Referee: 2

This is a nice study than analyses and compares important biases in PRIMAVERA models for high and low resolution. The ensemble is small but the authors spend sufficient time to discuss the differences between the models.

We thank the Reviewer for the thoughtful comments. In the following we answer each specific point (in blue).

Another issue is the separation between the impact of ocean and atmosphere resolution. The biases are often the result of coupled interactions and the present set of simulations make it difficult to separate the role of ocean and atmosphere resolution. I have sometimes the feeling that regarding the role of atmosphere or ocean resolution the authors jump too quickly to conclusions, thereby relying on other literature. I would prefer that they are a bit more cautious, if their conclusions cannot be without doubt supported by their own analyses. For instance I am not fully convinced that increasing atmosphere resolution is the main reason for the reduction of the warm bias in the upwelling regions as stated in the abstract. No clear analyses for that are provided.

Our conclusion on the benefit of increased atmosphere resolution derives from three models in particular, the ECMWF-IFS, HadGEM3-GC31 and MPI-ESM2-1. For the ECMWF-IFS model, C.D. Roberts et al. (2018b) show that the warm bias over the eastern tropical regions increases with increased ocean resolution between their MR (not included in our analysis) and LR resolutions but it is reduced with increased atmosphere resolution between the MR and the HR resolutions (see their Figure 3). For the HadGEM3-GC31, Roberts et al. (2019) compare model versions at different atmosphere and ocean resolutions, finding that the increase in atmosphere resolution from ~250 km to ~50 km reduces the warm bias over the eastern tropical regions, but the increase in ocean resolution from ~100 km to ~8 km shows no benefit there (see their Figure 7). Lastly, for the MPI-ESM1-2, only the resolution of the atmosphere increases between the LR (134 km) and HR (67 km) models in our analysis (Table 1 in the main manuscript) and therefore any bias reduction is due to increased atmosphere resolution. Bias reductions over the upwelling regions cannot be attributed to the increase in resolution in the other models (CNRM-CM6-1, EC-Earth3P) because i) the resolution increases in both the atmosphere and ocean between LR and HR, ii) there are no intermediate resolutions to test our hypothesis. This line has nonetheless been removed from the Abstract and has been clarified in the Discussion and Conclusions Section as “As for the increase in atmosphere resolution alone, it contributes to reducing the warm bias over the eastern tropical oceans in the ECMWF-IFS (Roberts C.D., et al., 2018b), HadGEM3-GC31 (Roberts M.J. et al., 2019) and MPI-ESM1-2 (this study) coupled models.”

One of my main concerns is the abstract which to my opinion does not reflect very well the main conclusions of is even in contradiction. In the conclusions it is written: "On average (i.e., in the

ensemble mean), the warm eastern tropical ocean, the double ITCZ, and the cold North Atlantic improve at higher resolutions, while the SO warm bias worsens or persists in some models, and a new warm bias emerges in the Labrador Sea in all the models as a result of excessive Atlantic ocean heat transport (Roberts M.J. et al., 2020b) and excessive ocean deep mixing in the Labrador Sea in NEMO models at a 0.25° resolution (Koenigk et al., 2021)." This is a fair summary of the results. I do not see this reflected in the abstract, instead it speculates too much about the role of the atmosphere or ocean and the need of eddy rich ocean modelling. This can be discussed in the discussion section, but the abstract should be mainly limited to the results obtained from the analyses. I urge the authors to modify the abstract and make it more coherent with the main text.

We thank the Reviewer for raising this issue. We have reworked the Abstract entirely to remove any discussion and highlight the results:

"**Abstract.** We examine the influence of increased resolution on four long-standing biases using five different climate models developed within the PRIMAVERA project. The biases are the warm eastern tropical oceans, the double Intertropical Convergence Zone (ITCZ), the warm Southern Ocean, and the cold North Atlantic. Atmosphere resolution increases from ~100–200 km to ~25–50 km, and ocean resolution increases from ~1° (eddy-parametrized) to ~0.25° (eddy-present). For one model, ocean resolution also reaches 1/12° (eddy-rich). The ensemble mean and individual fully coupled general circulation models and their atmosphere-only versions are compared with satellite observations and the ERA5 reanalysis over the period 1980–2014. The four studied biases appear in all the low resolution coupled models to some extent, although the Southern Ocean warm bias is the least persistent across individual models. In the ensemble mean, increased resolution reduces the surface warm bias and the associated cloud cover and precipitation biases over the eastern tropical oceans, particularly over the tropical South Atlantic. Linked to this and to the improvement in the precipitation distribution over the western tropical Pacific, the double ITCZ bias is also reduced with increased resolution. The Southern Ocean warm bias increases or remains unchanged at higher resolution, with small reductions in the regional cloud cover and net cloud radiative effect biases. The North Atlantic cold bias is also reduced at higher resolution, albeit at the expense of a new warm bias that emerges in the Labrador Sea related to excessive ocean deep mixing in the region, especially in the ORCA025 ocean model. Overall, the impact of increased resolution on the surface temperature biases is model-dependent in the coupled models. In the atmosphere-only models, increased resolution leads to very modest or no reduction in the studied biases. Thus, both the coupled and atmosphere-only models still show large biases in tropical precipitation and cloud cover, and in mid-latitude zonal winds at higher resolutions, with little change in their global biases for temperature, precipitation, cloud cover, and net cloud radiative effect. Our analysis finds no clear reductions in the studied biases due to the increase in atmosphere resolution up to 25–50 km, in ocean resolution up to 0.25°, or in both. Our study thus adds to evidence that further improved model physics, tuning, and even finer resolutions might be necessary."

Specific comments:

L55. Here it is clear which biases are analyzed. This was not clear from the abstract.
The biases are now mentioned in the Abstract as well.

For the upwelling regions, I miss a discussion about the role of ocean mixing. For the Atlantic upwelling region see for instance: https://doi.org/10.1175/JCLI-D-19-0608.1
We have reshaped Section 1.1.1 to include this and other important references.

"This bias has long been related to the underestimation of the cloud cover, which leads to warming because of excessive shortwave radiation reaching the surface (e.g., Huang et al., 2007; Hu et al., 2008). The warm bias, in turn, weakens the lower tropospheric stability and thus hinders the formation of the stratocumulus deck, which contributes to sustaining the surface warm bias. Other mechanisms have been proposed to explain this bias, including too weak equatorial and alongshore winds weakening upwelling (e.g., Richter et al., 2012; Koseki et al., 2018; Goubanova et al., 2019; Voldoire et al., 2019a), biases in regional atmospheric moisture (Hourdin et al., 2015), too weak offshore transport by ocean mesoscale eddies, and the misrepresentation of the coastal current system (Xu et al., 2014) or vertical mixing in the upper ocean (e.g., Hazeleger and Haarsma, 2005; Exarchou et al., 2018; Deppenmeier et al. 2020)."

L 180. I was surprised that ERA-Interim analyses were used and not the more recent ERA5.
ERA5 is now used for ERA-Interim. The results remain the same, which is mentioned in the revised manuscript: "Biases in SAT and zonal winds with respect to the ERA-Interim reanalysis (Dee et al., 2011) are very similar to those with respect to ERA5 (not shown)."

L 230. I do not think that on the basis of one eddy rich model and no dedicated analyses between increase of ocean and atmosphere resolution, you can make that statement.
As previously mentioned, this is supported by the analysis in Roberts M.J. et al., 2019, in which model versions of the HadGEM3-GC31 at different atmosphere and ocean resolutions are compared. The authors find that increased atmosphere resolution tends to lead to reduced temperature biases over the upwelling regions, but increased ocean resolution offers no major benefit (see their Fig. 7).

L 376. Why not mention this in the abstract?
This is now included in the Abstract.

L 420. This statement seems to contradict for instance with L364 and L376. The reduction of the cold bias in the sub-polar gyre in the North Atlantic is one of the strongest signals between LR and HR. Also, the reduction of the double ITCZ bias in the Pacific is a clear signal that the

authors partly attribute to the increase in the ocean model resolution (L290). So, this statement is not backed up by the authors own analyses. This statement is then lifted to the abstract, where it should be removed. It can be discussed in the discussion section with a reference to the statement at L370 where eddy-rich ocean models improve the Gulfstream separation and reduce the warm bias near the coast.
The line has been removed entirely.

Typo's
L216: … small impact… The structure of the sentence suggests positive is missing between small and impact.
Increasing resolution does not necessarily reduce global biases (positive impact); it can also increase the bias (for example, as for the surface temperature and precipitation in the ensemble mean). Nonetheless, increased resolution has most of the times a small impact in the global bias, as stated in the manuscript.

L303: Referring only to Fig. 1 and not also to Fig. S2 reads strange in the first sentence.
Corrected.

L348: ... compared to...
Corrected.

L425 ..helps to reduce..
Rephrased.