Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2016-26-RC1, 2016
© Author(s) 2016. CC-BY 3.0 License.

Earth System Dynamics
Discussions

Interactive comment on "Towards improved and more routine Earth system model evaluation in CMIP" *by* Veronika Eyring et al.

Anonymous Referee #1

Received and published: 21 June 2016

The authors advocate the very laudable goal of developing a set of diagnostic tools that could be applied during future phases of the Coupled Model Intercomparison Project (CMIP). Unfortunately, however, the authors remain rather vague regarding some elementary design features of the proposed framework in which these tools should be implemented. For example, the question whether the set of tools should be easily portable to users' platforms or whether it will be more or less tied to the ESGF framework is only answered somewhat implicitly on page 9, especially since "open source" clearly does not imply portability. Also, it would be interesting to know whether the proposed code can also be used to interpolate and/or extract data or whether it will only provide ready made plots. I think it would be good to answer these questions already at an early stage in the design of the tool. I also think an overview of the key requirements/specifications (or maybe it is better to say "goals" instead of "requirements" since

after all the authors are planning to provide another valuable and voluntary service to the community) in a list that might help the reader to understand what might be coming and also to serve as a guide for the further development of the framework would be helpful.

Specific comments and questions:

1) could the proposed code also be used to interpolate and extract data or will it only provide ready made plots? Interpolating and extracting seems to be one capability that is needed anyway in designing the evaluation tool, and that by itself would be of great benefit to the user community. Would it make sense to construct the tool in order to eventually do data extraction and/or interpolation on the server side and plotting on the user side?

2) in case a standard set of plots will be provided during CMIP, will the plots be archived permanently, will they be citable, and should they be copied and included in publications?

3) p. 3, line 26f: I find that there has been absolutely no lack of studies that have pointed out model deficits, and modelers do know where the problems are. One thing that has been missing is sufficient investments in climate model developers. My impression is that having a standard suite of evaluation tools will not do much to actually ameliorate the problem described in line 26f.

4) p. 11, line 7: could you either give a reference and/or describe what these "well-defined standard interfaces" might look like? (Please see also my "additional comment" below).

5) would it be better to use just a single language (python) that can in principle could replace the potpourri of all the other languages in the ESMValTool? Please discuss.

6) who should users contact if they are interested in contributing to the tool? Who will decide which diagnostics are to be included and which diagnostics are not to be

included in the framework? Will there be "standard" and "user supplied" diagnostics? I understand that some strategy still needs to be developed, but it would also be nice to know what the possible outcomes of these developments might be.

7) p. 10, line 2: "users can however make substantial use of the tools by downloading the open source versions and by running them locally on their machines" -> this seems to me a major design requirement and it should be mentioned already at an early on in the manuscript. Is the code meant to be portable? Or will it tied to the ESGF servers?

8) is it thought that individual users will eventually be able to adapt the code that they run on the ESGF machines? In other words, will users eventually operate their own version of the code on the servers in which they can adapt not only namelist settings, but also add diagnostics? Will it be possible to use additional data that might not be stored on the servers in these diagnostics? If yes, how could this be achieved? Would it make sense to do data extraction and/or interpolation on the server side and plotting on the user side as suggested in point 1 above?

9) p. 7, line 25: how are the groups supposed to use the tool during model development if it is run on the ESGF nodes? Will there be a stand-alone and an online version or will it just be one tool that can do both jobs? And also, how are you planning to deal with the dual requirements that the code should facilitate automatic processing while at the same time be user friendly, highly portable, and easily adaptable and expandable?

10) p. 2, line 26: I can not find any useful documentation of CMIP5 models under the web site given for ES-DOC. This reminds me of another project that has received funding for collecting meta-information on CMIP5 models, but that provided a poorly designed questionnaire and website to the model developers and as far as I can see has ultimately failed to be useful to users as well.

11) on p.8 in line 26 you are suggesting that the software will be able to acquire cache data from other servers. Will this cache data be kept for when one of the other servers is down? My experience has been that due to the distributed nature of ESGF it is

sometimes very difficult to have access to all the data sets one wants to analyze at a given time. Would it be useful to cache processed (interpolated/extracted) data on the user side once the tool is opened up to users?

12) p. 9 line 7f: "these supernodes have the necessary storage and computing resources". In line 17 it says: "requires the extension of current hardware" and in line 30, it says that the computing resources might not suffice for users to base their own analysis on this tool.

13) p. 10 line 7: "whereby new diagnostics developed by individual scientists can quickly and routinely" -> how will porting diagnostic tools be handled? Especially, what do the scientists have to do in order to port their diagnostic tools to the framework or to have them considered?

14) p. 15 line 15: to me it seems important that the data version should be somehow documented. Yet, this is not mentioned here. As far as I can see, with CMIP5, finding out the version of a data set can only be achieved via sending a query with a checksum to an ESGF server. Maybe the users' interest in version numbers for the data sets has been underestimated? Also, will old versions of the data be stored so that one can reproduce results later without having to keep a local copy of the data? With CMIP5 this is not clear to me.

15) p. 11, line 3: I think for all practical purposes, this would require either a new electronic data base format for citing the data or else summary doi's, which I don't think would work. I don't think that having 500 references to data sets each with its own doi would make much sense in something that might be printed on a printer, even if it would certainly be possible to automatically generate the corresponding list.

16) p. 12, line 2f: "Model evaluations must take into account the details of any model tuning" -> how? Are you planning to archive output from all the untuned model versions? I don't understand what this sentence and also the following sentences might mean in practical terms. I also don't quite understand why this might be useful at all.

I think that it might be nice to have output for the same model tuned in different ways (maybe as "physics options" p1, p2,..."). But the sentences in the manuscript sounds like you are advocating the archiving of data for untuned models? If yes, please explain what you expect to learn from this. I do not think that archiving the data of untuned models within the framework of a model intercomparison projects makes much sense. Untuned models do not generally simulate a realistic radiation balance at the top of the atmosphere, and I think it is save to discard them in for the sake of model-intercomparisons, especially since you are talking about comparisons with observations.

17) p. 14, line 15: "requires ongoing maintenance" -> very good point. How can this be achieved?

18) Fig 1: given this centralized approach, how can sufficient reliability and redundancy be achieved? Just recently, the ESGF nodes have been completely unavailable for several months.

19) Notwithstanding my criticisms above, I do think that the ESGF people have on the whole done a great job and that their efforts have been extremely useful to the community. I also very much appreciate the initiative for the standard model evaluation tool, and I am confident that it will ultimately be very useful as well. I was also glad to have find other sources of the CMIP5 data while the ESGF servers were down.

Minor Points:

1) p. 12, line 27f: for an "emerging constraint", one needs a relationship between climate sensitivity and a model diagnostic that varies between models but can be constrained by observations. I think the formulation in the manuscript is not entirely clear.

2) p. 12., line 31: "might" -> could be considered more likely to

3) p. 12, line 32: "A question raised ... " -> I don't understand what is meant here. Please re-formulate.

4) p. 12, line 33: "Moreover, ..." -> I think that this is a very good point.

5) p. 13, line 3: "studies need not lead to contradictory results" -> I don't understand this sentence. Please re-formulate.

6) p. 13, line 21: in my opinion, one key question might be how easily adaptable this platform is by individual users

7) p. 24, line 24: could you please specify what you mean by "revolutionary"?

Additional comment:

I am using an analysis framework in which placeholders such as "###(obs_data_path)###" are used for variables in analysis scripts (which are e.g. in ncl, R, python, etc) which are then inserted e.g. based on values specified in .xml files. In other words, the xml file and the analysis scripts are parsed by a preprocessor that then inserts whatever values are provided by the .xml file into the scripts (e.g. paths to data, etc.) before the scripts are automatically executed. I liked this more than the interface approach in which various interfaces are used for the various languages.

In my diagnostic package, one can combine diagnostics into packages by specifying the package name in the .xml file and then run a package of scripts. I do, however, sometimes ask myself whether I should convert to a language such as python that would make the whole construct more uniform.

Technical comments:

p.1 line 4: Scientifically more research -> nice pleonasm

p.9 line 7: was the list intended to be in alphabetical order?

p. 10, line 2: can -> could

p. 13, line 18 this is -> this would be

p. 13, line 19f in shared -> a shared

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-26, 2016.