Response to 'Technical questions', Eric Maisonnave

My main question concerns computational cost of the system, "comparable [...] to running in atmosphere-only mode". It is written that MC-KPP is "computationally inexpensive (<5% of the cost of the atmosphere ...)", and it is clear that this cost is smaller than for a full ocean model, but I would appreciate to know if the MPI communication library has been used to parallelize this mixed-layer model, to be able to evaluate its scalability at higher resolution. I also would like to know how much additional computing cores are necessary to run MC-KPP. To summarize, and have a clear idea of this computational cost (including coupler and coupling synchronization), may I suggest to the authors to provide two simple numbers for both atmosphere-only and MC-KPP coupled runs ? The number of simulated year per day and the number of core.hours necessary to simulate a year.

We do not use MPI communication within MC-KPP, but we have incorporated OpenMP threading within the model's one-dimensional physics routines to improve performance for large coupling domains or fine horizontal resolution.

The additional computational cost from coupling depends on the horizontal resolution and the construction of the supercomputing system. We used the ARCHER (Cray XC30) supercomputer to perform these simulations. On ARCHER, it is not possible to share a computing node between executables. There are 24 cores per node. For the MetUM-GOML simulations, we ran the MetUM atmosphere on four nodes, the OASIS coupler on one node (as a serial process) and the MC-KPP ocean model on one node. Timing diagnostics from our simulation indicate that MetUM CPU time was 98% of wallclock time, OASIS CPU time was 3% of wallclock time and MC-KPP CPU time was 7.5% of wallclock time. The difference between each of these numbers and 100% represents the amount of time that each component of the coupled model is waiting for data from the other components. We simulated 1.67 years per wallclock day and required 2073.6 core hours per model year.

Based on these figures, the MetUM atmosphere requires:
4 nodes x 24 cores x 0.98 (CPU time/wallclock time) / 1.67 (years/wallclock day) x 24 (hours/day) = 1352 core hours per model year (of actual CPU time)

The MC-KPP ocean requires
1 node x 24 cores x 0.075 (CPU time/wallclock time) x 1.67 (years/wallclock day) x 24 (hours/day) = 72 core hours per model year (of actual CPU time)

From these two numbers (72 divided by 1352), we conclude that MC-KPP requires approximately 5% of the computational cost of the atmosphere, measured by CPU time. We acknowledge that the design of the supercomputer does not allow this efficiency in practice, however.

The MetUM atmosphere-only simulations in this study used the same number of cores (144), but obviously all cores were devoted to the atmospheric model. Those simulations achieved 2.08 years per wallclock day and required 1661.5 core hours per model year. Based on these numbers, MetUM-GOML requires approximately 25% more wallclock time per model year.

We emphasize that these numbers depend strongly on the number of cores that the user decides to devote to the atmospheric model. The MetUM does not scale linearly, so varying the number of cores assigned to the atmosphere will alter the how "cheap" MC-KPP appears relative to the cost of *the atmosphere. Even at a horizontal resolution 25 times finer (i.e., 25 times more gridpoints) that the one used here, it is possible to run MC-KPP on only one node without slowing the coupled model simulation. In such a simulation, the atmosphere is typically configured with at least 32 nodes (768 cores), but MC-KPP still requires only one 24-core node.*

We have clarified the costs in the revised text, noting the difference between the cost in CPU time and the cost in wallclock time. We hope these details are useful to you. Please contact us if you would like further information.

My second question is related to flux correction technique (2.1.1 §). I wouldn't say that "constraining ocean temperature and salinity profiles in the coupled model produces small SST biases" "by construction" "in the resulting free-running simulation". In this free-running simulation, a correction is imposed "with no interactive relaxation": it means that the coupling could freely lead to substantially bigger anomalies than the imposed correction. It is not the case in your simulation, but it is not as obvious as "by construction" can suggest it. Could you describe with more details how this correction is calculated ? And could you precise if Figure 1b shows results of a free-running coupled simulation with correction made with a15-day relaxation run ?

The aim of section 2.1.1 is to describe the flux correction technique in general terms. The specific details of the setup used in the experiments which are the focus of the remainder of the paper are later detailed in section 2.2. This is purposefully done separately; the modelling framework is presented first in general terms then one particular application of that framework is described and analysed. Therefore, the comment is correct in saying that constraining ocean temperature and salinity profiles does not (always) 'by construction' lead to small SST biases. The text about small SST biases at the end of section 2.1.1 was linked to a particular example of this modelling setup in which observations were used as the reference climatology in the relaxation run to calculate the temperature and salinity corrections.

It has been made more obvious in the text that the comment about small SST biases and Figure 1 relates to an example (with observations as the reference climatology and a 15-day relaxation timescale). The subtitles in Figure 1 have been updated to show exactly which observational dataset the biases are calculated against and in the caption it has been made clear that this is from a free-running MetUM-GOML simulation in which the flux corrections have been calculated as described in section 2.2 (observations as a reference climatology and a 15-day relaxation timescale).

Another correlative question that seems not addressed in this document is the possible drift in the free-running simulations ("K"). Mixed-layer models are particularly sensible to flux correction calculation: small differences in imposed correction values lead, sooner or later, to temperature and salinity drifts in free-running simulations. I didn't find any information about this drift, since Figure 1b only shows an average value (of 10 years ?). This value is important to determine the maximum length of the free-running simulations (before reaching biases values comparable to that of full ocean coupled simulations).

To clarify, Figure 1b shows an average of 60 years of free-running simulation (calculated from 3x20 years of MetUM-GOML simulation). The global RMSE of SST over those 3x 20 years of simulation for the three K-O ensemble members is shown in the graph below. All three members remain below a RMSE of 0.38 for the entire 20 years, compared with an RMSE value of 0.82 for MetUM-NEMO (dark *red dot). There is no clear drift over the first 20 years. Coupled MetUM-GOML simulations have been continued to 60 years and do not indicate signs of drift.*

The subtitles and captions of Figure 1 have been updated to clarify what the Figure shows. A comment about there being no signs of drift in these simulations has been added to the end of section 2.1.1.

I am interested by this powerful and necessary framework and would like to know how easy it is to adapt MC-KPP to another atmosphere grid. Maybe a few sentences about how this grid is defined and how parallelization (if any) is organized would help me to evaluate the amount of work necessary for such operation.

MC-KPP defines its grid by reading the land/sea mask of the atmospheric GCM to which it is coupled. Since the columns are independent, their spacing and organization are arbitrary. As stated above, there is no parallelization in MC-KPP, only OpenMP threading.

One of the authors, Nicholas Klingaman, has successfully coupled MC-KPP to two other GCMs, the NCAR CAM and the NCEP GFS. Both of these are spectral models, which demonstrates that MC-KPP is easily adaptable to irregular grids.

Response to 'Exec Editor Comments'

"All papers must include a model name [AND] version number (or other unique identifier) in the title. "

The model name in the title has been changed from MetUM-GOML to MetUM-GOML1.

Response to Review #1

Minor comments:

1. P6175 L19: Did Straub and Kiladis (2003) specifically mention the MJO in relation to tropical-extratropical interactions? From the second paragraph in their Section 5b, it seems that their focus is on convectively-coupled Kelvin waves (I admit the line between CCKWs and the MJO can at times be blurry).

The comment is correct that the Straub and Kiladis (2003) paper does not explicitly mention the MJO, although, as stated above, the line between CCKWs and the MJO is indeed blurry. The text has been altered to only attribute the link between equatorward-propagating Rossby wave trains from the SH and CCKWs (not the MJO specifically) with the Straub and Kiladis (2003) study.

2. P6179 L6-11: It may also be worth mentioning that sensitivities can also arise due to choices of mixed-layer depth and relaxation timescale.

The reviewer is correct in suggesting that such sensitivities are missing from the discussion of slab ocean models. The part of section 1.2 discussing slab ocean models has been extended considerably in the manuscript. Further related modifications have been made to section 1.3.

These sections now outline the advantages of having a vertically resolved ocean over a slab on the representation of tropical variability (Woolnough et al. 2007, Klingaman et al. 2011, Tseng et al. 2014). Discussion has also been added, as suggested, about the sensitivity of simulated variability to the depth of the slab ocean. For example, the Maloney and Sobel (2004) and Watterson (2002) studies which systematically test the sensitivity of variability to different slab depths are discussed. The additional point is made about the inability, within a slab configuration, to store anomalies below the mixed-layer depth which can later re-emerge and affect the atmospheric circulation. Sensitivities to temperature correction techniques employed in a slab model are also mentioned.

3. P6186 L10-11, L23-25: It's quite difficult to assess 2K and 4K temperature differences based on the way that Fig. 3a is plotted (10K increments). Would it be worth using the shading for A-K31 minus ERA-Interim instead of shading total A-K31? I have a similar reaction to Fig. 4a.

The author s accept this comment. In Figures 3a and 4a the shading has been changed to A-K31 minus ERA-Interim rather than the total field for A-K31. This makes identifying biases compared with ERA-Interim much clearer. Following a comment by a separate reviewer, stippling has also been added to Figures 3 and 4 to indicate where differences are significant at the 95% level.

4. Fig. 6: I recommend enlarging the text on the Figure 6 panels. Also, it is difficult to see and compare the MJO spectral region.

The text on Figure 6 has been enlarged. A grey box has been added to highlight the MJO spectral region to make these comparisons easier. The spectral box is added for the region of 30-80 days, wavenumbers 1-3.

5. P6192-6193: I didn't feel that, given the commentary of previous studies presented in the last paragraph of Sec. 4.1, the authors brought enough closure to the discussion of extratropical storm tracks. Given the notable improvement to tropical intraseasonal variability (MJO), it was somewhat

disappointing to see that the storm tracks did not change much for the better. I encourage the authors to add a brief comment about this result – does this indicate that the connection between tropical and extratropical variability in the MetUM-GOML is not as robust as it is in nature? Is this representative of an inherent weakness in the MetUM regardless of whether air-sea coupling is active?

There are improvements in the representation of the MJO in the coupled MetUM-GOML configuration shown here compared with the MetUM atmosphere-only versions: slight improvement in spectral power associated with the MJO (Figure 6), improved intraseasonal variability in precipitation (Figure 7) and significant improvements in MJO propagation (Figure 8). However, there remain significant deficiencies in the representation of deep tropical convection in the MetUM which result in weaker-than-observed MJO activity in K-O. The amplitude of MJO power in Figure 6 (d) is still significantly less than in NOAA observations Figure 6 (a). Therefore, it may be that the improvements in MJO activity in K-O are not significant enough to influence the circulation response to the MJO in the extra-tropics which is why minimal changes are seen in the storm tracks by this measure (Figure 9). That said, all of the changes in 2-6 day variability in Figure 9 are significant at the 95% level and there are considerable changes to the representation of blocking frequency in the Euro-Atlantic region with the coupling (Figure 10). This suggests that there may be teleconnection patterns associated with the improvements in intraseasonal variability in the tropics. The tropical-extra-tropical teleconnections within these runs warrants further investigation.

Horizontal resolution may also play a role here in the connection between tropical and extra-tropical variability. Increasing horizontal resolution has been found to improve storm-track variability and blocking (e.g., Matsueda et al. 2009). Therefore, the N96 horizontal resolution used in the experiments in this study may be too coarse to capture this extra-tropical variability, no matter how good the tropical variability is represented.

In response to this comment the text at the end of section 4.1 has been changed to reflect that, although improvements are made in the representation of the MJO in K-O, significant deficiencies still remain. The text now states that if the improvements in MJO activity are large enough and the MetUM is able to accurately simulate the circulation response to increased MJO activity associated changes may be seen in extra-tropical variability. Further clarification has been added to the text at the end of section 4.2.1.

6. P6195 L7-10: This note may be more important than the authors seem to suggest and should be moved (or, better yet, repeated) earlier when the experimental setup is described in Sec. 2.3. The temporal resolution of the SSTs (1- vs. 31-day averages) may well impact atmospheric circulations whether or not air-sea coupling is active.

This note about the 31-day smoothed SSTs including the effect of increased, high frequency SST variability has been repeated at the end of the experimental setup section (section 2.3).

Possible typographical errors:

P6175 L19: have -> has

This has been changed as suggested.

P6176 L13: influence -> influences

This has been changed as suggested.

P6195 L28: indicate -> indicates

This has been changed as suggested.

Response to review #2

One aspect, in which the paper can be improved significantly, would be additional comparison with a simulation using the AGCM coupled to a slab mixed layer ocean model, in addition to the comparisons with the atmosphere only simulations. As the authors discussed in the introduction, the AGCM-slab ocean configuration is the most often used experimental design as the intermediate step between the AGCM-only and fully coupled configurations, and the most relevant one to the new MetUM-GOML. Therefore, it would be helpful for the future potential users to demonstrate the advantage of using MetUM-GOML over the AGCM-slab ocean.

It is true that the introduction in section 1.2 discusses slab ocean models as an intermediate step between fully coupled AOGCMs and AGCM-only models and that mixed-layer vs. slab comparisons could be potentially interesting.

Therefore, the discussion of slab ocean models in section 1.2 and the advantages of using the MetUM-GOML configuration over the AGCM-slab ocean model in section 1.3 have been modified and extended considerably. These sections now discuss the benefits of having a vertically resolved upper-ocean on the representation of variability (e.g., in the tropics for the MJO; Woolnough et al. (2007), Klingaman et al. (2011), Tseng et al. (2014)). They also discuss a further advantage of a mixed-layer over a slab ocean configuration: that within a mixed-layer configuration, temperature and salinity anomalies can be stored below the mixed layer and re-emerge in subsequent seasons as SST anomalies in the extra-tropics which can affect the atmospheric circulation (e.g., in the North Atlantic; Bhatt et al. (1998), Alexander et al. (2000), Cassou et al. (2007)). These advantages of the mixed-layer MetUM-GOML configuration over the intermediate AGCM-slab ocean experiment have been made clearer in the discussion.

These sections now also discuss the known sensitivities within a slab-coupled model configuration to the slab depth. For example, the representation of intraseasonal precipitation variability and propagation speed of the MJO have been shown to be very sensitive to the choice of slab depth (Maloney and Sobel (2004),Watterson(2002)). Given that the variability in the slab-coupled model can be tuned by varying its depth, it is not clear how to make a fair comparison between a slab and a vertically resolved ocean. This would certainly not be possible with a single slab configuration.

Therefore, because we already know that vertical resolution in the ocean is better than a slab and because it is non-trivial to fairly compare a slab and a vertically resolved ocean given the above-mentioned sensitivities to the slab depth, we believe further comparisons with AGCM-slab ocean configurations of the MetUM are beyond the scope of this particular study.

Overall, I recommend a minor revision of the manuscript. Additional comments are included below.

1. L.179-185, L6, L789: While the MetUM-GOML configuration is not one of the most commonly used types in climate modeling, it is not the first model with the full AGCM coupled to the 1-D multi-layer ocean mixed layer model. Please refer to the following papers for the similar previous experiments:

Bhatt, U.S., M.A. Alexander, D.S. Battisti, D.D. Houghton, and L.M. Keller, 1998: Atmosphere–Ocean Interaction in the North Atlantic: Near-Surface Climate Variability. J. Climate, 11, 1615–1632.

Alexander, M. A., J. D. Scott, and C. Deser, 2000: Processes that influence sea surface temperature and ocean mixed layer depth variability in a coupled model. J. Geophys. Res., 105, 16823–16842.

Alexander, M.A., I. Bladé, M. Newman, J.R. Lanzante, N.-C. Lau, and J.D. Scott, 2002: The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans. J. Climate, 15, 2205–2231

Cassou, C., C. Deser, and M.A. Alexander, 2007: Investigating the Impact of Reemerging Sea Surface Temperature Anomalies on the Winter Atmospheric Circulation over the North Atlantic. J. Climate, 20, 3510–3526.

Kwon, Y.-O., C. Deser, and C. Cassou, 2011: Coupled atmosphere–mixed layer ocean response to ocean heat flux convergence along the Kuroshio Current Extension. Climate Dyn., 36:11-12, 2295-2312.

The authors acknowledge that this is not the first model using a full AGCM coupled to a 1-D mixed-layer ocean model. All of the above papers are configurations coupling the NCAR Community Atmosphere Model (CAM) version 2 to a one-dimensional ocean model developed by Gaspar (1988) and the authors are aware of them.

In the beginning of section 1.2 "widely used" has been added to the description of current approaches to climate modelling to reflect the comment that there are others. At the start of section 1.3 the gap in modelling capability is referring to the "widely used" approaches outlined in section 1.2. This has been clarified in the text.

Further to this, the CAM2 mixed-layer ocean configuration has been explicitly mentioned in section 1.3 with reference to the studies mentioned above. The point is made, however, that there are only a handful of such studies and that they do not typically use a contemporary AGCM. Until now, there is no existing mixed-layer (or slab) coupled model configuration of the MetUM.

Incidentally, current work by one of the authors, Nick Klingaman, has involved coupling the MC-KPP mixed-layer model to a more recent version of the NCAR CAM model. It will be interesting to compare these simulations with the existing studies which are highlighted here.

Additionally, a number of these references have been added to the discussion in sections 1.2 in relation to the advantage of a mixed-layer over a slab coupled model configuration. Specifically in the ability within a mixed-layer ocean model to store anomalies below the mixed-layer depth which can later re-emerge as SST anomalies in subsequent seasons and affect the atmospheric circulation.

2. Figure 1 caption: Please explain which observational dataset is used to calculate the model biases.

In Figure 1 in the caption and subtitles it has been made clear that the bias is calculated relative to the Met Office ocean analysis of Smith and Murphy 2007.

3. L388-391: Please discuss a bit more detail on the sensitivity of the model simulation to the choice of the relaxation time scale, e.g. how the results change from 5-day to 90-day time scale, or what objective measure is used to determine the time scale.

We tested 5-day, 15-day, 30-day and 90-day timescales. The 15-day timescale produced the smallest SST biases in the free-running coupled simulation, so we chose that timescale for the simulations presented in this study. Longer timescales produced stronger SST biases since the relaxation was too weak to counter the SST drift in the forced simulation, which arises from the lack of ocean dynamics and biases in atmospheric surface fluxes. The 5-day relaxation timescale is analogous to forcing the atmospheric model with climatological SSTs. In the relaxation simulation, the atmospheric surface fluxes did not adequately adjust to the presence of coupling. This led to a substantial difference between the surface-flux climatologies of the free-running simulation and the relaxation simulation, for which the temperature and salinity tendencies could not correct, and hence larger SST biases than the simulation in which we used a 15-day relaxation.

Text has been added to section 2.2 in the model description to give more detail about what objective measures were used to determine the preferred timescale.

4. L423-424: Please briefly explain why the 31-day smoothing is applied.

The A-K31 experiment is designed to mimic the AMIP-style setup of forcing the atmospheric model with monthly-mean SSTs. In this case a 31-day running mean was applied because it produces a smoother SST timeseries than interpolating monthly means to daily values. This has been clarified and added to the text in the description of the experimental setup in section 2.3.

5. Figures 3-5: It would be worth adding one more panel showing the MetUM-fully dynamical ocean (used in Fig. 1a) minus observation to compare with A-K31 minus observation, which will show the typical biases in a fully coupled model.

The authors partly accept the comment about showing the typical biases of the fully coupled version of the model, appropriate changes have been made to Figure 5 (see details below). However, the authors do not feel that adding such a comparison to Figures 3 and 4 would aid in the discussion presented here. The purpose of this manuscript is not to demonstrate that biases in the MetUM-GOML model configuration are larger or smaller than MetUM-NEMO - indeed some biases increase in magnitude and some decrease, others are simply different – rather it is to investigate the role of air-sea interactions within a framework which has minimal effect on the mean state.

As suggested, a new panel has been added to Figure 5 to show the tropical precipitation bias of the MetUM-NEMO (MetUM coupled to a fully dynamical ocean model, as used in Fig.1a) compared with TRMM. A further panel has been added to show how the representation of tropical precipitation differs in A-K31 compared with MetUM-NEMO. In the case of tropical precipitation, the biases are of similar magnitude in A-K31 and MetUM-NEMO but exhibit different spatial structure. The different precipitation biases seen in Figure 5 are linked to the different SST biases between MetUM-GOML and MetUM-NEMO (Figure 1). Discussion has been added to the text to reflect this.

6. Figures 3-5, 7-9: Please test the statistical significance of the anomalies and discuss only when they are statistically significant.

The authors accept this comment and have tested the statistical significance of the anomalies shown in Figures 3-5 and 7-9 and made modifications to the figures and text accordingly, see details below.

In Figures 3 and 4 stippling indicates where the differences are significant at the 95% level. Figure 3a and 4a have also been changed such that the shading shows the bias following a comment by a separate reviewer.

Figure 5 has been modified such that only differences significant at the 95% level are shaded. Further panels have been added to Figure 5 to show the bias of the fully coupled MetUM-NEMO model compared with TRMM (Figure 5 (d)) and compared with the A-K31 (Figure 5 (b) ; following comment above). The relevant text has been changed accordingly. Additionally, a mistake in the JJA colour scales in Figure 5 panels (g) and (h) has also been corrected so that they are consistent with the colour scale shown on the multi-panel plot.

All changes in variance shown in Figures 7 and 9 and discussed in the manuscript are significant at the 95% level.

Stippling has been added to Figure 8 to show where the differences are significant at the 95% level.

7. Figure 6 caption: “interio-gravity” -> “inertio-gravity”

This change has been made to the figure caption.

8. Figure 8 caption: “130deg” -> “130degE”

This change has been made to the figure caption.