Below is a detailed response to each of the points made by the two referees, plus the revised paper with the modified text highlighted in red. We have also added RMS errors to Fig. 6 & 8, panel a to Fig. 15 and updated the references.

Response to referee #1

Reply to general comments

We thank the referee for his/her review and helpful suggestions. We note the point that readers will need to consult several papers for a full description of GC2. As the referee notes, this is unavoidable for large models, however we intend to form a special issue of GMD containing the 4 papers which together provide this full description.

Evidence for conclusion on ENSO

The referee raises a concern that insufficient evidence is given for the statements about ENSO being well simulated and also questions the reliability of the results given the length of simulation.

To address this, we have added a new table (Table 3) which contains 5 metrics: M1 and M2 are standard deviation of monthly SST anomaly for regions Nino3 (90W-150W, 5N-5S) and Nino4 (160E-150W, 5N-5S), M3 is the ratio of power in the 3-7 year range relative to 0-10 years for monthly Nino3 SST anomaly (%), M4 is a seasonality metric defined as the ratio of November to January and March to May standard deviation of Nino3 SST anomaly (Bellenger et al., 2014), M5 is the standard deviation of precipitation anomaly for Nino4 (mm day-1). We present these for the final 50 years of the CLIM experiment (as elsewhere in the paper), but also for the final 100 years of a second simulation which is equivalent to the CLIM experiment except for different initial conditions. The results are shown to be similar between the two. We have also added an extra panel to Fig. 15 showing the mean Pacific SST averaged 5N-5S.

The text on p537 lines15-21 has been revised as follows: *As a result of the improved wind-stress, improved MJO (which can be the source of westerly wind bursts; e.g. Lengaigne et al., 2004) and higher horizontal resolution of the ocean, ENSO is well simulated in HadGEM3-GC2 with a good spatial pattern (Fig. 15c&d). When assessed against a range of metrics (Table 3) we see that variability in SST agrees well with observations in the*

central-east Pacific although is somewhat weaker than observed near the dateline. A power spectrum analysis shows the frequency lies within the observed range (3 to 7 years), with no dominant short (e.g. 2 year) or longer period peaks. The model seasonality is good, with maximum (minimum) variability in boreal winter (spring). The standard deviation of precipitation in the central Pacific gives a measure of model capability for regional climate impacts and although slightly underestimated, is good in comparison with other climate models which tend to underestimate this quantity. Overall, HadGEM3-CG2 compares favourably with a range of CMIP5 models (Bellenger et al., 2014). The main observed ENSO teleconnections to remote precipitation anomalies (S. America, Sahel, India, E. Africa, etc.) are also present in the model (not shown).

Responses to Minor concerns

Clarity of experiment being used

As the referee suggests, the experiment being used (CLIM, SEAS, NWP) has been added to each of the figure and table captions.

Conservation in the coupling

We have expanded the text to make the description of the conservative coupling clearer. p526 lines5-11 have now been replaced with *To ensure energy conservation, the coupling part of the NEMO name-list is set to ensure that in most cases there are separate coupling fields received in NEMO as relevant to ocean (solar and non-solar heat fluxes; evaporation) or sea ice (top and bottom conductive heat fluxes as calculated in the JULES land surface model; sublimation). These fields are converted to mean values over atmosphere grid boxes before being conservatively interpolated by OASIS, and once received by NEMO are applied to the ocean or sea ice component as appropriate. Where necessary, CICE can pass any excess heat or freshwater fluxes back to NEMO – this may be required if the interpolation of coupling fields produces sea ice fluxes in ocean grid boxes without sea ice, or if the sea ice melts either between coupling exchanges or within a CICE time-step. The wind*

3

stress components provided from the atmosphere model are currently mean values which are assumed to apply equivalently to ocean and sea ice.

Quantification of model drift

As the referee suggests, the following has been added to p528 line24 *The largest global-mean ocean temperature drift over the 100 year simulation occurs at a depth of 563m with a rate of 0.11K/decade. Over the final 50 years this reduces to 0.08K/decade. Below 1000m the average drift is less than 0.02K/decade at all depths.*

Figure 15

As the referee suggests, the longitude axes of the panels on Fig. 15 has been reformed to use the same tick interval to make comparison clear.

Response to referee #2

Reply to general comments

We thank the referee for his/her review and positive comments.

Responses to Minor concerns

Comment #1

We fully accept the referee's point that the full set of papers is needed to have a complete description of GC2 and it is undesirable to have some available without others. We had intended for all of the papers to be submitted around the same time, however submission of the GA6 description paper (Walters et. al.) has been delayed due to unexpected operational issues taking the time of key authors of that paper. Nevertheless we expect it to be submitted within the next couple of months. All of the other papers (including the GSI6 paper by Rae et. al.) have now been submitted or accepted in GMD (references have been updated). We intend to form a special issue of GMD to collect together the 4 papers forming the description of GC2.

Comment #2

At the Met Office we make a distinction between the physical model which is is used across all timescales, and an earth system model (ESM) which uses one of our physical models but contains additional processes (interactive vegetation, carbon cycle, chemistry, etc.) which only become essential for climate change projections on centennial and longer timescales. Developing a full ESM model takes a considerable amount of human and computational resource and most centres who only develop a climate model will build a physical and earth system model around once every 6 years (generally to feed into the CMIP/IPCC process). At the Met Office, we require updating of the physical more often for systems making forecasts on shorter timescales (seasonal and NWP) hence our physical model is developed on an

annual timescale. Whilst it might be desirable to build a full ESM annually, the resource required make this impossible. We have added *Due to the additional resource required to build an earth system model...* before the sentence starting *The intention...* on p523 line27 in order to make our reasoning clearer.

Comment #3

As the referee suggests, the following has been added after p528 line9: *On 36 nodes of the Met Office IBM Power7 machine, HadGEM3-GC2 at N216-ORCA025 achieves 1.87 simulated years per wall clock day. Of the 36 nodes, 17 (544 processors) are used by the atmosphere, 18.75 (600 processors) by the ocean and the remaining 8 processors by OASIS.* Only N216 is given since this is the resolution presented in the paper, is what is being used operationally for seasonal forecasts and will be the resolution used for most of the climate work to be conducted with GC2.

Comment #4

As the referee suggests, the area weighted RMS differences have been added to Figures 6 & 8. The following has been added to p531 line 15: *Overall the area-weighted root mean square (RMS) error for the field is reduced from 2.02 to 1.55.* and the following to p532 line 18: *The RMS error is slightly increased (from 1.68 to 1.76), primarily due to an increased mean bias over the tropical west Pacific and east Indian Ocean.*

The Met Office Global Coupled model 2.0 (GC2) configuration

K. D. Williams[1]**, C. M. Harris**[1]**, A. Bodas-Salcedo**[1]**, J. Camp**[1]**, R. E. Comer**[1]**,
D. Copsey**[1]**, D. Fereday**[1]**, T. Graham**[1]**, R. Hill**[1]**, T. Hinton**[1]**, P. Hyder**[1]**, S. Ineson**[1]**,
G. Masato**[2]**, S. F. Milton**[1]**, M. J. Roberts**[1]**, D. P. Rowell**[1]**, C. Sanchez**[1]**, A. Shelly**[1]**,
B. Sinha**[3]**, D. N. Walters**[1]**, A. West**[1]**, T. Woollings**[4]**, and P. K. Xavier**[1]

[1]Met Office, Exeter, UK
[2]University of Reading, Reading, UK
[3]National Oceanography Centre, Southampton, UK
[4]Atmospheric, Oceanic and Planetary Physics, Oxford, UK

Correspondence to: K. D. Williams (keith.williams@metoffice.gov.uk)

Abstract

The latest coupled configuration of the Met Office Unified Model (Global Coupled configuration 2, GC2) is presented. This paper documents the model components which make up the configuration (although the scientific description of these components is detailed elsewhere) and provides a description of the coupling between the components. The performance of GC2 in terms of its systematic errors is assessed using a variety of diagnostic techniques. The configuration is intended to be used by the Met Office and collaborating institutes across a range of timescales, with the seasonal forecast system (GloSea5) and climate projection system (HadGEM) being the initial users. In this paper GC2 is compared against the model currently used operationally in those two systems.

Overall GC2 is shown to be an improvement on the configurations used currently, particularly in terms of modes of variability (e.g. mid-latitude and tropical cyclone intensities, the Madden–Julian Oscillation and El-Niño Southern Oscillation). A number of outstanding errors are identified with the most significant being a considerable warm bias over the Southern Ocean and a dry precipitation bias in the Indian and west African summer monsoons. Research to address these is ongoing.

1 Introduction

The Met Office produces forecasts across a range of timescales from numerical weather predictions (NWP) for days ahead or less, through monthly-seasonal-decadal forecasts, to climate change projections. For over 20 years, the framework of the Met Office Unified Model (MetUM, Cullen, 1993) has been used to produce models which between them span these timescales. Over the last few years, the development of the science within the MetUM has been made more seamless across timescales than ever before, with numerous benefits including greater scientific robustness of the model, improved ability to investigate model biases and a more efficient use of resources (Brown et al., 2012). Model development now progresses on an approximately annual timescale 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 each year for use across timescales by the Met Office and its collaborators (Walters et al., 2011).

The latest configuration of the coupled model, released March 2014, is known as Global Coupled model 2.0 (GC2). This is comprised of component configurations Global Atmosphere 6.0 (GA6.0), Global Land 6.0 (GL6.0), Global Ocean 5.0 (GO5.0) and Global Sea Ice 6.0 (GSI6.0). GA6.0 and GL6.0 are fully documented by Walters et al. (2015), whilst GO5.0 is described by Megann et al. (2014) and GSI6.0 by Rae et al. (2015). In this paper we provide a technical description of the coupling between the components and then present the coupled model performance in terms of systematic errors through a range of diagnostic techniques. We do not discuss predictions/projections from GC2 as these will be presented elsewhere (e.g. Senior et al., 2015). Currently, coupled models are used in Met Office systems on monthly and longer timescales, hence most of the results presented here will be for the seasonal forecasting system (referred to as GloSea5-GC2) and the climate model (referred to as HadGEM3-GC2). In each case, comparisons will be made against the current "operational" configuration which is GloSea5-GA3 for seasonal (MacLachlan et al., 2015) and HadGEM2-AO for climate (The HadGEM2 Development Team, 2011). It should be noted that, unlike the new GC2 configuration which is identical in the different systems, these two control configurations from which we are upgrading differ significantly. It is envisaged that future coupled configurations will also start to be used on shorter NWP timescales, hence a few results are also included from these timescales.

The "physical" model presented here does not include earth system components such as interactive vegetation or ocean bio-geochemistry (note, our definition of the physical model does include interactive aerosols). Due to the additional resource required to build an earth system model (ESM), the intention is that earth system components will be built on top of a subset of the annual physical model releases to form an ESM every six years or so. GC2 will not be a configuration to which earth system components will be added, although it is envisaged that the next coupled model release (GC3) will be developed into an ESM.

3

In the next section we provide details of the coupling and experiments subsequently presented. In Sect. 3 the climatological biases of the model are discussed, whilst systematic errors in mid-latitude variability are presented in Sect. 4, and tropical variability in Sect. 5. We summarise in Sect. 6.

2 Coupled model details

The GC2 configuration is defined by the combination of the component model scientific configurations (GA6.0, GL6.0, GO5.0, GSI6.0) and associated 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. (2015), Megann et al. (2014) and Rae et al. (2015), whilst the technical details of the coupling are described below.

Relative to GloSea5-GA3, GC2 has a significant revision to the atmosphere dynamical core and a number of parametrization revisions. HadGEM2-AO predates GloSea5-GA3, so relative to HadGEM2-AO, GC2 has additional changes including a new ocean model, new sea-ice model, new cloud scheme, and considerable revisions to all of the existing parametrization schemes.

The vertical resolution is set by the component 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) and 5 sea-ice thickness categories. The ocean resolution is 0.25° on a tri-polar grid. The GA6 science can be run over a wide range of horizontal resolutions on a regular latitude–longitude grid with no explicit changes to model parametrizations, however results presented here all use a horizontal resolution of N216 (60 km in mid-latitudes). The atmosphere and ocean horizontal and vertical resolutions presented here are an increase on HadGEM2-AO (which uses an N96 (135 km) L38 atmosphere and 1° L40 ocean) but the same as GloSea5-GA3.

2.1 Description of coupling

The atmosphere (UM) and land surface (JULES (the Joint UK Land Environment Simulator); Best et al., 2011) 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 and Lipscomb, 2004) 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. (2015) and Megann et al. (2014) respectively, so here the focus is on the coupling of GA6.0/GL6.0 with GO5.0/GSI6.0 using the OASIS3 coupler (Valcke, 2013). As already mentioned, although the atmosphere (plus land surface) science can be run over a wide range of horizontal resolutions, this is not true for the ocean (plus sea ice) configuration which is fixed at $0.25°$ (using the ORCA025 tri-polar grid; Madec, 2008). This means that GC2 coupled configurations are limited to those using an ORCA025 ocean. At present no resolution dependent choices have been made in the details of the atmosphere–ocean coupling although this will not necessarily be true in all future GC configurations.

The coupled model infrastructure remains essentially unchanged from that described by Hewitt et al. (2011). The atmosphere and ocean models run concurrently with OASIS3 (now at version 3.0) handling the exchange and interpolation of model fields between the two executables. OASIS restart dumps are not used and so all relevant fields to initialise 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 GC2.

The momentum, freshwater and heat fluxes passed from the atmosphere via OASIS to the ocean are largely as described for "HadGEM3-AO r1.1" in Hewitt et al. (2011). To ensure energy conservation, the coupling part of the NEMO name-list is set to ensure that in most cases there are separate coupling fields received in NEMO as relevant to ocean (solar

and non-solar heat fluxes; evaporation) or sea ice (top and bottom conductive heat fluxes as calculated in the JULES land surface model; sublimation). These fields are converted to mean values over atmosphere grid boxes before being conservatively interpolated by OASIS, and once received by NEMO are applied to the ocean or sea ice component as appropriate. Where necessary, CICE can pass any excess heat or freshwater fluxes back to NEMO – this may be required if the interpolation of coupling fields produces sea ice fluxes in ocean grid boxes without sea ice, or if the sea ice melts either between coupling exchanges or within a CICE time-step. The wind stress components provided from the atmosphere model are currently mean values which are assumed to apply equivalently to ocean and sea ice.

There are a number of minor changes since the configuration described by Hewitt et al. (2011). Firstly the coupling period is now 3 h for most GC2 simulations to allow the diurnal cycle to be better resolved in both atmosphere and ocean boundary layers (the NWP simulations use hourly coupling – this is something we intend to unify across timescales in future configurations and will 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 still time-averages but now over a 3 h (1 h for NWP) rather than a 24 h period. In addition, a constant field representing iceberg calving is now added to the run-off field within the atmosphere model before passing to OASIS. There has also been a change to the solar radiation field passed from the atmosphere to allow the use of the RGB (red-green-blue) penetrative radiation scheme in GO5.0.

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 re-examined for the higher resolution ORCA025 grid (although N216-ORCA025 is a comparable resolution combination to N96-ORCA1, so similar conclusions are expected to be valid). Hence, with the exception of vector fields which all use bi-linear interpolation, atmosphere to ocean fields are regridded using first order conservative interpolation (to avoid undershoots and overshoots for fields which must be positive everywhere) whereas second order conservative interpolation is used for ocean to atmosphere fields.

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 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 numbers are smaller than the internal conservation errors of some of the individual model components and are therefore not viewed as significant.

Although OASIS3 has the capability of generating interpolation weights at run-time, we continue to calculate these weights off-line using SCRIP (Jones, 1999). This is much more efficient, traceable and also allows some minor adjustments to be made where weights are otherwise calculated incorrectly due to complications caused by the north-fold of the tri-polar ocean grid. The method for coupling the ocean component with the UM atmosphere is such that the ocean grid determines the coast-line (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). A consequence of the way the atmosphere deals with ocean information on these fractional land grid-boxes is that when ocean fields are regridded to the atmosphere the OASIS3 "FRACAREA" option is used (rather than the standard "DESTAREA"). Equivalently when checking conservation for atmosphere to ocean fluxes, the atmosphere fields on coastal points must be multiplied by the land fraction.

The technical details of model set-up are dependent on the machine architecture being used, but typically when running with several hundred processors for both atmosphere and ocean components (e.g. on the IBM Power7 machine), the "pseudo-parallel" capability

of OASIS3 is used such that the various coupling fields are typically distributed between eight OASIS3 processes in order to reduced elapsed time for coupling exchanges. This has been shown to provide satisfactory performance without the coupling being a significant overhead on model run time. Given that atmosphere and ocean in GC2 run concurrently it is necessary though to ensure that the model is well "load-balanced" to minimise time when processors are standing idle. On 36 nodes of the Met Office IBM Power7 machine, HadGEM3-GC2 at N216-ORCA025 achieves 1.87 simulated years per wall clock day. Of the 36 nodes, 17 (544 processors) are used by the atmosphere, 18.75 (600 processors) by the ocean and the remaining 8 processors by OASIS.

2.2 Experimental design

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

CLIM 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 the Coupled Model Intercomparison Project 3, CMIP3). Where appropriate (e.g. for aerosol emissions), these forcings vary through the annual cycle. The ocean is initialised from EN3 climatology (Ingleby and Huddleston, 2007). The top-of-atmosphere radiative imbalance in a parallel atmosphere-only simulation is $0.8\,\mathrm{W\,m^{-2}}$, consistent with using present-day forcings, hence a small drift due to the net energy flux would be expected. Average results from the final 50 years of the simulation are shown unless otherwise stated. For those variability diagnostics using high temporal resolution (e.g. daily) data, the final 20 years of the simulation are used. The largest global-mean ocean temperature drift over the 100 year simulation occurs at a depth of 563m with a rate of 0.11 K/decade. Over the final 50 years this reduces to 0.08 K/decade. Below 1000m the average drift is less than 0.02 K/decade at all depths.

Results presented from SEAS are a mean of seasonal hindcasts for the years 1996–2009, each of 140 days in length. Within each year, there are 3 DJF hindcasts initialised on 25 October, 1 November, 9 November and 3 JJA hindcasts initialised on 25 April, 1

May, 9 May and each start date has a 3-member initial condition ensemble, resulting in 120 hindcasts being averaged for each of DJF and JJA. The ocean and sea-ice are initialised from Met Office Ocean Forecast analyses, the atmosphere from ECMWF analyses and soil moisture from a climatology of the land surface model used within GC2 forced with ECMWF analyses. More details on the initialisation can be found in MacLachlan et al. (2015).

The NWP experiment comprises 15 day hindcasts, run daily at 12Z for the period 2–14 December 2011. The atmosphere and land surface are initialised from Met Office NWP analyses, and ocean from Met Office Ocean Forecast analyses. Both NWP and SEAS use prescribed aerosol concentrations from a HadGEM2-AO AMIP (Atmosphere Model Inter-comparison Project) simulation, but with direct and indirect effects being calculated interactively as for CLIM (Walters et al., 2015).

3 GC2 mean biases

HadGEM2-AO is characterised by a cold SST bias over much of world, especially in the North Atlantic, with a slight warm bias over the Southern Ocean and Southern Hemisphere stratocumulus regions (Fig. 1). The change to the new NEMO ocean model and higher ocean resolution has resulted in GC2 SSTs being generally warmer, which is beneficial over most regions, but detrimental over the Southern Ocean.

A considerable amount of work is ongoing to investigate the Southern Ocean warm bias in the Met Office model (e.g. Bodas-Salcedo et al., 2012). To first order, the surface flux biases are similar in AMIP simulations parallel to HadGEM2-AO and HadGEM3-GC2 (Fig. 2a), both having a large downwards surface flux bias over the Southern Ocean which is only slightly worse in HadGEM3-GC2. However, the coupled SST (and upper ocean heat content) biases are much larger in HadGEM3-GC2 than HadGEM2-AO. This appears to be related to changes to both the lateral and vertical ocean heat transports associated with the change in ocean model and ocean resolution. The HadGEM3-GC2 errors also include a contribution associated with too shallow Southern Ocean summer mixed layers. A detailed analyses of this problem is currently underway which will be documented

separately, although it is believed that the primary problem is the atmospheric heat flux biases (e.g. Trenberth and Fasullo, 2010; Williams et al., 2013; Bodas-Salcedo et al., 2014). In GC2 both excess downward SW flux and too little upward heat transport from turbulent fluxes are thought to contribute to the net heat flux bias and these are the focus of our efforts to improve future configurations.

The warm bias over the Southern Ocean can be seen early in NWP simulations for the austral summer (Fig. 3). Hindcasts for the austral winter with an earlier configuration did not show such a rapid warming (not shown), providing further evidence that fast atmosphere processes are contributing, and that biases in the SW flux are likely to be significant.

The increase in SSTs in HadGEM3-GC2 is particularly notable over the North Atlantic to the south of Greenland where, in common with many climate models, HadGEM2-AO has a very large cold bias (Fig. 1). Here, the higher horizontal resolution of the ocean model leads to a significantly improved Gulf stream extension, accurately reproducing the northward turn around Newfoundland. Scaife et al. (2011) have shown the importance of this SST improvement for European climate variability. On seasonal timescales, these relatively small SST biases in the North Atlantic have been further improved between GloSea5-GA3 and GloSea5-GC2 by introducing aerosol indirect effects from the aerosol climatological concentrations, consistent with climate model simulations, rather than using fixed droplet concentrations for land and sea. As a result, the JJA hindcasts in particular have improved to now match the observed seasonal cycle very well (Fig. 4).

The cold SST bias in HadGEM2-AO impacted the atmosphere with DJF biases of over 6 K in the boundary layer at high latitudes, but also biases of 2 K extending through much of the troposphere in the Northern Hemisphere (Fig. 5). The improved SSTs in HadGEM3-GC2 mean that these biases have reduced, although the troposphere remains slightly cool. An exception is over the Southern Ocean where the warm bias has increased. HadGEM2-AO has a warm stratosphere with a 5 K temperature bias in the tropics at 70 hPa. This warm bias is of concern when developing ESMs with interactive chemistry as these processes are particularly sensitive to the stratospheric temperature and humidity. The stratospheric specific humidity is largely determined by the cold point temperature (Brewer, 1949), hence

minimising this bias is particularly important. It can be seen that this tropical tropopause warm bias has been considerably improved in HadGEM3-GC2 through a combination of many parametrization changes, changes to the dynamics and increased vertical resolution.

In common with many climate models (e.g. Klein et al., 2006; Ma et al., 2014), HadGEM2-AO develops a warm bias over mid-latitude continents in summer (Fig. 6). This bias is over 6 K in atmosphere-only simulations, but is mitigated in coupled simulations through the cold Northern Hemisphere SST bias in HadGEM2-AO. The summer warm bias is reduced in HadGEM3-GC2 through developments to the land surface (Martin et al., 2010) and parametrization improvements (including increasing the frequency of radiation calls from 3 hourly to hourly). Changes elsewhere are generally small, although a cold bias is now starting to develop at high latitudes, consistent with the troposphere remaining a little cold. Overall the area-weighted root mean square (RMS) error for the field is reduced from 2.02 to 1.55.

The accurate simulation of sea-ice extent and thickness in the present-day climate is of importance for the estimation of climate sensitivity (Hall and Qu, 2006), 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). For GC2, the sea ice parameters in the model were tuned within the range of observational uncertainty (Rae et al., 2015), and the sea ice simulation generally does a reasonable job of simulating the annual cycles of Arctic sea ice extent and volume (Fig. 7). The Arctic sea ice volume simulation is most accurate at N216 resolution, where throughout the year the ice is around 20 % thicker than at N96, on a spatial average. The warm SST bias over the Southern Ocean results in there being far too little Antarctic sea-ice, which further exacerbates the bias.

The mean value of the Atlantic meridional overturning circulation (MOC) at 26° N in HadGEM3-GC2 (over years 11–50) is 16.4 Sv. This is close to the observed value from the RAPID array, particularly since a downward trend has been observed since 2004 (average observed value 2004–2012 is 17.5 Sv; Smeed et al., 2013), and is an improvement over HadGEM2-AO in which the MOC was around 15.1 Sv. However, the depth of the North Atlantic Deep Water return flow remains too shallow. This is a common bias in z-level mod-

els which tend not to simulate overflows well as there is excess entrainment. The maximum ocean heat transport remains similar to HadGEM2-AO, being below 1PW and therefore low compared with observational estimates.

Turning to precipitation, the general structure of the mean biases remains similar to HadGEM2-AO, with a southward displaced Inter-Tropical Convergence Zone (ITCZ) over the Atlantic and Indian oceans (Fig. 8). This is consistent with the asymmetry in the SST bias (Kang et al., 2008), and results in lack of precipitation in the summer monsoon systems over west Africa and India. However, there are other processes contributing to reduced precipitation over west Africa and India since a dry bias exists here in AMIP simulations whereas the southward displacement of the ITCZ over the oceans does not. Unlike a number of CMIP5 models (Taylor et al., 2012), there is no pronounced split ITCZ in either HadGEM2-AO or HadGEM3-GC2, although a significant wet bias exists on the north side of the warm pool, Pacific ITCZ and South Pacific Convergence Zone (SPCZ). Whilst the geographical pattern is similar, mid-latitude precipitation biases are slightly reduced in HadGEM3-GC2, particularly the dry bias over the northern North Atlantic. The RMS error is slightly increased (from 1.68 to 1.76), primarily due to an increased mean bias over the tropical west Pacific and east Indian Ocean.

The accurate simulation of clouds is a particular strength of HadGEM2-AO (e.g. Klein et al., 2013; Jiang et al., 2012; Nam et al., 2012). GC2 includes a new prognostic cloud scheme (PC2, Wilson et al., 2008) which gives similar or even slightly improved cloud properties (amount, height and albedo) for optically thicker clouds, including a good simulation of cloud in marine stratocumulus regions which are of particular importance for climate sensitivity (e.g. Bony et al., 2006) (not shown). The main difference in the cloud simulation between the two models is for optically thin cirrus. The Smith (1990) scheme used in HadGEM2-AO had an implicit coupling between cloud fraction and optical depth, preventing high fractional coverage of very thin cirrus. This coupling does not exist in PC2 and consequently HadGEM3-GC2 has almost double the amount of cirrus, much of which is sub-visual. CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) is a space-borne cloud lidar and is particularly suited to detecting optically thin "sub-

visual" cirrus, in addition to thicker cirrus which can be detected with passive instruments (Winker et al., 2010; Chepfer et al., 2010). A comparison of the models with CALIPSO using the CFMIP (Cloud Feedback Model Intercomparison Project) Observational Simulator Package (COSP) (Bodas-Salcedo et al., 2011; Chepfer et al., 2008) suggest that the true amount of cirrus should be between what is simulated by the two models (Fig. 9). HadGEM2-AO also has a bias for the tropical cirrus to be too low in altitude. The altitude of the peak amount of cirrus has been improved slightly in HadGEM3-GC2, but still remains lower than observed.

4 Mid-latitude variability

One of the most recent changes in the development of GC2 was the inclusion of a significant revision to the atmosphere dynamical core – ENDGame (Even Newer Dynamics for Global Atmospheric Modelling of the Environment, Wood et al., 2014). Walters et al. (2015) discuss how one of the main aims of this change was to improve the accuracy of the semi-Lagrangian dynamics, so less implicit smoothing is used, with the effect of increasing synoptic variability. A weakening of synoptic variability, measured through the intensity of tracked extra-tropical cyclones, has been shown to be a general problem in many NWP models (Froude, 2010) and climate models in CMIP5 (Zappa et al., 2013). Walters et al. (2015) illustrate the improvement in cyclone intensities in GA6 on NWP timescales, and this carries through to the GC2 CLIM experiment (Fig. 10). In HadGEM2-AO there is a general negative bias in cyclone intensity (as measured by 850 hPa relative vorticity) in the storm tracks in both hemispheres, especially on the equatorward side of the storm tracks. This bias is largely eliminated in HadGEM3-GC2, although there remains a slight deficit in intensity at the end of the Atlantic storm track and the eastern hemisphere of the Southern Ocean storm track.

The winter jet over the North Atlantic follows a tri-modal structure (Fig. 11; Woollings et al., 2010). These jet positions have been shown to have some correspondence to primary blocking locations with the southernmost jet position being associated with

Greenland blocking, the central position with no blocking and the northernmost position being correlated with European blocking (although some studies suggest European blocking can exist as a separate regime) (Cassou et al., 2004; Woollings et al., 2010; Davini et al., 2014). GC2 reproduces the tri-modal structure well although there is a slight tendency for the jet to occupy the northernmost position too frequently (Fig. 11). At other times of the year, the jet is more uni-modal which GC2 captures, although again there tends to be a slight northward displacement (not shown). These results are robust in that a very similar structure is seen in different atmosphere resolutions, including when the model is run at the lower atmospheric resolution of N96.

The tendency to have a more favoured northward position of the jet is in contrast to most CMIP3 and CMIP5 models which either do not capture the tri-modal structure at all, or tend to have the southern jet location simulated too frequently (Hannachi et al., 2013; Anstey et al., 2012). Davini et al. (2014) indicate that a North Atlantic cold bias, resulting from a weak and displaced North Atlantic Drift, can result in the simulated jet favouring the southern position. The improved SSTs in HadGEM3-GC2 relative to other models may account for the more frequent simulation of the northern location. Despite the good distribution of jet latitudes, there is still around a 25 % deficit in European blocking in HadGEM3-GC2 (defined using a 2-dimensional "wave-breaking" index based on Tibaldi and Molteni, 1990; not shown), which is again insensitive to horizontal resolution and is the subject of ongoing research.

The North Atlantic Oscillation (NAO) is a leading mode of variability affecting Europe. Seasonal forecasts of the winter NAO are now starting to show skill (Scaife et al., 2014), hence an accurate simulation of the NAO in a proposed replacement model is important. Table 1 illustrates a further improvement in the pattern correlation and slight improvement in variability of the winter NAO in GC2 relative to the model currently used for seasonal forecasts. Some studies have suggested a link between the Quasi-Biennial Oscillation (QBO) and the winter NAO with a more positive NAO during westerly QBO events (Holton and Tan, 1980; Anstey and Shepherd, 2014). Figure 12 shows a composite of westerly minus easterly QBO events for the Arctic stratospheric vortex and pressure at mean sea-level (PMSL)

in the GC2 coupled control simulation. It can be seen that this relationship does exist in GC2, albeit somewhat weaker than observed and with some displacement of the PMSL pattern in the Euro-Atlantic sector. The simulation of this teleconnection is noteworthy given the potential implications for seasonal predictability, and is a relationship which is not present in all CMIP5 models.

5 Tropical variability

A significant improvement in GA6 relative to earlier configurations is in the simulation of tropical cyclones (Walters et al., 2015). Both the change to the ENDGame dynamical core and convection parametrization changes have contributed, resulting in more intense tropical cyclones and improved tracks. At N216 resolution (around 100 km in the tropics), GC2 is now able to simulate tropical cyclone central pressures which would be expected from category 3 storms (Fig. 13). However, the 10 m wind speed associated with these systems remains below category 1, suggesting that the storms are too large. It might be expected that increased horizontal resolution would improve the pressure–wind-speed relationship, however even increasing the resolution to N1024 (around 20 km in the tropics) in an atmosphere-only simulation only partially improves the bias (Fig. 13) (Roberts et al., 2015).

Seasonal forecasts of landfalling Atlantic hurricanes are an emerging product which relies on the accurate tropical cyclone track distributions in the basin. Figure 14 shows the track densities from the SEAS experiment for the existing seasonal forecast system and GC2. In the Atlantic basin, GloSea5-GC2 has more storms than its predecessor (GloSea5-GA3) and has a better distribution of tracks with more early recurvature into the central North Atlantic, more reaching the Caribbean and a broad peak making landfall on the US coast. The west Pacific has slightly too many storms and there is a lack of a clear break in storms in the central Pacific.

There have been a large number of changes to the convection parametrization between HadGEM2-AO and HadGEM3-GC2. Out of these, increases to the entrainment

and detrainment rates have been primarily responsible for an improved simulation of the Madden–Julian Oscillation (MJO), although the amplitude of the systems remains significantly weaker than observed. Model performance is measured using simplified metrics proposed by the the international MJO Task Force (which is under the WMO Working Group on Numerical Experimentation, WGNE) (Kim et al., 2009). One simple measure of the MJO is based on the space–time power spectrum of equatorial rainfall. The ratio of eastward to westward power (E/W ratio) at MJO time and space scales (zonal wavenumbers 1–3 and periods of 30–60 days) reveals the prominence of the eastward propagating intraseasonal variability relative to its westward counterpart and is a useful indicator of how prominent the MJO is relative to the background variability (Kim et al., 2014). On this metric, HadGEM3-GC2 has improved relative to HadGEM2-AO, although it remains below the observational range (Table 2). Another measure of MJO fidelity is R_{max}, proposed by Sperber and Kim (2012) which is the maximum correlation between the two time series obtained by projecting model outgoing longwave radiation (OLR) anomalies onto the leading pair of empirical orthogonal functions (EOFs) of observed OLR that capture the MJO. The MJO is deemed well simulated if the correlation between the two leading principal components (PCs) is strong at a lead time of about 10–15 days, thereby demonstrating coherent eastward propagation with appropriate spatiotemporal structure. Again, HadGEM3-GC2 performs better than HadGEM2-AO (Table 2).

Previous studies have suggested that atmosphere–ocean coupling is important for the propagation of the MJO (Klingaman and Woolnough, 2014). Relatively clear skies ahead of the MJO result in higher SSTs which encourage propagation of the MJO which in turn cools the SSTs due to the high cloud amounts and precipitation as the system passes. This is seen in the NWP experiment using a configuration similar to GC2 (Shelly et al., 2014). The coupled model maintains the observed lag of about 5 days of the outgoing longwave radiation (OLR) anomaly behind the maximum SSTs in the coupled simulation, whereas the convection moves over the maximum SSTs within the first few days in parallel atmosphere-only simulations and then remains static since the SSTs are not being updated with the cooling effect from the cloud. This is one example of why it is desirable to

move to coupled weather forecast models for even relatively short range predictions. Similar coupled model feedbacks might be expected to impact forecast tropical cyclone intensities (Schade and Emanuel, 1999).

A reliable simulation of the El-Niño Southern Oscillation (ENSO) is important for seasonal prediction and climate projections alike since it forms a leading mode of global variability and the major source of seasonal predictability. For many years, Met Office climate models have suffered from excess equatorial easterly wind stress. Improvement in this was a focus of HadGEM2-AO development and it has been further improved in GC2 with a contribution from a number of the science changes, most notably a change to the gravity wave drag scheme which reduces the coupling between the low level flow-blocking drag and gravity wave drag following (Vosper et al., 2009) (Fig. 15b). As a result of the improved windstress, improved MJO (which can be the source of westerly wind bursts; e.g. Lengaigne et al., 2004) and higher horizontal resolution of the ocean, ENSO is well simulated in HadGEM3-GC2 with a good spatial pattern (Fig. 15c&d). When assessed against a range of metrics (Table 3) we see that variability in SST agrees well with observations in the central-east Pacific although is somewhat weaker than observed near the dateline. A power spectrum analysis shows the frequency lies within the observed range (3 to 7 years), with no dominant short (e.g. 2 year) or longer period peaks. The model seasonality is good, with maximum (minimum) variability in boreal winter (spring). The standard deviation of precipitation in the central Pacific gives a measure of model capability for regional climate impacts and although slightly underestimated, is good in comparison with other climate models which tend to underestimate this quantity. Overall, HadGEM3-CG2 compares favourably with a range of CMIP5 models (Bellenger et al., 2014). The main observed ENSO teleconnections to remote precipitation anomalies (S. America, Sahel, India, E. Africa, etc.) are also present in the model (not shown).

Africa is a region where accurate predictions and projections of rainfall are particularly important for those living there and model simulations have generally been poor (Flato et al., 2013). Teleconnections from remote SST anomalies are primarily responsible for the large interannual variability of seasonal mean rainfall over many areas of the continent. Rowell

(2013) investigated the ability of CMIP models to accurately represent teleconnections from remote SST anomalies to African rainfall. Figure 16 is a reproduction of Fig. 10 from Rowell (2013) but with HadGEM3-GC2 added. It shows HadGEM3-GC2 has the joint highest proportion of the teleconnections accurately simulated compared with CMIP3 and CMIP5 models previously analysed, supporting the use of GC2 for seasonal predictions and climate change projections over the region.

6 Summary

In this paper we have presented the performance of the GC2 configuration of the Met Office Unified Model in terms of its systematic errors. The focus has been on seasonal and climate timescales since these are the timescales on which the model is to be used operationally for predictions/projections, and GC2 has been compared with models currently used in those systems. The results presented here should be considered alongside the atmosphere-only results in Walters et al. (2015) and ocean/sea-ice results in Megann et al. (2014) and Rae et al. (2015).

Overall GC2 provides a significant improvement in mean bias and variability over the coupled configurations currently used, with temperature biases in most regions, simulation of atmospheric regimes over the North Atlantic, simulation of tropical cyclones and ENSO being particularly notable. However, there are a number of systematic errors requiring further work, the highest priority being the Southern Ocean warm SST bias and low levels of rainfall over India and west Africa during the summer monsoons. Consequently, caution is required when considering predictions/projections from GC2 in these regions.

Climate change simulations using HadGEM3-GC2 are already in progress and will be reported by Senior et al. (2015), whilst GloSea5-GC2 is being used operationally for Met Office seasonal forecasts since 3rd February 2015.

Consistent with the annual development discussed in the Introduction, work is already underway to further develop GC2 to form GC3. It is envisaged that GC3 will subsequently have earth system components built on top of it to form UKESM1 (United Kingdom Earth

System Model version 1), which will be the UK's submission to CMIP6. Hence, the assessment presented here provides an initial picture of the model performance which will help inform the next stages of UKESM1 development.

Code availability

The MetUM is available for use under licence. A number of research organisations and national meteorological services use the MetUM in collaboration with the Met Office to undertake basic atmospheric process research, produce forecasts, develop the MetUM code and build and evaluate Earth system models. For further information on how to apply for a licence see http://www.metoffice.gov.uk/research/collaboration/um-collaboration. Version 8.5 of the source code is used in this paper.

JULES is available under licence free of charge. For further information on how to gain permission to use JULES for research purposes see https://jules.jchmr.org/software-and-documentation.

The model code for NEMO v3.4 is available from the NEMO website (www.nemo-ocean.eu). On registering, individuals can access the code using the open source subversion software (http://subversion.apache.org/). The revision number of the base NEMO code used for this paper is 3424.

The model code for CICE is freely available from the United States Los Alamos National Laboratory (http://oceans11.lanl.gov/trac/CICE/wiki/SourceCode), again using subversion. The revision number for the version used for this paper is 430.

A number of branches are applied to the above codes. Please contact the authors for more information on these branches and how to obtain them.

19

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., and Arkin, P.: The Version 2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), J. Hydrometeorol., 4, 1147–1167, 2003.

Allan, R. P., Liu, C., Loeb, N. G., Palmer, M. D., Roberts, M., Smith, D., and Vidale, P.-L.: Changes in global net radiative imbalance 1985–2012, Geophys. Res. Lett., 41, 5588–5597, doi:10.1002/2014GL060962, 2014.

Anstey, J. A. and Shepherd, T. G.: High-latitude influence of the quasi-biennial oscillation, Q. J. Roy. Meteor. Soc., 140, 1–21, 2014.

Anstey, J. A., Davini, P., Gray, L. J., Woollings, T. J., Butchart, N., Cagnazzo, C., Christiansen, B., Hardiman, S. C., Osprey, S. M., and Yang, S. T.: Multi-model analysis of Northern Hemisphere winter blocking: model biases and the role of resolution, J. Geophys. Res., 118, 3956–3971, doi:10.1002/jgrd.50231, 2012.

Bellenger, H., Guilyardi, E., Leloup, E., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: From CMIP3 to CMIP5, Clim. Dynam., 42, 1999–2018, doi:10.1007/s00382-013-1783-z, 2014.

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

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S., Zhang, Y., Marchand, R., Haynes, J. M., Pincus, R., and John, V. O.: COSP: satellite simulation software for model assessment, B. Am. Meteorol. Soc., 92, 1023–1043, doi:10.1175/2011BAMS2856.1, 2011.

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

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J.-L., Koshiro, T., Stevens, B., Wang, Z., and Yokohata, T.: Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models, J. Climate, 27, 41–56, doi:10.1175/JCLI-D-13-00169.1, 2014.

Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J. L., Hall, A., Hallegate, S., Holland, M. M., Ingram, W. J., Randall, D. A., Soden, B. J., Tselioudis, G., and Webb, M. J.: How well do we understand and evaluate climate change feedback processes?, J. Climate, 19, 3445–3482, doi:10.1175/JCLI3819.1, 2006.

Brewer, A. M.: Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere, Q. J. Roy. Meteor. Soc., 75, 351–363, 1949.

Brohan, P., Kennedy, J., Harris, I., Tett, S. F. B., and Jones, P. D.: Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850, J. Geophys. Res., 111, D12106, doi:10.1029/2005JD006548, 2006.

Brown, A. R., Milton, S., Cullen, M., Golding, B., Mitchell, J., and Shelly, A.: Unified modelling and prediction of weather and climate: a 25 year journey, B. Am. Meteorol. Soc., 93, 1865–1877, doi:10.1175/BAMS-D-12-00018.1, 2012.

Cassou, C., Terray, L., Hurrel, J., and Deser, C.: North Atlantic winter climate regimes: spatial asymmetry, stationarity with time, and ocean forcing, J. Climate, 17, 1055–1067, 2004.

Chepfer, H., Bony, S., Winker, D., Chiriaco, M., Dufresne, J.-L., and Sèze, G.: Use of CALIPSO lidar observations to evaluate the cloudiness simulated by a climate model., Geophys. Res. Lett., 35, L15704, doi:10.1029/2008GL034207, 2008.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J.-L., Minnis, P., Stubenrauch, C. J., and Zeng, S.: The GCM Oriented Calipso Cloud Product (CALIPSO-GOCCP), J. Geophys. Res., 115, D00H16, doi:10.1029/2009JD012251, 2010.

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

Cullen, M. J. P.: The unified forecast/climate model, Meteorol. Mag., 122, 81–94, 1993.

Davini, P., Cagnazzo, C., Fogli, P. G., Manzini, E., Gualdi, S., and Navarra, A.: European blocking and Atlantic jet stream variability in the NCEP/NCAR reanalysis and the CMCC-CMS climate model, Clim. Dynam., 43, 71–85, doi:10.1007/s00382-013-1873-y, 2014.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hölm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, P. K., Peubey, C., de Rosnay, P., Tavolato, C., Thê-

paut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of climate models, in: IPCC WGI Fifth Assessment Report, chap. 9, Cambridge University Press, Cambridge, UK, New York, NY, USA, 2013.

Froude, L. S. R.: TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems, Weather Forecast., 25, 819–836, 2010.

Hall, A. and Qu, X.: Using the current seasonal cycle to constrain snow albedo feedback in future climate change, Geophys. Res. Lett., 33, L03502, doi:10.1029/2005GL025127, 2006.

Hannachi, A., Barnes, E. A., and Woollings, T.: Behaviour of the winter North Atlantic eddy-driven jet stream in the CMIP3 integrations, Clim. Dynam., 41, 995–1007, doi:10.1007/s00382-012-1560-4, 2013.

Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S. R., Keen, A. B., McLaren, A. J., and Hunke, E. C.: Design and implementation of the infrastructure of HadGEM3: the next-generation Met Office climate modelling system, Geosci. Model Dev., 4, 223–253, doi:10.5194/gmd-4-223-2011, 2011.

Hodges, K.: Feature tracking on a unit sphere, Mon. Weather Rev., 123, 3458–3465, 1995.

Holton, J. R. and Tan, H. C.: The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb, J. Atmos. Sci., 37, 2200–2208, 1980.

Hunke, E. C. and Lipscomb, W. H.: CICE: The Los Alamos sea ice model, documentation and software. Version 3.1, Tech. Rep. LA-CC-98-16, Los Alamos National Laboratory, Los Alamos, NM, 2004.

Ingleby, B. and Huddleston, M.: Quality control of ocean temperature and salinity profiles – historical and real-time data, J. Marine Syst., 65, 158–175, doi:10.1016/j.jmarsys.2005.11.019, 2007.

Jiang, J. H., Su, H., Zhai, C., Perun, V. S., Genio, A. D., Nazarenko, L. S., Donner, L. J., Horowitz, L., Seman, C., Cole, J., Gettelman, A., Ringer, M. A., Rotstayn, L., Jeffrey, S., Wu, T., Brient, F., Dufresne, J.-L., Kawai, H., Koshiro, T., Watanabe, M., Lécuyer, T. S., Volodin, E. M., Iversen, T., Drange, H., Mesquita, M. D. S., Read, W. G., Waters, J. W., Tian, B., Teixeira, J., and Stephens, G. L.: Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA A-Train satellite observations, J. Geophys. Res., 117, D14105, doi:10.1029/2011JD017237, 2012.

Jones, P. W.: First and second-order conservative remapping schemes for grids in spherical coordinates, Mon. Weather Rev., 127, 2204–2210, 1999.

Josey, S. A., Kent, E. C., and Taylor, P. K.: The Southampton Oceanography Centre (SOC) ocean – atmosphere heat, momentum and freshwater flux atlas, Report 6, Southampton Oceanography Centre, European Way, Southamton, SO14 3ZH, UK, 55 pp., 1998.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–471, 1996.

Kang, S. M., Held, I. M., Frierson, D. M. W., and Zhao, M.: The response of the ITCZ to extratropical thermal forcing: idealized slab-ocean experiments with a GCM, J. Climate, 21, 3521–3532, doi:10.1175/2007JCLI2146.1, 2008.

Kim, D., Sperber, K., Stern, W., Waliser, D., Kang, I.-S., Maloney, E., Wang, W., Weickmann, K., Benedict, J., and Khairoutdinov, M.: Application of MJO simulation diagnostics to climate models, J. Climate, 22, 6413–6436, 2009.

Kim, D., Xavier, P., Maloney, E., Wheeler, M., Waliser, D., Sperber, K., Hendon, H., Zhang, C., Neale, R., and Hwang, Y.-T.: Process-oriented MJO simulation diagnostic: moisture sensitivity of simulated convection, J. Climate, 27, 5379–5395, 2014.

Klein, S. A., Jiang, X., Boyle, J., Malyshev, S., and Xie, S.: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach, Geophys. Res. Lett., 33, L18805, doi:10.1029/2006GL027567, 2006.

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., and Gleckler, P. J.: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, J. Geophys. Res., 118, 1329–1342, doi:10.1002/jgrd.50141, 2013.

Klingaman, N. and Woolnough, S.: The role of air–sea coupling in the simulation of the Madden–Julian Oscillation in the Hadley Centre model, Q. J. Roy. Meteor. Soc., 140, 2272–2286, doi:10.1002/qj.2295, 2014.

Landsea, C. W. and Franklin, J. L.: Atlantic hurricane database uncertainty and presentation of a new database format, Mon. Weather Rev., 141, 3576–3592, doi:10.1175/MWR-D-12-00254.1, 2013.

Lengaigne, M., Guilyardi, E., Boulanger, J.-P., Menkes, C., Delecluse, P., Inness, P., Cole, J., and Slingo, J.: Triggering of El Nino by westerly wind events in a coupled general circulation model, Clim. Dynam., 23, 601–620, 2004.

Liu., C., Allan, R.P., Berrisford, P., Hyder, P., Loeb, N., Smith, D., Vidale, P.-L., and Edwards, J. M.: Combining satellite observations and reanalysis energy transports to estimate global net surface energy fluxes 1985–2012, in preparation, 2015.

Ma, H.-Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., Douville, H., Fermepin, S., Medeiros, B., Tyteca, S., Watanabe, M., and Williamson, D.: On the correspondence between mean forecast errors and climate errors in CMIP5 models, J. Climate, 17, 1781–1798, doi:10.1175/JCLI-D-13-00474.1, 2014.

MacLachlan, C., Arribas, A., Peterson, D., Maidens, A., Fereday, D., Scaife, A. A., Gordon, M., Vellinga, M., Williams, A., Comer, R. E., Camp, J., Xavier, P., and Madec, G.: Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system, Q. J. Roy. Meteor. Soc., doi:10.1002/qj.2396, in press, 2015.

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

Martin, G. M., Milton, S. F., Senior, C. A., Brooks, M. E., Ineson, S., Reichler, T., and Kim, J.: Analysis and reduction of systematic errors through a seamless approach to modelling weather and climate, J. Climate, 23, 5933–5957, doi:10.1175/2010JCLI3541.1, 2010.

Megann, A., Storkey, D., Aksenov, Y., Alderson, S., Calvert, D., Graham, T., Hyder, P., Siddorn, J., and Sinha, B.: GO5.0: the joint NERC–Met Office NEMO global ocean model for use in coupled and forced applications, Geosci. Model Dev., 7, 1069–1092, doi:10.5194/gmd-7-1069-2014, 2014.

Nam, C., Bony, S., Dufresne, J.-L., and Chepfer, H.: The "too few, too bright" tropical low-cloud problem in CMIP5 models, Geophys. Res. Lett., 39, L21801, doi:10.1029/2012GL053421, 2012.

Rae, J. G. L., Hewitt, H. T., Keen, A. J., Ridley, J. K., West, A. E., Harris, C. M., Hunke, E. C., and Walters, D. N.: Development of Global Sea Ice 5.0 and 6.0 CICE configurations for the Met Office Global Coupled Model, Geosci. Model Dev. Discuss., 8, 2529–2554, doi:10.5194/gmdd-8-2529-2015, 2015.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, doi:10.1029/2002JD002670, 2003.

Roberts, M. J., Vidale, P. L., Mizielinski, M., Demory, M.-E., Schiemann, R., Strachan, J., Hodges, K., Camp, J., and Bell, R.: Tropical cyclones in the UPSCALE ensemble of high resolution global climate models, J. Climate, 28, 574-596 doi:10.1175/JCLI-D-14-00131.1, 2015.

Rowell, D. P.: Simulating SST teleconnections to Africa: what is the state of the art?, J. Climate, 26, 5397–5417, doi:10.1175/JCLI-D-12-00761.1, 2013.

Scaife, A. A., Copsey, D., Gordon, C., Harris, C., Hinton, T., Keeley, S., O'Neill, A., Roberts, M., and Williams, K.: Improved Atlantic winter blocking in a climate model, Geophys. Res. Lett., 38, L23703, doi:10.1029/2011GL049573, 2011.

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., Eade, R., Fereday, D., Folland, C. K., Gordon, M., Hermanson, L., Knight, J. R., Lea, D. J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A. K., Smith, D., Vellinga, M., Wallace, E., Waters, J., and Williams, A.: Skillful long-range prediction of European and North American winters, Geophys. Res. Lett., 41, 2514–2519, doi:10.1002/2014GL059637, 2014.

Schade, L. R. and Emanuel, K. A.: The ocean's effect on the intensity of tropical cyclones: results from a simple coupled atmosphere–ocean model, J. Atmos. Sci., 56, 642–651, 1999.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R.: Uncertainty in modeled Arctic sea ice volume, J. Geophys. Res., 116, C00D06, doi:10.1029/2011JC007084, 2011.

Senior, C. A., Andrews, T., Copsey, D., Graham, T., Harris, C., Ineson, S., Ridley, J., Ringer, M., and Rowell, D.: Idealised climate change simulations with a high resolution physical model, HadGEM3-GC2, in preparation, 2015.

Shelly, A., Xavier, P., Copsey, D., Johns, T., Rodriguez, J. M., Milton, S., and Klingaman, N.: Coupled versus uncoupled hindcast simulations of the Madden–Julian Oscillation in the Year of Tropical Convection, Geophys. Res. Lett., 41, 5670–5677, doi:10.1002/2013GL059062, 2014.

Simpson, J., Adler, R., and North, G.: A proposed Tropical Rainfall Measuring Mission (TRMM) satellite, B. Am. Meteorol. Soc., 69, 278–295, 1988.

Smeed, D. A., McCarthy, G. D., Cunningham, S. A., Frajka-Williams, E., Rayner, D., Johns, W. E., Meinen, C. S., Baringer, M. O., Moat, B. I., Duchez, A., and Bryden, H. L.: Observed decline of the Atlantic meridional overturning circulation 2004–2012, Ocean Sci., 10, 29–38, doi:10.5194/os-10-29-2014, 2014.

Smith, R. N. B.: A scheme for predicting layer clouds and their water content in a general circulation model, Q. J. Roy. Meteor. Soc., 116, 435–460, 1990.

Sperber, K. R. and Kim, D.: Simplified metrics for the identification of the Madden–Julian Oscillation in models, Atmos. Sci. Lett., 13, 187–193, 2012.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

The HadGEM2 Development Team: G. M. Martin, Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., Jones, C. D., McDonald, R. E., McLaren, A. J., O'Connor, F. M., Roberts, M. J., Rodriguez, J. M., Woodward, S., Best, M. J., Brooks, M. E.,

Brown, A. R., Butchart, N., Dearden, C., Derbyshire, S. H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J. M., Falloon, P. D., Gedney, N., Gray, L. J., Hewitt, H. T., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W. J., James, P. M., Johns, T. C., Johnson, C. E., Jones, A., Jones, C. P., Joshi, M. M., Keen, A. B., Liddicoat, S., Lock, A. P., Maidens, A. V., Manners, J. C., Milton, S. F., Rae, J. G. L., Ridley, J. K., Sellar, A., Senior, C. A., Totterdell, I. J., Verhoef, A., Vidale, P. L., and Wiltshire, A.: The HadGEM2 family of Met Office Unified Model climate configurations, Geosci. Model Dev., 4, 723–757, doi:10.5194/gmd-4-723-2011, 2011.

Tibaldi, S. and Molteni, F.: On the operational predictability of blocking, Tellus A, 42, 343–365, 1990.

Trenberth, K. E. and Fasullo, J. T.: Simulation of present-day and Twenty-First-Century energy budgets of the Southern Oceans, J. Climate, 23, 440–454, doi:10.1175/2009JCLI3152.1, 2010.

Valcke, S.: The OASIS3 coupler: a European climate modelling community software, Geosci. Model Dev., 6, 373–388, doi:10.5194/gmd-6-373-2013, 2013.

Vosper, S. B., Wells, H., and Brown, A. R.: Accounting for non-uniform static stability in orographic drag parametrization, Q. J. Roy. Meteor. Soc., 135, 815–822, doi:10.1002/qj.407, 2009.

Walters, D. N., Best, M. J., Bushell, A. C., Copsey, D., Edwards, J. M., Falloon, P. D., Harris, C. M., Lock, A. P., Manners, J. C., Morcrette, C. J., Roberts, M. J., Stratton, R. A., Webster, S., Wilkinson, J. M., Willett, M. R., Boutle, I. A., Earnshaw, P. D., Hill, P. G., MacLachlan, C., Martin, G. M., Moufouma-Okia, W., Palmer, M. D., Petch, J. C., Rooney, G. G., Scaife, A. A., and Williams, K. D.: The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations, Geosci. Model Dev., 4, 919–941, doi:10.5194/gmd-4-919-2011, 2011.

Walters, D. N., Brooks, M. E., Boutle, I. A., Melvin, T. R. O., Stratton, R. A., Bushell, A. C., Copsey, D., Earnshaw, P. E., Gross, M. S., Hardiman, S. C., Harris, C. M., Heming, J. T., Klingaman, N. P., Levine, R. C., Manners, J., Martin, G. M., Milton, S. F., Mittermaier, M. P., Morcrette, C. J., Riddick, T. C., Roberts, M. J., Selwood, P. M., Tennant, W.J., Vidale, P.-L., Wilkinson, J. M., Wood, N., Woolnough, S. J., and Xavier, P. K.: The Met Office Unified Model Global Atmosphere 6.0 and JULES Global Land 6.0 configurations, in preparation, 2015.

Williams, K. D., Bodas-Salcedo, A., Deque, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S. A., Senior, C. A., and Williamson, D. L.: The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models, J. Climate, 26, 3258–3274, doi:10.1175/JCLI-D-12-00429.1, 2013.

Wilson, D. R., Bushell, A. C., Kerr-Munslow, A. M., Price, J. D., and Morcrette, C. J.: PC2: a prognostic cloud fraction and condensation scheme. I: Scheme description, Q. J. Roy. Meteor. Soc., 134, 2093–2107, doi:10.1002/qj.333, 2008.

Winker, D. M., Pelon, J., Coakley Jr, J. A., Ackerman, S. A., Charlson, R. J., Colarco, P. R., Flamant, P., Fu, Q., Hoff, R. M., Kittaka, C., Kubar, T. L., Treut, H. L., McCormick, M. P., Mégie, G., Poole, L., Powell, K., Trepte, C., Vaughan, M. A., and Wielicki, B. A.: The CALIPSO mission: a global 3D view of aerosols and clouds, B. Am. Meteorol. Soc., 91, 1211–1229, doi:10.1175/2010BAMS3009.1, 2010.

Wood, N., Stainforth, A., White, A., Allen, T., Diamantakis, M., Gross, M., Melvin, T., Smith, C., Vosper, S., Zerroukat, M., and Thuburn, J.: An inherently mass-conserving semi-implicit semi-Lagrangian discretisation of the deep-atmosphere global nonhydrostatic equations, Q. J. Roy. Meteor. Soc., 140, 1505–1520, doi:10.1002/qj.2235, 2014.

Woollings, T., Hannachi, A., and Hoskins, B.: Variability of the North Atlantic eddy-driven jet stream, Q. J. Roy. Meteor. Soc., 136, 856–868, 2010.

Zappa, G., Shaffrey, L. C., and Hodges, K. I.: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones, J. Climate, 26, 5379–5396, 2013.

Table 1. Spatial correlation of the NAO pattern (the leading empirical orthogonal function of the winter mean sea-level pressure fields) and interannual variability in the CLIM experiment compared with NCEP (National Centers for Environmental Prediction) re-analyses (Kalnay et al., 1996). NCEP re-analyses are used here rather than ERA-I (ECMWF Interim Re-analyses; Dee et al., 2011) which are used throughout the rest of the paper since a longer record is needed to estimate the observed variability.

	GA3 coupled	GC2 coupled	NCEP re-analysis
Correlation	0.83	0.91	N/A
Variability	6.04 hPa	6.20 hPa	8.5 hPa

Table 2. E/W ratio in spectral power following Kim et al. (2014) and R_{max} following Sperber and Kim (2012) for the CLIM experiment. GPCP (Global Precipitation Climatology Project; Adler et al., 2003) and TRMM (Tropical Rainfall Measuring Mission; Simpson et al., 1988) data are used for the observational values.

	E/W ratio	R_{max}
Observations	2.09–2.73	0.67
HadGEM2-AO	1.17	0.57
HadGEM3-GC2	1.52	0.64

Table 3. Metrics for ENSO assessment. M1 and M2 are standard deviation of monthly SST anomaly for regions Niño3 (90° W–150° W, 5° N–5° S) and Niño4 (160° E–150° W, 5° N–5° S) (K), M3 is the ratio of power in the 3-7 year range relative to 0 to 10 years for monthly Niño3 SST anomaly (%), M4 is a seasonality metric defined as the ratio of November to January and March to May standard deviation of Niño3 SST anomaly (Bellenger et al., 2014), M5 is the standard deviation of precipitation anomaly for Niño4 (mm day^{-1}). CLIM is the final 50 years of experiment CLIM, CLIM2 is the final 100 years of a 150 year experiment equivalent to CLIM differing only in the initial conditions. SST observations are HadISST (1901–2000) and precipitation is GPCP (1979–2013).

	M1 ($\sigma_{\text{SST-N3}}$)	M2 ($\sigma_{\text{SST-N4}}$)	M3 (%$P_{\text{SST-3-7}}$)	M4 ($\sigma_{\text{SST-NDJ}}/\sigma_{\text{SST-MAM}}$)	M5 ($\sigma_{\text{PPTN-N4}}$)
CLIM	0.70	0.32	53	1.2	2.0
CLIM2	0.73	0.37	65	1.6	2.0
Observations	0.79	0.54	69	1.6	2.3
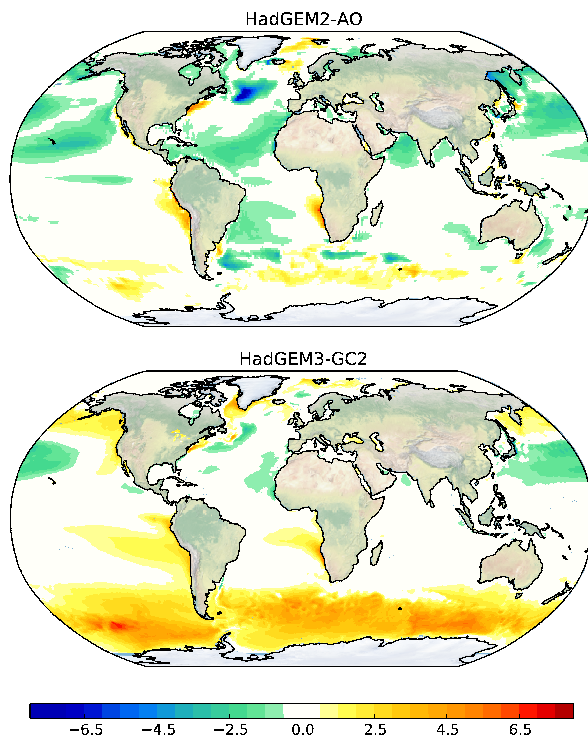
Figure 1. Mean SST bias (K) against EN3 for the CLIM experiment for HadGEM2-AO and HadGEM3-GC2.
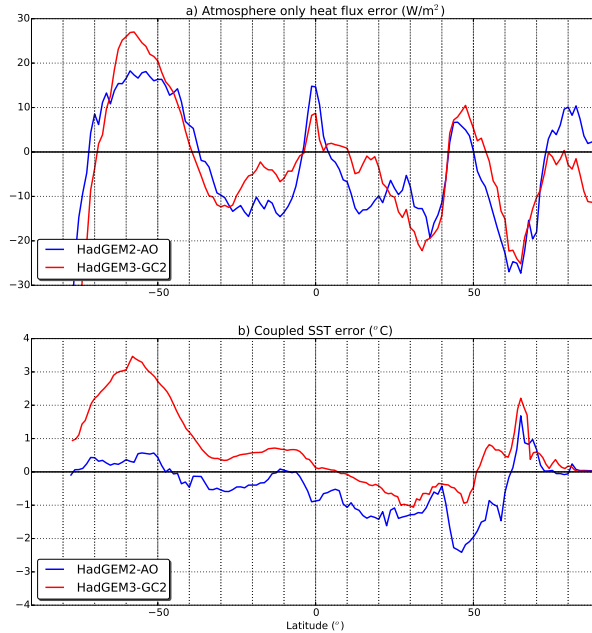
Figure 2. (a) Zonal mean net downward surface flux bias in AMIP simulations. The observed surface flux is a developmental version of the University of Reading surface flux product which combines satellite-based radiative fluxes (Allan et al., 2014) with re-analysis estimates of atmospheric column energy storage and horizontal divergences (Liu et al., 2015). **(b)** Zonal mean SST bias against EN3 for years 5–15 of the coupled model CLIM experiment (before any biases from the deep ocean influence the SSTs).
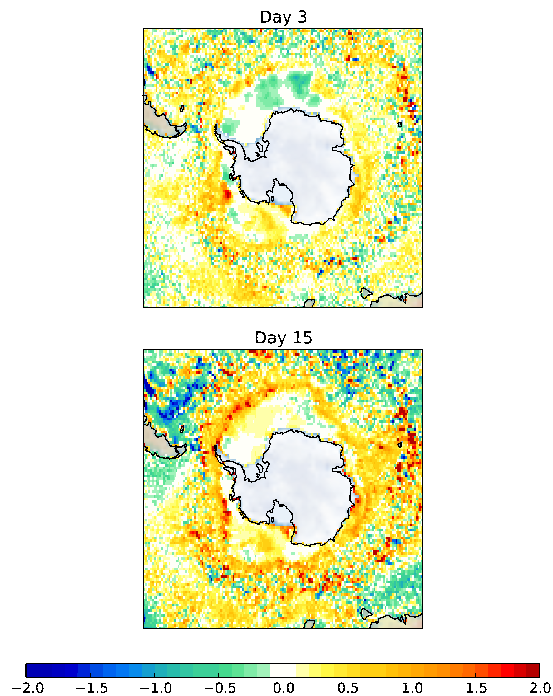
Figure 3. Mean day 3 and day 15 SST bias (K) against analyses in the NWP experiment.
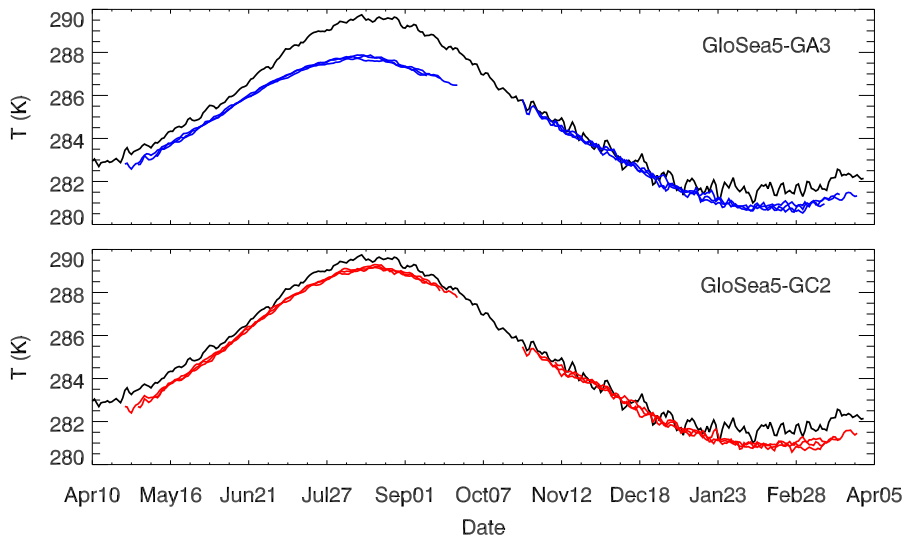
Figure 4. Mean 1.5 m temperature over the North Atlantic (10–50° W, 40–60° N) for the SEAS hind-casts (coloured). ERA-I is shown black.
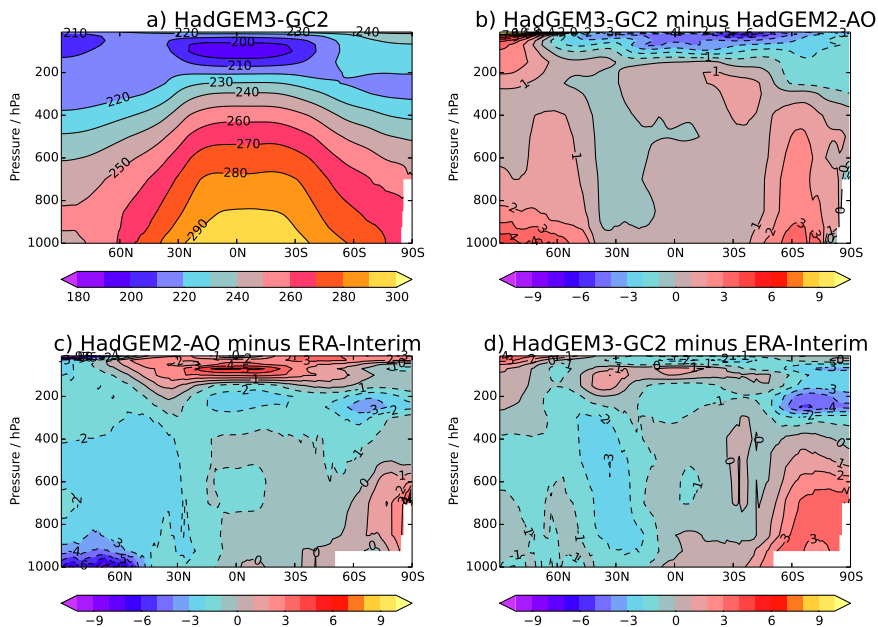
Figure 5. (a) DJF zonal mean temperature (K) from the HadGEM3-GC2 CLIM experiment, **(b)** HadGEM3-GC2 minus HadGEM2-AO, **(c)** HadGEM2-AO minus ERA-I, **(d)** HadGEM3-GC2 minus ERA-I.
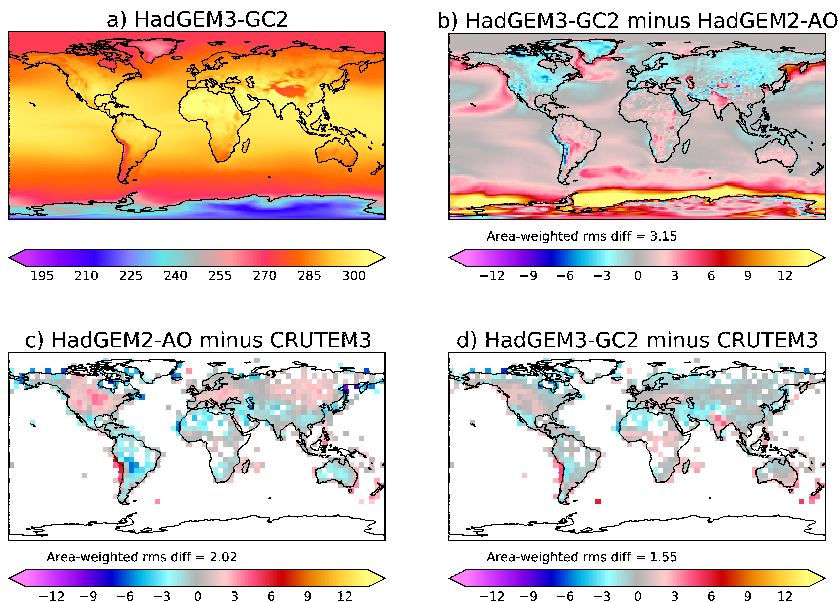
Figure 6. (a) JJA 1.5 m temperature (K) from the HadGEM3-GC2 CLIM experiment, **(b)** HadGEM3-GC2 minus HadGEM2-AO, **(c)** HadGEM2-AO minus CRUTEM3 observations (Brohan et al., 2006), **(d)** HadGEM3-GC2 minus CRUTEM3 observations.
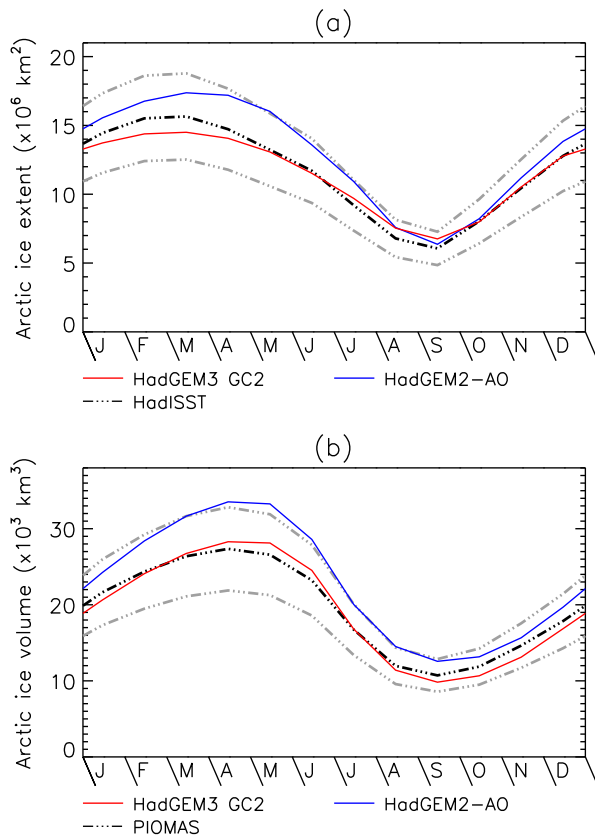
Figure 7. Annual cycle of Arctic sea ice **(a)** extent, **(b)** volume. Colours show HadGEM3-GC2 and HadGEM2-AO CLIM experiment and the black line is HadISST observations (Rayner et al., 2003) **(a)**, PIOMAS (Pan-Arctic Ice–Ocean Modeling and Assimilation System) analyses (Schweiger et al., 2011) **(b)**. Grey lines show 20 % intervals around the observations.
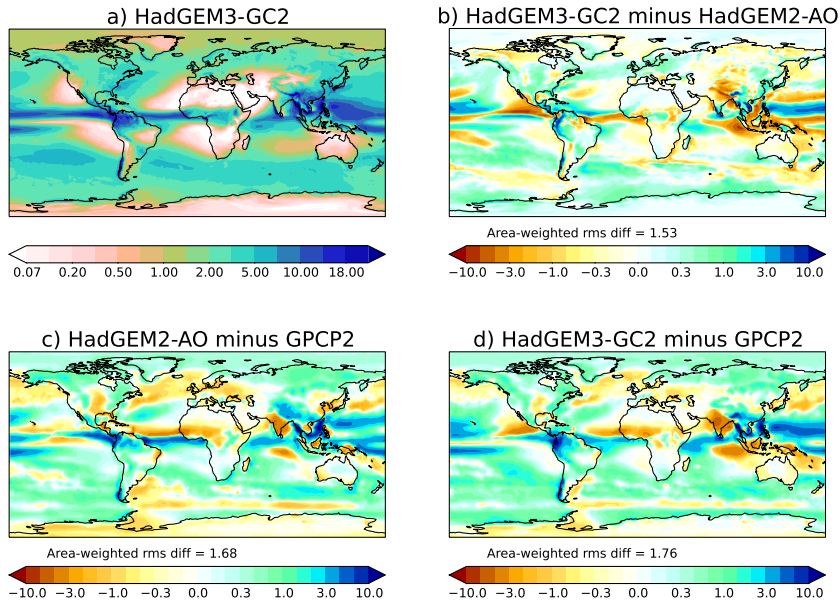
Figure 8. (a) JJA mean precipitation rate ($\mathrm{mm\ day}^{-1}$) from the HadGEM3-GC2 CLIM experiment, **(b)** HadGEM3-GC2 minus HadGEM2-AO, **(c)** HadGEM2-AO minus GPCP observations, **(d)** HadGEM3-GC2 minus GPCP observations.
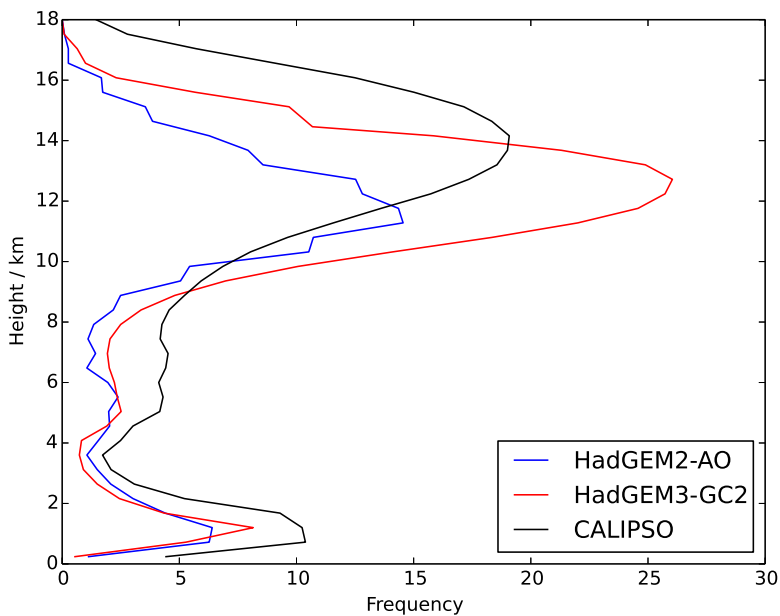
Figure 9. Tropical-mean ($20°$ N–$20°$ S) vertical profile of cloud frequency in atmosphere-only (AMIP) simulations of HadGEM2-AO and HadGEM3-GC2 using the CALIPSO simulator from COSP. The observed profile from CALIPSO is shown black.
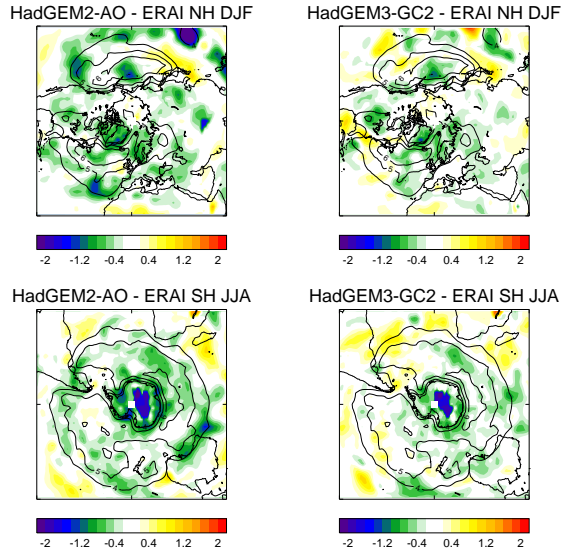
Figure 10. Bias in winter tracked cyclone intensity, as measured by 850 hPa relative vorticity, relative to ERA-I. Tracking is on the final 20 years of the respective model simulations and for the period 1988–2008 for ERA-I using TRACK (Hodges, 1995). Contours show ERA-I values of relative vorticity at $10^{-5}\,\mathrm{s}^{-1}$ and the colours indicate the bias on the same scale (contours at $10^{-5}\,\mathrm{s}^{-1}$ intervals) for the CLIM experiment.
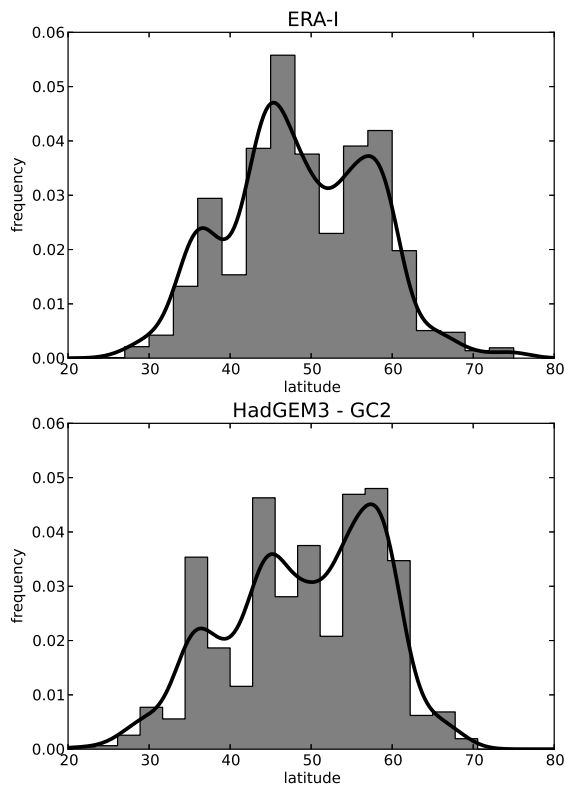
Figure 11. Normalised frequency distribution of DJF jet latitude defined as the maximum 850 hPa wind over the North Atlantic 60° W–0° E (following Woollings et al., 2010) for ERA-I 1979–2012 and the final 20 years of HadGEM3-GC2 CLIM experiment.
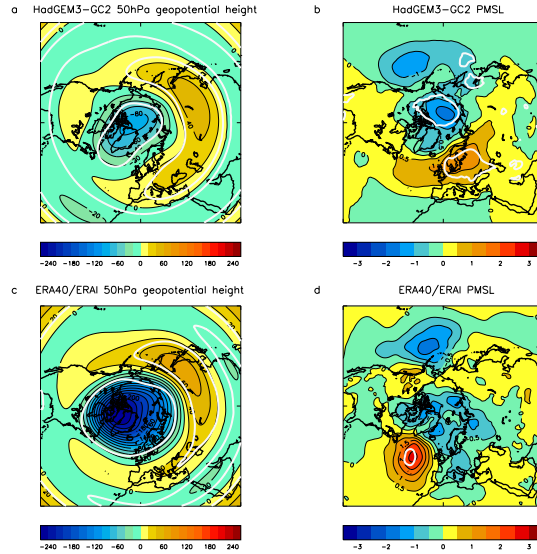
Figure 12. DJF composite mean 50 hPa geopotential height and PMSL for westerly minus easterly QBO events (mean 50 hPa equatorial wind, using a threshold of $\pm 5\,\mathrm{ms}^{-1}$) in HadGEM3-GC2 CLIM experiment **(a, b)** and a combined dataset of ERA40 (1957–1978) and ERA-I (1979–2013) **(c, d)**. White contours show significance at the 90 % level using a 2-sided t test.
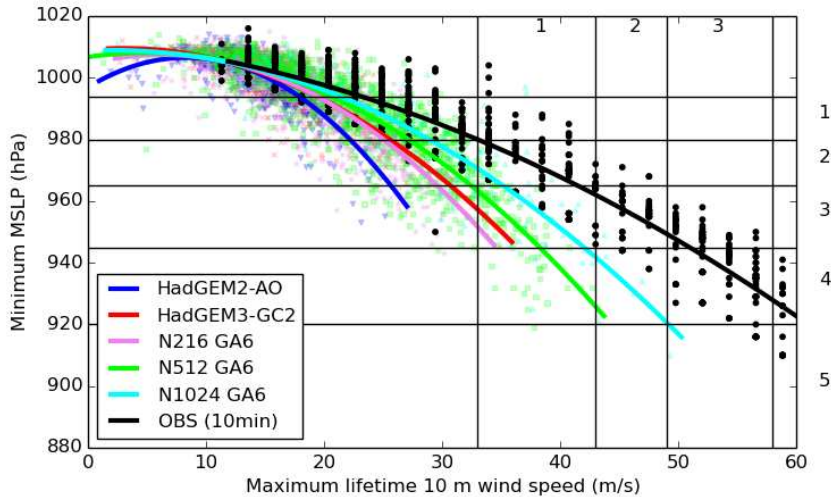
Figure 13. Scatter plot of 10 m wind speed vs. central PMSL for tropical cyclones in HadGEM2-AO and HadGEM3-GC2 CLIM simulations. Also shown are GA6 atmosphere-only (AMIP) simulations at N216, N512 and N1024. Hurricane Database (HURDAT, Landsea and Franklin, 2013) are black.
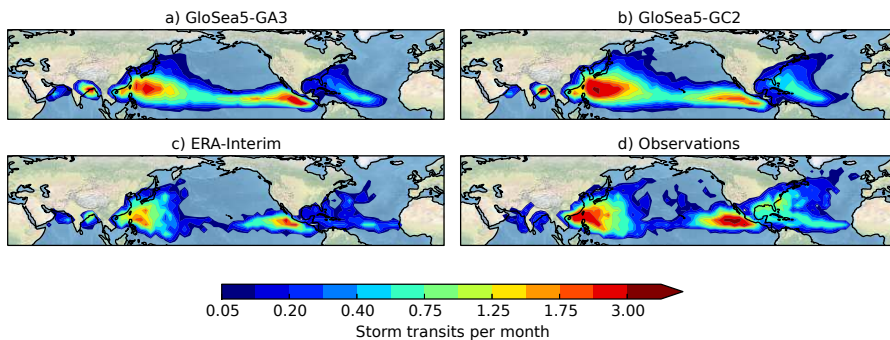
Figure 14. Seasonal hindcast (SEAS experiment) tropical cyclone track densities obtained using TRACK for JJA 1996–2009. **(a)** GloSea5-GA3, **(b)** GloSea5-GC2, **(c)** ERA-I, **(d)** observations (North Atlantic and eastern Pacific from HURDAT; western North Pacific and North Indian Ocean from the US Navy's Joint Typhoon Warning Centre (JTWC) best-track (Chu et al., 2002).
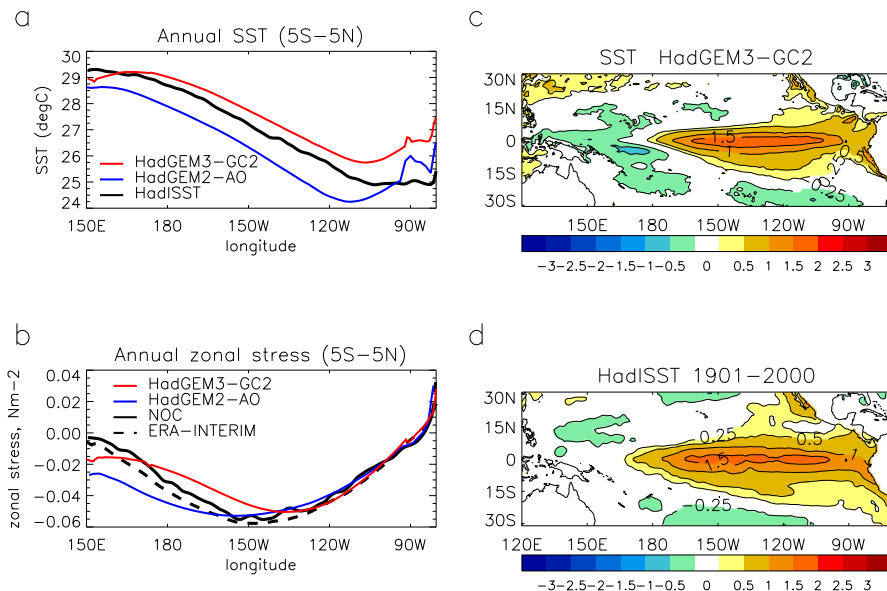
Figure 15. (a) Annual mean Pacific SST averaged over 5° N–5° S for the CLIM experiment. Black lines show HadISST. **(b)** Annual mean Pacific wind stress averaged over 5° N–5° S for the CLIM experiment. Black lines show ERA-I (dashed) and Southampton Oceanography Centre climatology (Josey et al., 1998) (solid) **(c)** El-Niño SST composite from HadGEM3-GC2 CLIM experiment. **(d)** El-Niño SST composite from HadISST observations. Composites are of events with a Niño 3.4 SST anomaly > 0.8 K.

Figure 16. Proportion of teleconnections to Africa in each of the five skill categories (details in in Rowell, 2013): green: at least reasonable model skill; yellow: marginal skill; pale brown: moderate and significant difference between model and observed teleconnection strength, dark brown and red: poor or very poor skill; and white: SST-rainfall associations of little practical interest. CMIP3 and CMIP5 models, together with HadGEM3-GC2 CLIM experiment, are ranked by the number of teleconnections that do not differ significantly from those observed at the 10 % level.