Response for Referee #2

We thank the referee for his comments and suggestions. We will strive to address each specific concern in detail.

Minor comments:

Comment 1: Diurnal cycle and discretization choices

If I get it correctly, for simulating CO you decided to choose a model setup that uses climatological forcing that does not resolve diurnal light cycle. Forcing data like NCEP or ERA interim however do resolve SW radiation fluxes on a 6 or 3 hourly basis. Given that you are using a sophisticated wave-length dependent CO production mechanism omitting on the other hand this feature seems like an odd decision on complexity versus simplification. As production and decay rates are highly uncertain, does resolving wave length dependency and omitting the day night cycle (thus also dark production, as described in Day and Faloona 2009, JGR Oceans 114) imply you expect the former to propagation stronger into the presented mean solution than the latter, i.e. more important to resolve in a realistic CO production scheme?

Reply: Indeed, forcing data like NCEP or ERA would enable us to resolve the diurnal cycle in terms of forcing fields. However, our choice for not including an explicit diurnal cycle for the CO photochemistry in our model set-up was mostly dictated by the consistency with the existing biogeochemical simulation (Aumont et al., 2015) in which we embed the CO module. In the standard version of NEMO-PISCES, as described in Aumont et al. (2015), the lack of an explicit diurnal cycle is justified by the fact that:

- the simple phytoplankton model that is used (constant stoechiometry for N and C) does not enable to explicitly represent the decoupling of the plankton internal carbon and nitrogen cycles that is a main feature of the plankton diurnal cycle (Flynn and Fasham, 2003).
- the vertical resolution in the upper part of the ocean (10 m for the first layers) does not enable to adequately represent the diurnal variability of the mixed layer depth.

We thought it would have been inconsistent to explicitly resolve the diurnal cycle for the photochemical part, but not for the biological part.

That said, because the photochemical production of CO is described, in our model, as a linear function of irradiance, the average photoproduction term over 24h is probably not very different from what we would have obtained from resolving the diurnal cycle.

These considerations will be added to the Discussion session. The potential inclusion of dark production is discussed in more details in the response to Referee 1. In brief, we plan to include the parameterization by Zhang et al. (2008) in our model set-up, to conduct some sensitivity experiments to be able to discuss the potential underestimation of CO production in our standard model set-up.

Comment 2: Seasonal cycle

How does the seasonal cycle of CO look like in the model? Is the quality of the model solution different for different seasons, i.e. how does it relate to the models' ability to represent the seasonal cycle of Chla?

Reply: We will add to the revised manuscript a brief description of the seasonal cycle of CO (figure and text).

For the evaluation of the Chla solution, we will recall in the discussion section the main conclusions from Aumont et al. (2015) and will include in the supplementary material a few plots comparing surface chlorophyll to estimates from available data-sets. Briefly, in Fig. 7 of Aumont et al. (2015), the simulated distribution is compared to GlobColour satellite observations for 2 seasons: Apr-May-Jun and Nov-Dec-Jan. The observed patterns are qualitatively reproduced by the model. Too low concentrations are simulated in the subtropical gyres, which is attributed to the lack of phytoplankton acclimation to oligotrophic conditions or to the assumption of constant phytoplankton stoichiometry. Chlorophyll concentrations are quite strongly underestimated in the equatorial Atlantic and in the Arabian Sea. In the latter region, mesoscale and submesoscale processes have been shown to be of critical importance. In two of the three main HNLC regions, i.e., the equatorial Pacific and the eastern subarctic Pacific, the model succeeds in reproducing the moderate chlorophyll concentrations. In the Southern Ocean, the third and largest of the principal HNLC regions, chlorophyll concentrations appear to be strongly overestimated by the model when evaluated against satellite-derived observational products, especially during summer. Furthermore, the increase in phytoplankton in late spring and early summer occurs too early.

Comment 3: In situ data processing

3.1. In the evaluation of the concentrations you are using model data collocated to observational data. Does this mean you are using individual grid cells? If so, are these representative for a larger surrounding area – did you consider averaging several grid cells, as physical features such as the extent of subtropical gyres, location of fronts etc. are not geo-

referenced, i.e. collocated with real world conditions during the ship cruises? In particular, for the vertical profiles it could be useful instead of showing one profile adding its variability taking into account several neighbouring cells (and eventually temporal standard deviations for within the averaging period).

Reply: The referee is right. We are using individual grid cells of the model output, which are co-located with the observations (this will be made explicit in the revised manuscript).

We did not consider averaging several grid cells as the model resolution is already very coarse (2°x2°). It would have been interesting to consider adjacent cells to capture specific physical features that are not exactly geo-referenced. However, this is limited first by the fact that we don't have access to the physical features associated to the in situ data and second by the fact that we are comparing model output which are climatological (with no inter-annual variability) with in situ data measured for a specific year and month.

3.2. Given that the compilation of observational data is presented to be unique and its averaging methods are very diverse I suggest to extend the section observations. E.g. it would be interesting to learn, also in light of the large temporal variability of CO concentrations, if certain months/seasons are better resolved in the observations than others. Now I can deduce this only from the tables, but do not get any direct information in the main text.

Reply: yes, we will better present and discuss the spatial and temporal coverage of the in situ data in the Materials section, instead of presenting it only in the tables.

3.3. It would be important to know on how many values per averaging period the temporal means are based, as e.g. the observed diurnal cycle is very strong and party not symmetric (afternoon maxima). Furthermore, I do not know if CO measurement techniques are comparable across the different observational sources wrt the limit of quantification / detection, or if there was development in the methodology from the 1970s to now. Are all of the published observational data equally reliable, in particular wrt to low CO values?

Reply: It is difficult to gather in situ data from literature as the quality of the metadata associated with measurements are very heterogeneous. For example, information about the time of the day at which the measure had been made is not always given, or in other cases the CO concentration given in the literature is already a mean upon a few measures taken during the day. Furthermore, it is true that the different data sets might not be equally reliable as the measurement's techniques have evolved since the late 70's, but this issue is not easily quantifiable. In the tables, we did our best to inform the reader about the origin and significance of the data used. We would like to emphasize that it is the first effort to gather observations from the first cruises in the 70' up to now in a CO observation-based data set.

Comment 4: Choices of the manuscript structure –

I struggled with the structure of the manuscript, presenting first simulation results and evaluation of the "standard" experiments followed by a separate discussion of the sensitivity experiments. Statements as "indicating again a possible bias in the production process" in the section on the standard experiment could be easily complemented by the results of the sensitivity studies on process parameterization, instead of having to collect this information later in the manuscript. Also this separation of the evaluation of surface concentrations leads to inconsistent level of discussion of potential sources of discrepancies: Whereas in the standard experiment it was argued that missing processes related to sea ice or a missing spatial variability of the decomposition rate might be causes of discrepancies in polar regions, only later in the section of CO production the authors state that also missing terrestrial CDOM sources might be a source of model data discrepancies. It is not clear to me how it was decided which of the parameterizations are chosen to be "standard" vs "alternative". For example, I understand that choosing the Launois et al. 2015 CDOM parameterization leads to high CO production and in combination with a consumption rate of 0.2 d-1 CO concentrations get too high compared to observations. On the other hand, using the "standard" parameterization together with a consumption rate of 1 d-1 lead to very low CO concentrations. The combination of the Launois et al. 2015 CDOM parameterization and a fast consumption rate was however not tested or presented without commenting on the reasons for this. If instead the authors would present all of the tested parameterizations in the methods part equally, present first an evaluation of the model results wrt to the range of parameters chosen and process parameterizations, and resulting from this discuss the source and sink budget, and emissions only for the most successful configuration, decisions taken and its consequences could become clearer.

Reply: We did initially chose a structure showing equally the different simulations with their respective evaluation. However, this structure was cumbersome as we had to evaluate against oceanic CO concentrations each simulation. Moreover, it did not permit to clearly highlight our "best guess" simulation (which we called "standard"), that we propose for use by the chemical atmospheric community. Indeed, considering the different tests performed and parameterizations, we think our standard simulation to be the best based on its CDOM parameterization. The section 'Sensitivity to alternative parameterizations' should be considered as a presentation of the possible range around the standard simulation. Simulations with a combination of the Launois et al. (2015) CDOM parameterization and a faster consumption rate, or with Preiswerk and Najjar (2000) with a slower consumption rate were indeed tested (but not included in the current version). Considering the present comment, we

plan to add and discuss these other tests in the next manuscript version.

Comment 5: Lack in the physical model evaluation

The model evaluation is lacking discussion of the simulated physical ocean solution, e.g. whereas contributions of MLD are mentioned in possible causes of discrepancies of modelled and observed CO concentrations these are not compared to the ship cruise or climatological T, S, or MLD data. I guess it would be possible to get the CTD data of the ship cruises and compare them. It would be useful to see in particular in the analysis of the vertical profiles, but also for the surface data whether how NEMO performs in regions with CO data.

Reply: As mentioned in reply for comment 2 for the Chla solution, in the supplementary material of the revised manuscript, we will include a few plots comparing SST and Mixed Layer Depth for different seasons to estimates from available data-sets. The dynamical state of NEMO used here is partly evaluated elsewhere: e.g., Mixed Layer Depth in Southern Ocean (Person et al., 2018), Water masses and transport (Iudicone et al., 2016), but no one publication offers an extensive evaluation of the physics used here. However, for the in situ data collected, it is difficult to gather the corresponding physical features as most of the time the CTD data of the ship cruises are not given along with the CO concentrations.

Comment 6: Constant atmospheric CO

The authors assume a constant homogeneous atmospheric mixing ratio of CO in their emission calculation. As major sources of CO (fossil fuel combustion, biomass burning) are on land and a major sink is reaction with of a large hemispheric and seasonal variation of CO mixing ratios in air due to the continental distribution and OH seasonality is expected. Is the over-saturation of the ocean indeed that strong that these variations can be omitted in the emission calculation?

Reply: Tests had previously been carried out on the atmospheric CO but were not shown in the submitted version of the manuscript. These tests show that the value of the atmospheric CO dry mole fraction is of little influence on the oceanic emission. For example, using a homogeneous and constant CO dry mole fraction of 45 pptv leads to a global oceanic CO emission of 3.7 Tg C yr^{-1} (against 3.6 Tg C yr^{-1} using 90 pptv). We will add these results in the next manuscript version.

Comment 7: Clarification of a few points

- p8 l23: ... same forcing fields as the ones in Aumont et al. 2015. Please help the reader to easily understand implications of your model setup into the results by repeating main

characteristics of that forcing (source, spatial and temporal resolution).

Reply: The main characteristics of the forcing will be better detailed in the next manuscript version.

- p10 l10: … all vertically integrated over the upper 1000 m. The vertical profiles suggest that below the euphotic zone there is not much CO left, why do you choose to integrate over 1000 m?

Reply: Yes, the integration upon 1000 m is nearly equivalent with an integration over the whole water column as almost no CO remains under that depth. This will be made clearer in the revised version.

- p12 l24 … which can be related to differences in the light penetration and mixed layer depth. Please be more specific… is the simulated MLD generally too low/high? Do you indicate a different mixing scheme would improve the profiles? How does the model's vertical resolution in the upper ocean affect the vertical profiles?

Reply: see reply to comment 5.

Comment 8: CDOM comparisons

P13 l17: … those minimums are best represented by the relation of Morel since the ones in Launois … give too high and too low values. Could you please comment on the simulated Chl field here, so that it gets clear that these CDOM parameterizations are responsible for the discrepancies rather than the simulated Chl. Furthermore, satellite derived observations are based on a number of assumptions (e.g. also wrt to light penetration depth in turbid and non-turbid waters) and models (bio-optical, atmospheric correction…), in particular in derived product as CDM. Furthermore, the discussion of the evaluation could be more detailed in discussion the quality of the Chla solution, which influences CO production.

Reply: we will better discuss the Chla evaluation as mentioned in the reply to comment 2. This will help to clarify the respective roles of CDOM parameterization versus Chla representation for the CDOM representation. In particular, we will complete Fig. 8 by adding a plot showing modelled and satellite-derived Chla as a function of latitude.

Comment 9: Inter-annual variability

Conclusions ii) […] the model does not consider the inter-annual variability of ocean physics and biogeochemistry […] Even with a climatological forcing both ocean physics and biogeochemistry solutions will show inter-annual variability due to e.g. fluid dynamics (wave propagation) and different plankton over-wintering stocks. Clearly this variability will be lower

than in simulations using transient forced boundary conditions or coupled atmosphere-ocean simulations. I suggest to rephrase this statement slightly.

Reply: We may not have been clear enough in the description, of the simulation protocol. In fact, we use an offline configuration of NEMO-PISCES model, i.e. PISCES is run using a climatological ocean dynamical state obtained from a previous NEMO physics-only simulation. We plan to better explain the experimental setup in the next manuscript version. In brief, the dynamical state of the ocean has been simulated with NEMO 3.2 (Madec, 2008) under climatological forcing as described in Aumont et al. (2015). The ocean model is first spun up for 200 years starting from the climatology of Conkright et al. (2002) for temperature and salinity. We then use the last year of these dynamical fields at 5-day temporal mean resolution (ocean currents, temperature, salinity, mixed layer depth, surface radiation…) to force the biogeochemical model, which is then spun up for 3000 years. We then turn on the CO module for an additional 2 years and analyse the last year of the simulation.

By construction, we do not simulate any inter-annual variability: the atmospheric forcing fields are climatological fields, the ocean dynamical state is also climatological as well as our biogeochemical simulation. In case of high spatial model resolution (> ¼°), it has been shown that an ocean-only simulation forced by climatological atmospheric forcing fields could indeed show some intrinsic variability at the inter-annual scale (Sérazin et al., 2018). This is clearly not the case at the typical resolution we're running our ocean-biogeochemical model.

Comment 10:

Fig.7: I find the colour coding of the lines confusing... is it time or location, what does it show in the subplot 'Swinnerton and Lamontagne 1973'?

Reply: Inside each subplot, one colour is for one profile. The dash line shows in situ measured profiles whereas full line shows the model profile co-located in space and time with the measured profile. This will be made clearer in the figure's legend.

Bibliography:

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci Model Dev, 8, 2465–2513, doi:10.5194/gmd-8-2465-2015, 2015.

Conkright, M. E., R. A. Locarnini, H. E. Garcia, T. D. O'Brien, T. P. Boyer, C. Stephens, and J. Antononov, World Ocean Atlas 2001: Objective Analyses, Data Statistics and Figures, CD-ROM Documentation, Tech. rep., National Oceanographic Data Centre, Silver Spring, MD,

USA, 2002.

Flynn, K.J., and Fasham, M.J.R. (2003). Operation of light–dark cycles within simple ecosystem models of primary production and the consequences of using phytoplankton models with different abilities to assimilate N in darkness. J Plankton Res 25, 83–92.

Iudicone, D., K. B. Rodgers, Y. Plancherel, O. Aumont, T. Ito, R. M. Key, G. Madec, and M. Ishii, 2016: The formation of the ocean's anthropogenic carbon reservoir. Sci. Rep., **6**, 35473, doi: https://doi.org/10.1038/srep35473.

Launois, T., Belviso, S., Bopp, L., Fichot, C.G., Peylin, P.: A new model for the global biogeochemical cycle of carbonyl sulfide - Part 1: Assessment of direct marine emissions with an oceanic general circulation and biogeochemistry model, Atmos. Chem. Phys., 15, 2295–2312, doi:10.5194/acp-15-2295-2015, 2015.

Madec, G.: "NEMO Ocean Engine", Note du Pôle de Modélisation 27, Institut Pierre-Simon Laplace (IPSL), France, http://www.nemo-ocean.eu, 2008.

Person, R., Aumont, O., and Lévy, M. (2018). The Biological Pump and Seasonal Variability of pCO2 in the Southern Ocean: Exploring the Role of Diatom Adaptation to Low Iron. Journal of Geophysical Research: Oceans 123, 3204–3226.

Preiswerk, D. and Najjar, R.G.: A global, open-ocean model of carbonyl sulfide and its air-sea flux, Global Biogeochem Cycles, 14, 585–598, doi:10.1029/1999GB001210, 2000.

Sérazin, G., Penduff, T., Barnier, B., Molines, J.-M., Arbic, B.K., Müller, M., and Terray, L. (2018). Inverse Cascades of Kinetic Energy as a Source of Intrinsic Variability: A Global OGCM Study. J. Phys. Oceanogr. 48, 1385–1408.