Answers to Reviewer 2

First of all, we would like to thank the anonymous reviewer #2 for his/her great review of our publication and provide here the answers to his/her questions.

> *The manuscript describes the newly developed SOCOL-MPIOM atmosphere-ocean GCM with coupled atmospheric chemistry and its behavior in a number of climate simulations. As the authors correctly point out, so far, for climate applications mostly models without coupling to a comprehensive atmospheric chemistry scheme have been used. On the other hand, so-called CCMs, i.e. Atmospheric general circulation and chemistry models are mostly used without a coupled full ocean, and hence not in climate mode. Therefore, this model and its performance are a highly interesting topic and very well suited for publication in GMD. It is very nice to see that the authors evaluate the model performance using numerous standard climate simulations. Nevertheless I have a couple of comments which I think will partly require rerunning the described experiments (unless it is just the description that is flawed) and I also think that the priorities in the model description and other parts of the text are not always well chosen so that I would consider the necessary revision as major. One suggestion would be to considerably shorten the rather long manuscript. I will make suggestions where one could cut the text, but I do not want to impose this. Finally, the authors have to decide what they think is relevant to communicate. I will list my concerns according to their appearance in the text. In addition I would like to see a careful reconsideration of the use of language in the manuscript which is often rather approximate. I will only give examples for this. 13 co-authors should be able to deal with language issues themselves. Abstract: The abstract seems somewhat too lengthy for my taste. However, this should be up to the editor.*

Thank you for your comments. We substantially shortened the manuscript ($\sim$ -10 pages) and removed the lengthy discussion of the solar forcing. The focus is now on the role of the interactive chemistry, the climate sensitivity, and the reasons for the simulated temperature increase in the modern period. Therefore, we reduced the number of experiments considered for the period 1600-2000 to two, by excluding the two stronger forced simulations. This allows also also to shorten the discussion of the solar forcing.

Please note, that the results of the sensitivity experiments also changed slightly. We found an error in the detrending of the experiments, which lead to an overestimation of the response in the order of 0.05 K. The temperature increases for the different experiment are therefore slightly reduced.

When the revised manuscript is ready, we will also carefully check the language. The abstract of the revised manuscript has been shortened as well and fits now on one page.

Introduction

> *Introduction: The introduction needs about 4 pages until the goal of the paper is mentioned. This should happen much earlier. The first 4 pages contain mostly something like a review on troposphere-stratosphere coupling. The reader is totally left alone with the question which part of the content is relevant for this paper. The introduction should only concentrate on those issues and make their relevance clear. By this I think the introduction could easily cut by half.*

Thank you for this comment. We restructured and shortened the introduction, so that it now covers 2.5 pages.

> *3015L9ff: "In recent years the stratosphere has become more and more important for our understanding and proper simulation . . . " The importance of the stratosphere has likely not changed much, rather the recognition of its importance.*

Changed to:

„In recent years, the stratosphere has received increasing attention for our understanding and proper simulation of climate variability and climate change."

3015L14ff: "importance of the vertical resolution" Most of these papers, as far as I remember, do not discuss the relevance of the vertical resolution in the stratosphere but of resolving the stratosphere at all.

In general all cited publications compare the effect of a 'well resolved stratosphere' to older versions, with the model top typically at 10 hPa. To make this more clear we changed the statement to:
"The importance of a well resolved stratosphere is highlighted in several studies."

3015L20: Why "furthermore"? Isn't the chemical composition part of the stratosphere?

We changed this to:
„Moreover, some of the recent changes in the surface climate can only be reproduced when stratospheric chemistry and changes in the chemical composition of the stratosphere are considered in GCMs."

3016L8: "underlying mechanisms are still debated" Mechanisms of what? Of wave propagation?

changed to „. . . the underlying mechanisms of stratosphere–troposphere coupling . . . "

The above four points exemplify the issues I have with the use of language. However, I will ignore all further language issues unless they really distort the content.

3018L13: Are we talking about a coupling of a GCM with a CTM, here or rather of the inclusion of a chemistry mechanism into a GCM?

In case of SOCOL this question is not easy to answer. It is rather the inclusion of the chemistry mechanisms of MEZON into ECHAM5. However, in this paragraph in the Introduction, we intended to give a general overview of chemistry-climate modelling, instead of describing SOCOL-MPIOM. Nevertheless, we rewrote this paragraph, to reflect both approaches:
„Coupled climate models have been shown to be an important tool for understanding processes and feedbacks between the different components of the climate system, e.g., between the ocean and the atmosphere. To consider interactions between the atmospheric chemistry and the physical component of the atmosphere, atmospheric chemistry modules needs to be included or coupled to GCMs. Most of the coupled chemistry–climate model (CCM) simulations so far were performed with prescribed sea surface temperatures (SSTs; e.g., Eyring et al., 2006) or simplified mixed-layer oceans (e.g., Stenke et al. 2013b). However, in both approaches the climate system is not able to simulate the full response to, e.g., a strong external forcing like volcanic eruptions, since interactions between atmosphere and ocean are not considered (Kirchner et al. 1999). Moreover, global SST data sets are only available for a few centuries. "

3018L18: "coupling of an interactive ocean model is preferable" Why? I guess that depends on the scientific question.

Agreed, we removed this sentence from the manuscript.

3018L26: "evaluated using . . . a pre-industrial control simulation" Such an evaluation is difficult due to the lack of comprehensive sets of observations. Please be more specific.

We replaced „evaluated" by „described".

3018L31: "Finally, we close . . . "

Changed to:
„Finally, the results are discussed and summarized."

Model description

3019L1: "The coupled model consists of . . . coupled to . . . "

Thank you.

3019L7: Is SOCOL version 3 used here?

Yes, it is SOCOL version 3, as stated in the manuscript.

3019L10: Are really need five references to MEZON needed? This seems like excessive self-citing to me. If all are needed it should be said what for.

We reduced the number of references for MEZON and cite now only Egorova et al. 2003.

3020L12: "Chemistry climate coupling: . . . " It sounds like the only coupling between MEZON and MA-ECHAM5 is done via the radiative effects of the trace gases. What about their transport? Isn't this done by the winds calculated in the GCM. And why have this point separate from SOCOL? Isn't that the coupled MEZON-MAECHAM5?

We restructured this part of the manuscript to improve readability and included the chemistry-climate-coupling part into the description of SOCOL. Furthermore, we briefly describe transport of chemical species by:
„The transport of the chemical species uses the advection scheme of MA-ECHAM5."

Later it is mentioned that chemistry is calculated only every two hours. It would be interesting to discuss the potential error introduced by this. In the abstract (where it may not belong) it is already mentioned that NOCHEM has issues because of the missing diurnal cycle in ozone. If one prescribes low daytime ozone this may not be a severe problem. Instead using the high nighttime ozone for up to two hours after sunrise could cause problems.

The chemistry is, similar to the radiation scheme, calculated every 2 hours. This indeed may lead to some biases. Prescribing daytime ozone only in the NOCHEM simulations could indeed improve the simulation without interactive chemistry and reduce the differences between CHEM and NOCHEM. However, so far the model version used in CHEM is configured to write out daytime averages of ozone for the NOCHEM forcing. Changing this to daytime averages would imply that the simulations would need to be redone, which would take a large amount of time and computational resources. Nevertheless, the usage of daytime averaged ozone is an promising step towards smaller differences between CHEM and NOCHEM and mention this possible improvement in the discussion section of the revised manuscript. Similarly, we discuss possible biases related to the 2h step in the radiation and chemistry schemes is discussed.

We also argue that the usage of daymean averaged ozone on NOCHEM is closer to the way ozone is handled in most GCMs today. When the ozone data set of Cionni et al. (2011) or Fortuin and Kelder (1997) is used, models are forced with monthly mean data. The diurnal cycle is missing in this data as well, and the mesospheric ozone concentrations in these data sets are probably closer to our daymean averages than to daytime averages.

I would also like to know more about the coupling details (the way of operator splitting for instance). Additionally about transport: Is this done via the transport scheme of ECHAM or of the MEZON CTM. How is water vapor dealt with, which experiences phase transitions (in the GCM) and chemical reactions (in the CTM part). Maybe all this is discussed in earlier SOCOL publications. In this case, references would suffice.

All these questions are indeed important, but where described previously in Stenke et al. (2013b). Our manuscript focuses on the coupling of SOCOL to MPIOM, which is why we keep the description of the model rather brief. The same applies to the ocean model MPIOM, which is described in Jungclaus et al. (2006).

To highlight this, we added the following two sentenced to the beginning of the SOCOL section:
„An in-depth description of the model and the parametrizations used in the chemical module is given in (Stenke et al. 2013). In the following we refer only to the most important fact that are needed to understand the characteristic of the coupled model SOCOL-MPIOM.“

> *3021L16ff “SOCOL does not use zonally averaged ozone concentrations . . . ” The meaning of this sentence gets only clear on page 3023 where it is said that even in the NOCHEM simulations 3D ozone is used? It is mentioned that the zonal structure is relevant for dynamics. However, prescribing 3D ozone means also that zonal structures in dynamics and ozone will be inconsistent occasionally. To me it is not a priori clear that this inconsistency is a priori preferable to the inconsistency arising from the prescription of zonally averaged ozone. This needs to be discussed.*

We clarified that we are talking about the NOCHEM configuration here:
„Additionally to the interactive chemistry mode the chemistry module can be deactivated, which disables chemistry-climate interactions in the model. . . . In contrast to many other models, SOCOL without interactive chemistry does not use zonally averaged ozone concentrations, as this leads to significant biases in the stratospheric climate and also affects tropospheric dynamics . . . “
We agree with Reviewer 2 that dynamics and ozone concentrations can be inconsistent when a 3D forcing is applied to the model. However, the same applies also when zonally averaged ozone concentrations (2d) are used as forcing. In both cases dynamics and ozone can be inconsistent, which is what we expect when we apply ozone as forcing in the model. The same effect is found after volcanic eruptions. Here, we also typically prescribe aerosol concentrations, but the shape of the aerosol cloud is independent of the dynamical changes simulated by the model.
We decided to use a 3D climatology for a different reason: it was shown by Waught et al. (2009) that a zonal averaged ozone data set leads to an underestimation of recent changes in the stratospheric temperatures trends and also the effect of these changes on the tropospheric dynamics in the SH.

> *3021L18: “ozone forcing” I find the use of the word “forcing” excessive in this document. Maybe I’m wrong, but I would use it only in the case where a quantity is changing. Please check all the document.*

Thank you, we carefully revised the manuscript and reduced the occurences of ’forcing’.

> *3022L18ff: repetition*

Some of the information about the spatial resolution was mentioned before in the introduction and is removed from the introduction in the revised version of the manuscript. Furthermore, we mention the spectral truncation of T31 also in the model description, when the different time steps of the calculation are described. However, we decided to keep this repetition, since we think that the time step information belongs into the model descriptions part and a full description of the spatial resolution is needed in the experiment section.

> *3022L24: What does “from scratch” mean?*

From scratch refers to present day conditions. Clearly, this state of the atmosphere differs from the 1600 state, but given the short adjustment time of the atmosphere we think that this effect is negligible.
We clarified this in the manuscript:
„The atmospheric and chemistry components are initialized by present day conditions, which adjust to the pre-industrial climate state within a few years.“

experiments

> *3023L6f: repetition*

We removed this sentence.

3023L8/12: 1367 or 1368 W/m2?

Thank you for this comment, 1367 Wm^{-2} is the correct value.

3023L9: "positive temperature drift" Which temperature?

The positive drift in the global mean surface temperatures. We corrected this in the manuscript.

3024L12: "TCR is ... temperature change in the 20yr period" Or the difference between the mean temperature during this period and the initial temperature? And again: Which temperature?

Again the global mean surface air temperature. We corrected this in the manuscript. The temperature change in these experiments were estimated relative to a control experiment fixed CO2 concentrations. However, this has changed now (see next answer).

3024L20: "All other forcings are constant at 1990 level" It was said earlier that the control run uses constant 1600 conditions. How can a climate sensitivity be calculated when boundary conditions other than CO2 are changed? I think these simulations need to be rerun. Or at least a very thorough discussion would be needed to argue why this doesn't matter.

Thank you for this comment. Obviously there were some misleading statements in our manuscript. We performed these experiments using 1990th restart files from one of the transient simulations, therefore the CO2 forcing increase is relative to 1990 and all other forcing are held constant at 1990 levels. From this restart file, a control simulation (fixed CO2) and the climate sensitivity experiment were started and the SAT change was estimated relative to this control simulation. However, this obviously led to confusions, since the transient experiments were not introduced at this point of the manuscript.

Furthermore, as we already discussed in the submitted manuscript, the possibility that the warm 1990 climate state might affect the estimated climate sensitivity of the model. For CMIP5 the sensitivity experiments were performed in a pre-industrial climate state, therefore the comparison may be biased. For the revised version of the manuscript we decided to repeat the simulations, but in a pre-industrial climate state with initial conditions from the CHEM control simulation. Furthermore, we performed all experiments (TCR and ECS) for all models (SOCOL with and without chemistry, as well as ECHAM5/MPIOM).

For the revised manuscript we rewrote the description of the experiments and the evaluation of the climate sensitivity.

Quoted from our answer to reviewer #1:

The new results show that the different climate state has strong effects on the estimated TCR and ECS, and that the TCR and ECS of SOCOL/MPIOM agree now much better with the estimates for the CMIP5 models.

	TCR [K]	ECS [K]
SOCOL_chem/MPIOM	1.8	3.8
SOCOL_nochem/MPIOM	1.8	4.0
ECHAM5/MPIOM	1.8	5.4

Given the differences to the former estimates, we hypothesize that the warm climate state leads to some amplified positive feedbacks in the former present day climate sensitivity experiments (compare Meraner et al., 2013). Furthermore, the transient experiment are characterized by a very strong positive surface air temperature drift around 1990, which might also affect the TCR/ECS estimates.

Since we are anyway planning another publication focusing on the differences in the model response to CO_2 and solar forcing, we will also analyse this difference in the response between the two climate states at a later point. In the revised manuscript we will include the pre-industrial climate sensitivity experiments only.

3025L7: "Emissions are based on ... concentrations" How is this done?

This statement was not correct, emissions are already given in the CMIP5 database.
„Emissions of ozone depleting substances (ODS) are based on the historical emissions from the CMIP5 database.“

3025L25ff: repetition.

Thank you, the discussion of the SSI differences between standard ECHAM5 and Shapiro is now included in the description of the Control experiments and has been condensed.

3028L16: In view of the discussion of climate effects from the different types of forcing it would be useful to state what aerosol effects are included in the model.

This description was indeed missing in the submitted manuscript. Thank you for the comment. We included the following paragraph in the model section of the manuscript:
„SOCOL considers the climatic effects of stratospheric sulphate aerosols. For the heterogeneous chemical reaction on the sulphate aerosols surface area densities (SAD) need to be prescribed. The optical properties of the aerosols are calculate offline, e.g., by a microphysical model. They include extinction coefficients, asymmetry factor, and single scattering albedo of the aerosols for each wavelength interval. In the troposphere, only the radiative effect of the aerosols is included. Therefore, 10 different species are considered, including carbon aerosols, dust particles, sea salt, and sulphate aerosols. “

3030L26: The MPI-ESM is based on ECHAM6 and has a preindustrial equilibrium temperature of about 13.5C. I guess the authors mean the ECHAM5/MPIOM model.

Thank you, we corrected this.

3031L3ff: I guess, again, not the MPI-ESM but ECHAM5/MPIOM is meant. Even then: Is it clear that the energy imbalance in the new SOCOL-MPIOM is the same? Could new imbalances have been introduced? E.g., through the coupling of water vapor?

Thank you for this comment. We tested the TOA radiation imbalance of SOCOL in a slab ocean experiment and found, that this model is characterized by an imbalance of $1.45\,\mathrm{W\,m^{-2}}$.
We therefore replaced the sentence referring to ECHAM5/MPIOM with the following:
„Tests with SOCOL coupled to a mixed layer ocean model reveal a TOA imbalance of $1.45\,\mathrm{W\,m^{-2}}$, a further, slight adjustment of the temperatures is therefore likely.“

3031L23ff. Why is it remarkable that absolute variability is small where absolute values are small? What is missing is a discussion of the maximum relative variability in the lower stratosphere.

We did not intend to state that low absolute variability in regions of low ozone concentrations is remarkable. In the revised manuscript we included a sentence on the variability maxima in the lower stratosphere and mesosphere:
„Variability in the troposphere and mesosphere is in general very small and is only reflected in the normalized anomalies. These variability maxima are found in the lower stratosphere, in particular in the tropics and polar latitudes, as well as in the polar mesosphere.“

3032L1: Do the authors really mean “intra-seasonal” variability? Or are they talking about inter-annual variability of seasonal means?

Thank you, we clarified this in the manuscript.

3033L1ff: This paragraph sounds like it is a general flaw of models without chemistry to produce too high temperatures in the mesosphere. Please see my earlier comments. I see no reason for prescribing daily averaged ozone concentrations instead of using daytime averaged values. On the other hand the potential flaws of calculating chemistry only every 2 hours should be discussed. (see comments above).

We do not fully agree with this comment. The above mentioned paragraph describes the behaviour of SOCOL–MPIOM to produce to high mesospheric temperatures when the chemistry module is disabled. This is explained by a combination of the ozone climatology used and the Lyman-α parametrisation. We did not state anything about other climate models.

Nevertheless, as stated above we think that the usage of a daymean ozone climatology is the best approach for a comparison of a model with chemistry-climate feedbacks to a model version without these feedbacks. Models without interactive chemistry typically apply a ozone climatology, which is based on monthly averages. These models may therefore indeed also simulate a similar temperature bias, in comparison to interactive chemistry models.

We extended the discussion about the 2h computational time steps in the revised version of the manuscript.

> *3033L6: It is claimed that during nighttime ozone levels increase due to upward transport. I would consider photochemistry as more important. Please provide references or discuss in more detail.*

We corrected this in the manuscript. During daytime the photolytical destruction of ozone dominates. This process is missing during the night, therefore ozone can build up.

„In the mesosphere ozone exhibits a pronounced diurnal cycle (Brasseur and Solomon, 2005). During daytime ozone is destroyed by UV radiation. In the night the photolytic ozone destruction is missing and ozone concentrations increase. Differences between night and day can reach up to 15 % in SOCOL.“

> *3033L14f: With interactive chemistry I would guess that another process causing differences is the photodissociation of water vapor.*

Thank you. We changed the sentence to:

„With interactive chemistry water vapour in the stratosphere further depends on a number of photochemical reactions and the oxidation of CH_4.“

> *3.1.3 Dynamics: It is an important question if dynamical variability is significantly altered by chemistry coupling. The results here are interesting because they seem to indicate that chemistry coupling plays a minor role. However, the discussion remains superficial and I would either suggest to make this into a major topic of this manuscript or to shorten it considerably.*

We prefer to keep the discussion of the zonal wind differences (mean state and variability) as well as the comparison of the NH polar vortex index and the sudden stratospheric warming. However, the major focus of this publication of the comparison of NOCHEM and CHEM and since the differences between these two are small for the dynamics, we do not want to make this a major topic.

In the revised manuscript we shortened this subsection.

> *3034L1ff: Interesting differences between CHEM and NOCHEM are the strengthened jets and lower polar temperatures. It is true that a stronger polar vortex would better isolate the polar air masses. On the other hand, winds and temperature are tightly connected through the thermal wind relation. So there could be a radiative origin of the stronger polar vortex. Why can one be certain that the stronger vortex comes first and what would be the cause for it?*

We analysed the contribution from different processes and parametrisations (e.g, Lymann-α, water vapour differences, GCR effects, ...) to the mesopheric temperature differences between CHEM and NOCHEM in a set of sensitivity experiments (atmosphere only with fixed SSTs). In these experiments we can find a cooling response in the polar stratosphere due to the water vapour differences (stratospheric water vapour is higher in CHEM and the anomalies have a regional maximum in the polar stratosphere). The cooling may therefore be partly related to the stratospheric water vapour differences. The cooling may, furthermore, be related to the differences in the high altitude cirrus between CHEM and NOCHEM (which are related to the water vapour differences), since these clouds reduce the amount of outgoing LW

radiation. However, it may also be possible that the water vapour and cloud differences are a consequence of the dynamical differences.

With the current setup, we can not really state, whether the dynamic response leads to the colder polar stratosphere or whether the cooling has some radiative origin, which leads to the vortex amplification. We state this in the revised version of the manuscript:

„In the lower and middle polar stratosphere on both hemispheres, the temperatures are significantly reduced during the winter and spring seasons. This cooling is accompanied by enhanced water vapour mixing ratios in the lower polar stratosphere in CHEM, which could explain the regional cold anomalies. However, with the current setup of the experiment, we can not rule out the possibility of a dynamical origin of the cold anomalies, related to the changes in the polar vortex intensity explained in the following."

3034L16ff: Tropical variability should not be discussed as it is mainly determined by the nudging.

In these paragraph we refer several times to the paper by Driscoll et al. (2012), who evaluated the response of several CMIP5 models to strong volcanic eruptions. They concluded that the bad performance of many models might be related to the low tropical variability or the too stable NH polar vortex. Therefore, we decided to discuss both indices in our manuscript as well, since in both indices, the performance of SOCOL/MPIOM is improved in comparison to Driscoll et al. The description of the tropical variability is already very brief (3 sentences) and directly states that the good agreement between reanalysis and model in the tropical region is related to the nudging.

We therefore decided to keep this description.

3035L8ff: The discussion of the SSW frequency is lengthy including some unnecessary introductory remarks on their relevance for tropospheric climate. I would summarize the result saying that statistics for CHEM and NOCHEM are both very similar to those for ECHAM5. This can be shown and said in a much shorter way. When comparing preindustrial simulations to ERA data the different periods/climate states should however somewhere be mentioned as a caveat.

We have considered you comment and the introduction of the SSW analysis has been substantially shortened. Furthermore, we added a reference to the SSW frequency in MAECHAM5 and mention the caveat of the different periods again in the SSW paragraph.

3036L14ff: Also the part on differences in zonal wind variability should either be shortened or the discussion should be done more thoroughly. One may e.g. speculate that polar ozone provides a positive feedback for the vortex strength in spring. While results for the NH seem to confirm this, the SH results do not. Why?

We have shortened this paragraph in the revised manuscript.

3.2 Tropospheric and surface changes: The discussion of the cirrus cloud effects in the SH high latitudes is interesting. Are there any other studies on the potential influence of stratospheric water vapor on high-latitude cirrus?

Thank you for this question. We could not find any studies directly related to our findings. In general, there are a few publications focusing on the role of higher water vapour concentration for the formation of polar stratospheric clouds (PSC) (e.g., Hervig et al., 1997, Kirk-Davidoff et al. (2002)). Others also suggest linkages between an increase in PSCs and temperature increases at the surfaces in polar latitudes (e.g., Sloan et al. (1998), Kirk-Davidoff et al. (2002, 2008)). In our results, however, the additional clouds formate at altitudes between 250 and 100 hPa in the Southern polar stratosphere, which is too low for PSCs.

- Hervig, M. E., Carslaw, K. S., Peter, T., Deshler, T., Gordley, L. L., Redaelli, G., Biermann, U., et al. (1997). Polar stratospheric clouds due to vapor enhancement: HALOE observations of the Antarctic vortex in 1993. Journal of Geophysical Research, 102, 28,185–28,193.

- Kirk-Davidoff, D. B., Schrag, D. P., & Anderson, J. G. (2002). On the feedback of stratospheric clouds on polar climate. Geophysical Research Letters, 29(11), 1–4.
- Kirk-Davidoff, D. B., & Lamarque, J.-F. (2008). Maintenance of polar stratospheric clouds in a moist stratosphere. Climate of the Past, 4(1), 69–78. doi:10.5194/cp-4-69-2008
- Sloan, L. C., & Pollard, D. (1998). Polar stratospheric clouds: A high latitude warming mechanism in an ancient greenhouse world. Geophysical Research Letters, 25(18), 3517–3520. doi:10.1029/98GL02492

The rest of the section where mostly small differences between CHEM and NOCHEM are reported could be shortened significantly. In particular the last part on surface variability could become one sentence.

We shortened this part.

3.3 Climate sensitivity: As said in earlier: If really the 1%CO2 and 4xCO2 simulations differ in boundary conditions other than CO2, they cannot be used to estimate climate sensitivity. If the climate sensitivity stays high in corrected experiments it would be interesting to discuss reasons for the discrepancy to the ECHAM5/MPIOM. A potential candidate are cloud feedbacks. This could be analyzed easily by calculating cloud radiative effects.

As described above, the boundary conditions of the former climate sensitivity experiment were consistent, but represented a present day climate state. For the revised manuscript the re-ran the simulation in a pre-industrial climate state and found again that ECHAM5/MPIOM and SOCOL/MPIOM differ in their ECS. For the ECS estimates we performed 4× CO_2 simulations with both models. In agreement with Lie et al. (2012) we estimated an equilibrium temperature response of 10.8K (ECS: 5.4 K) for ECHAM5/MPIOM and 8.0 K, 7.5 K, for SOCOL/MPIOM without and with interactive chemistry, respectively (ECS: 4.0 K, 3.8 K). Restart files for these experiments are taken from the CHEM control simulation for the SOCOL/MPIOM experiments and from the control run of the millennium simulations (Jungclaus et al. 2010) for the ECHAM5/MPIOM experiment. The stronger temperature increase in ECHAM5/MPIOM is related a stronger positive albedo feedback (sea ice loss in the NH). It may be possible that differences in the NH sea ice extent at the beginning of the experiments explain a part of the differences in the responses. CHEM is warmer and NH sea ice is reduced in comparison to the millennium control simulation. Furthermore, we also found differences in the cloud response, with a stronger cloud cover reduction in ECHAM5/MPIOM, in particular in tropical latitudes. An other possible explanation for the difference, may be found in the vertical resolution of the experiments. The SOCOL experiment are conducted with 30 levels (up to 0.01 hPa), while for the ECHAM5 experiment, a lower resolution with 19 levels (up to 10 hPa) was selected.

We discuss these results and caveats in the new results section:

„A feedback analysis Andrews et al. (2012) reveals larger contribution of the SW component to the overall feedback in ECHAM5–MPIOM, which are related to a stronger reduction of sea ice, in particular in the NH, and a stronger cloud cover reduction. The differences in the amount of sea ice loss may partially be related to differences in the initial state of the experiments, with more NH sea ice in ECHAM5–MPIOM. The stabilizing LW feedback is also smaller in ECHAM5–MPIOM. The comparison of SOCOL-MPIOM and ECHAM5-MPIOM may also be biased by the differences in the vertical resolution in the atmosphere. The ECHAM5-MPIOM simulations were performed with 19 vertical levels (up to 10 hPa), while SOCOL-MPIOM used 39 levels (up to 0.01 hPa).“

3040L2: “second double CO2 simulation” It is confusing to name the 1simulation “double CO2” (also in table 1).

Thank you, we will rewrite this subsection of the manuscript and avoid the term “double CO2 simulation”.

3040L8: Again: ECHAM5/MPIOM is not equal to MPI-ESM.

Thank you, we will consider this for the revised version.

3040L21: The title of this section is almost equal to the title of subsection 4.2.1.

Thank you, we changed the titles. Since section 4.1 has been removed we could reduce the number of subcaptions and merged the former section 4.2.1 with the former 4.2 (now 4.1).

3040L24: "surface air temperature" In earlier sections the terms "2m temperature" or just "temperature" was used. I would prefer to talk about SAT, but anyway, please be consistent.

Thank you for this comment. We revised the manuscript and use now the term SAT where surface air temperatures are described.

Section 4.1: I'd suggest to cut this section. It is not without interest to see that the large difference between the solar forcings of the L and M simulations do not really matter much, but the section is very descriptive and I think the aim of the paper to present the new model wouldn't suffer from omitting this part.

We decided to remove the description of the pre-industrial period from the manuscript and focus on the performance of SOCOL–MPIOM in the control experiments and the industrial period.

3044L20: "The UV variability is not important for the surface climate." It is often claimed that the UV variability would impact the surface response pattern by topdown mechanisms. In case the authors decide to keep this section, a more thorough analysis of this phenomenon (and why it seems not to act in this model despite the large solar forcing) would be needed.

In this paragraph we refer to results from Anet et al (2013 a,b). These two studies lead to the conclusions, that UV variability is not important for the surface climate. This is not a finding of this publication, therefore we do not think that this publication is the right place to analysis this phenomena more in detail. However, since the description of the pre-industrial period has been substantially shortened, we also removed the discussion of the results from Anet et al (2013 a,b).

3044L25: It is a very strong statement that "a significant anthropogenic influence on global mean temperature starts ... with the beginning of the industrialization". Attribution studies using fingerprint methods have often indicated that the anthropogenic influence is only evident since about the 1980s.

Thank you, we agree that this statement was too strong and too simplified. We removed it and shortened the introductory paragraph of this section.

3045L25 "TOA": This may cause confusion. Isn't TSI usually given for 1AU? Radiative forcing is in general defined for TOA. Maybe it would be useful to already at this point estimate the radiative forcing from this TSI difference.

We not fully understand this comment. What we describe with 1.7 Wm^{-2} is the difference in the TSI change between the early and late 20th century between the M and the L forcing. This is the differences in the incoming SW at the TOA. However, we removed this paragraph from the manuscript, since the revised version includes only the (former) M solar forcing. The experiment forced by the L solar forcing have been excluded from the analysis.

3045L25f: This is interesting. Is there any reason why the ocean heat uptake is so different?

The differences in the heat uptake between the L and M forced simulations may be related to the forcing differences and seem to compensate their effect on the SAT to some extent. However, differences between the members are also pronounced and it may also be possible that the differences are related to internal processes in the ocean.

In the revised version of the manuscript we no longer consider the M and L forced simulations and focus only on the two M experiment. Therefore, we did not extend the analysis of the ocean heat uptake.

3046L14: Why give the temperature change in K/100yr. Why not simply as a difference between the selected periods?

Thank you for this comment, we corrected this in the revised manuscript.

3047L3: It is very interesting that the 20th century warming simulated by SOCOL/MPIOM is much larger than the average simulated by CMIP5 models while apparently the future projected warming is not. Why doesn't the high sensitivity act in the future?

We argue in the manuscript, that the pronounced temperature increase in the 20th century is due to a combination of several processes (sensitivity, solar, ozone, and maybe underestimated aerosol effect). The new estimates for the TCR indicate, that the sensitivity of the model is even less important for the 20th century warming and therefore related to the remaining forcings. These forcings (solar, ozone, and aerosols) are less important in the future and the forcing differences between Anet et al. (2013) and CMIP5 are smaller, than in our simulations (e.g., solar amplitude). Non-GHG forcings therefore do not substantially affect the temperature increase of the 21th century and the results of Anet et al. agree better with the CMIP5 projections.

3048L3ff: All relative contributions to warming are probably estimated with some error margin. Could it be that the error margin of the total 87% includes 100% so that no additional feedbacks would be required?

Thank you for this comment, this is indeed the case. We state now:
„All individual forcings (solar, ozone, GHG, aerosols) add up to only 70 % of the full forcing experiment, but given the large uncertainties in the estimates this difference is not significant."

3048L27ff: I guess an albedo of 0.3, not 0.7 was assumed. In the estimation of radiative forcing, was stratospheric adjustment considered? I guess that quite some percentage of the TSI variability comes from the UV that does not necessarily warm the troposphere.

Thank you, indeed 0.3 was used for the albedo. Stratospheric adjustment was not accounted for. We mention this in the revised version of the manuscript.

3049L9: Unit of sensitivity is wrong. Please check all over the document.

Thank your for this comment, the unit was indeed not correct. In the revised manuscript we removed this paragraph from the manuscript and extended the discussion of the simulated temperature trends due to differences forcing and the comparison to the literature.

3049L17: What does "inclusion of the additional RF from ozone" mean? Additional with respect to what was used in other CMIP5 models? Many of those have used the Cionni et al. ozone climatology which includes a trend. How different is the forcing in SOCOL from this standard dataset?

We had a discussion of the simulated ozone changes in the discussion section of the submitted manuscript. While SOCOL simulates an increase of roughly 15 DU in tropospheric column ozone, other studies indicate changes below 10 DU and Cionni et al. gives an change of approximately 5 DU (between 1890-1919 and 1970-1999). The simulated tropospheric ozone changes are therefore probably overestimated, which applies also to the associated SAT changes.
We changed the structure of the manuscript and moved the discussion of the simulated ozone changes into the results section.

3051L4f: Please check the bracket.

Sorry for that. The sentence was accidentally pasted in.

3051L23: Is the ECS really "commonly assumed to be independent of climate state"? I wouldn't think so. At least for very different climate states I don't know of this common assumption.

We were not specific enough at this point. What we meant is that there was not always an agreement, whether the climate sensitivity experiments should be performed in a preindustrial or present day climate state. For CMIP3 for example, both approaches were possible [1].

[1] `https://www-pcmdi.llnl.gov/ipcc/standard_output.html#Experiments`

However, since we decided to start new climate sensitivity experiments in a preindustrial climate state, the discussion of the climate sensitivity will also be updated and we will clarify this statement, when the experiment are finished.

> *3051L27ff: Is there an explanation for the increased TCR caused by chemistry coupling in the SOCOL-MPIOM? If the ozone change is less than 1/20 of that described in the cited paper, it must have a different cause in SOCOL-MPIOM.*

The comparison between Dietmueller et al (2014) and our results was not completely fair in the submitted manuscript, since we compared a 1%/year CO_2 experiment (our study) to an abrupt 2× CO_2 experiment (Dietmueller), ran into equilibrium. Therefore some differences in the response are expected.

For the revised manuscript, we performed abrupt 4× CO_2 simulations for the SOCOL/MPIOM with and without interactive chemistry and found a reduction of the feedback parameter of 7% (Dietmueller: -8.4%). The agreement is therefore much better with the new simulations. Still, the simulated ozone changes are smaller than in Dietmueller et al., therefore a weaker negative feedback is expected. Weaker anomalies in our results may be related to the fact that our coupled atmosphere-ocean model is still not in equilibrium after 150 years, while Dietmueller et al. reach the equilibrium within 50 years with their slab-ocean setup. Anomalies in the stratospheric water vapour concentrations on the other side, are larger and these may balance the weaker feedback due to the ozone changes. SOCOL/MPIOM seems to simulate a stronger increase in the water vapour transport from the troposphere to the stratosphere in the 4× CO_2 experiment than ECHAM/MESSy. Overall, the net-feedback is then similar,