**Answers to Reviewers**

**Referee #2:**

*The title of the manuscript is misleading. It should reflect the fact that only one model was used to run several scenarios. Thus, the part ". . . in global model studies" should be replaced by ". . . in a global model simulations" or ". . . in simulations using the MOZART model".*

The title has been changed to "On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations" to address the concerns of both reviewers. We keep the plural of models in the title, because we wish to emphasize that this is a common problem of global model simulations.

*The paper is too long and repetitive. It looks like a part of a project report.*

We shortened the paper in several sections, particularly in the introduction and skipped several links to the MACC project including the detailed reference to the MACC reanalysis and original Figure 2.

*Specifically, the last sentence in the Abstract needs to be re-written; the first 6 pages of the Introduction should be synthesised to a single page; the actual introduction starts on page 253.*

We re-wrote the abstract and shortened the introduction significantly. Nevertheless we found it important to give a comprehensive overview on the actual knowledge of the global CO budget in the introduction, as there is little information available up to now.

*Section 2.3 does not fit under the general title of "Model description"; sections 3 and 4 are not logically organized.*

Original section 2.3 is shifted to new section 4 ("MOZART sensitivity simulations"), i.e. the presentation of the model sensitivity simulations is now done after the summary of the evaluation datasets. All results from the model runs are presented in new section 5 ("Results"). A few contents were shifted in between these sections and the sub-sections in new section 5 were titled differently.

*The first paragraph of Conclusions would better fit in the Abstract; conclusions are presented in the last two paragraphs of the paper on page 272.*

The abstract starts now with the (modified) first paragraph of the conclusions. The first part of the conclusions has been shortened.

*Also, the authors should better explain the concept of the "Optimized approach" (section 4.1.3). As it stands, the reader does not know what was done.*

To address the concerns of both reviewers, we changed the new section 5.4 significantly. In particular, we added text on our initial model simulations with added traffic emissions world-wide without seasonality or for selected regions only. All these simulations revealed too high CO mixing ratios in summer or for certain regions, particularly over Asia, and needed unrealistically high additional global emissions. We decided not to show these results in the manuscript, because we believe that these additional simulations do not have the potential to significantly alter our conclusions and would make the manuscript less readable.

*Specific comments:*

*In all places the authors should change the noun "concentration" with the term "mixing ratio" or use proper units for concentrations (i.e. mass per unit volume).*

We changed "concentration" to "mixing ratio" wherever applicable. Sometimes the misleading noun also referred to "total column". In a few cases we kept "concentration", when no specific variables were addressed.

*247-5: there is something missing in this sentence*

The whole abstract has been re-arranged.

*247-27: limited – lower* - done

*248-7: caused – contribute* - done

*249-1: MEGAN is not an inventory* - done

*249-4: production of CO from . . .* - done

*250-5: the lower values – lower values, pls. remove 'the'* - done

*252-8: missing emissions – underestimation of emission fluxes?* - done

*253-9: how far – to what extent* - done

*253-21: Inspired by the MACC – Following the MACC?* - skipped

*254-8: model version presented – model version used?* - done

*254-13: accomplish a horizontal resolution – run at horizontal resolution* - done

*255-17: the finally published – pls. remove "finally"* - done

*255-21: distributed quickly – diffused rapidly?* - done

*256-18: conducted several – pls. state how many* - done

*256-27: determined – controlled?* - done

*258-2: In total – pls. remove* - done

*258-7: delivered – provided?* - done

*258-9: erratic or locally determined . . . - what does it mean?*

We changed the sentence to "Data was averaged over large-scale areas as depicted in Figure 3 to minimize the influence from local pollution or meteorology for single stations."

*258-9: we built – we calculated* - done

*259-4: acronym MOPIIT was already explained*

We skipped first appearance, so MOPITT is still explained here.

*259-4: flying on board – pls. remove flying* - done

*267-2: GAV – never explained* - done in section 3.1

*267-3: December to May – should it be January?*

Our statement is in-line with new Figure 8. We applied a flexible y-axis to Figures 6 / 8 to enhance their legibility.

*271-3: rough sensitivity – what does it mean?* - skipped “rough”

*271-7: Nevertheless – pls. remove* - done

**Referee #3:**

*The paper is generally extremely well-written if somewhat lengthy.*

Following the remarks from both reviewers we shortened the paper in several sections, particularly in the introduction.

*My greatest concern is the optimised approach as it is based on constructing a missing winter source of emissions. Although the general idea is well-argued the derivation of these specific scaling factors and within season variation in Table 4 are not outlined. With some improved description here this concern would be addressed.*

We changed new Section 5.4 significantly. See our answers to referee #2.

*Major comments:*

*1) Adding the word “found” before “in global models studies” may be appropriate.*

The title has been changed to “On the wintertime low bias of Northern Hemisphere carbon monoxide found in global model simulations” to address the concerns of both reviewers.

*2) The introduction is rather long –possibly not all the numerical values given are essential and the text could be more concise. In particular the text on anthropogenic emissions and MACC could be shortened.*

The introduction has been shortened significantly, particularly the reference to the MACC reanalysis including old Figure 2. Nevertheless we found it important to give a comprehensive overview on the actual knowledge of the global CO budget in the introduction.

*The methods also could be shortened where it is repetitive in a few places as outlined in specific comments.*

We addressed this point in the specific comments.

*The evaluation data section again contains many instrument details that could perhaps be shortened.*

We further condensed the evaluation data description in Section 3.

*3) The dry deposition velocities simulated with the two schemes seem to differ greatly both in terms of magnitude and variability. It would be useful to comment further on this.*

We already introduced the two schemes and their differences in detail in the introduction. We've now extended our implementation description of the Sanderson scheme in Section 2.2 and give a reference to Section 1 as well as an outlook on the expected effect on CO.

*4) Evaluation is performed often over three large continental regions. However these regions cover almost equivalent areas of land and ocean. For North America and Europe the sampling stations are mostly over land so it would seem appropriate to apply a land mask or to evaluate over land and ocean separately. Due to the long lifetime of CO, this may only be an issue for the lower model levels but still worth clarifying.*

The evaluation is performed over large-scale regions. For simplicity these regions have been named after the continents which they cover. It was not our intention to elaborate on differences between land and ocean, as we don't expect any substantial result. Furthermore, surface stations and MOZAIC airports are mostly over land, but sometimes close to the coast. Land/Sea differences also pose problems to the satellite retrievals which are therefore hard to interpret. We now avoided the term "continental-scale areas".

*5) It would be helpful to have more than one year of simulations to compare to observations and to draw conclusions from sensitivity studies. Although the anthropogenic emissions may not change dramatically in recent years, biomass burning and natural emissions may. It would be useful to at least comment on this aspect.*

Of course it would be beneficial to have model results from more than one year, but this was not feasible with our resources. The simulations analyzed in the manuscript already summed up to 10.5 simulation years, and several other simulations were done in the preparation phase. Even on the ECMWF and FZ Jülich supercomputers such a task consumed weeks of computer time. Moreover, we chose a year when no extraordinary natural emissions or meteorological conditions occurred (see also Figure 6 in Kaiser et al. (2012) and  Figure 2 in Inness et al. (2013), citations as in the manuscript). NH natural or biomass burning emissions predominately occur in the summer months and a global trend for these emissions has not been reported yet. We also noted in the new Section 5, that the interannual variability of the CO burden has been calculated with a similar range than the difference between our simulations MI and MI-OPT.

*6) Since simulated NH low biases are the key feature of this paper it would be helpful to either replace the current panels of maps in Figure 6 (since this is one of the few figures that is a map) with anomalies plots relative to one of the satellites or to add extra panels depicting anomalies to Figure 6.*

We added difference relative plots (100x(MOPITT-MOZART)/MOPITT) to new Figure 5 and added "(see relative difference panels in Figure 5)" in the text.

*7) Initial background text in Section 4.1.3 on the optimised approach and Figure 10 although interesting appear somewhat tangential to this study and could be condensed. This would allow for less general background and more specific background to the origin of the scaling factors in Table 4 and their variation. With regards to these scaling factors was some preliminary work done to derive the "optimised" concentrations? Optimized suggests the emissions were derived to match the concentrations but there is no description of the optimisation process and Table 4 appears somewhat out of the blue.*

In new section 5.4 we skipped the discussion of the US NEI. We condensed the text on the anthropogenic emission trends and added some information on our initial test simulations with globally increased traffic emissions and increased traffic emissions for specific regions, but without seasonality. See also our answers to referee #2.

*8) I'm find the text describing Fig 13 rather misleading in a number of aspects especially in having a comparison for July. The first sentence in the text describing Figure 13 results refers to differences of 170 ppb in December between MI-OPT and MI-DEP. December is not shown in this figure and the differences in January do not look like they exceed 100 ppb at most. Area average values in Figure 9 for Europe look to be < 50 ppb. Moreover, is there any merit in showing any panels for July since the scaling factor was 1, hence we would expect identical(?) results for MI-OPT and MI-DEP? Lastly, the zonal-mean OH results are illegible in the Figure and confusing displayed as is. Could these be new panels in fig 13?*

Thanks to this helpful comment we re-designed our new Figure 12 and its description. July results are not shown anymore. Zonal mean OH has become a new panel. All results now refer to January. The first sentence in the description was misleading and referred to an older version of the Figure. For January, we found indeed differences of up to 160 ppb in Central Europe with one grid point value reaching 167 ppb in the Po valley. We changed the text accordingly.

*9) "Overestimations attributed to natural emissions". This conclusion is important and could be phrased more clearly when discussing Figure 6 so it can be referred to in section 4.1.3. However, the argument constructed can also be applied to a certain extent to the MI+AN simulations for East Asia.*

We now refer more clearly to Figure 6 in new section 5.2. Also in new section 5.4 a reference to Figure 6 is now given.

*Specific Comments:*

*Page 247, line 5 "finally CO concentrations" – this text would be better placed earlier in the abstract where emission perturbation results are described.* - done

*Page 249 Line 18: Text discusses considerable uncertainties about the global budget of CO and refers to Table 1, but the relationship between Table 1 and uncertainties is not given. Further text should explain the connection.*

We changed the first sentence of the paragraph to make clearer that we address both, sources and sinks. Table 1 is an illustration of the uncertainties in anthropogenic emission inventories, followed by more discussion in the text on other budget terms.

*Page 249: Line 7 add "of CO" after "methane oxidation".* - done

*Page 250: line 17 remove "however".* - done

*Page 251: line 16 remove "also"* - done

*Page 253: line 22 remove "Inspired by MACC . . .set-up"* - skipped

*Page 255: beginning of page- line 15 can the anthropogenic, biomass and biogenic emissions data used in this study be referenced to Tables 1 and 2?*

We inserted references to Tables 1 and 2.

*Page 256: Line 21-: is it correct to say "differ in the underlying emission inventories" or rather is it that the emissions in these same inventories have been perturbed?*

The paragraph is now in Section 4 ("MOZART sensitivity simulations") and the formulation has been improved. The sentence is now: "These sensitivity simulations differ in their specific perturbation of the emission inventories and dry deposition velocities".

*Page 266: lines 22-24 and page 257 lines 17-18- this text is repeats text in earlier sections- is it necessary?* - skipped

*Page 258: line 2 – remove "In total".* -done

*Page 258: line 8 "Data was averaged over continental areas"- were these the blue boxes in Figure 4? If so these span both land and ocean hence averaging may not be appropriate.*

See our answer to the reviewer's major comment.

*Page 258: line 25- it would be clearer to discuss the European stations and the USA stations in a separate sentences as the latter only are then referred to in the following sentence.* - done

*Page 259: line 1: "interpolated, single profiles combined to monthly means" this is rather unclear. What temporal resolution data is interpolated? What is the relevance of "single"- which usually refers to location rather than time?*

We improved the sentence to: "The 3-hourly model results were interpolated to the times and locations of the MOZAIC observations, afterwards the profiles of each airport were combined to monthly means."