Geoscientific
Model Development
Discussions

Interactive comment on "Current status on the need for improved accessibility to climate change models" *by* Juan Antonio Añel et al.

Anonymous Referee #2

Received and published: 28 December 2019

In the framework of computational scientific reproducibility, this manuscript attempts to explore the status of CMIP5-class climate model accessibility, following a methodology based on web access attempts and emails. The question addressed here is very important, and an appropriate discussion could help the climate model community improving its code sharing practices. Authors show that only very few of these models comply to a full accessibility. Such a result should rock the boat of the climate modelling community and ignite discussion on reproducibility within successive CMIPs. Unfortunately, in its present sate, that is basically the only result this manuscript provides, and the text suffers from a lack of discussion and perspective, together with far too many general and un- or ill-referenced statements. I think the main results should be followed by a discussion both on the relevancy of code sharing policy for CMIPs, and

more suggestions to improve it. Also an historical perspective of how code availability was considered in the successive CMIP projects would be interesting to evaluate how things are evolving in the climate modelling community.

Many questions come in mind reading the ms, I am not requesting authors to answer them, but I think they could help building a discussion section: I acknowledge authors suggestion to setup "frozen" versions of the codes, but in more details, what could be suggested, at the international level, to improve sharing policy of the codes/experiments ? What has been done in the past? Why did it not work until now ? Do we have information regarding the ongoing CMIP6 experiments ? What are the specificities of climate model intercomparison projects when compared to other massive modelling works ? Making codes available is one thing, but does it make sense to make a million-line-ish code available without any support ? Can modelling group follow a standard for documenting the setup of a typical CMIP experiment ? Is it possible ? How the ultimate goal of reproducibility could be reached for CMIP models ? On what machine ? Should the community think of compiling/running all the CMIP models on one single machine ? What are the limits of reproducibility in that case ? What was the situation for previous CMIPs, when there were less models involved ? How thoughts on code accessibility did evolve ?

On the methods: Although I am not an expert in surveying methods, I found it puzzling not to have more details about the emailing methodology, i.e. who was contacted in the different modelling group: engineers, researchers ? Climate/ earth system models are massive codes, developed by many people. Did the authors contact, for each model, responsible for each compartments (vegetation, atmosphere, ocean, etc.) or did they just take one contact from each model web page ?

On the results : I did not find the geographical (fig 1) approach relevant. What conclusion can be drawn from that ? Although it is more complicated, I think having a licensing history of each model would be more relevant to connect to their accessibility.

Specific comments :

P1,l7: "There are other reasons that justify the need for access to the codes of climate models used in scientific research. One of the most important is to prevent the loss of knowledge on the cycles of development of these models. Some of them nowadays rely on 'legacy' code that was written up to five decades ago, and new developers must understand why some decisions on implementation were undertaken so long ago. Âż Although I am convinced by the need of improving code sharing policy, I am not sure that it will help reducing loss of knowledge. From my experience it seems that many steps to improve a climate model code are either recorded internally, i.e. within the institutes documents, or through successive publications. I don't see how code sharing will improve that.

P1 L13: "The complexity of the problem, where in some cases scientists may be unaware of some of the determinants, or may make subjective judgements that have little to do with the most appropriate from a scientific point of view (Joppa et al., 2013), or may even fail to make the correct assessment, makes it necessary to consider a range of issues (Añel, 2017), including legal aspects.".

I must confess I don't understand this sentence.

P2l13 "It could be said that adequate sharing and documentation is not necessary if the code used in the models includes appropriate comments, but it is generally the case that climate models do not comply with what would be the ideal level of programming practice." That is a really strong statement that should be underpinned by appropriate reference.

"It is widely acknowledged that some scientists are reluctant to share code because of the perceived potential damage to their reputations." I am really surprised by this statement. Is it supported by any survey ?

"Given that many scientists have no formal training as programmers, it may be pre-

sumed that they consider that their code may not comply with the standards of excellence that they usually pursue in their main fields of knowledge. Indeed, it has been clearly documented that some climate scientists acknowledge that imperfections in climate models exist, and they simply address them through continuous improvement without paying too much attention to the normal techniques of software development (Easterbrook and Johns, 2009). Nevertheless, all scientists must believe that their code is good enough (Barnes, 20 2010) and that there are thus no reasons not to publish it (LeVeque, 2013)." This paragraph, as the previous sentence, suggest climate scientists, aware of their code imperfections, would be reluctant to share it. It must be supported by a reference or a survey, if not it is just a feeling.

"Barriers to code-sharing through licensing, imposed by e.g., government bodies, cannot be an excuse and when contributing to scientific studies and international efforts where collaboration and trust are critical, such practice is not acceptable." This sentence is more an open-ed-like statement that what is expected in a scientific journal. Questioning licensing is appreciable but it should be made in a more rigorous way.

"For cases where we obtained the code of a given model, we were not provided with a reason for the license behind it. In fact, in some cases despite getting the code we did not see a license explaining clearly the terms of use." Indeed it would have been crucial to obtain, for every model and every component, the license terms used. That would have helped a lot to discuss accessibility.