

Response to Anonymous Referee #1

The manuscript describes a coupled emulation approach called CLISEMv1.0, which builds two separate emulators for an ice sheet model (AISMPALEO) and a climate model (HadSM3) – The outputs from these two emulators provide inputs to each other, enabling synchronous simulation of ice sheet evolution and climate changes. The authors conducted several sensitivity analyses regarding how the two emulators are built (e.g. depending on how the ice sheet input for the climate emulator is defined), how long the coupling time is, and how the lapse rate is adjusted to account for the elevation difference between the climate model grid and the ice model grid. While the coupled emulation approach itself is scientifically highly important and perhaps long overdue, the current coupled emulation results shown in Sections 3 and 4 need a lot of further clarification before the manuscript can be considered for being published in GMD. More specific comments are listed below.

Author's response: We thank the reviewer for his critical analysis, which has definitely helped us to improve the clarity and quality of the manuscript. In the revised version, we give more explanations about why certain model set-up choices are made. Also, we have been trying to be more precise about the influence of this choices on the model performance.

Major comments:

1. The description about the three different experimental design setups (EMULATOR_70, EMULATOR_100a, and EMULATOR_100b) in Section 2.3 and the result of coupled experiments described in Section 3.1 are contradicting to each other. According to Section 2.3, “EMULATOR_70 has a good spread between the different ice volumes and ice areas” but somehow the results described in Section 3.1 and also shown in Figure 9 indicate that the design EMULATOR_70 seems to have some serious flaw. The other two are described as designs with some notable flaws in Section 2.3, but somehow lead to better results. The manuscript gives some brief description on this issue in Section 3.1, but the authors did not really get to the bottom of the issue – In fact I cannot find any good rationale for how the ice model settings for EMULATOR_100a, and EMULATOR_100b are determined at the beginning – Why did the author decided to let EMULATOR_100a have “more small ice sheet geometries (ice volumes) compared to EMULATOR_100b (Figure 4) and has a good spread for the ice area of the input ice sheet geometries” and EMULATOR_100b be “poorly defined by ice sheet area as there are several experiments with the same ice sheet area yet different ice sheet geometry, but is well defined for ice sheet volume”? Are they some data of opportunity from some other experiments? Or did the author gradually add more model runs to these to designs until they give some sensible results shown in Figure 9? I think the authors need to describe their decision making process behind these design points in detail.

Author's response: We agree that the explanation for the ‘bad’ and ‘good’ performance of the emulator was not stated in a way that was clear enough for the reader. The performance of the emulator is primarily dependent on the number of ice sheet geometries the emulator is calibrated on and on the spacing of the ice sheet geometries. Other tuning factors are of secondary importance. Even though EMULATOR_70 has a good spread for the different ice sheet geometries concerning ice volume and ice area, the eight different input ice sheet geometries are not enough to initiate the transition to a continental scale glaciation. EMULATOR_100a and EMULATOR_100b - both calibrated on precursor climate model runs including twelve different ice sheet geometries appeared to be sufficient to induce the transition to glacial conditions.

To increase the clarity of the manuscript with respect to the different model choices made and the model performance, we introduce in the revised manuscript another emulator in which 100 climate model runs are run based on 20 different input ice sheet geometries with a good spread for both ice volume and ice area. This way, it is tested whether adding more ice sheet geometries is improving the model performance. The number of different input ice sheet geometries is limited by the climate model resolution, adding more geometries only adds information if a grid point of the climate model changes from tundra to ice. Using 20 different input ice sheet geometries, we capture the latitudinal grid spacing of 2.5° or 277.5 km for a span of maximum 4500 km (the longitudinal grid spacing at 60° S - the edge of the Antarctic continent in our simulations - is 55.6 km and decreases towards the pole). The names of

the different emulators have also been changed to EMULATOR_8, EMULATOR_12a, EMULATOR_12b and EMULATOR_20, indicating the number of ice sheet geometries used for the precursor climate model runs instead of the number of climate model runs performed (which is less relevant). The reason to include another emulator is that the number of ice sheet geometries has a quite large influence on whether the glacial inception will occur or not. While using a different number of ice sheet geometries, it is tested how many of these prescribed ice sheet geometries are needed to make the emulator work properly. Since the number of climate model runs is not crucial in the performance of the model, EMULATOR_70 with 8 input ice sheet geometries and based on 70 climate model runs is replaced by EMULATOR_8, also with 8 input ice sheet geometries and based on 100 different climate model runs.

2. Related to the above point, it is hard for me to understand why the coupled emulation based on EMULATOR_70 leads to such poor results. For example, why does it lead to ice volume change that is largely unresponsive to the CO₂ concentration change when ice volume is used? Similarly, to me it is hard to figure out the true reason for the poor results in Figure 9c for all three design schemes. If an emulator based on two parameters (ice volume and area in this case) leads to a worse result than an emulator based on only one parameter for the *same perturbed physics ensemble*, the only possible explanation is that the emulator with the two parameters failed to capture the behavior of the original simulator. I suspect the poor results stem from the poor emulation accuracy when the emulators are built on both ice volume and ice area. In fact, Table 1a shows that there are only **11** design points for the two ice parameters (ice area and ice volume) and, to make the problem worse, these two parameters are highly correlated. I think any emulation approaches are destined to fail with such a small number of design points that are highly correlated with each other.

Author's response: The answer is partially given in the previous response, but extended here. EMULATOR_70, calibrated on eight different initial ice sheet geometries appeared to be insufficient to induce the glacial transition. The climate simulated with a certain ice sheet geometry has a very sharp boundary at the edge of the ice sheet because of the large difference in albedo between ice and tundra. If you have insufficient ice sheet geometries the emulator will not simulate a climate that is cold enough in order to reach the climatic imprint as simulated by the next, larger ice sheet geometry.

'Figure 9a shows the ice sheet evolution for the four emulators calibrated on ice volume. EMULATOR_12a, EMULATOR_12b and EMULATOR_20 show the transition towards a continental scale ice sheet in a very narrow CO₂ interval of 845 to 875 ppmv. On the other hand, EMULATOR_8 does not seem to show any sensitivity to the CO₂ forcing during the 3 million year-long simulation (and also not on a longer timescale). The reason is that the prescribed ice sheets are separated too much in the initial climate model runs. Because of the large difference in albedo between ice and tundra, the prescribed ice sheets create a sharp boundary at the ice sheet margin that is visible in the temperature field. If insufficient prescribed ice sheet geometries are used, the ice sheet does not grow enough to reach the next input ice sheet geometry and the emulated temperatures remain too warm at the ice sheet margin. It appears that the threshold on the number of needed ice sheet geometries is somewhere between 8 and 12. Using 20 input ice sheet geometries does not lead to a significant improvement in the model performance.'

'Another option is to calibrate the emulator based on the ice area, which has a direct influence on the albedo. The glaciation threshold for EMULATOR_12a and EMULATOR_20 shows a very similar sensitivity to CO₂ of about 860 ppmv (Figure 9b). EMULATOR_8 grows immediately to a medium sized ice sheet and cannot grow further towards a continental scale ice sheet for the reasons already quoted. EMULATOR_12b was poorly defined on ice area (several ice sheet geometries had a similar area, but different geometry) and the ice sheet grows immediately towards a continental scale. Therefore, in addition to having sufficient ice sheet geometries, a good spacing of the ice sheet parameter is another requirement to use an emulator for coupled ice sheet-climate simulations.'

Table A1 shows **100** design points (each design point has a value for the orbital parameters, CO₂ and the ice sheet parameter) and contains **12** values for the ice sheet parameter. Indeed there is a strong correlation between the ice sheet parameters ice volume and ice area, but not for every ice sheet geometry as ice sheets can have similar volumes for different areas or vice versa. The reason why the emulator based on ice area and ice volume is performing worse compared to the other emulators, is because the tuned correlation length scales are much lower for the ice sheet parameters and therefore less decisive in the climate signal compared to the CO₂ forcing and the orbital forcing (because of the contradictory values with similar ice sheet volume, yet different ice sheet area).

It is true that the poor results stem from the poor emulation accuracy. But the poor emulation accuracy originates from the ambiguous signal of the ice sheet parameter: when ice volume increases with a certain magnitude, the ice area will respond non-linear. When the emulator is tuned based on both ice area and ice volume, there are 6 variables while only 5 parameters are actually different in the forcing of the climate. We think it is an interesting point to raise that more information on the ice sheets is not leading to a better performance of the emulator.

‘When the coupling is based on both the ice volume and the ice area, the simulated ice sheet volume mostly follows the insolation pattern (Figure 9c). Remarkably, the overall simulated ice volumes stay small. The emulator calibrated on both ice area and ice volume has a low correlation length scale for the ice sheet parameter in comparison to the orbital parameters and CO₂, suggesting that there is a weak relationship between temperature changes and ice sheet size. The poor emulation accuracy originates from calibrating the emulator based on 6 variables, while actually only 5 input forcing parameters are variable. The additional information on the ice sheet parameter is strongly correlated in most cases, but arises to give an ambiguous climate signal because the spread between ice volume and ice area is not equal. Thus, calibrating the emulator with additional information on the ice sheet parameter does not lead to a better performance of the coupled ice sheet-climate simulations.’

Additionally, we devoted a paragraph in the discussion about the problems that a double ice sheet parameter induces.

‘A common problem for the emulator and the matrix look-up table method, where the ice sheet parameter is defined by a single number, is that there is no control on the regions where ice starts to grow. There are no obvious solutions to this problem. Adding an additional regional ice sheet parameter, that would receive a certain value depending on whether the region is ice-covered or not, would give rise to an ice sheet parameter that is defined in multiple ways. It has been shown that calibrating the emulator on two parameters (ice area and ice volume) that are strongly correlated and not independent, induces problems with the emulation accuracy. The best solution to overcome the lack of spatial control on ice sheet growth is to use enough prescribed ice sheet geometries in the model design, in order to give limited degrees of freedom to the emulator.’

3. I am not sure why Section 4 is called ‘Bayesian’ sensitivity analysis because nothing in the section seems to be particularly ‘Bayesian’. There seems to be no consideration on uncertainty in the model parameters in the form of posterior densities, which is typically done in Bayesian analysis. In addition, the authors somehow decided to throw away the emulators and build a new time series model for uncertainty quantification. Is there any particular reason behind this decision? I am also wondering if there is any particular reason that only the first-order autocorrelation is considered here.

Author’s response: In section 4, the sensitivity of the ice sheet evolution is tested based on Bayesian statistics, i.e. the climate is predicted, taken into account the climate at the previous time step. In that sense, we believe that the use of Bayesian sensitivity analysis is legitimate. We do not investigate the influence of the climate evolution based on different model parameters, but the sensitivity to having prior knowledge about the climate that was simulated one coupling time step before. The reason to only include the first order auto-correlation is for simplicity and justified because the first order auto-correlation captures the behaviour well (when the climate was cold in the previous time-step, the chance is high that the climate was also cold in the time step before because the climatic changes occur gradually).

The emulator is akin to the emulators shown in section 3, with the emulation of precipitation and January temperature exactly the same way as in the other experiments. Since only the uncertainty of one of the emulated fields can be investigated at once, the parameter most decisive for ice sheet growth is chosen. It appears to be temperature, especially summer temperature and hence uncertainty to January temperature is emulated using a Bayesian approach.

‘The additional value of the use of an emulator for coupled ice sheet-climate simulations is that the mean climate predictions come with the estimate of its variance. In that sense, the Bayesian framework of the emulator can be used to update the covariance between two climate prediction points to predict the variance of the next point, knowing the variance of the current state. In this section, the Bayesian framework of the emulator is used to explore the uncertainty of the emulator by performing 50 Monte Carlo simulations including the variance on EMULATOR_12b calibrated with ice volume.’

‘The emulator is exactly similar to EMULATOR_12b, except for the inclusion of the variance for the temperature fields. The uncertainty is explored, taking into account the correlation between the temperature field at the previous time step and the current emulated temperature field. The Gaussian process emulator has an exponential decaying correlation function. Most of the covariance structure is captured by the first-order autocorrelation. The climatic changes occur gradually, a cold climate state will not be preceded by a very warm climate and therefore, the first-order autocorrelation captures the behaviour well.’

Minor Comments:

1. Lines 204-215: Related to the Major Comment 1 above, I think describing EMULATOR_100a and EMULATOR_100b as ‘bad’ emulator seems to be weird. I think this part can be improved by clarifying how EMULATOR_100a and EMULATOR_100b are actually designed; otherwise, readers may wonder why the authors decided to use ‘bad designs’. Later, in Section 3, they will be surprised by the fact that these ‘bad designs’ led to better results.

Author’s response: We tried to clarify what is investigated with the different emulators and what is a ‘good’ design and a ‘bad’ design. Clearly, bad designs do not lead to good results and oppositely, good designs (in terms of ice sheet spacing) can only lead to bad results when insufficient ice sheet geometries are included.

Lines 239-240 “Therefore, they might be doing a poor job in reconstructing the simulated temperatures well.”: I think ‘well’ should be deleted.

Author’s response: Done.

2. Line 259-260: “The notion of ice sheet parameter as an emulator input is introduced in previous studies to be an integer ranging from 1 to the number of ice sheet geometries”: The sentence does not make much sense. Please revise.

Author’s response: The sentence is revised to: *‘The ice sheet parameter is a single number representing the shape and area of the prescribed ice sheets. In previous studies (Araya-Melo et al., 2015; Lord et al., 2017), it is defined as an integer, representing the number of different ice sheet geometries.’*

3. I feel that the overall writing quality of Sections 3 and 4 are notably worse than that of the other Sections. Hopefully the authors can improve the texts in the revised version.

Author’s response: We acknowledge that most comments were related to the model set-up and the model performance. The text has been revised by stating why certain decisions have been made or what certain decisions have learned us. We have tried to improve the clarity of the manuscript by including an additional emulator with 20 prescribed ice sheet geometries, since the number of prescribed ice sheet geometries has a main influence on the emulator performance. We also rewrote the critical paragraphs from section 3, including the results with the additional EMULATOR_20 and gave arguments why the use of the Bayesian sensitivity analysis in section 4 was justified.