Model evaluation expectations of European ESM communities: first results from a survey

Jérôme Servonnat, Zofia Stott, Klaus Zimmermann, Remi Kazeroni, Eric Guilyardi, Kim Serradell, Axel Lauer, Fanny Adloff, Marie-Claire Greening, and Javier Vegas

IS-ENES3 WP3 T4

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

THE CONSORTIUM

Coordinated by **CNRS-IPSL**, the IS-ENES3 project gathers **22 partners** in **11 countries**

cnrs with cea Sorbonne Université

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Met Office

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

BSC

CERFACS

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

SMHI

UK Research and Innovation

UC UNIVERSIDAD DE CANTABRIA

DLR

netherlands eScience

Norwegian Meteorological Institute

METEO FRANCE
Toujours un temps d'avance

MANCHESTER 1824
The University of Manchester

ΔΗΜΟΚΡΙΤΟΣ

WAGENINGEN
UNIVERSITY & RESEARCH

CHARLES UNIVERSITY

NORCE

LINKÖPING UNIVERSITY

IS-ENES3 WP3 T4

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°824084

Our website
https://is.enes.org/

Follow us on Twitter !
@ISENES_RI

Contact us at
is-enes@ipsl.fr

Subscribe to the **IS-ENES3 H2020** Youtube channel !

Context: climate model evaluation tools

ESMValTool
Earth System Model Evaluation Tool

… to coordinated development
Since 2010s

CMEC
PMP

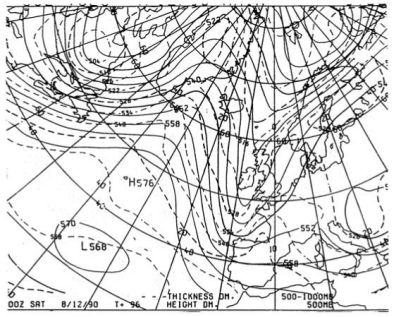
Standardization 2000s-2010s

ESGF
Earth System Grid Federation

(outputs, experiments, file format, metadata)

From individual efforts…

1970s-1990s



Cullen (1993)

Context: community tool vs in-house solution

- Community
  - Reduce duplication of effort on often repeated tasks
  - Promote standardisation and enable meaningful cross comparisons, eg between ESMs, ESMs and data
  - Critical mass to create a support and collaboration community in institutions and between institutions ➔ sustainability of the tool
  - More efficient use of resources (funding, staff)

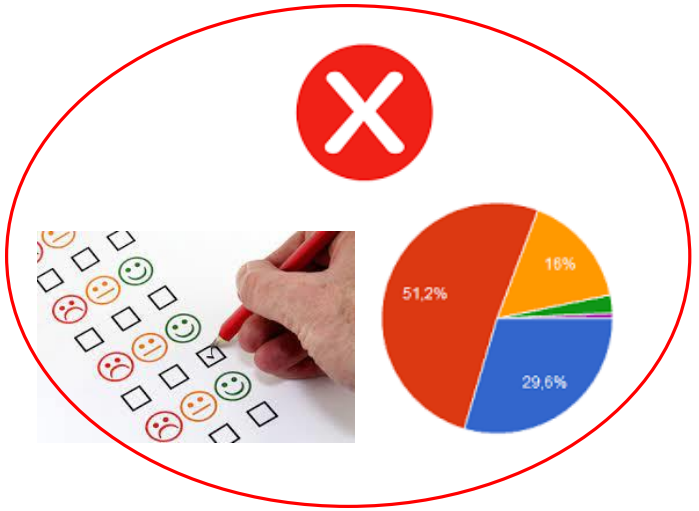
- In-house
  - Freedom: Scientists like to do things their own way and need to be convinced to use "off-the-shelf" tools
  - Heritage: Force of habit and previous investment prevent convergence on common tools

The roots of the survey

- In 2018 (writing of the IS-ENES3 project), different evaluation tools were moving forward as potential solutions for the modelling groups
- One front runner supported by the European community with IS-ENES3: the ESMValTool
- Good practice: if we want to develop tools for the community, give a chance to the potential users to say what they need
- Why a survey?
  - Going beyond the lobby discussion
  - A context to favour freedom of speech
  - Possibility to understand peoples choices

ASSIMILA

Survey: the process

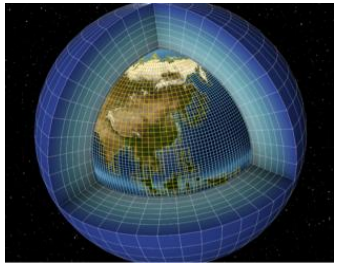


- An investigation
  - Open-ended questions ➔ interest/experience of the interviewees are paramount
  - Not a 'box ticking' survey with final statistics on items
  - Not designed to be a representative statistical sample with a clear answer

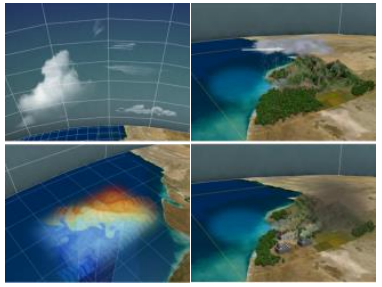
- The interpretation is partly subjective: hard to keep an un-biased eye on the content of the interviews

Who was interviewed?

Modellers



Process scientist



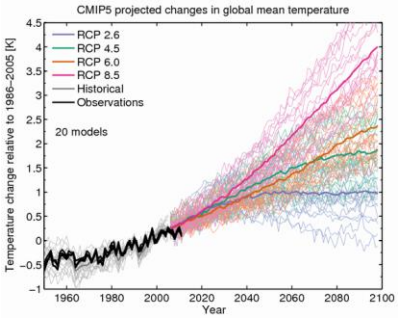
Software developers

ESMValTool
Earth System Model Evaluation Tool

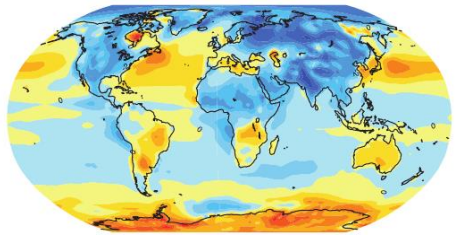
CliMAF
Sharing – Simplifying – Optimizing

NCAR UCAR | CGD's Climate Analysis Section
Climate Variability Diagnostics Package

Scientists /analysts



Evaluation



Data scientists

ESGF
Earth System Grid Federation

Impact community

- Senior scientists/professors, postdocs, industry representatives
- Working scientists with direct hands-on interests and those with strategic interests
- France, Germany, Spain, Sweden, UK, USA

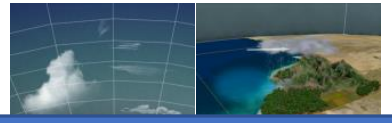
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

ASSIMILA

Who was interviewed?

Modellers

Process scientist

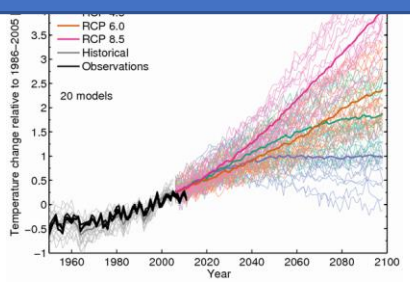


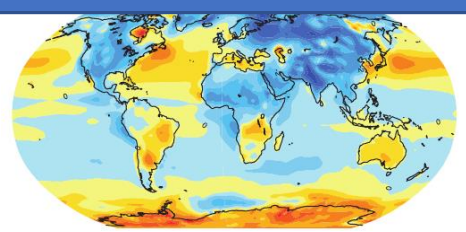
Software developers

ESMValTool

CliMAF

- 41 requests for interviews issued by Steering Group and Assimila
- 20 interviews held
- + 5 email exchanges





ESGF
Earth System Grid Federation

- Senior scientists/professors, postdocs, industry representatives
- Working scientists with direct hands-on interests and those with strategic interests
- France, Germany, Spain, Sweden, UK, USA

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

ASSIMILA

Tools reported by the interviewees

- Study showed a range of “ESM evaluation tools” in use, including:

Diagnostic packages and frameworks

Auto-assess

PMP

CliMAF
Sharing – Simplifying – Optimizing

ESMValTool
Earth System Model Evaluation Tool

Model Diagnostics Task Force
Diagnostic Package

freva

Data operators:

- Climate Data Operator (CDO) (Max Planck Institute), NCO (NetCDF Climate operators), Ncl (NCAR), Python libraries: Xarray, Iris

Consulting existing results

- “Toolboxes” in ESA’s Climate Change Initiative and Copernicus Climate Change Service

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

ASSIMILA

Results: ways to use a community evaluation tool

- Take it as it is: standard model evaluation, benchmarking – model development, collection of IPCC diagnostics ➔ minimum technical effort to use the scientific content
- Adopt the tool and extend its use for evaluation: science of model evaluation, process studies ➔ dive deeper in the technical aspects
- Take the core and use it do your own science ➔ adopt the technical solution (“wish I had this during my PhD”)
- Consulting the results – impact studies – decision making ➔ visualization / access to the results

Results: scientific content – user needs

- From context specific (general, ENSO, etc.) certified core set of diagnostics…
  - Providing "standard" evaluations at end of model runs to provide mark of quality
  - Approved diagnostics for specific questions

- … to a rich collection of diagnostics = pick the ones you like
  - Available diversity + possibility to choose
  - Accept you may need to adapt or (re)write code to get the exact diagnostic/plot you need
  - don't let community tools become a dumping ground for everyone's favourite diagnostic

➔ who decides what goes in? Governance / science of model evaluation

Results: identified technical user needs

Flexible: tuned/tunable to wide range of scientific needs

- Model development: Standard diagnostics to compare different versions of a particular model and against observations
- Model analysis, process studies : Tailored, complex diagnostics – for publications

Provide technical solutions

- Finding the data: stop need for data wrangling (model intercomparison)
- Core pre-processing functions
- Growing data volumes (becoming problematic for evaluation): High temporal and spatial resolution simulations
- Use different grids

Efficient execution

- Typically mins/~day

Interoperable with other tools

- I'm not stuck with one tool <-> ways to connect the tools
- Generic/standard (code and output)

Results: experience and trust

Experience

- Good documentation, support, training
  - Make it easy for any particular user to find/get what they want
  - Local support for community tools valued by those who have it
- Transparent and traceable: no “black boxes”, provenance of information easy to track
- Easy to get first result: typically hours/~day to get first plot
- Providing GUIs (cf IS-ENES 3 plans), APIs, click and play, toolbox – not just command line

Trust

- Reliable, tested: certified
- Sustainability: maintained and developed => governance

Feedbacks and implications for ESMValTool

- ESMValTool held in high regard by those interviewed

- "Flexible: tuned/tunable to wide range of scientific needs"
  - Significant progress in making the tool accessible and user friendly over the last few years

- "Good documentation, support, training"
  - Github repo + documentation + training sessions
  - Communication: avoid misunderstanding
  - success stories should be generated to convince the research community

- "Efficient and easy to use"
  - Most recipes run between minutes to hours
  - A couple hours to get first result (if you use conda)

Additional thoughts on community tools

- Need to recognize/acknowledge that we are in a continuous development = this is a journey (and a long one), not a destination!

- There has to be sufficient determination from the user ('stubbornness') to get his/her hands on the tool and overcome the errors: do not give up!

  ➔ Getting the scientists and software developers together!

Take-home points

Community tool

- Range of use cases : from community approved scientific evaluation packages to individual usage
- Technical solution to efficiently serve that range
- Buy in and trust: documentation & traceability, sustainability & governance – long term investment

Discussion is just starting!

Thank you for your attention

Subjects for the long term

- Areas for improvement
  - Machine learning and AI for ESM evaluation
  - On the fly post-processing and diagnostics while the model is running (cf developments at BSC)
  - Geographically distributed computing to eg reduce burden of data transfer
  - Address the data volume problem, eg greater use of cloud computing

About the interviews

- All notes have been checked by the interviewees.
- The notes are confidential to the project review panel and have been made available by the interviewees on the understanding that they will not be circulated further.
- The purpose of the notes is to review the findings of the study and identify key points for the final report.

Notes on governance

- Open source
- Transparent community
- More members, more active/implicated, more animation
- Governance is about the processes, not about the software development. There has to be a strong understanding of why you are doing something and how it will be financially maintained.

Survey: the process

- List of interviewees provided by the steering commitee
- One hour interviews
- A core set of open ended subjects/questions - keeping a lot of freedom in the structure of the interview to favour the expression of opinions and ideas based on interviewee experience and priorities

The questions / subjects

- Context: job description of interviewee
- What are the challenges of model evaluation tools?
- Which model evaluation tool do you use, or contribute to (eg as developer, or data provider)?
  - What is your experience with the ESMValTool?
- Explain how you work (workflow)? How are tools used today? Pros and cons of different approaches
- Most important aspects of these evaluation tools
- What future developments for the tools / challenges?
- Governance of a community tool with particular reference to ESMValTool?
- Other points arising?