Response to referee # 1

I thank the referee for his/her comments and I respond to them point by point hereafter.

Many sections such as the simulations of the last millennium remain purely descriptive. I suggest a restructuring or a severe shortening of the "Preliminary Experiments" and "Last Millennium" sections. I would even favor moving them to the "Supplementary Material" for people who are really interested in details of this specific model.

Sections 3, 4 and 5 will be moved to the Appendix.

The aim of the paper is not clear after reading the abstract. It needs to be explained that the correct model parameters should be assessed, model uncertainties in recent future predictions, etc. It also does not read fluently because thoughts/sentences are not connected.

The abstract will be modified to insist more on key points of the paper, i.e. the assessment of the model parameter sets and the importance of the uncertainty on the model parameters for estimating a plausible range of climate change for the future. Moreover, we will include some key results in the abstract.

Line 6ff: It remains vague how model uncertainties have been assessed in previous studies and if the measure by Gleckler et al. (2008) is different from the measure used here.

Gleckler et al. (2008) analyse a number of climate variables simulated by several coupled ocean-atmosphere general circulation models. They evaluate the model performance according to 2-D climatic fields simulated for a given climate state. Here we assess parameter sets (or model versions) in transient simulations (i.e. last century) for global climate variables. Therefore we cannot apply strictly the metric proposed by Gleckler et al. (2008). Rather we adapted it to our purpose. Instead of the RMS errors used by Gleckler et al., we use the error in the trend. Like Gleckler et al. (2008), we favour the use of the median to define the metric. Gleckler et al.'s methodology will be described in more details in the revised paper.

Last paragraph of the Introduction could be shortened by including the references already in the preceding paragraphs.

This will be done.

Here is is not clear who the experts are and how they have chosen the parameters. Can this choice cover the whole range of uncertainties and does the choice that is made to produce "contrasted responses" (line 22, page 716) influence the results?

Our purpose when choosing the parameter sets was really to cover a large range of model responses to an increase in atmospheric CO_2 concentration as well as to an input of freshwater in the North Atantic Ocean but we did not aim at covering all the possible values of each parameter. For example, IPCC (2007) suggests that the equilibrium climate sensitivity is likely 2.1°C to 4.4°C (based on GCM studies). With a range of climate sensitivities of 1.6°C to 4.1°C, our parameter sets cover indeed the likely range suggested by the IPCC (2007). Figure 4 (of the discussion paper) displays the range of climate and MOC sensitivities obtained with the selected parameter sets. Indeed our parameter sets are only a sub-set of all the possible parameter sets. They do not cover the full range of sensitivities but widely cover the range suggested by other studies. Moreover, the purpose of this paper is to obtain realistic present day and past climate simulations. For that purpose, we performed more than one hundred simulations of the present-day climate modifying the nominal value of one or several parameters and we kept only those that (1) provide a realistic present-day simulation and (2) cover the best the phase space (climate sensitivity vs MOC sensitivity). Those amongst the authors of this paper working since long with LOVECLIM identified the most relevant parameters according to previous studies (Goosse et al. 2007, Goelzer et al., 2010). This will be explained in the revised version of the paper.

The entire overview over the huge amount of experiments is confusing. One big table with all experiments and all codes/names of the experiments could help.

More that 200 experiments were performed and are discussed in this paper. It is indeed a huge amount. Therefore it is not possible to present all of them in one single big table. However, we propose two tables explaining (1) how the names of the experiments are constructed and (2) what are the main features of each experiment category. Furthermore, all the simulations presented in the revised version are performed with the same model components activated (ECBilt, CLIO, VECODE and LOCH).

	Name formation	
A	E	Either pre-industrial or two times CO_2 scenario or freshwater flux scenario
	M	Last millennium
xy	(11,12,21,22,31,32,41,51,52)	Climatic parameter sets; x is related with the climate sensitivity of the model for the given climatic parameter set and y is related with the MOC sensitivity to a freshwater perturbation (MOC sensitivity) for the model.
z	1 – low 2 – medium 3 – high	Carbon cycle parameter sets
END	2CO, HYS, S1Conc, S1Efor, S2Conc, S2Efor	Further specification for the experiment (see table below)

Table : Summary of the name formation for the different simulations discussed in the paper. The general name is AxyzEND. For each character of this name (column 1) , the table gives its possible value (column 2) and its meaning (column3).

Experiment name	

Exyz	Pre-industrial equilibrium: No volcanic eruption, GHG as in 1750, TSI[1] = 1365 Wm^{-2}
Exyz2CO	Two times CO_2 scenario: Starting from Exyz Forcings as in Exyz except for the atmospheric CO_2 concentration (see figure 2 of the discussion paper).
ExyzHYS	Water hosing simulation: Starting from Exyz Forcings as in Exyz except for a freshwater perturbation applied in the North Atlantic (see figure 3 of the discussion paper).
MxyzS1Conc	Simulation of the last millennium (transient simulation): Starting from an equilibrium state at 500 AD Forcings[2]: orbital parameters, land use changes, volcanic activity, solar activity, changes in concentration of GHGs other than CO_2, sulphate aerosols (S1), diagnostic mode for atmospheric CO_2 concentration
MxyzS1Efor	Simulation of the last millennium (transient simulation): Starting from an equilibrium state at 500 AD Forcings: orbital parameters, land use changes, volcanic activity, solar activity, changes in concentration of GHGs other than CO_2, sulphate aerosols (S1), prognostic mode for atmospheric CO_2 concentration
MxyzS2Conc	Simulation of the last millennium (transient simulation): Starting from an equilibrium state at 500 AD Forcings: orbital parameters, land use changes, volcanic activity, solar activity, changes in concentration of GHGs other than CO_2, sulphate aerosols (S2), diagnostic mode for atmospheric CO_2 concentration
MxyzS2Efor	Simulation of the last millennium (transient simulation): Starting from an equilibrium state at 500 AD Forcings: orbital parameters, land use changes, volcanic activity, solar activity, changes in concentration of GHGs other than CO_2, sulphate aerosols (S2), prognostic mode for atmospheric CO_2 concentration

Table : Summary of the major features of the different types of simulations discussed in the paper.

Do not use abbreviations for words that are used only a few times like Conc or Efor. This makes it just complicated for the reader to search for the place it was used first.

We are sorry that the referee feels uncomfortable with the names of the simulations that we performed. We hope that the additional table will help the referee to understand them. The names we gave to them (e.g. Conc and Efor) is unfortunately the best we found to avoid an arbitrary numbering of them.

---

[1] TSI = Total Solar Irradiance

[2] The reader is referred to the main text for a detailed explanation of the different forcings

As NO conclusions for the model performance assessment are drawn from this and the previous experiments/sections, I suggest to move them to the Supplementary Material or to shorten them severely.

The section will be moved to the Appendix.

6. Last century:

Now again new ensembles of simulation with new forcings are introduced. I would suggest to have a section and comprehensive table about all experiments. Otherwise it is far to complicated.

See response above.

The last paragraph of 6.1 rather belongs to the introduction. It makes the impression as if a new paper stats here and actually it would be a good choice to have the paper starting directly with the last century as all conclusions are drawn from discussions done in that second half of the manuscript.

The paragraph will be moved to the introduction

Section 6.2 to 6.4 are mainly descriptive results while in sections 6.5 a discussion of the results is added at the end. Here a consistent split into results and discussion would be better or the sections 6.2 to 6.4 should also include some interpretations.

We put all the discussion into a separate section according to the referee's advice.

What does the variability of MOC strength depending on the parameter set tell us? Are model comparisons useless because each model itself already has large uncertainty related to its tuning?

What we may offer is to model a range of reasonable responses. Of course, it would be much better if we could constrain this range with data but we don't have. Still, even though we cannot decide which parameter set is the best, we are able to give a range of uncertainty and response, which is already interesting by itself.

What is the message of this paper, to find a best parameter configuration or to show the possible range and thus model uncertainty?

Clearly the purpose of the work reported in this paper is not to find the best parameter configuration. Rather, our aim is twofold. We want to identify the model uncertainty through a range of possible responses to given climate forcings. Moreover, we want to be able to eliminate parameter sets that are giving unrealistic behaviours over the past in order to better constrain the range of climate change projections (not shown in this paper).

What metric did other studies use before, e.g. Gleckler et al. 2008?

Gleckler et al. (2008) calculated RMS errors for each model and each variable, using two references. For a given variable and reference, they define a 'typical' model error, i.e. the median of the RMS error calculations. For each model, variable and reference they normalize the RMS error to this typical model error. Therefore, they obtain a measure of how well a given model compares with the typical model error. Their methodology will be briefly described in the revised paper.

It remains unclear why the metric is degraded to indexes, especially if the threshold is a critical value as discussed on page 735, line 6ff.

As we explained above, our approach, although different from Gleckler et al.'s one (2008), is based on theirs. Instead of providing a measure for each variable, we provide an overall score corresponding to each parameter set. Therefore, we set a minimum threshold to be reached for each variable to be considered as 'doing well'. Indeed, we cannot afford to take a parameter set as good if the poor results for all but one variable are over-compensated by the excellent result of that single one.

Does the result of not finding a "perfect" parameter set indicate that there is none for this model or just that it was not found by the small amount of sets that could be tested?

We do not know any theoretical proof in favour of one or the other hypothesis. Obviously, there is no climate model able to perfectly reproduce all the complexity of the climate system. Some are doing better for some components or processes, others are doing better for other components and/or processes. Moreover, all climate modellers (or modellers in general) know that there is a large number of parameters in models. Therefore, it is almost impossible to test the combination of all the values for all the parameters (even though the experience by climateprojection.net tends to go in that direction). Thus, we can only give a honest answer to referee, i.e. in this paper and in many previous studies, we tested a large number of values of the different parameters of the model. This fine tuning of the model gives rise to 'a best' version although it still includes some bias. Therefore, we think that with the model as it stands, we will not be able to provide a perfect parameter set. LOVECLIM (as well as other climate models) does have a strong structural uncertainty (related to the choice made during the build-up of the model). Hence, improving the model probably requires improving the physics rather than (or in addition to) improving the parameters. This comment will be taken into account in the revised paper.

A major conclusion is that the ocean heat uptake of the LOVECLIM is too large. First that could also be mentioned in the final conclusions. But second the reader would like to know if there is a parameter or another option to correct the model for the bias which was found here.

This work allows us highlighting the problem of ocean heat uptake, unfortunately, not to find the solution. Work is still needed to fix it (see also referee #2).

Figures: There should be a legend in each figure. It is not practical to always search for a color code in a different figure on another page.

Will be added in the revised version

Detailed comments:

Abstract:

Page 712, Line 6: parameters seem to be rather "selected" than "identified"

Will be taken in into account in the revised version

Page 712, Line 11: the sentence "Climate simulations . . ." is not connected to the text before and after

Will be taken in into account in the revised version

Page 712, Line 17: wording: "set of parameter sets"

Will be taken in into account in the revised version

Page 712, Line 18ff: not clear, why are parameters useful that are not able to reproduce past climate?

When performing this study, we had in mind to simulate the evolution of the ice sheet and sea level in the future. The diagnostics are based on only a few variables, slightly oriented towards our final purpose. However, if the purpose of the study is strongly modified and if criteria are changed, less good parameter sets might give a much better score (see also response to referee #2). Moreover, if we want to study the very extreme climate situations (maybe less likely), it might be interesting to keep parameter sets giving rise to them.

Page 713, Line 21: vague, how large is "quite large"?

The sentence is taken verbatim from Knutti's (2008) paper and his published work do not allow to be more specific. Clearly, the structural uncertainty (related to the model) and the systematic bias of the model are strongly linked.

Page 713, Line 21ff: sentence too long and unclear.

The sentence will be cut.

"Among all those possible sources of uncertainty, we focus here on the parameter uncertainty in LOVECLIM, a global three-dimensional Earth system model of intermediate complexity (Goosse et al., 2010). Specifically, the overall goal of this study is to identify a reasonable number of parameter sets that meets two major constraints. First, each parameter set should yield past and present climate simulations coherent with observations. Second, the various parameter sets should lead to a wide range of possible climate and sea level change scenarios over the next millennia. Thus, they will provide a reasonable sample for quantifying the uncertainty of future climate changes in forthcoming studies."

Page 713, Line 28: ideas for forthcoming studies would better fit in an outlook

Will be taken in into account in the revised version

Page 714, Line 21: full stop missing

Will be taken in into account in the revised version

Page 715, Line 3: here it is said that parameters are "selected" and not "identified". Make consistent

Will be taken in into account in the revised version

Page 715, Line 27: there are a lot of simplifications in the model, why is especially the cloudiness mentioned to be a limitation?

Clouds are probably one of the most important components of climate models and it is not yet very well known. Consequently, it may be a source of uncertainty that could potentially play an important role on climate evolution. Unfortunately, we are unable to investigate the role of the cloud feedback in LOVECLIM because the representation of clouds is too crude, which is also partly related to the simplified atmospheric component of the model. Although we haven't much latitude to alter the cloud representation, we manage to compensate for this weakness through other parameters intervening in the computation of the radiative transfer. For example, we use extreme values for the representation of the water vapour effect in the radiative scheme.

4.1 The pre-industrial climate:

Page 721, Line 8: observations for 1979-2000 are compared with pre-industrial model simulation as if they should be the same

We indeed clearly underline that "Pre-industrial and present-day climates are slightly different. " As explained in the paper, we do not want to demonstrate that the simulated pre-industrial equilibrium climate reproduces perfectly the real pre-industrial climate. We do not have enough observations to do so and even if we had, there could still be a difference between equilibrium and transient climates. Rather, as explained in the manuscript, we want to show first, that all the parameter sets lead to similar pre-industrial climates and, second, that these simulated pre-industrial climates are reasonable.

4.3 Sensitivity to water hosing:

Page 722, Line 4: "anomalous"? Is there usually a normal amount for freshwater added? Where is the freshwater taken from/compensated?

We replace the word 'anomalous' by 'additional'.

4.4 Sensitivity of the carbon cycle:

Page 723, Line 17: not surprising that the model follows the GHG concentrations if it is constrained by them.

The model is indeed constrained by changes in concentration of greenhouse gases **except CO_2**. Therefore it is meaningful to check that the model follows the reconstructed CO_2 concentration.

Page 724, Line 13: also say which sensitivity Frank et al. (2010) found

It is not our purpose here to discuss the carbon climate sensitivity as defined by Frank et al (2010). Rather, we aim at assessing the impact of our carbon cycle parameter sets on this index (in line of what we did for climate and MOC sensitivity).

Page 724, Line 14: which third digit? The reader is only confronted with 2 digits for most of the paper. The entire over view over the huge amount of experiments is confusing. One big table with all experiments and all codes/names of the experiments might help.

We hope that the tables above will help the referee. Moreover, all the simulations presented in the revised version are performed with the same model components activated (ECBilt, CLIO, VECODE and LOCH). Thus, all the experiment names will include exactly three digits.

5. Last Millennium:

Page 725, Line 4: There is a reconstruction of land cover changes during the last millennium which might have been better than assuming a simple linear evolution (J. Pongratz, C. Reick, T. Raddatz and M. Claussen: A reconstruction of global agricultural areas and land cover for the last millennium. GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 22, GB3018, doi:10.1029/2007GB003153, 2008) Is there a reason why it was not used?

We are of course aware of this reconstruction. Pongratz et al. (2008) estimated the extent of cropland and pasture from 800 AD to 1700 AD. Their reconstruction does not differ from Ramankutty and Foley's one (1999) after 1700 AD. Moreover, before 1700 AD, the difference between Pongratz et al.'s reconstruction (2008) and the linear reconstruction remains small. Therefore, and because previous studies were performed with this linear reconstruction, we decided keeping it here. Furthermore, the referees asked to focus only on the last century in the revised version, a time interval over which Pongratz et al.'s reconstruction (2008) is identical to Ramankutty and Foley's one (1999).

7. Performance of the parameter sets:

Page 731, Line 3: what does "configuration" mean here?

We agree with the referee that the choice of the word 'configuration' is unfortunate and unnecessary in this context. This will be deleted in the revised version.

Page 733, Line 14: confusing, now the 3-digit experiment names are used for the first time

We hope that the tables above will help the referee. Moreover, all the simulations presented in the revised version are performed with the same model components activated (ECBilt, CLIO, VECODE and LOCH). Therefore, all the experiment names will include exactly three digits.

Page 733, Line 17: Its major ...

→ Its only major ...

Will be taken in into account in the revised version

Page 734, Line 16ff: correct this sentence

Will be done in the revised version

Page 734, Line 24: A weaker CO2 (what ???) generates...

A weaker CO_2 concentration generates… This will be corrected.

Page 734, Line 26ff: Meaning of last sentences is unclear

Lower CO_2 level generates a weaker carbon emission (and vice versa) due to deforestation since the emission is calculated on the basis of the potential growth of trees, which is favoured by higher CO_2 levels. This may explain the change in performance between Conc and Efor experiments.

Figure 13:

- Codes on the x-axis are not explained and different form the 2/3-digit code used in the text.

The 0 digit means that the corresponding component (ISM) has not been activated and therefore no parameter set is used for that component. As said before, all the simulations presented in the revised version are performed with the same model configuration. Consequently all the experiment names will include exactly three digits.

- This figure should also not be split over two pages because the legend has to be visible

We will forward this request to the publisher.

The revised paper will be re-organised according to the referee's suggestion. The other comments will be taken into account in this revised manuscript if still relevant.

References:

Frank, D.C, Esper, J., Raible, C.C., Büntgen, U., Trouet, V., Stocker, B., and Joos, F. (2010). Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. Nature, 463, 527-532 (doi:10.1038/nature08769).

Gleckler, P.J., Taylor, K.E., and Doutriaux, C. (2008). Performance metrics for climate models. J. Geophys. Res., 113 (D6), D06104, doi: 10.1029/2007JD008972.
Goelzer, H., Huybrechts, P., Loutre, M. F., Goosse, H., Fichefet, T. and Mouchet, A. (2010). Impact of Greenland and Antarctic ice sheet interactions on climate sensitivity. Clim. Dyn., doi: 10.1007/s00382-010-0885-0
Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Jongma, J., Huybrechts, P., Mouchet, Selten, F.M., Barriat, P-Y., Campin, J-M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M-F., Morales Maqueda, M. A., Opsteegh, T., Mathieu, P-P., Munhoven, G., Pettersson, E.J., Renssen, H., Roche, D., Schaeffer, M., Severijns,C., Tartinville, B., Timmermann, A. and Weber, N. (2010). Description of the Earth system model of intermediate complexity LOVECLIM version 1.2, Geosci. Model Dev. Discuss., 3, 309-390, http://www.geosci-model-dev-discuss.net/3/309/2010/gmdd-3-309-2010.html
IPCC (2007). Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.
Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G G. A., Meinshausen, M., Murphy, J. M., Plattner, G.-K., Raper, S. C. B., Stocker, T. F. Stott, P. A., Teng, H., and Wigley, T. M. L. (2008). A Review of Uncertainties in Global Temperature Projections over the Twenty-First Century. J. Climate, 21, 2651-2663. DOI: 10.1175/2007JCLI2119.1
Pongratz, J., Reick, C., Raddatz, T. and Claussen, M. (2008). A reconstruction of global agricultural areas and land cover for the last millennium, Global Biogeochem. Cycles, 22, GB3018, doi: 10.1029/2007GB003153.
Ramankutty, N., and Foley, J. A. (1999). Estimating historical changes in global land cover : Croplands from 1700 to 1992, Global Biogeochem. Cycles, 13(4), 997 -1027.