



IFS-NEMO developments

Carlos Peña

HPC for Earth Sciences, BSC

Earth Sciences
Department



**Barcelona
Supercomputing
Center**
Centro Nacional de Supercomputación

IFS–NEMO development scope

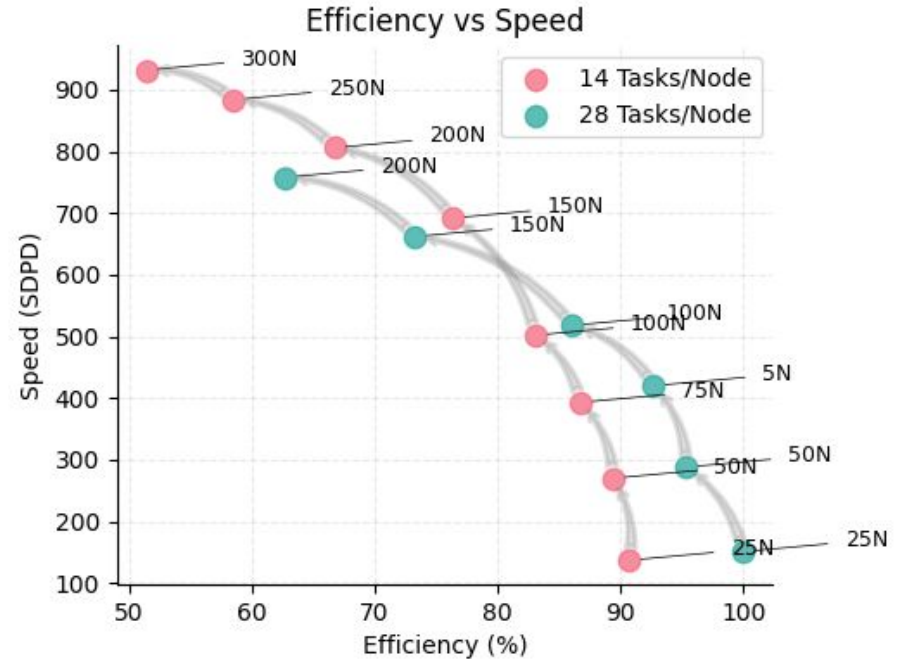
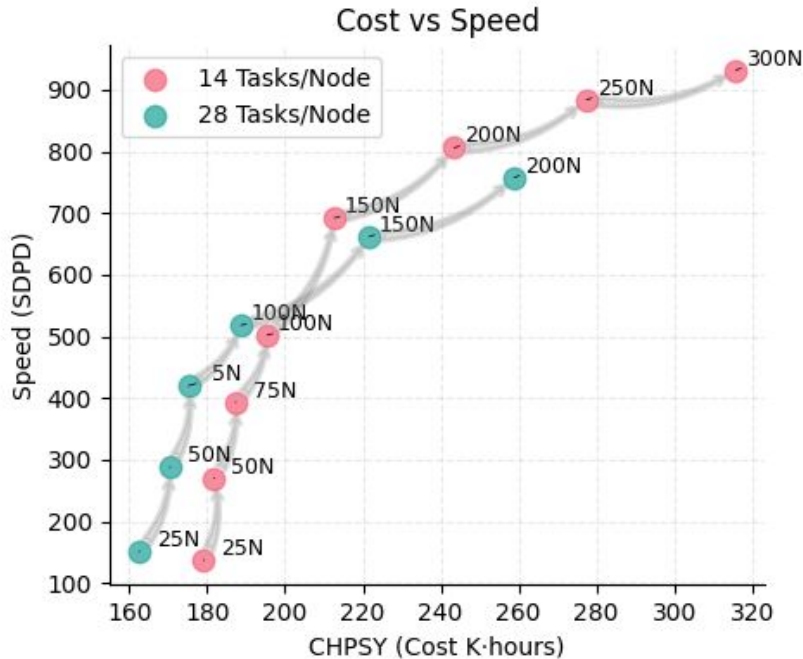
Numerical kernels, parallel execution and coupled scalability

Area	Development focus
IFS atmosphere	Vectorisation, communication and MPI imbalances
NEMO ocean and ice	System tuning, full ICE GPU port

Default test platform: MareNostrum 5 (MN5)

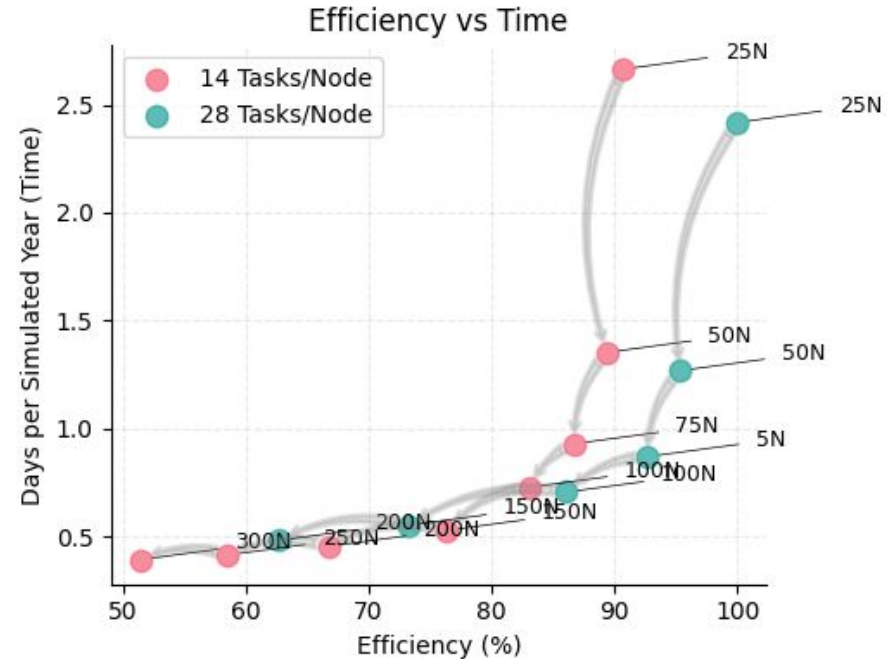
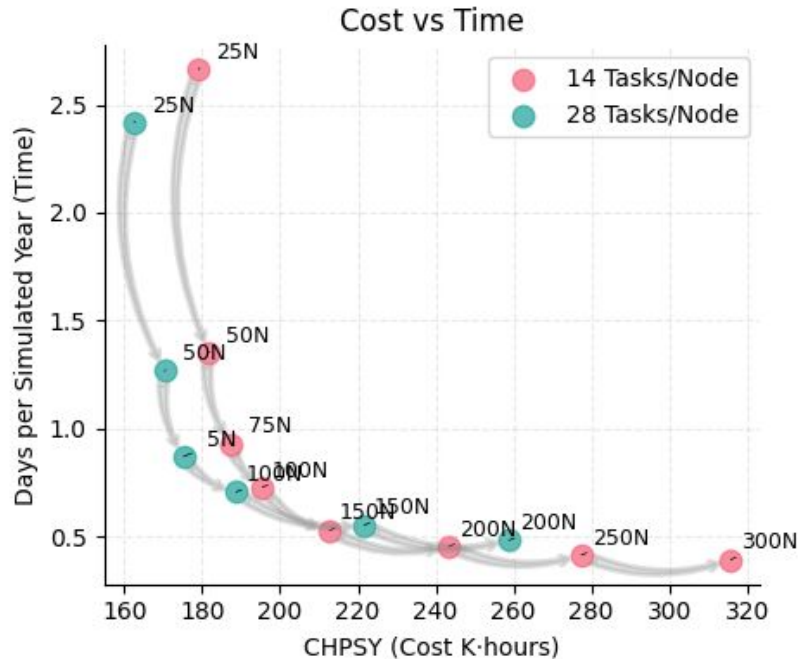
IFS-NEMO scalability: TCo1279-eORCA12

MN5: throughput versus cost and relative efficiency



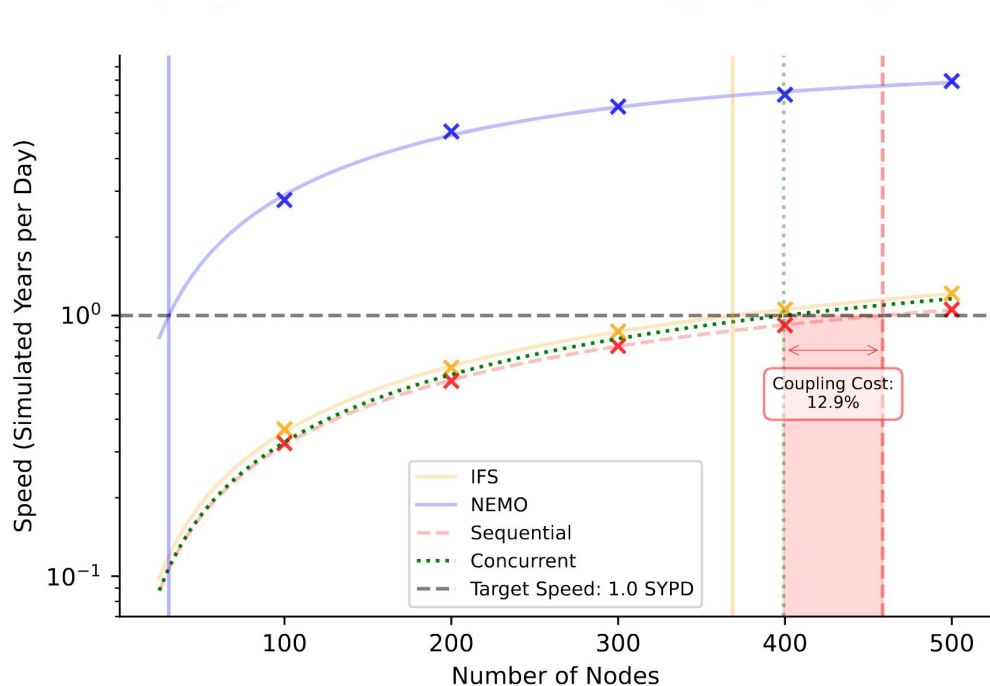
TCo1279–eORCA12: simulation time and cost

MN5: days per simulated year for 14 and 28 MPI tasks/node



System-Level Impact: Serial Coupling Cost

- Setup: ~5 km (IFS) vs ~10 km (NEMO)
- Same node scale → inefficient NEMO scaling
- More nodes required vs ideal decoupled execution (*shaded region*)



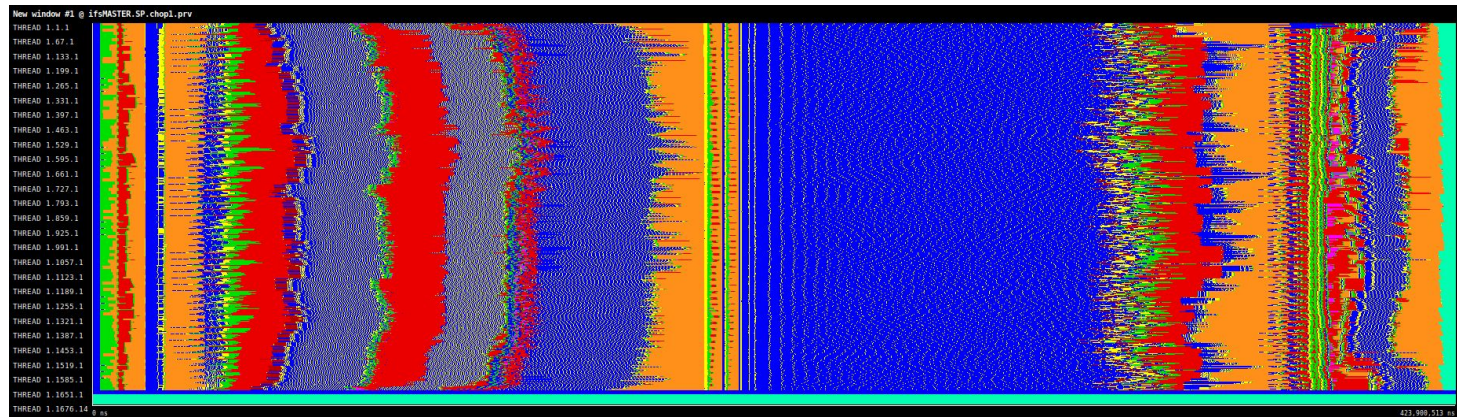
IFS

Profiling and implementation

Atmospheric kernels and repeated communication

IFS profiling: execution trace

MN5, TCo1279, 118 nodes, 14 MPI $\times$ 8 OpenMP per node



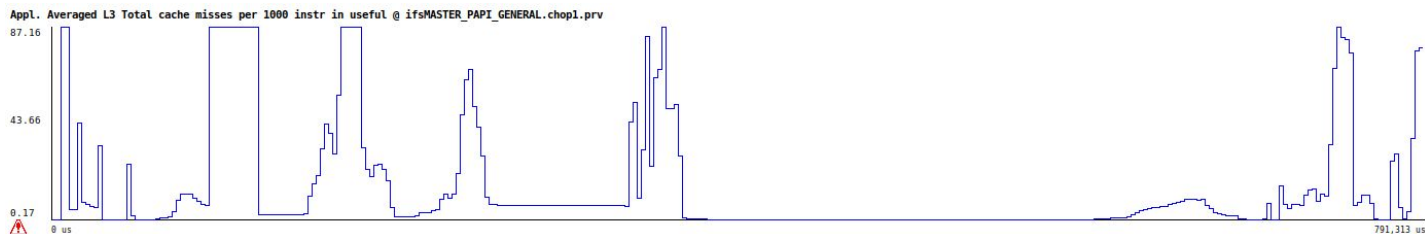
CORMASS DRY INVERSE TRANS Dynamics Physics_EC DIRECT TRANS SEMI IMP WAM

Trace evidence locates numerical work and communication phases

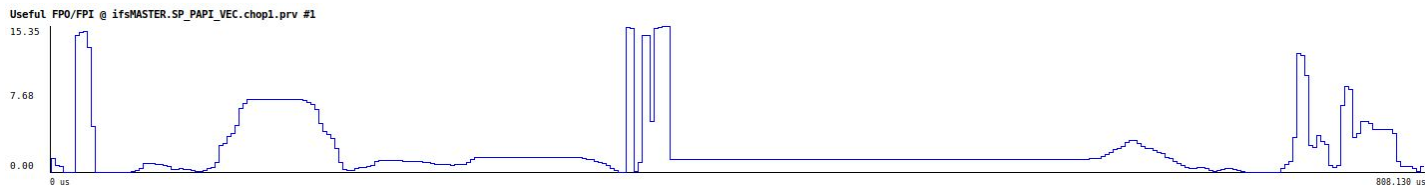
IFS profiling: memory access and SIMD

Hardware counters expose low SIMD use in key hotspots

L3 cache misses per 1,000 instructions



FP SIMD operations ratio



The profiling evidence motivated compiler and execution-policy tests

IFS: FP exception policy and vectorisation

Strict exception handling constrained SIMD execution in physics

Coupled throughput in simulated days per day



1.13×

**Higher coupled
throughput
in the measured example**

FP exception trapping remains a robustness policy choice

IFS: communication improvements

Reducing recurring communication in diagnostics and transport

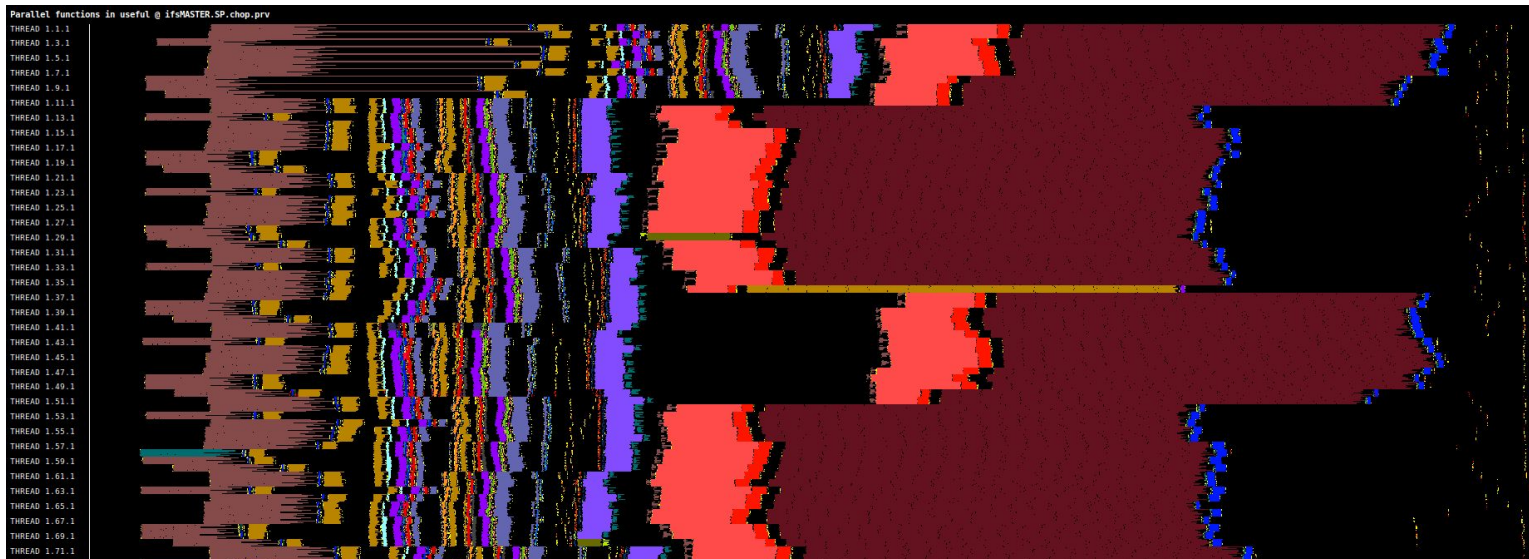
WAM diagnostics

Remove repeated norm gathers when those diagnostics are unnecessary.

Semi-Lagrangian halos

VMAX2 tuning reduces communication volume.

IFS: a resolved LT_DIR scaling bottleneck



- Special path for subset of ranks (Double precision needed for conservation)

IFS: a resolved LT_DIR scaling bottleneck

- Repeated matrix conversion slowed $m = 0$ ranks.

TCo1279-eORCA12

2%

TCo2559-eORCA12

8%

Share of total IFS-NEMO runtime lost to imbalance before the fix

Fix: retain both precisions of the Legendre polynomial matrices

NEMO

CPU tuning, precision and GPUs

Ocean and sea-ice implementation work

NEMO: CPU developments

Changes already applied in production simulations

Target	Change	Deployment
Diagnostics	Remove unnecessary dia_hsb MPI_Allreduce	Applied in production
dyn_ldf	Add missing OpenMP parallelism	Applied in production
CPU tuning	Tune AVX-512 and prefetcher aggressiveness	Applied in production

Compiler tuning complements OpenMP parallelism

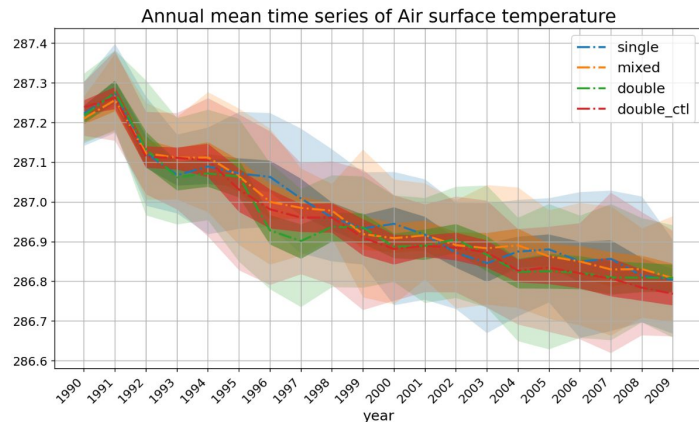
NEMO: reduced precision

Lower data volume improves use of the memory hierarchy

NEMO runtime relative to double precision



-11.5% mixed -28.2% single



25-member ensembles
TCO399-eORCA025 validation

NEMO: complete ICE GPU implementation

Full sea-ice porting coverage achieved

Numerical component	GPU region
Dynamics: rheology	icedyn_rhg
Dynamics: advection	icedyn_adv
Dynamics: ridging / rafting	icedyn_rdgrft
Thermodynamics	icethd

NEMO: ICE GPU speedups on MN5-ACC

ORCA12, CPU/GPU speedup at equal node counts

Region	8 nodes	32 nodes	64 nodes
icestp	3.9×	1.3×	0.7×
ice_stp	4.9×	1.5×	0.8×
ice_dyn	4.3×	1.2×	0.7×
icedyn_rhg	3.0×	0.8×	0.4×
icedyn_adv	8.7×	6.2×	4.0×
icedyn_rdgrft	4.9×	1.9×	1.1×
icethd	4.6×	1.6×	0.9×
extra	8.1×	3.1×	1.9×

16 MPI ranks/node, 4 ranks/GPU. Ratios below 1× indicate slower GPU execution.