Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-100-RC1, 2020

Geoscientific Model Development Discussions

Interactive comment on "ESM-Tools Version 4.0: A modular infrastructure for stand-alone and coupled Earth System Modelling (ESM)" *by* Dirk Barbi et al.

Anonymous Referee #1

Received and published: 20 September 2020

GENERAL COMMENTS

The manuscript by Barbi et al. presents the ESM-Tools v4.0, which is a modular, open-source software to "use" Earth System Models more efficiently, by providing a standardised infrastructure that covers, orchestrate, and wraps around all aspects of working with ESMs from obtaining the model's source code, building on different HPC systems, run control, monitoring, to post-processing.

This topic is highly applied and very relevant for the geoscience modelling community, it is in the scope of GMD and appropriate, presenting novel tools for, very broadly speaking, model system maintenance. The software is well designed and follows a

logic and concise concept. It is well maintained as community code on github.

Overall, however, in the opinion of the reviewer, the manuscript in its current form needs a careful restructuring within the individual sections and many refinements throughout in order to help the reader fully appreciate and understand the capability and potential of the ESM-Tools. Many parts lack a framing into the bigger overall context or even the ESM-Tools themselves. Also, some technical aspects should be presented in more detail to allow the readers to decide whether investing time in testing and test-implementing makes sense for their respective situation.

Apart from this general criticism, the manuscript by and large reads well and fluent. The Abstract and Appendix are OK, but as the rest of the manuscript need reworking, see above and detailed comments below. A more detailed schematic on a ECM-Tools workflow could be aided by more examples in the appendix; or those examples provided should at least be integrated more tightly. Also, more references could be provided to provide more background information.

SPECIFIC COMMENTS

(P: page, L: line or lines)

1. While the referee agrees with most of what is stated in the introduction, the introduction is lacking references, e.g., on current developments of ESMs, in the context also of HPC development, which makes using coupled model systems especially complex, by making, e.g., full use of emerging heterogeneous HPC systems. References that point into this direction might be, e.g., Bauer et al. (2015, Nature, doi: 10.1038/nature14956), Schulthess et al. (2019, Computing in Science Engineering, doi:10.1109/MCSE.2018.2888788), or Fuhrer et al. (2018, GMD, doi:10.5194/gmd-11-1665-2018), to name a few.

2. The sentence on P2 L26ff implies that ESM-Tools in fact aids in implementation of the the coupling itself; however, the lacking modularity of the ESMs component models

and difficulty of coupling them is not solved for models which are not part of the ESM-Tools, which are not a generic tool to help in the coupling as such (later on this is also clearly pointed out). Perhaps rephrase these parts so the actual scope of the software, which indeed can substantially aid in setting up, maintaining, using coupled model systems, becomes more clear. Maybe it already helps by moving the ESM-Tools notion out of the first paragraphs and further to the end of the introduction and focus in the first two paragraphs of the introduction on characteristics of state-of-the-art ESMs and HPC systems and challenges in efficiently developing and implementing and maintaining ESMs (the paragraph on the workflow involved) as a state-of-the-art and definition of the problem. In this part of the manuscript, it might help to clearly identify and list the different challenges along the workflow. This would put the listing of similar tools such as the ESM-Tools, which comprises most of the remainder of the introduction, into a better perspective.

3. The term "provenance" is not mentioned although the ESM-Tools offer a very nice auto-documenting functionality by means to YAML files, which eventually describe the setup and configuration very nicely. With these features a provenance tracking of complex numerical model runs becomes possible and lays the foundation for "reproducibility".

4. The manuscript could better or earlier-on in the text elaborate, that a prerequisite of the ESM-Tools usage is, in case of a multi-model ESM, that either modified source code repositories are used or some sort of source code patching is done, in case official source codes are used, so that the specific model coupling procedures are taken care of. This related to comment Nr.2 above, the crux of ESM software modularity is (among other things) the coupling as such. This is mentioned very clearly in the Discussion, but the manuscript would be more clear if these aspect were more explicitly addressed beforehand, where the manuscript seems a bit imprecise.

Additional comments to different parts of the manuscript:

"Abstract" - P1 L4f: I agree that the number of coupled setups increases, but I cannot see why this leads to a growing number of available versions of component models, except it means that they all need to be equipped with the respective coupler technology. - P1 L5f: Specify "system-dependent", it is HPC-system dependent; also aside from obtaining and operating, it is also "building" and "configuring", as mentioned further on in the introduction. - The content of the footnote should be more emphasised by putting this information in the introduction after the reader can appreciate why a focus on coupled model system is more rewarding, perhaps at the end of the introduction when the focus and scope of the manuscript is given.

"Introduction" - P1 L20: Perhaps expand the list here, not only climates are studied but all kinds of interacting and related processes in the Earth system for a multitude of reasons - P1 L21: Also "hydrology" should be mentioned in the list and the list refers mainly to "compartments" - P2 L23: Is it the ESM that lacks modularity or rather the individual component models, which make up the ESM, and which have been developed for a specific purpose? - P2 L48: Mentioning DSLs, cite, e.g., Lawrence et al. (2018, GMD, doi: 10.5194/gmd-11-1799-201), see also the general comments. - P2 L44ff: In case tools like MESSy are mentioned, wouldn't then also tools such as the Earth System Modelling Framework (ESMF)? - P3 L83: Which "specific criteria" are you refering to exactly?

"ESM-Tools Description" - This is a central part of the manuscript; as mentioned above, a restructuring and adding more information might help in better understanding the functioning of the tools. - P4 L93ff: As the introduction only gives little systematic overview of the individual steps and tasks involved in eventually getting an ESM to run ("to use"), I feel this paragraph is too detailed and lacks some introduction. I would suggest to introduce the structure, the overarching concept, the design principles of the ESM-Tools first. For example in on P4 line 93: which functions (I assume the individual tools that comprise the ESM-Tools and their functionalities)? - P4 L98: Rephrase: this reads as if the reader has to know the previous versions of the ESM-Tools. - P4

L110: What can be done with the plugin manager in detail, which functionalities do the ESM-Tools authors envision? - P5 L113ff: As above, without prior knowledge of the ESM tools, this section appears on the one hand side very detailed and on the other hand side it lacks introductory information: for which aspects of the complete workflow do YAML files exist? - P6 L137ff: This section reads as if different component models can be integrated and coupled with ESM-Tools. But in fact only those component models and combinations thereof, for which a dedicated coupling has been implemented already, irrespective of the ESM-Tools, purely driven by the coupling paradigm and software used (e.g., YAC, OASIS-MCT, MESSy, ESMF, etc.), can be integrated into the ESM-Tools and "managed" and re-arranged in an efficient way. This is how the reviewer understands the ESM-Tools functionality. See also the general remarks on this issue above. This is not a shortcoming of the ESM-Tools but should perhaps just be more clearly and explicitly stated – as in the Discussion. - P7 Fig1: Instead of this schematic, perhaps provide a workflow overview, which shows which tools use which YAML files etc., maybe split into different aspects of model usage: retrieving source code, building, configuring, etc. Also the further esm-master options should be further explained if they are mentioned. - P8 Fig2: The figure might be replaced by a listing, or a figure more in coherent with the schematic of Fig.1 - P7 L165ff: As much as the reviewer understands that it makes sense to provide a fully integrated system which covers all aspects of the "use" of an ESM, in how far do ESM-Tools compare run time functionalities as covered by the esm_runscripts compare to tools such as ecFlow, cylc, or JUBE? Do the esm_runscripts offer a specific benefit? - P8 L175f: Why is esm_siz not mentioned in the overview table 1; again here: given the rather complex implementations of cron jobs etc. to run the esm_viz, a schematic showing the overall structure of the ESM-Tools or a workflow would be very helpful - P8 L185ff: Does the GUI automatically pick up which setups, configurations etc. are possible depending on the available YAML files as part of an ESM-Tools installation? - P9 L192ff: A crucial and important aspect seems to be where the model systems come from and how they are provided: Just retrieving any component model from its official reposi-

tory without the extra very time-consuming step of coupling this component model to other compartment models does not help; i.e. the component models either have to be modified a-priori and are provided through specific repositories or official model codes are retrieved and then patched to provide the coupling functionality. This is the lacking ESM component model modularity, wich is mentioned in the introduction and also tackled in several model development approaches towards unified model systems with interchangeable components (e.g., the SIMA efforts by NCAR). - P9 L200: Perhaps more HPC information should be given as this seems a major problem in using ESMs, to efficiently set them up (i.e. compile, arrange data flow-paths, processing chains, etc. wich incorporate the intricacies of the respective HPC system) - P10 Tab3: I think this table should also include information on the coupler used, as this is determining how truly modular component models can be combined. - P11 Fig4: Given the Tab2 and Tab3, the reviewer cannot see how Fig.4 provides extra information. - The examples of the YAML files in the Appendix should be included more in the ESM-Tools Description. Good way of maintaining reference implementations.

"ESM-Tools Development" P11 L209: The ESM-Tools seem to follow various best practices in modern software development. Does a formalised software development plan exist? ESM-Tools are provided via github, this could also be mentioned here. Also: Can you provide any recommendations in software development, which are worth being shared?

"Discussion" - P13 L252ff: See my comments above, just to avoid confusion, this aspect should be mentioned earlier on. - P14 L295ff: This is a concise summary and despite the fact that it is mentioned that ESM-Tools provides a standardised way of working with ESMs it could be emphasised more before. - Perhaps information could be provided for interested modelling groups how long it takes approximately to get a new model system or a single component model integrated into the ESM-Tools.

TECHNICAL CORRECTIONS

- All text: Check acronyms and abbreviations. - L3: replace "...components of the Earth..." with "...compartments of the Earth..." - L131: oasis should be in capital letters, OASIS - L287: object-oriented

---

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-100, 2020.