Shruti Nath
Climate Analytics
Ritterstrasse 3, 10969, Berlin, Germany
Phone: +49 (0) 30 259 22 95 44
E-mail: shruti.nath@climateanalytics.org

Geoscientific Model Development editorial board
Berlin, 14th May 2023

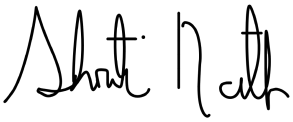
2nd Revisions: TIMBER v0.1: conceptual framework for emulating temperature responses to tree cover change

Dear Dr. Müller,

We are pleased to submit the second iteration of reviews for the paper on TIMBER v0.1. We have addressed the minor revisions as specified by Reviewer 2. Mainly, we clarify some remaining methodological questions, and provide more comprehensive discussion on the implications certain design choices have on TIMBER's modes of applications. A point by point response is furthermore provided in this document.

We are confident that these changes strengthen the paper and improve its depth. We hope these revisions address all remaining concerns but also welcome any additional suggestions.

Yours Sincerely,



Shruti Nath
(on behalf of all co-authors)

1) It isn't clear what temperature change estimates are being used in the results sections. It is possible that each section could be using a different set of estimates. The calibration results could be using estimates from different sets of lambda/function count, based on leaving out different blocks. The evaluation results could be using estimates from different sets of model function weights, but with the same lambda/function count, based on leaving out different blocks (and the blocks could be climate or tree frac). Section 4.2 appears to use a different setup that isn't discussed anywhere. The 2m temp results could use a single setup - the same lambda/function count and function weights - but which set of function weights?

We have provided further clarification on these as suggested in the specific suggestions and comments.

2) I am thoroughly disappointed that the scenario results are not compared with related model outputs of temperature response to lulcc. This discussion is essential to this effort. The authors do spend time explaining why these results should not be compared to such outputs. But then what is the point of TIMBER? As an emulator it is only useful if it can reliably emulate what it is trying to emulate. Forest cover changes are likely the largest factor in local temperature change, and so even at this stage TIMBER should be showing patterns similar to the corresponding model estimates for these scenarios, at least for the min/max/mean temperature. If not, it at least gives a platform for sections 5.1 and 5.2.

We recognize this problem and have further added some discussion on it in Section 5.2.

3) The Hooker parameterization for 2m temp is clearly inadequate. It would be nice a comparison with this equation parameterized for one of the models, as there may not be adequate data to parameterize it on observed tmax and tmin values. As this is the final step in the process and is thoroughly examined, the paper is probably fine without re-parameterizing the hooker equation.

This is a good point and we have added discussion on it in Sections 5.1 and 5.2

Specific suggestions and comments

Abstract

1. Introduction

2. data

3. methods

line 212:
what is training here? I assume it is determining the model function weights using the same lambda and function set.

We train on the full training dataset using the lambda and number of basis functions optimized for in the block cross-validation. We now clarify this and furthermore explicitly define to the fully trained Gamma_m as fully calibrated

line 257:
95% of the data is not an interquartile range. The inter-quartile range is the 50% of data between q1 and q3.

This was a typo and we now simply refer to it as the interquartile range

lines 270-305:
what is the final model you are using? presumably the best lamba and function set chosen during calibration. But which set of tuned model function weights? You do this tuning for several sets of data using two different block configurations. Do you do a final tuning with the entire global data set and use that?

When predicting into new scenarios we always use the final model as now specified in L212.

4. results

lines 319-347 (and figure 3):
How do you obtain RMSE per grid cell? The calibration procedure calculates an RMSE across cells within a block. If you are using the optimal GAM and one change in tree fraction per cell, how do you have multiple temperature change estimates per cell to calculate an RMSE? Is this based on the parametric uncertainty described in section 3.4.1 to get multiple temperature estimates? Or is this based on the different tunings with blocks left out? Or is this based on the different model fits for lambda and number of functions?

This is obtained as the RMSE from the sampled responses (i.e. that represent the parametric uncertainty). We have now clarified this in the text.

lines 348-366:
What do the RMSEs represent here?
They could be from each cell, irrespective of which block the cells are in, in which case the same question above applies.
They could be the RMSEs of each block present in the latitude band, but blocks can span latitude boundaries, and are there enough blocks to get these stats?

Similar to previous point, we use the samples drawn and this is clarified in the text.

lines 368-385:

What do the RMSEs represent here?
Now there are stats for each block within each latitude band, which suggests that the values are from each cell, in which case the same question above applies.

Similar to previous point, we use the samples drawn and this is clarified in the text.

line 376-377:
You may want to clarify that this relationship is for deforestation in the tropics only, as it currently sounds like a more general statement that forest change data is more important for skill than climate analog data.

This is a good point, and we have now clarified it.

lines 388—392:
This sounds like an interval based on the distribution across space of single-cell estimates. But then what is the mean of? Cells with the same fractional change? Is this the case? These last two lines may help with understanding, if made more clear sooner. It sounds like here you apply the temp change model to the whole globe many times, each with a different tree frac change that is the same across all grid cells. But with the single best calibration set and one set of tuned model weights?
Technically, you should explain this process in the methods, but since it could be a short explanation you may be able to include it here.

We realise that it may indeed be confusing, so for clarification we have added a short explanation:

“A select tree cover change value is applied to all grid points, and Gamma_m then used to predict the temperature responses for that tree cover change.”

line 399:
this doesn’t appear to be a valid general statement. In some cases deforestation show warming, in others cooling, and in others little to no change.

This is a good observation, we have now removed this statement

lines 425-426:
in my experience in temperate north america tmin is usually closer to sunrise and tmax is later than 1300, particularly in summer. This is one reason why I don’t think the Hooker parameterization is adequate. If you want to maintain independence from the ESMs with respect to diagnosing 2m temp, you need to re-do the hooker relationship using actual tmin and tmax values.

This is a valid point, we now add more discussion on this in Sections 5.1 and 5.2. Specifically, we note that there may be a discrepancy between night/day and min/max temperatures. Bearing in mind that we are looking at relative changes, we furthermore note that this does not affect the biases so much (as demonstrated in Figure 7). However, for select months and regions (e.g. winter in North America and Europe), it does seem to have an impact and we bring this to light within Section 5.2.