Context

- Study commissioned by IS-ENES3: *Initial requirements on model evaluation*

- Survey of broad section of climate community
  - ~30 study participants from
    - Climate scientists/evaluation tool developers (interviews): BSC, DKRZ, DLR, IPSL, Met Office, NCAR, SMHI, STFC, University of Reading (NCAS, NCEO)
    - Commercial/impact groups (email): Climate Adaptation Services, Netherlands; OceanNext, France; Climate Analytics, Germany

Types of use

- "Standardised" model intercomparison diagnostics and metrics, particularly in the context of international developments such as CMIP, CORDEX and for IPCC (figures for chapters scientists..)
- Model development and benchmarking, particularly comparing successive model versions and comparing a model with data
  - Tailored diagnostics – "getting under the bonnet of the model"
- Complex diagnostics from tailored model runs for process science
- Climate impact community – early days for using model evaluation tools

High level requirements – why use community evaluation tools

- **Reduced duplication** of effort on often repeated tasks
- **Promotion of standardisation** and hence enabling meaningful cross comparisons, eg between ESMs, ESMs and data
  - CMIP has provided important impetus for progress in Earth System Models and their intercomparison by promoting standardisation: experiments, formats, naming conventions, diagnostics and metrics. In turn this has promoted development and use of evaluation tools.
- **Critical mass** to create a support and collaboration community in institutions and between institutions
- **More efficient use of resources** (funding, staff)

But…

- **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

Requirements for community evaluation tools

- **Flexibility:** tuned/tunable to wide range of scientific needs
- **Easy for any particular user to find/get what they want**. Focus on the "user experience"
- **Efficient and easy to use**: comparing set up time and run time of existing tools versus developing own tools
- **Good documentation, training and support**
- **Transparent and traceable**: no "black boxes", provenance of information easy to track
- **Reliable, tested:** certified
- Confident that it will be **maintained and developed**
- **Solution to growing data volumes** (becoming problematic for evaluation): High temporal and spatial resolution simulations
- **Interoperable with other tools – Python** and other languages
- **GUIs, APIs, click and play, toolbox** – not just command line
- **Open source**
- **Ability to pick and choose, adapt and contribute**

More/some effort needed?

- Assessment metrics for model variability on various timescales (eg MJO, ENSO, NAO…)
- Analysis of extremes
- Analysis of model biases
- Which models – not just CMIP, eg Coupled Chemistry Intercomparison Project, more observational data
- Evaluation of regional/local models, downscaling (not just global)
- High temporal and spatial resolution simulations (eg precipitation was mentioned by several interviewees where spatial and temporal granularity was important to advise impact communities)
- Not just means – higher order moments
- Catering for new/different grids
- Machine learning and AI for ESM evaluation
- On the fly post-processing and diagnostics while the model is running
- Providing "standard" evaluations at end of model runs to provide mark of quality
- Making tools usable by different communities, eg impacts, adaptation, paleoclimate…but with caution. There was concern that the complexity of ESM diagnostics and metrics could easily mislead
- Fast cycles of releases: 2-3 minor releases per year to keep evaluation software up to date
- Improved user engagement, communications, training, tutorials, help desks, promoting success stories. At the most basic level many potential users of existing tools are not aware of their availability

Closing remarks

- Keep a close eye on CMIP 7 developments and requirements

- *"I wish ESMValTool had been available, at least in its current form when I was doing my PhD – would have saved a lot of time – got results in half the time."* Study interviewee

Thank you

zof.stott@assimila.eu