I strongly support the goal of this paper, to better constrain the oceanic N2O flux using optimization techniques based on a compilation of datasets of N2O and related N cycle variables, combined with process based models. However, the methodology is difficult to follow and it is hard to believe that all 4 data-based approaches converge to basically the same answer and have the same relatively narrow range of uncertainty, which is governed primarily by uncertainty in piston velocity. There is also no overall sense of what sets this paper apart from earlier efforts, since it seems to be based heavily on what is largely the same delta pN2O data set used before. While I support publication in principle, I think there are many details that should be clarified and explored before this paper is ready for publication.

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

Although four databases were used in our paper, two of these databases, the nitrification rate and the NH4 concentration, were used to formulate a model that was as realistic as possible. They were not used to calculate the N2O flux. We were gratified that the two other databases of depth resolved N2O concentration and surface ΔpN2O converged on the same answer of a low contribution of N2O production in the low O2 region. We have rewritten the end of the introduction to clarify how the four databases are used.

While the literature on N2O fluxes is growing, the only earlier estimate of global N2O flux based on ΔpN2O was by Nevison et al. 1995, 22 years ago, before the MEMENTO database was available. It used interpolation rather than a mechanistic model to obtain fluxes where there are no observations. Therefore our analysis goes beyond existing publications and uses a larger and more complete dataset and process modelling.

Specific comments

L24 Typo: It (is) also currently estimated as the dominant contributor

Corrected.

L32 It's worth mentioning (up front) that this wide range is governed in large part by the possibility of very large coastal and estuarine fluxes. Later on line 87 we learn that the dataset resolution used here is 1x1 degree or 1.1 x 2 degree (plankTOM10.2, line 164), i.e., probably not good enough to resolve these coastal areas.

We have commented on the ambiguity about whether emissions from estuaries are included in oceanic emissions or not in the introduction:

"Part of the uncertainty in the oceanic emissions is whether estuaries are included, which could emit as much as 2.3 - 3.6 Tg N y-1 (Bange et al. 1996)."

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).

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."

L43 although there are additional pathways, such as (please give brief list).

We initially wrote it like this because the reference for this statement, Klawonn et al. 2015, measured a large number of N-transformation reactions, so that a list would not be brief. However, to respond to the reviewer's comment that the statement needed more detail, we have replaced it with ", although denitrification may be significant in the anaerobic centres of

large marine snow particles in oxic waters", because that is the most important pathway identified by Klawonn et al.

L72 probably should mention up front that the deltaN2O/AOU data are based on MEMENTO. Otherwise, it's a bit confusing to know the basis of this calculation.

This information was moved from the Fig. 3 legend to the main text: "The globally averaged ΔN2O/AOU ratio was calculated from the MEMENTO database (Bange et al. 2009) as 81.5 ± 1.4 µmol/mol (Fig. 3)."

L76 NPP is estimated at 58 +/- 7 PgC/yr based on what? An ocean model? Satellite data? Even at the lower end of 51 PgC/yr, this is on the high side of satellite-based estimates.

This estimate is based on our previously published work (Buitenhuis et al. 2013a). It uses the same methodology as in the present paper. We have added the following text to clarify the origin of the estimate: "based on ^{14}C primary production measurements (n=50,050), parameter perturbations of a previous version of the model used here, and Eq. 3". Our estimate is broadly within satellite algorithm estimates, which range from 38-70.7 Pg C/y. We have found that the vertically integrated primary production from our model reproduces the observations better than the best satellite algorithm. It also had the second lowest error (root mean square difference) of vertically integrated primary production relative to observations in the Arctic only out of 21 biogeochemical models (Lee et al. 2016 doi:10.1002/2016JC01193). Depth resolved primary production constrains global NPP even better than that at 58 +- 7 PgC/y., as discussed in Buitenhuis et al. 2013a.

L83, please list relevant references rather than just saying "(see Introduction)".

We have replaced this by a reference to Fig. 4 (was Fig. 11) which includes all the references.

Line 92-93. Since pN2O is close to equilibrium in much of the ocean, it seems important to consider the quality of these pN2O measurements. For example, surface measurements made with underway systems tend to have much higher precision than analyses based on bottle collection. Was the uncertainty comparable across the MEMENTO database and if not, were the differences in data quality accounted for in the subsequent calculations?

Annette Kock (who does the technical support for MEMENTO in the group of H. Bange) informally estimates that more than 95% of the surface pN2O data entries included in MEMENTO have been measured with underway systems. In addition, comparison experiments between underway and discrete measurements show an overall good agreement between both methods (Arévalo-Martínez et al., 2013, doi: 10.5194/os-9-1071-2013). We have added a caveat to the manuscript: "Since there is at present no formal quality control beyond that performed by individual contributors to the MEMENTO database and a check by the database administrators that the values make physical sense (Kock, A., and Bange, H. W.: Counting the ocean's greenhouse gas emissions, Eos, 96, 10-13, 10.1029/2015EO023665, 2015), we have taken the MEMENTO database at face value."

L125 Ocean models often do a poor job of reproducing observed O2. Suntharalingam 2012 used WOA O2 rather than model O2 for that reason. Presumably, the sensitivity to light, temperature and O2 described here is based on values from plankTOM10.2 (if not, please clarify). How well does plankTOM10.2 reproduce O2 relative to observations? (Note: I saw later that my question was addressed in the Results on lines 184-188. That material belongs up front in the methods description.

We have added this information to section 2

"As will be described more fully in Section 3.1, we used observed O2 concentrations in the simulations (Bianchi et al. 2012) rather than interactively modelled O2, to minimise the impact of model biases in simulated O2 fields (Suntharalingam et al. 2012)."

L132 paragraph starting here. This paragraph could be written more clearly, especially the sentence spanning L137-138. What is a variable N quota? Is the model using Michaelis Menten kinetics?

The paragraph was reworded, and references were added that provide further documentation of the model formulation and the contrast between a quota model and a Michaelis Menten kinetics model:

"The model uses a fixed C:N:O2 ratio for all organic matter of 122:16:-172, and Michaelis-Menten kinetics for growth rate based on inorganic N uptake by phytoplankton (Buitenhuis et al. 2013a, supplementary material Eq. 8, 9). We therefore need a K1/2 for growth rather than for uptake to be consistent with the fixed C:N ratio (Morel 1987)."

On line 144, a low cost function of 3.3 is better than the cost functions of >4 described for the previous model, correct? Yet, the sentence beginning on L142 with "However" suggests a large uncertainty.

This is indeed confusing (yes, 3.3 is better than >4), and it was rewritten to give a more equal balance between the small scale unexplained variability and the large scale pattern that is well reproduced by the model:

"Due to the highly dynamic nature of NH4+ turnover, the ability of the model to reproduce the observed NH4+ concentrations at the same times and places was by no means perfect, but the large scale pattern of surface NH4+ concentration shows an increase with latitude, consistent with the observations (Fig. 6), which translates into a cost function of 3.0."

L155-156 What is meant by "The slopes of the three processes" ?

We have changed this from slopes to ratios, and added an explanation 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."

Section 2.2-2.8. General question. Do the 4 databases described in section 2.2 correspond to the specific sections 2.4-2.7? If so, where does section 2.8 fit in? Is Equation 3 an alternative cost function to Equation 2 described in Section 2.3?' The apparent switch from Equation 2 to Equation 3 as the optimization technique is confusing and unclear.

The 4 databases correspond to sections 2.4, 2.5, 3.1 and 3.2. It was split in that way because the nitrification and phytoplankton NH4 use are necessary model developments before we can implement the prognostic model, but they are not part of the main result of the paper, which is an estimate the ocean-atmosphere N2O flux and its confidence interval. The switch between Eq. 2 and 3 is split in the same way. We've been using Eq. 2 in multiple previous publications because it gives equal weight to biases in small and large numerical values, and is therefore is therefore more appropriate for optimising a global model that spans a range of conditions. We have added this to Section 2.8:

"In previous versions of the PlankTOM model (Buitenhuis et al. 2006, Buitenhuis et al. 2010, Buitenhuis et al. 2013a) we have used Eq. 2 to evaluate the model because it minimises relative error, which we have found to be more appropriate when the observations span several orders of magnitude. Unfortunately, statistical confidence intervals have only been defined for χ^2-statistics such as Eq. 3 and 4, which minimise absolute error, so that we end up with two cost functions (Eq. 2, 3), depending on the application."

Line 190 and Figure 6. The model substantially underestimates N2O in the most important hotspots of production. Doesn't this mean it will tend to underestimate global N2O production?

It seems like this concern is dismissed somewhat casually with handwaving arguments, e.g., the text starting on line 289.

We have given considerable attention to the underestimation of N2O at depth in the low O2 regions, and discuss it from different angles in paragraphs 3, 6 and 7 of the discussion.

We have added Table 2 and a paragraph in Section 3.2 (second paragraph, starting "High N2O fluxes"), that analyses the contribution of N2O hotspots to global N2O flux.

We have rewritten paragraph 3 (starting with 'This lack of knowledge") of the discussion 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. On balance, including the new Table 2, our analysis still suggests a small global significance of the hotspots. This conclusion was also reached by Freing et al. (2012), and we added a reference to this: "Freing et al. (2012) also estimated a small fraction of 7% of the global total contributed by denitrification/low O2 N2O production."

Line 201-202 Please clarify how these results link together. Line 201 says that both hypoxic production AND CONSUMPTION were optimized. The subsequent results mention GROSS production of 0.33 TgN/yr, then optimized N2O production of 0.12 or 0.16 TgN/yr. Are the latter results net production? Can we infer that about 0.17- 0.21TgN/yr is consumed in suboxic zones?

Yes the results are for net production and we have added the word net to clarify this. Yes, the optimised consumption is 0.21 Tg N/y. We have also added clarifications throughout the manuscript on which processes occur in suboxic, hypoxic and oxic waters.

Line 202-203 Total production of 0.12-0.33 TgN y-1 in low O2 is only about 10% on average of total production. This is much lower than the 33% suggested by Suntharalingam et al. 2012. Does this mean that the authors are concluding that the OMZs are only responsible for a small fraction of global N2O production? Please expand on this point and call it out more explicitly in the Discussion.

We are not confident that the lower attribution to denitrification produced by our current model version is better than published by Suntharalingam et al 2012. This is detailed in the Discussion, to which we've added references:

"Both the diagnostic and the prognostic models assign a small percentage of the total N2O production to the denitrification pathway, 6 and 4% respectively. However, because of the large bias between the observed and modeled N2O concentration depth profiles (Fig. 6) these may be underestimates (Suntharalingam et al. 2012, Arevalo-Martinez et al. 2015)."

We have revised the text to discuss the possible implications of this shortcoming, which we argue do not significantly affect our results for the total global N2O flux. See our reply to the comment above on Line 190 and Figure 6, that outlines changes we made to Section 3.2 and paragraph 3 of the Discussion to more clearly present the balance of evidence.

Line 204 perhaps add a clause clarifying that the .0017 molN2O/molO2 slope is about 20 times the mean deltaN2O/AOU ratio of 8.15e-5 from line 82.

The .0017 slope is the gross production slope in the low O2 regions. The 8.15e-5 slope is the net slope averaged over the whole ocean. It is therefore to be expected that the former is larger than the latter. See changes made in response to the next two comments.

Line 205 production for the prognostic model is given in units of umol N2O (mol NO3)-1. Does this represent NO3- consumed by denitrification, or produced by nitrification? Can you provide an estimate of how this relates to the previous units of mol N2O/mol O2?

We have added clarifications on which slopes apply to denitrification and to suboxic, hypoxic and oxic waters. Since NO3- consumption and O2 consumption are spatially separated, stating a

N2O/O2 slope would be confusing, but we have added the model N:O2 ratio in Section 2.5, so that the magnitudes can be placed in context.

Line 216 Please use consistent N2O (mol O2)-1 slope units. Here the units are umol/mol. On line 78 they were nmol/mol. On line 204, they were mol/mol.

All slopes have been converted to a denominator in mol, and a numerator in µmol, or in mmol if it was >1000 µmol.

Line 217 How does this nitrification slope in units of umolN2O/molNH4+ relate to the

"N2O production slope" on line 206 in units of mol N2O/mol NO3-?

We have added clarifications on which slopes apply to denitrification and to suboxic, hypoxic and oxic waters.

Line 219-220 Are all measurements really of deltapN2O, or are most of pN2O in the surface ocean? In the latter case, what is the uncertainty in pN2Oatm, e.g., from Eq 1?

We have added in Section 2.2 that:

"The average absolute difference relative to the global average pN2Oatm data from the NOAA/ESRL Global Monitoring Division (ftp://ftp.cmdl.noaa.gov/hats/n2o/combined/HATS_global_N2O.txt) is 0.5 ppb between 1977 and 2014 and 0.3 ppb between 2000 and 2014." See also the reply to the comment by reviewer 1 on equation 1, in response to which we ran the optimisation using the NOAA/ESRL data, and got essentially the same results.

Line 220 On what basis was this 1978 estimate made? Is there updated information that could be used?

It is cited in Cohen and Gordon 1978 as a personal communication from Weiss, who calculated the solubility we used based on published data. No further details are given, but because it's an order of magnitude less than the uncertainty in the piston velocity, this doesn't strike us as a problem. We are not aware of more recent measurements. The Sarmiento and Gruber textbook (2006, Ocean biogeochemical dynamics, ISBN: 9781400849079) also gives the solubility we used as the most up to date one.

Line 233 Typo or confusing sentence.

Based on the comments of reviewer 1, we hyphenated "N-cycle-based".

Line 248-250. It would be good to provide references to support these claims.

We have added references to the C-cycle data, and refer to the figures with observational data for the N-cycle data.

Line 282 – paragraph starting here. This exercise, combined with large data gaps in Figure 9a, including in both the ETSP and ETNP, suggest to me that the authors are overstating the degree of certainty in their confidence interval. There are large areas of the ocean with no data, including in the most important hotspots for N2O production.

The observations are in fact fairly evenly spread (Table 2). There are observations in the ETSP and ETNP. The observations include upwelling regions. The analysis we present of a hypothetical undersampling of high values suggests that constraining the piston velocity would narrow the confidence interval more than making more pN2O measurements. As mentioned in response to the reviewer's comment on L32, we have added an analysis of N2O flux in coastal seas, the deep offshore and East equatorial Pacific oceans to quantify the contribution of these hotspots.

Line 303-304. This is the first mention of the fact that the model produces low fluxes from the Southern Ocean. Can you cite the relevant figure here and call attention to this point earlier in

the Results section? (Figure 9b,c,d all seem to indicate a substantial flux from the Southern Ocean.)

It is the atmospheric inversions that produce low fluxes in the Southern Ocean, rather than the process models presented here. We have rewritten these sentences to make this clearer:

"However, N2O emissions from inversions in the Southern Ocean are lower than the priors (Hirsch06, Huang08, Thompson14, Saikawa14). These low Southern Ocean emissions (0.02 – 0.72 Tg N y-1) are consistent with our results (0.68 – 0.79 Tg N y-1). South of 30S, 88% of the Earth surface is ocean, resulting in a clearer attribution in the inversions of the atmospheric N2O anomalies to ocean fluxes. We suggest that the higher emissions estimates from inversions could be due to a combination of overestimated priors of ocean fluxes in combination with insufficient observational constraints at latitudes North of 30S to allow correct partitioning between land and ocean fluxes."

Line 308-310. The neglect of estuaries is indeed a key uncertainty, which needs to be mentioned much earlier, i.e., in the Introduction. It is also debatable whether coastal areas are adequately represented in the models presented here, which 2x1 or 1x1 resolution.

See reply to question on L32.

L487 Figure 2. This figure suggests very high f-ratios, e.g., of 0.8-0.9 in the northern subtropical gyres, that are a little hard to believe. The global mean looks to be on the order of 0.4! Are these generally accepted values or are they biased by measurements in highly productive coastal waters?

This one point of 0.8 is the average of 2 measurements (0.8 ± 0.1) made at 14N21W and 21N21W, about 4° West of Africa (Varela et al. 2005). In the Discussion we point out (first paragraph) that the f-ratio (global mean = 0.29 ± 0.18) is the largest contributor to the uncertainty in N2O production we estimate from the N-cycle-budget. We also explicitly state that further (synthesis of) observations would help constrain this uncertainty. We think that trying to ensure a representative mean by weighting some values more (e.g. open ocean measurements) would be too subjective / sensitive to the exact methodology used.

L505, "Model results are for the same months and longitudes as the observations." What about latitudes?

We have added the clarification to the panel D legend:

"Latitude y-axis to the left of panel A."

L527 This plot is dominated by the error bars and somewhat obscures the focus on the mean value, which arguably is the more important quantity. The current study makes a much more detailed effort to quantify uncertainty than most of the previous studies (some of which make no effort at all). Could a separate panel with a narrower Y-axis range be plotted to better compare the mean value of the fluxes? And can you please provide some discussion of the main factors contributing to the differences in mean value?

Error bars are important when comparing different estimates. We have decreased the aspect ratio of the figure to 1, so that differences along the y-axis become easier to read.

We have added a discussion of the two main factors contributing to the different N2O flux in Nevison95:

"Because of differences in methodology it is not possible to provide reasons for why our estimate is lower than the more recent estimates. We can, however, compare our estimate to that of (Nevison95), because it is also based on a database of ΔpN2O. Compared to their high end estimate using the piston velocity of Wanninkhof of 5.2 ± 3.6 Tg N y-1, our estimate is lower because we use the more recent 13% lower estimate of piston velocity of (Sweeney et al. 2007), and because our ΔpN2O of 7.6 ± 18.1 ppb is 25 - 28% lower compared to 10.55 natm in Nevison

(1995) (the range is calculated based on the water vapor correction for conversion between ppb and natm, which increases from 0.6 - 4.1% at temperatures from 0 - 30 °C, which brings the values slightly closer together)”