Response to reviewer's comments

We thank both reviewers for their useful and insightful comments. Our responses follow.

Reviewer 1

General comments

The infrastructure described in this paper is for the physical model only of HadGEM3 (HadGEM3-AO) which will in future be the underpinning model for the Earth System version of the model (HadGEM3-ES). We propose to make this clear in section 1 of a revised manuscript. To make this distinction clearer we could use HadGEM3-AO throughout (a naming convention consistent with HadGEM2).

The land model is run as part of the atmosphere model (i.e., in the same executable). We propose to include discussion of this in section 2 of a revised manuscript. However, since we have not changed the coupling between land and atmosphere moving from HadGEM1/2 to HadGEM3 we do not propose to discuss this in greater detail. In the 'Summary and Discussion' we propose to add brief discussion of the implication of the development of JULES as a sub model (which will make MOSES-II redundant) and how the method of building a single executable from the UM and JULES source codes (as in the NEMO-CICE case) can be applied.

We appreciate the reviewer's dilemma regarding whether this is a science document or a technical note. However, it is our understanding that this paper fits the remit of GMD. While some of the details listed may seem obvious, it is the authors' experiences that model exchanges are not well-documented and this can make developments difficult (eg, including multilayer thermodynamics in the sea ice model of this configuration or development of a generic surface flux interface for NEMO). We have found that there is real value in having model exchanges well-documented. Indeed we have already used the draft of this paper to enable collaborative work on similar models to take place.

We agree with the reviewer's comments regarding the detail of SCRIP and we propose reducing the detail in sections 4.1 and 4.2 in a revised manuscript. We can also reference a Hadley Centre technical note for more technical details.

We would be keen to maintain discussion in the paper regarding the implementation of the model on different architectures as it was key to some of the decisions we made and ones we will face in the future.

The model development strategy for HadGEM3-AO is different to previous model development projects (HadCM3, HadGEM1 and HadGEM2). For HadGEM3-AO we are not intending to work towards one fixed version of the model. Instead, model development will continue in cycles with updated versions of the model released in a similar manner to the operational weather forecast model development cycle. Seasonal forecast models are continually upgraded on an operational cycle and it is for that reason that the model can be used for seasonal forecasting. However, we have not reached a stage in the development cycle of HadGEM3-AO where we are ready to commit to a version that can be used for long climate runs in the style of IPCC. We propose to explain this more clearly in the sixth paragraph of section 1.

Specific comments

- Since there is no mention of how water and energy are routed from the atmosphere to ocean via land it is hard to evaluate the statements on heat and freshwater conservation.

The statements about heat and freshwater conservation refer specifically to conservation across the OASIS coupler between the atmosphere (including land surface and river routing components) and the ocean-ice system. We propose to make this clear in section 2.

- P1862L20: The pole problem is not a "new problem of going to high resolution" it is a problem at historically low resolution because the meridians converge and filtering is already required for lat-lon grids.

We agree that the 'pole problem' is not a new one at high resolution, but as resolution increases, technical constraints on the filtering have previously been found to make the filtering proportionately more expensive and to limit scalability (which is an increasingly important consideration at higher resolution). In addition, the removal of the need for a polar island is highly desirable. We propose to reword the text slightly to clarify these points.

- P1863L25: Again there is no mention of a land model in the system and how that is to be coupled through OASIS. Some readers may not know what "AMIP-style" means so an explanation about prescribing SSTs would be useful.

We propose to add the following text after paragraph 3, section 2: "The land surface hydrology scheme used is the Met Office Surface Exchange Scheme (MOSES-II, Cox et al, 1999 need to add to reference list) which is the same as that used by HadGEM1 and HadGEM2 but with updated soil properties and soil hydraulics. The river routing scheme used is the Total Runoff Integrating Pathways scheme (TRIP, Oki and Sud, 1998 need to add to reference list) also largely unchanged from HadGEM1 and HadGEM2. Both the land surface hydrology scheme and the river routing scheme run as subcomponents of the atmospheric sub model as they are included in the same executable and code base as the rest of the atmospheric code. Because of this all future references to the coupling between the atmospheric sub model and the ocean and sea-ice sub models include the land surface and rivers as part of the atmospheric sub model." We also propose to change "AMIP style atmosphere models" to "prescribed SST atmosphere only models".

- P1864L7: "In particular…" I am not sure what this means, is this referring to HadGEM3 or previous GEMs?

See general response. This refers to previous models not HadGEM3. We suggest rewording to reflect this.

- Fig 1: Again I don't know what to think about this figure in the absence of information about the budgets in the land (e.g. river routing etc). If this already unchanged from a self contained formulation within the UM then it needs to be said explicitly.

See general comment above. We propose to explain in section 2 that the land and river scheme are embedded within the atmosphere model.

- P1865L10: Is freshwater formally conserved i.e. through changes in mass in the ocean that impact the equations of motion? I am pretty sure this is not the case and the conservations are via implied salt changes from freshwater extraction or addition processes.

We propose to clarify this in a revised manuscript (section 2 re figure 1b; section 4.2 re figure 6b). The model uses a linear free surface and freshwater fluxes are converted to salt fluxes for the ocean. Ocean salt content will be conserved within the approximations of the free surface.

- P1865L29: Are aerosol fluxes across components dealt with given that aerosols are in the UM? i.e. flux of aerosols onto sea-ice causing darkening and flux of dust from the land to the atmosphere. This is important from an earth system perspective.

Since this is not the full Earth System model, aerosol fluxes are not passed across the interface.

- P1866L19: I am pretty sure that the north pole singularity has not been stretched as described in the text. Fig 1b indicates that singularity has been move to the middle of Canada and Siberia to avoid pole convergence problems.

We will change 'stretched into' to 'replaced by' (in both section 2 and Appendix C) of the revised version of this manuscript.

- P1866L29: Is surface sea-ice temperature really calculated in the atmosphere? That makes no sense as TS is mainly a function of the sea-ice column thermodynamic calculation. Do you really mean surface air temperature?

The surface sea-ice temperature is calculated in the atmosphere model using zero-layer thermodynamics. The conductive flux through the ice is also calculated implicitly with the surface temperature in the atmosphere model. The surface heat flux and the conductive flux are then passed to the sea ice model where they are used to grow or melt the ice. Therefore, the full sea ice column temperature calculation takes place in the atmosphere model but uses the zero layer assumption. The given reference of McLaren et al. (2006) explains this in more detail. We suggest a text change to clarify this: "Surface sea ice temperature, atmosphere to ice fluxes *and the conductive heat flux through the ice* are calculated in the atmosphere component…."

- I am not sure all the appendices are required for all the mapped variables between components as this is these are well established variables and practices, unless this is meant to serve as model technical documentation.

We think that the appendices act as a useful understanding of the model for future development/model intercomparisons as described above in the general comments.

- P1870L8: "takes PLACE …"

This is a typo and will be corrected in a revised manuscript.

- 4.1 : Is anything specified here specific to HadGEM3 and not just standard SCRIP techniques as it seems references would suffice?

We agree with the reviewer and would propose reducing sections 4.1 and 4.2 accordingly.

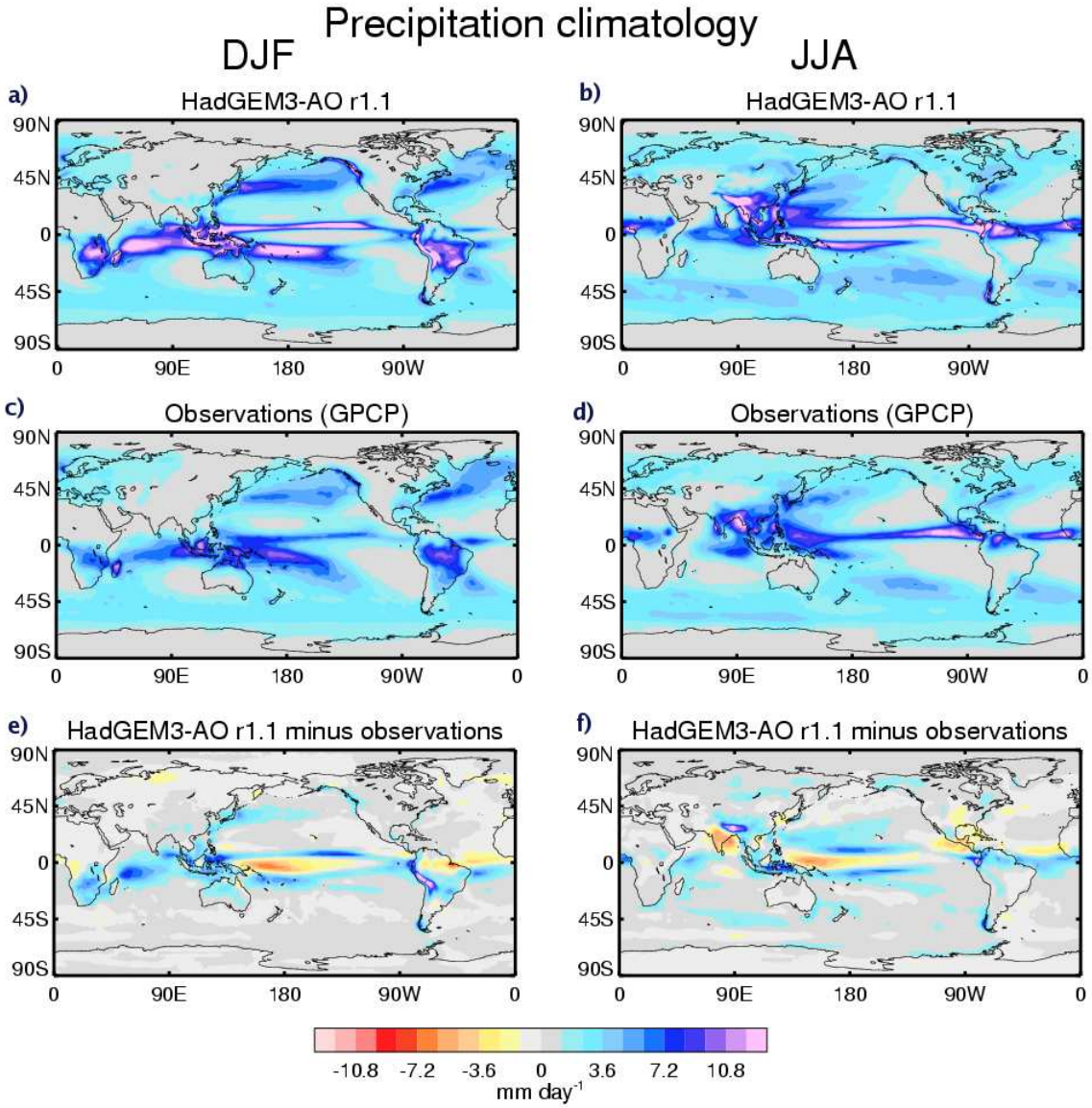
- 5.1 – I am not sure this section is really required as it appears a very transitory result.

We propose to keep 5.1 as it was relevant to some of the development choices and how the model may be implemented differently in the future.

- Fig 7f is wrong and looks like a copy of 7e. Also how long is the averaging period from the model.

We thank the reviewer for spotting this. There was an error in the plotting routine which resulted in one of the plots being duplicated. This has been corrected (see

below). The averaging period from the model is 30 years and this information will be added to the plot description.

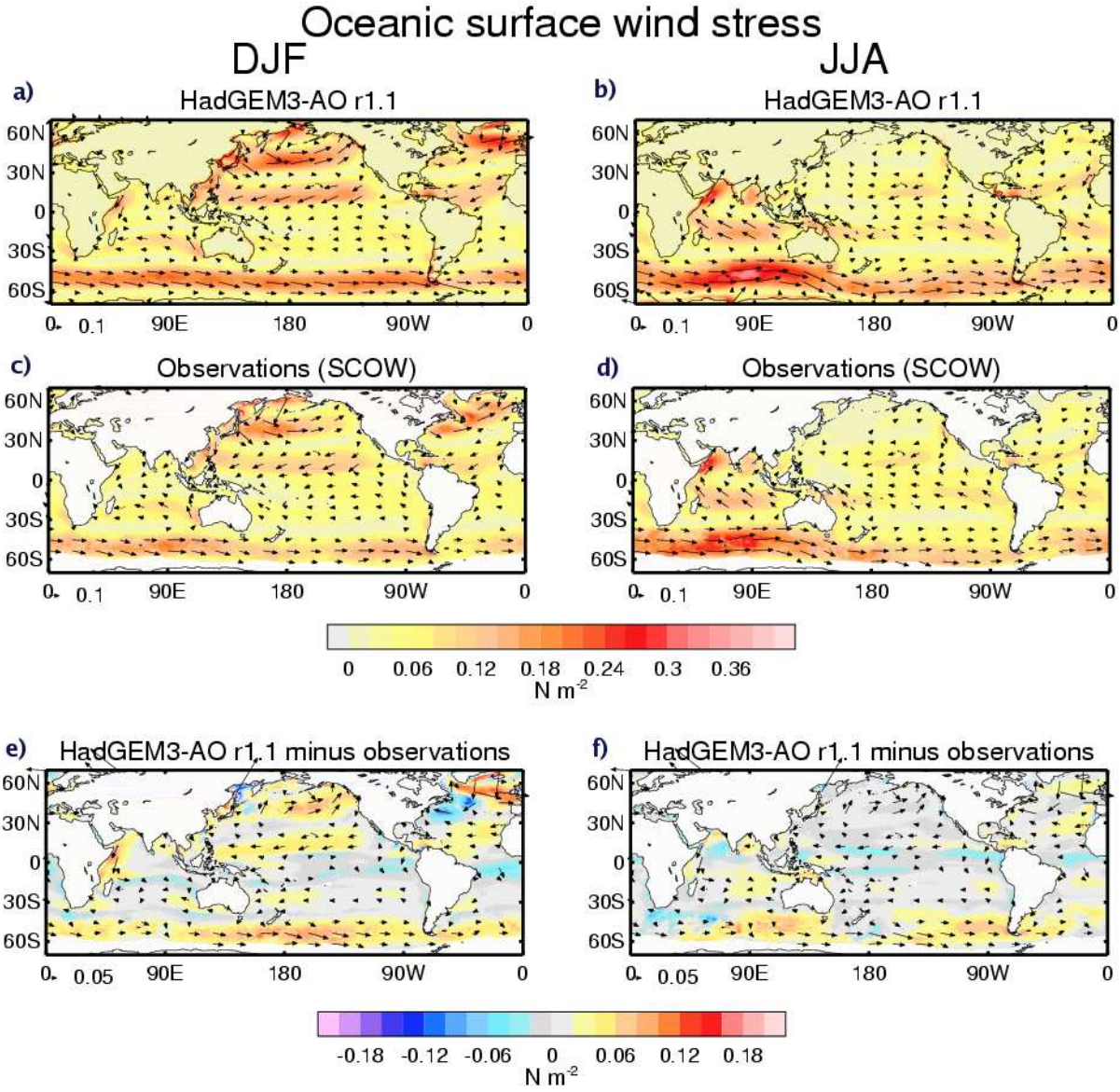


- P1878: Some of the descriptions of the model performance here appear to set the bar exceedingly low e.g., the storm tracks are in the right place. A more appropriate comparison, even at this early development phase, would be against HadGEM2 or maybe a HadGEM3 AMIP experiment where flux mapping issues between components shouldn't be a problem. A comparison against HadGEM1 is discussed but no figures are shown I think it would be good to include those figures.

As described in the general responses, HadGEM3 is continually under development. In this paper, we aim to show that the model performance compares reasonably with observations rather than how it benchmarks against HadGEM2.

- Given that the paper primarily seeks to validate the coupling it would be more appropriate to show surface stress rather than u850mb.
- P1879L10: This comment would also be more substantiated by showing surface stresses instead of u850mb.

We attach the plot of the oceanic surface wind stresses relative to the Scatterometer Climatology of Ocean Winds (SCOW). This plot shows similar biases to the 850hPa wind biases although the biases on the equator have reduced and the biases in the Southern Ocean have increased a little. The surface wind stress bias plots are only available over the oceans so these plots contain less information than the 850hPa wind plots.



- P1879: Again I think the same applies here as to the atmospheric diagnostics. What are the changes compared to HadGEM1/2, given that there hints that they are pretty good in the text?

See response above.

- P1880L20: One could also argue that the SST bias would be improved if the sea-ice extent was better.

This is a valid point, however, further investigation into the SST bias has found it to be related to a bias in the atmospheric forcing and the vertical mixing scheme used in NEMO and not directly linked to the low sea-ice extent.

Reviewer 2

1. The MetOffice releases code updates for the UM on a fairly regular basis. These have designation of form UM6.X, UM7.X , etc. It would be helpful if the authors were to note the code updates from which the HadGEM3 revision 1.1 featured here can be configured. Are there any code updates for which it is the default configuration?

HadGEM3 r1.1 was originally setup using UM7.1. Since then it has been upgraded to UM7.2, UM7.4 and UM7.6 but still using the same physics as that described in this paper. We could add this comment to the paper.

2. The Met Office performs standardised assessments on its latest version of the atmospheric and coupled model routinely every 6 months ("March 2009 assessment", "Sept 2009 assessment", etc.). It would be helpful if the authors were to note the assessment corresponding to the HadGEM3 r1.1 as featured here.

We will add a footnote indicating this is essentially the version from the March 2009 assessment (but with a couple of minor changes to make it more suitable for seasonal use).

3. Apparent slight inconsistency of the terminology: p4, lines 15-16: ". . . r1.1 . . ..is not, in principal, fixed to any particular resolution." P5, lines 29 – p6, l1: ". . . r1.1 . . . has a horizontal resolution of . . .." Suggest reword to clarify.

We would reword this for clarity.

4. p7, line 10: Useful to mention up front here that the T grid for the ocean and sea ice models align (though this is discussed later).

We would make this suggested change.

5. p7, line 14: "need for efficient load balancing". Should mention here that this is only for the (vector) platform used at the time. As discussed later, this is not so clear for massively parallel systems now widely used.

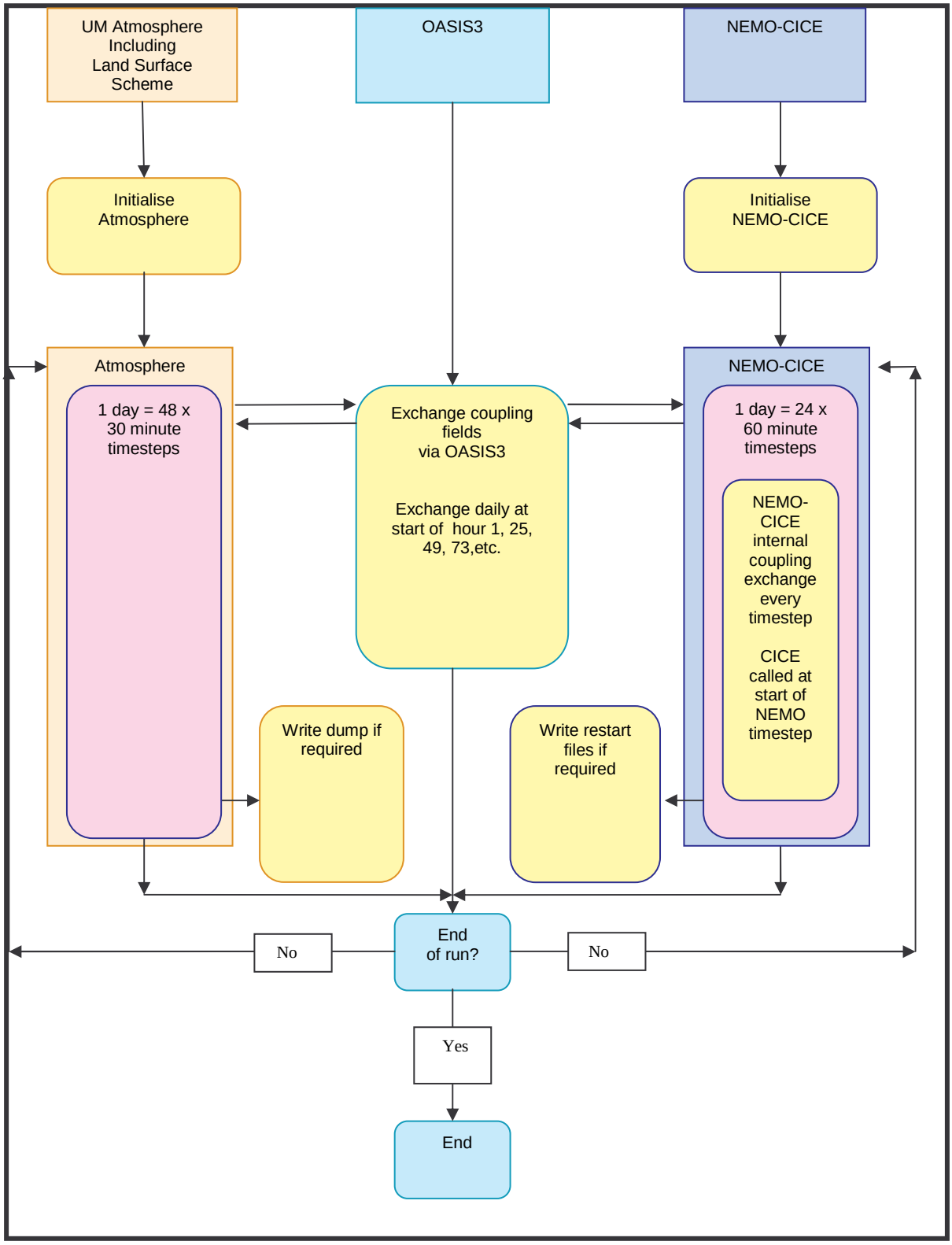
We propose to change this sentence to read: "The need for efficient load balancing, the fact that NEMO and CICE employ the same grid and the fact that the model was developed on a vector machine employing relatively few CPUs led to the decision that NEMO and CICE,.....

6. The use of a B grid for the sea ice and a C grid for the ocean means that the land sea boundaries for the U and V points for the ocean and sea ice will not quite match and so an interpolation is required (p 15). Do any modelling issues arise as a result (e.g., effect on coastal polynyas)?

While the reviewer is correct to say that there is a potential for problems around the coastline, we have not seen any evidence of this. This difference in the grids also occurs when NEMO is run with LIM2 and there are no documented problems with this model. We also mention in section 2 about avoiding single grid-box channels for the land. We could also argue that given that the wind stress (the most important dynamic forcing) is on yet another grid at lower resolution, that the differences in the ocean/ice grids are less important.

7. It would be useful to show a schematic clearly illustrating the time-stepping arrangement, with time axis along the page and showing the timing of the ocean, sea ice and atmospheric steps and the coupling times.

We do not think an extra figure is required but propose to modify Figure 3 (attached below) slightly to make the timing of coupling exchanges slightly clearer, and to complement the information already in section 3 and figures 4 and 5.



8. Suggest explain why CICE was chosen in preference to LIM, the existing NEMO sea ice model, given that the choice introduces the additional complexity of the differing grids (B and C).

The decision to use CICE rather than LIM was made for two main reasons: 1. At the start of the HadGEM3 project CICE was more scientifically advanced than LIM as it used multiple thickness categories; 2. We had incorporated several CICE components into HadGEM1 so the model science was maintained by moving to the CICE model. This would be explained in a revised manuscript.

9. p 13 line 13: Define _j, _j.
This would be defined in a revised manuscript.

10. p 13, lines 25-25: Note whether any issues are likely to arise as a result of possible non-conservation of momentum.
We do not believe that there are any serious issues arising from the non-conservation of momentum (since this was also used in HadGEM1 and HadGEM2).

11. p 14, line 1: Reword to “same scalar grid”.
This would be done in a revised manuscript.

12. p. 16, lines 3-7: The change in global ocean heat content in Fig. 6a actually appears to differ quite significantly from the global net ocean surface heat flux. Why is this? This would seem to suggest significant non-conservation issue somewhere in the ocean model. Similarly, in Fig. 6b, why does the change in global ocean freshwater content not exactly match year-by-year the global net ocean surface freshwater flux?
The apparent discrepancies are partly because the ocean heat content and freshwater volumes, which were used to calculate the “rate of change” components of figures 6a and 6b, are annual means, rather than instantaneous ones (for reasons of practical availability only). Since the fluxes are annual means, we should really be comparing these with the latter. The purpose of figure 6 is mainly to illustrate conservation across the OASIS coupler and, in addition, that the variation of integrated heat/freshwater contents follows that of the driving fluxes.

13. Suggest explain why 1980s forcings are used at the same time as 2000s initialisations are used. The radiative forcing 1980s to 2000s is somewhat different.
Unfortunately, while less than ideal, such inconsistencies are common place in climate modelling. For example, “pre-industrial control experiments” are often initialised with present day conditions but then forced with pre-industrial radiative forcing. We propose to add some discussion to a revised manuscript around this point.

14. p 18, lines 7-9: “simulates locations of the storm tracks correctly”. Reword – much more analysis would need to be done to make this claim.
We propose rewording to “The model also generates fairly realistic mid-latitude precipitation, with local precipitation maxima located off the eastern coast of the USA and to the east of Japan.”

15. Fig. 10: Do the calculated poleward transports include all components, including diffusive transports?
Yes, all components are included. We will add this to the caption.

16. Is there a maximum allowed depth set for the sea ice? If so, then this should be noted.
There is no maximum allowed sea ice depth.

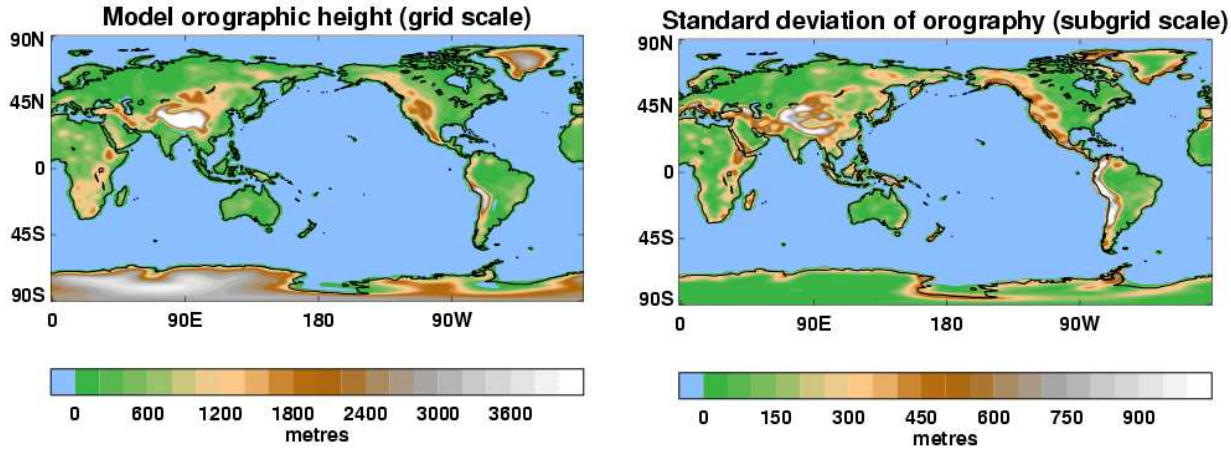
17. The Appendices feature a detailed comparison between HadGEM1 and HadGEM3 r1.1. This comparison is useful. However, even more useful would be a comparison between HadGEM2 and HadGEM3 r1.1, especially since HadGEM2 forms the immediate predecessor to HadGEM3, is the basis for the Met Office contribution to the IPCC AR5/CMIP5 modelling program, and is well documented in a series of technical reports. The utility of this paper would be improved by inclusion of a comparison between HadGEM2 and HadGEM3 r1.1.
HadGEM2 is very similar to HadGEM1 in terms of atmospheric physics (apart from convective detrainment) so it would be quite easy to compare the UM atmosphere between HadGEM3 and HadGEM2 instead of HadGEM1. HadGEM1 was originally chosen as at the time of writing as there was no HadGEM2 documentation so it made sense to compare our model with the older but documented model.

18. p 23, line 18: Claim of “a realistic QBO” would require substantiation. Reword to, e.g., “an improved stratospheric Quasi-biennial Oscillation”. Note that QBO needs spelling out.
This would be reworded in a revised manuscript

19. Section A4. There is little said about the land surface scheme. At least note the scheme used (MOSES?).
This is covered in our responses to reviewer 1’s comments.

20. A figure showing the global model orography would be useful. Also, a little more detail on the steps in the development of the model orography from the original 30” set would be helpful. Was use of envelope orography (to get the mountain ridges to be of realistic height) considered?
A figure of the grid scale orography and sub grid orography is attached below. The Raymond (1988) filter has been used to process the data from the raw 30” dataset using a filter length scale of 6km. Envelope orography was not used as the UM uses a flow blocking parameterization that includes the effects of sub grid scale ridges onto the grid scale flow. More information on the generation of the model orography and the flow blocking parameterization can be found in Webster et al (2003).

Raymond (1988). High-order low-pass implicit tangent filters for use in finite area calculations. Mon. Weather Rev., **116**, 2132-2141
Webster, S., Brown, A.R., Cameron, D.R. and Jones, C.P. (2003). Improving the representation of orography in the Met Office Unified Model. Q.J.R. Meteorol. Soc.,**129**, 1989-2010.



21. p 30, lines 3-5: Presumably this then means that those upper levels of thickness less than 20 m simply were not subject to partial stepping? Also, please specify what is the minimum allowed ocean depth.
22. p 30, line 14: “This is scaled . . .” Do you mean “aht0 was scaled ..”?
23, p 30, line 19: What scheme was used to vary the coefficient in the Gent-McWilliams term.
24. p 31: What, if any, bottom boundary layer/downslope flow scheme is used? If none, then was there any special treatment to improve the simulation of the North Atlantic Deep Water formation?
25. p 32, lines 21-24: The treatment of latitudinal resolution in the Sothern Hemisphere ensuring isotropic grid boxes is notable. Is this treatment carried all the way to the Antarctic coast (through _75_ S), which would result in a large concentration of rows at such latitudes?
(Comments 21-25): These would all be clarified in the revised manuscript. Upper levels can be subject to partial stepping because in that case the minimum of 20m and 0.1*fulldepth will be less than 20m. The minimum allowed ocean depth is 30m. ‘This is scaled’ does refer to aht0. The exact scheme for the GM coefficient calculation isn’t really referenced in the NEMO documentation/code. The model described did not have a BBL scheme or any other special treatment to improve NADW formation. There *are* a relatively large number of rows in the Southern Ocean.

25. p 35, line 13: “albedo being set to 0.61”. What was it in HadGEM1?
The bare ice albedo in HadGEM1 was 0.57. This was increased to 0.61 in HadGEM2 in conjunction with a code correction to the albedo scheme. Suggest new text: “This is the same scheme that was used in HadGEM1 (McLaren et al., 2006) with one exception of the snow-free sea ice albedo *being increased from 0.57* to 0.*61 following a change made in HadGEM2 (Martin et al., 2011).”*

26. p 35, line 23: “in thickness space.” Help the reader by explaining what this means, or reword.
Suggest new text: “The calculated thermodynamic growth or melt rates are then used in the linear remapping scheme of Lipscomb (2001) to transport *the ice between thickness categories.”*

27. p 40, line 3: “allow reproducible restarts”. Please explain just what this means. Does this mean that a model simulation can be broken up into segments of arbitrary length and still yield the same identical bit-wise solution? Please clarify whether the HadGEM3 r1.1 actually has this property, which is highly desirable.
We propose to change the paragraph: "OASIS3 provides a mechanism to allow reproducible restarts…." To: “It was essential for traceability reasons that models could be stopped and restarted at arbitrary times, producing exactly the same results as if they had run continuously for an equivalent period (“bit reproducibility”). OASIS3 provides such a mechanism via use of its own form of netCDF restart file. However, since coupled model file management already involved numerous files of various types and formats, then avoidance of the need to manage a further file type was considered desirable. Each component model dump contained, or could be modified to contain, sufficient information to achieve this bit reproducibility by rendering restart data files self-consistent. i.e. each could be used as the basis to restart the appropriate component without recourse to OASIS3 restart files and without the attendant data management issues.

28. p 40. line 21: Define “GCOM” more clearly.
Add explanation of GCOM thus: “GCOM (General COMmunications) is an interface library used by the UM which provides an interface or "wrapper" for MPI calls within the code. The presence of the GCOM wrapper around MPI communications in the UM caused some difficulty....”

29. p41 line 2: “UM was the . . . controlling model component” spell out what this means.
We propose to change the paragraph: "GCOM also assumed that the MPI_COMM_WORLD MPI communicator was solely for use by the UM component, i.e. that the UM was the sole and controlling model component....." To: "GCOM worked on the assumption that it was employed within a model component which had no need to communicate with any external components. Thus, it employed the MPI_COMM_WORLD MPI communicator directly for all its operations. This conflicted with the need for OASIS3 to be able to define MPI sub-communicators for each model component. Hence modifications were made to GCOM, allowing it to use communicators provided by OASIS3 in place of MPI_COMM_WORLD."

30. p41 lines 15-17: Passage obscure – consider rewording.
We propose to change the paragraph: "Hence any change in .." To: “Hence any change in processor arrangement required modifications to the source code followed by recompilation. This procedure was automated through the use of numeric values for relevant parameters supplied via cpp keys. These changes (with the aid of specially developed scripts embedded in the UMUI and environment variables) defined and updated critical information, such as the processor configuration, prior to compilation.”

31. Fig. 1: Some of the individual heat and freshwater exchanges would be better represented by arrows going both ways up and down.

The arrows on this figure follow the flow of information in the coupling, from the atmos to the ocean/sea-ice rather than the flow of energy or mass in the model (or even reality). For example, the evaporation/sublimation would almost always be upward in reality, but figure 1b shows that it's an atmosphere flux that contributes to the SSH and the SSS. Reversing the arrows may cause confusions but we could change the caption to explain it shows the direction of information flow in the model.

32. Figs. 4, 5 and 8: Hard to read aspects of these figures in the printout due to small font or arrow size (but ok to read on line).

Unfortunately we do not think there is much scope to improve these figures substantially. However, for reference purposes we usually print figures 4 and 5 on A3.