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

ASSIMILA

Context: climate model evaluation tools

ESMValTool
Earth System Model Evaluation Tool

… to coordinated development
Since 2010s

CMEC
PMP

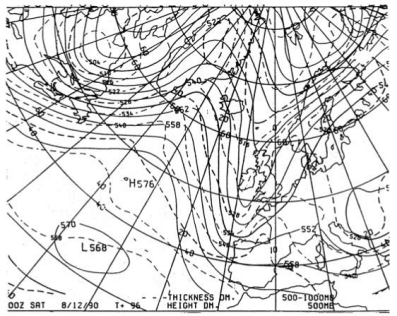
Standardization 2000s-2010s

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

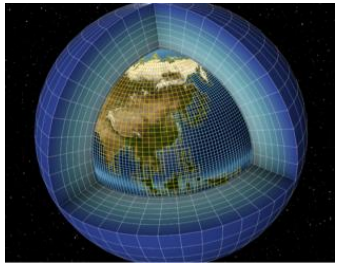
- In 2018 (writing of the IS-ENES3 project), different evaluation tools were moving forward as potential solutions for the modelling groups
- One « front runner » supported by the European community with IS-ENES3: the ESMValTool
- Good practice: if we want to develop tools for the community, give a chance 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

Survey: the process

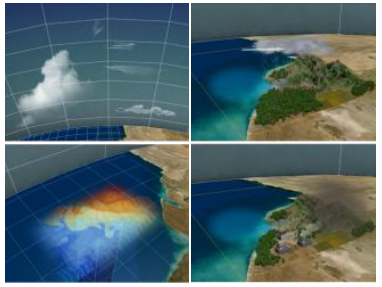
- An investigation
  - Open-ended questions ➔ interest/experience of the interviewees are paramount
  - Not a ‘box ticking’ survey with final statistics on items
  - Not designed to be a representative statistical sample with a clear answer

Who was interviewed?

Modellers



Process scientist



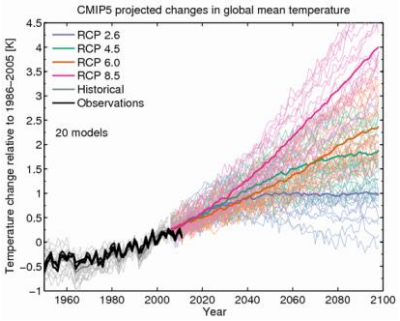
Software developers

ESMValTool
Earth System Model Evaluation Tool

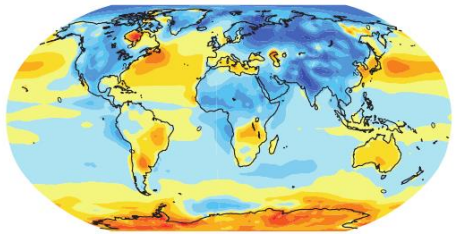
CliMAF
Sharing – Simplifying – Optimizing

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

Scientists /analysts



Evaluation



Data scientists

ESGF
Earth System Grid Federation

Impact community

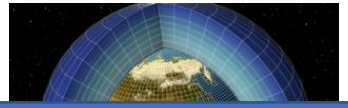
- Senior scientists/professors, postdocs, industry representatives
- Working scientists with direct hands-on interests and those with strategic interests
- France, Germany, Spain, Sweden, UK, USA

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

ASSIMILA

Who was interviewed?

Modellers



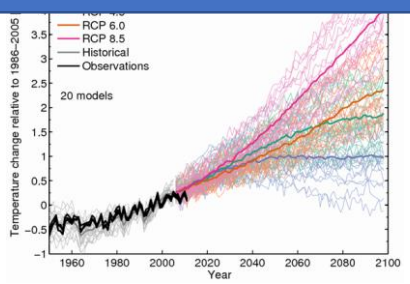
Process scientist

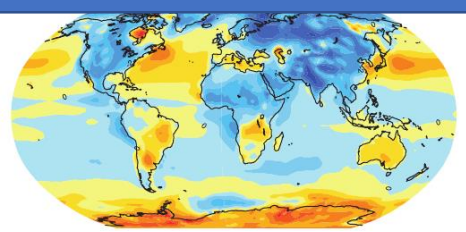
Software developers

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
  - 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

Thank you for your attention

ASSIMILA