Reply to editor and reviewers

Again, we thank the comments by the reviewers. We have prepared a new version of our paper addressing the minor issues pointed out. We have made an effort to tone down the language in this new version. We hope that the new version is suitable to be published in Geosc. Mod. Dev.

Reply to 'Topical Editor Decision: Publish subject to minor revisions (review by editor) (12 Dec 2020) by Richard Neale'

Comments to the Author:

Following the second round of reviews this article can be published as is subject to addressing the reviewers' minor revisions. I would certainly ask that you heed them closely as there is a significant amount of speculative and absolutist language used in the paper that still remains from your original submission. This topic is indeed important, representing the core principles that underlie GMD mission that is rare among journals, and so important to GMD. But be aware that GMD reviewers are not social scientists and cannot speak to whether the techniques you have employed in this study are well designed or scientifically accepted. Therefore we are relying on the importance of the subject rather than any absolute certainty in the appropriateness of the methods used.

Regards
Rich Neale

Dear Editor,

As requested, we have addressed all the concerns of the reviewers. We have also fixed some minor issues in the manuscript that we have detected during this new iteration. We have removed the term 'climate change models' from the title, that now reads 'climate models'. Moreover, we have attempted to tone down the language and to make less questionable statements.

Many thanks.

Reply to “Anonymous Referee #1”

Review of “Current status on the need for improved accessibility to climate change models code” by Juan A. Añel and colleagues.

This is my second review of the manuscript by Añel et al. As already mentioned in my original review I find the topic of the manuscript to be highly relevant for the scientific community in general and for the climate model community in particular. Personally I still think some of the rather strong statements about the need to make code publicly available for everyone under any circumstances could be framed differently and be better discussed as suggested in my original review but this is up to the authors. Apart from that I found several statements which still need to be clarified as detailed below.

Specific Comments:

Title: I maintain that CMIP5 models are not “specifically intended” to investigate climate change but also many other aspects of the climate system such as feedbacks, uncertainties, etc even in the absence of climate change (e.g., using the piControl runs). But if the authors are convinced that climate change models is more accurate in the title this is fine.

We have reflected on this issue again. Indeed, the CMIP5 is probably better known for its contribution to studies on climate change. However, it is also true that Geosc. Mod. Dev. is a journal with a readership that probably knows this difference.
To summarize, we have decided to accept the reviewer's suggestion and remove the word 'change' from the paper title. Probably it does not change the final impact of this work, and the message that we want to transmit will reach the same audience.

Page 1, Line 4 & p2, l20: “Climate Model Intercomparison Project (CMIP5)” CMIP5 actually stands for “Coupled Model Intercomparison Project” https://esgf-node.llnl.gov/projects/cmip5/

Fixed. Thank you for pointing it out.

p1, l13-15: For readers (and reviewers) who are not experts in the field of code reproducibility it would be helpful to have a short introduction what CSR means here (and in general) and what makes it so complex. I assume it is not enough to just put code in some repository? And also where does it end? E.g., climate model output is (to my understanding) not bit-by-bit reproducible, what does that mean in this context?

As suggested, we have included in the text a definition. In this case, we have chosen the one published in 2019 by the U.S. National Academies of Science, Engineering, and Medicine. Also, we cite the work. Now the text reads:

“*CSR, as defined by the U.S. National Academies of Science, Engineering, and Medicine, means 'obtaining consistent results using the same input data, computational methods, and conditions of analysis’*”

We note that in our previous version of the manuscript, although later in the manuscript, we already cited the web page of the ACM where it is defined (ACM. 2018).

P2, l23-26: I've raised some of these points in my first round of comments already and I'm raising them again here because I think it is crucial to be precise with such rather controversial statements. Without having done any study myself my intuition is to agree with the authors that in several (many) instances climate model code is probably not following an "ideal level of programming practice".

BUT: This must not be generalized as the authors do it here. I'm equally convinced that there are climate models out there that can serve at best practice! Therefore I do not think the word "generally", should be used here.

We have removed the word, as requested by the reviewer.

Several things the authors should be precise about:

- not all climate models are in CMIP (think of a simple 1D energy balance model for educational purposes – this might be perfectly coded, documented and licensed) I assume the authors do not refer to such models here even though they use the generic term "climate model"?

We consider that this is obvious. Also, in our view, adding a mention to 1D energy balance models, single column, etc. would mess up the readers. Moreover, along the manuscript, we clarify several times that we refer specifically to CMIP5 models. Only those not familiar with the field of climate simulation would doubt about the complexity of such models. For such people, the barrier to understanding our work hardly would be the difference between 1D and GCMs.

Therefore, in this case, we have opted for not including additional explanations in our paper. We hope that our reasoning for it is understood.

- "ideal level of programming practice" is fairly abstract. The authors introduced the CSR earlier, why not continue to use it (here an in other instances)? For example, if a model is not published how would the authors know if it follows coding and documentation standards? Or is code publication part of the programming practice so that any unpublished model automatically does not follow such practice?

We do not use CSR because with 'programming practice' we refer to a different thing: Standards on coding, comments and documentation. The programming practice is just one of the multiple barriers or faces involved in full CSR. Failing in programming practice makes it more difficult to comply with CSR, but it could be less compromising than, for example, legal restrictions imposed by a license or do not sharing the code.
However, it is interesting that after several versions of our manuscript, this point can be misunderstood. To make it clear, we have added now a brief explanation in this line, that reads:

"*(i.e., coding standards, number of comments, documentation, etc.)*"

- "García-Rodríguez et al. (2020) show how programmers have tended to perform very poorly in this regard in particular, and the incidence of comments throughout the code of CMIP5 models is very low." I'm sorry but I'm not able to find any paper under this citation merely a software tool. Am I missing something or is this what the authors want me to look at? In any case I argue it is impossible to state that "the incidence of comments throughout the code of CMIP5 models is very low" as the code is not available for all CMIP5 models (based on the results of this very study!).

Maybe because this comment by the referee was submitted several months ago, this could be true.

The manuscript pointed out was submitted to SoftwareX for more than one year, and very recently the editors decided that it was out of the scope of the journal. We have just submitted it to IEEE Access. However, the preprint and the Additional Information that contains the evaluation of the code of CMIP5 models are available there. We include here the direct link, in case that the reviewer continues to have problems to access it:
http://fortrananalyser.ephyslab.uvigo.es/docs/AdditionalInformation.pdf

About the second issue, it is true that we have not got access to all the models and that therefore, our statement is a generalization that we can not do. We have rewritten the sentence to be more precise, now it reads:
"*the incidence of comments throughout the code of some CMIP5 models is very low*"

P3, l9: Maybe the authors can already here link table 1 as it also lists all models involved?

Done.

P4, l3-6: I believe the contact information field is available in all NetCDF files for all models as it was required by CMIP5 not only in five models. I would therefore argue that this might even be the preferred place to look at.
This is a minor point but if the reason for publishing code is scientific reproducibility, it seems not unreasonable to me to require knowledge of the NetCDF standard from someone who is trying to run a climate model (which arguably requires way more expertise than opening a binary file).
Do I understand correctly that the five models mentioned were indeed NOT contacted in the end?

Again, it is interesting to see how some misunderstandings persist. The reviewer's argument on running the model would be right to study the results' 'replicability'. To assure reproducibility, it is unnecessary to run the model. However, it is necessary to have the tools, and the availability and information to do it. In this way, anyone without need or capability to run a model can verify its computational scientific reproducibility. A completely different thing is assessing the replicability of the results. However, the point is that the models should be available in open repositories without contacting anyone.

About the five models: We have clarified that the contact information is available in the NetCDF files for "another" five models. In this way, we make clear that we look for the information for all the models, but we only find it for five.

About contacting the groups developing these five models: The interpretation by the reviewer is right. We did not reach the groups. In our view, to make additional requests during the review process, with different timelines, would remove part of the value of our analysis, trying to reach all the groups or research centres at the same time and under the same conditions. In our view, they initially failed to communicate on their contact and therefore to make available the code.

P4, l14: "This analysis is relevant." Delete this sentence?

Done.

P3, l25-, Table 1, and Appendix A
I've mentioned this in my last review already and the references to the different mails are still unclear to me. Here is an attempted summary:

- Mail 1 & 2: "anonymous" requests; given in A1
- Mail 3: request explaining the research; this seems to be missing as A2 which is labelled as "Third mail" seems to be something else
- Mail 4: this is the survey send if access was denied (but presumably not if there was no answer?); I assume this is the "Third email" in A2? Even though the text in A2 is oddly specific and seems to be taken from a longer exchange??

To avoid further confusion, now we identify in the text and the caption of Table 1, what Appendix corresponds to each email.

About the text in the third email: The reviewer was right. In our previous version, we included an email very similar to the third one that we sent out after not getting a reply. It was part of the exchange with one of the groups that denied access to the code. As said, the cause of our confusion when including it in the manuscript was the similarity of both emails. We include now the correct version of the email. Thank you for pointing it out.

In addition table 1 still lists only Mail 1 & 2, it is unclear to which mails that refers.

As said before, in this version, we clarify it in the caption.

Reply to Anonymous Referee #2

We thank to the referee for the insight provided and the relevant questions pointed out.

I thank the authors for their detailed replies to my comments and acknowledge the rewriting of the conclusions. I still find an historical context about code sharing policy in the climate modelling community would have been valuable, but I understand authors points that it could be the scope of another study (and a lot of work).
I suggest to accept the manuscript for publication after some revisions.

We want to thank the reviewer for their supportive comments. Given that we find the interest in the models' historical development, we have begun to extend our study to earlier CMIP versions. However, at the moment, it is far from a point where we can present sensible results. We hope to publish it in the future.

The two points that still puzzle me as a reviewer are :
1/ Authors refer, in their replies, and also twice in the ms, to Garcia-Rodriguez 2020 manuscript submitted to SoftwareX, which we cannot access to, annihilating my ability to understand/evaluate their point both in their replies and in the text.

We can not understand what happens with this. In the list of references, we include the webpage to download Garcia-Rodriguez et al. 2020. It is there, it is online, hosted in our university serves. To the best of our knowledge, it has not suffered sensible downtimes. Both the manuscript and the Additional Information (that describe the models' evaluation) are there, linked at the top of the webpage. We include here the direct links to the content, in case it is necessary:
http://fortrananalyser.ephyslab.uvigo.es/docs/softwarex_article_template.pdf
http://fortrananalyser.ephyslab.uvigo.es/docs/AdditionalInformation.pdf

Also, as we point out because of comments by Reviewer#1, the manuscript is now to be submitted to IEEE Access after being rejected in SoftwareX. In the end, after a second review, the editors considered that was not totally in the scope of the journal.

2/ As in my first review, I am still puzzled with the strength/tone of some statements like "such practice is not acceptable" "cannot be an excuse" that are uncommon in the scientific papers I am used to read/write/review. I guess this particular manuscript aims at making a strong point to improve code sharing in the climate community, and somehow there is a necessity to make such statements. But I feel uncomfortable with such sentences, while confessing being unable to really evaluate their relevance.

We understand the point by the reviewer. However, sometimes things are black or white. Or we are complying with the scientific or we are not. But there is no such thing as 'we do our best'. Moreover, this depends on human decisions to make information available, not on technical or scientific problems.

That said, we have modified some of the sentences. We think that we can translate the same message without to sound aggressive. Instead of "*such practice is not acceptable*", now the text reads "*such practice severely undermines the validity of the results*".

The part of "*cannot be an excuse*" has been rewritten and now reads: "*Barriers to code-sharing through licensing, imposed by, e.g., government bodies, are artificial barriers that depend only on human decisions. They are not due to technical difficulties or scientific reasons.*"

Other minor comment :

Page 4, line 23. "This analysis is relevant" I suggest deleting this sentence which seems to be cut from the previous version of the ms.

Done.

PAge 7, line 20: "It must be had". Please rephrase.

Now the text reads 'It must be considered'.