

## ***Interactive comment on “ESMValTool v2.0 – Extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP” by Veronika Eyring et al.***

**Veronika Eyring et al.**

veronika.eyring@dlr.de

Received and published: 19 March 2020

### **Reply to Anonymous Referee #2**

We thank the reviewer for the very detailed and helpful comments. We have now revised our manuscript in light of these and the other review comments we have received. A pointwise reply is given below.

#### **General comments:**

C1

**This manuscript entails a monumental effort in attempting to facilitate the development and evaluation of climate models. Examples for analysis reproducibility, particularly output figures from IPCC chapters is commendable. A pathway to expand this to output figures in the literature is also evident. Model performance metrics, diagnostics for the evaluation of processes in different realms are presented in great detail along with the corresponding recipes. Example figures as a result of integrating community metrics is also shown in the manuscript. The flow and the content could be more consistent so the focus of ESMValTool goals and the impact in doing that is delivered as intended. Some level of brevity, citing references for details, providing more example figures from recipes, pointers to additional recipe documentation – should be made available through an external reference and/or supplementary material. Scalability and interoperability aspects can also be briefly touched upon, providing guidance to the community, making interoperability and practicality – a key to expanding the audience. There is scope for condensing and merging certain sections. Some key points to help improve readability is furnished below in specific comments. Overall, thank you for the contributions. Please see more comments below.**

Thanks for your suggestions. Please see our responses to the specific comments below how we have addressed them.

#### **Specific comments and technical corrections:**

**Page1, Ln 58 Reproducibility - Specifics and explicit wording is required here, as to what aspect is targeted.**

C2

We have rephrased accordingly.

**Page2, Ln 85 There may be more references that need to be cited while discussing data standardization for CMIP.**

**E.g. [https://pcmdi.llnl.gov/mips/cmip5/CMIP5\\_output\\_metadata\\_requirements.pdf](https://pcmdi.llnl.gov/mips/cmip5/CMIP5_output_metadata_requirements.pdf)?**

More references added.

**Page 3, Ln 92-83: The line about “full rewards of the effort. . .” should be reworded to provide more of a positive tone to the available observations and model output in standardized format. Expanding what is meant by “full rewards” will be very helpful in this line, rather than the subsequent paragraph.**

The paragraph has been reworded.

**Page2, Ln 96-97: Please cite or provide links to appropriate references w.r.t data volume estimations for CMIP. Also, what is the database that is being referred to here?**

References added and the database clarified.

**Page 2, Ln 100- I like the addition of “creativity” here.**

Thanks.

**Page 2, Ln 107- “that provides results. . .” - Substitute results with something more specific. E.g. analysis products/output?, so it better connects with Ln 108 (This is realized through..)**

Changed as suggested.

**Page 3, Ln 115- Does ESMValTool preserve the netCDF metadata (global attributes from input datasets) in output products? How is data provenance established? ( Ln 142 may answer this, please clarify)**

C3

The ESMValTool does preserve the netCDF metadata including the global attributes. These metadata are also written to the products (netCDF and plots) using W3C-PROV (Python package prov v1.5.3). Details can be found in the technical ESMValTool description paper Righi et al. (2020 that we now explicitly refer to here).

**Page 3 Ln 118- Consider stating “Figure 1 from their paper, or from Righi et al.. rather than “their Figure 1”**

Changed as suggested.

**Page 3, Ln 121- The flow from the introduction to companion papers and the present one can be better. Example- Precede the sentence “the use of the tool is demonstrated..” with “In the present paper,..”**

Changed as suggested.

**Page 3, Ln 124-125, Avoid too many conjunctions (and) here. . “Diagnostics and performance metrics and the variables and observations used”.**

Sentence rewritten for clarity.

**Page 3, Ln 129: What does “partly also with CMIP3” mean here?**

Partly refers to the fact that not all diagnostics can be run with CMIP3 data because for some diagnostics not all required variables are included in the CMIP3 data request.

**Page 3, Ln 130: Is CF-compliance and CMOR-compliance required? Please, also cite CF and CMOR references, expand acronyms. The sentence could be changed to - tool is compatible with any CF and CMOR compliant model output? Please change this as needed so users understand what is ready to be plugged in to ESMValTool, and what requires additional work.**

The ESMValTool requires that input data are following the CMOR standard. CMOR is based on the CF conventions but defines some additional metadata on top of it.

C4

For details, see Righi et al. (2020). CF and CMOR references are now added in the revised manuscript and abbreviations are defined.

**Page 3, Section 2. Ln 131-136. The data descriptions in this section are not satisfactory, especially where the manuscript reads “observation from other sources..” . obs4mip publications should be cited here. It will be nice if ty the different observation datasets used in recipes can be listed and cited thoroughly. Also, this section could be merged into the final section 6-7 on Code and Data availability.**

We now refer to Table 3 in Righi et al. (2020), which contains a list of observational data from external sources', i.e. observations that are not from obs4MIPs.

**Page 4, Ln 154-**

- **Reproducing IPCC chapter figures is impressive. Are these diagnostics-and-recipes written working directly with the IPCC authors? What is your advice to the IPCC authors to make this effort a success for CMIP6?**

Thanks! For the IPCC AR5 Chapter 9 diagnostics and recipes, they were written by the ESMValTool development team after the publication of the AR5. For AR6 Working Group I, several chapters are using the ESMValTool to produce their figures. The diagnostics and recipes are then written by the chapter scientists or the lead or contributing authors of the chapter, with support from the ESMValTool core development team. Our advice to the authors of the IPCC AR6 is to write a recipe for each chapter, so that figures can be reproduced any time. This would be a huge present to those involved in a possible AR7. It would enable a direct and prompt comparison to new model generations.

- **How are the recipe names constructed- is there a recommended naming convention?**

For recipes reproducing the analysis of a refereed publication, we use [first-C5

author] [year] [journal-abbreviation], for example "recipe\_lauer13jclim.yml". The same applies to IPCC chapters, e.g. "recipe\_flato13ipcc.yml". For other recipes there is no strict rule, but the authors are advised to give appropriate names reflecting the recipe content, e.g. "recipe\_heatwaves\_coldwaves.yml"

- **How resilient is ESMValTool to changes like the metadata conventions, DRS, etc from CMIP5 to CMIP6, or say another [inter]national assessment?**

The ESMValTool has been developed in a flexible way allowing defining the DRS structure via a configuration file. Metadata conventions are imported from the obs4mips/CMIP3/CMIP5/CMIP6 tables and can be easily extended with new tables.

**Page 4, Ln 160-161: Check and correct line,word spacing.**

Corrected.

**Page 4, Ln 164: How does one add an alternative observation dataset? One of the companion papers might be addressing this? Page 4, Ln 165: How can additional variables be added? Is it the same as the first version of the tool? Following the citation link here, I still could not get information in two hops.**

Alternative observational datasets are specified in the "recipe" (if supported by the diagnostic). Additional variables can be added by custom CMOR tables similarly to ESMValTool v1.0. In case of derived variables, Python scripts have to be provided to do the actual calculations. This has changed compared to version 1.0 in which these variable derivation scripts were written in NCL. We refer to the extensive ESMValTool documentation for more details: <https://esmvaltool.readthedocs.io/en/latest>.

**Page 5, Ln 174: Can there be a reference here to the regridding tools used? Why 4x5?**

Regridding is done by the Python Iris package, which offers different regridding

schemes. We have added a reference to the Iris user's guide at (<https://scitools.org.uk/iris/docs/latest/index.html#>). The grid resolution of 4x5 degrees has been chosen to be as consistent as possible with the equivalent diagnostic used in IPCC AR5.

### Section 3:

- **Throughout the Overview of recipes, under each sub-section, there can be more consistency. Example: For each recipe, one could ensure these are specified throughout: input (include time-frequency requirement consistently as well), output specifications, source, purpose and significance of the metrics, relevant citations to metrics calculations, summary of the recipe, a sample result. Sticking to this consistently can also condense the text.**

We have already defined a common structure how to discuss each of the recipes in each of the subsections. This is already described in the manuscript: "In each subsection, we first scientifically motivate the inclusion of the recipe by reviewing the main systematic biases in current ESMs and their importance and implications. We then give an overview of the recipes that can be used to evaluate such biases along with the diagnostics and performance metrics included, and the required variables and corresponding observations that are used in ESMValTool v2.0. For each recipe, we provide 1-2 example figures that are applied to either all or a subset of the CMIP5 models."

- **Suggest just pointing to references like how it was done for CVDP to get more information metrics.**

We already include many references in the submitted version of the manuscript. For CVDP this is more straight forward than for other recipes, as an externally developed tool exists.

C7

- **[3.3.4] Sea Ice, for instance, can be rewritten to condense text.**

Changed as suggested and sea ice section shortened.

- **Are the recipes part of the github repositories? Where can one find them?**

All recipes are included in the ESMValTool repository on GitHub. The directory structure of the ESMValTool is outlined in the technical description paper by Righi et al. (2020). All recipes can be found in the directory <https://github.com/ESMValGroup/ESMValTool/tree/master/esmvaltool/recipes>.

- **Though line 145 reads that the intent of the focus of the manuscript is not an assessment of CMIP5 or CMIP6 models, the construction of section 3 is not completely aligned with this. The message needs to be reiterated. If this manuscript is intended to be a documentation paper for the diagnostics and recipes used in ESMValTool, the length could be justified to an extent. Otherwise, some sections could be rewritten so focus is retained.**

The goal of the manuscript is indeed to document the diagnostics and recipes available in the ESMValTool. Assessing the CMIP models is indeed not the scope of the paper, nevertheless Section 3 presents a few examples to show how the ESMValTool output could support the scientific interpretation. We would like to stick to this structure as it turned out to be useful for users and developers, but we have followed the reviewer 1's comment to move the sentence on the results from the caption to the text which hopefully also addresses this point. We have also further shortened the paper by following specific comments from both reviewers.

- **Is this manuscript the single source for documentation for all the metrics and recipes?**

No, additional documentation will be provided in the companion papers Lauer et al. (description of diagnostics for emergent constraints and future projections from Earth system models in CMIP) and Weigel et al. (description of diagnostics for extreme events, regional model and impact evaluation and analysis of Earth

C8

system models in CMIP). The manuscript refers to these additional papers.

- **Page 21, Section 4.1. Automatic execution of ESMValTool at DKRZ sounds like a nice step to interface with more users. How scalable is this process? Is the idea to expand this to other nodes in ESGF? Is data replication of such huge CMIP6 volumes something that needs to be kept an eye on, leveraging distributed data access protocols or the cloud?**

The automatic execution of the ESMValTool has been tested only at DKRZ so far, but thanks to the tool's flexibility it can be easily ported to other ESGF nodes. Data replication is certainly an issue, as discussed in detail in Eyring et al. (ESSD, 2016).

**Page 22: Ln 843-844: Section 4**

- **When new plots are created, is there a step that incorporates a basic automated quality assurance conducted?**

The ESMValTool includes checks for data availability and CMOR compliance of all input datasets which provides some first basic quality control. However, at the moment the output of diagnostics (such as plots) has to be checked manually by a scientist to identify anything beyond missing or badly formatted data.

- **Is there a testing suite for each recipe?**

We have chosen not to ask scientists to implement unit tests for their recipes, because this would make the threshold to contribute diagnostics to ESMValTool too high for many scientists. ESMValCore, the part of the tool that runs the diagnostic scripts and does data quality checks, pre-processing, and records provenance, is rigorously tested with unit tests, to ensure that the tool can be used reliably. In addition to that, we are working on setting up automated tests that regularly run all recipes on a server at DKRZ, both during the development of new recipes

C9

as well as after incorporation into ESMValTool, to ensure that recipes are working and produce consistent output. At the moment this is still a manual process, but we hope to make progress on this in the next few months, as the hardware becomes available.

- **Ln 845: The result browser looks good. Steps to reproducing figures viewed from ESMValTool result browser should be made clearer. This is probably the place where the provenance information captured by ESMValTool will come handy?**

Yes, this is the typical target application of the provenance system, as detailed in the technical overview paper (Righi et al., 2020). The steps to produce the figures in the ESMValTool result browser are now further detailed in the revised manuscript.

- **How is the performance of running ESMValTool on CMIP data in an automated fashion, and in general from a disconnected sandbox, regardless of ESGF.**

The ESMValTool performance has little to do with the automatization. Again we refer to the technical overview for more details on the tool's performance.

- **How is the concept of data versioning incorporated within the automated generation of plots using ESMValTool in ESGF at DKRZ ? When there is bad data retracted on ESGF, and a newer version of data becomes available, what is the current implementation like at the ESMValTool-end or the result-browser to notify its users? If there is no mechanism to notify automatically or not-show-the-corresponding-plots, what is the recommendation to the users? In general, what kind of users does the ESMValTool aim to target?**

The figures on the result browser can be sorted by recipe used to produce this figure. By clicking on the figure, also all input datasets can be listed. This

C10

information should be enough as a starting point for reproducing a given figure for anyone familiar with running the ESMValTool. If specific dataset versions are crucial, the whole metadata provided with the plots have to be read and taken into account. By default, the ESMValTool always uses the latest version of a dataset available. The ESMValTool and the result-browser are not capable of notifying users of any changes. Recommendation to any user is therefore to check the result-browser frequently and/or get into contact with the ESMValTool team as stated on the webpage.

**Ln 857: How does the metadata w.r.t the software version get mapped to the actual source code in GitHub? With data DOIs/data citations widely prevalent for CMIP6, does ESMValTool automatically add data citations to the output figures/files? If not, please provide a pathway to achieve this. Page 24, Ln 943: Please provide an example for "preprocessor settings". Page 24, Ln 948-949. Unable to follow this line "...and tags (i.e. what is reported) ". I think these lines are not adding much value at this point.**

The exact version number of the tool used to produce a plot and written to the metadata corresponds to the release tag on GitHub, e.g. "2.0.0b6". Data citations are only added for observational data, while for model data all metadata are preserved. This typically includes the "tracking id" that can be used to exactly identify the dataset. The preprocessor settings are discussed in detail in the technical overview paper by Righi et al. (2020). We therefore do not see any use of repeating these here.

**Page 24, Ln 959. Identifying errors in the simulations early on is a key factor that is penned down as future work here. Even if there are no web-based capabilities, please address if ESMValTool can independently be installed and run by an individual user at different stages in model running. An idea or vision here to draw more inspiration and motivation for using ESMValTool can be**

C11

**provided.**

The ESMValTool can be installed within the user space, so each user can install and develop the tool independently.

**Ln 955, Again, enhancing quality control is a great use-case, but having ESMValTool run on published data on ESGF does not satisfy this use-case. Stand-alone, this tool seems to work towards QC. Please clarify.**

Analyzing first CMIP data published on the ESGF showed that there are still many errors in the metadata and/or the actual data. Examples of such errors include, for instance, wrong units, wrong coordinates (e.g. time) and entire fields consisting of missing values. We therefore do think that running the ESMValTool on these data can be regarded as a quality control process, admittedly on a higher level than the initial quality control done by the modeling groups.

**Section 5. Font size seems to be mixed up in the Summary section - lines till 950 and after 950 are different.**

Corrected.

**Page 25, Ln 965, Sections 6 and 7 should be condensed into one section. Addressing data citations briefly in Data availability will add more value to the CMIP and ESMValTool efforts.**

We have removed the data availability section.

**Comments on Figures:**

**Adding some of the figures to a supplementary or appendix should be considered.**

C12

Since we have one example figure for each of the recipes, we would like to keep them in the main manuscript. This also supports the general structure the reviewer is calling for.

**Verify that there is not much redundancy in the text in captions (e.g. Section 3, 5, Figures text). Avoid redundancy where possible.**

This has been cleaned up, see also our responses to Reviewer 1.

**I find the captions in figures helpful, especially the reference to the corresponding sections. The captions are mostly like IPCC-chapter and documentation paper style. A short caption in bold followed by the description is something that will make the figures stand out.**

Using bold font in the figure caption is unfortunately not supported by the journal standards.

**While specifying OBS in figures, please specify names of OBS in the figures.**

Changed as suggested.

**Name the variables/fields corresponding to the figures, example Figure 4,5,6 - zonal wind,air temperature,precipitation? respectively.**

Changed as suggested.

**How are the color palettes picked in general and what flexibility ESMVal-Tool allows w.r.t color palettes?**

Default color palettes are picked by the diagnostic authors. Some diagnostics allow for setting it as an option. Implementing such a feature, however, is also up to the diagnostic authors and therefore not supported by all diagnostics.

C13

**Expand the acronym QBO in Figure 4, although the section covers it.**

Done.

**Better labeling on the figure itself needed for Fig 15,16 especially.**

Done.

**What is “j” in r1i1j1 in several figure captions - e.g. Fig 22.**

This was typo, thanks for spotting! Changed to r1i1p1, see <https://portal.enes.org/data/enes-model-data/cmip5/datastructure>

**In Fig 20, use r1i1p1 to be consistent, not r1p1i1.**

Changed as suggested.

**Is Figure 26 a reproduction of Fig 9.14 from AR5, Chapter 9. (Including chapter helped me since there is some ambiguity looking up for Fig 9.14 from AR5).**

Yes it is reproducing Fig 9.14 from Chapter 9 of AR5. We have added "Chapter 9 of" before "AR5" for clarity.

**In Figure 30, “typo” - “ether” vs either;**

Corrected

**In Figure 34- typo: predictand, not predictant. Please use long names on the figure themselves, not the short CMOR names (example Figure 35, Y axis); Units missing in some of the figures, e.g Figure 39.**

Corrected.

**In the summary section– Given the challenges of CMIP6 (and beyond) and the scientists all over the globe working on multiple research areas, this**

C14

**manuscript should include something along the lines of the role and future of ESMValTool in the community as a whole and how it can be interoperable with overlapping efforts. The ability to cross-function using tools like ESMValTool and making them more inter-operable is a key challenge. The experience from developing ESMValTool in the form of these manuscripts is helpful to the community, and it can also be helpful for the expansion of metrics-and-recipes used in ESMValTool.**

Briefly expanded on this point. However, since inter-operability with other tools requires discussion and agreement with the developers of the other tools who are not authors on this paper, we cannot define this in detail here. There are also license issues to be considered. We are however very open for this collaboration and have put a lot of effort into actively seeking coordination with other tool developers in the past years. Having different and complementary tools might however well be desirable, see also Eyring et al. (2019).

## References

Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., Guilyardi, E., Jousaume, S., Kindermann, S., Lawrence, B. N., Meehl, G. A., Righi, M., and Williams, D. N.: Towards improved and more routine Earth system model evaluation in CMIP, *Earth Syst. Dynam.*, 7, 813-830, doi:10.5194/esd-7-813-2016, 2016.

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., Collins, W. D., Gier, B. K., Hall, A. D., Hoffman, F. M., Hurtt, G. C., Jahn, A., Jones, C. D., Klein, S. A., Krasting, J. P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G. A., Pendergrass, A. G., Pincus, R., Ruane, A. C., Russell, J. L., Sanderson, B. M., Santer, B. D., Sherwood, S. C., Simpson, I. R., Stouffer, R. J., and Williamson, M. S., Taking climate model evaluation to the next level. *Nature Climate Change*.

C15

doi:10.1038/s41558-018-0355-y, 2019.

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., Lorenz, R., Pérez-Zanón, N., Righi, M., Schlund, M., Senftleben, D., Weigel, K., and Zechlau, S.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – diagnostics for emergent constraints and future projections from Earth system models in CMIP, *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2020-60>, in review, 2020.

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview, *Geosci. Model Dev.*, 13, 1179–1199, <https://doi.org/10.5194/gmd-13-1179-2020>, 2020.

Weigel, K., Eyring, V., Gier, B.K., Lauer, A., Righi, M., Schlund, M., Adeniyi, K., Andela, B., Arnone, E., Berg, P., Bock, L., Corti, S., Caron, L.-P., Cionni, I., Hunter, A., Lledó, L., Mohr, C.-M., Pérez-Zanón, N., Predoi, V., Sandstad, M., Sillmann, J., Vegas-Regidor, J. and von Hardenberg, J.: ESMValTool (v2.0) – Diagnostics for extreme events, regional model and impact evaluation and analysis of Earth system models in CMIP. *Geosci. Model Dev.*, in prep., 2020.

Interactive comment on *Geosci. Model Dev. Discuss.*, <https://doi.org/10.5194/gmd-2019-291>, 2019.

C16