



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



PRIMAVERA and ESMValTool

**Making climate metrics
available to everyone**

Javier Vegas-Regidor

27/11/2017

ECMWF

The issue

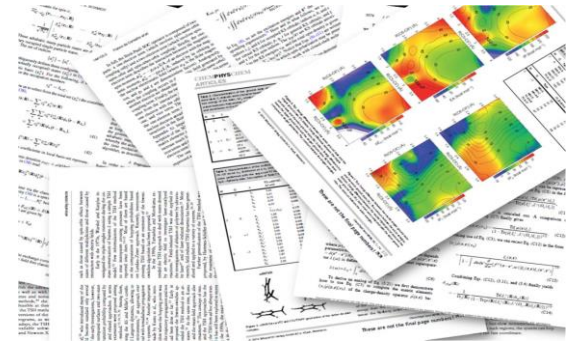


**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

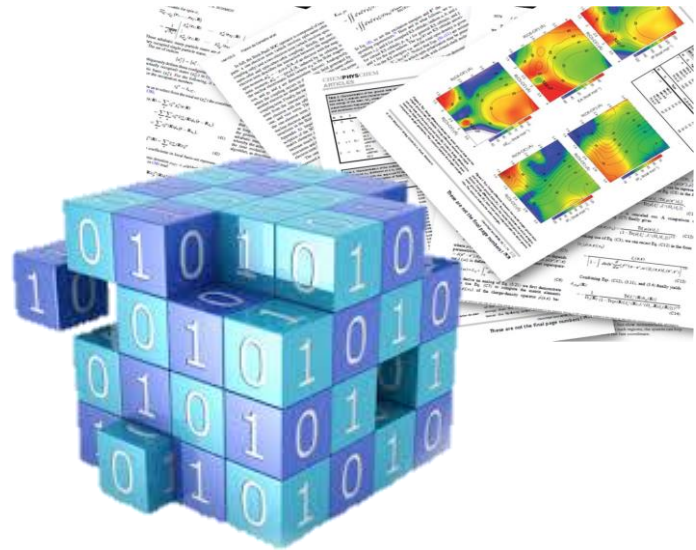
Scientific project's life cycle

- Steps:
 - Generate data
 - Analyse data
 - Publish results
 - Get a new project and repeat
- What the community receives:
 - Papers
- Reproducibility
 - What are you talking me about?



Improving slightly

- Steps:
 - Generate data
 - Analyse data
 - Publish results
 - Publish data
 - Get a new project and repeat
- What the community receives:
 - Papers
 - Data
- Reproducibility
 - You will need to program again the metrics, but at least its possible



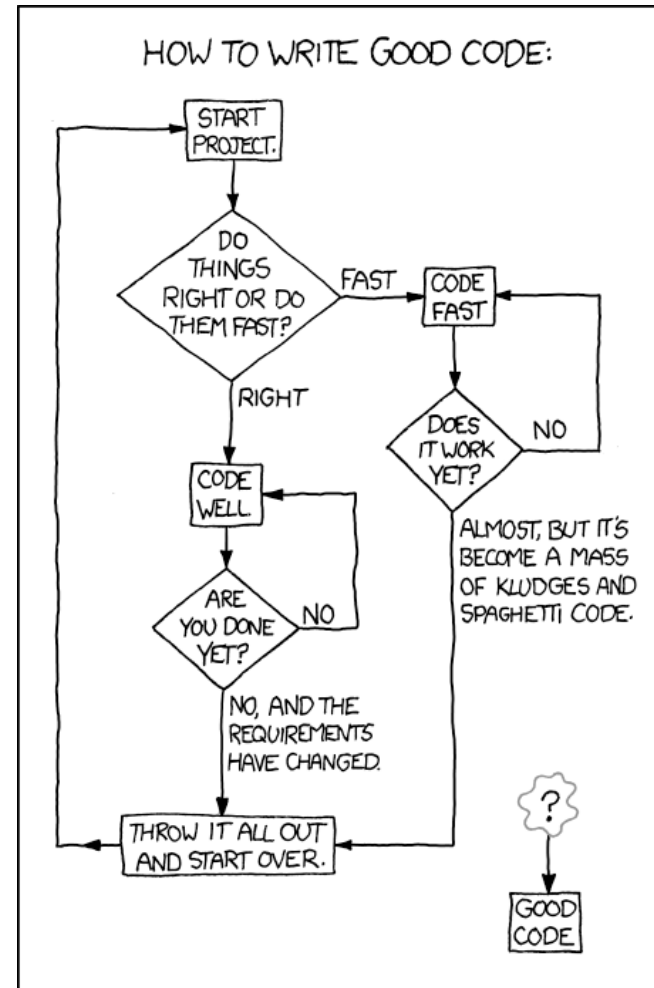
Even better

- Steps:
 - Generate data
 - Analyse data
 - Publish results
 - Publish data
 - Publish code
 - Get a new project and repeat
- What the community receives:
 - Papers
 - Data
 - Code
- Reproducibility
 - It should be straightforward



It is easy to do bad code sharing

- Poor documentation
- No maintenance
- Works only for a specific model
- Works only for a non-standard data format
- Difficult to configure
- And, over all, Climate scientists are not computer scientists: they know how to code but not how to code well



<https://xkcd.com/844/>

PRIMAVERA's vision



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

What is PRIMAVERA?

- PRIMAVERA is a Horizon 2020 project funded by European Commission. The project is a collaboration between 19 European partners, led by the Met Office and the Reading University in the UK.
- Aims to develop a new generation of advanced and well-evaluated high-resolution global climate models, capable of simulating and predicting regional climate with unprecedented fidelity, for the benefit of governments, business and society in general.

PRIMAVERA metrics

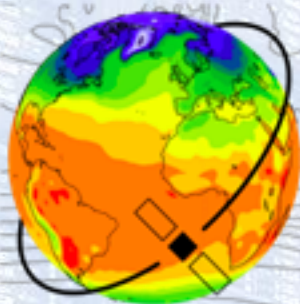
- As part of the PRIMAVERA project, we are going to develop a number of metrics to analyse the results
- They will have to work with the huge high resolution output we are going to get from our model runs (we are talking about several petabytes)
- This is too big an effort to waste it: PRIMAVERA metrics should survive the project and benefit the whole community in the long term

How to reach our vision (1/2)

- Avoiding all the flaws we talked about earlier
- How? Make it a two step process
 - Climate scientists develop the metrics
 - Software engineers enhance and correct them
- Climate scientists only worry about getting the metrics scientifically correct
- Software engineers get them working with all the models and ensure they have a good enough performance
- You can also use this process to try to teach some good practices to the climate scientists

How to reach our vision (2/2)

- Find a metrics package to integrate our metrics
- It must be:
 - Open source
 - Based on standards (specially data standards are crucial)
 - Open to collaboration
 - Maintained in the long run
 - Able to integrate metrics implemented with different languages and libraries
 - Respectful with our scientists for their work by giving them the credit they deserve
- We looked for alternatives and choose the best we found



ESMValTool

Earth System Model eValuation Tool



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

What is ESMValTool

- The Earth System Model eValuation Tool (ESMValTool) is a community diagnostics and performance metrics tool for the evaluation of Earth System Models (ESMs) that allows for routine comparison of single or multiple models, either against predecessor versions or against observations.
- The ESMValTool is a community effort open to both users and developers encouraging open exchange of diagnostic source code and evaluation results from the CMIP ensemble.

Key features

- Standardized model evaluation can be performed against observations, against other models or to compare different versions of the same model.
- Wide scope: includes many diagnostics and performance metrics covering different aspects of the Earth System (dynamics, radiation, clouds, carbon cycle, chemistry, aerosol, sea-ice, etc.) and their interactions.
- High flexibility: new diagnostics and more observational data can be easily added.
- Multi-language support: Python, NCL, R...
- CF/CMOR compliant: data from many different projects can be handled.

Key facts

- Supported by many projects: H2020 Crescendo and APPLICATE, CMIP6-Dicad, C3S-MAGIC
- Even better: most of the partners involved in its development see it as a long-term strategic asset
- Going to be installed at the ESGF nodes for CMIP6 and used to perform routine evaluations on the data as it is uploaded
- Continuous improvements in all areas:
 - Preprocessing
 - Automatized data download
 - Data provenance
 - And many more

Our journey together



**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

ESMValTool development

- We have joined the core development team
- We are part of the team that is creating the new preprocessor engine
- In particular, I am in charge of the CMOR checking and automatic data fixes for the new preprocessor...
- ... but I also help with other issues as needed

Metrics development

- Climate scientists are developing their metrics in Python or R
- We are not prescribing packages to them
- When metrics are ready and running for a subset of PRIMAVERA's data, they submit them to the project SVN repository
- At that point my job start: I have to test that the metrics actually run
- After that I integrate the metrics into ESMValTool

Integrating the metrics

- The first step is just to make it run within the framework but over the same data
- Next to clean the metrics as much as possible
 - Improve performance
 - Improve readability of the code
- Run it for all the other sample data we have for the project
 - This usually implies generalization, it will rarely run as it is
- Extract as much parameters as you can to the configurable section
- Final optimizations and cleaning

Results so far

- New preprocessor is now working
 - About 10x faster than the previous one
 - We are still improving it
- A couple of metrics have been added to the 2.0 version
 - Sea-ice drift strength: python-based, a couple of numbers per model as a result
 - Sea-ice edge: R-based, results are map plots

What's next

- A test set of PRIMAVERA metrics will be integrated into ESMValTool 2.0
- They will be tested against the high resolution output coming from the project's model runs
- I expect this to expose some limitations of the current framework, which we will need to address

Bonus track: some lessons from the metrics



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

Just some unrelated facts

- Scientists are going to use whatever they feel more comfortable with
- If you want your package to be widely use by the community you need two things:
 - A remarkable feature to make them willing to test it
 - An easy way to integrate it with their legacy code
- Most ocean models have irregular grids, but most packages have trouble with irregular grids.
- We need to be respectful with metadata but not too much



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación



**EXCELENCIA
SEVERO
OCHOA**

Thank you

Javier.vegas@bsc.es