D. Royer (Referee)

van de Wal present here a novel way to calculate temperature and CO_2 for the last 20 Myrs and explore the implications of this association, especially with regards to climate sensitivity. This topic is of significant interest to many, and the approach taken by the authors is novel and compelling. Overall, the manuscript is in pretty good shape except for the discussion of climate sensitivity. I begin with some comments with the broadest significance, followed by a list of more detail-oriented comments.

We are glad that the reviewer recognizes the importance of the work and the approach taken. As you will see below we have rephrased and improved the discussion of the climate sensitivity.

Abstract: The abstract revolves around the second sentence ("The lack of transient climate models and in particular the lack of high-resolution proxy records of CO_2, beyond the ice-core record prohibit however a full understanding of the inception of the Northern Hemisphere glaciation, as well as the mid-Pleistocene transition"). The authors need to return to this key sentence at the end of the abstract. Do their data support the statement? Why or why not?

We are generating no new proxy data in the classical sense, but merely derive a continuous series for CO_2 from benthic oxygen isotopes, as such a step forward is made as no continuous transient data existed. We leave it to the reader to decide whether the NH glaciation and MPT can be understood better now, at least the prerequisite of a continuous CO_2 series is fulfilled and a CO_2 threshold for the NH inception is indicated explicitly and the magnitude of the change of CO_2 over the MPT is indicated (see abstract last two sentences).

As a whole, the abstract misses the mark. It doesn't focus on the areas that are emphasized most in the main body of the text (especially climate sensitivity).

The focus in the text is on the reconstruction of the CO_2 series as the paleoclimate sensitivity is very easily abused in the interpretation of ongoing recent climate change. (see comment Rapp). This method is not appropriate enough to claim that our insights in the climate sensitivity as formulated by Köhler et al. 2010 needs to be rephrased. The basic reason is that errors in the different components cancel each other as they are all multiplied with each other in equation 5. For this reason we have an important line in the abstract claiming that we find no evidence for a change in the climate sensitivity other than the expected decrease following from saturation of the absorption bands. We have explained this in more detail in the climate sensitivity section by adding an additional paragraph and prevented the focus on climate sensitivity by rephrasing the text.

Also, on line 7, I don't think it is fair to say that the authors produced a "continuous high-resolution CO2 record". It is a modeling effort, and in essence what the authors have done is "filled in the gaps" between existing proxy records assuming an averaged transformation between temperature and CO2 (Figure 5).

We have changed the somewhat ambiguous term record to series and tuned down the phrase as a whole. We fully agree it is a modeling effort, where the trick lies in the fact that we convert d18O into temperature so that we can derive a relation between CO2 and temperature, which is used to find a CO2 record, which is continuous and consistent with the benthic record. More proxy data in the future will show whether there are really strong deviations from the relation between CO2 and temperature as we find it. The number of data points and the accuracy of the proxies (particular for the warm part of the record) does not justify more detail in this stage.

p. 440: Please be clear here and throughout (including Figures 1, 4 & 5) what you mean by reconstructed temperature. Is this deep-water temperature or surface temperature? And if it's surface temperature, is it sea surface temperature or land+sea surface temperature? Is this temperature fully-integrated from equator to pole? »>Ahh, I see the answers to my questions on p. 444. This information needs to be stated back on p. 440!! And it should be given in abbreviated form in the captions for Figures 1, 4,& 5. Just saying "temperature" or "NH temperature" is not sufficient.

Thanks for pointing this out it is very important that the reader is aware of what we mean with temperature. We have now defined it in the introduction.

p. 445, lines 21-24: The two Pagani alkenone records should be combined: they come from the same author, following the same methodology, etc.

We have considered this but what counts for our purposes is the slope between ΔT_{NH}

and CO_2. For both records this is way out of what follows from ice cores and several of the other proxies, hence we have to reject the data from our further analysis. The data sets are treated separately because their mutual slope (one negative, one positive) varies widely. The data would also be rejected after first grouping them.

Some key stomatal-based estimates are missing. They are from: Kürschner, W.M., van der Burgh, J., Visscher, H., Dilcher, D.L., 1996. Oak leaves as biosensors of late Neogene and early Pleistocene paleoatmospheric CO2 concentrations. Marine Micropaleontology 27, 299-312. Once these are added, the "stomata" regression slope should be steeper.

Thanks for pointing out the missing stomatal data we have now included those which are reported in the recent commentary by Beerling and Royer, 2011. It is true that the effect is that the relation between ΔT_{NH} and CO_2 gets steeper. The gradient is still rather low but we decided to include them now because the number of data points for the warm period is rather low. Figure 3 and 4 are changed in accordance with the new data.

And finally, the Pearson and Palmer (2000) data can be rejected for reasons related to diagenesis, use of incorrect fractionation factors, and poor modeling of seawater alkalinity and _11B (Foster, G.L., Ni, Y., Haley, B., Elliott, T., 2006. Accurate and precise isotopic measurement of sub-nanogram sized samples of foraminiferal hosted boron by total evaporation NTIMS. Chemical Geology 230, 161-174. Klochko, K., Cody, G.D., Tossell, J.A., Dera, P., Kaufman, A.J., 2009. Re-evaluating boron speciation in biogenic calcite and aragonite using 11B MAS NMR. Geochimica et Cosmochimica Acta 73, 1890-1900. Klochko, K., Kaufman, A.J., Yao, W.S., Byrne, R.H., Tossell, J.A., 2006. Experimental measurement of boron isotope fractionation in seawater. Earth and Planetary Science Letters 248, 276-285. Lemarchand, D., Gaillardet, J., Lewin, É., Allègre, C.J., 2000. The influence of rivers on marine boron isotopes and implications for reconstructing past ocean pH. Nature 408, 951-954. Pagani, M., Lemarchand, D., Spivack, A., Gaillardet, J., 2005. A critical evaluation of the boron isotope-pH proxy: The accuracy of ancient ocean pH estimates. Geochimica et Cosmochimica Acta 69, 953-961. Royer, D.L., Berner, R.A., Beerling, D.J., 2001. Phanerozoic CO2 change: evaluating geochemical and paleobiological approaches. Earth-Science Reviews 54, 349-392).

We thank the reviewer for the adequate explanation of the failure of the Pearson data and have added the explanation in the text. As we use $\delta^{11}B$ from Hönisch we maintained for completeness the Pearson and Palmer data in figure 3 and 4 but again did not use them for the further analysis. So results are not affected by those data.

p. 450, lines 24-26: This doesn't make sense because the slope of the "ice" line in Figure 4 (red line) is steeper than the overall regression, implying a smaller change in CO2 for a given change in temperature.

We have rephrased the sentences to explain better that the stomata data and the other data do not support the idea that alfa is much smaller in warm climates.

Climate sensitivity section: This section is difficult to understand for several reasons. First, the authors present on p. 449 that the climate sensitivity calculated from presentday and late Pleistocene observations (i.e., "Charney" sensitivity; equations 4-5) is similar to the sensitivity calculated from paleo-CO2 and modeled temperature (Figures 4-5). Thus, it is difficult to figure out where the authors' conclusion for higher sensitivity in the ancient record comes from (see also previous comment).

We have caused confusion with the sentence that the paleo sensitivity is higher. We only intended to say that as the sensitivity (dT/dCO2) depends on the CO2 concentration itself one has to realize that the sensitivity is different for LGM conditions. We revised the last part of section 5 considerably. The Charney sensitivity is of course constant.

Second, the calculated climate sensitivity for doubled CO2 is _30K, or 12K for global surface temperature, for both the first principles calculation (equations 4-5) and the empirical calculation (Figure 5). This makes no sense because Charney sensitivity is typically around 3K.

Adopting a value of alfa=2.5 and CO2 is 576 yields indeed a large temperature change of 10.7 ˚C. It is the Charney sensitivity not the overall sensitivity which is much more restricted. The large value in the temperature change is obviously related to the large values of f included in the calculation. This implies to us that the feedback factor f is likely not constant. Likely f decreases for warmer temperatures, but this might coincide with a decrease in one of the other components, such that C keeps more or less constant. However we do not use f or C as such, we derive C from the temperature calculations and the CO2 data directly. Those data are binned in figure 5 to prevent too much bias to the cold period with much more data points. From the black dots at the right hand side of the figure it is clear that we have no reason to assume that the relation is non-linear and from that we derive C to be 39. Note furthermore that the value of f as derived by Köhler et al. 2010 is based on paleo data and is not intended to be used for 2*CO2. The text has been improved on this point.

These problems require addressing. Further, the manuscript would be much clearer and compelling if the section on how paleo-sensitivity compares to Charney sensitivity was expanded. This is a topic of great interest to many, especially if the data are also presented in terms of doubled CO2 and global surface temperature. Along these lines, the papers of Hansen, Lunt, and Pagani should be discussed. They find climate sensitivities of 4-6+K per doubled CO2 for the late Cenozoic glaciation. (Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., Pagani, M., Raymo, M., Royer, D.L., Zachos, J.C., 2008. Target atmospheric CO2: where should humanity aim? Open Atmospheric Science Journal 2, 217-231. Lunt, D.J., Haywood, A.M., Schmidt, G.A., Salzmann, U., Valdes, P.J., Dowsett, H.J., 2010. Earth system sensitivity inferred from Pliocene modelling and data. Nature Geoscience 3, 60-64. Pagani, M., Liu, Z., LaRiviere, J., Ravelo, A.C., 2010. High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. Nature Geoscience 3, 27-30.)

We have added a paragraph exploring the climate sensitivity more. As we follow the analysis made by Köhler et al. 2010. It is merely a comparison of Köhler et al. 2010 and Hansen et al. 2008, although our independent temperature estimates in fact support the feedback factor found by Köhler rather than the lower value deduced by Hansen et al. implying that excluding the feedback factor has a larger effect on the short-term climate sensitivity for our reconstruction than presented by Hansen et al. 2008.

Third, the authors should explore the potential for variable climate sensitivity within their paleo-record. For example, if the CO2 proxies are correct, the authors over predict CO2 from _7-10 Myrs, and perhaps from 2.5-10 Myrs too if the “Alk.+_11B” estimates are incorrect. This means that climate sensitivity during this period was higher than the mean sensitivity calculated by the authors. This pattern may make physical sense because it would mean that sensitivity dropped going back into the mid-Miocene climatic optimum, when glaciers were at their minimum extent (and thus the ice-albedo feedback weakest).

It is an appealing thought to have a variable climate sensitivity, we already given that some thoughts, but rejected the idea to include it. Hargreaves et al. 2007 suggested a 15% smaller sensitivity between LGM and 2*CO2. We assign an error of 10% to our sensitivity value in Fig. 5. In combination with the fact that the number of data points with high CO2 concentrations (say >400 ppm), from proxies, which are qualified (so excluding, alkenones, d11Bp) is small, this is not feasible. The limited number of points for high concentration is in fact the reason for binning the data in Figure 5. More reliable proxy data for warmer climates are a prerequisite to study any change in the sensitivity. Our temperature records offer the framework to do it, but more CO2 proxy data are needed.

Minor comments

p. 438, lines 16-17: "We find no evidence for a change in climate sensitivity other than the expected decrease following from saturation of the absorption bands for CO2." Climate sensitivity accommodates for the saturation effect (i.e., it is cast in log space), so the second half of the sentence is misleading and should be cut. The first half should be revised too, given that the authors conclude that climate sensitivity was higher than the present-day during their paleo-interval (!)

rephrased. We imply to say that C remains constant, but it is not clear to the reader what C is, which is not the Charney sensitivity so we circumvented it now by using not the word sensitivity any longer.

p. 438, line 20: "Finally it might be noted that we observe" This is awkward; change to something like "Finally, we note".

Done

p. 438, lines 21-22: Why should the reader care about only minor CO2 changes during the mid-Pleistocene transition? A follow-up sentence is needed to put these data in proper context.

Done

p. 438, line 23: This statement is somewhat misleading because it implies that the climate has steadily cooled. This of course is not true, for example during the mid-Miocene climatic optimum.

Done

p. 439, line 19: “we forced by then: : :” Not sure what this means.

Changed

p. 440, line 4: “It” should read “it”

ok

p. 441, lines 22-24: Why are the authors restricting themselves to calculating Northern Hemisphere temperature if Southern Hemisphere ice sheet data are now available? This choice needs to be defended.

We do not use an explicit climate model where Southern Hemisphere temperature is independent to at least some extent from the Northern Hemisphere temperature. Here Southern Hemisphere temperature is parameterized as a function of the Northern Hemisphere temperature (details. De Boer et al. 2010). We clarified this point in the text.

p. 442, line 16: Instead of saying “Miocene”, give the numerical age you are using for your point of comparison. During the mid-Miocene climatic optimum, for example, your modeled temperature change is close to 30K, much larger than the 12K figure you cite here! “Miocene” is too ambiguous.

It is the temperature difference between the average over 14.5 and 15.5 Myr compared to the average over the last Myr.

p. 443, line 20: “gradually” should read “gradual”

ok

p. 444, line 16: “Intriguing is the question: : :” Bad language.

Ok

p. 445, line 15: “there is a relation between CO2 and temperature”. This point should be made more nuanced, as the authors are comparing CO2 to temperatures adjacent to northern hemisphere ice sheets.

ok

p. 447, lines 3-4: Or a problem with your model: : :

We don’t believe this can be a problem of the model. The modest benthic values seem to be inconsistent with the low CO2 values. It is way out the uncertainties of the model. We maintain this finding. We are not in the position to note whether it is due to the benthic record or the B/Ca.

p. 449, line 7: “ïA¸nˇ ”should read “ïA¸gˇ ”. Also, the authors say here f = 0.72 but on the previous page say 0.71.

correct 0.72 is a typo, and labda should be gamma

p. 450, lines 12-17: This section is not highly related to the first part of the same paragraph. It would be much clearer if this section was its own paragraph.
p. 450, lines 14-16: This sentence is too opaque. The associated CO2 change is needed. Ditto for the present-day scenario.

A more elaborate paragraph is replacing this part

p. 450, line 25: remove comma and “which are”

ok

p. 451, line 21: remove comma

ok

p. 452, line 1: The authors should clarify here that the mean changing sensitivity relative to the present-day, not changing sensitivity within their paleo-time series.

ok

Figure 3: It's very difficult to figure out what's going on in this plot. I see several problems. First, the vertical scale is so compressed that it is difficult to see patterns in CO2. An obvious solution is to combine all subplots into a single unified plot. Second, why are there horizontal lines at 300 ppm? A more intuitive choice is to have the horizontal lines dividing each subplot. Third, it is difficult to figure out the magnitude of the minor tick marks. Yes, it is stated in the figure caption, and yes, a few of the subplots have "400" written on the right-hand side (why only a few? Why not all?), but if I am any measure of a typical reader it will take a few minutes to figure this out. This delay is not desirable at all. Again, a solution is to combine all subplots. Fourth, why are multiple ice-core studies combined into one subplot but not for the other records. For example, the authors split the two Pagani studies into two separate subplots. Why? At first pass, the figure gives the impression that alkenone records show a marked decline at 5 Myrs. Not good!! Finally, what is the star marked "100"? This is not described anywhere in the figure or figure caption.

Ok changed. The idea behind the plot was that it should be easier to grasp the number of data points for each proxy as well as the time period they cover. This is at the expense of the details of the values. We now turned this around to plot simply all proxies in one plot.

Figure 5: The red dots are virtually impossible to see. Enlarge them.

ok

Anonymous Referee #2

This manuscript aims to estimate atmospheric CO2 of the past 20 million years from benthic foraminiferal oxygen isotopes, where the ice volume component of d18O (i.e. sealevel) was used to force the ice-sheet volume, and temperature was subsequently estimated to match independent sealevel-observations.

We would phrase it differently; The ice volume component and temperature are solved at the same time such that the benthic oxygen record is followed.

This is an interesting study that should get published, however, much more information is needed for the reader to evaluate the quality of the reconstructions. For instance, on page 439 line 27 ff the favorable comparison with independent sealevel and temperature records is mentioned, however, it is not mentioned that the sealevel records are restricted to the Pleistocene only,

We have added that the sea level records have been compared with the Miller and Muller records over the last 20 Myrs as done by de Boer et al. 2010, fig 9.

whereas Lear et al. (2000) also estimated both sealevel and bottom water temperature from benthic foraminiferal d18O and Mg/Ca for the entire Cenozoic. I would like to see a comparison of this new reconstruction with published estimates, a discussion why sealevel was not compared with the longer Lear et al. (2000) record, and consideration of the carbonate ion effect on benthic Mg/Ca (e.g. Yu & Elderfield 2008, EPSL, Sosdian & Rosenthal 2009). How do the estimates compare to Lear et al. 2010 (Paleoceanography), during the Miocene? A
demonstration of the "favorable comparison" is necessary to evaluate this new reconstruction.

We have added the comparison with the work by Lear for the longer time perspective and noted that Bintanja et al 2008 compared their data with the record by Lawrence et al. Evolution of the eastern tropical Pacific through Plio-Pleistocene glaciation. Science, **312**, 79-83 (2006).

We have improved the manuscript by a note on the comparison with the work by Lear et al. 2000 and 2010.

The sea level record as published by Sosdian and Rosenthal over the MPT seems not very reliable see de Boer et al. 2011, Figure 6. We do not want to stress that again in the present manuscript.

In addition, how is deep-sea temperature compared to surface temperature? Please explain briefly the procedure outlined in Bintanja et al. 2005b. How has this parameterization been validated?

The method section has been extended here with more details extracted from our previous work to improve the readability and make this paper more stand alone.

Page 442, line 2ff: What is the effect of a lack of the bipolar seesaw on the climate system and how could the result of this study be changed if the bipolar seesaw and Dansgaard/Oeschger events were included?

The smoothed record of benthic d18O does not allow for this detailed temporal studies. One might envisage that an ocean model (not included here) including the bipolar see-saw might result in differences in the results of temperature and ice volume changes between Northern and Southern Hemisphere. We don't foresee that the general overall picture changes. We extended the method section and changed the title of section 3 in order to clarify this.

Page 444, line 3/4: is that air or surface water temperature?

Air, clarified

Page 445, line 14ff: Temperature is used to select CO2 records that are consistent with a temperature/CO2 relationship comparable to ice cores. This procedure seems circular and dangerous, as it assumes that the CO2/T relationship in the past was comparable to the Pleistocene.

It is true that assuming the temperature/CO2 relation to be comparable is dangerous (it is not circular). This is however supported by the data from the B/Ca, $\delta^{11}B_h$, Alk+$\delta^{11}B_s$. Those data are not affected by the $\delta^{18}O$ and our model assumptions and hence can be considered as independent justifying the temperature CO2 relation. From a physical point of view it is difficult to imagine that the relation between temperature scales inversely as one of the Pagani records shows. This point has been clarified in the text.

In particular the B/Ca reconstruction by Tripati et al. 2009 falls in this same trap, as B/Ca varies little over their study period, and variations in their CO2 estimate are largely driven by secondary corrections such as temperature variations.

We do not agree with the argument that the B/Ca ratio varies little over their study period. Figure 4 clearly shows that this is not the case. We are also aware that the method is criticized, though we cannot find this in the literature and based on our own analysis we have no reason at all to reject the data. Hence we have to include them. Having said that it has been tested that excluding those data does not significantly change our results.

Page 445, line 21: it is argued that Pearson and Palmer used multiple species for their reconstruction, where they actually used predominantly a single species for the past 20 million years. In order to remove the species argument, only Pearson & Palmer's single species data could be compared to independent CO2 estimates. However, it should also be kept in mind that Pearson & Palmer used smaller size classes in their earlier samples, which may reflect a deeper growth habitat at lower pCO2, and that they applied a correction for the isotopic composition of seawater, and modeled alkalinity, all of which are debatable.

We rephrased the argument why the Pearson and Palmer data can not be used for our purpose following the argument presented by D. Royer.

I find the overall treatment of proxy data in this comparison rather questionable, as proxies that extend the CO2/T relationship observed in ice cores are deemed more reliable than others. While the authors are modelers and may not fully understand the pitfalls of each proxy reconstruction, I find the conclusion that "the various CO2 proxies can be understood in the broader framework of long-term climate change" (page 451, line 22/23) rather bold and haphazard. Although this comparison may identify proxies that deviate from the average, it does not help to identify whether one proxy estimate is better than the other, or whether a proxy may find a reasonable answer for the wrong reasons. Such conclusions should be left to decide by the proxy community, not by consistency with a modeling estimate.

We have tuned down the statement in line with the reviewers suggestion. It is not the consistency with a model estimate which counts but consistency between paleo co2 data and the marine benthic record.

Page 446, line 7/8: Please provide a figure or further evidence of how the different CO2 proxies individually affect the modeled temperature and CO2 estimates.

The effect of omitting one proxy effects the key parameter at most by 6%. This is achieved if the B/Ca ratio data are omitted, this is within the error bar of 10% which is assigned to the uncertainty in C. We have added this information to the text

Page 446, Line 20-25: It should be mentioned that Hönisch et al. 2009 specifically selected glacial/interglacial extremes for their reconstruction and found stable interglacial pH and pCO2 values before 1 Ma, followed by a decrease in both G/I extreme pCO2 between 0.8-0.6 Ma (comparable to ice cores). Although the average pCO2 across the MPT appeared to decrease (which would be comparable to this modeling study), the overall similarity of interglacial pCO2 was taken as an indication that Carbon was not generally removed from the active carbon reservoirs but only temporarily stored e.g. in the deep ocean. To better evaluate the quality of each proxy, it might be useful to consider the sampling strategy of different proxy reconstructions (random or G/I extremes) in comparison to their CO2 estimates.

This is an interesting idea, but we have not enough information of the data to do so. We have added the info on the Hönisch data

Page 446/Line 25/26: Please make sure that the reader does not get the impression the "combined d11B and alkenone record" is a new technique. Seki et al. 2010 studied both proxies independently and compared their results. Seki et al. 2010 argued that their sampling strategy may have favored interglacials, which may explain the high CO2 values estimated for the past 1.5 Ma.

We have rephrased this, see also reviewer 1.

Page 447, line 4: please specify that B/Ca was measured on planktic foraminifers. It seems that the authors hint at something but do not complete the thought.

Done

Page 447, line 11: please specify which fast and slow feedbacks have been considered.

We indicated that they are defined somewhat later in the text.

Page 447, Line 21: "functional relationship between DeltaT and CO2" sounds like one driving the other. Please rephrase, possibly using "quantification of the covariation between Delta T and CO2"

The sentence before explains that CO2 is not always the driver

Page 449, Line 1-3: How does the estimated NH temperature change compare to terrestrial proxy estimates?

We simply use the data by de Boer et al. 2011. We do not attempt to verify those data in more detail than already done in the earlier papers as explained much more extensively in the manuscript by now.

Page 449, equation 5: Please explain the parameters used for the calculation of climate sensitivity.

They are all explained on page 447-449. Unfortunately there was one typo alfa in the equation should be gamma this is corrected.

Page 449, Line 19: please add references for the potential change in meridional temperature gradient

Ok

Page 449, Line 22-25: How much higher should/could CO_2 have been? How do d11B and alkenones compare to that expectation?

From equation 5 it follows that C scales linearly with alfa. A reduction of C by 50% would imply for mid Miocene conditions a CO_2 value of ±730 ppm. That is way out what proxies suggest. So a very strong decrease in alfa does not seem to be likely.

Page 451, line 16: What is meant by "the trend in CO_2 before the inception is strong"? Please rephrase.

Rephrased

Page 452: Please rephrase this paragraph. This reads as if proxy data cannot be trusted, when what I believe the authors would like to say is that absolute proxy data have to be considered in the framework of specific Earth system parameters at that time.

Rephrased

Figure 1: What determines "zero" in panel b?

Zero is the pre-industrial temperature. The caption is extended

Figure 2b: The individual contributions to sealevel seem to match the integrated 3D estimates at sealevel >0m but a large part of the variation is missing at sealevel <0m. What makes up this difference? Please specify that sealevel validation with independent estimates was only done for the Pleistocene. A) "thick lines in lower panel" is confusing, please rephrase.

Caption adjusted

Figure 3: Seki et al. 2010 did not "combine" d11B with alkenone estimates, they used both techniques and compared the results.

Rephrased

Figure 6: This figure should be plotted larger, the turquoise symbols can barely be seen. The horizontal bar on the middle panel may be better replaced by a vertical bar laid behind the data.

Caption b) should read "NH glacial inception"

Adjusted

In summary, this is an important contribution that may likely attract many readers. Although several explanatory studies are referred to, I think it would be useful if this manuscript could be easier understood without having to read multiple secondary papers describing the modeling methods and constraints applied.

We have added more of the key results of the earlier work to improve the readability and extended the methodology so that the paper can be read without digging through all our previous papers where we used and developed this methodology.

The introduction and treatment of proxy data should be somewhat revised and conclusions about their reliability toned down.

We have changed the treatment of proxies by including more Stomata data (see rev. D. Royer). We have changed the explanation of the failure of Pearson and Palmer (see rev. D. Royer) and explained the last paragraph better.

Anonymous Referee #3

This manuscript attempts to model atmospheric CO_2 concentrations from the Zachos et al. (2001) benthic _18O curve, a series of 1-D ice sheet models and four parameters relating bottom water temperature, atmospheric temperatures and CO_2. Reduced complexity models, such as this one, have a role in trying to understand processes and reduce highly complex systems down to their fundamentals. For all such models, it is important they are formulated in such a way to retain enough information or predictive ability to be able to accurately reproduce all the essentials of the system. To show this they need to reproduce known independent data, i.e. data not used in the calibration of the model. While this work is undoubtedly interesting and I'm sure that there is much to be learnt from these simulations, the efficacy of the model formulation and assumptions to accurately reproduce climate and CO_2 over the last 20 million years is questionable, particularly with the lack of testing against relevant independent data.

We have extended the modeling approach and the testing of it with the independent data, which was published in Bintanja et al. 2005(a,b), Bintanja and van de Wal 2008 and de Boer et al. 2010 and 2011. By extending this part we hope to achieve that there is less need for a reader to go through already published work based on this methodology.

Probably the most problematic feature of the model is the amount of information that is required for the reconstruction of 20 million years of climate history and atmospheric CO_2. Essentially the only information driving the model is the benthic _18O curve. For this to be a sufficient proxy for the whole climate system and atmospheric CO_2 levels requires a number of assumptions, including that the partitioning of heat in the climate system is fixed, that all temperature change over the last 20 million years is coupled to atmospheric CO_2, that the ice sheet only responds to surface air temperature, that all the other factors affecting _18O (ocean circulation, ocean gateways, salinity, isotope fractionation etc.) are unchanged and that any climate change is globally coherent.

We do not fully agree with reviewer here. Input for the model is the benthic curve. This is generally accepted to be determined by global temperature changes and ice volume changes. Hence the model can produce information about global temperature. We do by no means claim more detailed climate change. We furthermore only assume a priori that there is a relation between CO2 and T. Results show that we have no evidence that grosso modo this relation changes over time (Figure 3). Although we realize that we need to be more careful with the warmer part of the record due to limited proxy CO2 data. The fact that ocean gateways are important is not denied and is in fact the reason that we not bluntly extend the method back to the Eocene. At the end of the discussion we have one paragraph addressing the possible effects of other geological processes on the temperature CO2 relation

Each of these assumptions can be relatively easily disproved, for example see Dowsett et al. (2010), Lunt et al. (2009), Schoof (2007), Spero et al. (1997) and Raymo et al. (2006) respectively. The problems of such assumptions are best illustrated by the middle Miocene (14 million years ago). Although this was a period of particular warmth and significant ice sheet retreat, no-one proposes that the Antarctic ice sheet returned to a pre-glacial state as in the Eocene. However, middle Miocene _18O values were similar to those found in the Eocene and hence this model shows an almost total collapse of the Antarctic ice sheet and a return to a pre-glacial climate. Testing the results of this model is a particular problem, as the use of Pleistocene glacial-interglacial contrasts and the Eocene-Oligocene transition to calibrate the model leaves only the much less well-known Oligocene and Neogene for model evaluation. However, the predictions of the model for the Pliocene and Miocene do not seem to match existing knowledge of the climate and ice sheet in these periods, e.g. Dowsett et al. (2010), Denton et al. (1984) and Talarico and Sandroni (2009).

We thank the reviewer for pointing out this point. As in the Zachos et al. 2008 curve the d18O values around the Mid-Miocene are comparable to the level at the E-O transition our model will lead to a strong reduction of ice volume during the Mid-Miocene. We therefore tested our approach by using the newer d18O compilation by Cramer et al. 2009. That data set discusses in more detail interbasin changes in d18O and gets to a different stacked record. As a consequence our model yields more ice during the Mid-Miocene. This is reflected in a new paragraph in the discussion section.

Furthermore some of the implications of the equations derived for this model require the authors to be absolutely sure of the model efficacy. From their equations 4 and 5 it is easy to see that the sensitivity of palaeoclimate temperatures to a doubling of CO_2 (_ Earth System Sensitivity) is 27_C for the Northern Hemisphere and (from the authors reply to D. Rapp we see) this corresponds to 11_C for the globe. This is almost double the next highest estimate of Earth System Sensitivity from data and models (Hansen et al., 2008; Pagani et al., 2009; Lunt et al., 2009).

Any paper that is going to suggest alternative Cenozoic climate history and suggest much larger values than generally held for the sensitivity of the Earth System to increases in atmospheric CO_2, definitely needs to be well evaluated and tested and currently this paper does not achieve this.

There is a lot of misunderstanding on sensitivity, which we tried to prevent in the revised version by comparing our results to Hansen et al. 2008, which is more widely known than the paper by Köhler et al. 2010.
First of all our reconstructed CO2 record does not depend on equation 4 and 5, it follows directly from the relation between proxy data and reconstructed temperature. Eq. 4 and 5 are introduced to show that the value we find for C is in agreement with paleo evidence. In fact we follow the work by Kohler et al. 2010, which provides a careful analysis of the climate sensitivity. From that work it follows that the factor f is considerably stronger than in the estimate by Hansen et al. 2008. As a result our short term climate sensitivity (setting f to zero) is smaller than for Hansen et al. 2008. Combining this with the fact that observation of global temperature are in agreement with our short term climate sensitivity (see figure) suggest that our results are not too bad. The problems are f (and alfa) for paleoclimate, not on beta, gamma or Sc. (gamma might be an issue for short specific intervals).

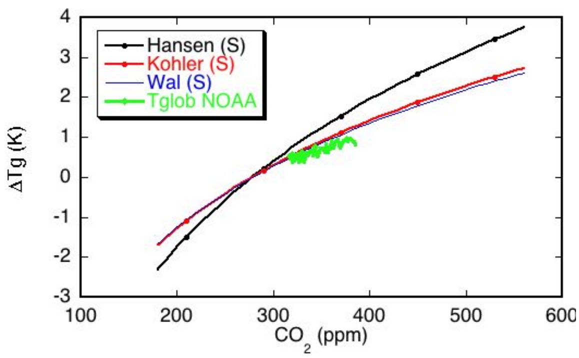

Figure: plot for the temperature change excluding the long-term feedback factor.

The model presented here seems too simple to reproduce CO_2 and climate over the last 20 million years. Either this paper needs to be reframed as an interesting sensitivity study or the authors need to provide much more justification for their results.

Of course the paper is no more than an interesting sensitivity study, which tries to bridge the paleo evidence from the benthic record on temperature and the paleo records of CO2. Eventually a more elaborate climate model including 3D atmosphere, ocean, ice at high resolution over 20 Myrs is needed in combination with a wealth of reliable proxy CO2 records. For the time being we have to work with a more modest toolbox, as there are still limits to computer power.

As an absolute minimum the authors would have to do the following:
_ justify the use of Pleistocene relations between bottom-water temperature, Northern Hemisphere temperature, Southern Hemisphere temperature and CO_2 for periods with hugely different boundary conditions (e.g. no large-scale Northern Hemisphere glaciation, open seaways in Panama and Tethys etc.).

We have extended the manuscript with the validation with other proxies of temperature and sea level as carried out in earlier papers to make this paper more readable.

_ assess from a purely data perspective the assumption that these relations are constant over geological time.

Assessing the relation between CO2 and temperature requires reliable CO2 records (we used most of the available records) and requires a temperature proxy. We believe that the best continuous temperature proxy can be derived from the benthic record if one can correct for ice volume changes, which is precisely what we do. Section 5 is merely a justification of the C value we find, it is not a prerequisite for the CO2 reconstruction.

_ assess the model predictions of temperature, sea level and CO_2 against the available independent data, i.e. not Pleistocene or Eocene-Oligocene transition.

More information is added in the manuscript.

_ explain why this analysis of ice core CO_2 records produces such a different relationship to that found by Hansen et al. (2008), who showed a longer term or Earth System sensitivity of 6_C (not 11_C).

We addressed this in more detail in order to prevent confusion

_ much greater analysis of the relationship between temperature and CO_2, including the following:

_ improved analysis of ice core CO_2 record, including showing scatter and uncertainty in temperature relationship.

We have attempted to provide error bars to both CO2 records and benthic records as far as possible. We now stress in more detail that the largest scatter is caused by the limited data for warm climates.

_ justification of linear relationships for proxy CO_2 records when some seem to show a break in gradients, little relationship etc.

We reject some records because of this reason, the other records show a linear relation as shown in Figure 3, albeit with large scatter, which by the transient nature of the climate system may be expected

.

_ why records that show very different development over the last 5 million years seem to have the same CO_2-temperature relationships.

It is not clear to us what the reviewer mean here.

_ how ice core record can rule out stomatal records, when there is no overlap either in time or CO_2 concentrations.

We have revised this point after we included more stomata data. The slope between temperature and stomata increases and we now have no reason any longer to reject the stomata data from the analysis. Including them does not really affect the outcome as expected. The scatter of the stomata data is large, but that is no reason to reject them

_ discussion of the implications of the formulation of the model and the retrodictions of past climate and CO_2.

We try to discuss this in section 5 and 6, which are extended and clarified

(1) The authors never defined exactly what they mean by "Northern Hemisphere". They use this term six times in the paper but never define it. In my view, the "Northern Hemisphere" is the entire region of the Earth north of the equator including the tropics north of the equator. However, I have a suspicion that when the authors use the term "Northern Hemisphere" they might mean the area north of 60°N. This needs to be clarified.

Just above the start of paragraph 4 we define deltaTNH. "In order to interpret the results one has to bear in mind that the reconstructed temperatures are strictly only valid in the continental areas where ice sheets develop in the NH (deltaTNH) being mid to subpolar latitudes (Bintanja et al. 2005) implying that they are therefore not necessarily representative for the entire globe (deltaTg)." So we defined delta TNH already in the manuscript and explained this in our earlier reply. It is land temperature not ocean and it is at mid to subpolar latitude. To improve the manuscript this is defined now earlier in the manuscript to prevent confusion

(2) In their response to my comment, the authors claim that ?T(NH) = 15.1°C corresponds to ?T(global) of 6.1°C, and these temperatures correspond to a CO2 concentration of 390 ppm. Since we have already reached 390 ppm of CO2 and temperature nowhere near the values calculated by the authors, clearly the authors are dealing in fantasy.

As explained in our first reply the difference is whether you include or exclude slow feedback mechanisms as explained in the text and expressed in equation 5

(3) Furthermore, the ratio ?T(NH)/?T(global) = 2.5 seems incredibly high, unless the authors mean by "Northern Hemisphere" only the area north of 60°N.

see point 1 yes it is land temperature at high latitudes

(4) The authors do not refer to a relevant paper: Hansen, J., R. Ruedy, M. Sato and K. Lo (2010) "Global surface temperature change"
http://data.giss.nasa.gov/gistemp/paper/gistemp2010_draft0601.pdf.
In this paper, the authors compare conditions at the last glacial maximum (LGM) with those prevailing just prior to industrialization of the Earth when CO2 was ~ 280 ppm. In the following figure, I have arbitrarily set delta-T = 0 at 280 ppm and plotted that point. According to Hansen and Sato, the global average temperature was 6.8°C colder when CO2 was 180 ppm at the LGM, and I plotted that point. Hansen and Sato estimate that if and when CO2 goes to 560 ppm, delta-T will be 3°C, so I also plotted that point. Then I joined the 3 points with a curve. For high NH latitudes, the changes will be significantly greater but the shape of the curve is likely to be similar (I suspect). Although this is a plot of delta-T vs. CO2, other changes occur on the earth as delta-T and CO2 change, and the total delta-T change is not entirely due to changes in CO2.

The basis for these estimates by Hansen and Sato are the forcings they estimated:

It seems evident that the temperature changes estimated by van de Val are much too extreme.

We will add a paragraph addressing the work by Hansen et al. and explaining that our work does not come to different conclusion with respect to the climate sensitivity as we explained in our first rebuttal, the difference lays in which feedbacks are included or excluded. See also our reply to reviewer 3.

(5) The relationship between ?T and ?CO2 remains one of the most important unknowns in climatology. Many papers have been written on the subject but there remains considerable doubt. Adding more noise to the weak signal that we presently have is not constructive.

This is the first attempt to our knowledge to reconstruct a self-consistent and continuous picture of delta18O marine benthic, sea level, temperature and CO2 over the last 20 Myrs. It will not be the last. What needs to be emphasized in more detail in the paper is that we do not come to substantial different climate sensitivities as previous studies, like the work by Hansen. See also point 4 and the first rebuttal.