Response to Referee #2

In the present manuscript, the authors summarized basic model performance/drifts of EC-Earth3P-HR in comparison with lower resolution version, EC-Earth3P, together with optimization procedure of model code, data handling, how to post-process. The manuscript is well-organized, and basically I consider that the present manuscript will be worth publishing in GMD. In general, however, physical explanations on causes of model biases, drifts, and differences between EC-Earth3P-HR and EC-Earth3P are quite limited throughout the manuscript. After minor revisions in order to answer the suggestions and comments listed below, the manuscript will be more suitable for publication.
We thank the reviewer for these positive remarks.

Comments
L. 373-374: How did you generate atmospheric temperature perturbations? Gaussian random noise with a certain amplitude? Please specify the method.
3D temperature perturbations are random samples from a uniform distribution over [-5e-5, +5e-5] degree. We have added this information in the manuscript.

L. 380: How did you change the oceanic mixing parameters? Please describe more details.
We added more details:

….It was therefore decided to change the ocean mixing parameters, which improved the AMOC. The main difference compared to the first ensemble member of EC-Earth3P is that the parameterization of the penetration of turbulent kinetic energy (TKE) below the mixed layer due to internal and inertial waves is switched off (nn_etau=0; Madec et al. 2016). The mixing below the mixed layer is an ad-hoc parameterization into the TKE scheme (Rodgers et al. 2014,) and is meant to account for observed processes that affect the density structure of the ocean's boundary layer. In EC-Earth3P, this penetration of TKE below the mixed layer caused a too deep surface layer of warm summer water masses in the North Atlantic convection areas which lead to a breakdown of the Labrador Sea convection within a few years and a strongly underestimated Atlantic Meridional Overturning Circulation (AMOC) in EC-Earth. An additional minor modification compared to ensemble member 1 is an increased tuning parameter rn_lc (=0.2) in the TKE turbulent closure scheme that directly relates to the vertical velocity profile of the Langmuir Cell circulation. Consequently the Langmuir Cell circulation is strengthened.

Rodgers, K. B., O. Aumont, S. E. Mikaloff Fletcher, Y. Plancherel, L. Bopp, C. de Boyer Montégut, D. Iudicone, R. F. Keeling, G. Madec, and R. Wanninkhof, 2014: Strong sensitivity of southern ocean carbon uptake and nutrient cycling to wind stirring. Biogeosciences, 11 (15), 4077–4098, doi:10.5194/bg-11-4077-2014, URL HTTP://www.biogeosciences.net/11/4077/2014/.

Madec and the NEMO team 2016: NEMO ocean engine version 3.6 stable. Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace No 27, ISSN No 1288-1619.

L. 399-408: There are almost no explanations on cause of model biases described here. Please give possible reasons or speculations for the biases which may be arisen from, for example, deficiencies in parameterizations for cloud microphysics, (deep,

shallow, strat) cumulus, insufficient horizontal resolution, albedo parameterization of snow, sea-ice, etc.
Without a detailed analyses of the origin of the biases it is difficult to know the causes. We have, however, included a discussion about the possible causes.

L. 411: Why the MSLP over Antarctica is higher (worse) in EC-Earth3P-HR than ECEarth3P? Please give possible reasons or speculations for the biases. In addition, if the biases in stationary eddies (Fig. 9) MSLP (Figs. 7 and 10) are evaluated, you may want to show wintertime storm track activity defined as subweekly eddy meridional temperature flux at the 850 hPa for EC-Earth3P-HR than EC-Earth3P, which may be useful for interpreting differences of model biases between two models.
We agree that this difference is remarkable and unexpected. More analyses is required to fully understand this. The difference is most strongly in the austral winter, which suggest that it is related to the dynamics of the polar vortex that is sensitive to the horizontal resolution. We have added a few lines in the text.

We agree that storm track activity is a useful diagnostic, but because this article focuses on basic validation this will be explored more in detail in future papers.

Figure 6, 7, and 8: In order to evaluate model errors quantitatively, please calculate root-mean-squared errors (RMSE) for EC-Earth3P-HR with respect to observations/ reanalysis and add the RMSE to somewhere in the corresponding figures, for example, just right of the figure title as "Diff DJF SAT EC-Earth3-HR ERA-Int (0.8 K)".
The global mean RMSE have been calculated and are mentioned in the figure captions, together with the global mean values.

Figure 10: Panels showing difference between EC-Earth3P-HR and EC-Earth3P may be replaced by the errors between EC-Earth3P and observations/reanalysis as in Figs. 6-8. And, RMSEs for EC-Earth3P may be given.
Because EC-Earth3P-HR and EC-Earth3P have similar error patterns compared to observations/reanalysis we have chosen to show the differences between EC-Earth3P and EC-Earth3P-HR to highlight the impact of resolution. The similarity in error between EC-Earth3P and EC-Earth3P-HR is also reflected in similar RMSE values. These are now mentioned in the figure captions and are also discussed in the text.

Figure 12: Top label "2040-2049 minus 1950-1959" may be wrong.
The label may indeed be confusing. We have removed it and expanded the figure caption to remove this confusion.

L. 444: Why does not the activation of deep convection in EC-Earth3P-HR occur and why do global-mean SAT and AMOC transport keep stable in EC-Earth3P-HR?
We do not know the cause yet. A few possible reasons are the impact of ocean resolution on the sea-ice dynamics and deep ocean convection. But also changes in the ocean temperature and salinity structure might play a role. This is presently under investigation. We have added a few lines in the text to discuss this.

Figure 14: For comparison, Figs. 14c and 14f may be replaced by observations. Also Z500 anomalies regressed onto NINO3.4 index, which can be superimposed onto Fig.

14de by contours, are useful for evaluating atmospheric teleconnection pattern.
Thanks for these two points. We have added both to Fig. 14, namely the analysis of observations (HadISST for SST and ERA-Interim for Z500) and the regression of Z500 onto Niño3.4 to report the ENSO teleconnection to the Northern Hemisphere. We have also modified the text accordingly.

L. 473-476: SST variability is closely related to the frontal structure seen in the climatic mean SST. So, you may want to add DJF climatic-mean SST to Figs. 14a-c by contours.
This is also a very relevant point; thanks. We have now included in Fig. 14-top the winter SST climatology for both, the model versions and HadISST.

Figure 15: Again, please add the corresponding panels for observations/reanalysis.
We have added the analysis of ERA-Interim to Fig. 15, and modified the text accordingly. Note that we have removed the citation of the observational paper in that passage.

L. 512: Please capitalize "Rapid".
Done.

L. 525: ERA-Interim is just reanalysis data, not observations. You may want to redraw the green lines in Fig. 18 based on observations, for example, HadCRUTv4.4 with keeping consistency of undefined grid point between observations and models. And then, please rewrite Section 4.2.4.
The undefined grid points of HadCRUTv4.4 at 1950 cover a large fraction of the globe as shown below. We argue that comparing the model over this limited region is not representative for the global warming of the model during the hist-1950 run. Although we recognize that ERA-interim is just reanalysis data it provides a reasonable estimate of the global warming. We have replaced the ERA-interim data by ERA5 in Fig. 18 for obtaining the most up to date reanalysis estimate.

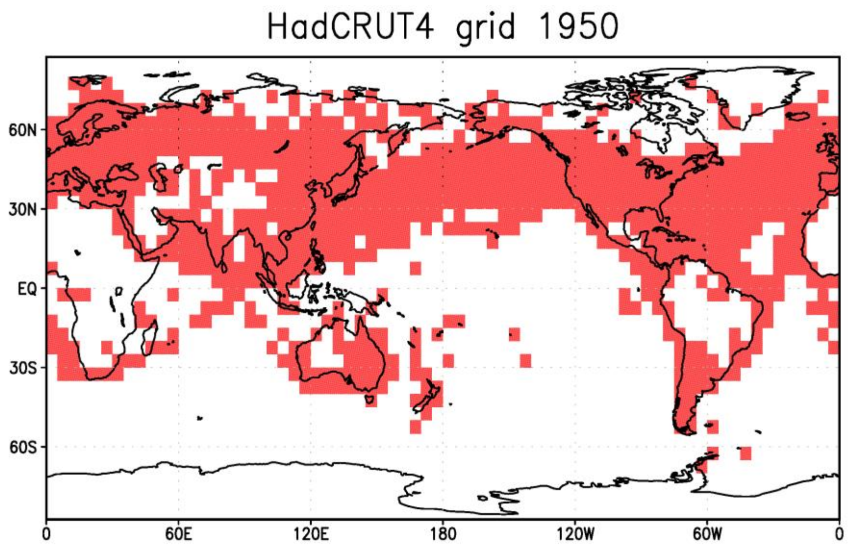

Figure R2. Red: Grid points with T2m observations of HadCRUT4 for 1950.