Impact of research software engineering by natESM in climate and weather domain

Aleksandar Mitic[1], Lakshmi Aparna Devulapalli[1], Wilton Jaciel Loch[1], Jörg Benke[2], Sergey Sukov[2]
1-Deutsches Klimarechenzentrum (DKRZ), 2-Forschungszentrum Jülich (FZJ)

Background

The national Earth System Modeling Strategy (natESM) project was funded as part of a sustainable strategy for the future of the Earth system community. It provides targeted support for the development and optimization of Earth System Models for HPC systems. Started in November 2021, a growing team of research software engineers worked on 4-6 months individual sprints in collaboration with domain scientists.

Sprints

What is a Sprint?

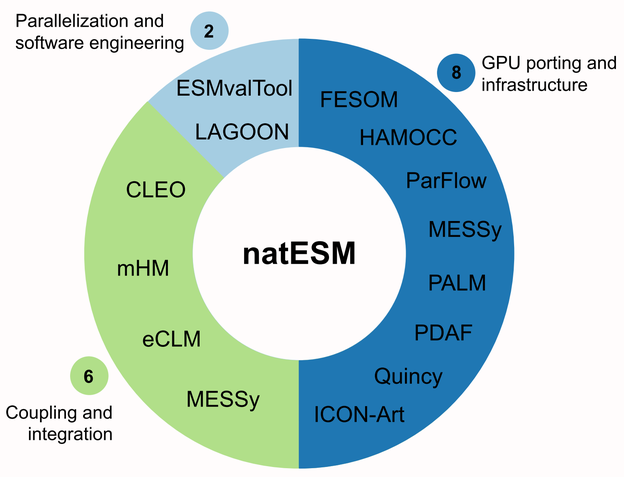
- A sprint consists of a goal-oriented package of work executed by a dedicated RSE on a selected ESM model during a defined amount of time.
- The natESM support team's expertise compounds with each sprint. Every sprint offers a unique challenge and a learning opportunity for the RSE team. These learnings are useful for future sprints.

Sprint subjects

GPU porting: Adapting the CPU compatible code to work on the GPU architecture by porting the kernels and implementing appropriate data movement between the host and the device.

Parallelization and software engineering: Improving computation time and memory efficiency by utilizing different (parallelization) paradigms like MPI, Dask etc.

Coupling and integration: Integrating two or more Earth System Models to expanding functionalities and improve modularity.

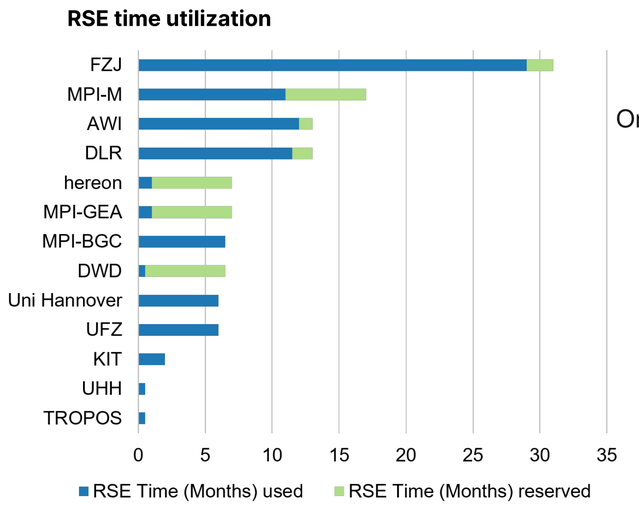


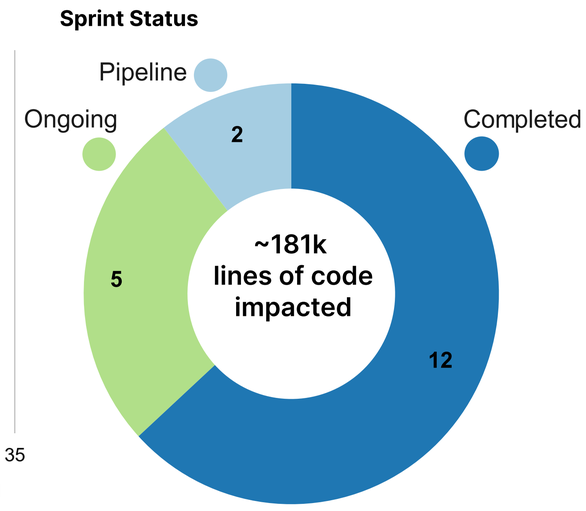
1

Results

RSE time utilization

The time of a highly skilled RSE is valuable and hence the sprints are managed and organized by the natESM's steering committee in an efficient manner. A structured timetable outlining a set of tasks aligned with the sprint goal is carefully planned and implemented. The chart shows the RSE time utilization by community members so far [3].





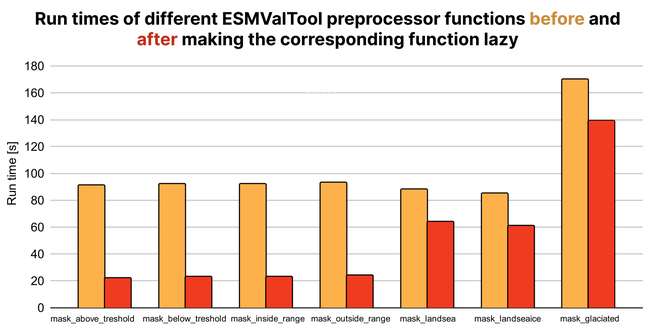
Sprint status

The quantification of the natESM sprints gives a clear picture of the current and future sprints status. The RSE knows which future sprint is in his/her pipeline and gets an opportunity to form connection with the scientist.

3

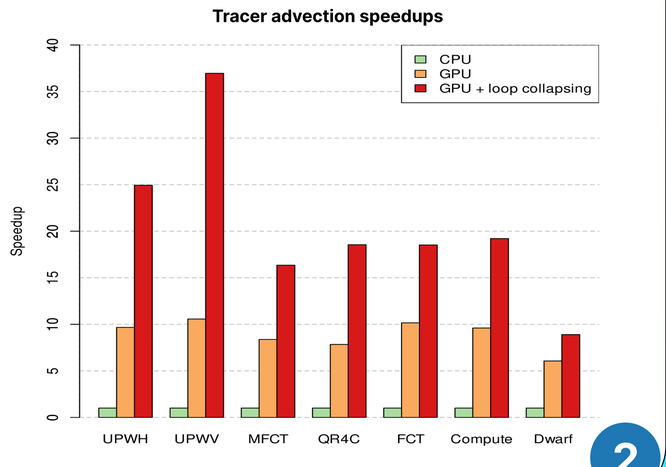
Sprint#6 : Improving ESMValTool memory efficiency

On the right figure, the results of the Dask porting on the example of a test case can be seen. Several preprocessor functions were applied to 65 years of daily 3D data of the CMIP6 model CanESM5. The tests were performed on a single compute node on DKRZ's Levante with 128 cores and 256 GB of RAM [1].



Sprint#3 : FESOM GPU porting

Excellent speedup on GPUs was achieved as part of the FESOM sprint. The initial porting served as a reference for future porting. Adjacent graph shows the GPU speedups in a node-to-node comparison on Levante for tracer advection. The average compute speedup achieved for the tracer advection is around 10 times for the ported version and around 20 times with the addition of loop collapsing [2].

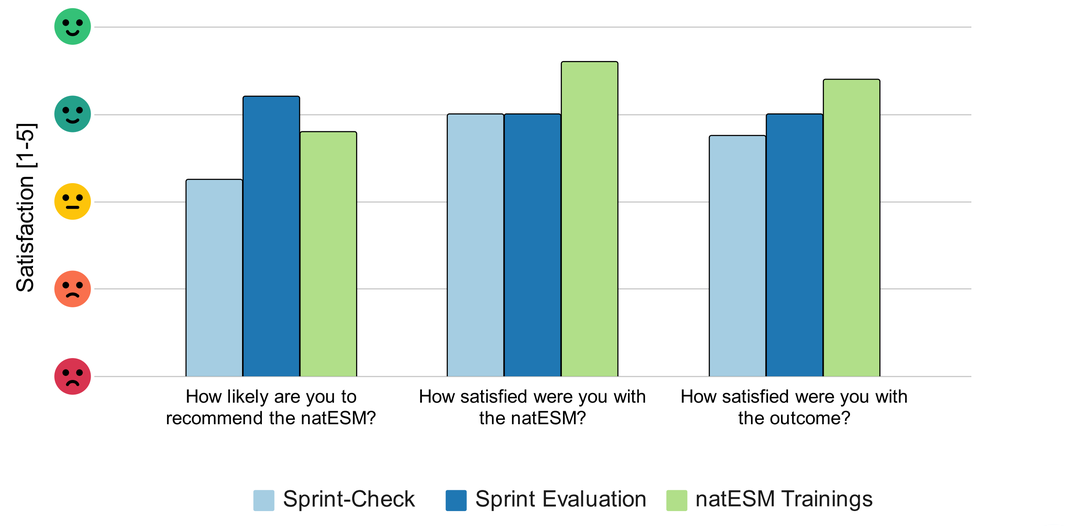


2

natESM community feedback

The natESM steering group conducted a survey on community satisfaction with the sprint process and our training offerings. The survey results shown below look promising for the efforts put in by natESM for building a stronger ESM community.

natESM's positive impact on the ESM community resulted in the preparation of a second phase for the project. The project proposal has been submitted and the community is excited for another successful project phase.



4

Summary

- The natESM project sprints are a building block of growing importance in advancing Earth System Modeling through, for instance, innovative GPU optimization and the seamless coupling of complex Earth system models.
- By accelerating computational efficiency and improving model interactions, we are better equipped to understand and predict the impacts of environmental change. Together, these efforts lay the foundation for more accurate simulations, ultimately supporting informed decision-making for a sustainable planet.
- Keeping up with the fast pacing world moving towards Artificial Intelligence and Machine Learning, natESM plans to invest in the Machine Learning capabilities for Earth System Modeling in phase 2.

References:
[1] ESMValTool Sprint. [Online] Available: https://www.nat-esm.de/services/support-through-sprints/accepted-sprints/20240513-natesm_sprint_docu_esmvaltool_final.pdf (accessed 19-Feb-25)
[2] Booster for FESOM 2.1 Sprint [Online] Available: https://www.nat-esm.de/services/support-through-sprints/accepted-sprints/20230721_natesm_sprint_docu_fesom_booster_v03.pdf (accessed 19-Feb-25)
[3] natESM Community Workshop 2025. [Online] Available: https://www.nat-esm.de/services/trainings/events/ws5_2025 (accessed 19-Feb-25)