Response to referee #2

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

Taken at face value, the second quote as to the overall goal of the study suggests that the majority of the paper, in its current form, might be a better fit to another EGU journal, Geoscientific Model Development (www.geosci-model-dev.net). This, I feel, would be an appropriate place to document the existence and evaluation of the different parameter-set versions of LOVECLIM without more extensive analysis and discussion. With such a target, the paper would need little revision.

We feel that the aims and scope of CP fits with our paper better than GMD. Indeed, our paper is centred on climate, more specifically, the climate of the last millennium with a strong emphasis on the last few decades. In this paper, we are comparing the ability of various parameter sets to simulate past climate changes. However, our purpose is clearly not to evaluate the components of the climate system. Rather, we aim at presenting the readers how we tried to select a range of parameter sets that yield a range of simulated climates in agreement with observations and that could be used for providing a range of plausible climate change projections. In other words, we aim at building a scientifically sound assessment of the range of past and future climate changes/variations. The focus of this paper is really climate. Parameter sets are only a tool to reach that goal. Therefore, we feel that the readership for our paper is more towards CP than GMD.

p715:26 The prescribed nature of the cloud forcing in ECBilt is passed over lightly here, and not mentioned again. Given that cloud feedbacks in models are one of the primary sources of uncertainty in determining climate sensitivity, if general conclusions are to be drawn from a parameter-uncertainty study with LOVECLIM then I think this limitation of the model deserves a more considered discussion.

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.

p716:16,20 Frustratingly little detail is given as to why these parameters were chosen, what ranges of uncertainty were considered for each and how the particular sets considered were, prior to the study, "chosen to produce reasonable simulations of present-day climate [...]".

Were all other potential combinations tried and rejected under a metric as outlined in section 7?

We did not set an experimental setup like Climateprediction.net. We run ourselves all the experiment and we were therefore limited by the number of simulations. Moreover, we aimed at keeping some control on each experiment individually. Nevertheless, 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.

p717:2 It is noted that Goosse et al. (2007) has considered a number of the climate parameter sets previously, yet the results of this paper aren't mentioned again - it would seem that a useful comparison of the model response in that study could probably be made with the present results.

Goosse et al. (2007) used five parameter sets corresponding to 112, 212, 312, 412 and E512. They performed transient simulations from 8 kyr BP to 2100 AD, starting from an equilibrium state at 8 kyr BP. After 1AD, Goosse et al (2007) used the same forcings as those used in the millennium simulations described in our paper. Moreover, they conducted five ensemble members for each parameter set over the period after 1851 AD. All the simulations display a decrease in summer sea ice extent throughout the 20th century. Compared to observations covering the second half of the 20th century, 112 and 212 seriously underestimate the decline in summer sea ice, while 312 slightly underestimates it. Goosse et al (2007) concluded that 112 and 212 are incompatible with the observed record. Moreover, they also showed that the model results for the parameter set 512 are not in agreement with the very few reconstructions of the summer sea ice extent during the early Holocene. Those previous results will be added in the revised version of the paper.

The descriptions of how each parameter is used in the model very useful though.

We thank the referee for acknowledging the usefulness of our description of the model parameters.

As alluded to in the overall comments, this whole section feels rather vague and purely descriptive, without drawing anything in the way of conclusions.

The purpose of this section is to establish the framework in which the subsequent simulations are performed. First, we define the equilibrium state from which the transient simulations start. Second, we compute the response of the model to a change in atmospheric CO_2 concentration and in freshwater flux in the North Atlantic according to the different parameter sets and we describe the model behaviour in response to different parameter sets for the carbon cycle. This second step is meant to characterise the general behaviour of the model for each parameter set.

p720:12-14 suggests that this is deliberate, but I'm not sure I buy it. The section does demonstrate that they have achieved a "reasonable" range of "reasonable" climate models, but just leaves me asking "why?" (not to mention "what is reasonable?". Why do the different parameter sets end up with different climate sensitivities? What parameter variations are responsible for the different responses to MOC hosing? The carbon cycle sensitivity is at least subdivided into land and marine parameters, but more explanation of the land parameter effects would be interesting.

The referee asks why we obtained the phase distribution described in the paper. As already mentioned, we performed more than one hundred simulations of the present-day climate modifying the nominal value of one or several parameters. We chose amongst them, a small number of simulations (and corresponding parameter sets) that cover the phase space.

This kind of analysis would be really interesting, and is the sort of thing I think should be done in order to get the most out of this work. A multi-panel figure that simply shows parameter value against climate sensitivity for each parameter might be a useful place to start, although one would of course expect a degree of non-linear interaction between the different parameter variations in the sets.

We understand that the referee would like that we indicate/compute the impact of each parameter on the simulated climate. However, this is strongly different from what we aim for in this paper. Indeed, here we try to stick as much as possible to realist climate simulations. The approach suggested by the referee would request much more computer time and, more importantly, would lead for many combinations of parameters to unstable simulated climates or non realistic mean climate states. It would definitely be a completely different work although certainly very instructive as well.

Is there a particular reason carbon-parameter set 2 is used for 4.1, 4.2 and 4.3 rather than the others? This information should also be in the text, not just in table 3. Which climate-parameter set is used for 4.4?

The carbon cycle parameter set 2 is used to analyse the role of the different climatic parameter sets on the simulated climate because it corresponds to the values used in previous studies. For the same reason, the climatic parameter set 11 is used for describing the effect of the different carbon cycle parameters. This information, included in the table caption, will be added in the revised version of the main text.

I'm afraid I don't find the experiment acronyms (first digit for climate sensitivity, second for MOC sensitivity third for CO2 sensitivity) very clear. The third digit seems to only be used in passing in section 7 anyway.

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,

[1] TSI = Total Solar Irradiance

	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.	

p730:19-20 CO2 sensitivity is clearly affecting the ocean heat content more than the MOC, but I think it's going too far to say on this evidence alone that CO2 sensitivity is dominating "ocean behaviour" as a whole.

We were maybe too optimistic in the wording of this section. Such a general conclusion can indeed not be drawn from the results of a single model. Even for a single model, it should be checked if the conclusion is robust, for example, for other forcings.

Would the authors recommend their metric for application to other model ensembles?

We would be glad if our metric proved to be useful to give a measure of how well the simulated trends fit the observationally-based estimates for other models than ours. However, we emphasize that the selected variables were subjectively chosen, depending on the goal of the study. Therefore, although the principle of the metric may be used for other studies, the metric itself must be adapted to the purpose of the study. For example, the temperature over Greenland or its melting rate could be used in the metric, some weight could also be given to each variable, although, at this stage, weight could only be given subjectively.

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

Could the metric score be used to usefully weight the results of individual members of a perturbed parameter ensemble, retaining information from all their parameter sets?

We do not have enough simulations to perform a weighted average of them. Moreover, we do not have a statistical method to attribute weight to each simulation and/or to each variable of a simulation. Finally, our subjective choices of the parameter sets and of the variables included in the definition of the metric do not allow us to perform appropriate statistics. However, our goal was much more simple. We aim at identifying (an eventually reject) parameter sets leading to completely unrealistic climate simulations.

p733 implies a significant issue with the heat uptake of the ocean in LOVECLIM. I would echo referee#1's comment in wondering whether a parameter-variation study such as this one might at least speculate on a parameter tuning that might alleviate the problem.

Unfortunately none of the parameter sets used in this study gives us a clue on how to improve the model results, as far as the heat uptake issue is concerned. Of course, it does not mean that it is impossible. Nevertheless, it tends to suggest that the problem is more intrinsic to the model than a 'simple' parameterisation. This is why we wrote "Varying the values of key physical parameters of LOVECLIM cannot solve clearly identified drawbacks as underlined by some systematic biases present with all the parameter sets. In other words, further tuning will most probably be relatively ineffective to improve the model behaviour to simulate past climates, at least for some variables and in some regions." So far, we diagnosed the problem but we have not yet found the solution. We haven't identified yet neither the mechanism that leads to this behaviour of the heat uptake nor its origin. Is it global or local (e.g. in the Southern Ocean)? This point requires further research.

Additional typographical comments

Will be corrected.

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:

Goosse, H., Driesschaert, E., Fichefet, T., and Loutre, M.F. (2007). Information on the early Holocene climate constrains the summer sea ice projections for the 21st century. Clim. Past, 3, 683-692.