IS-ENES3 Deliverable D9.2

ESMValTool version enabling model development usage

Reporting period: 01/07/2020 – 31/12/2021

Authors: Alistair Sellar (Met Office), Klaus Zimmerman (SMHI), Jan Griesfeller (Met Norway), Javier Vegas (BSC), Stéphane Sénési, Jérôme Servonnat (CNRS-IPSL)
Reviewers: Martin Schupfner (DKRZ), Julie Deshayes (CNRS-IPSL)
Release date: 30/07/2021

ABSTRACT

This report covers developments to support the incorporation of ESMValTool into the model development workflows of five European modelling centres. These developments include enhancements to ESMValTool itself, and adaptations to local workflow tools to allow them to use ESMValTool. The enhancements to ESMValTool include underpinning infrastructure to support the reading of native model output data, model-specific configuration information, and new diagnostics and metrics that are needed within one centre's workflow. A new release of ESMValTool has been made available (version 2.3) and all the ESMValTool enhancements described here have been made available in this release, apart from one diagnostic that will be included in the next release.

Dissemination Level		
PU	Public	X
CO	Confidential, only for the partners of the IS-ENES3 project	

Revision table			
Version	**Date**	**Name**	**Comments**
Release for review	01/07/2021	Alistair Sellar	
Responses to internal review	30/07/2021	Alistair Sellar	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Table of contents

Executive Summary

This report covers developments to support the incorporation of ESMValTool into the model development workflows of five European modelling centres. These developments include enhancements to ESMValTool itself, and adaptations to local workflow tools to allow them to use ESMValTool. The enhancements to ESMValTool include underpinning infrastructure to support the reading of native model output data, model-specific configuration information, and new diagnostics and metrics that are needed within one centre's workflow. A new release of ESMValTool has been made available (version 2.3) and all the ESMValTool enhancements described here have been made available in this release, apart from one diagnostic that will be included in the next release.

The modelling infrastructures differ across these centres, and as a result we have differing strategies for how ESMValTool will ultimately be used in our systems. For some centres, for example the Met Office and Met Norway, the full implementation of these strategies will take a number of years and a large investment of effort, while for others such as BSC and IPSL this journey is more complete. In all cases, the journey has been accelerated by the work reported here, which ultimately will unlock the great potential of the many ESMValTool recipes which have been developed by the international Earth system modelling community, to enhance the scientific scope of our model development evaluation.

Immediate next steps for these centres include continuing engagement with users, provision and enhancements of the process and guidance for ESMValTool development. For the Met Office, the next steps additionally will include the porting to ESMValTool of further AutoAssess metrics, and adaptation of our CMORisation tool. One clear recommendation from the IPSL implementation is the need to speed up the loading of data from multi-variable files, and while a workaround has been implemented, discussions will continue with the Iris and ESMValTool development teams on a long-term solution.

1. Background and objectives

Model evaluation is a critical component of the model improvement process at all stages of development. Proposed model enhancements must be assessed individually for their impact on model performance, and once changes are packaged together, new candidate versions of a model must be evaluated against a previous version. The method by which model changes are tested and evaluated is known as the model development workflow.

There are two primary roles for evaluation in the model development workflow.

- The first is to compare a completed "experiment" simulation (testing an individual change or a package of changes that may form a candidate release) against a "control" model version, which is usually a previously released version of the model. This evaluation must be detailed and thorough to avoid improvements in some aspects to be achieved at the expense of degradation in others.
- The second role is to monitor the performance of a small number of model quantities while the run is progressing, so that problems (or successes) can be detected at the earliest possible stage without the need to wait for a long run to complete. Such capability is sometimes referred to as "quick-look" evaluation.

Model development workflows are typically unique to individual modelling centres, adapted to their own particular needs. The modelling centres in WP9.5 of IS-ENES3 have recognised the efficiency gains that can be realised from adopting a shared community tool to enhance or replace their existing tools. The efficiencies come in two main areas: facilitating collaboration on development, implementation and maintenance of new diagnostics that reflect new understanding in process-based evaluation; and technical aspects such as release testing and maintaining dependencies on powerful but evolving computational packages. The plan is to use ESMValTool to enhance the development of our next generation of climate models which will be used in CMIP7; given the long development timescales for these models, this requires incorporation into model development workflows over the next 1-2 years. The objective of WP9.5 is therefore to work towards incorporating ESMValTool into our respective model development workflows. This has required underpinning developments to ESMValTool that could be used by any model, and has involved changes to the workflows themselves.

It has also required engagement with model-specific communities of users. Change management theory and experience tells us that in order for any technical change to be successful, users must be aware of, desirous of, and knowledgeable about, the proposed change. The model evaluation community is no different from other scientific groups in being comfortable with an existing set of tools and needing good evidence before investing in change. Within this task we have therefore engaged users through a variety of means, ranging from discussion forums to an evaluation hackathon.

Since the modelling infrastructure and user needs differ significantly from one centre to another, the work to incorporate ESMValTool into these workflows has therefore been carried out in parallel by each partner, and we report on the progress of this work at each institute in turn in section 3. However, much of this progress has depended on underpinning infrastructure developments also carried out within WP9.5 and we firstly outline those developments in section 2. Note that the centre-specific sections refer to individual modelling centres, but in many cases multiple partners have contributed to that work, thanks to the collaborations facilitated by IS-ENES. In section 4 we draw overall conclusions and make recommendations for the next steps.

2. Underpinning developments and release management (SMHI)

2.1 Loading native model data formats

A common challenge for all modelling centres is that ESMValTool was primarily designed to work with input data from the Climate Model Intercomparison Projects (CMIP) in CMIP-standard netCDF format, which is not in general how models provide their output natively. Some centres have chosen to address this challenge by incorporating a native-to-CMIP-standard conversion (aka CMORisation) step into their evaluation workflow alongside ESMValTool to allow analysis of the native model output. For other centres it is beneficial to read the native model data directly into ESMValTool without an intermediate post-processing step; this mode of operation has required underpinning developments.

To enable this second mode of operation, SMHI developed and integrated infrastructure for the support of native model output in ESMValTool, in close collaboration with IPSL and with support from the NLeSC. This makes it possible to read native model output in any format supported by the Iris library, among others netCDF4 and GRIB files, aligning the model output on-the-fly with the corresponding standardized variable from the CMIP project. This capability comes in two parts:

- First is the location of the data, that in CMIP is separated with one file containing only one variable; contrary in native model output often many variables are stored together in a single output file, even with varying frequency.
- Second is the adaptation of the model data to follow the established standard as closely as possible. This is facilitated to a large extent by the existing ESMValTool Data Finder and its Fix mechanism. This general framework is easy to adapt to new models while at the same time guaranteeing data homogeneity from the diagnostic perspective.

This was used successfully to make native model output from the IPSL-CM model available to ESMValTool (see IPSL contribution, section 3.3). It was also applied to EC-Earth4, the next generation of the EC-Earth model that is developed by an international consortium with participation of SMHI. However, since EC-Earth4 is still in an early development stage and the output format is not fully settled, this has not yet been integrated into the mainline code of ESMValTool.

2.2 ESMValTool release

Finally, SMHI has managed the release of ESMValTool 2.3.0, making these developments, and those described below, available to the general ESMValTool community. This release is publicly available at https://github.com/ESMValGroup/ESMValTool/releases/tag/v2.3.0.

3. Incorporation of ESMValTool into model development workflows

3.1 Met Office

The Met Office uses a tool called AutoAssess for evaluation of climate model performance in its model development workflow. AutoAssess compares two simulations (an experiment and a control) and calculates hundreds of metrics that quantify the performance across a wide range of scientific domains, e.g. tropospheric circulation, radiation, monsoons, modes of oscillation. The Met Office has embarked on a multi-year project to adopt ESMValTool as its climate model evaluation tool, for both interactive analysis and for automated evaluation in the model development workflow. The work in IS-ENES3 task 9.5 (reported here) has accelerated this implementation through facilitation of collaboration with ESMValTool developers. For the part of the project related to the model development workflow, there are two primary strands of work:

1. Porting diagnostics and metrics from AutoAssess to ESMValTool. In addition to benefiting the wider community by making new metrics available, this is required in order to ensure that ESMValTool can replace the current functionality of AutoAssess.
2. Incorporating ESMValTool into Met Office systems and model development tools. This includes ensuring interoperability with development environments, and interfacing ESMValTool recipes with the evaluation workflow engine (Rose/cylc).

We report on progress with each of these strands in turn.

Following previous implementation of AutoAssess metrics for the stratosphere (Eyring et al, 2020), three new diagnostics have been ported to ESMValTool, covering permafrost, soil moisture and surface radiation, with support from the University of Reading. The full set of diagnostics and metrics are documented here:

- Permafrost
- Soil moisture
- Surface radiation

In this implementation process, it is important to verify that the results are consistent with those produced by the original AutoAssess code. The ESMValTool implementation works on model data that has been converted to CMIP-standard netCDF, and in some cases the AutoAssess diagnostic code had to be adapted to use CMIP standard physical variables instead of native model variables, because there is not a one-to-one mapping between these two sets of variables. This leaves a number of opportunities for implementation errors, and the consistency test is one way to prevent these. Thus, we compared the ESMValTool results against AutoAssess results from the same model runs. An example of this is shown in Figure 1, and Table 1. Relative differences in metrics between the AutoAssess and ESMValTool implementations, arising for example from differences in regridding algorithms, are less than 1%, which is much smaller than the typical differences in these metrics between models, which are generally of order 10%.

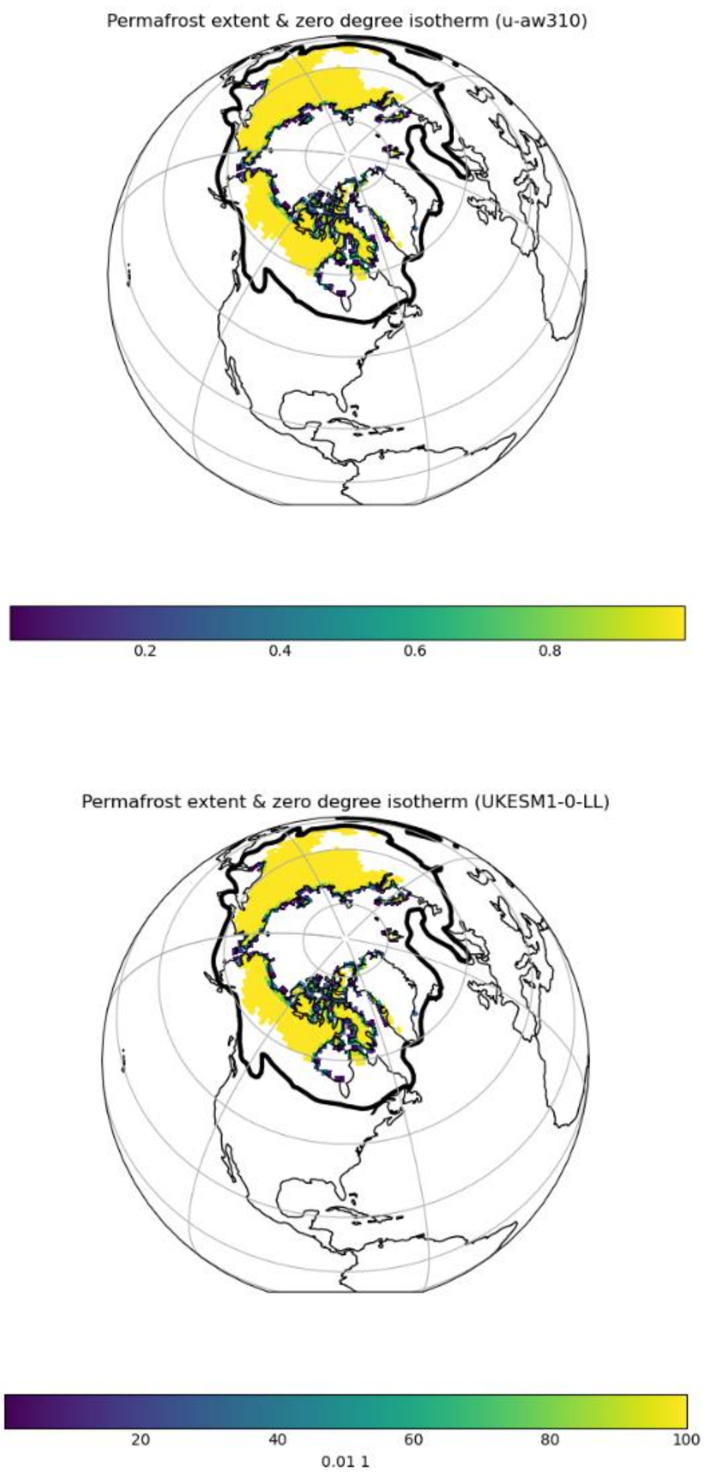


Figure 1: Permafrost extent for UKESM from AutoAssess (left) and ESMValTool (right), using data from the same model simulation in native format (for AutoAssess) and CMIP format (for ESMValTool).

Metric short name	AutoAssess	ESMValTool
attenuation 1m over surface	0.496	0.496
attenuation surface over air	0.493	0.492
fraction area permafrost over zerodeg	0.289	0.290
offset 1m minus surface	0.947	0.947
offset surface minus air	7.71	7.67
permafrost area	13.7	13.5

Table 1: Comparison of permafrost metrics between AutoAssess and ESMValTool, using data from the same model simulation in native format (for AutoAssess) and CMIP format (for ESMValTool).

A key output from AutoAssess is a set of "normalised assessment metric plots". Because model assessment metrics use a wide range of methods and are applied to a wide range of model variables, they span a wide range of magnitudes and differing units, which poses a challenge in visualising multiple metrics in a single image. The approach taken by the Met Office is to normalise each metric for an experiment by the value of that metric in its control simulation. Thus an experiment that only differs slightly from the control will have a value close to 1, and most metrics can typically be plotted in a range between 0 and 2. An example of such a plot, produced by ESMValTool to display the new metrics added in the current work, are shown in figure 2 for two non-Met Office models to demonstrate that the metrics can be applied to any CMIP model. The work in WP9.5 has made these normalised assessment metrics plots available in ESMValTool for 20 new metrics.

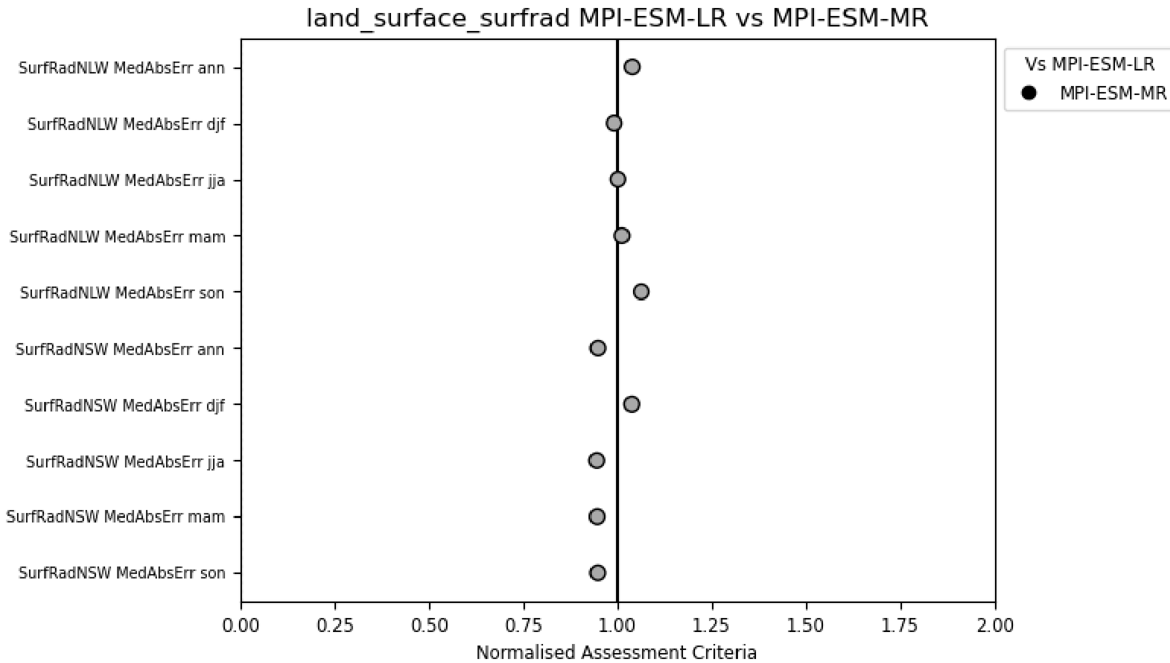


Figure 2: Normalised assessment metrics plots produced by ESMValTool for surface radiation metrics. These metrics are "zero-oriented", meaning that zero would be a perfect score; thus values smaller than 1 indicate that the experiment is an improvement over the control.

An important part of the Met Office development cycle is a phase of assessment by domain experts. Once a candidate release version of the model is defined, an extended set of experiments is performed and the results of these are shared outside the model development teams with scientists in other parts of the Met Office and our partners. These experts use established tools and must be engaged with any major change to the development workflow. We have organised discussions with a representative sample of these experts to understand their requirements and any barriers to change. A clear message from this user engagement is the need for support from RSEs in the more technical aspects of development, particularly management of development environments. We are evaluating a number of options for how this support could work.

ESMValTool has been installed on the Met Office IT estate, and we have checked that the process for developing a new diagnostic is possible on our systems. We have documented the Met Office particulars of this process for use by scientists, making use of the excellent central ESMValTool tutorial wherever possible. However, the local Met Office documentation will need to be revised to reflect whichever decision is taken regarding the development environment options discussed above.

3.2 BSC

ESMValTool has been incorporated at BSC both in and out of the EC-Earth Workflow. This incorporation has been facilitated by the previous adoption of a strategy to CMORise all the output of our own models, even if that meant an extra step in the post-processing chain. This change was done, among other considerations, to facilitate the development of analysis software by supporting only one kind of format.

ESMValTool is now used at BSC in three different ways:

1. As a standalone tool for analysis. In this mode of operation, scientists use ESMValTool within their own scripts for interactive evaluation and analysis of model output data.
2. Integrated with the ecFlow workflow manager, to compute diagnostics in a chunk by chunk basis and store them as new variables after the models have run. This allows users to compute new variables after the model runs in a fast and easy way. Derived variables and metrics computed this way include sea ice extent for each pole (computed from the sea ice concentration), yearly total carbon mass produced in the ocean (computed from the 2D flux) and some regional indices like Niño3, among others.
3. Integrated in the workflow used at BSC to run the experiments, which is orchestrated by Autosubmit. ESMValTool is used to compute diagnostics and to produce near real time plots of the model runs to monitor them (quick-look capability). Example monitoring outputs using ESMValTool are shown in Figure 3. This workflow organises every step of the run: deployment and compilation of the model in the HPC facility; the run of multiple start dates and / or realizations, usually divided in several chunks; post-processing, CMORisation and long-term storage of the outputs; and the first analysis of the results, which is the part that involves ESMValTool.

These three options are used both in development and production experiments, as they are not treated differently at BSC as far as the tools are concerned. In fact, production experiments are usually derived from one of the test cases that we continuously test.

The plots used to monitor the model runs are generated by a single diagnostic directly from pre-processed output that can be configured using a styles file made in the yml configuration format. This diagnostic is mature and nearly ready for merging to the main branch of ESMValTool, but was not quite ready in time for ESMValTool v2.3; it is expected to be available in the next ESMValTool's release (see https://github.com/ESMValGroup/ESMValTool/pull/2184 to check the current status).

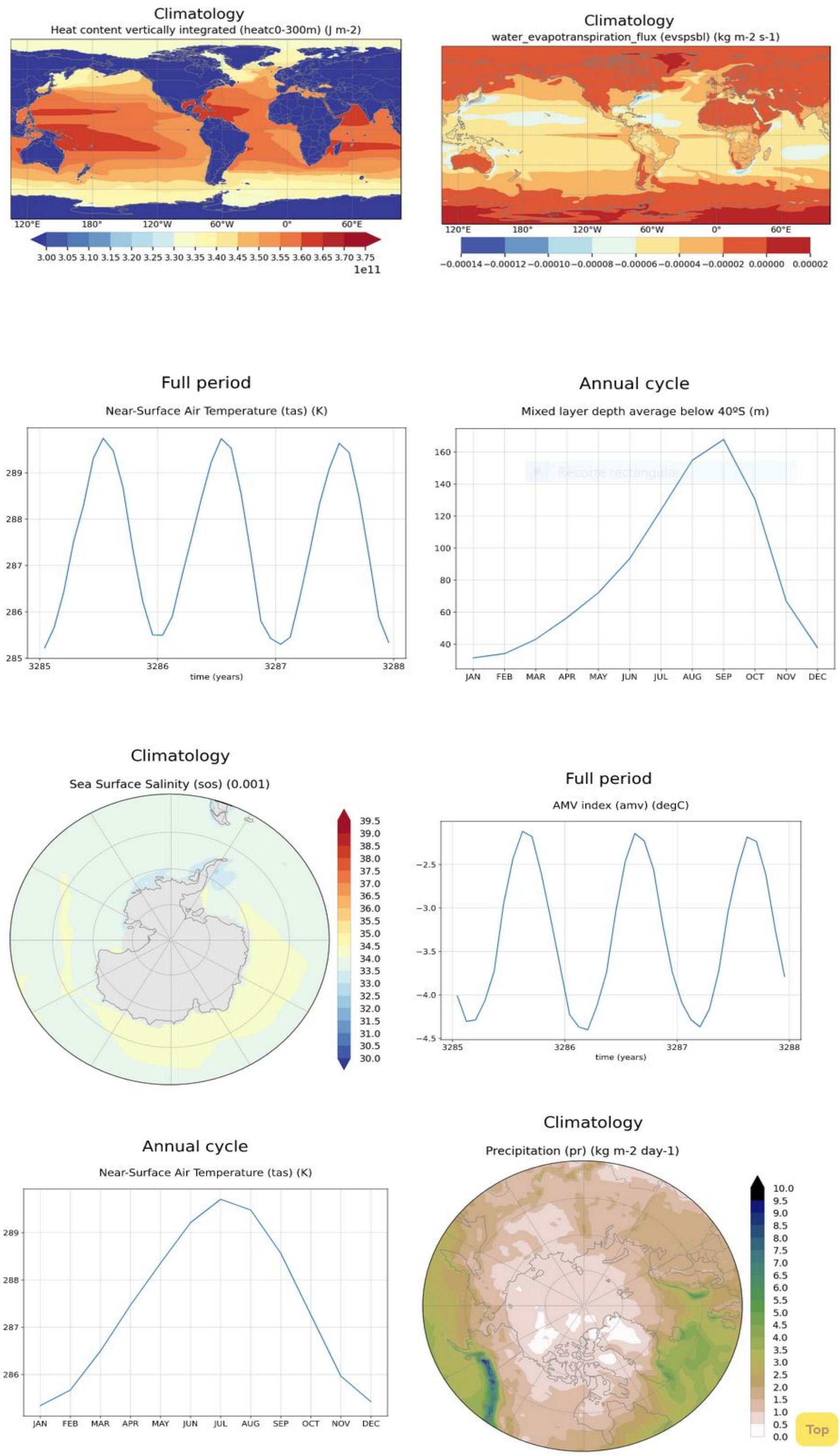


Figure 3: Examples of the output of the BSC monitoring tool for 3-years of an ongoing simulation.

3.3 CNRS-IPSL

We have installed the ESMValTool on the IPSL "meso-center" computing facility, which provides both a large disk capacity, an archive of major climate simulations (CMIP3, CMIP5, CMIP6) and observational datasets, and a computing facility using a cluster. This centre also provides fast access to the file system of the two major HPC centres used for climate simulations, one of which (TGCC) is a Tier-1 European HPC centre. This allows users to access simulation outputs readily for analysis.

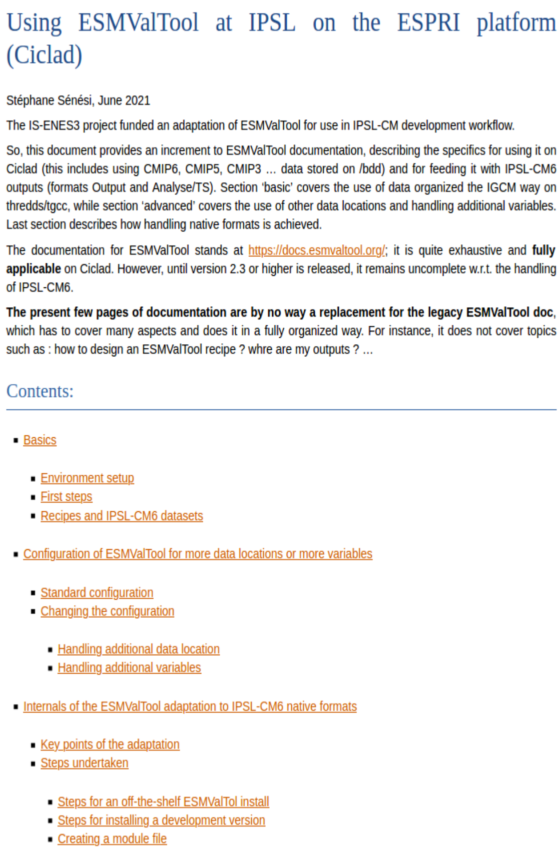
ESMValTool was first configured to access CMIP3, CMIP5, CMIP6, CORDEX and Obs4MIPS data on this centre according to the local data organization, thus directly providing IPSL scientists with the full synergy between ESMValTool diagnostics set and world class datasets.

In order to tackle the inclusion of ESMValTool in the model development workflow, a number of topics have been addressed; all deal with avoiding to re-format IPSL-CM model outputs for feeding the tool, in order to ensure acceptability[1]; first, there was the need for ESMValTool to know about the IPSL-CM variable naming convention, and second to know about the way to identify which data file should be searched for each given variable. These issues were addressed by a close cooperation with the ESMValTool core development team, which developed a generic feature that allows the provision of customized key/value pairs for each variable.

A further topic was the efficiency of Iris, the library that underlies ESMValTool for reading and handling data, in loading data from multi-variable files: the original assumption for ESMValTool is that data files are organized with one variable per file (the CMOR way), while one of the IPSL-CM native output formats groups up to 300 variables per file. This showed that Iris (up to version 3.0.1) is slow in such cases; we had exchanges with the Iris development team at the Met Office on this subject, which proposed a fix to Iris, yet not sufficient for efficient use. Therefore, as a temporary measure, the module added in ESMValTool for reading IPSL-CM data optionally uses an external, IS-ENES supported, utility (CDO) for speeding up the process. Work is still on-going in the Iris team to propose much better performance, but this might require a re-design of the ESMValTool data access API.

The new IPSL-CM module in ESMValTool also includes a number of fixes to data with respect to CMIP conformance. Last, the documentation for the specifics of using ESMValTool at IPSL for simulation analysis has been developed (see figure 4, and [IPSL website]).

[1] Although XIOS allows IPSL-CM to output CMIP formats, IPSL policy remains to produce such outputs only for CMIP contributions, and on top of the so-called ‘native’ format. Therefore, this adaptation of ESMVaTool was necessary



Using ESMValTool at IPSL on the ESPRI platform (Ciclad)

Stéphane Sénési, June 2021

The IS-ENES3 project funded an adaptation of ESMValTool for use in IPSL-CM development workflow.

So, this document provides an increment to ESMValTool documentation, describing the specifics for using it on Ciclad (this includes using CMIP6, CMIP5, CMIP3 ... data stored on /bdd) and for feeding it with IPSL-CM6 outputs (formats Output and Analyse/TS). Section 'basic' covers the use of data organized the IGCM way on thredds/tgcc, while section 'advanced' covers the use of other data locations and handling additional variables. Last section describes how handling native formats is achieved.

The documentation for ESMValTool stands at https://docs.esmvaltool.org/; it is quite exhaustive and **fully applicable** on Ciclad. However, until version 2.3 or higher is released, it remains uncomplete w.r.t. the handling of IPSL-CM6.

The present few pages of documentation are by no way a replacement for the legacy ESMValTool doc, which has to cover many aspects and does it in a fully organized way. For instance, it does not cover topics such as : how to design an ESMValTool recipe ? whre are my outputs ? ...

Contents:

Figure 4: Screenshot of documentation for use of ESMValTool at IPSL.

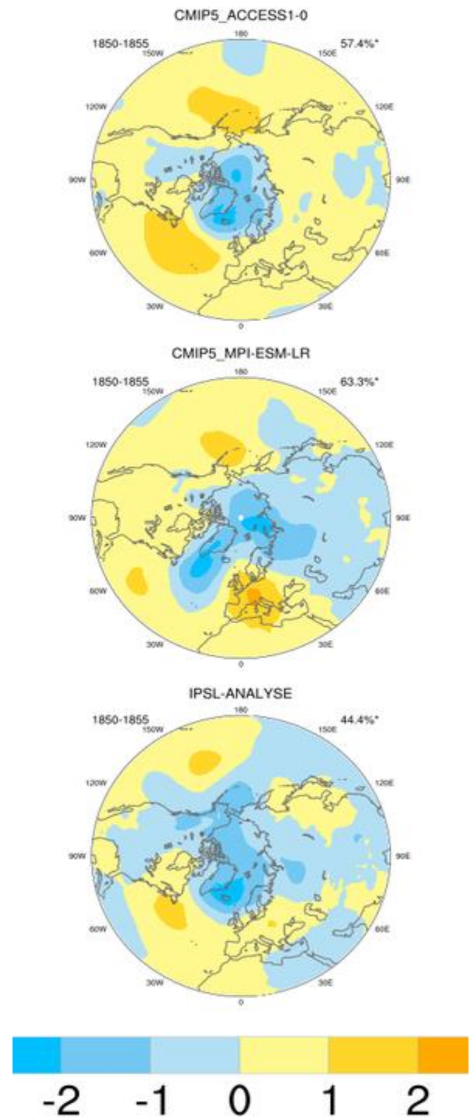


Figure 5: Maps of teleconnection patterns for the North Atlantic Oscillation index for two CMIP5 models using CMIP netCDF input data (top, centre) and an IPSL test experiment using native IPSL model input data.

These changes were included in ESMValTool V2.3, and now provide IPSL scientists with the full capacity of ESMValTool for assessing IPSL-CM results in model development, besides CMIP results, in a fully consistent way. Figure 5 shows a result obtained on IPSL native format with the Climate Variability Diagnostics Package from the ESMValTool.

3.4 Met Norway

ESMValTool has been installed and made available to the NorESM community on the Norwegian National Infrastructure for Research Data (NIRD), where the post-processing for NorESM usually takes place. This infrastructure also carries a 180TB local mirror of many ESGF datasets which is accessible by the ESMValTool.

On 12 March 2021 a Nordic ESM hackathon was organised together with other projects to widen the user base of the ESMValTool and to make the NorESM community aware of the benefits of incorporating the ESMValTool into the standard modelling workflow. This hackathon consisted also of a user session where actual users could try out and modify recipes predefined by pilot NorESM users.

In order to use the ESMValTool for model development, ESMValTool has either to be able to read the native model output or the CMORisation has to be efficient and user friendly. NorESM has chosen the CMORisation approach because NorESM is based on CESM and we judged the effort to write a native CESM reader too large for the resources in this task. We therefore invested effort in making the NorESM CMORisation tool, which had been hard for scientists to employ, more user friendly and to build the infrastructure to make CMORisation also possible during run time of the model. The improved code to perform this conversion is openly available via github: https://github.com/NorESMhub/noresm2cmor.
In addition, a user workshop for noresm2cmor took place in June 2021 to educate the users. We expect that these developments and user engagement will enable future projects to make wider use of ESMValTool for NorESM evaluation.

3.5 SMHI

ESMValTool was installed on two HPC clusters at Sweden's National Supercomputer Centre (NSC). Here, it was set up to use the Rossby Centre Data Library, a partial replica of CMIP5, CMIP6, and CORDEX data as distributed via the ESGF.

These HPC clusters also serve as the principal computing resources for SMHI's climate modelling effort, providing the capacity used for model runs. Thus, the ESMValTool installation permits the analysis of model runs right at the source and as part of the model development process.

4. Conclusions and Recommendations

Five European modelling centres have made progress towards incorporating ESMValTool into their model development workflows. Changes have been made to ESMValTool in order to facilitate this, and all but one of these changes have been made available in v2.3 of ESMValTool, which was released on 27th July 2021 (the remaining change will be made available in the next release). For some centres these changes leveraged underpinning infrastructure enhancements that have been made to ESMValTool's core, and these enhancements are also included in v2.3.

The modelling infrastructures differ across these centres, and as a result we have differing strategies for how ESMValTool will ultimately be used in our systems. For some centres, for example the Met Office and Met Norway, the full implementation of these strategies will take a number of years and a large investment of effort, while for others such as SMHI, BSC and IPSL this journey is more complete. In all cases, the journey has been accelerated by the work reported here, which ultimately will unlock the great potential of the many ESMValTool recipes which have been developed by the international Earth system modelling community, to enhance the scientific scope of our model development evaluation. No significant increases in computational cost have been noted by the modelling centres.

Immediate next steps for these centres include continuing engagement with users, provision and enhancements of the process and guidance for ESMValTool development. We will share any outcomes from further user engagement, particularly where it relates to technical support for evaluation scientists. For the Met Office, the next steps additionally will include the porting to ESMValTool of further AutoAssess metrics, and adaptation of our CMORisation tool. One clear recommendation from the IPSL implementation is the need to speed up the loading of data from multi-variable files, and while a workaround has been implemented, discussions will continue with the Iris and ESMValTool development teams on a long-term solution.

In the current implementations, centres have retained their existing web interfaces and used ESMValTool to generate the plots that are displayed by these web interfaces. ESMValTool itself does not yet include a web interface for interactively browsing a large collection of plots; this could be considered for future development and it would be very useful for centres that do not have such a system. However, it would be a major undertaking to design it in a generic way that performed well for the needs of multiple centres.

References

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., Arnone, E., Bellprat, O., Brötz, B., Caron, L.-P., Carvalhais, N., Cionni, I., Cortesi, N., Crezee, B., Davin, E. L., Davini, P., Debeire, K., de Mora, L., Deser, C., Docquier, D., Earnshaw, P., Ehbrecht, C., Gier, B. K., Gonzalez-Reviriego, N., Goodman, P., Hagemann, S., Hardiman, S., Hassler, B., Hunter, A., Kadow, C., Kindermann, S., Koirala, S., Koldunov, N., Lejeune, Q., Lembo, V., Lovato, T., Lucarini, V., Massonnet, F., Müller, B., Pandde, A., Pérez-Zanón, N., Phillips, A., Predoi, V., Russell, J., Sellar, A., Serva, F., Stacke, T., Swaminathan, R., Torralba, V., Vegas-Regidor, J., von Hardenberg, J., Weigel, K., and Zimmermann, K.: Earth System Model Evaluation Tool (ESMValTool) v2.0 – an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP, Geosci. Model Dev., 13, 3383–3438, https://doi.org/10.5194/gmd-13-3383-2020, 2020.

Abbreviations

BSC: Barcelona Supercomputing Centro
CMIP: Climate Model Intercomparison Project
CNRS: Le Centre National de la Recherche Scientifique (National Centre for Scientific Research)
CORDEX: Coordinated Regional Downscaling EXperiment
ESMValTool: Earth System Model eValuation Tool
Iris: Python package for data analysis which underpins ESMValTool. https://github.com/SciTools/iris
GRIB: General Regularly-distributed Information in Binary, a binary data format
HPC: High Performance Computing
netCDF: Network Common Data Form, a binary data format
SMHI: Sveriges Meteorologiska och Hydrologiska Institut (Swedish Meteorological and Hydrological Institute)