in the midlatitudes. Following the methodology of Scherrer et al. (2006) to identify blocks based on a reversal of the equator-pole 500 hPa geopotential height gradient applied between 35°N and 75°N, three main blocking regions can be identified during the boreal winter: N. Pacific, Greenland and N.W. Europe (Figures 16a–16c). Both HadGEM3-GC2 and HadGEM3-GC3 capture blocking in these regions, and the frequency of blocks simulated in Pacific sector is reasonably good (although there is a slight deficit in HadGEM3-GC3), whereas both models have notably too little over N.W. Europe. HadGEM3-GC3 may be regarded as a slight improvement in that the excess blocking over the eastern Atlantic (to the S.W. of the UK) is reduced, and more Greenland blocking is present. However, the significant N.W. Europe blocking bias remains. This drop in the frequency of European blocking develops in the second week of initialized forecasts and may be connected with the hemispheric-scale PMSL errors discussed in the previous section. During the boreal summer, blocking is mainly over the N. Pacific and Scandinavia (Figures 16d–16f). Again, HadGEM3-GC2 and HadGEM3-GC3 represent the Pacific blocking reasonably well (although there is a slight deficit) whereas blocking over Scandinavia occurs with a frequency of less than half that in the reanalysis. In contrast, there is excess summer blocking over the eastern Atlantic and central Europe, possibly connected with the northward displaced jet noted above.

Moving to the tropics, the El-Niño Southern Oscillation (ENSO) is a leading mode of tropical variability with impacts around the globe. Williams et al. (2015) show the El-Niño SST composite and equatorial wind stress in HadGEM3-GC2 which both compare favourably with observations. This has remained the case in

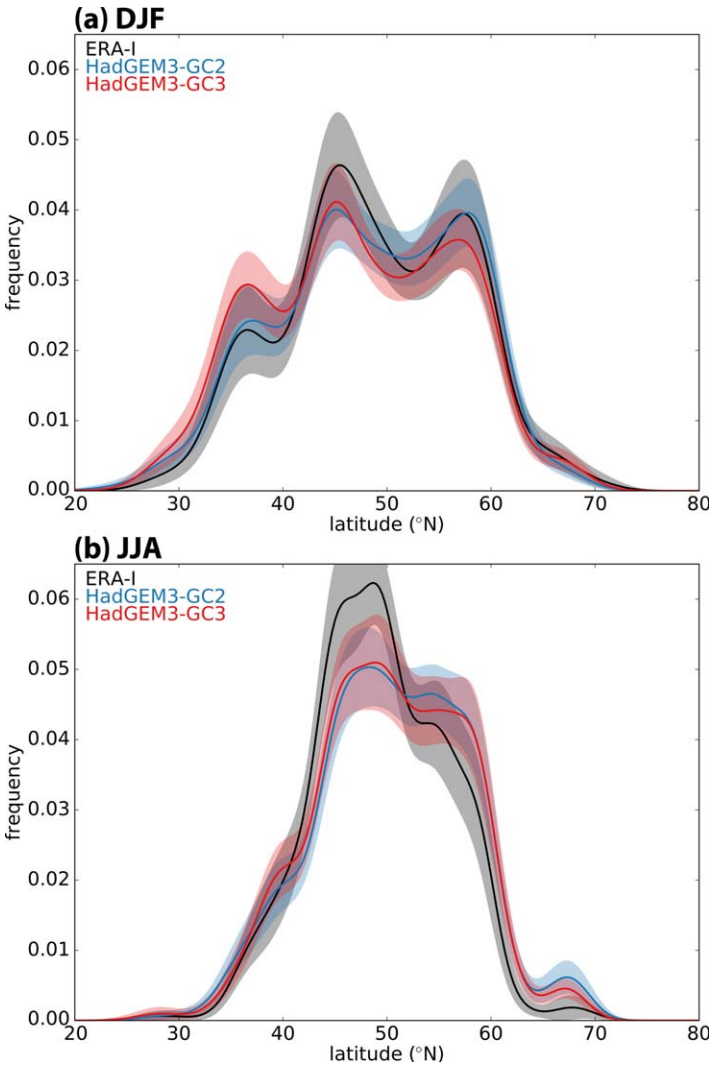

Figure 15. Normalized frequency distribution of (a) DJF and (b) JJA jet latitude defined as the maximum 850 hPa wind over the North Atlantic 60°W–0°E (following Woollings et al., 2010) for ERA-I (years 1979–2016), HadGEM3-GC2 and HadGEM3-GC3.0. The shading represents an estimate of uncertainty due to internal variability. This is calculated using a bootstrap methodology in which all days within a season are used within a 50 repetition bootstrap and then a 1,000 repetition bootstrap is used across the different years. The intraseasonal looping is to account for the fact that the jet can stay in one regime for much of a season. The shading represents ±2 standard deviations of the resulting distribution.

HadGEM3-GC3 and a variety of performance measures for ENSO are provided in Table 3. Since HadGEM2-AO there have been notable improvements in the power spectrum time scale, mean Niño3 SST, mean Niño4 wind stress and Niño4 standard deviation in precipitation. The Niño4 standard deviation of SSTs deteriorated between HadGEM2-AO and HadGEM3-GC2 but has recovered again in the development of HadGEM3-GC3.

The main mode of tropical variability on intraseasonal time scales is the Madden-Julian Oscillation (MJO). In addition to being a direct source of precipitation variability for regions such as the Maritime Continent and northern Australia, teleconnections from the MJO have been shown to affect other parts of the tropics and variability in the midlatitudes (e.g., Cassou, 2008). In common with many models, the UM has traditionally struggled with MJO propagation across the Maritime Continent, preferring to keep it as more of a stationary mode of variability within the confines of the Indian Ocean basin. MJO propagation has significantly improved in HadGEM3-GC3 with more systems now crossing the Maritime Continent region and into the western Pacific (Figure 17). OLR and winds both have coherent propagation in the observations. In HadGEM3-GC2, there is little propagation of OLR anomalies yet the winds show weaker but faster eastward propagation. In HadGEM3-GC3 the propagation of convection is much improved and so is the correlation with the winds, however, the propagation speed remains faster than observed. This improvement has arisen primarily due to a change in the SST bias which, although generally being improved in the tropics, now has a local warm anomaly of around 1K in the waters surrounding the Maritime Continent. An AMIP simulation using the coupled model SSTs reflects this propogation whereas it is not apparent in the standard AMIP experiment with observed SSTs. The increased lower tropospheric humidity resulting from the warm SST is believed to be key to the improved propagation. While this indicates that the improvements to the MJO propagation in HadGEM3-GC3 might not be for the correct reasons, it provides hope that if the lower tropospheric humidity were correctly increased over the islands (e.g., through increased moisture convergence; Rashid & Hirst, 2016) it would significantly help the MJO simulation.

More generally the tropical precipitation power spectra has been improved in HadGEM3-GC3 (Figure 18). The number of eastward propagating Kelvin waves increased and more closely matches the spectra in the parallel AMIP simulation (not shown), probably due to the improved tropical SSTs. There is some increase in the power of Kelvin waves at around 30 days and the error dipole between this and 60 days for zonal wave numbers of 1–2 (bottom right plot in Figure 18) is consistent with the faster phase speeds seen in the lead-lag correlations.

The introduction of the ENDGame revision to the dynamical core and GA6 physics package (Walters et al., 2017a) resulted in a large improvement in the frequency, track, and intensity of tropical cyclones across all time scales and was evident in HadGEM3-GC2 (Heming, 2016; Williams et al., 2015). Using the TRACK software package (Hodges, 1995) to track tropical cyclones, the track densities for HadGEM3-GC2 and HadGEM3-GC3 are shown in Figure 19. Generally, the number of tropical cyclones is increased in HadGEM3-GC3 leading to improved track density maxima over the west and east North Pacific, where the model is now comparable with the observed frequency. In the Southern Hemisphere, the simulated frequency of tropical cyclones was already excessive in HadGEM3-GC2 and this has become slightly worse in HadGEM3-GC3. Over the North Atlantic basin, HadGEM3-GC2 simulates too few tropical cyclones in the eastern Atlantic. This has been partially improved in HadGEM3-GC3 but remains below what is observed.

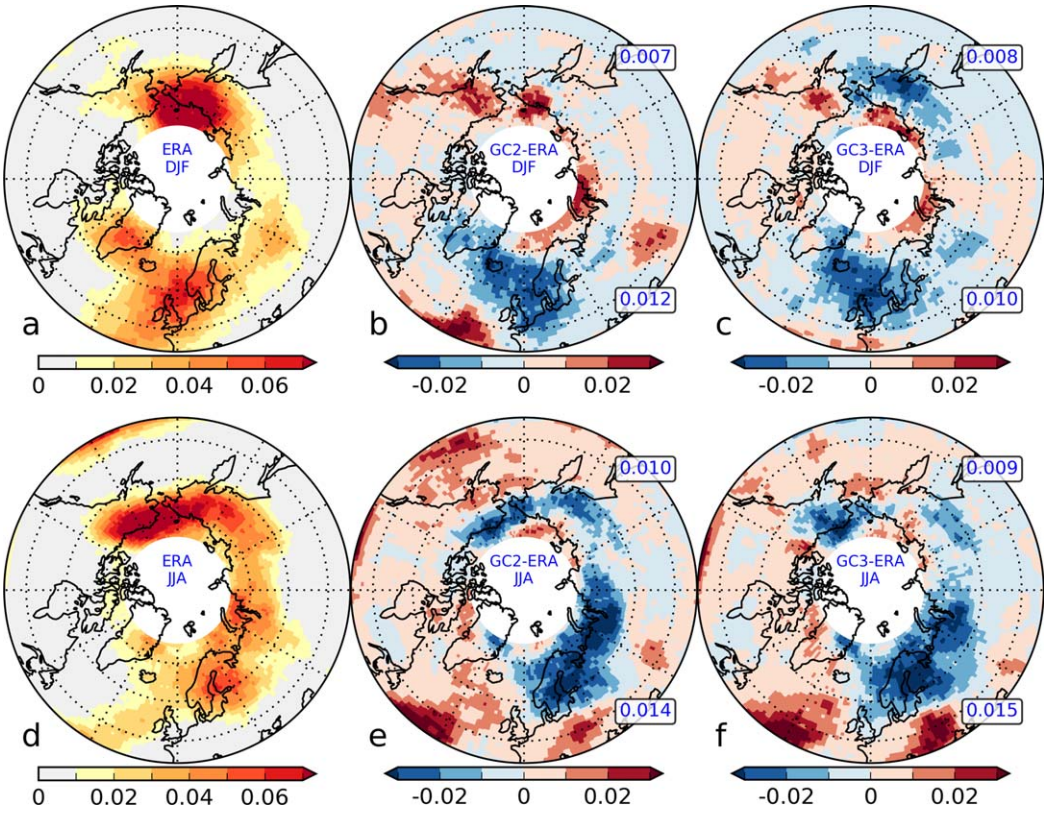

Figure 16. Northern Hemisphere blocking frequency as defined following Scherrer et al. (2006) for (a–c) DJF and (d–f) JJA. Fifty years of concatenated ERA-I and ERA40 (as described in Schiemann et al., 2017) are used (Figures 16a and 16d) and the bias in HadGEM3-GC2 and HadGEM3-GC3 against this is shown (Figures 16b, 16c, 16e, and 16f).

African Easterly Waves (AEWs) are a key generation mechanism for Atlantic tropical cyclones and hence their accurate simulation is important for tropical cyclones and Sahel rainfall alike. Walters et al. (2017b) show the improvement in the frequency of African Easterly Waves in the AMIP simulation, mainly attributable to the introduction of stochastic physics (Sanchez et al., 2016), and this improvement is also reflected in the coupled model. The frequency of AEWs is now comparable to reanalyses (not shown), however, Atlantic tropical cyclones require not only the correct frequency of waves, but also the correct distribution of vorticity and accurate positioning and strength of the African Easterly jet. We believe that the

Table 3
Selection of ENSO Measures

	M1:N3 σ_{SST}	M2:N4 σ_{SST}	M3:time scale of N3 SST power	M4:σ_{SST} (NDJ/MAM)	M5:N3 SST	M6:N4 wind stress	M7:N4 $\sigma_{Precipitation}$
Observations	0.79	0.54	**3.5**, 5.2	1.6	25.7	−0.034	2.3 (2.7)
HadGEM2-AO	0.87	0.48	3.5, **6.7**	1.1	25.1	−0.046	1.9
HadGEM3-GC2	0.67	0.31	**2.9**, 3.3	1.4	26.5	−0.029	1.9
HadGEM3-GC3.0	0.89	0.51	2.8, **3.3**, 4.0	1.3	26.4	−0.033	2.4

Note. M1 and M2 are the standard deviation of monthly SST anomaly for regions Niño3 (90°W–150°W, 5°N–5°S) and Niño4 (160°E–150°, 5°N–5°S) (K). M3 shows power spectrum time scales for monthly Niño3 SST anomaly (years, dominant time scale is in bold). M4 is a seasonality metric defined as the ratio of November–January and March–May standard deviation of Niño3 SST anomaly. M5 is annual-mean SST for Niño3 (K). M6 is the annual-mean zonal wind stress for Niño4 (N m^{-2}). M7 is the standard deviation of precipitation anomaly for Niño4 (mm d^{-1}). Model years 20–100 are assessed. SST observations are HadISST 1901–2000, precipitation is GPCP 1979–2013 (CMAP 1979–2015; Xie & Arkin, 1997), and wind stress is Southampton Oceanography Centre climatology (Josey et al., 1998).

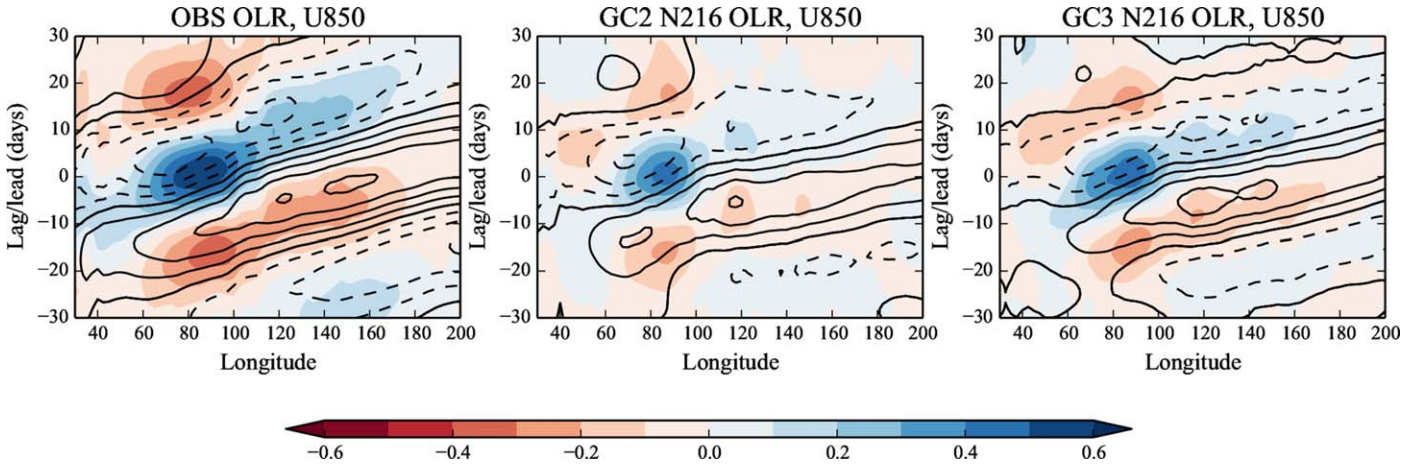

Figure 17. Hovmoller plot of OLR correlation coefficient (colour shades) at all grid points with a reference time series of OLR averaged over 70°E–100°E, 10°S–10°N at various lead/lag. Similar correlation of U850 with the OLR reference series is plotted as contours. The correlation fields are averaged 10°S–10°N. Positive values are represented by solid contours and negative values by dashed contours. Same contours levels are used for both OLR and U850.

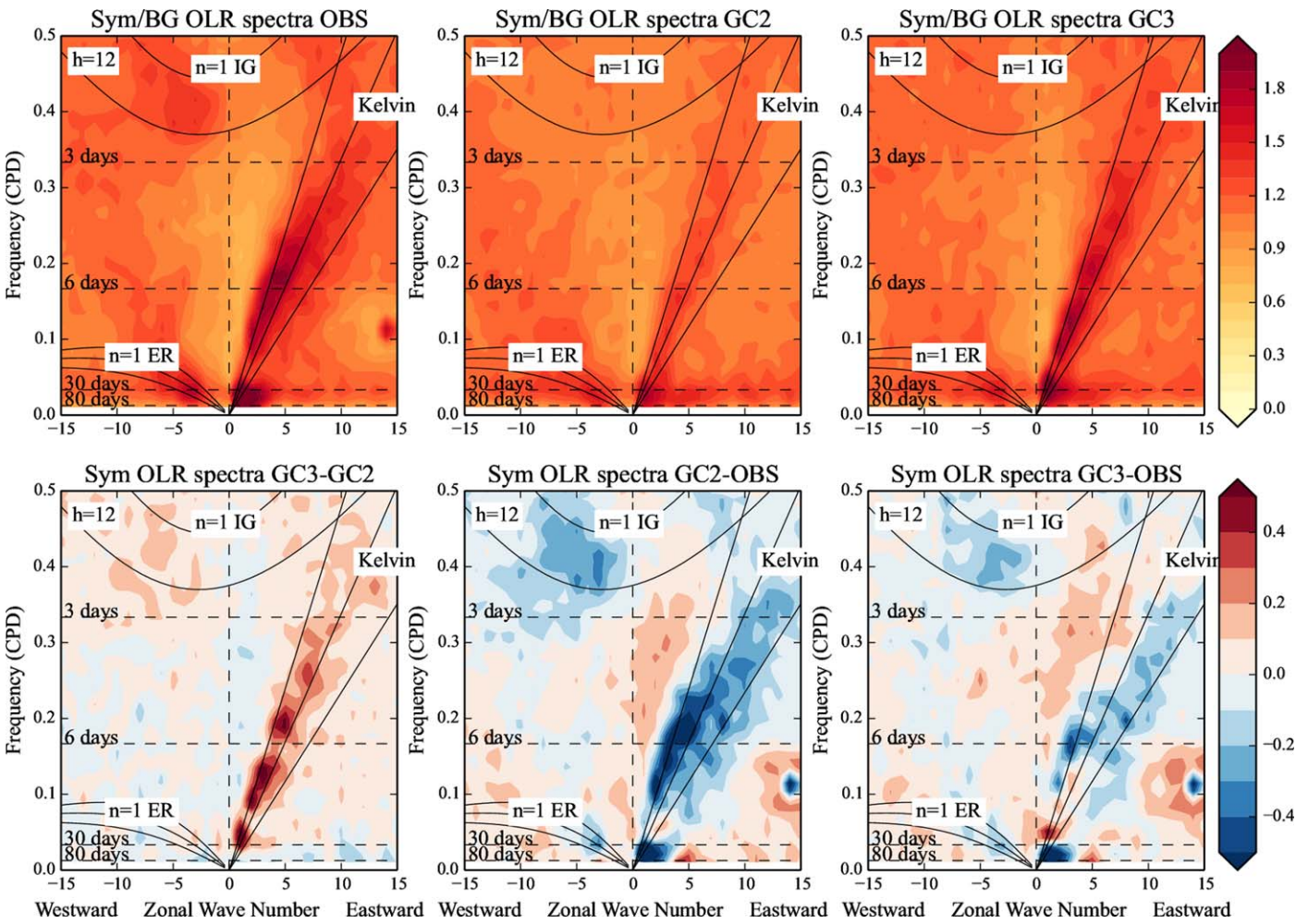

Figure 18. (top row) The ratio between raw power spectrum of OLR and a background spectrum from (left) obs, (middle) HadGEM3-GC2, and (right) HadGEM3-GC3. (bottom row) The differences in raw power spectrum between the models and the differences of models with the observations. All values here are log of power. See Wheeler and Kiladis (1999) for details.

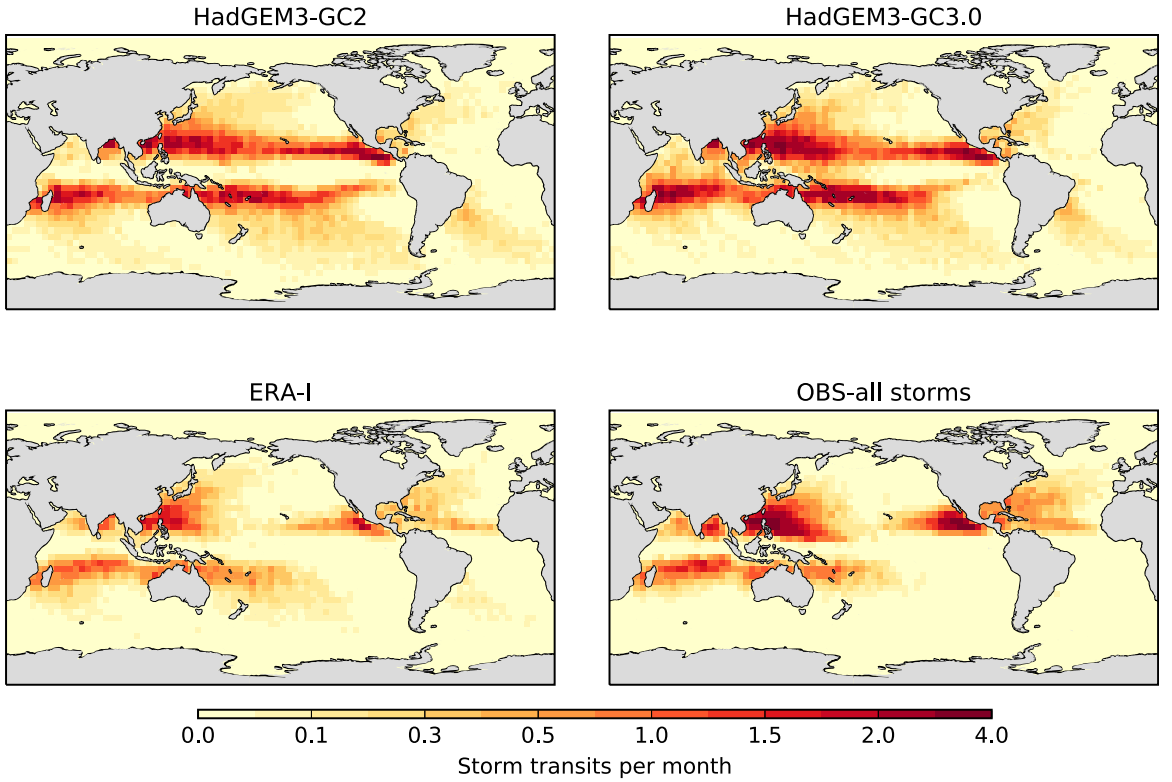

Figure 19. Tropical cyclone track densities (transits per month per unit area equivalent to a 4° spherical cap) obtained using TRACK for the final 30 years of HadGEM3-GC2, HadGEM3-GC3.0, ERA-I, and observations (North Atlantic and eastern Pacific from HURDAT [Hurricane database; Landsea & Franklin, 2013]; western North Pacific and North Indian Ocean from the U.S. Navy's Joint Typhoon Warning Centre [JTWC] best-track; Chu et al., 2002). The northern hemisphere period is set to May–November and the Southern Hemisphere period is set to October–May.

combination of these elements is still not sufficiently well simulated to provide a reliable frequency of Atlantic tropical cyclones (Tomassini et al., 2017).

5. Summary

In this paper, we have presented the performance of the GC3 configuration of the Met Office Unified Model in terms of its systematic errors. The differences between GC3.0 and GC3.1 are relatively small and, with the exception of the Southern Ocean, have localised impact. Here we have only highlighted where the differences are notable, otherwise the performance of the two can be regarded as comparable. The focus has been on the free-running climate simulation since HadGEM3-GC3.1 will be the physical model submitted to CMIP6 and underpinning UKESM1. However, the Met Office develops configurations for use across time scales and seasonal and NWP simulations have also been assessed. The results presented in this paper should be considered alongside the atmosphere-only results in Walters et al. (2017b) and ocean/sea ice results in Storkey et al. (2017)/Ridley et al. (2017).

Relative to its predecessor GC2, GC3 has improvements in a number of key systematic errors. Most significantly, the large Southern Ocean warm bias has been reduced by 75% between HadGEM3-GC2 and HadGEM3-GC3.1. There have also been improvements in the simulation of cloud, MOC, MJO, sea ice, and northern hemisphere tropical cyclone frequency. In the case of the Southern Ocean warm bias, while the net heat flux into the ocean remains too high, the combined magnitude of the error in the flux components is now well below previous UM configurations and comparable with the best models in CMIP5. This implies that many models achieve reasonable SSTs over the Southern Ocean through large compensating errors in the atmosphere fluxes.

Relative to HadGEM2-AO which was submitted to CMIP5, HadGEM3-GC3 provides a large improvement in many systematic errors. This includes reductions in temperature biases in most regions, good simulation of

midlatitude synoptic variability and notable improvements in the simulation of tropical cyclones and ENSO. However, there are a number of systematic errors requiring further work, including excess global-mean precipitation and its regional biases, the outstanding interhemispheric SST bias and insufficient European blocking.

Currently, the Met Office does not run climate change simulations which involve feedbacks (such as historical simulations with time-evolving forcing) during its model development process, only effective radiative forcing tests. Hence, there has been no deliberate model tuning to affect the climate sensitivity. Climate change simulations using HadGEM3-GC3 are now in progress and will be reported in subsequent papers. For technical reasons associated with issues around NWP implementation of GA7, the GloSea5 seasonal forecast system will not upgrade to GC3 operationally, however, the system is expected to upgrade in the future to GC4 and hence incorporate changes described here at that point.

Consistent with the quasi-annual development cycle discussed in section 1, work is already underway to further develop GC3 to form GC4. Since all of the changes in GC3.1 relative to GC3.0 are believed to be physical improvements, they will be included in GC4 along with a routine basket of parametrization changes, the most notable of which will be introduction of prognostic convective entrainment. Hence, the assessment presented here also provides a baseline for evaluating this future development.

Appendix A: Description of Coupling

The atmosphere uses the Unified Model (UM; Cullen, 1993) and land surface uses JULES (the Joint UK Land Environment Simulator; Best et al., 2011). These component models run on the same grid and as part of the same model executable so can be considered to be "tightly coupled," passing data where necessary by subroutine arguments or shared data arrays. Similarly the ocean (NEMO [Nucleus for European Modelling of the Ocean]; Madec, 2008) and sea ice (CICE; Hunke et al., 2015) models are compiled into a single executable and are "tightly coupled" on the same grid (with the caveat that CICE uses an Arakawa B grid placement of velocities in contrast to the C grid in NEMO).

Any relevant details of the UM-JULES and NEMO-CICE coupling are largely covered by Walters et al. (2017b) and Hewitt et al. (2011), respectively, so here the focus is on the coupling of GA7-GL7 with GO6-GSI8 using the OASIS3 coupler (Valcke, 2013; Valcke et al., 2015).

The atmosphere and ocean models run concurrently with the OASIS coupler. Compared with GC2, GC3 uses the more scalable MCT version of OASIS3 which provides routines to be compiled with the atmosphere and ocean executables rather than using its own executable. OASIS restart dumps are not used and so all relevant fields to initialize the component models at start-up are stored in their restart dumps. Given that OASIS fulfills a technical and (relatively) simple interpolation task, it might be envisaged that the same coupled scientific configuration could be reproduced using an alternative coupler. This may theoretically be true but currently details of the way models are sequenced, along with interpolation options available, mean that OASIS3 (although not necessarily the specific code version) is considered to be part of the definition of GC3.

Table A1
Fields Exchanged Through the Coupler Between the Sea Ice Component and the Atmosphere-Land Component of HadGEM3-GC3

From UM-JULES to CICE	From CICE to UM-JULES
X component of wind stress (grid-box-mean; N m^{-2})	Ice area fraction (per category)
Y component of wind stress (grid-box-mean; N m^{-2})	Ice thickness (per category; m)
Rainfall rate (grid-box-mean; kg m^{-2} s^{-1})	Snow thickness (per category; m)
Snowfall rate (grid-box-mean; kg m^{-2} s^{-1})	Top layer ice temperature (per category; K)
Ice sublimation flux* (per category; W m^{-2})	Top layer effective conductivity (per category; W m^{-2} K^{-1})
Ice top melting* (per category; W m^{-2})	Melt-pond fraction (per category)
Ice conductive flux* (per category; W m^{-2})	Melt-pond thickness (per category; m)
Ice surface skin temperature (per category; K)	X component of sea ice velocity (grid-box-mean; m s^{-1})
	Y component of sea ice velocity (grid-box-mean; m s^{-1})

Note. Fields denoted * are subject to semiimplicit coupling. Sea ice velocities are combined with ocean velocities before passing through the coupler as described in Hewitt et al. (2011).

The momentum, freshwater, and heat fluxes passed from the atmosphere via OASIS to the ocean are largely as described in Hewitt et al. (2011). To ensure energy conservation, fluxes provided from the atmosphere are time-mean values over the model coupling interval and where necessary are scaled within the ocean or sea ice models to take account of the latest ice fraction. There are separate heat flux coupling fields for ocean and sea ice, however, the wind stress components provided from the atmosphere model are currently mean values which are assumed to apply equivalently to ocean and sea ice. When the grids are markedly different (e.g., N96 and ORCA025) there is an imprint of the atmosphere grid seen in some ocean fields. A second order conservative regridding of selected fields has been introduced for GC3 which smooths the data passed through the coupler, reducing this imprinting (although it is not completely removed). Only fields which are already relatively smooth can use second order regridding as otherwise any noise in fields can be increased. Therefore, second-order regridding is only used for evaporation, sublimation and net heat flux. The solar radiation field passed from the atmosphere allows the use of the RGB (red-green-blue) penetrative radiation scheme in the ocean model.

The coupling for the sea ice differs from that described in Hewitt et al. (2011), with the fields now exchanged listed in Table A1. Previously, with zero-layer sea ice thermodynamics, the JULES land surface model (which handles surface fluxes everywhere, not just over land) would calculate the diffusive heat flux all the way through the sea ice. For GSI8 with multilayer thermodynamics, CICE evolves a temperature profile, performing all calculations within the ice, and JULES calculates the surface exchange and the diffusive heat flux into the top ice layer only. To calculate the surface exchanges with multilayer thermodynamics, JULES requires extra coupling fields from CICE, top layer temperature and effective conductivity, to act as the lower boundary condition for the surface exchange calculations (West et al., 2016). The addition of prognostic melt-ponds at GC3 required additional coupling changes. Although the melt-pond evolution is calculated in CICE, the pond fraction and depth are used within the JULES surface exchange scheme to determine surface albedo. In previous configurations, the surface exchange fluxes were calculated for a single sea ice category. These calculations are now performed separately for each of the five ice thickness categories, as well as for open water. Finally, changes have been made to the way in which energy fluxes are passed from JULES to CICE to ensure energy is apportioned more realistically across the sea ice in underlying cells, while still conserving energy. Grid-box-mean fluxes in JULES are transformed to be "over sea ice fluxes" prior to being passed through the coupler; the reverse transformation is then performed once the fluxes are received in CICE. These coupling changes are described further in Ridley et al. (2017).

The coupling period is 1 h in GC3, compared with 3 h in GC2, to allow the diurnal cycle to be better resolved in both atmosphere and ocean boundary layers, and aid in reducing the inherent lag in ocean forcing fields as a result of running atmosphere and ocean components concurrently. To ensure conservation, coupling fields passed from atmosphere to ocean are 1 h time-averages.

Snow amounts on land ice gradually increase over time. In the real world, this is balanced through iceberg calving. In GC2, a uniform fresh water flux was applied to the extratropical oceans to balance the snow increase. In GC3, snow amounts over land ice are passed though the coupler to NEMO's prognostic iceberg scheme and ice shelf melting scheme, maintaining instantaneous water balance for both the ice sheets and ocean. This is a simple approach to ensure that in long-term equilibrium the water content of ice sheets and ocean do not drift. This surface mass balance and geographical distribution of calving will then be used to calibrate a constant iceberg calving for use in transient climate change simulations where it is expected that there will be a substantial buildup of snow which is not immediately discharged as icebergs.

Coupling fields (sea surface temperature, surface velocities, ice fraction, ice, and snow thickness) passed from the ocean to the atmosphere are instantaneous fields, but again at the new coupling frequency. Consistent with the treatment of momentum fluxes described above, the surface velocities passed to the atmosphere model are simply meaned ocean and sea ice values, weighted according to ice fraction.

Hewitt et al. (2011) described some of the choices made for the interpolation schemes for atmosphere to ocean and vice versa. These were made based on detailed assessment of regridding between the N96 atmosphere grid and the ORCA1 ocean grid and have not been reexamined for the higher resolution ORCA025 grid (although N216-ORCA025 is a comparable resolution ratio to N96-ORCA1, so similar conclusions are expected to be valid).

For long climate integrations, energy and freshwater budgets are clearly critical and so conservation of both heat and freshwater across the coupler has been checked in the GC2 and GC3 configurations and found to be accurate to within around 10^{-4} W m^{-2} (equivalent top of the atmosphere flux) and 10^{-5} Sv, respectively. These residual imbalances are smaller than the internal conservation errors of some of the individual model components and are therefore not viewed as significant.

Although OASIS3-MCT has the capability of generating interpolation weights at run-time, we continue to calculate these weights off-line using SCRIP (Jones, 1999). This is more efficient, traceable and allows some minor adjustments to be made where weights are otherwise calculated incorrectly due to complications caused by the north-fold of the tripolar ocean grid. The method for coupling the ocean component with the UM atmosphere is such that the ocean grid determines the coastline (with land fractions in all grid-boxes as either 0 or 1) but the atmosphere model then uses "coastal tiling" allowing the grid-box land fractions around the coast to take a value between 0 and 1 (calculated by interpolating the ocean land-sea mask onto the atmosphere grid). The atmosphere deals with incoming ocean data on fractional land grid-boxes by normalizing the values, based on the nonmasked source cell areas (as opposed to the alternative approach of normalizing based on the target cell areas). Equivalently when checking conservation for atmosphere to ocean fluxes, the atmosphere fields on coastal points must be multiplied by the ocean fraction.

The models use versions 2.0 (GC3.0) and 3.0 (GC3.1) of the "fully parallel" OASIS3-MCT coupler which has been demonstrated to be scalable for the number and size of subdomains deployed. Since OASIS3-MCT coupling adopts a library-based approach whereby the coupling code is linked directly to the component models there is no separate coupler executable as such. The cost of coupling transformations is thus included directly in the cost of the components themselves. Use of OASIS3-MCT makes load balancing and model configuration relatively simple since there are only two executable components to consider and only a single OASIS control file to set up.

Acknowledgments
This work was primarily supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). Part of the work was undertaken with National Capability funding from NERC for ocean modeling. The source code for the models used in this study, UM, JULES, and NEMO are available to use. To apply for a license for the UM go to http://www.metoffice.gov.uk/research/collaboration/um-collaboration and for permission to use JULES go to https://jules.jchmr.org. NEMO is available to download from www.nemo-ocean.eu. For more information on the exact model versions and branches applied, please contact the authors. Data from the simulations are archived at the Met Office and available for research use through the Centre for Environmental Data Analysis JASMIN platform (http://www.jasmin.ac.uk/); for details please contact UM_collaboration@metoffice.gov.uk referencing this paper. Identifiers for the data sets in the archive are HadGEM2-AO:ajpdr, HadGEM3-GC2:anqjm, HadGEM3-GC3.0:u-ab635, HadGEM3-GC3.1:u-ai599, GloSea5-GC2:u-ad686, GloSea5-GC3:mi-an938, NWP-GC2:mi-ab598_cplfcst_ctl, and NWP-GC3:u-ae499.

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). *Journal of Hydrometeorology*, *4*, 1147–1167.

Allan, R. J., & Ansell, T. J. (2006). A new globally complete monthly historical gridded mean sea level pressure data set (HadSLP2): 1850–2003. *Journal of Climate*, *19*, 5816–5842.

Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., & Vidale, P.-L. (2014). Changes in global net radiative imbalance 1985–2012. *Geophysical Research Letters*, *41*, 5588–5597. https://doi.org/10.1002/2014GL060962

Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., . . . Yang, S. T. (2013). Multi-model analysis of Northern Hemisphere winter blocking: Model biases and the role of resolution. *Journal of Geophysical Research: Atmospheres*, *118*, 3956–3971. https://doi.org/10.1002/jgrd.50231

Behrangi, A., Stephens, G., Adler, R. F., Huffman, G. F., Lambrigtsen, B., & Lebsock, M. (2014). An update on the oceanic precipitation rate and its zonal distribution in light of advanced observations from space. *Journal of Climate*, *27*, 3957–3965. https://doi.org/10.1175/JCLI-D-13-00679.1

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Mnard, C. B., . . . Harding, R. J. (2011). The joint uk land environment simulator (Jules), model description—Part 1: Energy and water fluxes. *Geoscientific Model Development*, *4*, 677–699. https://doi.org/10.5194/gmd-4-677-2011

Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., . . . Storkey, D. (2014). Recent development of the Met Office operational ocean forecasting system: An overview and assessment of the new global FOAM forecasts. *Geoscientific Model Development*, *7*, 2613–2638. https://doi.org/10.5194/gmd-7-2613-2014

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S., . . . John, V. O. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Bodas-Salcedo, A., Williams, K. D., Field, P. R., & Lock, A. P. (2012). The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: The role of midlatitude cyclone clouds. *Journal of Climate*, *25*, 7467–7486. https://doi.org/10.1175/JCLI-D-11-00702.1

Brewer, A. M. (1949). Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere. *Quarterly Journal of the Royal Meteorological Society*, *75*, 351–363.

Brown, A. R., Milton, S., Cullen, M., Golding, B., Mitchell, J., & Shelly, A. (2012). Unified modelling and prediction of weather and climate: A 25 year journey. *Bulletin of the American Meteorological Society*, *93*, 1865–1877. https://doi.org/10.1175/BAMS-D-12-00018.1

Cassou, C. (2008). Intraseasonal interaction between the Madden-Julian oscillation and the North Atlantic oscillation. *Nature*, *455*, 523–527.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J.-L., Minnis, P., . . . Zeng, S. (2010). The GCM Oriented Calipso Cloud Product (CALIPSO-GOCCP). *Journal of Geophysical Research: Atmospheres*, *115*, D00H16. https://doi.org/10.1029/2009JD012251

Chepfer, H., Bony, S., Winker, D., Chiriaco, M., Dufresne, J.-L., & Sèze, G. (2008). Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model. *Geophysical Research Letters*, *35*, L15704. https://doi.org/10.1029/2008GL034207

Chu, J.-H., Sampson, C. R., Levine, A. S., & Fukada, E. (2002). *The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000* (Tech. Rep. NRL/MR/7540-02-16, 112 pp.). Washington, DC: Naval Research Laboratory.

Cullen, M. J. P. (1993). The unified forecast/climate model. *Meteorological Magazine*, *122*, 81–94.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, *109*, C12003. https://doi.org/10.1029/2004JC002378

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Eade, R., Smith, D. M., Scaife, A. A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, *41*, 5620–5628. https://doi.org/10.1002/2014GL061146

Ferranti, L., Corti, S., & Janousek, M. (2015). Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, *141*, 916–924. https://doi.org/10.1002/qj.2411

Furtado, K., Field, P. R., Boutle, I. R., Morcrette, C. J., & Wilkinson, J. M. (2016). A physically-based, subgrid parametrization for the production and maintenance of mixed-phase clouds in a General Circulation Model. *Journal of the Atmospheric Sciences*, *73*, 279–291. https://doi.org/10.1175/JAS-D-15-0021.1

Ganachaud, A., & Wunsch, C. (2003). Large-scale ocean heat and freshwater transports during the World Ocean Circulation Experiment. *Journal of Climate*, *16*(4), 696–705.

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*, 6704–6716. https://doi.org/10.1002/2013JC009067

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

Heming, J. T. (2016). Met Office Unified Model tropical cyclone performance following major changes to the initialization scheme and a model upgrade. *Weather and Forecasting*, *31*, 1433–1449. https://doi.org/10.1175/WAF-D-16-0040.1

Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., . . . Hunke, E. C. (2011). Design and implementation of the infrastructure of HadGEM3: The next-generation Met Office climate modelling system. *Geoscientific Model Development*, *4*, 223–253. https://doi.org/10.5194/gmd-4-223-2011

Hodges, K. (1995). Feature tracking on a unit sphere. *Monthly Weather Review*, *123*, 3458–3465.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. M. (2015). *CICE: The Los Alamos sea ice model, documentation and software, version 5.1* (Tech. Rep. LA-CC-06–012). Los Alamos, NM: Los Alamos National Laboratory.

Ingleby, B., & Huddleston, M. (2007). Quality control of ocean temperature and salinity profiles—Historical and real-time data. *Journal of Marine Systems*, *65*, 158–175. https://doi.org/10.1016/j.jmarsys.2005.11.019

Jiang, J. H., Su, H., Zhai, C., Perun, V. S., Genio, A. D., Nazarenko, L. S., . . . Stephens, G. L. (2012). Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations. *Journal of Geophysical Research: Atmospheres*, *117*, D14105. https://doi.org/10.1029/2011JD017237

Jones, P. W. (1999). First and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, *127*, 2204–2210.

Josey, S. A., Kent, E. C., & Taylor, P. K. (1998). *The Southampton Oceanography Centre (SOC) ocean—Atmosphere heat, momentum and freshwater flux atlas* (Rep. 6, 55 pp.). Southampton, UK: Southampton Oceanography Centre, European Way.

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Gleckler, P. J. (2013). Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *Journal of Geophysical Research: Atmospheres*, *118*, 1329–1342. https://doi.org/10.1002/jgrd.50141

Landsea, C. W., & Franklin, J. L. (2013). Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, *141*, 3576–3592. https://doi.org/10.1175/MWR-D-12-00254.1

Laxon, S., Giles, K. A., Ridout, A., Wingham, D. J., Willatt, R., Cullen, R., . . . Davidson, M. (2013). Cryosat-2 estimates of Arctic sea ice thickness and volume. *Geophysical Research Letters*, *40*, 732–737. https://doi.org/10.1002/grl.50193

Liu, C., Allan, R. P., Berrisford, P., Mayer, M., Hyder, P., Loeb, N., . . . Edwards, J. M. (2015). Combining satellite observations and reanalysis energy transports to estimate global net surface energy fluxes 1985–2012. *Journal of Geophysical Research: Atmospheres*, *120*, 9374–9389. https://doi.org/10.1002/2015JD023264

Liu, Y., Daum, P. H., Guo, H., & Peng, Y. (2008). Dispersion bias, dispersion effect, and the aerosol–cloud conundrum. *Environmental Research Letters*, *3*, 045021. https://doi.org/10.1088/1748-9326/3/4/045021

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Kato, S., Wong, T., Smith, G. L., . . . Manalo-Smith, N. (2009). Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, *22*(3), 748–766. https://doi.org/10.1175/2008JCLI2637.1

MacLachlan, C., Arribas, A., Peterson, D., Maidens, A., Fereday, D., Scaife, A. A., . . . Madec, G. (2015). Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, *141*, 1072–1084. https://doi.org/10.1002/qj.2396

Madec, G. (2008). *NEMO ocean engine* (Tech. Rep. 27). Paris, France: Note du Pole de modélisation, Institut Pierre-Simon Laplace (IPSL).

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

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., . . . Sinha, B. (2014). GO5.0: The joint NERC-Met Office NEMO Global Ocean model for use in coupled and forced applications. *Geoscientific Model Development*, *7*, 1069–1092. https://doi.org/10.5194/gmd-7-1069-2014

Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E. K., Bulgin, C. E., Corlett, G. K., . . . Donlon, C. (2014). *ESA sea surface temperature climate change initiative (ESA SST CCI): Analysis long term product version 1.0*. Oxford, UK: NERC Earth Observation Data Centre. https://doi.org/10.5285/878bef44-d32a-40cd-a02d-49b6286f0ea4

Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, *143*(11), 4631–4644.

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The 'too few, too bright' tropical low-cloud problem in CMIP5 models. *Geophysical Research Letters*, *39*, L21801. https://doi.org/10.1029/2012GL053421

Rae, J. G. L., Hewitt, H. T., Keen, A. B., Ridley, J. K., West, A. E., Harris, C. M., . . . Walters, D. N. (2015). Development of the Global Sea Ice 6.0 CICE configuration for the Met Office Global Coupled model. *Geoscientific Model Development*, *8*, 2221–2230. https://doi.org/10.5194/gmd-8-2221-2015

Rashid, H. A., & Hirst, A. C. (2016). Mechanisms of improved rainfall simulation over the Maritime Continent due to increased horizontal resolution in an AGCM. *Climate Dynamics*, *49*, 1747–1764. https://doi.org/10.1007/s00382-016-3413-z

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

Ridley, J., Blockley, E., Keen, A., Rae, J., Schroder, D., West, A., & Schroeder, D. (2017). The sea ice model component of HadGEM3-GC3.1. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-2017-212, in press.

Sanchez, C., Williams, K. D., & Collins, M. (2016). Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society*, *142*, 147–159. https://doi.org/10.1002/qj.2640

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., . . . Williams, A. (2014). Skillful long-range prediction of european and north american winters. *Geophysical Research Letters*, *41*, 2514–2519. https://doi.org/10.1002/2014GL059637

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., . . . Slingo, J. (2017). Tropical rainfall, rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, *143*, 1–11. https://doi.org/10.1002/qj.2910

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., . . . Williams, K. (2011). Improved Atlantic winter blocking in a climate model. *Geophysical Research Letters*, *38*, L23703. https://doi.org/10.1029/2011GL049573

Scherrer, S. C., Croci-Maspoli, M., Schwierz, C., & Appenzeller, C. (2006). Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *International Journal of Climatology*, *26*, 233–249. https://doi.org/10.1002/joc.1250

Schiemann, R., Demory, M.-E., Shaffrey, L. C., Strachan, J., Vidale, P. L., Mizielinski, M. S., . . . Jung, T. (2017). The resolution sensitivity of northern hemisphere blocking in four 25-km atmospheric global circulation models. *Journal of Climate*, *30*, 337–358. https://doi.org/10.1175/JCLI-D-16-0100.1

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., & Kwok, R. (2011). Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research*, *116*, C00D06. https://doi.org/10.1029/2011JC007084

Storkey, D., Megann, A., Mathiot, P., Sinha, B., Calvert, D., Hewitt, H., Blaker, A., Kuhlbrodt, T., Graham, T., & Hyder, P. (2017). UK Global Ocean GO6 and GO7: A traceable hierarchy of model resolutions. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-2017-263

Tomassini, T., Parker, D. J., Stirling, A., Bain, C., Senior, C. A., & Milton, S. F. (2017). The interaction between moist diabatic processes and the atmospheric circulation in African Easterly Wave propagation. *Quarterly Journal of the Royal Meteorological Society*, *143*, 3207–3227. https://doi.org/10.1002/qj.3173

Valcke, S. (2013). The OASIS3 coupler: A European climate modelling community software. *Geoscientific Model Development*, *6*, 373–388. https://doi.org/10.5194/gmd-6-373-2013

Valcke, S., Craig, T., & Coquart, L. (2015). *OASIS3-MCT 3.0, OASIS3-MCT User Guide TR/CMGC/15/38*. CERFACS. Retrieved from http://www.cerfacs.fr/oa4web/oasis3-mct_3.0/oasis3mct_UserGuide.pdf

Vosper, S. B., Wells, H., & Brown, A. R. (2009). Accounting for non-uniform static stability in orographic drag parametrization. *Quarterly Journal of the Royal Meteorological Society*, *135*, 815–822. https://doi.org/10.1002/qj.407

Walters, D., Boutle, I., Brooks, M., Melvin, T., Stratton, R., Vosper, S., . . . Xavier, P. (2017a). The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations. *Geoscientific Model Development*, *10*, 1487–1520. https://doi.org/10.5194/gmd-10-1487-2017

Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D., . . . Williams, K. D. (2011). The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations. *Geoscientific Model Development*, *4*, 919–941. https://doi.org/10.5194/gmd-4-919-2011

Walters, D. N., A., Baran, I., Boutle, M. E., Brooks, P., Earnshaw, J., Edwards, K., . . . Zerroukat, (2017b). The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-2017-291, in press.

Waters, J., Lea, D. J., Martin, M. J., Mirouze, I., Weaver, A., & White, J. (2015). Implementing a variational data assimilation system in an operational 1/4 degree Global Ocean model. *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 333–349.

Watterson, I. G., Bathols, J., & Heady, C. (2014). What influences the skill of climate models over the continents? *Bulletin of the American Meteorological Society*, *95*, 689–700. https://doi.org/10.1175/BAMS-D-12-00136.1

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., . . . Watanabe, M. (2017). The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, *10*, 359–384. https://doi.org/10.5194/gmd-10-359-2017

West, A. E., McLaren, A. J., Hewitt, H. T., & Best, M. J. (2016). The location of the thermodynamic atmosphere-ice interface in fully coupled models—A case study using JULES and CICE. *Geoscientific Model Development*, *9*, 1125–1141. https://doi.org/10.5194/gmd-9-1125-2016

Wheeler, M., & Kiladis, G. N. (1999). Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber–frequency domain. *Journal of the Atmospheric Sciences*, *56*, 374–399.

Williams, K. D., & Bodas-Salcedo, A. (2017). A multi-diagnostic approach to cloud evaluation. *Geoscientific Model Development*, *10*, 2547–2566. https://doi.org/10.5194/gmd-10-2547-2017

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

Williamson, D. L., Blackburn, M., Hoskins, B. J., Nakajima, K., Ohfuchi, W., Takahashi, Y. O., . . . Stratton, R. (2012). *The APE atlas* (Tech. Note NCAR/TN-484+STR). Boulder, CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6FF3QBR

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., & Morcrette, C. J. (2008). PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Quarterly Journal of the Royal Meteorological Society*, *134*, 2093–2107. https://doi.org/10.1002/qj.333

Winker, D. M., Pelon, J., Coakley, J. A., Jr., Ackerman, S. A., Charlson, R. J., Colarco, P. R., . . . Wielicki, B. A. (2010). The CALIPSO mission. A global 3D view of aerosols and clouds. *Bulletin of the American Meteorological Society*, *91*(9), 1211–1229. https://doi.org/10.1175/2010BAMS3009.1

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

Woollings, T., Hannachi, A., & Hoskins, B. (2010). Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, *136*, 856–868.

Xie, P., & Arkin, P. A. (1997). Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bulletin of the American Meteorological Society*, *78*(11), 2539–2558.