The manuscript by Buitenhuis and Coauthors describes the results of simulations with an ocean biogeochemical model that includes different parameterizations of N2O production, constrained with available N2O observations. The main finding is a net N2O outgassing to the atmosphere of 2.4 ± 0.8 TgN/year, which is substantially lower than many of the estimates previously reported, and also less uncertain. A very small proportion of the N2O production comes from denitrification-associated pathways in suboxic waters. The estimate also appears robust to the choice of the parameterization of N2O production.

This is a short and potentially useful paper, although not particularly original. But I think that, if better supported, the results will push other scientists to reconsider estimates of N2O emissions from the ocean (as well as from other sources) in light of the low values suggested. That said, I also think the paper is poorly written, in particular when it comes to the description of the methods employed - for example the model equations, the rational and choices behind the parameterizations, the steps behind the optimization. Furthermore, I have some additional concerns about the results that prevent me from fully supporting publication of the manuscript as is.

We thank the reviewer for the comments. We have tried to clarify the methodology throughout the manuscript.

Specific comments:

- The model formulation is quite hard to follow, partly because equations are not show. This makes it difficult at times to judge the validity of the model's assumption. I strongly encourage the Authors to show all the pertinent equations, either in the main text, or in an appendix.

We have added the full set of equations for nitrous oxide to the model description as supplementary material. This has all the equations, parameter values, and how it is set up. We have also made multiple clarifications in the text following the reviewer's comments. Please see point-by-point responses below (and in response to the other reviewers).

- The choice of some of the model equations and parameterization is unclear and could be better justified. The Authors could do a much better job explaining why certain functional forms have been utilized, and what consequences these choices may have, if any. For example, looking at Table 1, line 3 lists an equation that uses the logarithm of O2. Is there any reason for this form? The logarithm will expand the range at very low O2 concentration - do we trust O2 measurements there? Further, this form breaks down at O2=0; does this ever happen in the model/observations, and is there any limit applied to prevent it? Finally, all of these equation should represent a process ultimately limited by O2. Is there any limitation as O2 goes to zero?

On the specific justification for the use of logarithm of O2, the choice is due to a better fit to the data. We now explained this in the text:

"A logarithmic function fit the data better than linear, exponential, or power functions."

We have also added an explanation why the N2O consumption equation and parameters were used:

"The functional form of the O2 dependence of N2O consumption (suppl. Eq. 71) was the same as that of denitrification (suppl. Eq. 67), and with an O2 response function that is 1.5 µmol L^{-1} lower than that of denitrification, which is similar to that used by Babbin et al. (2015). We independently optimised the ratios of N2O production and consumption from denitrification (Section 3.1), which controls the net N2O production as a function of O2 concentration. There is not enough information at present to optimise the O2 concentration parameters of denitrification and N2O consumption as well."

The choice of model equations for the preferential algal uptake are justified in Vallina and Le Quéré (2008).

The lowest O2 concentration in the Bianchi et al. 2012 database (after regridding onto the model grid) is 1.15 µmol O2/L. The lowest concentration for which there is a yield measurement is 5.4 µmol O2/L. It is therefore true that the logarithm extrapolates the N2O/AOU ratio from 232 µmol/mol at 5.4 µmol O2 to 251 µmol/mol at 1.15 µmol O2. Given the variability in the measurements this is an insignificant extrapolation beyond the range of the measurements. Also, nitrification rate decreases with O2, so that the N2O production rate is low.

- I found the distinction between the prognostic and diagnostic model (for N2O) somewhat confusing. In both models N2O is carried as a prognostic tracer - except in the first model it does not depend on other N-cycle tracers, and is not consumed, but only produced and passively advected until it outgassed from the surface. What makes one model diagnostic and another prognostic?

In this manuscript we use the distinction between diagnostic and prognostic model to mean that the former is based on statistical relationships with observations, while the later is based on process understanding and representation. The N2O field from models are indeed transported in the same way. We clarified this in Section 2.6 as follows:

"N2O production is implemented as two distinct submodels. The diagnostic submodel is based on statistical relationships of DeltaN2O/AOU ratios taken from observations and has previously been published {Suntharalingam00,Suntharalingam12}.",

"The prognostic submodel presented here is based on process understanding and explicitly represents the primary N2O formation and consumption pathways associated with the marine nitrogen cycle (Fig. 1)."

and "The N2O concentrations from both the diagnostic and the prognostic model are transported in the same way by physical transport and the formulation of their gas exchange is also identical."

- Regarding the prognostic model - the Authors say that it explicitly represent the redox transformations that lead to the conversion of NH4+ to eventually N2O (actually only a subset, as for example, the model does not include NO2-), but the model seems to still parameterize them heavily. For example the current understanding is that N2O is an obligate intermediary during heterotrophic denitrification, so that one should expect a gross N2O production comparable to the denitrification rate (i.e. ~70 Tg N/year)

However, the Authors indicate a suboxic gross production of only 0.33 Tg N year – a very small rate in comparison. This may be explained by the use of "slopes" in the prognostic model that relates N2O to other tracers (more on these slopes in the next comment). This implicitly assumes a tight coupling between production and consumption at suboxic levels, with only a fraction of the N cycled by denitrification escaping to the water column. It may be fine as a parameterization - especially since it is one that is optimized against observations. However it may be problematic if the model is to be used under varying circulation or climate - the coupling between production and consumption may vary, and given the large gross N2O fluxes this may impact net production and accumulation of N2O.

The model only represents the process of denitrification, it does not represent a state variable for denitrifiers (first sentence of Section 2.4). Therefore, reactions that happen intracellularly in denitrifiers are not represented either, and gross production from denitrification represents N2O production that is exuded/leaks into the surrounding seawater and stays there long enough to leave a measured increase in N2O concentration. The small net production rate is a result of the optimisation against observations, there is no implicit assumption built into the model that production and consumption should be tightly coupled. We don't present climate change simulations here, so we cannot comment on whether using the present model for climate change projections would be more or less problematic than using any of the other available models.

- It is unclear what “slopes” are used in the prognostic model. Are these slopes actual yields (e.g. N2O production per NH4+ oxidized), relationships with O2 consumed, or just empirical relationships based on data syntheses? And what is then the slope of the third step of N2O cycling (consumption of N2O by denitrification)? Is it a relationship with O2, with NH4+ or with NO3-? (and specifically, is there explicit denitrification in the model, so that one could relate N2O production to NO3- deficit?).

We have changed slopes to ratios, and explained our reasoning in using the word ratios rather than yields at the end of the second paragraph of the introduction:
“Throughout the manuscript we will refer to N2O stoichiometries relative to O2, NH4 and NO3 as ratios, because they have been optimised against global databases of concentration measurements, rather than from microbiological yields. Using the latter would be more mechanistically satisfying, but the relevant yields are at present insufficiently constrained by observations.”

Yes, denitrification is explicitly represented, as stated in Section 2.4. N2O consumption is therefore proportional to NO3- consumption, as stated in Section 3.1. Yes, it would be possible to calculate a NO3- deficit, such as N*. We judged this to be outside the scope of the paper, because denitrification can be accompanied by both N2O production and consumption, so model validation of denitrification rate against observations of N* would not help constrain the N2O budget. We added supplementary material to this paper which contains a detailed description of the biogeochemistry model equations (taken from Le Quere et al. 2016, and now updated with a description of both N2O submodels (section 6.5 and 6.6).

- The lack of spinup in the model is worrisome: the model was apparently initialized in 1965 and run for 49 years through 2014. This is a short running time, and it completely misses a spinup phase. It may very well be the case that the N2O inventory of the ocean over the last 5 years is still adjusting from the initial condition, in a way that could bias the outgassing estimates. For example, there seems to be a substantial accumulation of N2O in the deep ocean - if this is still ongoing after 49 years, then the outgassing estimated by the Authors could be a lower estimate. A comparison between the total net production in the interior and the outgassing could give a sense of any disequilibrium. Note that a similar modeling study by Martinez Rey et al., 2013, BGS (incidentally finding about 4Tg/year emissions) suggested a 150-year spinup was not enough to eliminate drifts in N2O and other biogeochemical variables. Any drift should be discussed in the paper, and the consequences assessed.

We have added an analysis of the optimised N2O flux for the 2000-2004 and 2005-2009 periods, which show the same result. We note in Section 3.1 that N2O production below 1600 m, where there is an increase in concentration, is only 5% of the total production. Given the slow ventilation of the deep sea, this accumulation will have a negligible effect on the optimised flux, and keeping the simulation short actually helps with this. The frequency distribution of ΔpN2O in the submodels closely matches that in the observations (Fig. 12), which supports the conclusion from the small error attributed to the model-observation ΔpN2O mismatch, that the model does not have a major bias.

Martinez-Rey et al. do climate change simulations, and spin up the model so that they don’t have to include control simulations and present the climate impacts relative to the control. Our study is different, where we initialise from the available observations and optimize model parameters using the available observations to derive the present day oceanic N2O flux. See also our reply to the comment of Gianna Battaglia on Line 166.

- I found the description of the optimization steps very unclear. It took me a while to figure out what steps the Author follow and how the model is actually compared to the data, and I’m still not sure about them. Now my understanding is that a first optimization is carried out for the NH4-cycle using nitrification rates and NH4+ concentrations; then a second optimization is performed with interior N2O data to determine parameters for low-O2 pathways (but does this apply to both the prognostic

and diagnostic model?); and finally a third optimization (presumably with some parameters fixed by the previous steps?) using surface Delta-pN2O data for the global source terms, used to determine the final air-sea fluxes. That's my understanding but I am still not sure I got it right, and some aspects remain puzzling. I think this could be much better explained from the start, for example by a method section outlining the optimization strategy in more detail.

We do carry out 3 optimisations, but we split the presentation into two parts, one where we develop the model so that we can implement the prognostic model, and the other where we use the model to optimise the model to the two N2O datasets. We do not include NH4 concentration database in the optimisation because the high turnover rates and the many processes that are involved would make this a process that would require a whole paper by itself, which is outside the scope of the present paper. Fortunately, the many processes turn out to be reasonably well constrained by observations we present in this and previous papers (Buitenhuis et al. 2006, 2010), so that we judge the resulting NH4 concentration distribution to be fit for the purpose of optimising the N2O cycle which we undertake here. We have more explicitly described the progressive steps of how we use each observational dataset at the end of the introduction, from model development of the N-cycle in Section 2 to identifying N2O rates that best fit the observations in Section 3. See also the reply to the first comment of reviewer 2. We have clarified the legends of Fig. 9 and 11, see reply to reviewer 1 comment on Eq. 3. See also the reply to reviewer 2 on Section 2.2-2.8.

- Related to the previous comment, the equations for the optimization are absolutely opaque and unclear. They need to be substantially clarified: ideally anyone should be able to apply them after reading the paper, which is not the case. For example equation (1) is not very specific: instead of "average", "model", "observations" the actual mathematical form could be given - this would also help knowing how the average was done, wether the in situ or gridded data were used, how the model was sampled etc.

We have changed the mathematical form of Eq. 3 and 6 to replace average by the sum divided by the number of observations. We added Eq. 6 to give the actual mathematical form of the model and observation data used. We have added that the model was converted to the same grid as the observations, and sampled where there are observations in Section 2.3. See clarification added in response to Gianna Battaglia's comment on Line 199.

Similarly I am completely at loss with section 2.8, and I could not trace back the steps applied by the Authors based on this description alone. How is RSS/RRS_min (equation 3) used, how does it relate to the quantities shown in Fig. 8 and 10, and why does it only contain the number of observations but no information on the actual variables?

We explained how Eq. 5 (was Eq. 3) relates to Fig. 9 and 11 (was Fig. 8 and 10) in the legends of these figures: "MSE_{min} was obtained as the minimum of a second order polynomial fit (black lines). The 1σ confidence interval, where MSE equals the value calculated from Eq. 5, is indicated by the horizontal lines." We have added Eq. 6 to show how the actual variables (observations and model results) are included in the calculation of MSE (=RSS/n).

What does the "phi" term (equation 4) represent, and how is it actually used?

Because the paper we discussed only tested sample sizes that were more than 2 orders of magnitude smaller than our database, we decided to delete this equation and text.

- Regarding the final estimate of N2O air-sea flux, I think it could be couched much better into the context of previous estimates (also, a table would help), and what could be behind the potential discrepancies in light of the substantially lower revision. This could be especially interesting given that many modeling studies use a similar approach. The Authors also present an "observational" estimate of N2O production whose central value (4.6 TgN/year) is quite different than the final model estimate - this discrepancy could be added to the discussion. I am not particularly surprised by the lack of sensitivity of N2O production to the choice of diagnostic and prognostic models, since

both are optimized versus observations. Surface pN2O should be a quite powerful constraint to outgassing fluxes. However, one may still expect different sensitivities to interannual variability and climate change, so this is not a strong argument in favor of not resolving complex pathways that characterize the low-O2 N2O cycle.

We present the context of previous estimates in Fig. 4. We have added a discussion of the discrepancy with the Nevison et al. 1995 estimate using Wanninkhof piston velocity in the 4th paragraph of the discussion (paragraph starting "Despite these shortcomings"). Because the methods of other previous estimates are different, we can't give specific reasons why our results are different from the other estimates. The observational estimate in Section 2.1 is similar to (NOT quite different from) the combined model-observation estimate: the confidence interval of that estimate completely overlaps our better constrained estimate in Section 3.2. We have added this to the discussion:

"This estimate of global marine N2O production derived from analyzing the N cycle is statistically indistinguishable from the N2O flux derived from DeltapN2O observations, but has a much larger error."

We note that our estimate of the optimised N2O flux is sensitive to the observational dataset used, but not to the details of the model. Since our model parameters are optimised using a database spanning multiple years, and not on a year to year basis, we note that this model specification is more suited to estimating long-term or climatological fluxes, and not interannual variability.

- The model is biased in its representation of export and remineralization, as well as N2O distribution. The discussion of the effect of these biases (e.g. lines 289-300) is not especially thorough - so the conclusion, in particular regarding the narrower range of the new estimate, is not very convincing. Furthermore, there are hidden resolution biases. For example, the model can not resolve low-O2 coastal upwelling regions, which have been shown to be powerful conduits to N2O outgassing (e.g. Arevalo Martinez et al., 2015, Nature Geosciences). The abstract/conclusions could be more cautious with respect to the real uncertainties.

We have rewritten the discussion of the bias due to the too deep remineralisation, to more explicitly present the balance of evidence whether or not the underestimate of N2O concentrations at 500-1000m depth (Fig. 7) influences N2O flux at the surface:

"it should also be noted, first, that the optimization using surface ΔpN2O agrees with the optimization using N2O concentration that the contribution of the low O2 N2O production needs to be low (Fig. 11). Second, the error contribution from the model vs. observed ΔpN2O comparison is low, with confidence intervals of 0.3 Tg N y −1 for both submodels. Third, ΔpN2O is equally well modelled above the low O2 regions as in the rest of the ocean (Fig. 10, 12), and the contribution of the coastal and deep offshore ocean are nearly proportional to their surface areas (Table 2). These three features are supporting evidence for our results that suggest that the low O2 regions make a small contribution to the global ocean N2O production. They should be balanced against the model bias of the vertical distribution of N2O concentrations, which suggests a larger contribution from the low O2 regions. Freing et al. (2012) also estimated a small fraction of 7% of the global total contributed by denitrification / low O2 N2O production."

We have added a calculation of the contribution of coastal seas, the deep offshore and East equatorial Pacific oceans to N2O flux in Section 3.2 and the Discussion (Table 2 and associated text). And we have expanded on the information about coastal seas and estuaries at the end of the discussion:
"The largest coastal seas are resolved in our model although the processes related to specific coastal environments are not, such as the interactions with sediments and with tides. Our results do not include emissions from estuaries."

- Line 43: The reference to Klawonn et al., 2015 is missing.

The reference was added.

- Line 95, equation 2: More information should be given on this equation, and how it was used in the model/observation comparison. Does using this equation mean that the N2O flux is calculated for a specific period, or that it varies in time? This is unclear.

We have clarified that the model/observations comparison is done at places and months where there are observations. See reply to Gianna Battaglia's comment on Line 199.

Also, there number of significant digits in the various coefficients is way larger than any believable uncertainty associated with the measurements the equations should fit.

We reduced the number of significant digits in Eq. 1. See reply to the same comment on Eq. 1 by reviewer 1.

- Section 2.4, Table 1. Maybe some effort can be done to evaluate the improvements associated withe each model: by adding terms the cost function decreases minimally - is the improvement significant? Does it justify the increase in the model degrees of freedom?

We have used model representations that have relatively few parameters (=degrees of freedom), because the observational data that has been synthesised on a global scale cannot constrain more parameters. Because the prognostic model is an explicit part of the model N-cycle processes, the representations of which are independently constrained by additional observations, it actually has one parameter less (4) than the diagnostic model (5). Akaike's Information Criterion (a criterion that quantifies whether models with more degrees of freedom are "justified" by their increased predictive power $AIC=1/(n_{observations}-n_{parameters})*\log(RSS)+2n_{parameters}$) of the prognostic submodel is 5.9 lower than the diagnostic submodel. This is in the range (2-10) where there is more support for the prognostic model, but there is still some support for the diagnostic model (Burnham,K.P., and D. R. Anderson (1998) Model selection and inference, a practical information-theoretic approach. Springer).

- Line 133-134. The equation could be shown.

We have added: "(Eq. 9 in the supplementary material)" and have also added references to the other relevant equations in the supplementary material in the rest of the Materials and Methods.

- Section 2.6. The slopes (of what, with respect to what?) and relationships used for the model should be clarified with equations, and maybe with corresponding figures (e.g. the observational constrains used). Also, what is the range from which the various slopes were drawn in order to run the different model versions for the optimization?

How were they determined? What values were actually used? Finally, there must be concentration thresholds associated to the transitions between different slopes (e.g. O2). How were these thresholds determined? Were they also optimized for?

The equations, optimised ratios, and range of values tested are given in the supplementary material, we have added references to the relevant equations in Sections 2.4 – 2.6.

- Line 211-212. The reasoning is unclear: an increase in outgassing for a given atmospheric concentration should be driven by a parallel increase in surface concentrations, since the flux is proportional to the concentration (or pN2O) difference. For example, in the limit of removing the saturation N2O concentration, a doubling of the interior production of N2O should double both the outgassing and the surface concentration.

No, a doubling of production leads to a doubling of ΔpN2O, but ΔpN2O/pN2O is small in most of the surface ocean, and the surface concentration increase is proportional to pN2O, not ΔpN2O, so we are correct in stating that a doubling in production leads to only a small increase in surface N2O concentration. We have added this clarification to the manuscript.

- Lines 242-247. This entire paragraph is very unclear, please clarify.

We have clarified this paragraph:

"further observational constraints could not only reduce the error, but also further our understanding of the whole N cycle, including the option of evaluating their model representation against observations, and not just the part that N2O plays in them. Such further constraints are also likely to provide the most productive way to reduce unexplained variability that is found in the observations but not in the present models. E.g., we have shown that both the N2O and NO3 are underestimated at ~300 - 1500 m depth and overestimated below ~2000 m (Fig. 6, 7). Thus, improved representation of mesopelagic remineralisation might lead in improved representation of the N2O depth distribution. However, this falls outside the scope of this study."

- Lines 270-271. Constraining remineralization backwards from N2O production seems a bit far-fetched, given how hard it is to even constrain processes like denitrification alone.

Our point is that the current lack of constraints is not cast in stone. Addressing questions concerning the nitrogen cycle from different angles and integrating the different sources of information in a falsifiable model is more robust than constraining it from the more usual angles of export and nutrient concentrations alone. We added to the end of this paragraph: "Although there are relatively few N2O concentration observations, nitrification and denitrification respond to specific environmental queues (in particular O2 concentration), so that they could contribute a relatively large observational constraint over the full range of environmental conditions."

- Lines 279-281. Please clarify.

See reply to question about Eq. 4 above: Because the paper we discussed only tested sample sizes that were more than 2 orders of magnitude smaller than our database, we decided to delete this paragraph from the discussion.

- Lines 294-297. The issue of biases in model circulation could be assessed by using ventilation tracers, e.g. CFCs. Are they available for this mode?

We are currently including CFCs in our model but this will require time for the development, tuning and validation. The results will not be available for the current study but will inform follow up developments.

- Line 308: do the Authors really think their model can capture costal N2O dynamic, and the massive air-sea fluxes observed there (see Arevalo Martinez et al., 2015),especially in eastern boundary upwellings?

We have added separate analysis of the main N2O hotspots: coastal seas, deep offshore, and East equatorial Pacific oceans. This analysis shows that our two submodels are able to reproduce the observations (see in particular the close correspondence between both submodels and the observations in the high end tail in Fig. 12). Arevalo-Martinez et al. (2015) use the mean N2O flux to represent the whole Peruvian upwelling region. This is similar to linear interpolation with correlation length-scales of the whole region and the whole year. Since their plots suggest that the N2O fluxes are not linearly distributed, this could lead to overestimation of the N2O flux. Therefore we believe our mechanistic model is much more likely to capture realistic N2O dynamics, including in the hotspots, than previously published estimates. For further details see replies to reviewer 2's questions on L32 and L282.