HANAMI

Scalar metrics for evaluating the scientific skill of Earth System Models

Marta Alerany, Kai Keller, Leo Arriola, Mario Acosta, Hisashi Yashiro, Chihiro Kodama

Japan Geoscience Union Meeting 2025 - Chiba (Japan)

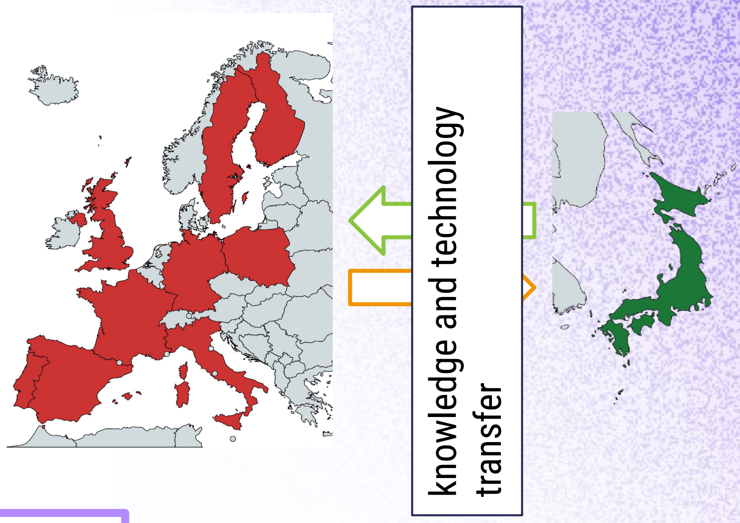
Funded by the European Union

EuroHPC Joint Undertaking

This project received funding from the European High Performance Computing Joint Undertaking (EuroHPC JU) under the European Union's Horizon Europe framework program for research and innovation and Grant Agreement No. 101136269. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or EuroHPC Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

HANAMI

Hpc AlliaNce for Applications and supercoMputing Innovation: the Europe - Japan collaboration

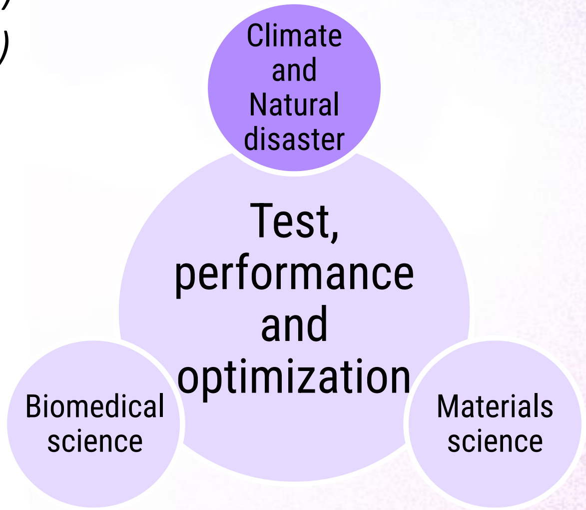


Three Scientific Pillars:

- *WP4 - Climate and natural disaster*
- *WP5 - Biomedical science*
- *WP6 - Materials science*

WP4 has three projects:

- *Earth-system-model performance assessment*
- *Earth-system-model benchmark suite (HPCW)*
- *Large-eddy cloud simulation model (UWLCM)*

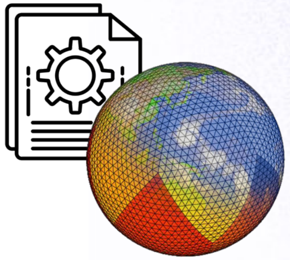


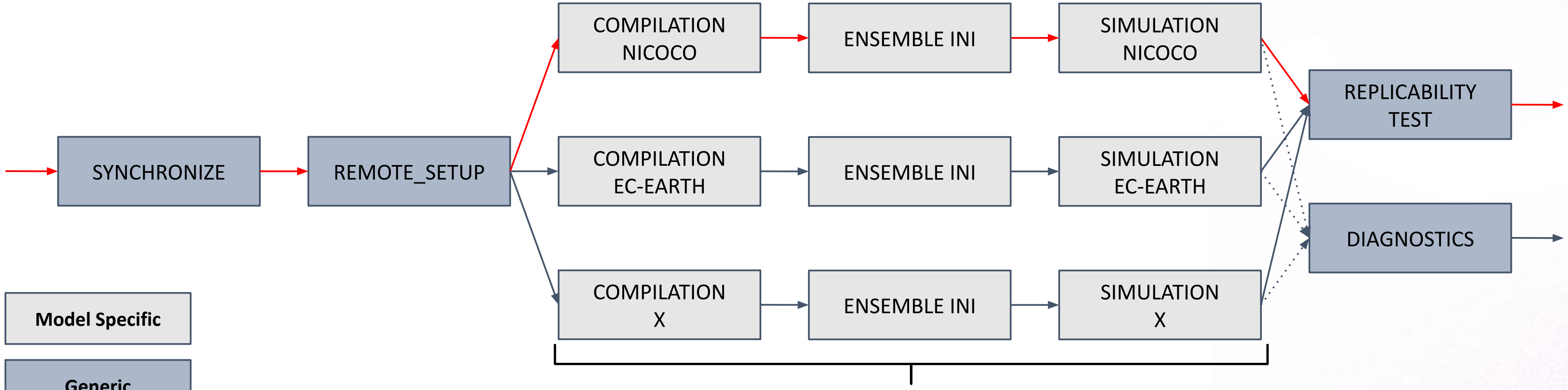
Generic workflow

Modular and flexible

Multiplatform support

High degree of data traceability





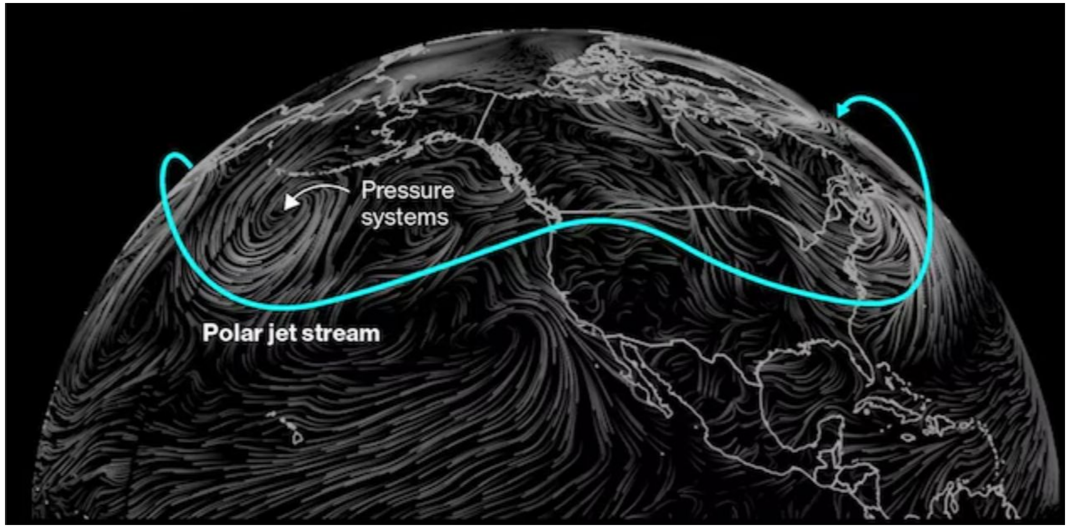
Model specific scripts in git submodules

GitLab

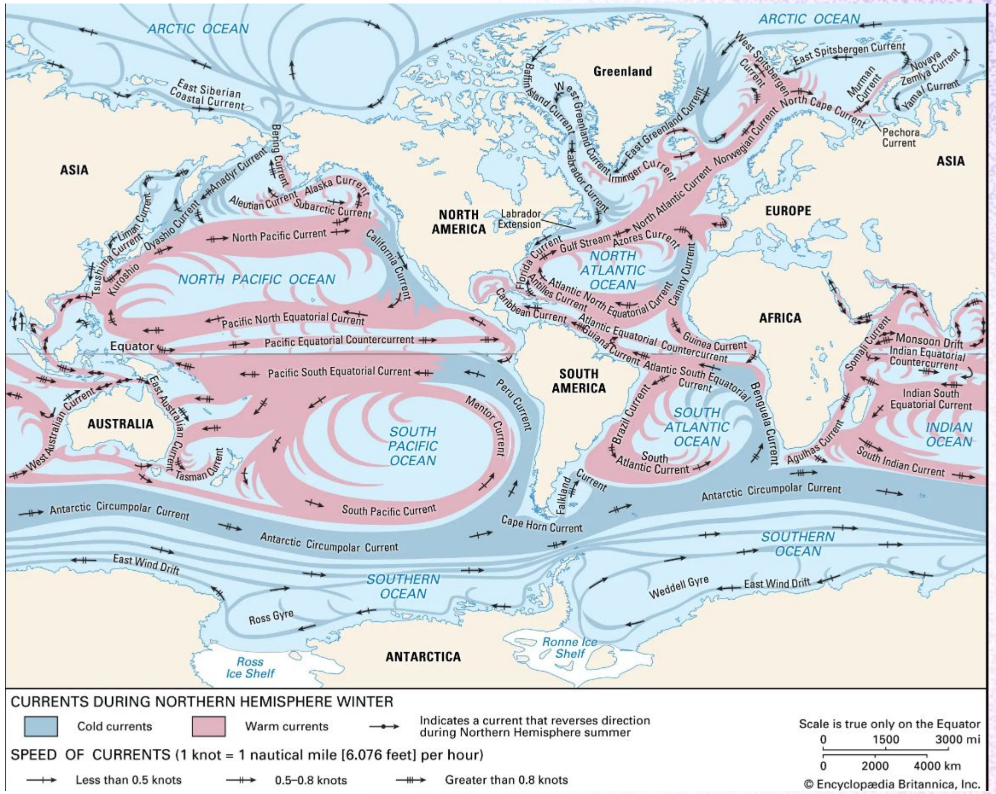
What is Earth System Models (ESMs) "skill"?

Ability to accurately predict various aspects of the climate system:

- Simulate observed **climate patterns**
- Capture complex **interactions** and **variability**
- Forecast future **climate changes**



[1]



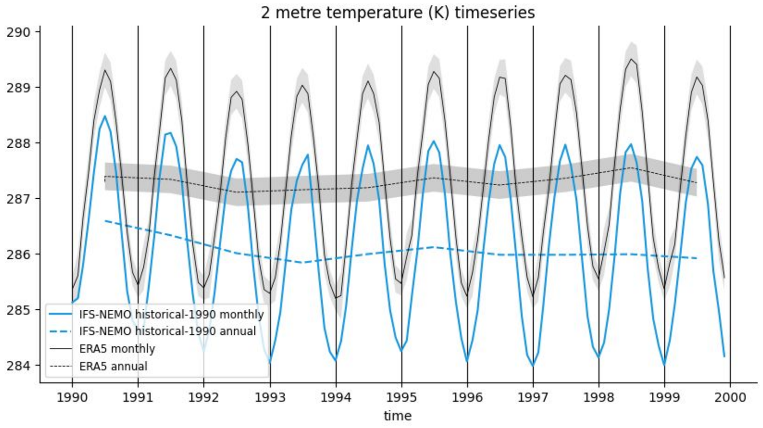
[2]

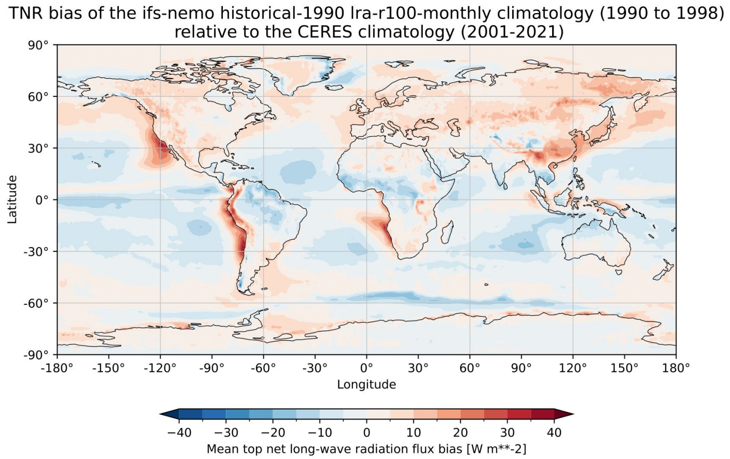
[1] L. Rosenthal & B.K. Sullivan (2024): *Severe flooding across globe linked to climate-fueled rain storms*
[2] Encyclopædia Britannica (Accessed: May 2025): *Major ocean current systems of the world*

Why scientific skill matters?

Current metrics

Commonly used: **spatial/visual metrics** → susceptible to subjective or biased interpretation



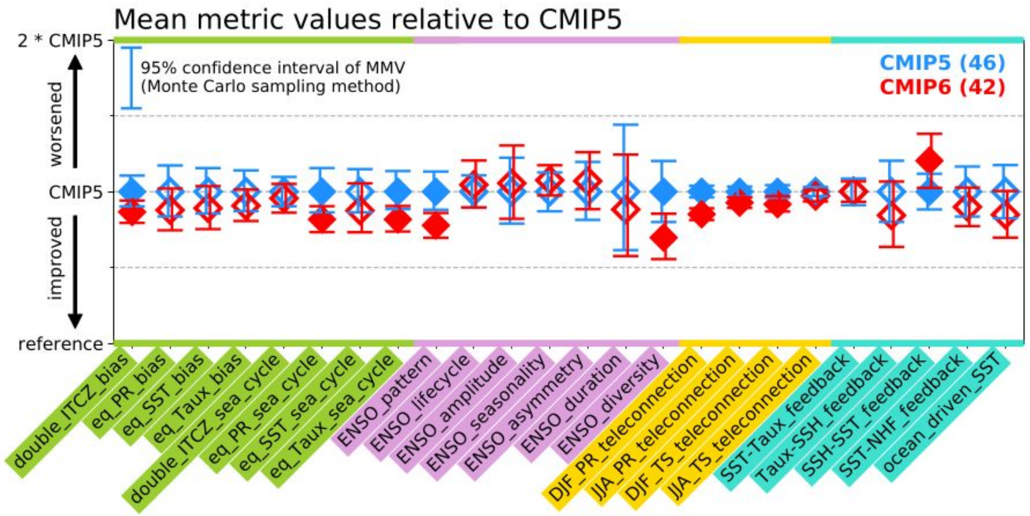


[3]

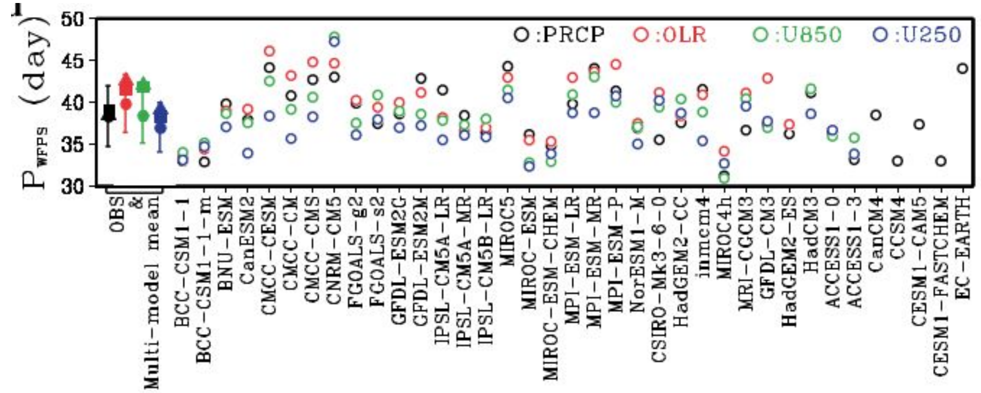
[3] M. Nurisso et al. (2024): *AQUA: A Python Framework for High-Resolution Climate Model Diagnostics*

Our goal

Set of **scalar metrics** → quantitative and quick evaluation



[4]



[5]

ENSO	24 metrics for ENSO performance, teleconnections and processes (DOI: 10.1175/BAMS-D-19-0337.1)
Precipitation	11 metrics benchmarking precipitation (DOI: 10.5194/gmd-16-3927-2023)
Madden–Julian oscillation (MJO)	8 metrics derived from the wavenumber -frequency power spectrum and the combined empirical orthogonal function (DOI: 10.1007/s00382-017-3558-4)
Annual cycle of monsoon	5 metrics related to rainfall accumulation and fractional accumulated rainfall (DOI: 10.1007/s00382-014-2099-3)
Extremes	6 precipitation and temperature indices (DOI: 10.5194/gmd-17-3919-2024)

[4] Y.Y. Planton et al. (2021): *Evaluating Climate Models with the CLIVAR 2020 ENSO Metrics Package*
[5] M.-S. Ahn et al. (2017): *MJO simulation in CMIP5 climate models: MJO skill metrics and process-oriented diagnosis*

High-resolution models

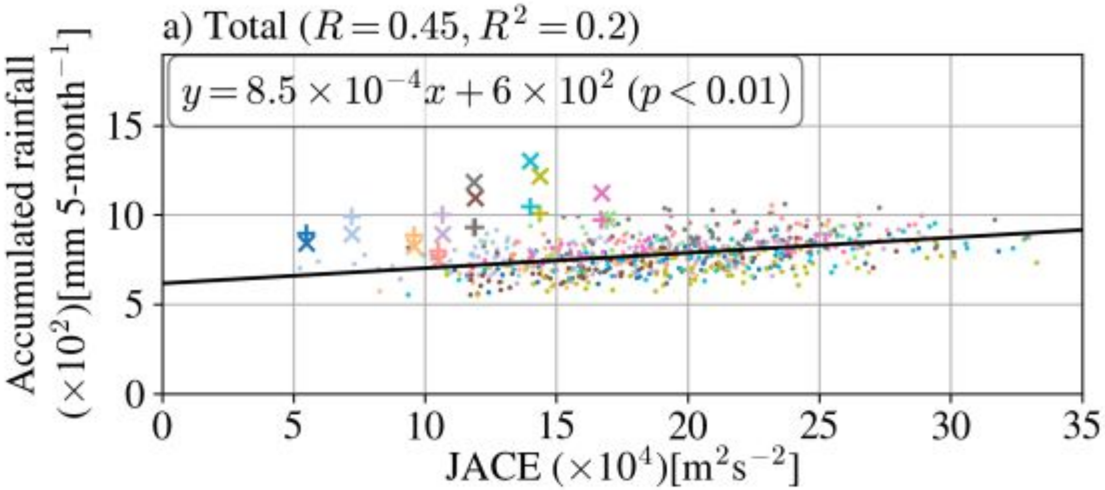
Evaluation of **tropical cyclones**:

- Frequency and intensity
- Relation between accumulated rainfall and cyclone energy (ACE)
- …

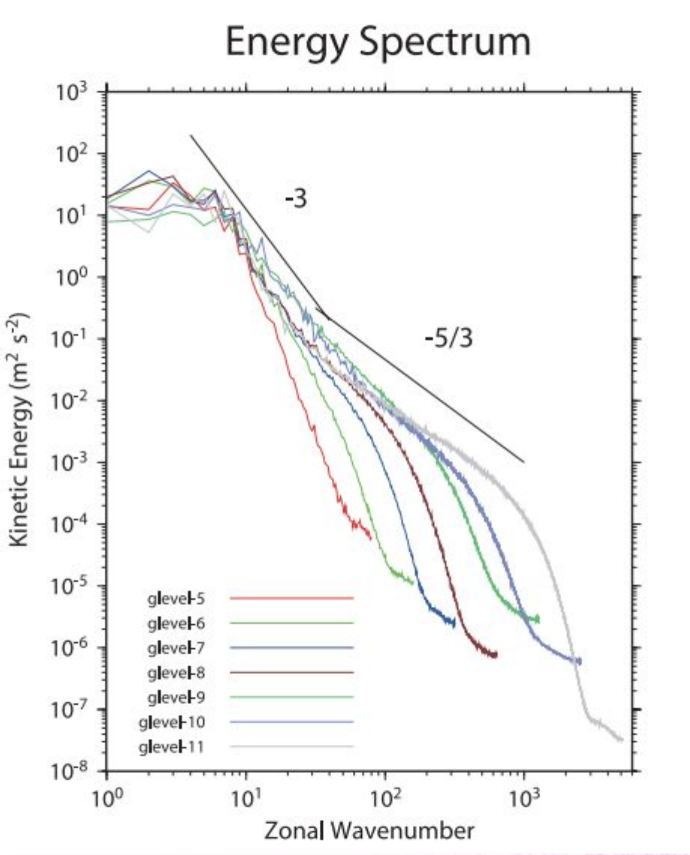
Energy spectrum in the atmosphere

Number of clouds

…



[6]



[7]

[6] Y. Yamada et al. (2025): *Evaluating the Relationship between Seasonal Tropical Cyclone Activity and Rainfall in Japan using a Large-Ensemble Simulation*
[7] K. Terasaki et al. (2009): *Characteristics of the Kinetic Energy Spectrum of NICAM Model Atmosphere*

Integrated into the workflow

HANAMI

- Set of widely applicable and scientifically relevant **scalar metrics**

Run simulations

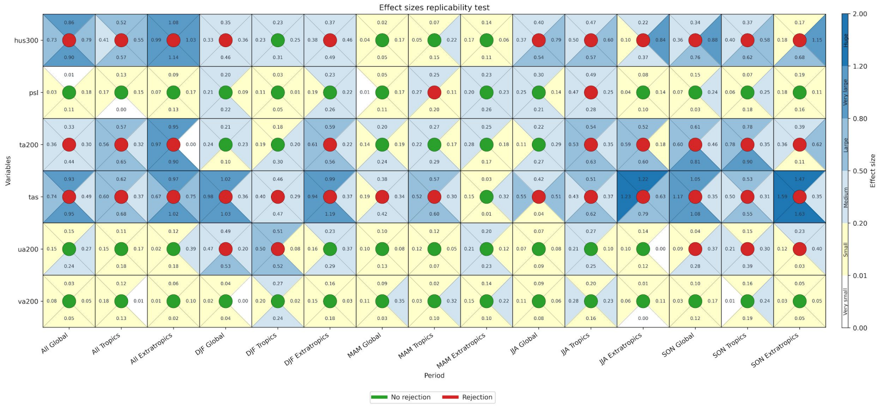
EC-EARTH

NICOCO

+

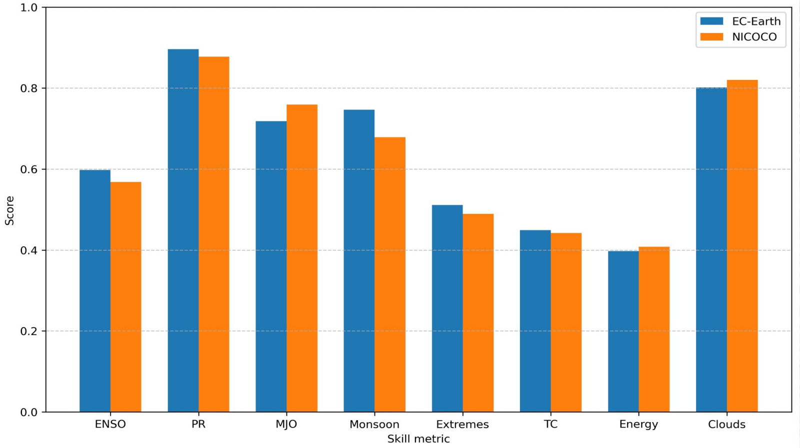
Apply diagnostics

Replicability test



+

Scientific skill



pyhanami Python package

pyhanami

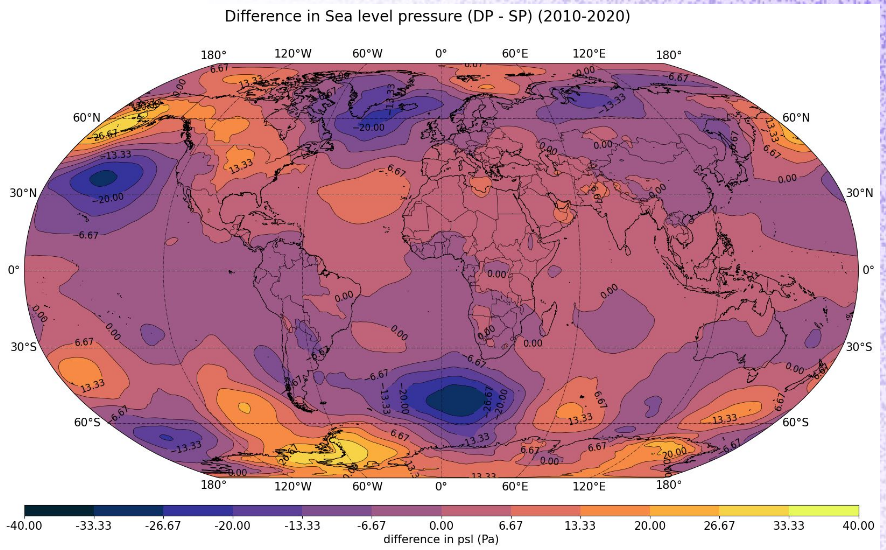
A pip-installable tool to evaluate the replicability and scientific skill of ESMs

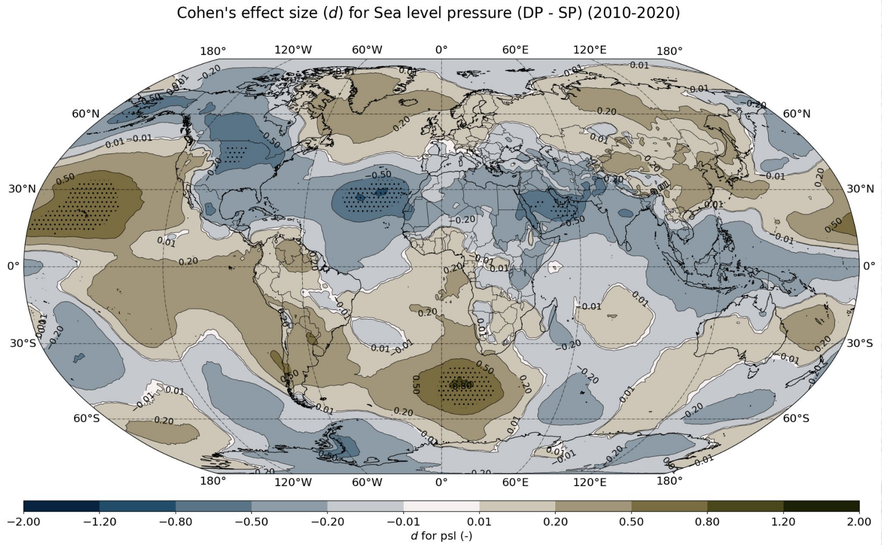
```python
import pyhanami as hnmi

# Retrieve data from catalogue interface
ref = hnmi.SimulationData('interface_path_ref', name='ref')
test = hnmi.SimulationData('interface_path_test', name='test')

# Create plots
diags = hnmi.DataDiagnostics(ref, test)
diags.time_series_plots('output_path')
diags.spatial_plots('output_path')

# Run test and report
tester = hnmi.ReplicabilityTest(ref, test)
tester.matrix_plot('output_path')
tester.report('output_path', time_series=True, spatial=True)
```



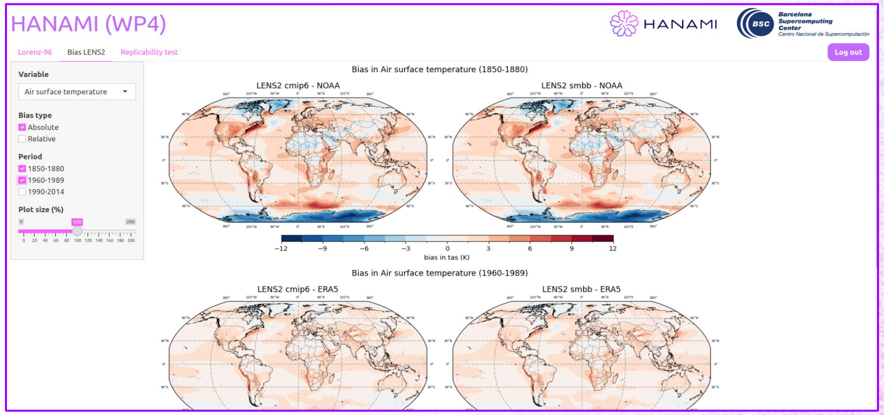


HANAMI

ありがとうございます , thank you

marta.alerany@bsc.es

Take a look at our work!



https://earth.bsc.es/shiny/hanami/

Supplementary slides

Earth System Model (ESM) performance assessment

- Model performance in three aspects
  1. Computational performance
  2. Scientific model skill
  3. Model replicability

Acosta, M. C., Palomas, S., Paronuzzi Ticco, S. V., Utrera, G., **Biercamp, J.**, Bretonniere, P.-A., Budich, R., Castrillo, M., Caubel, A., Doblas-Reyes, F., Epicoco, I., Fladrich, U., Joussaume, S., Kumar Gupta, A., Lawrence, B., Le Sager, P., Lister, G., Moine, M.-P., Rioual, J.-C., … Balaji, V. (2024). The computational and energy cost of simulation and storage for climate science: Lessons from CMIP6. Geoscientific Model Development, 17(8), 3081–3098. https://doi.org/10.5194/gmd-17-3081-2024

Computation Performance for Model Intercomparison Project (CPMIP)	
METRIC	DESCRIPTION
Simulation Year Per Day (SYPD)	Simulated years per day in a 24 h period, collected by timing a segment of a production run of usually one year.
Core-hours Per year Simulated (CHSY)	Simulated years produced with respect to the number of parallel resources used
Complexity (Cmpx)	number of prognostic variables per component
Actual SYPD (ASYPD)	how queue time and interruptions affect the complete experiment duration
Parallelization (Paral)	total number of cores allocated for the run
Energy Cost Per Year (JPSY)	Energy in Joules needed per year of simulation
Memory Bloat (Mem B)	ratio between actual and ideal memory size
Data Output Cost (DO)	time and resources used for performing I/O. The value is given as the percentage added to the simulation without outputs. For example, 1.05 means that DO is 5%.
Data Intensity (DI)	amount of data produced in GB per compute-hour
Coupling Cost (Coup C)	time and resources used in the execution of the coupling algorithm as well as load imbalance among model components. The value is given as the percentage represented comparing to the simulation of the components without coupling. For example, 0.05 means that Coup. C. is 5%.
Simulation Year Per Day (SYPD)	Simulated years per day in a 24 h period, collected by timing a segment of a production run of usually one year.

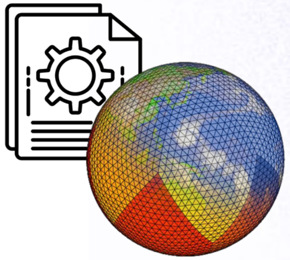
Generic workflow

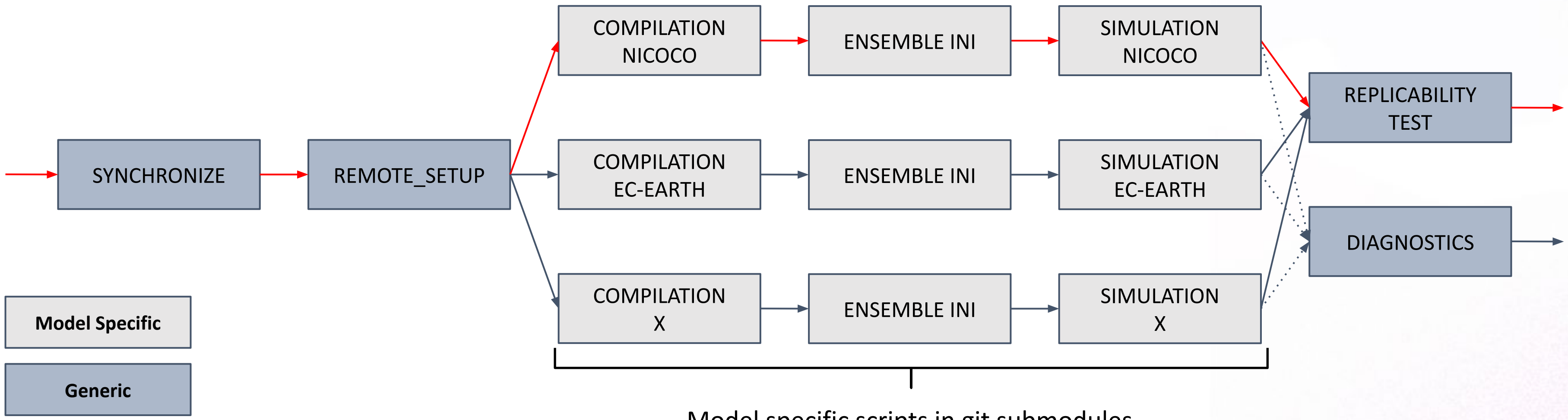
Modular and flexible

Multiplatform support

High degree of data traceability

AUTOSUBMIT





Model specific scripts in git submodules

GitLab

Disadvantages of manual workflows

Excessive user intervention is error-prone. This causes higher usage of computational resources and re-adjustment time.

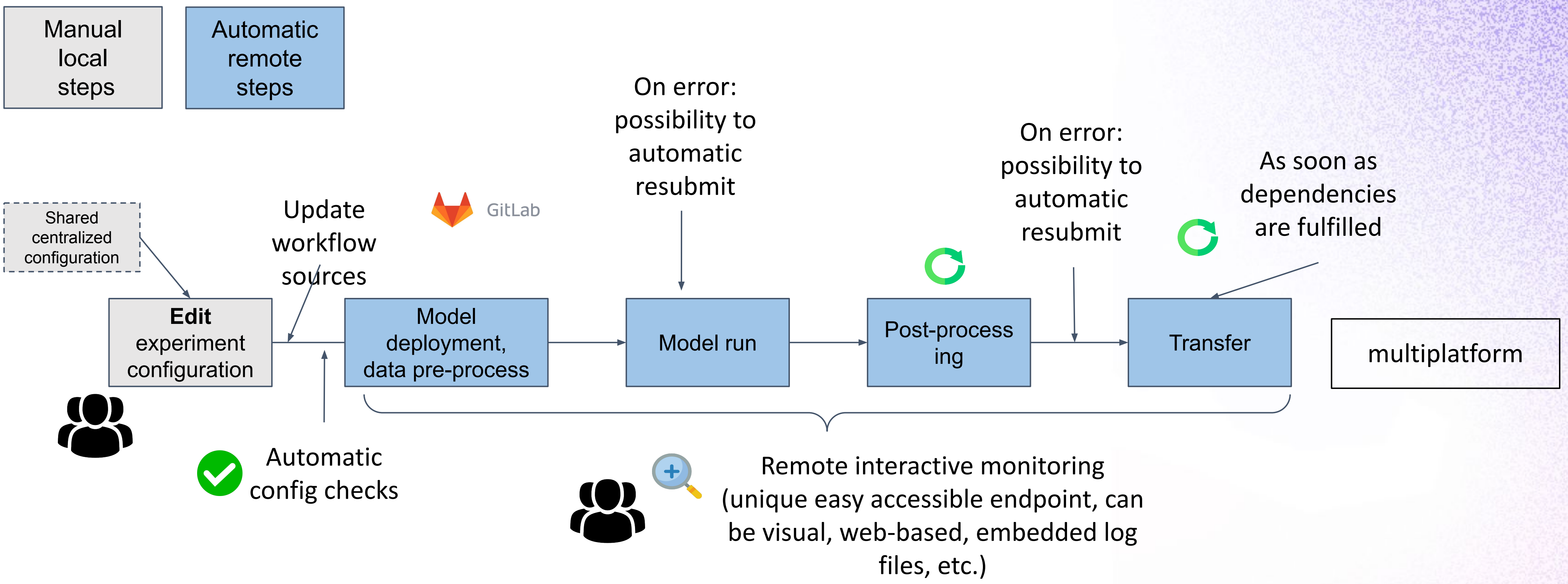
Submitting time chunked long simulations or multiple members is time-consuming.

Decentralized management of sources for running experiments reduces reproducibility and facilitates introducing errors.

Monitoring is time-consuming, specially with unstable environments where machine failures are common.

HANAMI

Using a workflow manager



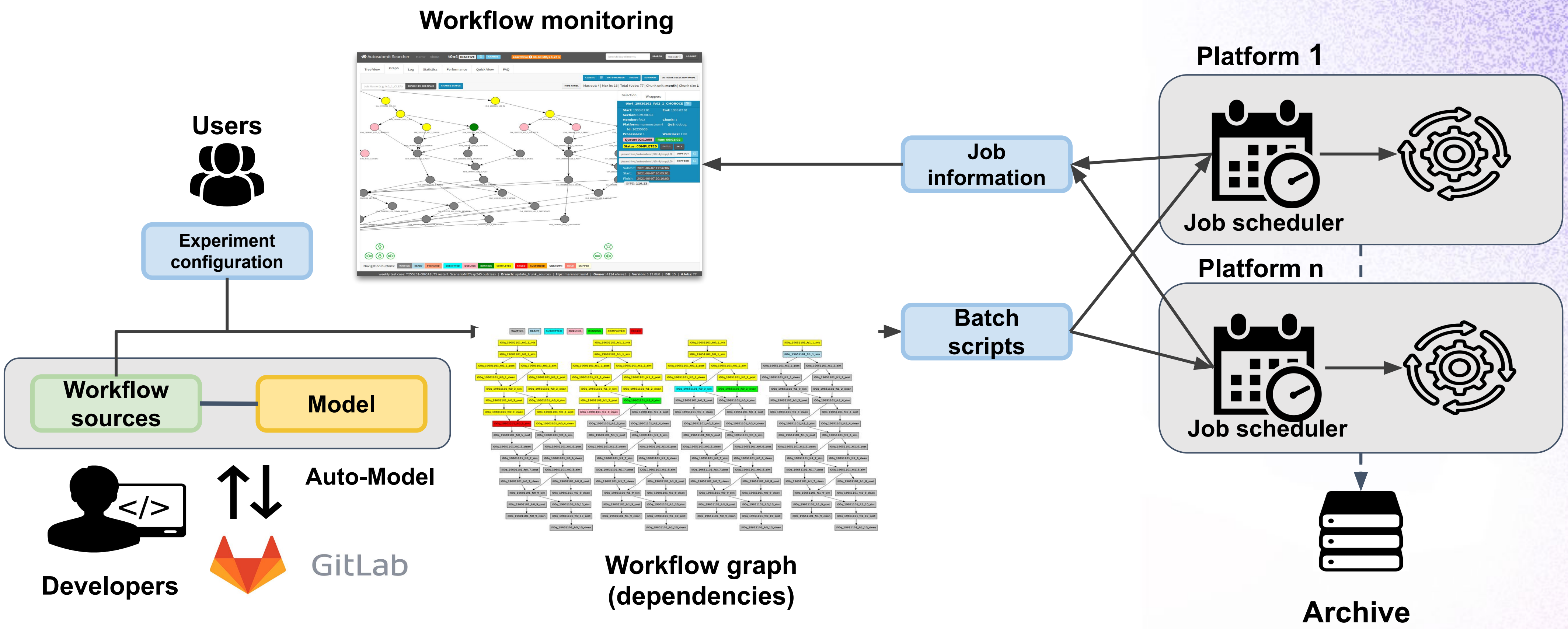
HANAMI

Using a workflow manager

- Automatization and error management, specially on unstable environment.
- Centralized endpoint for monitorization
- Version-controlled experiment configurations and sources
- Increased data provenance
- Increased experiment reproducibility

HANAMI

Autosubmit workflow manager



HANAMI

Used infrastructure