THE UK EARTH SYSTEM MODEL

PROJECT REPORT

Recommending a resolution for UKESM-LO

Till Kuhlbrodt, Colin Jones, Alistair Sellar and Robin S. Smith

May 20, 2015

Summary

To determine the best resolution for UKESM-LO, the low-resolution version of the UK Earth System Model, several test configurations have been built, tested and analysed. We focus primarily on two different atmospheric resolutions, N48 and N96.

The analysis of the scientific and computational performance of these two resolutions shows several significant deficiencies in the science of the model at N48, whilst the scientific performance at N96 scores well across a range of standard tests. Many assessment metrics of the N48 configurations are clearly worse than at N96, and there is a marked lack of transient eddy kinetic energy (TEKE) in the mid-latitude jet stream regions of both hemispheres at N48, an indication of an unrealistic simulation of atmospheric variability and low meridional heat transport. Initial efforts to increase the TEKE at N48 have shown no improvement. Running the current version of the Unified Model (UM) at N48 also poses major technical difficulties, as it is not one of the standard resolutions supported by the Met Office. Although the N96 model is more computationally expensive, the aim of achieving a throughput of at least 10 years of simulation per day has almost been reached already with these test configurations.

We have also evaluated two resolutions of the NEMO ocean model, ORCA1 and ORCA025, in coupled UM-NEMO configurations. In a coupled N96-ORCA1 configuration, the scientific performance is even better in some regards than in the standard HadGEM3 GC2 configurations N96-ORCA025 and N216-ORCA025. Improving the computational efficiency of the model through lengthening the time step of the atmosphere component does not significantly affect the overall scientific performance. The coupled N96-ORCA1 configuration can currently be run at nearly 10 model years per day, with further acceleration envisaged. We therefore recommend that N96-ORCA1 be the standard resolution for UKESM-LO.

1 Introduction

The UK Earth System Model (UKESM) is currently under development, with a planned release in Summer 2016. UKESM will exist in two versions, with high (UKESM-HI) and low (UKESM-LO) resolutions. While UKESM-HI will be the flagship version for key centennial simulations, UKESM-LO will be used to complement UKESM-HI to make ensembles of simulations in cases where UKESM-HI is too expensive—for example, CMIP6 simulations such as future projections in scenarioMIP, or detailed climate-chemistry simulations made for AerChemMIP. For this reason we are concerned to establish that UKESM-LO and UKESM-HI have similar science performance and projection responses, at least for averages on longer time scales and larger spatial scales. In addition, UKESM-LO is designed to be used for long simulations (e.g. millennial timescales) and for developing and testing parameterizations in a model framework that is similar to UKESM-HI but quicker to work with. All of this means that UKESM-LO needs to have scientific performance characteristics that are similar to UKESM-HI.

Several resolutions for the atmosphere and ocean components of UKESM-LO are being considered. For the atmosphere, these are the N48 and N96 resolutions (grid spacing approx. 300 km and 150 km) of the Unified Model (UM). The choice of these resolutions results from the need to find a balance between an acceptable computational performance for ensemble runs and millennial simulations on the one hand and an acceptable climate and scientific performance on the other hand. A resolution lower than N48 was not considered since that would be lower than that of the long-standing coupled model HadCM3 (*Gordon et al.*, 2000). A resolution greater than N96 was not considered as it would be computationally too expensive for the simulations that UKESM-LO is intended to conduct. The vertical resolution was left unchanged at 85 levels in order to maintain traceability to HadGEM3.

The ORCA1 and ORCA2 resolutions (approx. 1° and 2°) of the ocean model NEMO (Nucleus for European Modelling of the Ocean) are under consideration here for similar reasons. Based on these resolutions, the present report gives an overview of the UKESM-LO test configurations and their scientific and computational performance with a view to recommending a preferred standard model resolution.

The UKESM core group (the development team) has consulted with the UK academic community about the development of UKESM-LO through various activities. We mention here an NCAS Forum, held at the University of Reading in March 2014, where the questions and challenges raised in the present report were comprehensively discussed. A brief report of the Forum can be found at http://collab.metoffice.gov.uk/twiki/bin/view/Project/UkesmLoResolution . As a follow-up to this NCAS Forum, a survey on the evaluation criteria and the scientific objectives of UKESM-LO was conducted among the participants of the Forum. A report on this survey is also available on the above website.

UKESM as a whole is being developed as a set of community model versions that are planned to be operational for the UK community from approximately mid 2016 until the end of 2022. Hence UKESM is a model that will be in use for at least 6 to 7 years from now. We expect high-performance computing (HPC) resources to increase over this period. Such a forward perspective should be borne in mind when thinking about what model resolution and complexity is appropriate for UKESM1-LO.

2 Test configurations

Several atmosphere (UM) configurations with resolutions of N48 and N96 have been developed, run and evaluated. For all the configurations, the starting points of the development were the standard HadGEM3 configurations in the GC2 release (*Williams et al.*, 2015). We use UM version 8.6 throughout. N96 is one of the standard resolutions of HadGEM3, hence only small changes had to be applied for our purposes. In contrast, the N48 resolution is not part of HadGEM3, and significant technical work was required to simply run even a basic version of this configuration.

The model runs were forced with a fixed present-day carbon dioxide concentration. The atmosphere was started from restart files of the GC2 standard configurations (see the overview on https://code.metoffice.gov.uk/trac/GA/wiki/GADocumentation/GA6.0), while in the coupled configurations the ocean and sea ice models were started from climatologies, as spun-up states compatible with the atmosphere were not available for these tests.

The test configurations use only the physical models (atmosphere, ocean, sea ice, land [AOIL]), due to the difficulty of generating ancillary files for the Earth system components at multiple resolutions. Our expectation is that the implication of resolution differences on ESM components will be reasonably clear from the physical model results. We invite the ESM community to investigate these results from the perspective of their component model and feed back the importance of the differences in the physical model tests. We do plan to run tests using UKCA's age of air tracer in order to measure the performance of the stratospheric circulation.

2.1 N48e AMIP

The semi-implicit, semi-Lagrangian dynamics used in the UM since the introduction of New Dynamics (ND) in the UM does not generally perform as well on coarse grid resolutions as the explicit schemes used in the past (*Stratton*, 2004).

A number of scientific performance issues have been previously identified in the N48 ND UM, such as a lack of mid-latitude transient eddy kinetic energy (TEKE) and a very weak Atlantic meridional overturning circulation (AMOC). Such issues led to the Met Office withdrawing support for the N48 configuration. We wanted to ascertain whether the N48 climate biases were improved with the current ENDgame dynamics, given theoretical arguments that the atmosphere should be more energetic under this scheme.

The N48e AMIP (i.e. atmosphere-only) configuration was developed from the standard N96e AMIP GC2 job antie. The suffix ”e” to the resolution denotes the ENDgame dynamical core. Since the transition from New Dynamics to ENDgame involved a change in the grid, to run an N48e simulation meant that all ancillary files had to be regenerated for these tests. This was extremely time-consuming, as there is no comprehensive scripted process—supporting this is only one of the jobs that would likely have to be taken on on an ongoing basis by the UKESM team should UKESM-LO use N48 since the Met Office do not support N48 anymore. Producing a set of ancillary files for a coupled configuration would be an even bigger task than that undertaken for the atmosphere-only model. Since many of the climate biases found with the N48 ND UM were apparent without coupling to an ocean, we chose to avoid this additional effort initially.

Difficulties in creating the necessary ancillaries also meant that our N48e runs use prescribed climatological aerosols rather than the interactive aerosols (CLASSIC) used in the N96e configurations. To see whether this different treatment of aerosols contributes to the differences seen between N48e and N96e, we compared the impact of interactive vs prescribed aerosols in two N96 configurations (dkeus vs. djtkn). The conclusion of this comparison is that the different treatment of aerosols in N48e and N96e is largely insignificant with respect to the differences in scientific performance we outline later.

Once a working N48 ENDgame configuration existed, it was found that the TEKE remained very low in the mid-latitudes of both hemispheres. Earlier work (*Sanchez et al.*, 2013) with the N48 ND UM has shown that this lack of mid-latitude TEKE results from the high diffusivity of the Semi-Lagrangian advection scheme compared to Eulerian advection as used in HadCM3, and suggested it might be improved somewhat by a vorticity confinement scheme, amongst other parameterisations. Unfortunately the preliminary work on the vorticity confinement scheme has not been ported to ENDgame, and there are no plans at the Hadley Centre to do so. The amount of work required to fully develop such a scheme for the ENDGame dynamics is substantial, and beyond the scope of our present effort. However, we did develop and analyse several further N48e configurations using other pre-existing parameterisations aimed at enhancing the model TEKE, including (a) using quintic interpolation in the Ritchie departure point scheme and (b) using the full stochastic physics that *Sanchez et al.* (2015) developed with the aim to enhance the small-scale atmospheric variability of the UM at N48. Although some differences were found, these parameterisations had no major impact on the TEKE or on the assessment metrics in our runs. The fact that quintic interpolation does not substantially increase the TEKE is in line with the results from *Stratton* (2004) and *Sanchez et al.* (2013). More details on these tests are available on the collaboration twiki.

2.2 N96e-ORCA1

The N96e-ORCA1 configuration was developed starting from the N96-ORCA025 GC2 vn8.6 configuration (anqjo) of the Met Office Hadley Centre. Several new branches and hand edits are applied to overcome failures for technical reasons (e.g. properly initializing individual variables or allocating processors). Some ancillary files had to be re-created (e.g. the lake fraction file). The coupler OASIS is now run on a single processor. Parameter settings for NEMO were updated to be appropriate for ORCA1, based on advice from physical ocean modellers at the Met Office and at the National Oceanographic Centre (NOC). The NEMO namelists for both the present N96-ORCA1 configuration and the N96-ORCA025 configuration can be found in the Attachments of the Collaboration TWiki page http://collab.metoffice.gov.uk/twiki/bin/view/Project/UkesmLoResolution .

Two stable versions of N96e-ORCA1 have been developed. The first version (run ID: xjpdu) has an atmospheric time step of 20 min (the standard N96 time step in HadGEM3-GC2), an ocean time step of 45 min and a coupling frequency of 3 h.

The second version (xkyld) was built with the aim to speed up the model in terms of CPU time. To this end, the atmospheric dynamics time step was extended to 30 minutes. The radiation time step was extended to 3 hours, with 1 hour solar zenith angle updates to the shortwave increments as described in *Manners et al.* (2009). The intermediate 1 hour updates do not include absorption by oxygen or other gases, but calculate increments due to changes in cloud.

Both versions have been run for more than 50 years. The full output is available on JASMIN and MASS and we present some analysis below.

Initially it was planned to test ORCA2 as well. Since the atmosphere uses the bulk of the computing time (around 80%), in terms of coupled model throughput there is little computational gain in degrading the model resolution from ORCA1 to ORCA2. We have therefore focused our efforts on speeding up the UM. In addition, there is substantial experience in the UK with the ORCA1 configuration at the Met Office, at NOC and at other institutions, compared to the little-used ORCA2.

3 Computational performance

The N96e-AMIP configuration currently runs at just under 10 model years per wallclock day on nearly 400 PEs, while the N48e-AMIP configuration approaches 15 years/day on 128 PEs (ignoring any idle time in queues on the high-performance computers). See Fig. 1 for details.

Configura-tion	Run ID	R.T./ mon [min]	yrs/ day	Δ [per cent]	PEs
N96e-ORCA1	xjpdu	28.1	4.3		198+1+ 24=**223**
	xjpdy: nn_pdl=0	21.2	5.7	-25%	396+1+ 24=**421**
	xkyld: dt=30 min, dt_rad=3h (1h incr)	18.0	6.7	-36%	420+1+ 48=**469**
	xkyli: dt=30 min, dt_rad=3h (1h incr)	12.3	9.8	-56%	420+1+ 128=**549**
N96e-AMIP	xkpzo: dt=20 min, dt_rad=1h	27.2	4.4		198
	xkpzp: dt=30 min, dt_rad=1h	21.3	5.6	-22%	
	xkpzq: dt=30 min, dt_rad=3h (1h incr)	16.8	7.1	-38%	
	xkpzt: dt=30 min, dt_rad=3h (1h incr)	12.4	9.7	-54%	396
N48e-AMIP	xjwoc: dt=30 min	8.1	14.8		128
	xjwod: dt=20 min	10.0	12.0	+24%	

Figure 1: The computational performance of several test configurations of UKESM-LO. Explanation of the abbreviations: R.T.: run time on MONSooN, Δ: change in R.T. per month, PEs: processing elements on MONSooN (UM [+ OASIS + NEMO = total]). In the Run ID column, nn_pdl is an ocean mixing parameter, dt is the atmospheric dynamics time step, dt_rad is the radiation time step, and incr stands for diagnostic increments in the radiation scheme.

Although the N48e configuration is cheapest, this comes at a significant cost in terms of scientific performance, as will be shown later. By lengthening the timestep, an N96 model can achieve two-thirds of the N48 throughput, although this requires substantially more processors to be used in parallel. The fastest coupled N96e-ORCA1 configuration achieves just under 10 years/day. There are various promising avenues for further increases in model speed being investigated:

(i) It may be possible to run the UM at N96 with a time step of one hour, and this is currently being investigated in N96e atmosphere-only configurations. We have already demonstrated that the N96 UM is stable with a 45min timestep, although due to technical problems that configuration does not produce usable diagnostic output.

(ii) It is very likely that NEMO (ORCA1) can also be run with a time step of one hour, increased from the standard 45 min. This should give a further CPU time gain Δ of roughly 19%.

(iii) More PEs could be used for the two main components of UKESM-LO, with improved load balancing. Initial work along these lines has led to an additional 20% speedup. This will be investigated further.

Our present target for UKESM-LO, guided by the NCAS Forum participants, is a computational throughput of at least 10 simulated years per day. We are confident that this can be achieved with the coupled physical version (AOIL) of UKESM-LO. However, the addition of ocean biogeochemistry (MEDUSA) and atmospheric chemistry and aerosols (UKCA) will add more cost to the overall model. Indications at present suggest the same overall throughput may be achievable by an increased use of computational cores (e.g. increased parallelization), although this can only be fully assessed once all components are fully coupled. Some work has been initiated in the UKESM core group to investigate modifying the call frequency for UKCA from the standard 1 hour to a longer call frequency (e.g. up to 3 hourly). This will speed up UKCA as a component of the full model and may be scientifically justifiable if the radiation time step in UKESM1-LO is also 3 hours. This work is in the very early stages.

4 Scientific performance

The analyses were performed with current Met Office tools for evaluating the climate model: these are "Maverick" for the atmosphere, "assess_seaice.pro" for the sea ice and ”ocean_assess.py” for the ocean. The AutoAssess module of Maverick produces assessment metric plots which present large numbers of standardised metrics (scalar values, many based on root mean square differences between the model and an observational dataset), grouped by phenomena and regions. In these plots (for an example see Fig. 2), an ”experiment” and a ”control” run are compared with respect to these metrics. The position of the markers (dots) shows the value of the metrics for the ”experiment”, normalised by the metric for ”control”. The solid grey bars indicate the size of the model internal variability in each metric, as a proportion of the control metric value (centred on unity). These are included to indicate whether a change in the metric is significant or could just be due to internal variability. The thin black ‘T’ bars indicate the size of the uncertainty in observations, as a proportion of the ”control” metric value (centred on zero). Finally, for each marker there is a colour code:

- Green: Magnitude of model metric in ”experiment” is less than observational uncertainty
- Amber: Metric in ”experiment” is outside the range of observational uncertainty but within the range of variability in the ”control”, so any change may not be significant
- Red: Magnitude of model metric in ”experiment” is significantly worse than that of ”control”

For more details on these metrics see *Williams et al.* (2013).

A comprehensive analysis of the GC2 runs used in this section can be found at http://collab.metoffice.gov.uk/twiki/bin/view/Development/GC2Assessment .

4.1 N48e AMIP

We have compared the N48e AMIP configuration with several other configurations:

- N96e AMIP GC2
- HadAM3, the atmosphere-only version of HadCM3
- N48 AMIP GA3.0 New Dynamics

As outlined in sec. 2.1, we decided to first build an N48e AMIP configuration because many of the climate biases seen in previous N48 configurations can be seen in atmosphere-only runs, and building a coupled N48e configuration would have entailed substantially more technical effort. The N48e runs shown here use prescribed aerosols and omit the additional parameterisations we tested in an attempt to enhance model variability (see sec. 2.1 and the collaboration twiki pages for more details); neither of these factors significantly alters the results shown.

The comparison with N96e AMIP directly assesses the performance of N48e AMIP in comparison to the current version (GC2) of HadGEM3. The other two comparisons assess the performance of N48e AMIP with previous N48 configurations, namely: (1) The atmospheric component (HadAM3) of the HadCM3 coupled model, a successful low-resolution climate model predating HadGEM; and (2) a more recent version of N48 using the New Dynamics rather than the ENDgame dynamical core.

4.1.1 N48e vs. N96e AMIP GC2

We first present standard performance metrics, before looking at some features of the scientific performance in spatial plots (derived from the Met Office Validation Notes). Using averages over the second 10 years of the runs (year 10 to 20), we find N48e clearly displays worse performance metrics than N96e. For instance, the metrics for the Global Troposphere (Fig. 2), Radiation (Fig. 3), mid-latitude blocking (Fig. 4) and mid-latitude storm tracks (Fig. 5), are in the vast majority of cases clearly degraded in moving from N96e to N48e.

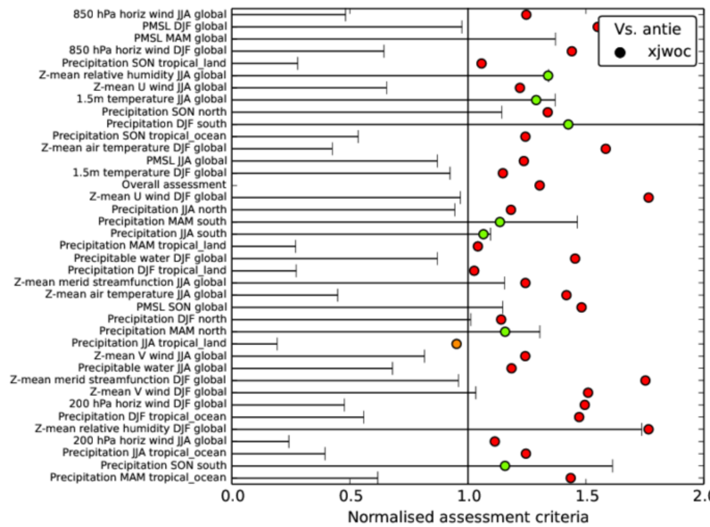

Figure 2: Global troposphere metrics calculated by AutoAssess for the comparison of N48e AMIP (xjwoc) with N96e AMIP GC2 (antie). Based on an average of year 10 to 20 of the runs.

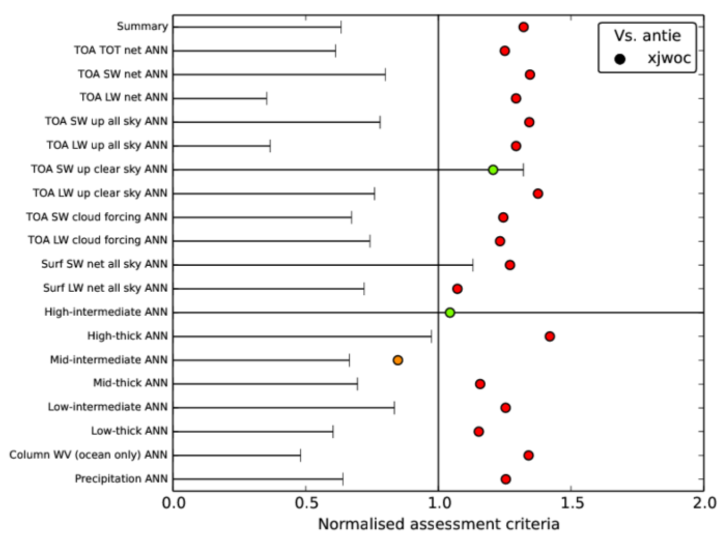

Figure 3: Radiation metrics calculated by AutoAssess for the comparison of N48e AMIP (xjwoc) with N96e AMIP GC2 (antie). Based on an average of year 10 to 20 of the runs.

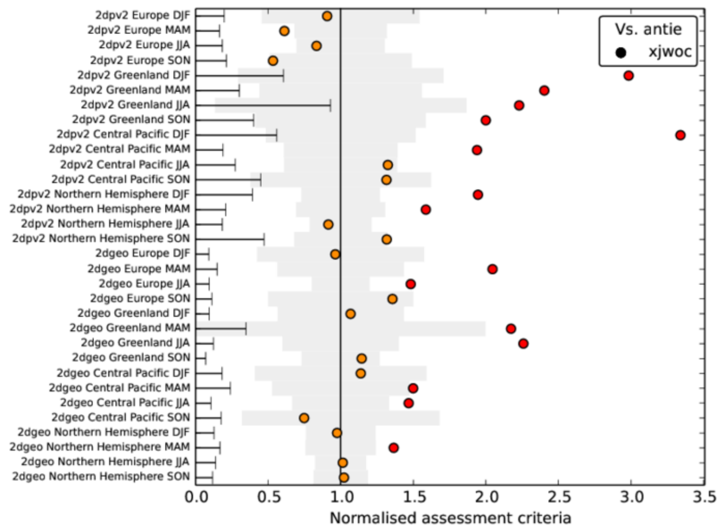

Figure 4: Mid-latitude blocking metrics calculated by AutoAssess for the comparison of N48e AMIP (xjwoc) with N96e AMIP GC2 (antie). Based on an average of year 10 to 20 of the runs.

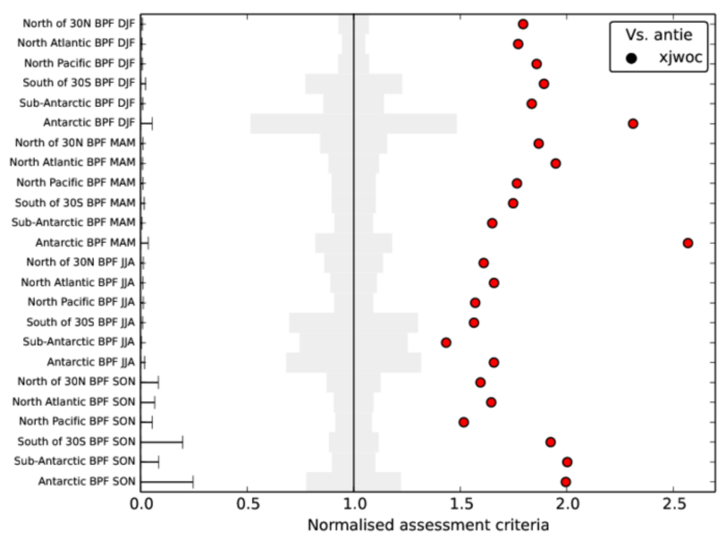

Figure 5: Mid-latitude stormtrack metrics calculated by AutoAssess for the comparison of N48e AMIP (xjwoc) with N96e AMIP GC2 (antie). Based on an average of year 10 to 20 of the runs.

The results show clear seasonal mean biases in mean sea level pressure (SLP) and cross sections of atmospheric temperature and precipitation in the N48e simulation. There is also a large positive bias in Arctic SLP (Fig. 6 and Fig. 7), accompanied by a negative bias over Antarctica and in several mid-latitude regions (e.g. the North Pacific in DJF, Fig. 6).

The temperatures around 200 hPa in both high-latitude regions are too low by several Kelvin in N48e (Fig. 8), while the tropical tropopause is too warm—here the existing bias in N96e is increased in N48e by roughly 1 K. Such tropopause-region biases will negatively impact important chemistry processes in UKESM1-LO and are known to be difficult to improve due to the numerous atmospheric processes that are involved. Extensive work to improve tropical tropopause temperature and humidity biases in HadGEM3 is currently being undertaken by a UM process evaluation group (PEG). With regards to precipitation, the dry bias in N96e over the Maritime Continent is exacerbated in N48e (Fig. 9) while over the tropical Pacific a wet bias in N96e is also made worse in N48e.

In line with the temperature and precipitation biases, N48e has strong wet biases compared to N96e in relative humidity in both high-latitude regions just below 200 hPa, as well as in the Antarctic troposphere. These biases are strongest in northern hemisphere winter (Fig. 10), but are also evident in the southern hemisphere. The same applies for the dry bias throughout the tropical troposphere. In a general sense, humidity biases of either sign that exist in N96e are increased in N48e.

As mentioned earlier, there is a large negative bias in transient eddy kinetic energy in N48e in the mid-latitudes of both hemispheres (Fig. 11). In this case, the equivalent bias in N96e is very small. Hence the large negative bias seems to be a feature specific to the N48e resolution.

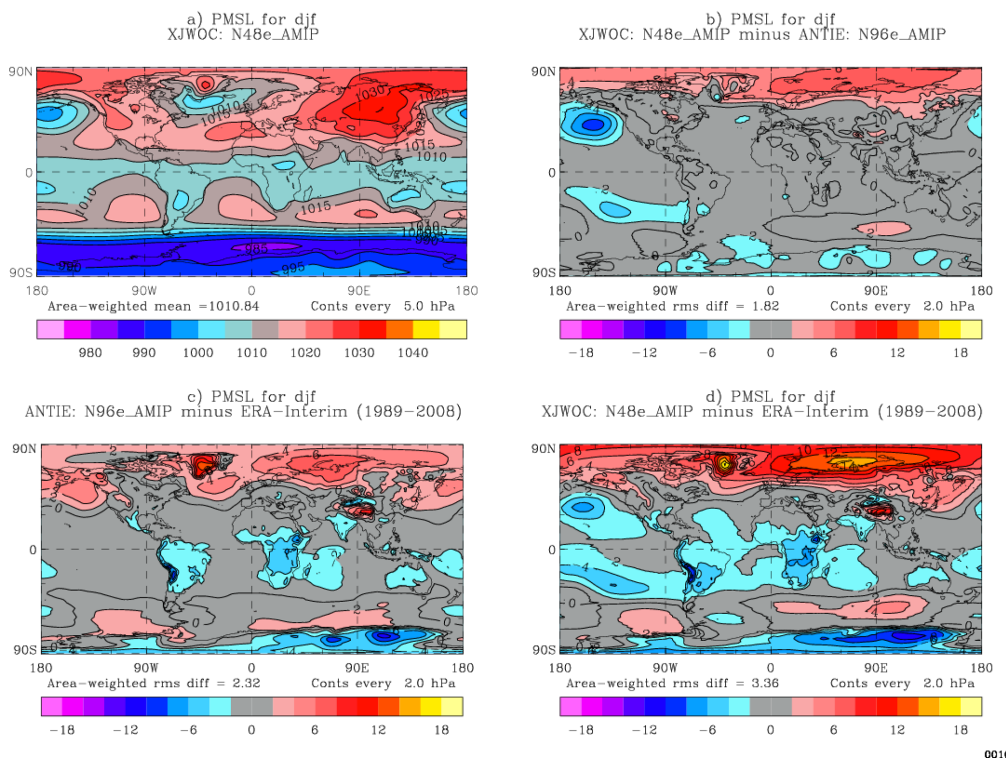

Figure 6: Mean sea level pressure in N48e (xjwoc) and N96e (antie) in northern hemisphere winter (DJF) in comparison with observations (ERA-interim).

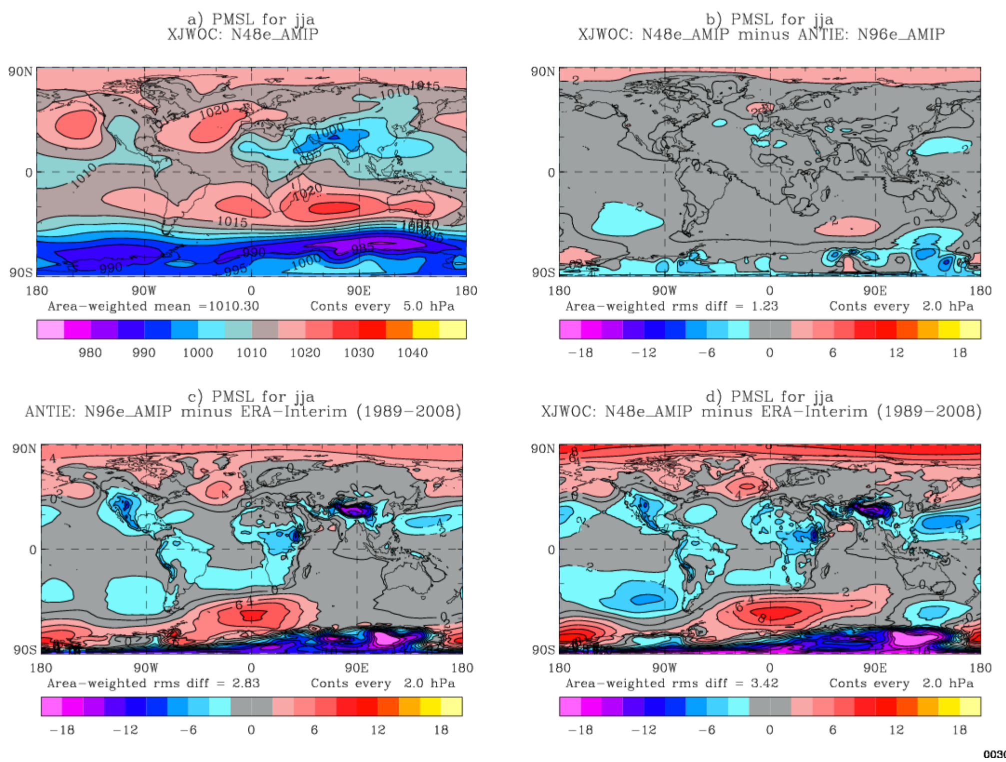

Figure 7: Mean sea level pressure in N48e (xjwoc) and N96e (antie) in southern hemisphere winter (JJA) in comparison with observations (ERA-interim).

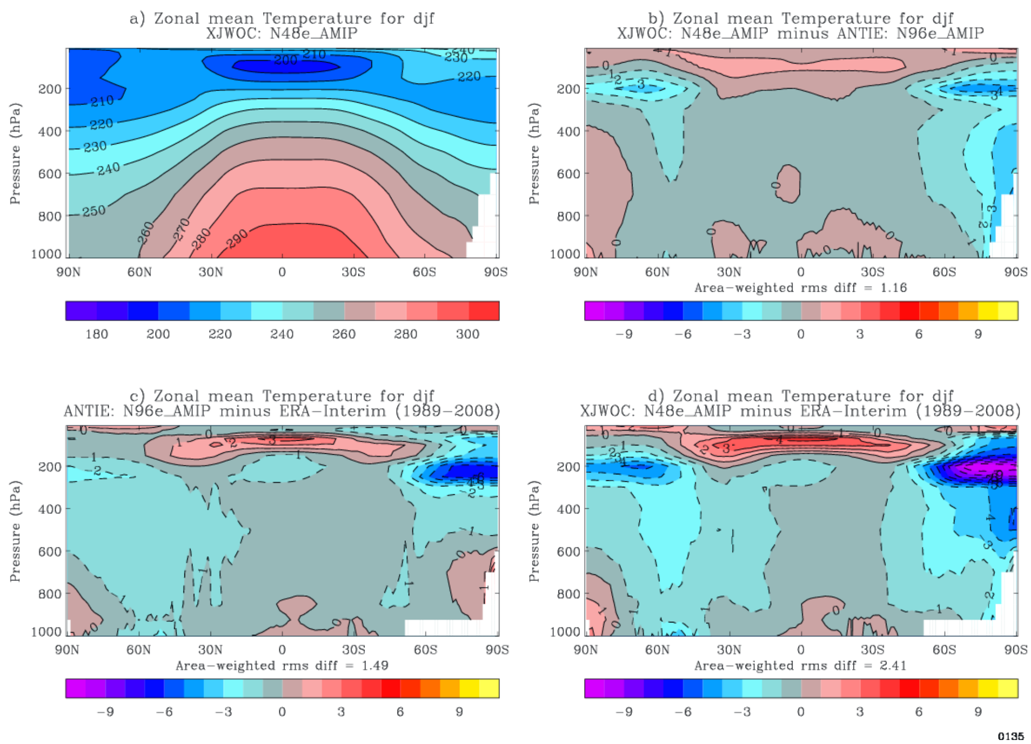

Figure 8: Zonally averaged temperature (in K) in N48e (xjwoc) and N96e (antie) in northern hemisphere winter (DJF) in comparison with observations (ERA-interim).

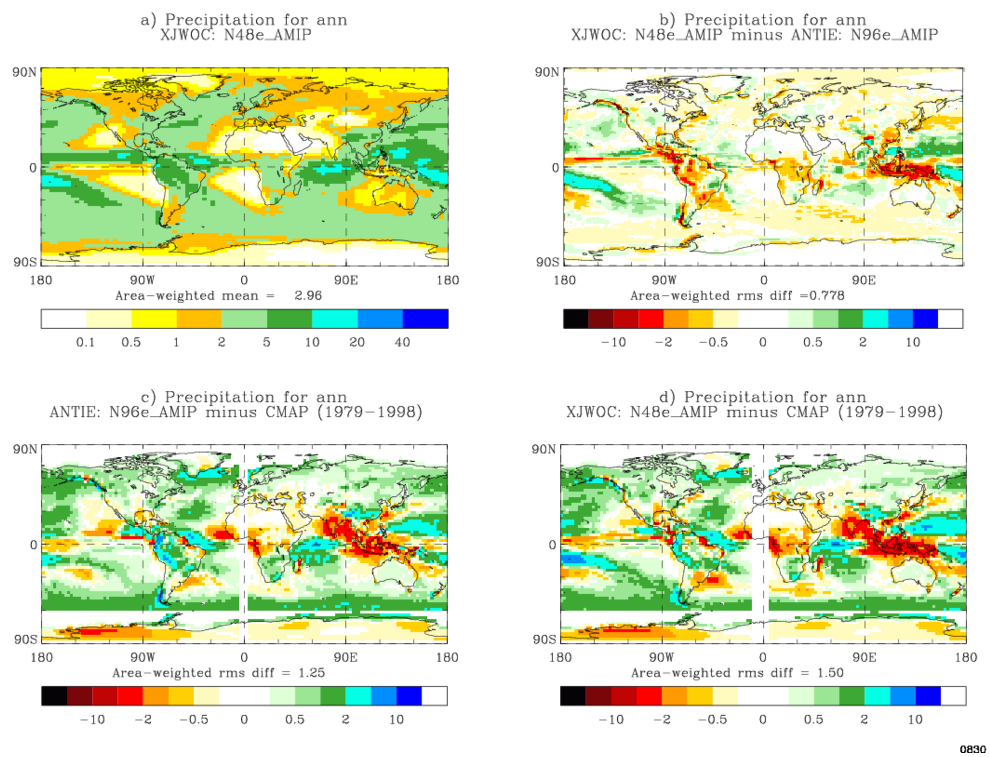

Figure 9: Annually averaged precipitation (in mm/day) in N48e (xjwoc) and N96e (antie) in comparison with observations (CMAP).

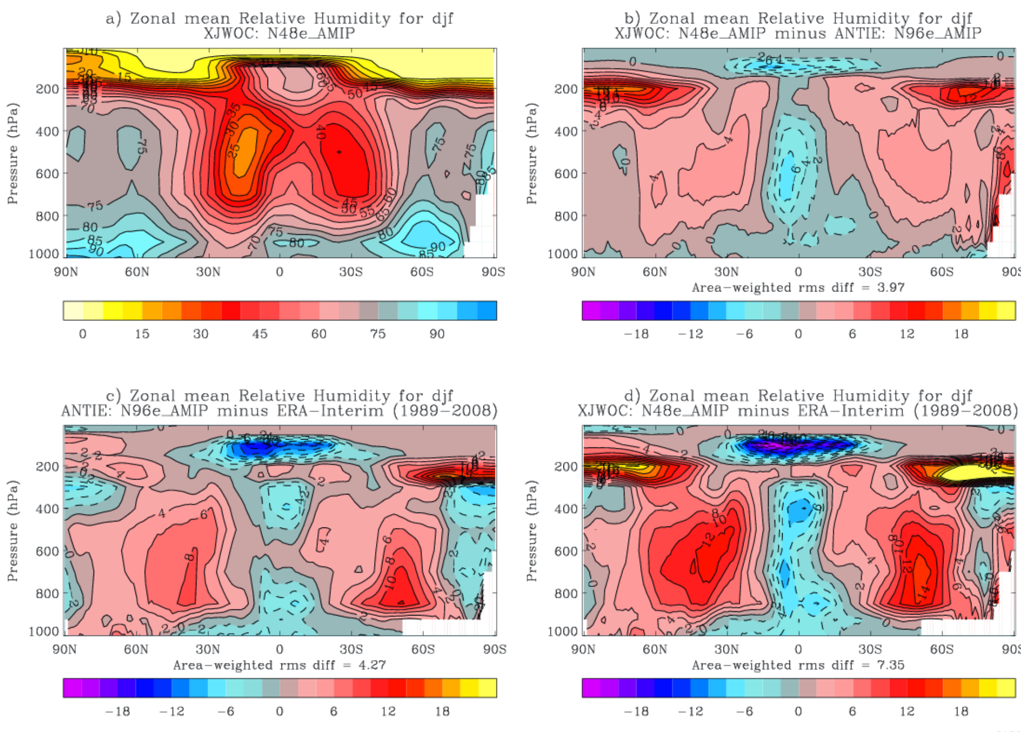

Figure 10: Zonally averaged relative humidity (per cent) in N48e (xjwoc) and N96e (antie) in northern hemisphere winter (DJF) in comparison with observations (ERA interim).

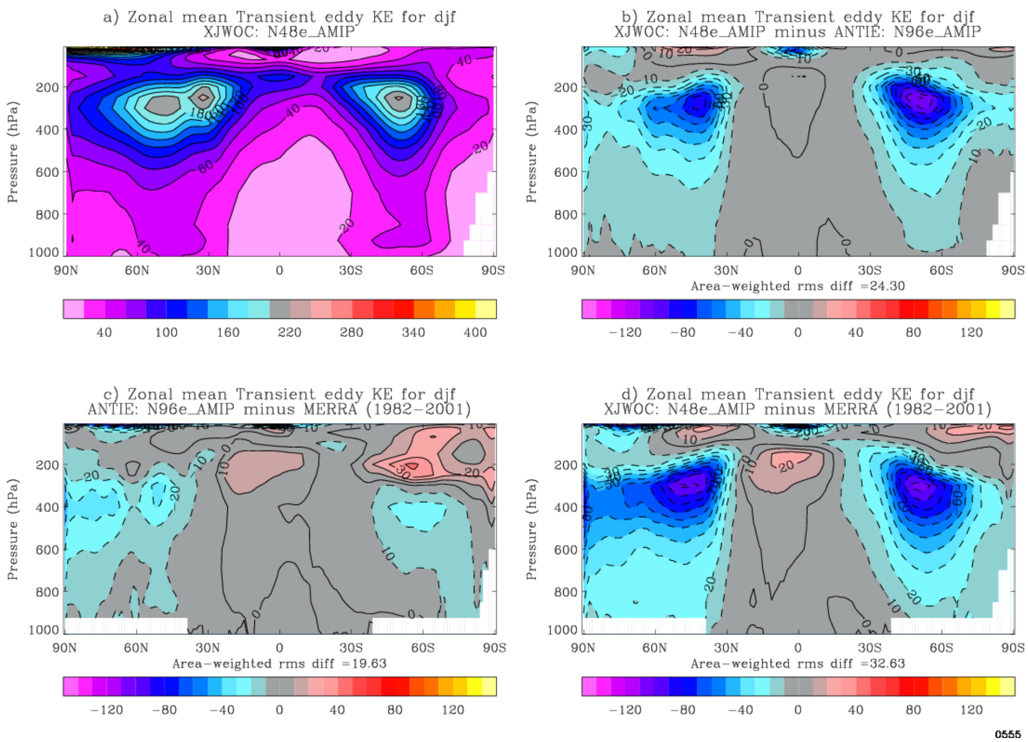

Figure 11: Transient eddy kinetic energy in N48e (xjwoc) and N96e (antie) in northern hemisphere winter (DJF) in comparison with observations (MERRA).

4.1.2 N48e vs. HadAM3

One of the strong requirements that emerged from the NCAS Forum on UKESM-LO was that performance of UKESM-LO should be better than HadCM3. From a practical point of view users are unlikely to move from HadCM3 to UKESM unless it performs better scientifically, given that UKESM is computationally more expensive, even at N48 resolution. With HadCM3, typically 20 model days per year can be achieved on just 20 PEs, while for our N48e AMIP configuration we needed 128 PEs to achieve roughly 15 years/day (cf. Fig. 1). That said, work on extending the dynamical time step and radiation call frequency in N96e would both likely be transferable to N48e and could result in a throughput of approx. 25 years/day or better.

The computational advantage of HadCM3 over UKESM-LO comes largely from the fact that HadCM3 has simpler physics and only 19 vertical levels, compared to the 85 levels used in all UKESM configurations. The more sophisticated science and greater integration of Earth System components in UKESM means that UKESM-LO will be able to be used to address a much wider range of scientific questions than HadCM3 can, as well as having the advantage of being directly linked to a high resolution flagship coupled model, UKESM-HI. Of particular note, while the lower number of vertical levels in HadCM3 directly increases model speed, the stratosphere in HadCM3 is very poorly resolved and its ability to investigate stratosphere-troposphere and stratospheric chemistry-climate interactions is severely limited.

Since we run N48e in atmosphere-only mode (AMIP), we compare it with HadCM3 in its atmosphere-only configuration HadAM3 (aawei). Due to the differences in the grid (HadAM3 using the old Eulerian dynamical core and N48e with ENDgame) only a limited number of plots could be created by Maverick. In comparison to HadAM3, N48e again shows strong positive biases in sea level pressure and 500 hPa height over the Arctic (not shown). For annual precipitation, there are dry biases in the tropics over land, notably in Africa, India and the Maritime Continent, while many tropical oceanic regions are too wet (Fig. 12). In line with *Stratton* (2004), we find that the TEKE at N48e is clearly lower than in HadAM3 (Fig. 13, see also the discussion in sec. 2.1).

The 200 hPa winds in JJA are much too strong over the tropics in N48e AMIP. However the 600 hPa winds in JJA fare better in N48e AMIP, especially over the Southern Ocean, because the jet stream is not located too far equatorwards as in HadAM3.

Some features related to cloud cover (improved over large regions of the

ocean, Fig. 14) and planetary albedo are closer to observations in N48e AMIP than in HadAM3, with reduced biases in 1.5 m temperature over land in the Americas and Africa. These are likely due to the significant developments in model physics between HadCM3 and HadGEM3, particularly cloud and boundary layer physics. For more details on the comparison of N48e (xjwoc) with HadAM3 (aawei), see the documentation on
http://collab.metoffice.gov.uk/twiki/bin/view/Project/UkesmLoResolution .

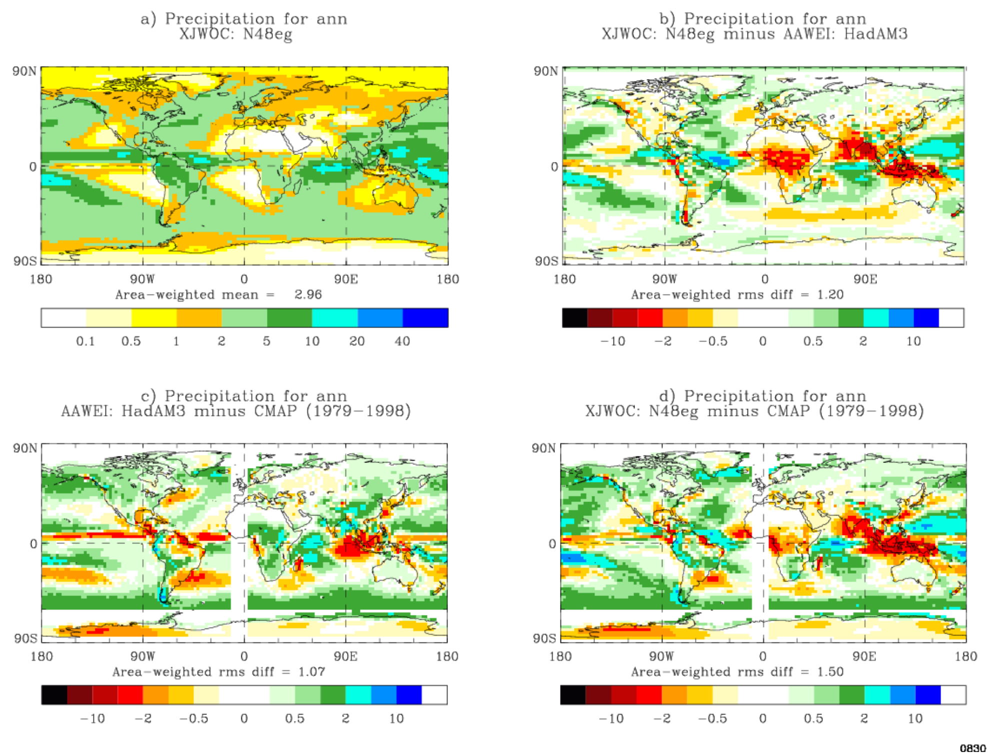

Figure 12: Annual mean precipitation in N48e (xjwoc) and HadAM3 (aawei) in comparison with observations (CMAP).

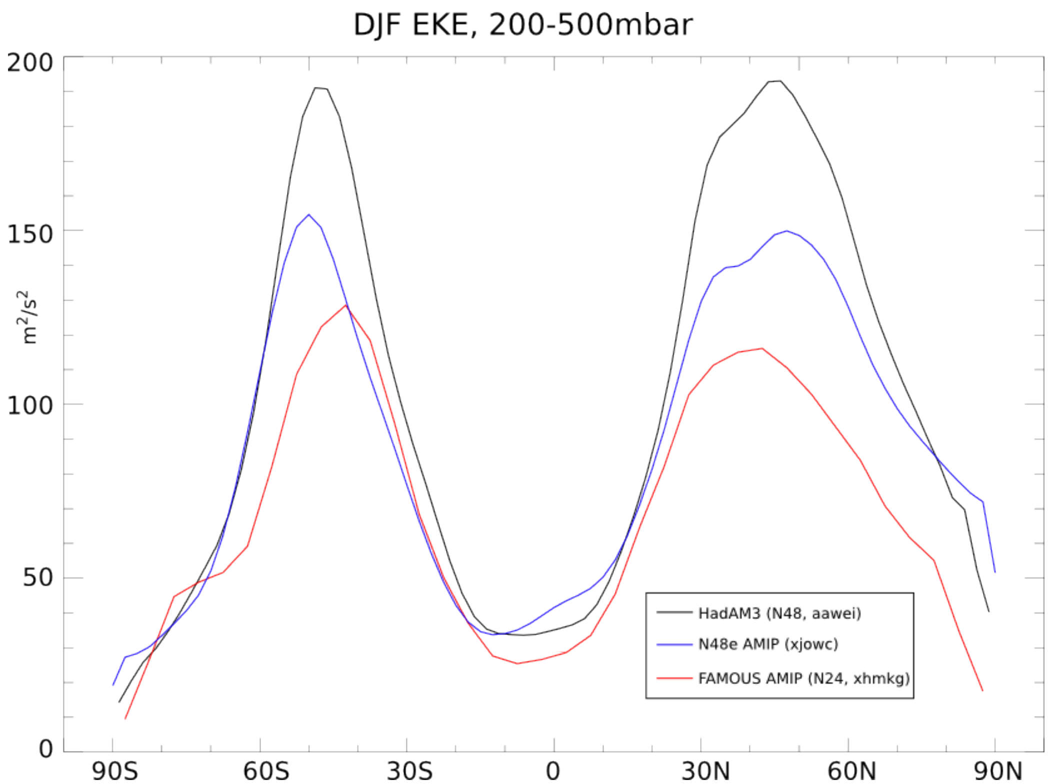

Figure 13: Transient eddy kinetic energy in N48e (xjwoc), HadAM3 (antie) and FAMOUS-A (xhmkg, an N24 version of HadAM3) in northern hemisphere winter (DJF).

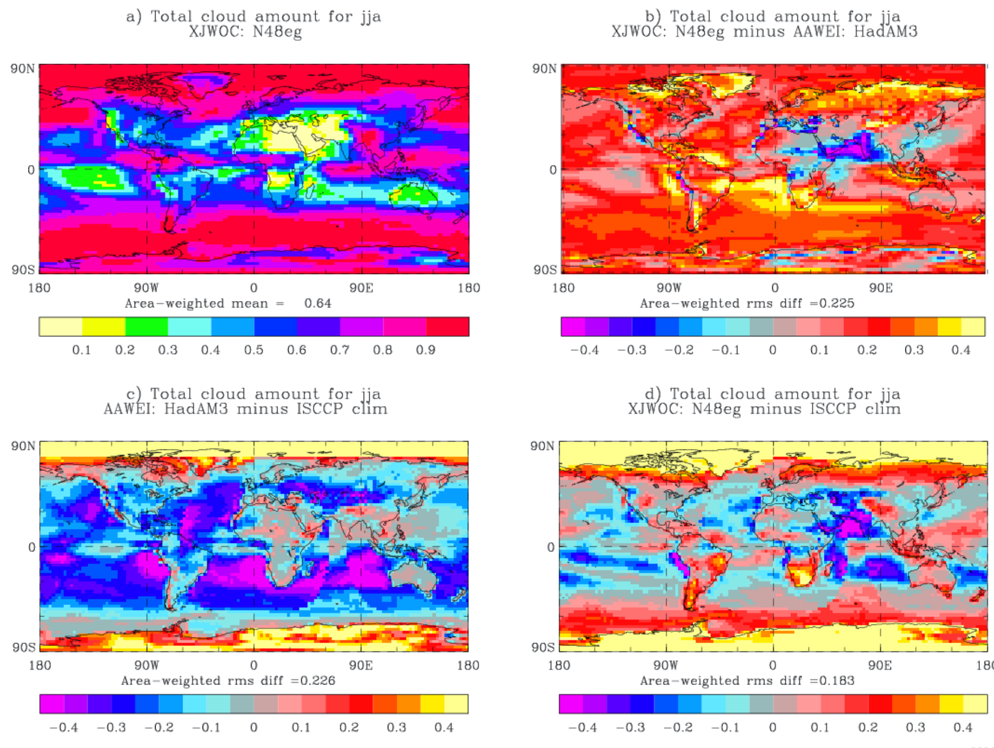

Figure 14: Total cloud amount (in fractions) in N48e (xjwoc) and HadAM3 (aawei) in northern hemisphere summer (JJA) in comparison with observations (ISCCP).

4.1.3 N48e vs. N48 AMIP GA3.0 New Dynamics

We particularly wanted to compare the ENDgame dynamics at N48 with New Dynamics (ND) because there is some evidence from tests at higher model resolution that ENDgame dynamics is more energetic than ND and may perhaps lead to a reduction in the known negative bias in TEKE seen at N48. If this improvement were not found (as unfortunately proved to be the case) then this would be a strong signal that an N48e coupled model, would have similar problems as N48 with ND, such as a weak ocean overturning circulation.

Between the two tested configurations (N48e AMIP [xjwoc] and N48 AMIP GA3.0 ND [akrdm]), there are several other significant differences on top of the ENDgame/ND change that might lead to the differences in scientific performance. Given the available data for N48 AMIP GA3.0 New Dynamics (akrdm), we could only compare an average over the first ten years of the run.

We find the following features in N48e AMIP in comparison with N48 AMIP GA3.0 New Dynamics:

- A slightly deteriorated bias in TEKE (Fig. 15)
- Worse metrics for Global Troposphere (Fig. 16), Stormtracks, ENSO and Radiation
- Slightly improved 200 hPa winds: less bias in the southern mid-latitude stratosphere

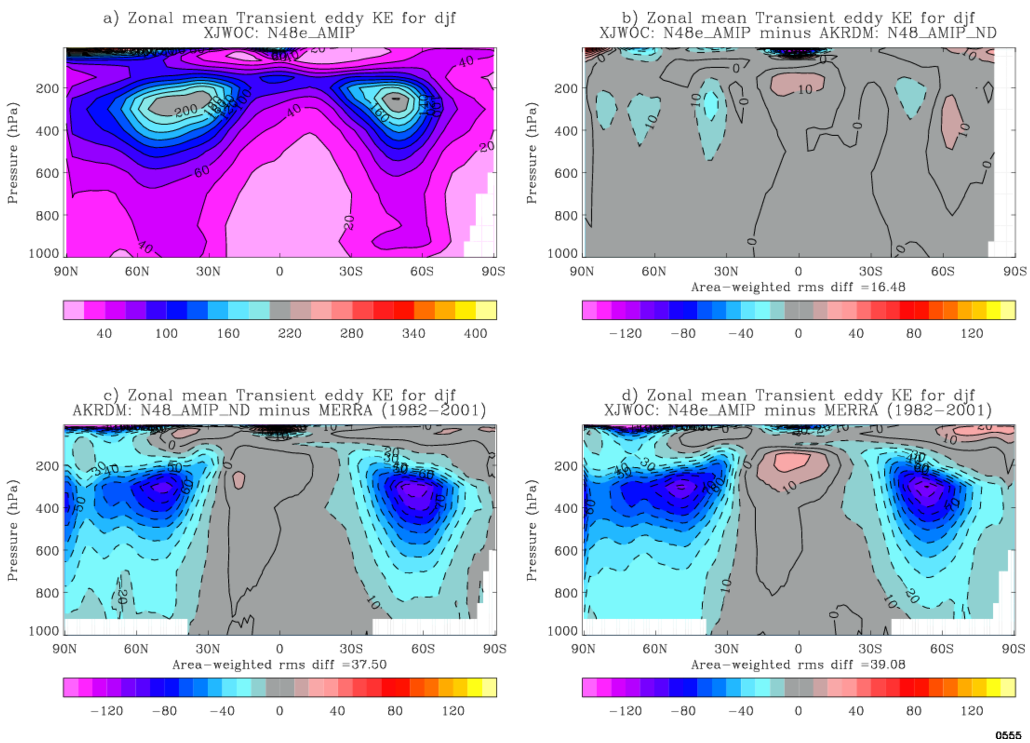

Figure 15: Transient eddy kinetic energy in N48e AMIP (xjwoc) and N48 AMIP GA3.0 (akrdm) in northern hemisphere winter (DJF) in comparison with observations (MERRA). Based on an average of year 0 to 10 of the runs.

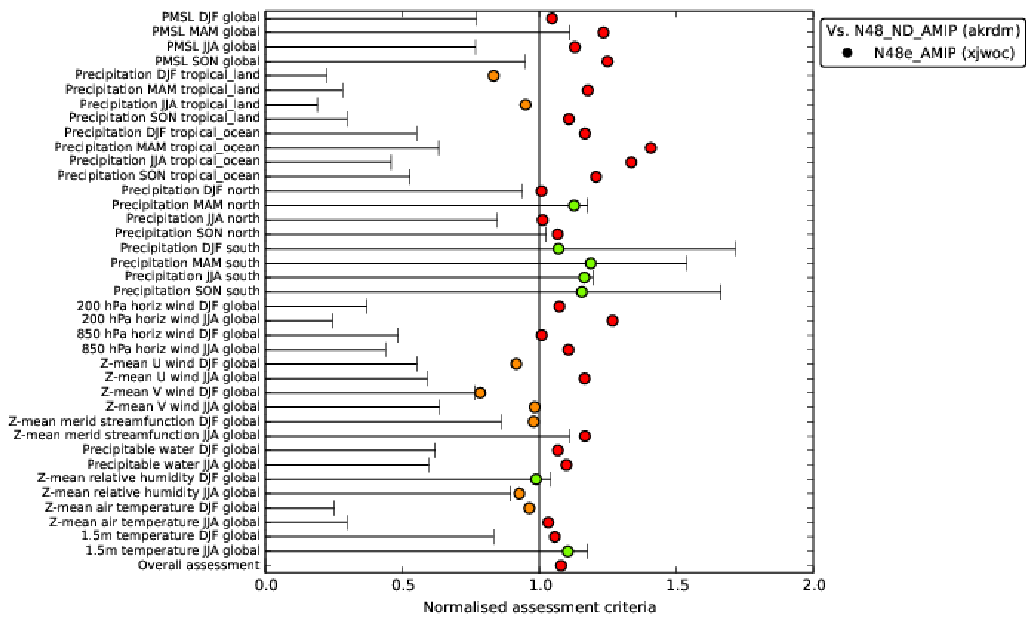

Figure 16: Global Troposphere metrics calculated by AutoAssess for the comparison of N48e AMIP (xjwoc) with N48 AMIP GA3.0 New Dynamics (akrdm). Based on an average of year 0 to 10 of the runs.

To sum up, the current N48e-AMIP configuration does not perform well scientifically in comparison to the N96e-AMIP configuration. Our efforts to improve on the marked lack of TEKE in N48e-AMIP did not yield a substantial improvement. In comparison with HadAM3, the current N48 version does not seem to perform much better, despite the increased computational cost of N48e. We also find the step from the New Dynamics dynamical core to ENDgame does not improve simulated TEKE. Due to these science performance results and the additional major complication of producing suitable ancillary files, we decided not to build a coupled N48e configuration. Instead, we focused on improving the computational performance of the UM at N96.

4.2 N96e-ORCA1

N96e-ORCA025 and N216-ORCA025 are standard HadGEM3 GC2 configurations. The latter resolution will be used for UKESM-HI. By comparing N96e-ORCA1 with N216-ORCA025, we compare physical models with the likely resolution of UKESM-LO with the resolution of UKESM-HI. By comparing N96e-ORCA025 and N96e-ORCA1 the effect of coupling N96e to a lower-resolution ocean model is assessed.

4.2.1 N96e-ORCA1 vs. N216-ORCA025 GC2

Two important applications of UKESM1-LO will be (i) to complement UKESM1-HI simulations in terms of increased ensemble members and (ii) efficient parameterization development and assessment. For both applications it will be important that the science performance of UKESM1-LO and HI have some degree of commonality (traceability). Comparison of these two resolutions is an initial step in assessing this requirement. We use the N96e-ORCA1 configuration with a 30 min time step (xkyld) for these comparisons.

Atmosphere We compare years 10 to 20 of both runs (the N216-ORCA025 run, anqjp, is only 20 years long). Most metrics in N96-ORCA1 are roughly on par with those from N216-ORCA025 GC2 (see example in Fig. 17). Exceptions are temperature and radiation (Fig. 18), where N96-ORCA1 fares slightly less well.

In terms of net downward radiation at the top of the atmosphere (TOA), N96-ORCA1 loses too much energy (-0.16 W/m^2) in comparison to observations (CERES EBAF; +0.5 W/m^2), while N216-ORCA025 gains too much energy (+1.67 W/m^2). From the components of the radiation budget (Table 1) it appears that N96-ORCA1

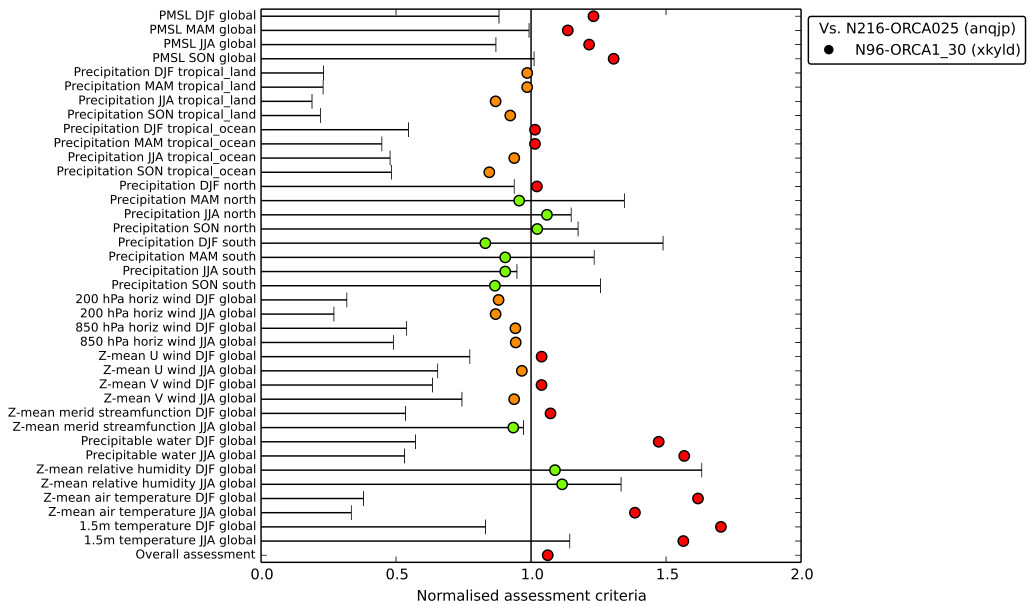

Figure 17: Global Troposphere metrics calculated by AutoAssess for the comparison of N96-ORCA1 (xkyld) with N216-ORCA025 GC2 (anqjp). Based on an average of year 10 to 20 of the runs.

is brighter at the top of atmosphere than N216-ORCA025, reflecting more shortwave radiation and emitting less longwave radiation. The increase in shortwave reflection occurs in areas of excessive sea ice cover.

These moderate radiation imbalances are not a major issue in themselves since there will be final tuning of the model in pre-industrial mode to ensure radiative balance. In development of the physical GC model a range of +0.25 W/m^2 to +1.5 W/m^2 is targeted for present-day coupled runs, which is designed to encompass observational uncertainty and internal variability of 20-year means. However, if there are large differences between the radiative balance of UKESM-HI and UKESM-LO this would imply distinct final tunings for each, with possible implications for traceability between the two.

The reason for the radiation bias is mainly a cold bias in the Northern Hemisphere that is directly related to the overly large sea ice area. Fig. 19 displays the strong cold bias of N96-ORCA1 in the northern high-latitude troposphere. At the same time, the warm bias of N216-ORCA025 over Antarctica is greatly alleviated in N96-ORCA1. Lastly, the standard deviation of near-surface temperature (not

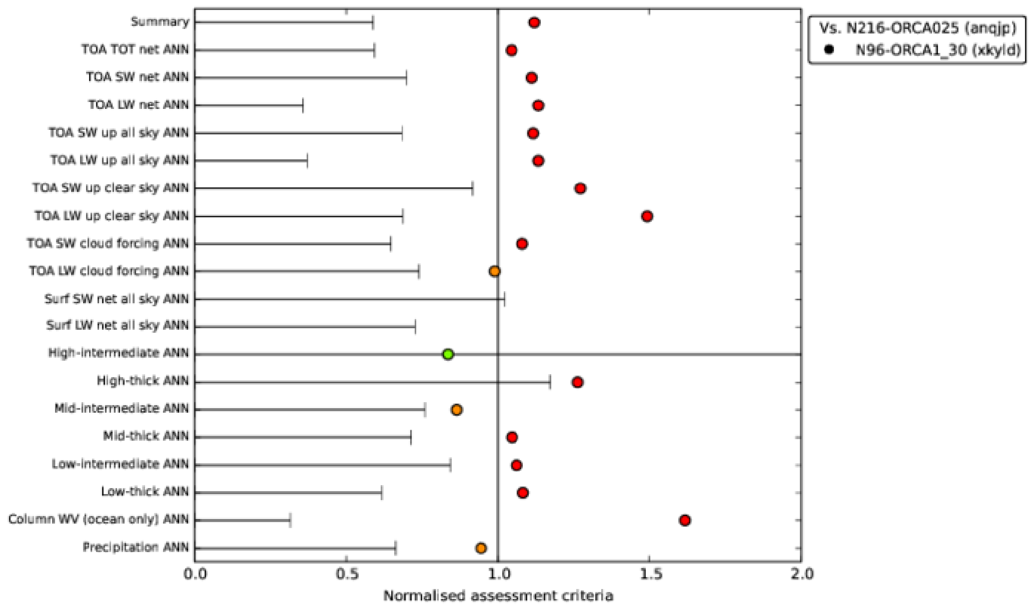

Figure 18: Radiation metrics calculated by AutoAssess for the comparison of N96-ORCA1 (xkyld) with N216-ORCA025 GC2 (anqjp). Based on an average of year 10 to 20 of the runs.

shown) is closer to observations over several land areas in N96-ORCA1.

The cold bias in the Arctic region comes with a positive bias in mean sea level pressure. N96-ORCA1 also has stronger dry biases than N216-ORCA025 in the Arctic lower troposphere and in several regions around the tropopause: the two polar regions and the tropics.

Concerning the zonally averaged zonal wind, the jet stream on the northern hemisphere in winter is located further north than in N216-ORCA025, and thus N96-ORCA1 is in better agreement with observations.

Ocean In N96-ORCA1, the Atlantic Meridional Overturning Circulation (AMOC) at 26N decreases from approx. 17 Sv to 12 Sv over the first thirty years of the run (Fig. 20) and then stays at that value. This contrasts with N216-ORCA025 where the AMOC is stable, with some interannual and decadal variability, between approx. 16 Sv and 18 Sv. While an AMOC volume transport of 12 Sv is a small value compared to observations, it is often found in ocean climate models of this resolution (*Danabasoglu, G., et al.*, 2014).

Table 1: Globally averaged radiation budget at the top of atmosphere (TOA) for N96-ORCA1 (xkyld) and N216-ORCA025 (anqjp), averaged over year 10 to 20 of the two runs. The observed value is from CERES EBAF.

Quantity (W/m^2)	Observed	N96-ORCA1	N216-ORCA025
Net downward radiation	0.5	-0.16	1.67
Incoming shortwave radiation	340.2	340.39	340.39
Outgoing shortwave radiation	100	103.94	99.68
Outgoing longwave radiation	240	236.61	239.04
Planetary albedo (dimensionless)	0.294	0.3054	0.2929

The sea surface temperature field (Fig. 21) mirrors the temperature bias seen in the atmosphere. While the UM's longstanding Southern Hemisphere warm bias is still present (cf. *Williams et al.* (2015); a PEG team is addressing this issue), it is smaller than in N216-ORCA025 (by roughly 1°C in Southern Ocean SST). Fig. 22 shows the SST bias in N216-ORCA025. In the northern hemisphere, particularly at subpolar latitudes, the strong cold bias of N96-ORCA1 is not present in N216-ORCA025.

The question is to what extent the cold SST bias in the North Atlantic is related to the weaker AMOC in N96-ORCA1. An earlier version of N96-ORCA1 (aljyr, GA4.0 instead of the current GA6.0) shows a rather similar decrease of the AMOC at 26N, while the AMOC maximum (irrespective of the latitude) is at roughly 16 Sv. This older version of N96-ORCA1 has smaller SST biases both in the Southern Ocean (positive) and in the North Atlantic (negative). In the present version of N96-ORCA1, we speculate that the parameters for sub-gridscale oceanic mixing are not optimal for a resolution of 1°, affecting the simulated mixed layer depth and leading to a lack of convective overturning in the northern North Atlantic. More analyses will be conducted to improve on this with an aim to both increase the AMOC strength and reduce the North Atlantic cold SST bias.

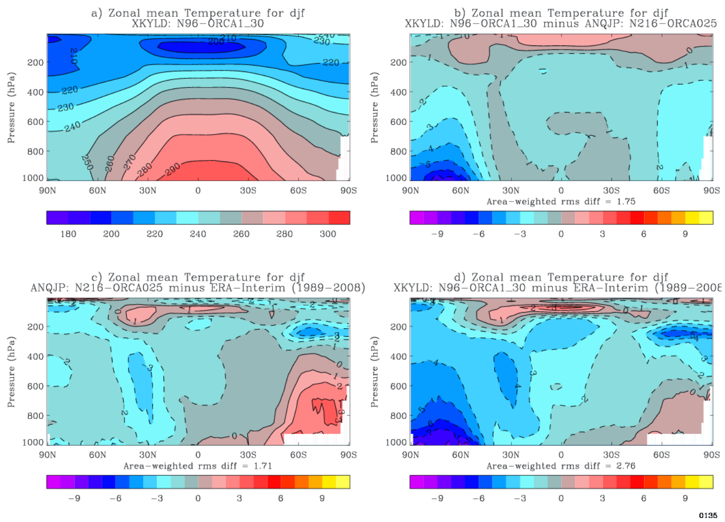

Figure 19: Zonal mean temperature (DJF) in the two model runs N96-ORCA1 (xkyld) and N216-ORCA025 GC2 (anqjp) as well as in observations (ERA40). Shown are averages over years 10 to 20 of the runs.

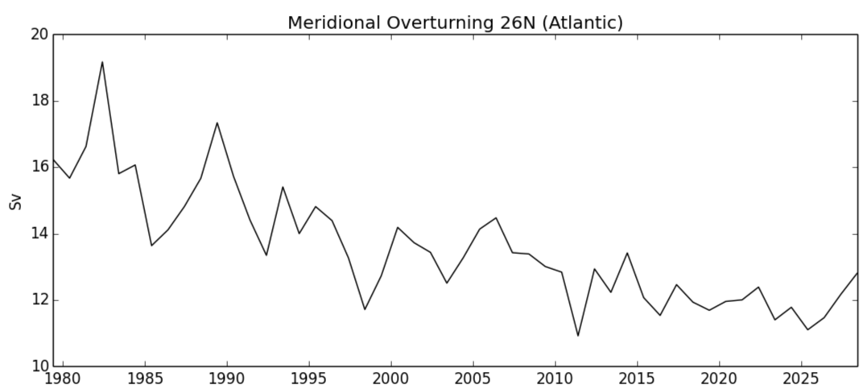

Figure 20: Time series of the AMOC at 26N in N96-ORCA1 (xkyld).

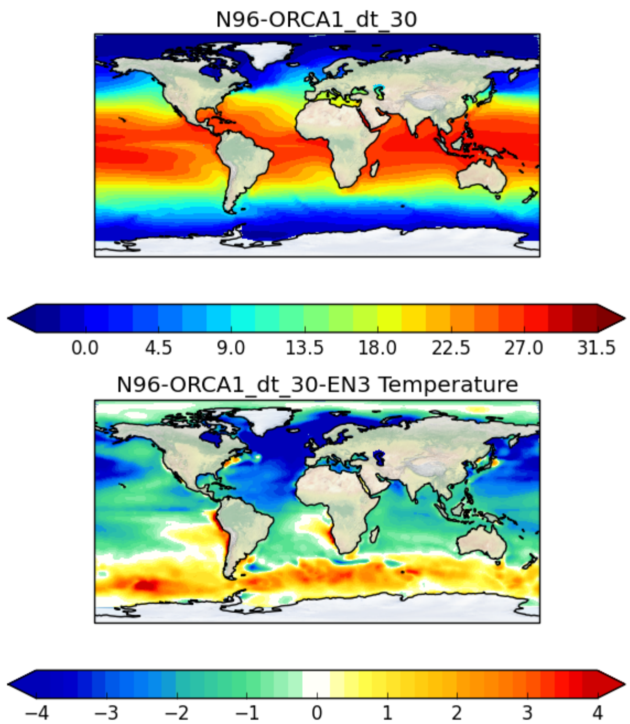

Figure 21: Upper panel: Annual mean sea surface temperature (in °C) in N96-ORCA1 (xkyld), averaged over the years 10 to 20. The bias in comparison with observations (in °C) is shown in the lower panel.

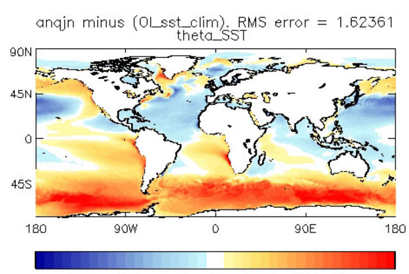

Figure 22: Bias of annual mean sea surface temperature (in °C) in N216-ORCA025 (anqjp), averaged over the years 10 to 20.

Sea ice In the Arctic Ocean, N96-ORCA1 has an approx. 55% larger sea ice area than N216-ORCA025. In this regard, the performance of N96-ORCA1 is worse that that of N216-ORCA025, since in the latter the sea ice area in the Arctic is in line with observations (HadISST 1980-1999).

In the Southern Ocean however, N96-ORCA1 performs better than N216-ORCA025. In N96-ORCA1, the southern hemisphere sea ice area in the austral winter is roughly 80% larger than in N216-ORCA025. In austral summer the northern hemisphere ice area in N216-ORCA025 is extremely small: 126,000 km^2, or less than 5% of the observed value. N96-ORCA1, by contrast, has an ice area of 1.1 million km^2, or just under 40% of the observed value. These differences in sea ice area are strongly linked to temperature biases in both high-latitude regions.

Fig. 23 shows time series of the sea ice extent. While sea ice area is defined as the product of grid cell area and sea ice concentration, the sea ice extent is defined as the total area of all grid cells with a sea ice concentration greater than 0.15.

The circulation of the sea ice in the Arctic in Summer is cyclonic in observations (SSMI). N216-ORCA025 shows this cyclonic circulation, but it is anticyclonic in N96-ORCA1. In winter, both models and observations show an anticyclonic circulation. In the Southern Ocean, the sea ice velocity field in austral Winter is too zonal and cyclonic in both models in comparison to observations.

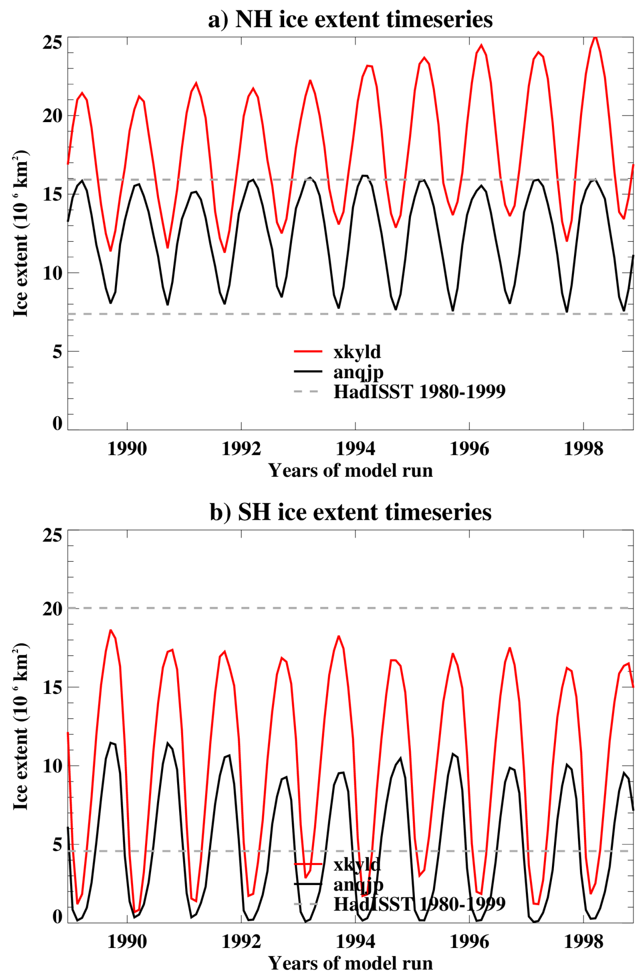

Figure 23: Time series of the sea ice extent on both hemispheres in summer and winter in N96-ORCA1 (xkyld) and N216-ORCA025 GC2 (anqjp). Dashed lines indicate the range of observations (HadISST). Shown are years 10 to 20 of the runs.

4.2.2 N96-ORCA1 vs. N96-ORCA025 GC2

The results from this comparison are roughly similar to the comparison with N216-ORCA025, implying that this change in ocean resolution, rather than the jump from N216 to N96 in the atmosphere, has the more significant impact on the large-scale climate metrics we look at here. The detailed analyses can be found on the collaboration twiki.

4.2.3 N96-ORCA1: dt=30 min vs. dt=20 min

This pair of runs was analysed in order to assess the effect of moving to a longer time step in the atmosphere. At the same time, the radiation time step was extended to 3 h (see sec. 2.2 for details).

Atmosphere We focus on differences from end of the simulations, between years 31 and 50. The AutoAssess metrics appear very similar in the two model versions, apart from some biases in temperature and radiation.

Mirroring differences in sea ice extent, N96-ORCA1 (30 min) shows a larger cold bias in the Arctic troposphere than N96-ORCA1 (20 min). Around the Antarctic tropopause the cold bias in the 20 min version is alleviated in the 30 min version. Likewise, the warm bias in the lower Antarctic troposphere is also alleviated in the 30 min version (Fig. 24).

The upper atmosphere winds (200 hPa and 700 hPa) are closer to observations in the 30 min version, especially over the Tropics and the Southern Hemisphere.

The net downward radiation at TOA, in the global average, is quite different in the two versions: +0.15 W/m^2 in N96-ORCA1 (30 min) and +1.07 W/m^2 in N96-ORCA1 (20 min). This is not surprising, given the longer timestep in the radiation scheme. The radiation bias appears to come mainly from the regions of the increased sea ice cover: the Nordic Seas in clear sky outgoing longwave radiation (DJF; Fig. 25), and the Southern Ocean in clear sky outgoing shortwave radiation at TOA (DJF). In other regions the differences between the two runs are small.

Ocean In N96-ORCA1 (20 min) the AMOC at 26°N declines from 16 Sv to 10 Sv over the 50 years of the run. N96-ORCA1 (30 min) shows an improvement with a smaller decline; the AMOC at 26°N stabilizes at 12 Sv (Fig. 20).

There are considerable drifts of temperature and salinity in the 20 min version (these have not yet been analysed in the 30 min configuration). The global SST

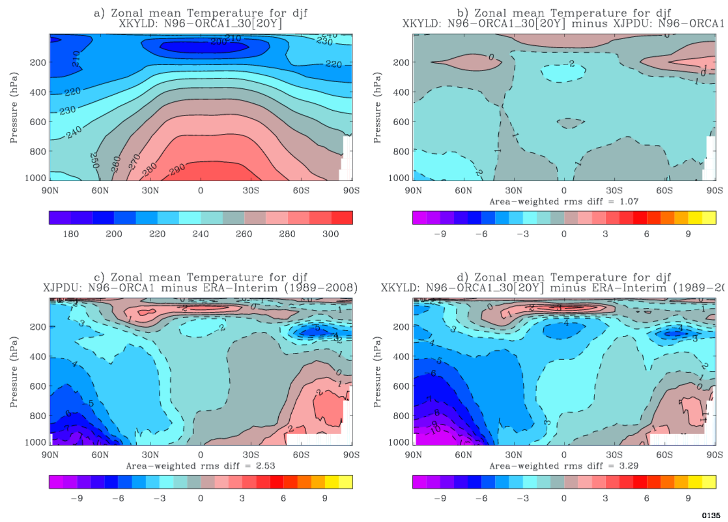

Figure 24: Zonal mean temperature (DJF) in the two model runs N96-ORCA1 (30 min) and N96-ORCA1 (20 min) as well as in observations (ERA40). Shown are averages over years 31 to 50 of the model runs.

decreases by approx. 0.4°C over the first 10 years of the run and then stabilizes around 18°C. This cooling happens in the Northern Hemisphere, but not in the Tropics or the Southern Hemisphere. At the same time there is a net heat flux into the ocean of approx. 1.5 W/m^2. This heat flux is established after about four years into the model run and is then stable, with some interannual variability.

Sea surface salinity (SSS) falls by approx. 0.25 psu over the 50 years in the global mean. This trend continues throughout the run and only decreases slightly in magnitude. As opposed to SST, this freshening happens everywhere, with the Arctic the only exception. It has not been assessed yet to what extent these drift can be attributed to the fact that the ocean was started from climatology, and not from a spun-up state. In addition, once a configuration is decided, the parameterisation of iceberg freshwater flux which compensates the build up of ice on land will be re-calibrated, which will address the salinity drift.

Sea ice In both hemispheres sea ice extent is larger in N96-ORCA1 (30 min) than in N96-ORCA1 (20 min). This means that in the 30 min version the positive area bias in the Arctic is exacerbated while the negative area bias in the Southern Ocean is alleviated. The differences between the 30 and 20 min versions are how-

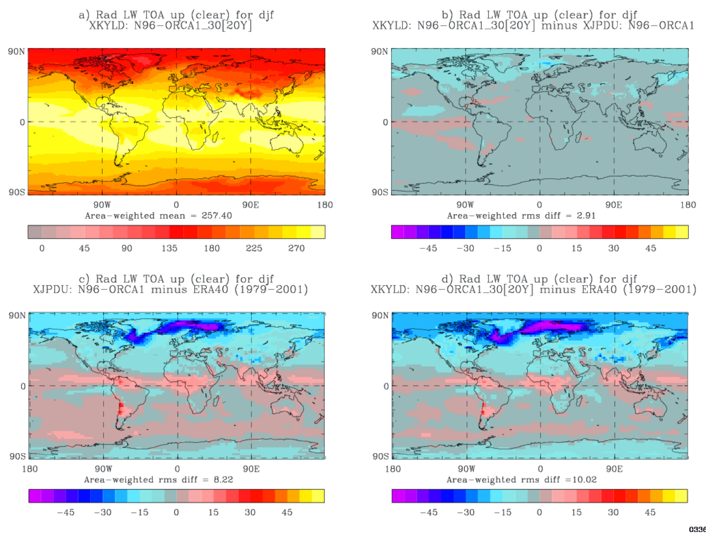

Figure 25: Clear sky outgoing longwave radiation (DJF) in the two model runs N96-ORCA1 (30 min) and N96-ORCA1 (20 min) as well as in observations (ERA40). Shown are averages over years 31 to 50 of the model runs.

ever clearly smaller than those between any of the N96-ORCA1 runs and coupled configurations with ORCA025. Both the 30 min and the 20 min configurations circulate the sea ice in the wrong direction during Arctic summer (Fig. 26).

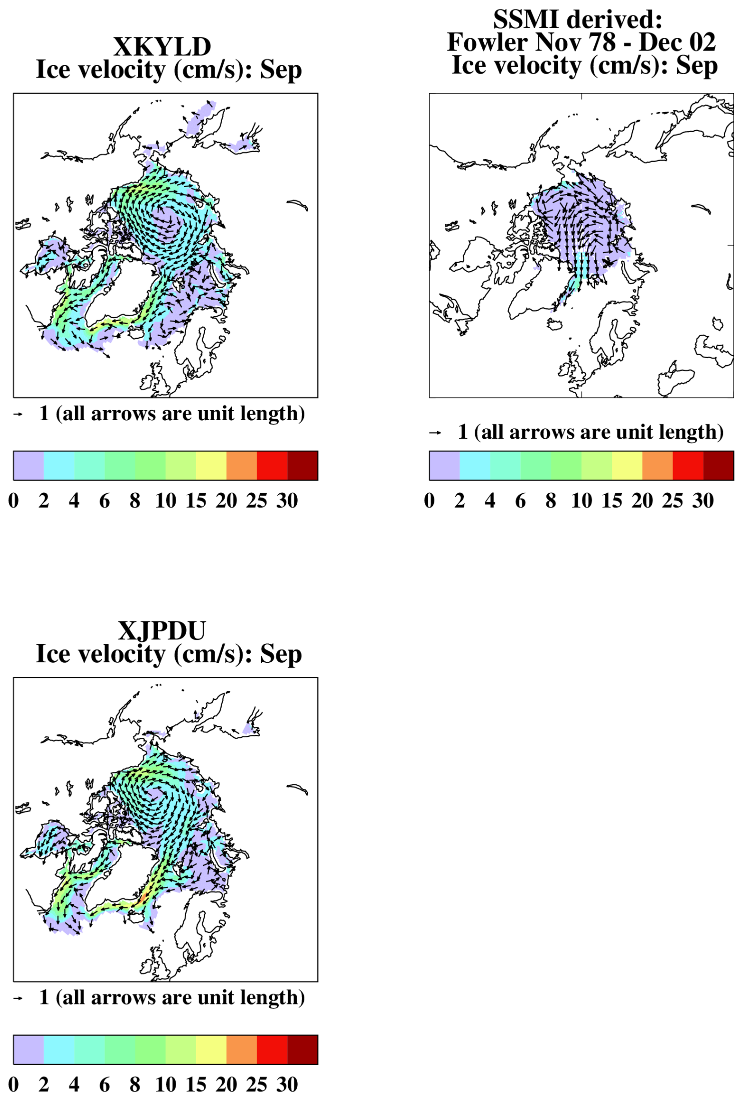

Figure 26: Sea ice velocities in the Arctic in summer in the two model runs (left) and in observations (right). Shown are averages over year 10 to 20 of the model runs.

5 Available data and documentation

Detailed documentation of the evaluation of the test configurations is available at http://collab.metoffice.gov.uk/twiki/bin/view/Project/UkesmLoResolution .

Further details about the test configurations, along with the location of the model output on MASS, can be found on the same website, following the links under the header "UKESM-LO test configurations". The namelist and FPP keys configuration file for NEMO are available in the Attachments on that website.

The model output of the most important runs discussed in this report is available to the community on the JASMIN facility. We invite the community to perform their own analyses of these data.

6 Conclusions

This report evaluates several test configurations of the UM at the N48 and N96 resolutions with regards to their suitability for UKESM-LO. Both computational and scientific performance are taken into account. Differences between coupled configurations at ORCA1 and ORCA025 resolutions are also considered.

While being computationally faster, the current UM version at N48 poses several problems. To begin with the technical viewpoint, the N48 resolution is no longer supported by the Met Office Hadley Centre. This caused major practical difficulties in simply producing a limited set of ancillaries (boundary condition files) necessary for this report, and implies very significant technical effort from the UKESM team on an ongoing basis were this configuration to be adopted for UKESM-LO. The evaluation of atmosphere-only configurations of the N48 model shows that the significant biases found in a version of the model with New Dynamics (many of which are related to weak transient eddy kinetic energy, TEKE) have not been improved by the introduction of the ENDgame dynamics. In comparison with N48 using old dynamics (HadAM3, the atmosphere-only version of HadCM3), N48 ENDgame did not perform markedly better. It is known from earlier work (*Stratton*, 2004; *Sanchez et al.*, 2013) that in terms of TEKE, the performance of N48 ENDgame is worse than in HadAM3 due to the inherent diffusive properties of the semi-Lagrangian advection scheme. Some methods to improve TEKE in N48 with New Dynamics are not available for the current ENDgame dynamical core, and those methods we could test showed no improvement. Furthermore there were major differences in TEKE and a substantial number of other metrics between N48 ENDgame and N96 ENDgame (i.e. HadGEM3 at the GC2 release), with N96 clearly faring better and manifestly producing performance metrics closer to N216 (UKESM-HI).

Hence we argue that there are three factors for not continuing to develop N48 ENDgame: (i) the difficulty in producing N48 ENDgame coupled ancillary files (and this problem would propagate into all future ancillary files for different applications of UKESM-LO, e.g. in CMIP6, causing a drain on core group resources which could otherwise be devoted to improving model performance and understanding traceability); (ii) N48 ENDgame displays a clearly worse scientific performance than N96 ENDgame, and the known problems in N48 New Dynamics have not improved with ENDgame; (iii) N96 ENDgame coupled to ORCA1 shows a scientific performance that is close to N216-ORCA025—N96 looks closer to N216 than it does to N48. Therefore it is more likely that N96 has traceable performance to N216. This can be exploited productively in all the planned applications of UKESM. Improving the scientific performance of N48 and making it traceable to UKESM-HI (N216) would require a large effort from the UKESM core group with no guarantee of success.

As a result, we recommend very clearly that N96-ORCA1 be the resolution for UKESM-LO. N96-ORCA1 compares very well with two of the HadGEM3 GC2 configurations, N216-ORCA025 and N96-ORCA025, of which N216-ORCA025 is the resolution of UKESM-HI. Work to improve the cold bias in the North Atlantic (on top of other ongoing Met Office developments) is warranted and further efforts of the UKESM core group will be dedicated to increasing the computational efficiency of N96-ORCA1 and addressing some of the primary climate biases.

As a first step, we have increased the time step length of the model dynamics and the radiation scheme, which appears to not significantly degrade the scientific performance at N96 resolution. A further lengthening of these time steps, along with the time step in NEMO, is currently being investigated, with some promising indications. While N96 obviously needs more computational resources than N48, we re-emphasize that both UKESM-LO and UKESM-HI are expected to be used until about 2022. While N96-ORCA1 may seem an ambitious model for UKESM1-LO now, by 2020 this is less likely to be the case in the view of increasing availability of computing resources.

Acknowledgments

We thank David Walters, Claudio Sanchez, Pat Hyder, Dan Copsey and several other staff of the Met Office Hadley Centre for their support with model develop-

ment and evaluation. An earlier report by Simon Read and Alistair Sellar was a useful source for a description of the AutoAssess metrics plots.

References

Danabasoglu, G., et al. (2014), North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states, *Ocean Modelling*, *73*, 76–107, doi:10.1016/j.ocemod.2013.10.005.

Gordon, C., C. Cooper, C. Senior, H. Banks, J. Gregory, T. Johns, J. Mitchell, and R. Wood (2000), The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments, *Clim. Dyn.*, *16*, 147–168.

Manners, J., J.-C. Thelen, J. Petch, P. Hill, and J. Edwards (2009), Two fast radiative transfer methods to improve the temporal sampling of clouds in numerical weather prediction and climate models, *Quart. J. Roy. Met. Soc.*, *135*, 457–468, doi:10.1002/qj.385.

Sanchez, C., G. Shutts, K. Williams, T. Hinton, R. McDonald, C. Senior, and N. Wood (2013), Towards the development of a traceable model hierarchy: Investigation of dynamical limitations at low resolution and possible solutions, *Quart. J. Roy. Met. Soc.*, *139*, 75–84, doi:10.1002/qj.1971.

Sanchez, C., K. Williams, and M. Collins (2015), Improved stochastic physics schemes for global weather and climate models, *Quart. J. Roy. Met. Soc.*, *submitted*.

Stratton, R. A. (2004), The energetics of a semi-Lagrangian dynamical core, Hadley Centre Technical Note No. 52, Met Office, UK.

Williams, K., D. Walters, S. Milton, and C. Senior (2013), Strategy for evaluation of MetUM global coupled configurations, Technical Report, Met Office, UK.

Williams, K. D., et al. (2015), The Met Office Global Coupled model 2.0 (GC2) configuration, *Geosci. Model Dev. Discuss.*, *8*, 521–565, doi:10.5194/gmdd-8-521-2015.