Evaluating climate model performance with various parameter sets using observations over the recent past

M. F. Loutre¹, A. Mouchet², T. Fichefet¹, H. Goosse¹, H. Goelzer³, and P. Huybrechts³

¹Université catholique de Louvain, Earth and Life Institute, Georges Lemaître Centre for Earth and Climate Research (TECLIM), Chemin du Cyclotron, 2 bte L7.01.11, 1348 Louvain-la-Neuve, Belgium
²Laboratoire de Physique Atmosphérique et Planétaire, Université de Liège, Allée du 6 août, 17, Bâtiment B5c, 4000 Liège, Belgium
³Earth System Science & Departement Geografie, Vrije Universiteit Brussel, Pleinlaan, 2, 1050 Brussels, Belgium

Received: 6 April 2010 – Published in Clim. Past Discuss.: 29 April 2010
Revised: 19 April 2011 – Accepted: 20 April 2011 – Published: 17 May 2011

Abstract. Many sources of uncertainty limit the accuracy of climate projections. Among them, we focus here on the parameter uncertainty, i.e. the imperfect knowledge of the values of many physical parameters in a climate model. Therefore, we use LOVECLIM, a global three-dimensional Earth system model of intermediate complexity and vary several parameters within a range based on the expert judgement of model developers. Nine climatic parameter sets and three carbon cycle parameter sets are selected because they yield present-day climate simulations coherent with observations and they cover a wide range of climate responses to doubled atmospheric CO₂ concentration and freshwater flux perturbation in the North Atlantic. Moreover, they also lead to a large range of atmospheric CO₂ concentrations in response to prescribed emissions. Consequently, we have at our disposal 27 alternative versions of LOVECLIM (each corresponding to one parameter set) that provide very different responses to some climate forcings. The 27 model versions are then used to illustrate the range of responses provided over the recent past, to compare the time evolution of climate variables over the time interval for which they are available (the last few decades up to more than one century) and to identify the outliers and the “best” versions over that particular time span. For example, between 1979 and 2005, the simulated global annual mean surface temperature increase ranges from 0.24 °C to 0.64 °C, while the simulated increase in atmospheric CO₂ concentration varies between 40 and 50 ppmv. Measurements over the same period indicate an increase in global annual mean surface temperature of 0.45 °C (Brohan et al., 2006) and an increase in atmospheric CO₂ concentration of 44 ppmv (Enting et al., 1994; GLOBALVIEW-CO2, 2006). Only a few parameter sets yield simulations that reproduce the observed key variables of the climate system over the last decades. Furthermore, our results show that the model response, including its ocean component, is strongly influenced by the model sensitivity to an increase in atmospheric CO₂ concentration but much less by its sensitivity to freshwater flux in the North Atlantic. They also highlight weaknesses of the model, in particular its large ocean heat uptake.

1 Introduction

Policymakers are facing a wide range of possible scenarios for long-term climate and sea level evolution without knowing precisely why they differ and how reliable they are (e.g. IPCC, 2007; Knutti et al., 2008; Stainforth et al., 2007). There are indeed many sources of uncertainty in modelling experiments used in climate projections. Among others, there are uncertainties in the future anthropogenic emissions of greenhouse gases and aerosols (e.g. Nakicenovic and Swart, 2000; Meehl et al., 2007), and uncertainties in the boundary and initial conditions (e.g. Knutti et al., 2008). Moreover, climate models, and in particular the physical parameterisations they are using, are far from being perfect and the values of many physical parameters themselves are often poorly known (e.g. Stainforth et al., 2005; Murphy et al., 2004).
Several strategies can be used to assess those uncertainties. Model results can be analysed to quantify the errors in the simulations. For example, Gleckler et al. (2008) proposed objective measures of coupled ocean-atmosphere general circulation model performance according to several climatic variables. They evaluate the model performance according to 2-D climatic fields simulated for a given climate state. Therefore, they calculate the root mean square (RMS) errors for each model and each variable using two references. They define a “typical” model error, i.e. the median of the RMS error calculations for each climatic variable, which is used to normalize the RMS error for each variable. So, they obtain a measure of how well a given General Circulation Model (GCM) compares with the typical model error. In parallel, the modelled responses to different external forcings are utilised to illustrate the uncertainty related to the non-perfect knowledge of the forcing (e.g. Crowley, 2000; Bertrand et al., 2002). Another example of strategy to assess model uncertainty is given by Murphy et al. (2004) and Stainforth et al. (2005). They used the same model and the same forcings with varied values of key physical parameters to identify the range of the climate response to a CO2 doubling related to parameter uncertainty. Lastly, Knutti et al. (2008) gave another example of strategy. Based on several emission scenarios and coupled GCMs, they concluded that the contribution of structural uncertainties (i.e. the error related to the choices made in the model structure that would remain even if all the parameters were perfectly known) to temperature projection over the next century is quite large (Knutti et al., 2008).

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). The overall goal of this study is to design several alternative model versions that provide a wide range of climate responses to a climate forcing. Therefore, we identify a reasonable number of parameter sets that yield present climate simulations coherent with observations. Moreover, the various parameter sets should lead to a wide range of possible climate responses to increase in atmospheric CO2 concentration and to freshwater discharge in the North Atlantic. They will thus provide a reasonable sample for quantifying the uncertainty of future climate changes in forthcoming studies. This approach has been chosen rather than a systematic random variation of important parameters because the latter would imply a very large number of long simulations, which is not affordable even with a relatively fast model like LOVECLIM. Moreover, preliminary tests clearly showed that most parameter combinations lead to unrealistic present-day climate and therefore would be useless for the purpose of this study. In addition, using a restricted number of parameter sets allows a better knowledge of their characteristics and thus potentially offers a better understanding of the different responses.

After a brief model description (Sect. 2), the remainder of the paper is divided into two major parts. First, we present the design of the alternative model versions (Sect. 3). Twenty-seven combinations of key physical parameter values of LOVECLIM that have a large impact on the model results are selected and utilised to carry out transient experiments over the last millennium. In the second part (Sect. 4), we use these model versions to analyse the range of responses to given forcings over the recent past. In particular, we focus on the ability of the model to simulate the trend of key global climate variables over the last century and, therefore, we design a metric (i.e. a scalar measure) to quantify this ability. The last century was chosen because it corresponds to a period when relatively accurate observations are available for model-data comparison. Furthermore, as LOVECLIM has been (and is still) mainly used in process studies focused on mid- and high latitudes, we select variables that potentially have a direct or indirect impact on the evolution of sea level, on the stability of the North Atlantic meridional overturning circulation (MOC) and on the future of the climate of polar regions. We also select global variables that give a global view on climate change. Therefore, in addition to atmospheric CO2 concentration and surface temperature, we specifically assess the ability of the model to reproduce the observed trends in the Northern Hemisphere sea ice extent and global ocean heat content.

2 The climate model – description

LOVECLIM1.1 (further termed LOVECLIM) is a 3-D Earth System Model of Intermediate Complexity (EMIC). It consists of five components representing the atmosphere (EC-Bilt), the ocean and sea ice (CLIO), the terrestrial biosphere (VECODE), the oceanic carbon cycle (LOCH) and the Greenland and Antarctic ice sheets (AGISM). The ice sheet model AGISM (Huybrechts, 1990, 1996; Huybrechts and de Wolde, 1999) is not activated in this study because of the negligible influence of ice sheet-climate interactions on the climate evolution over the last century. Rather, its influence on future climate simulations is investigated in a separate study (Goelzer et al., 2010). The previous model version (LOVECLIM1.0) is described in Driesschaert et al. (2007), while version 1.2, which differs only very slightly from version 1.1, is presented in Goosse et al. (2010).

ECBilt (Opsteegh et al., 1998) is a quasi-geostrophic atmospheric model with 3 levels and T21 horizontal resolution that explicitly computes synoptic variability associated with weather patterns. It includes simple parameterisations of the diabatic heating processes and an explicit representation of the hydrological cycle. Cloudiness is prescribed according to present-day climatology. CLIO (Goosse and Fichefet, 1999) is a primitive-equation, free-surface OGCM coupled to a thermodynamic-dynamic sea ice model. Its horizontal resolution is 3° × 3°, and there are 20 levels in the ocean.
VECODE (Brovkin et al., 2002) is a reduced-form model of the vegetation dynamics and of the terrestrial carbon cycle. It simulates the dynamics of two plant functional types (trees and grassland) at the same resolution as that of EC-BILT. LOCH (Mouchet and François, 1996; Mouchet, 2011) is a comprehensive oceanic carbon cycle model that includes an atmospheric module to represent the evolution of CO$_2$, $^{13}$CO$_2$, and $^{14}$CO$_2$ in the atmosphere. LOCH is fully coupled to CLIO and runs with the same time step and on the same grid. LOVECLIM has been utilised in a large number of climate studies (e.g. Driesschaert et al., 2007; Goosse et al., 2007; Menviel et al., 2008a,b) and was part of several model intercomparison exercises (e.g. Braconnot et al., 2002, 2007a,b; Dutay et al., 2004).

3 Alternative model versions

3.1 Introduction – parameter sets

Several physical parameters of the model may significantly impact the model response to an external perturbation. We performed more than one hundred simulations using combinations of parameters. These simulations were designed to lead to contrasted responses to a doubling of CO$_2$ concentration and to additional freshwater flux in the North Atlantic, and to induce different responses of the carbon cycle model. Amongst them, we selected those that produced reasonable simulations of the present-day climate. Eventually, we kept nine climatic parameter sets and three carbon cycle parameter sets, which makes 27 parameter sets (see Sect. 1 of the Supplement for the description of the parameter sets; a three digit code identifies the parameter set, the first two digits correspond to the climatic parameter sets, and the third one to the carbon cycle parameter set).

3.2 Sensitivity

3.2.1 Sensitivity to CO$_2$ concentration

A first sensitivity experiment (prefix E, suffix 2CO, Table 1) is performed starting from the equilibrium state simulated under pre-industrial conditions (Sect. 2 of the Supplement). The atmospheric CO$_2$ concentration is increased by 1% per year from the pre-industrial value until doubling, i.e. after 70 years. It is thereafter held constant (Fig. 1, left). This experiment provides a clear and strong climate signal as well as a good insight into the response of the atmosphere under perturbed conditions. Figure 1 (right) displays the temperature evolution during the first 2000 year time interval of the experiment. The rate of change is largest over the first 70 years of the simulation, when atmospheric CO$_2$ concentration is increasing. The increase in global annual mean surface temperature after 1000 years in this sensitivity experiment is translated into an index to characterise the model in terms of response to the prescribed perturbation (climate sensitivity). Its value ranges from 1 to 5, corresponding to a temperature increase from less than 2.0 °C to more than 3.5 °C (by step of 0.5 °C) in the experiment described above.

**Fig. 1.** Atmospheric CO$_2$ concentration in the perturbation scenario (left) and time evolution of the global annual mean surface temperature in response to this perturbation according to the selected model parameter sets (right). Temperature is presented as deviation from the initial value. The colour code for the parameter sets is given in the figure.
3.2.2 Sensitivity to water hosing

In a second sensitivity experiment (prefix E, suffix HYS, Table 1), freshwater is added in the North Atlantic (20° -50° N) with a linearly increasing rate of $2 \times 10^{-4} \text{ Sv yr}^{-1}$. This results in a freshwater perturbation of 0.1 Sv after 500 years, 0.2 Sv after 1000 years, and 0.3 Sv after 1500 years (Fig. 2, left). This simulation, which allows assessing the stability of the North Atlantic MOC, provides a good insight into the response of the ocean under perturbed conditions and can be compared with simulations performed with other models in similar conditions (e.g. Rahmstorf et al., 2005; Weber et al., 2007). The percentage of decrease in the maximum value of the meridional overturning streamfunction below the Ekman layer in the Atlantic Ocean after 1000 years in this water hosing experiment (at the time the perturbation reaches 0.2 Sv)
is chosen to characterise the response of the model to this perturbation (MOC sensitivity). The MOC sensitivity is reflected in the second digit of the name of the experiments: 1 for a decrease in the maximum value of the meridional overturning streamfunction of less than 50 %, and 2 otherwise. LOVECLIM with parameter set 112, i.e. the closest to LOVECLIM 1.0 used in Driesschaert et al. (2007), simulates a 20 % reduction in the meridional overturning streamfunction after 1000 years. This decrease ranges from 19 to 56 % for the other parameter sets (Table 2).

Lastly, Fig. 3 confirms that the phase space (MOC sensitivity vs. climate sensitivity) of our set of experiments is rather homogeneously covered as required by our initial objective.

### 3.2.3 Sensitivity of the carbon cycle

We assess the sensitivity of the atmospheric CO₂ level to the choice of carbon cycle parameters by performing a prognostic CO₂ experiment (prefix E, suffix TRA, Table 1) for each of the three parameter sets. This transient simulation starts from an equilibrium state corresponding to the conditions prevailing in 1750 AD (all years are in AD). It runs until year 3000 and is constrained by changes in the Earth orbital parameters (Berger, 1978) and in concentrations of greenhouse gases (GHGs) except CO₂. In addition, the model is forced by anthropogenic emissions of CO₂, including both fossil fuel and deforestation fluxes. Over the historical period (1750–2000), the GHG concentrations (Houghton et al., 2001) and carbon emissions (Marland et al., 2003; Houghton, 2003) follow the historical records. From 2000 to 2100, we use the SRES A2 scenario (Houghton et al., 2001) for both carbon emissions and GHG concentrations. After 2100, concentrations of all GHGs (except CO₂) are kept fixed to their 2100 values, while CO₂ emissions from land use are set to zero and fossil fuel emissions decrease according to a bell-shaped curve so that they reach zero a few decades after 2200 (Fig. 4, top left).

The three carbon cycle parameter sets (Table 3) lead to contrasted responses of the atmospheric CO₂ to the identical forcing (Fig. 4, top right). Maximal values of the atmospheric CO₂ concentration differ by up to 169 ppmv between carbon sets 1 and 3 (Table 3). By year 2500, they still differ by 133 ppmv, i.e. a relative difference of about 11 %. With carbon cycle parameter sets 1 and 2, the land CO₂ uptake outpaces the ocean uptake (Fig. 4, bottom left), while the reverse happens with carbon parameter set 3.

The parameters related to the continental vegetation processes explain up to 87 % of the difference in atmospheric CO₂ response between the various experiments. On such time scales, changes in the rain ratio or in the export production within the ocean have a much smaller impact on the atmospheric CO₂. The contribution of the rain ratio to

---

### Table 2. Main features of the model climate using different parameter sets (first column).

<table>
<thead>
<tr>
<th>Name</th>
<th>TCR</th>
<th>CS</th>
<th>Ceff</th>
<th>Equi</th>
<th>MOC sensitivity</th>
<th>MOC</th>
<th>Ts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>°C</td>
<td>°C</td>
<td>°C</td>
<td>°C</td>
<td>%</td>
<td>Sv</td>
<td>°C</td>
</tr>
<tr>
<td>112</td>
<td>0.7</td>
<td>1.6</td>
<td>1.6</td>
<td>1.6</td>
<td>−19</td>
<td>28.4</td>
<td>16.1</td>
</tr>
<tr>
<td>122</td>
<td>0.8</td>
<td>1.9</td>
<td>1.9</td>
<td>2.0</td>
<td>−52</td>
<td>17.3</td>
<td>15.8</td>
</tr>
<tr>
<td>212</td>
<td>0.8</td>
<td>2.1</td>
<td>2.2</td>
<td>2.2</td>
<td>−29</td>
<td>25.6</td>
<td>15.8</td>
</tr>
<tr>
<td>222</td>
<td>0.9</td>
<td>2.2</td>
<td>2.2</td>
<td>2.3</td>
<td>−56</td>
<td>21.5</td>
<td>15.6</td>
</tr>
<tr>
<td>312</td>
<td>0.9</td>
<td>2.6</td>
<td>2.7</td>
<td>2.9</td>
<td>−20</td>
<td>25.1</td>
<td>16.4</td>
</tr>
<tr>
<td>322</td>
<td>1.0</td>
<td>2.9</td>
<td>3.1</td>
<td>3.3</td>
<td>−55</td>
<td>20.9</td>
<td>15.7</td>
</tr>
<tr>
<td>412</td>
<td>1.1</td>
<td>3.2</td>
<td>3.5</td>
<td>3.7</td>
<td>−45</td>
<td>24.0</td>
<td>15.9</td>
</tr>
<tr>
<td>512</td>
<td>1.5</td>
<td>4.0</td>
<td>4.4</td>
<td>4.5</td>
<td>−25</td>
<td>23.9</td>
<td>16.1</td>
</tr>
<tr>
<td>522</td>
<td>1.5</td>
<td>4.1</td>
<td>4.8</td>
<td>4.8</td>
<td>−54</td>
<td>19.9</td>
<td>15.5</td>
</tr>
</tbody>
</table>

(1) increase in global annual mean surface temperature after 70 years from the pre-industrial equilibrium value in the doubling CO₂ experiment;
(2) increase in global annual mean surface temperature after 1000 years from the pre-industrial equilibrium value in the doubling CO₂ experiment;
(3) the effective climate sensitivity according to Gregory et al. (2002) (see also Goelzer et al., 2010);
(4) the equilibrium response in global annual mean surface temperature is computed after 2000 years for the parameter sets 112, 122, 212, and 222; and after 3300 years for the parameter sets 312, 322, 412, 512, and 522;
(5) percentage of decrease in the meridional overturning streamfunction after 1000 years in the water hosing experiment;
(6) strength of the meridional overturning streamfunction in the North Atlantic (Sv) at equilibrium in the pre-industrial experiment;
(7) annual mean global surface temperature (°C) at equilibrium in the pre-industrial experiment.
Fig. 4. CO₂ emission scenario (top, left) used to assess the sensitivity of the carbon cycle to the different carbon cycle parameter sets (see description of the scenario in the text). It includes both fossil fuel emission and fluxes related to land use change. Evolution of the annual mean atmospheric CO₂ concentration (ppmv) with time (top, right), terrestrial carbon inventory versus ocean carbon inventory (both in GtC) (bottom, left), and atmospheric CO₂ versus the global annual mean surface temperature (bottom, right) for the different carbon cycle parameter sets. The dashed line in the bottom left panel represents the 1:1 slope. Inventories are presented as anomalies with respect to the control run. The same colour code is used in each panel, i.e. black for parameter set 111, green for set 112, and red for set 113.

Table 3. Model parameter sets for the carbon cycle and their effect on the CO₂ response. These parameters influence the continental vegetation fertilization effect ($\beta_{g}$ and $\beta_{t}$; columns 2 and 3), the vertical flux of POM ($\alpha_{\text{diatom}}$ and $\alpha_{\text{others}}$, columns 4 and 5), and the buildup of calcium carbonate shells ($\Psi_{\text{zoo}}$, column 6). Columns 7 and 8 give the maximum value of the annual mean atmospheric CO₂ concentration and its value at year 2500 from the transient simulations (see text) with the three carbon cycle parameter sets.

<table>
<thead>
<tr>
<th>Carbon parameter set</th>
<th>$\beta_{g}$</th>
<th>$\beta_{t}$</th>
<th>$\alpha_{\text{diatom}}$</th>
<th>$\alpha_{\text{others}}$</th>
<th>$\Psi_{\text{zoo}}$</th>
<th>Max</th>
<th>2500 AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.14</td>
<td>0.50</td>
<td>$-0.750$</td>
<td>$-0.950$</td>
<td>1146</td>
<td>11146</td>
<td>877</td>
</tr>
<tr>
<td>2</td>
<td>0.36</td>
<td>0.36</td>
<td>$-0.858$</td>
<td>$-0.858$</td>
<td>22</td>
<td>1202</td>
<td>918</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>0.22</td>
<td>$-0.648$</td>
<td>$-0.648$</td>
<td>22</td>
<td>1315</td>
<td>1010</td>
</tr>
</tbody>
</table>

The maximum value of the atmospheric CO₂ range is about 10%, while changes in oceanic remineralization depth explain about three percent. Such small changes (a few ppmv) are within the variability produced by the model and cannot be ascertained yet. All together, the three parameter sets allow us to obtain a change in the carbon climate sensitivity (as defined in Frank et al., 2010) of the order of 7% (Fig. 4 bottom right).

The third digit in the experiment name refers to the carbon cycle parameter set with relatively low (1), medium (2), or high (3) changes in atmospheric CO₂ in response to the same emission scenario.

4 The climate of the last century

4.1 The simulations

In this section, we study the climate simulated over the last century using all combinations of the different parameter sets (see Table 4 for the name of the different experiments). Our purpose is to quantify the ability of each parameter set to simulate climate changes over the last century, or over shorter periods, for which accurate observations are available. For the analysis of the simulated climate changes, we consider the average over an ensemble of five members in order to reduce the impact of internal variability. Each member consists
of a simulation of the climate of the last century starting in 1900 from the state at 1900 of a climate simulation of the last millennium performed with the same parameter set (Sect. 3 of the Supplement). The members of one ensemble differ only in their initial conditions. To do so, we have introduced a very small perturbation in the quasi-geostrophic potential vorticity the first day of the simulation, as described in Goosse et al. (2007).

The evolution of the atmospheric CO$_2$ concentration is either diagnostic or prognostic. In the diagnostic mode (Conc), the atmospheric CO$_2$ concentration is prescribed according to Enting et al. (1994) until 1990, and then according to GLOBALVIEW-CO2 (2006) (Fig. 5). For the prognostic mode (Efor), the atmospheric CO$_2$ concentration is computed by forcing the model with emissions of CO$_2$ from fossil fuel burning (Fig. 5, Marland et al., 2003). Both simulations also take into account land use changes related to human activities as in Goosse et al. (2005) (percentage of crops; Ramankutty and Foley, 1999; Pongratz et al., 2008). We assume that croplands replace only forests, as long as there is a forest fraction. Furthermore, desert and forest (except for the part replaced by crops) keep their original extent at year 500. This scenario was previously used in a model intercomparison exercise aiming at analysing the response of six EMICs, including ECBilt-CLIO-VECODE, to historical deforestation (Brovkin et al., 2006).

In addition to the atmospheric CO$_2$ concentration, either prescribed or computed by the model (see above, Enting et al., 1994; GLOBALVIEW-CO2, 2006; Marland et al., 2003), the transient simulations are forced by the volcanic activity (Crowley, 2000), the solar activity (Muscheler et al., 2007), the Earth orbital parameter changes (Berger, 1978; Bretagnon, 1982), and changes in concentrations of GHGs other than CO$_2$ (Prather et al., 2001; Houghton et al., 1990 and updates).

The effect of sulphate aerosols is accounted for through a modification in surface albedo, as suggested by Charlson et al. (1991) (scenario S1). For 150 years, human activities have increased the sulphate aerosol load in the troposphere (Houghton et al., 2001). However, the magnitude of
its effect on the Earth climate is difficult to estimate. The radiative forcing computed by LOVECLIM for the present day with respect to the pre-industrial era related to the sulphate aerosol load is $-0.4 \text{ Wm}^{-2}$ in the reference situation (climatic parameter set 112). However, there is a large uncertainty in this quantity. IPCC AR4 (Forster et al., 2007) reported a direct radiative forcing due to sulphate aerosols of $-0.40 \pm 0.2 \text{ Wm}^{-2}$. The overall aerosol direct radiative forcing (i.e. radiative forcing values associated with several aerosol components) was estimated to $-0.50 \pm 0.40 \text{ Wm}^{-2}$. In addition to a direct effect, aerosol particles also affect the formation and properties of clouds. IPCC AR4 gives a median value of $-0.70 \text{ Wm}^{-2}$ for the cloud albedo radiative forcing due to aerosol influence on clouds. Therefore, we decided to perform a second set of simulations for which the radiative forcing related to sulphates is doubled, $-0.8 \text{ Wm}^{-2}$ in the reference situation (climatic parameter set 112) (scenario S2).

4.2 Results

4.2.1 CO$_2$ concentration

The comparison of the simulated time evolution of the atmospheric CO$_2$ concentration over the last century with data shows that some parameter sets display a poorer agreement than others (Fig. 6). In particular, the simulated increase in CO$_2$ concentration obtained with carbon cycle parameter set 3 is of the order of 10 ppmv larger than in the corresponding observations over the 20th century. In contrast, the simulated increase in atmospheric CO$_2$ concentration remains close to the measured one for carbon cycle parameter sets 1 and 2.

Similar conclusions can be reached by analysing the rate of increase in CO$_2$ concentration over different periods. Between 1959 and 2008, it varies between 1.35 and 1.47 ppmv yr$^{-1}$ for carbon cycle parameter sets 1 and 2, with the nominal (S1) sulphate forcing. Furthermore, the rate is higher with carbon cycle parameter set 3 ($\approx 1.58 \text{ ppmv yr}^{-1}$) as well as when the S2 sulphate forcing is applied (by about 0.03 ppmv yr$^{-1}$). It is in reasonable agreement with the corresponding value in the Mauna Loa record (NOAA ESRL, 2009) of 1.44 ppmv yr$^{-1}$. A comparison with another
4.2.2 Surface temperature

The increasing trend in global annual mean surface temperature computed from HadCRUT3 time series (Brohan et al., 2006) is 0.0168 °C yr\(^{-1}\) over the last 35 years (1979–2005) and 0.0071 °C yr\(^{-1}\) over the last century (1901–2005). Some parameter sets lead to an underestimate of this increasing trend. This is, for example, the case for the climatic parameter sets 11, 21, and 22, especially with the S2 sulphate forcing; while other climatic parameter sets yield an overestimate of this trend, e.g. 51 and 52, especially with the S1 sulphate forcing.

4.2.3 Minimum sea ice extent

Most of the simulations, either with S1 or S2 sulphate aerosol forcing, experience a too small decrease in Northern Hemisphere minimum sea ice extent between 1979 and 2006 compared to observations (Fig. 7). This is especially the case for those simulations with low climate sensitivity (climatic parameter sets 11, 12, 21, 22). For higher sensitivities, the type of simulation (Efor or Conc), the sulphate aerosol load (S1 or S2) as well as the sensitivity to the carbon cycle may play a role in the simulated trend. However, larger sulphate aerosol concentrations do not systematically lead to lower or higher trend in Northern Hemisphere minimum sea ice extent.

4.2.4 Oceanic variables

Most of the simulations overestimate the estimated warming of the global ocean in the 700 m upper layer over the last 50 years (Levitus et al., 2009) when the S1 sulphate forcing is used (Fig. 8). This overestimation is strongly reduced for S2 sulphate forcing. Indeed, in that case, only the simulations with high climate sensitivity (climatic parameter sets 51, 52, as well as 32 and 41 for some experimental setups) exhibit an ocean heat content increase significantly larger than in the real world.

The modelled ocean circulation does not experience major changes during the last century. Over this period, all the simulations show a reduction of less than 4 Sv in the
Fig. 8. Trend in ocean heat content in the upper 700 m \((10^{22} \text{ J yr}^{-1})\) wrt trend in sea surface temperature \((\text{C yr}^{-1})\). Each symbol corresponds to one simulation, either for S1 (left) or S2 (right) sulphate aerosol forcing. Trends are computed as the slope of the regression line through the annual values between 1955 and 2007. The colour code refers only to the climatic parameter sets, i.e. the first two digits of the parameter set name. Full symbols are for carbon cycle parameter set 1, half-filled symbols are for carbon cycle parameter set 2, and empty symbols are for carbon cycle parameter set 3. The full name of the parameter set is obtained by appending the number corresponding to the carbon cycle parameter set (i.e. 1, 2, or 3 according to the symbol) to the number corresponding to the climate parameter set given by the colour code. Squares (triangles down) correspond to Efor (Conc) simulations. The full black line represents the trend computed from observation (Levitus et al., 2009). The dashed line represents the uncertainty related to the variability in the data (one standard deviation on the slope of the linear regression through observation.

The surface temperature changes simulated over the last century obviously depend on the climate sensitivity. The parameter sets corresponding to the lowest climate sensitivity (such as climatic parameter sets 11, 21, and 22) lead to small temperature changes over the last century and those with the largest climate sensitivity (e.g. climatic parameter sets 51 and 52) lead to a large temperature increase over the last century. Moreover, using a larger sulphate aerosol forcing tends to shift the simulated temperature increase over the last century towards smaller values because of the radiative cooling effect of those aerosols. Still, the discrepancy between simulated global annual mean surface temperature and observations remains small (within one standard deviation) in many cases.

Moreover, although the deviation from observations of the simulated atmospheric \(\text{CO}_2\) concentration is of the order of 10 ppmv over the 20th century (Fig. 6) for carbon cycle parameter set 3, this discrepancy is not large enough to drive the surface temperature towards larger values than for carbon cycle parameter set 1 or 2. Therefore, most of the simulations with carbon cycle parameter set 1 or 2 remain close to temperature observations, while those using carbon cycle parameter set 3 display only a small disagreement.

The simulations performed here display an approximately linear relationship between the increase in the upper ocean heat content and the increase in sea surface temperature (Fig. 8), i.e. when temperature increases, in particular sea surface temperature, the ocean captures more heat. We speculate that a too large ocean heat uptake leads to a deficit in energy available at the ocean surface for melting the sea ice simulated with several parameter sets.

The relationship between increase in atmospheric \(\text{CO}_2\) concentration and the North Atlantic MOC was studied in several GCMs and EMICs (including LOVECLIM) (Gregory et al., 2005). These authors performed partially coupled integrations to evaluate the influence of heat and freshwater in each of the models. They pointed out that heat flux changes generally contribute more than freshwater flux changes to weakening the MOC for all models. We also find an approximately linear relationship between the sea surface temperature and the North Atlantic MOC intensity (Fig. 9). In contrast, there is no clear relationship between the change in upper ocean heat content and MOC sensitivity. Therefore, the climate sensitivity has a stronger effect on the ocean over
the 20th century than MOC sensitivity. In other words, even though we selected parameter sets in a large phase space, the ocean is responding more to the atmospheric forcing than to its intrinsic characteristics over the last few decades. The initial states of the ocean, that are different depending on the parameter sets, do not induce large changes in the upper ocean heat content either. Of course we should verify that this conclusion, drawn only from LOVECLIM simulations, is robust for both other models and other forcings.

4.3 Performance of the parameter sets

Although none of the selected parameter sets is able to yield a climate simulation in the range of observations for all the variables examined, some parameter sets perform better than others. The purpose of this section is to characterise (and rank) them according to their performance. Therefore, we designed a metric that quantifies the ability of a simulation (i.e. a given parameter set and a given configuration) to simulate the observed climate change over the last century. This metric is a measure of how well the simulated trends fit the observationally-based estimates of several climatic indicators during the 20th century. Indeed, as long as we are interested in climate change, it is more important to simulate a correct evolution of the variables, rather than a correct value of any time. The metric is based on the same variables as those discussed in the previous section (global annual mean surface temperature, atmospheric CO₂ concentration, minimum sea ice extent in the Northern Hemisphere, and ocean heat content of the upper 700 m of the global ocean). The design of the metric is explained in Sect. 4 of the Supplement. Each simulation (i.e. given parameter sets, sulphate forcing, and setup) receives a score. None of the simulations received the maximum score of four (Conc) or six (Efor) points. The best simulations received a total score of three (Conc) and four (Efor) points (Fig. 10).

Simulations with the carbon cycle parameter set 3 do not properly reproduce the observed atmospheric CO₂ increase. Still, the deviation from observations remains less than 10 ppmv over the last 50 years and this does not prevent simulation of temperature increase in agreement with observations. Moreover, none can simulate simultaneously a correct time evolution for the ocean heat content in the upper 700 m and for the Northern Hemisphere sea ice extent. Goosse et al. (2007) studied the time evolution of the Northern Hemisphere sea ice extent in transient simulations from 8 kyr BP to 2100 AD, starting from an equilibrium state at 8 kyr BP, and using five parameter sets corresponding to 112, 212, 312, 412, and 512. They showed that, compared to observations covering the second half of the 20th century, parameter sets 112 and 212 seriously underestimate the decline in summer sea ice extent, while parameter set 312 slightly underestimates it. Therefore, Goosse et al. (2007) concluded that parameter sets 112 and 212 are incompatible with the observed record. This is well in line with our analysis.

The aerosol forcing scenario (S1 or S2) has a strong impact on the skill of a parameter set to reproduce the climate change for a given variable. For example, more parameter sets perform well in reproducing the ocean heat content trend under S2 than S1. On the contrary, the temperature increase
Fig. 10. Summary of the performance of the Conc (top) and Efor (bottom) simulations to reproduce the observed trend of the time evolution for different climate variables for each parameter set under S1 (left) and S2 (right) sulphate aerosol forcings. The variables and the time intervals are described in Sect. 4 of the Supplement. The x-axis lists all the parameter sets. Colour bars indicate the variables (see colour code) simulated with a good skill (according to our metric), i.e. $R$ above the threshold (see text). A single colour bar is used for sea ice and upper ocean heat content. Ts stands for global annual mean surface temperature either over the interval 1901–2005 or 1979–2005. CO$_2$ is for atmospheric CO$_2$ concentration either over the time interval 1901–2005 or 1979–2005. Sea ice extent trend is computed either between 1979 and 2006 or 1979 and 2007. Trend in ocean heat content in the upper 700 m of the ocean is computed over either the time interval 1955–2007 or 1950–2003. See also the Supplement for references.

over the 20th century and the last decades is better simulated with S1 than S2.

When atmospheric CO$_2$ concentration is prescribed (Conc), most of the parameter sets, except those with the lowest climate sensitivity, are able to reproduce the observed temperature trend over the 20th century. The trend in upper ocean heat content remains within 66% of the median of the deviation from observations for the low climate sensitivity parameter sets when S1 aerosol forcing is used. It is also true for a few more parameter sets when S2 aerosol forcing is used.

Generally speaking, simulations with high climate sensitivity (climatic parameter sets 32, 41, 51, 52) have a better global score than simulations with low climate sensitivity. Amongst the simulations ranking the highest (i.e. a final mark of 3 for Conc and 4 for Efor), parameter set 321 is the only one performing well for both Conc and Efor, as well as for both S1 and S2 sulphate aerosol forcings. Moreover, other parameter sets (322, 511, 512) also display good performance for both Conc and Efor, but only either for S1 or S2 sulphate aerosol forcing.

Parameter set 321 performs particularly well. Its only major weakness is the simulation of the evolution of the upper ocean heat content in the case of S1 sulphate forcing and long-term temperature (and CO$_2$) in the case of S2 sulphate aerosol forcing. Simulating an increase of the upper ocean
heat content in line with observations is also a major problem for the other “good” parameter sets (except for parameter set 322 under Conc-setup with S2 sulphate aerosol forcing).

The difficulty of simulating properly the increase in the upper ocean heat content is a rather general feature of all the simulations, especially those with high climate sensitivity. The parameter sets selected as having a “good skill” to reproduce the 20th century climate trend are those allowing a good simulation of the atmospheric temperature increase of the last century and the last decades of that century. However, their skill in reproducing the increase in the upper ocean heat content is much poorer. Conversely, the parameter sets leading to a good representation of the trend in upper ocean heat content lead to a too weak global warming over the last century and last decades.

It is worth mentioning that a good skill over the last decades, as measured by the metric, does not guarantee a good skill over the entire last century. This is particularly true for the temperature changes for most of the parameter sets. On the other hand, as already underlined, most of the parameter sets do not allow accurately capturing the CO2 trend over the last decades, although the deviation is small. Moreover, Efor experiments have two additional degrees of freedom compared to Conc. In Efor, the atmospheric CO2 is prognostic as well as the carbon emissions resulting from the land-use changes, while, in Conc, the latter flux is imposed and is the same for all climatic parameter sets. Since different climatic parameter sets lead to different vegetation distributions, the CO2 emissions in Efor may differ among parameter sets. This results in different atmospheric CO2 levels, which in turn result in a different vegetation response since the latter also depends on CO2 levels through the fertilization term. Lower atmospheric CO2 concentration generates a weaker carbon emission (and vice versa). Indeed, the emission is calculated on the basis of the potential growth of trees, which is favoured by higher atmospheric CO2 concentrations. This may explain the change in performance between Conc and Efor experiments, such as for parameter 412 (with S2 forcing), which is very poor in the Conc-simulation, while it performs well in the Efor-simulation.

For each variable, the metric gives only a binary result, either good or bad agreement between simulation and data. However, the discrepancy may be weak or strong, even for parameter sets that exhibit a good overall skill. For example, parameter set 321 (under Conc-setup), which has a good global score, displays a very strong disagreement for the upper ocean heat content with S1 sulphate aerosol forcing. Conversely, parameter set 211 displays a poor global skill, although the disagreement between model and data is only weak for most of the variables. This highlights how critical the choice of the threshold value is that separates between “good” and “poor” agreement.

5 Conclusions

This work is part of a study that aims at the quantification of uncertainties in modelling experiments used in climate change projections. Different approaches could be used, such as using various models or different external forcings. Here, we used different values for selected parameter sets of a particular model (LOVECLIM). In this way, we create alternative versions of the model. More precisely, we selected 27 parameter sets (nine climatic parameter sets and three carbon cycle parameter sets) according to their ability (1) to cover a large range of potential climate behaviours over the next millennium, and (2) to properly simulate the major features of the present-day climate. This small and manageable number of parameter sets was selected because: (i) we broadly knew the individual effect of each parameter on the modelled climate; (ii) we knew that each parameter set would lead to realistic simulated present-day climate; and (iii) we knew for sure that they would yield a range of different model behaviours according to the model sensitivity to an increase in atmospheric CO2 concentration, to its response to a freshwater hosing, and to its sensitivity to carbon cycle.

We designed a metric in order to quantify the skill of the different parameter sets to reproduce the climate change over the last decades of the 20th century, with a specific focus on global annual mean surface temperature, atmospheric CO2 concentration, minimum Northern Hemisphere sea ice extent, and upper ocean heat content. Indeed, when designing this metric, we had in mind to simulate the evolution of the ice sheets and sea level in the future (Goelzer et al., 2010), and our metric is therefore based on variables chosen in line with this final purpose. However, another set of variables, for example giving more weight to the ocean or the ice sheet component of the system, could give rise to a slightly different conclusion about the skill of the parameter sets.

We then rank the model versions according to their ability to simulate the past climate changes. None of the parameter sets are able to reproduce the observed trend of all the chosen key variables of the climate system (e.g. surface temperature, atmospheric CO2 concentration, sea ice extent, ocean heat content). Nevertheless, parameter sets 321, 322, 511, and 512 display good performance for more than one simulation setup (i.e. S1Conc, S2Conc, S1Efor or S2Efor). Moreover, parameter sets 321 and 322 are able to simulate a relatively good agreement for surface temperature and ocean heat uptake trends. Other parameter sets (e.g. 311, 412, 521, 522, 523) have only a slightly less good score. Therefore, we can use this work to reduce the number of parameter sets, keeping only the best ones in simulating climate with LOVECLIM and the model will still cover a wide range of realistic responses to given forcings. Alternatively, we can keep them all, although it must then be kept in mind that some yield a less realistic behaviour, at least for some components of the climate system. This second alternative allows coverage of more extreme cases that must be assessed in light of
the simulation of past climates. Another possibility to be explored is to give a weight to each simulation, with less weight given to a simulation performed with a low skill parameter set. However, in that case, additional simulations should be performed and more variables should be included in the design of the metric.

We also noted that the climate sensitivity seems to have a stronger impact on the simulated climate, and even on the ocean behaviour, than the mean ocean state or the model response to a freshwater hosing. Of course, this conclusion applies for LOVECLIM, within the framework of the forcing and parameter study performed here. It should be checked whether it is robust for other models, other parameterisations, or other forcings.

By using one single model, we did not address the structural uncertainty (related to the choice made during the build-up of the model) that can also be a major source of discrepancy between model results. Ideally, both types of uncertainty should be addressed together, in addition to the ones associated with the forcing, but this long-term goal is clearly outside the scope of the present study. Even though we tested a large number of values for different key physical parameters of the model (many more than finally used in this study), we were unable to clearly solve identified drawbacks as underlined by some systematic biases present with all the parameter sets, such as the strong ocean heat uptake. Therefore, we are convinced that further tuning will be relatively ineffective to improve the model behaviour to simulate past climates, at least for some variables and in some regions. Hence, improving the model probably requires improving the physics rather than (or in addition to) improving the values of its parameters. Given those biases, it is clearly inappropriate for the ensemble we have built to be used for making sound estimates of uncertainty in climate predictions/projections at the decadal-to-century time scale. Nevertheless, we feel that this ensemble is diverse and realistic enough to test the effect of the differences in model sensitivity, which are poorly constrained and vary largely among GCMs and EMICs, on the long-term response of the Earth’s system to the greenhouse gas forcing. In Goelzer et al. (2010), for instance, it has been utilized to investigate the impact of fully interactive Greenland and Antarctic ice sheets under greenhouse warming conditions on the climate sensitivity at the millennial time scale.

Supplementary material related to this article is available online at:

Acknowledgements. The Belgian Federal Science Policy funded this research within its Research Programme on Science for a Sustainable Development (SD/CS/01). We thank Pierre-Yves Barriat for helping in developments of the computer code. H. Goosse is Research Associate with the Fonds National de la Recherche Scientifique (Belgium). Computer time was partly made available by UCL (Fonds spéciaux de recherche) and FNRS (Fonds de la recherche fondamentale collective) FRFC N 2.4502.05: “Simulation numérique. Application en physique de l’état solide, océanographie et dynamique des fluides”. We thank Julia C. Hargreaves and two anonymous reviewers who provided helpful comments on the manuscript. We also thank the CISM-team.

Edited by: V. Rath

References


