Responses to Reviewer 1

General comments:

"It meets the GMD criteria of a model description paper, but does not advance the science of ocean modeling. Overall, I felt it is lacking even as a simple description. ....
"...In short, there is little of interest here to anyone except those using this particular version of the model, and then only as a supplement to the user guide. I find little reason to recommend publication."

The authors are surprised by the reviewer's stance on this point. The paper describes a configuration of NEMO that will be used widely in applications that are high profile, and there are few ocean modelling configurations that will have as much user interest. This model will be used as part of the GMES Marine Core Services (delivering daily forecasts to European taxpayers), it is being developed as part of a coupled Numerical Weather Predictions system; it has been incorporated into the UK's seasonal forecasting system that is widely used within government advice and policy; and it will be a substantial stepping stone on the road to producing the UKs contribution to IPCC AR6. Additionally it is made available as the global ocean model for UK research institutes. It is essential that the users of this configuration have a baseline description of the configuration and its performance.

So, even if this paper is only of interest to users of this particular configuration, that is a significant (and in our view unusually significant) group which alone makes this paper worthy of publication.

Additionally, NEMO is a widely used model and the assessment of the changes from NEMO3.2 to NEMO3.4, and the significant impacts that will have upon model performance, are not published to our knowledge. This result will, therefore, be of significant interest to those beyond the users of this configuration.

"There is no motivation for what improvements were sought in the changes implemented in the model, just an unprioritized list."

We have improved the description of the sensitivity studies. Section 5.3 is now divided into subsections and paragraphs which deal with the individual changes to processes and parameters. In each case we have provided a rationale for the change, its expected effect (even if it is expected to be negligible). Some changes are included simply because they are more physically correct (such as the improved bathymetry and the salinity-dependent freezing point) irrespective of whether or not they improve the simulation. We have also provided more rationale for the changes made in Sections 3 and 4.

"*The main conclusions are that the main improvements from the prior version of the model GO1 are due to a.) a bug fix, and b.) some changes in some parameters in the vertical mixing code. The variable names from the code for these parameters are provided, but they are not defined in terms of the model equations or what physics they control. The justification for these later changes is a citation of a grey-literature report (Calvert and Siddorn, 2013).*"

We accept that the physical significance of the key mixing parameters was not made clear, and we have made changes to the text to improve this. In particular the main results of Calvert and Siddorn (2013) are now presented in Section 3 along with a justification for the changes made and what the expected improvement were (mainly in Sections 3 and 5.3). The function of each of the parameters is also described in this section. Our use of the term "bug-fix" was perhaps unfortunate – it is better to say that the model does not now feed the mixing due to convection scheme back into the turbulence scheme, which we consider to be more physically realistic. We have clarified the text describing the changes between versions, as well as giving descriptions of the physical role of each of the parameters.

<u>Specific Comments</u>

pg 5752, line 11-14: "*What if any smoothing or interpolation is used to go from the 1/60 deg. topography data to 1/4 degree model grid?*"

The derivation of the Drakkar bathymetry is described by Barnier et al (2006). Initially, each model grid cell was assigned the median of all observations falling within the boundaries of that grid cell. The initial estimate was then modified by application of two passes of a uniform Shapiro filter and finally hand editing was performed in a few key areas. We have clarified this in the text,

pg 5752, line 18 *"reducing polewards": at what rate?*"

The bilaplacian horizontal viscosity is proportional to the cube of the maximum horizontal grid dimension at each grid cell; thus at 60°N the horizontal viscosity is approximately 1/8 of its value at the equator. We have now added this clarification to the text.

pg 5752, line 20 "*no parameterization of eddy mixing is used" This is false. Isonuetral diffusion is a parameterization of eddy mixing.*

We have replaced this with the more accurate statement: "The isopycnal mixing scheme of Gent and McWilliams (1990) is not used in this configuration."

pg 5753, line 23 "*180 days" This should be expressed as a piston velocity (independent of vertical resolution) to be compared with other published values.*

This is "rn_deds" in the namelist, equal to -33.33 mm/day/psu. We have added this clarification to the text.

pg 5754, line 24: *The results of the Calvert and Siddorn study need to be summarized here, or a standard citation provided. These are shown to be among the key changes in the model.*

We have added some sentences summarising the results of Calvert and Siddorn in the first paragraph of Section 3.

pg 5754, line 27 *"a number of additional parameterizations": These need to be enumerated if this manuscript is meant to be a documentation of the model design.*

We have added the following text to the first paragraph of Section 3: "Additionally, the NEMO implementation of the scheme includes a number of parameterisations to represent additional unresolved turbulent processes, including surface wave breaking (Craig and Banner, 1994) and Langmuir turbulence (Axell, 2002). A further parameterization represents the enhanced mixing due to breaking of near-inertial waves as an additional source of TKE exponentially decaying from the surface."

pg 5755, line 15 "*use of in situ salinity": This will cause a global non-conservation of salt. What is the magnitude of this drift in the solutions presented?*

This was an erroneous statement in the first version of the paper: in fact NEMO v3.2 and v3.4 both use the in-situ surface salinity to calculate salt fluxes. With this in mind, given that this drift will be almost the same in all the model configurations that we discuss here (and given that the SSS relaxation and river runoff also contribute to salinity non-conservation) we feel that analysis of this is outside the scope of the paper.

pg 5756, line 21-22: *What are these parameters? What physics do they control?*

This refers to the TKE scheme parameters *rn_ebb*, *rn_mxl0* and *nn_htau*. Text has now been added to the relevant part of Section 3 to clarify the purpose of these parameters and their expected effects.

pg 5756, line 24: *What is the "lake parameterization" How does it affect the open ocean?*

The lake modification adds inland seas and lakes ready for a coupled configuration and has no effect on the global ocean basins in an ocean-only simulation. In our simulations the lake modification was not enabled (we apologise for our previous lack of clarity on this). This means that the sensitivity experiment originally described as "bottom boundary layer and lakes", in fact only added the bottom boundary layer, and we may therefore ascribe the changes described entirely to the BBL. We have accordingly removed references to the lake scheme.

pg 5757, line 25 *“cool bias over most of Northern Hemisphere” This is not at all clear from the figure. There is a strong positive bias over the western boundary currents and the California Current system for example. What is the hemispheric mean bias?*

We have revised and corrected the description of the surface biases as follows: "There is overall a warm bias over most of the global ocean, with a global mean bias of +0.72°C, and with the largest biases (of over 1°C) in the tropics, the Southern Ocean, the subpolar North Atlantic and over the separated western boundary currents in the North Atlantic and North Pacific. There are cool biases of 0.25-0.50°C extending over much of the subtropical North Atlantic and North Pacific."

pg 5758, line 20-25: *These statements do not agree with what is shown in the figures, nor do the figures make sense. For example, the minimum MLD in the tropical Pacific shown in Fig 2a appears to exceed the maximum MLD shown in Fig 2c. The figure also lacks latitude and longitude axis labels.*

The legend on Figures 2(a) and 2(b), showing the minimum MLD, was incorrect: they are plotted with a different scale (with a range of 0-100m) from that for the maximum MLD, but the same legend was erroneously used for all panels. This has now been corrected, removing the inconsistency noted by the reviewer, and we have also added lat/long axis labels.

pg 5758, line 27 *“deep mixing extending from the Weddell Sea”: The observations at emitting in this region, how can you compare with the solution?*

The large Weddell Sea Polynya that emerges after year 20 of GO5.0 has some similarities with the feature observed in 1976-78, although its persistent recurrence in the latter part of the run is probably unrealistic. We have cited observations of winter mixing in excess of 500m by Arthun et al using Weddell seals, but have acknowledged the paucity of winter observations.

pg 5759, line 10 *“Note the...” Why should we note these? There is no discussion of the density field illustrated in Fig. 3.*

The isopycnals were added to illustrate the position of the biases with respect to the main pycnocline. We have revised the text, including an explanation of the contour lines on this figure, and have rephrased this sentence.

pg 5759, line 26 *“higher end”: No, they are beyond the range of the observations.*

We have replaced the phrase ...at the higher end..." with "...significantly stronger than...".

pg 5761, line 20 *“Gibraltar Straits” The net flow through Gibraltar should be exactly zero when using a virtual salt flux b.c.*

No figure for Gibraltar throughflow was included in Table 3, and we have removed reference to it from the relevant text, which now only refers to Bering Strait throughflow.

pg 5766, line 2: *“ probably a signature of ENSO” No need to speculate. Test this.*

We have revised this sentence, noting that the subsurface variability is not well correlated with the surface variability, so is unlikely to be directly related to ENSO.

Pg 5766, line 27-28: *Provide values for the spatial RMS.*

We have added global mean RMS SST errors for JJA (reduction from 0.93°C in GO1 to 0.65°C in GO5.0), DJF (reduction from 0.79°C in GO1 to 0.67°C in GO5.0), which confirm the improvements in the fidelity of the annual cycle in GO5.0.

pg 5768: *“larger effect .. where the surface salinity is saltier ..”: This is false. The constant reference salinity is probably around 35 psi. The error where they salinity is very high, say 38, is about 10%. The error where the salinity is 0 psi is 100% (the salt flux goes to zero when using the locally references salinity).*

This is indeed true, but we have now removed the relevant text because (as mentioned earlier) the surface salinity forcing formulation is in fact identical in the two model versions.

pg 5770, *line 1-2: No idea what a lake parameterization is or why it should produce changes similar to that of a BBL scheme?*

We have confirmed that the lake modifications are disabled in the experiments described here, so any changes between these two experiments may unambiguously ascribed to the BBL. We have accordingly removed all mention of the lakes scheme.

<u>Technical Corrections</u>
pg 5758 line 15-16 *realistically... realistically. Remove first instance.*

Done.

pg 5764, line 23: *There is no Fig 6i*.

This has been corrected to refer to Figure 6h.

pg 5765, line 16: *What do you mean by a “drift” in the climatology? A long-term trend?*

We have replaced "drift" with "trend".

Responses to Reviewer 2

General comments

"The paper focus on the surface layer and there is in the text few comments or remarks about the atmospheric forcing, I suggest adding some figures to illustrate impact of the atmospheric forcing which is the main driver of the ocean and which can explain lot of bias discussed in the paper. For example, it will be useful to know how the surface fluxes, computed with the bulk formulae, are impacted in the experiments. Do we have a large difference in the mean fluxes between GO1 and GO5.0?".

We have added an extra figure (Fig. 10) showing zonal mean heat fluxes from GO1 and GO5.0 alongside those from the CORE2 dataset, and the 2-D surface difference in the fluxes between GO1 and GO5.0, as well as an extra subsection (5.2.3) discussing them.

" The more interesting part could be the sensitivity study with the comparison between the two simulations GO1 and GO5.0 and the sensitivity experiments. Unfortunately these sections (5.2 and 5.3) are really poor in term of results, interpretations, conclusions, recommendations and we can't find real contribution in the understanding of modelisation approach or parameterization. I suggest that authors complete this section using the sensitivity experiments to quantify improvements or to explain in a more detail way process and/or numeric involved in these changes. Authors shall also justify in a more precise and scientific way the sensitivity tests performed and described in the paper. I am not sure that there is sense to talk about some tests as geothermal heat flux or lakes for example. If yes, explain what is expected, how it works and justify why there is no impact. I suggest adding almost a new sensitivity experiment to understand the differences changing NEMO version separating vertical mixing parameterization (or mistake in parameter) and computation of salt flux.".

We have improved the description of the sensitivity studies. Section 5.3 is now divided into subsections and paragraphs which deal with the individual changes to processes and parameters. In each case we have provided a rationale for the change, its expected effect (even if it is expected to be negligible). Some changes are included simply because they are more correct (such as the improved bathymetry and the salinity-dependent freezing point) irrespective of whether or not they improve the simulation. We have also provided more rationale for the changes made in Sections 3 and 4.

On the last point (as mentioned elsewhere in our responses), there is in fact no difference in salt flux formulation between NEMO v3.2 and v3.4, so the only physics change between versions is in the TKE scheme.

Specific comments:

Section 3

More information about Calvert and Siddorn (2013) will be useful. What kind of sensitivity experiments? Which configuration? What kind of validation? How conclusion obtains with 1° resolution model are available at 1/4°?

We have added some text summarising the results of Calvert and Siddorn in the first paragraph of Section 3.

For vertical mixing, which is treated as a 1D model, direct impacts are to first order horizontal resolution independent. Clearly, results from a 1° model may not translate directly to a 1/4° model (or indeed to any other model setup), which is why we also tested the impacts in the ORCA025 model to ensure they had the expected impact. As the paper shows, they did.

There is a long list of differences between the GO1 and GO5.0 simulation but it is not clear what are exactly the differences between GO1 (based on NEMO 3.2) and N3.4 (based on NEMO3.4). I understand that there is a bug correction in TKE and new way to convert freshwater flux in salt flux. This is briefly discussed in section 5.3 but there is no way to separate the 2 effects.

We have confirmed that the reference to a difference in the salt flux formulation between the two model versions was in error, and we have accordingly removed all mention of this. The one remaining significant difference between the versions is in the implementation of the TKE scheme.

Section 4:

How do you initialize the 10-year sensitivity experiments? From rest or from another experiment as GO1?

All the integrations are started from rest with T and S fields taken from an average of years 2004-2008 of the EN3 monthly objective analysis, as stated in Section 2. We have now explicitly stated that all the runs start from the same state. The details of initialisation and forcing are now moved to the Experimental Design section, since they are not part of the GO5.0 configuration itself.

More information about the namelist parameters in the NEMO TKE implementation will be useful. What are the expected effect of rn_ebb, rn_mxl0, nn_htau?

We have added brief descriptions of the functions of each of these parameters, as well as the motivations for changing them, to the summary of the results of Calvert and Siddorn, which has been added to Section 3.

How do you justify the sensitivity experiment changing at the same time *geothermal heat flux and double diffusion of tracers? What are the expected effects of these parameters? Can you provide reference or can you explain what is done?*

In response to the reviewer's comment, a separate experiment was carried out in which only the geothermal heat flux was added in order to isolate the effect of each process. It is physically correct to include the effects of geothermal heating and double diffusion and for this reason we included them. On the timescales of our model simulations and expected applications we do not expect either process to be important on ocean basin scales, but for the sake of completeness we have investigated their effect. As expected, neither process resulted in significant changes.

Same remark for the bottom boundary layer and lakes. If we can understand that BBL can impact your solution and especially some really important diagnostic/process described in the following, what about the lakes?

We have confirmed that the lake modifications are disabled in the experiments described here, so any changes between these two experiments may unambiguously ascribed to the BBL. We have accordingly removed all mention of the lakes scheme.

P5727 l10, Last paragraph: reader needs more explanation to understand what are these parameters. Is the 10-year simulation performed with N3.4 and 0.01 one of the sensitivity tests listed in table 2?

This refers to the parameters *rn_mxl0* (and *rn_lmin* (actually on page 5757). The first of these is the minimum permitted surface mixing length, as is now stated earlier in this section; *rn_lmin* is an interior minimum length scale for the TKE scheme in NEMO v3.2, but not used in v3.4. We have clarified this in the text.

We have added the extra experiment, which we denote N3.4_mxl0, to the summary table for the sake of completeness, although we do not discuss it further in Section 5.

<u>Section 5:</u>
P 5758 l24: *more information about the MLD climatology in the deep convection area (WEDDEL sea and North Atlantic) is needed to validate the simulation. What are the uncertainties in this area in the climatology? Most of the observations are not deeper than 2000m. There is no observation during austral winter in the climatology in the Weddel sea (figure 2).*

The large Weddell Sea Polynya that emerges after year 20 of GO5.0 has some similarities with the feature observed in 1976-78, although its persistent recurrence in the latter part of the run is probably unrealistic. We have cited observations of winter mixing in excess of 500m by Arthun et al using Weddell seals, but acknowledged the paucity of winter observations.

P5759 l23: *Can you compute this correlation? Is there a time lag between the density in the Labrador sea and the MOC at 26_N?*

We have revised the text, noting the specific trends in both indices at the start and end of the time series, but have removed any claim that they are statistically correlated.

P5760 l8: *If 30% of the AMOC variability is explained by the meso scale, can you discuss this percentage for the 1/4° model which clearly underestimate the meso scale variability. Can you discuss which part of the variability is explained by the atmospheric forcing, by the general circulation, by other process and phenomena. What is expected in this kind of simulation in term of correlation between MOC observation and simulation?*

The value of 30% is taken from the work of Hirschi et al (2013), in which forced 1/4° simulations similar to those described in the present paper were analysed. In these simulations about 70% of the AMOC variability is determined by the surface forcing, and 30% from intrinsic ocean variability. The text has been revised to include these points.

P5762 l11: *You suggest that the overestimation of the transport at Bering is due to the SSH slope. But why there is this too strong slope between Pacific and Arctic Ocean? Can you suggest some hypothesis about local wind fields, water masses, sea ice, geometry of the strait...?*

We have tentatively ascribed the strong SSH gradient across the Bering Strait with the positive salinity biases in the Chukchi Sea and eastern Arctic, relating this to the hypothesis, based on observational data, of Aagard et al (2006).

P5764 l28: *You write that there is no atmosphere ocean feedback in your simulation. As you use bulk formulae there is a feedback and it will be useful to illustrate flux differences between the simulations.*

The changed surface flux does not correspond to any change in surface air temperature, but we acknowledge that there is limited feedback in the form of a damping term towards the surface air temperature, rather than a true feedback. In this sense the surface flux error diagnoses the tendency of the model to drift away from the climatological SST. We have inserted a new figure (Figure 10) showing the zonal mean heat fluxes and the difference in surface heat flux between GO1 and GO5 and added a subsection (5.2.3) discussing this.

P5765 l2: *I don’t understand the last sentence of this section. What do you mean with “significant reduction in model skill”?*

We have clarified this sentence as follows: "However, in a fully coupled model atmospheric dynamics might cause a significant effect on regions remote from the ice-covered oceans."

P5767 l24: *Why do you choose an empirical criterion for significance? It will be more convincing with an objective criteria based on the variance of the signal*.

We are looking at the differences between five year mean fields and judge visual comparison (given that the variance of these fields is not available form these integrations) to be a satisfactory method of discriminating whether the anomalies are significant, based on the size and large-scale spatial coherence of the T and S difference fields. We have now stated this explicitly at the beginning of Section 5.

P5768 l5: *I think you can't say that the upgrade of version has significant effect, effect is due to difference in the way to use several option in the two versions of the code. I suggest to better document this part and I think you can't discuss this point if you don't have at least one simulation with in a first step the modification in the vertical mixing and in a second step the salt flux computation. It is also your conclusion of this part.*

As stated earlier, the description in the original version was incorrect: the version change does not in fact involve any change in salt flux, so that the differences can be ascribed entirely to the TKE scheme.

p5770 l5, *you have to identify what are the differences between this two versions which induced these largest differences in the results.*

As we have noted, the version upgrade only in fact involved a single physics change (namely the TKE modification)

P5769: *If there is no illustration and no more comment on these sensitivity experiments I don't think it is useful to keep them in the paper. Can you argue on the choice for the sensitivity tests, what are the expected impacts, can you quantify the impact at global scale and in interested areas, what changes are significant?*

We have now divided this section into subsections and paragraphs, each dealing with a single change, and explaining the rationale behind the change and its expected effect and comparing with the observed effects from the sensitivity studies.

P5769 l1: *what are the modification in the bathymetry?*

The derivation of the Drakkar bathymetry is described by Barnier et al (2006). Initially, each model grid box was assigned the median of all observations falling within the boundaries of that grid box. The initial estimate was then modified by application of two passes of a uniform Shapiro filter and finally hand editing was performed in a few key areas. We have added these clarifications to the text.

P5769 l11: *Why are you not sure that the background term is always smaller that the explicit mixing term? You don't have the vertical mixing coefficient in the output files?*

This is true by construction (the background value is a minimum). We have clarified this text as follows: "In the upper ocean the explicit representation of mixing processes by the TKE scheme dominates the background term, while it is also likely that over much of the ocean the numerical mixing in the model's advection scheme is at least as large as that associated with the 1.2 x 10^{-5} m^2s^{-1} explicit background diffusivity, as discussed in Griffies et al., (2000) and Lee et al. (2002)."

P5769 l26: *where is the freshening in the Atlantic Ocean? In the bottom layer? Is a 0.05psu decrease really negligible?*

This remark concerns the effect of double diffusion and geothermal heat flux (the effects of which we have now separated). The freshening we refer to is in the surface layer, which is consistent with increased upper-ocean mixing.

We wrote "Adding double diffusion (also not shown) again has relatively little effect on the surface temperature, apart from a small localised cooling along the path of the ACC by 0.05°C, but does produce a freshening of 0.05 psu over much of the Atlantic and the subtropical Pacific."

P5770 l3: *Is there really a change in the pathway of the deep western boundary current with BBL. In this case you have to illustrate that point.*

We have removed this sentence, since it would require further and quite complex analysis to justify and we think this is beyond the scope of the present paper.

<u>Section 6:</u>
P5773 l1: *can you explain what is a PEG. What kind of implication on which performance? can you add comment on this interesting point?*

We mention here that "Process Evaluation Groups (PEGs) have been set up within the JOMP programme specifically to address issues relating to the two aforementioned regions".

We have removed the reference to the diapycnal mixing PEG, since its topic is unlikely to have much effect on the time scales discussed in this paper.

<u>Technical corrections:</u>

P5766 l9: *the later configuration is GO5.0?*

(this is actually on P5765). Yes, we have amended the text to state this explicitly.

P5768 l26: *V3.2 instead of 3.4*.

This has now been corrected.

P5785 table2: *what is the column "UM job id"? Add in the legend definition of the NEMO namelist parameters.*

We have added the following to the caption: "The UM job id is a unique identifier for each experiment within the Met Office Unified Model system, and allows any given configuration to be replicated by another user."

We have also added to the caption brief descriptions of all the parameters in the table.

P5786 table 3: *sign of the transport in the indonesian throughflow*.

This was incorrect in the original draft, and has now been corrected. We have adopted the convention of positive sign for northward and eastward flows, so the sign is negative in both model and obs. We have mentioned this in the text.

P 5790 Fig4: *no data for GO1 from 1984 to 1991? what is the time period for the mean on figure b ?*

Unfortunately some velocity output files are missing for years 1985-1990 inclusive of GO1), and we have noted this in the figure caption.

Panel b is a mean over the last 10 years (1996-2005). This is now stated clearly in the caption.

Other changes

We have made the following additional changes to the paper:

Acknowledgments: the following sentence was added, consistent with policy: "We acknowledge use of the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council."

We have added the following reference for the geothermal heat flux data used: Stein and Stein (1992); and two additional references to describe the TKE scheme used in the model: Axell (2000) and Craig and Banner (1994).

We have moved the two paragraphs describing the initialisation and surface restoration and the surface forcing used in the integrations from Section 2 (Model Description) to Section 4 (Experimental Design), since these are specific to the ocean-only experiments described in this paper, rather than being intrinsic to the GO5.0 model configuration. We have also two subdivisions to Section 4, on Model initialisation and forcing, and Model integrations, for clarity.

We have removed the reference to the older NEMO manual (Madec et al, 1998), since the newer one (Madec, 2008) also includes the full description of the TKE scheme.

We have carried out an extra 10-year integration of the model to test the individual effects of the geothermal heat flux and the bottom boundary layer, and have amended the relevant text, as well as adding this run to Table 2.