ITT380 Benchmarking

Benchmaking Working Group

itt380-benchmarks@groups.ecmwf.int

ECMWF

Disclaimer

- The information contained herein represents the proposed benchmarking methodology for ITT380 and is provided solely to assist prospective tenderers in preparing their benchmarking activities for ITT380
- ECMWF expressly reserves the right to change, modify, or revise some or all of the aspects of the benchmarking methodology and related requirements presented herein at any time prior to the release of ITT380
- ECMWF will not reimburse any expenses, costs, or charges incurred by (prospective) tenderers or any third parties in connection with the benchmarking activities for ITT380
- ECMWF accepts no liability for any actions, decisions, or costs resulting from the use of this information

New ECMWF HPCF

- The new HPCF is intended to meet ECMWF's high performance computing requirements for both **physics-based** and **data-driven** modelling
- Four expected node types as part of the new HPCF but only three are relevant to benchmarking activities:
  - **SIM** nodes for physics-based modelling and simulation using the IFS, likely split across four partitions
  - **MLM** nodes for mixed and general purpose machine learning workloads including time-critical operations (e.g., inference and training) hosted in the Bologna DC, likely split across two partitions
  - **MLT** nodes tuned for training machine learning models that can be hosted either in the Bologna DC or off-premise, likely in a single partition
- Minimum performance requirement for **SIM** nodes of **1.5** relative to the current ECMWF HPCF
- ECMWF expects **SIM** nodes to be based on traditional general purpose CPUs but leaves this decision with the vendor
- Even split expected of resources/capability between **MLM** and **MLT** nodes

Timeline of RAPS releases

Version	Date	Release Notes
RAPS23	August	new IFS cycle (CY50R3), AIFS forked from main (branch itt380)
RAPS23	August	data assimilation and IFS/AIFS updates from vendors
RAPS23	September	GPU capabilities (radiation, advection) and IFS/AIFS updates
RAPS23	October	general fixes and final release

Benchmark code modifications

- No code modifications permitted to IFS and AIFS after final benchmark release in October except for those required to be able to run the benchmark
- Compared to previous procurements, we only ask for a minimum performance factor for **SIM** nodes
- ECMWF wants to achieve this performance uplift thanks to new hardware and not a faster benchmark code
- Code optimisations can be performed prior to final RAPS23 release in October in collaboration with ECMWF
- However, they will be visible and available to all other vendors that take part in ITT380
- Cut-off date for accepting code modifications and optimisations will be two weeks prior to the official ITT release.
- Changes and updates to the the underlying software stack are allowed during ITT380 (compiler, math libraries etc)

ITT380 benchmarks

- Physics-based modelling based on the **IFS** $\rightarrow$ F_{sim}
  - IFS **research** (res) benchmark
  - IFS **time-critical** (tc) benchmark
- Data-driven modelling based on **AIFS** $\rightarrow$ F_{ai}
  - AIFS-ENS **inference** (inf) benchmark
  - AIFS **training** (train) benchmarks

IFS research benchmarks

- **4 km (TCo2559L137) analysis benchmark → IFS-4DVAR:**
  - Run **30** copies of IFS-4DVAR concurrently on one **SIM** partition
  - Performance factor: $F_{\text{an}} = \frac{T_{\text{ecmwf}}}{T_{\text{an}}}$
  - T_{an} is time to successfully complete all copies and T_{ecmwf} is reference time of running the same number of copies of IFS-4DVAR on one ECMWF HPCF parallel partition
- **4 km (TCo2559L137) ensemble forecast benchmark → IFS-ENS:**
  - Run **60** copies of IFS-ENS forecasts concurrently on one **SIM** partition
  - Performance factor: $F_{\text{fc}} = \frac{T_{\text{ecmwf}}}{T_{\text{fc}}}$
  - T_{fc} is time to successfully complete all copies and T_{ecmwf} is reference time of running the same number of copies of IFS-ENS on one ECMWF HPCF parallel partition.
- **The number of copies for IFS-4DVAR and IFS-ENS are provisional and might change**
- IFS research benchmark performance factor:

$$F_{\text{res}} = \frac{2}{\frac{1}{F_{\text{fc}}} + \frac{1}{F_{\text{an}}}}$$

30 is the number of IFS-4DVAR copies and **60** the number of IFS-ENS copies that ECMWF were able to fit into one of their current ECMWF HPC parallel partition. These are preliminary numbers that are given for guidance and might change for the actual ITT.

IFS time-critical benchmark

- Components:
  - 4 km (TCo2559L137) analysis benchmark (IFS-4DVAR)
  - 4 km (TCo2559L137) ensemble forecast benchmark (IFS-ENS)
- Current ECMWF operations: 50% analysis and 50% model forecast
- Time critical operational benchmark:
  - Run **12** copies of IFS-4DVAR concurrently followed by **12** copies of IFS-ENS
  - **The number of copies is subject to change and should be treated as provisional**
  - T_{an} is time to successfully complete all IFS-4DVAR copies in time-critical configuration
  - T_{fc} is time to successfully complete all IFS-ENS copies in time-critical configuration
  - Ideally, all copies of IFS-4DVAR and IFS-ENS will complete successfully within one hour each
  - A maximum of 75% of **SIM** nodes can be used across two of the four partitions

$$F_{\text{tc}} = 1.5 \times \frac{3600}{T_{\text{tc}}} \tag{1}$$

$$T_{\text{tc}} = \frac{T_{\text{an}} + T_{\text{fc}}}{2} \tag{2}$$

F SIM

$$F_{\text{sim}} = \frac{2}{\frac{1}{F_{\text{res}}} + \frac{1}{F_{\text{tc}}}} \geq F_{\text{min}}$$

$$F_{\text{min}} = 1.5$$

AIFS-ENS inference benchmark

- Target:
  - 100 4 km AIFS-ENS members
- Run all AIFS inference members concurrently on up to **42%** of available GPUs of **MLM** nodes (i.e., approx. 85% of one **MLM** partition)
- N_{used} is number of **GPUs** used to run the benchmark, N_{total} is total number of **GPUs** available and T_{ecmwf} is reference time provided by ECMWF

$$F_{\text{inf}} = \frac{T_{\text{ecmwf}}}{\left(\frac{N_{\text{used}}}{N_{\text{total}}}\right) \times T_{\text{inf}}} \quad \text{where } \frac{N_{\text{used}}}{N_{\text{total}}} \leq 0.42 \tag{3}$$

AIFS training benchmark

- 4 km AIFS training benchmark
- Fixed data parallelism dimension and ensemble dimension
- Maximize F_{train} compared to ECMWF provided reference

$$F_{train} = \frac{T_{ecmwf}}{T_{train}}$$

- Additional training benchmark at TCo799 resolution, equivalent to upcoming ERA6, is under development that will target both **MLM** and **MLT** nodesets
- F_{train} will be defined as a composite of F's for 4ḱm and 14ḱm training

F AI

$$F_{\text{ai}} = \sqrt{F_{\text{train}} \times F_{\text{inf}}}$$

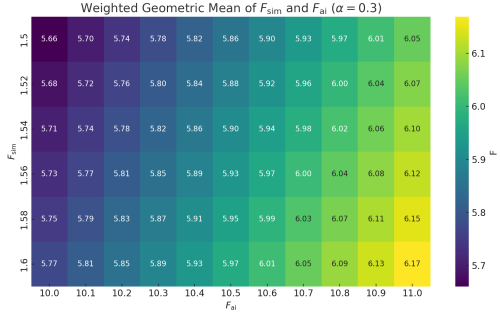
The F factor

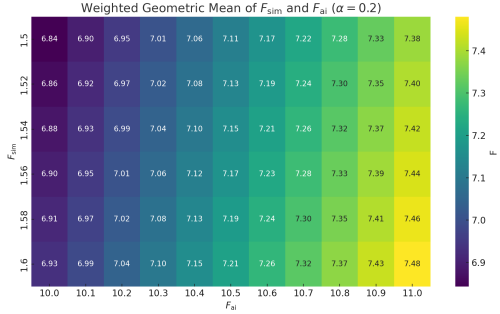
$$F = F_{\text{sim}}^{\alpha} \times F_{\text{ai}}^{1-\alpha}$$

- Weighted geometric mean to combine F_{sim} and F_{ai} into a single scalar
- α will be set such that increases in F_{ai} have a larger impact on the overall F compared to increases in F_{sim}
- ECMWF wants the vendor to focus fully on the **MLM** and **MLT** nodes once the minimum performance requirement for **SIM** nodes is met
- Overshoots in F_{sim} are still rewarded although best way to improve overall F is to increase F_{ai} once $F_{sim} \geq 1.5$
- α likely to be between 0.2 and 0.3

The F factor

- Example of possible values for α and their impact on the final **F**





I/O

- Compared to last procurement, we have removed most of the I/O from the benchmarks with the goal of making life easier for the vendors
- For the IFS time-critical benchmarks, the IFS-ENS models will still perform post-processing and send data to I/O servers, which will also perform grib encoding and compression, but no I/O will be sent to the filesystem
- We are also strongly considering to remove initialisation time for all IFS workloads when accounting for the walltime
- For AIFS inference, plan is to also do data aggregation, and grib encoding via multIO but not write anything to the filesystem
- I/O for both IFS-ENS and AIFS ENS will be modelled using standalone tool
- The only relatively large I/O that we cannot remove is in the AIFS training benchmark when reading training data set via data loaders

RAPS23 July release

- First RAPS23 release for ITT380
- **IFS:**
  - First release based on CY50R3
  - 4 km IFS-ENS in both research and time-critical configurations
  - Model resolutions from 128 km to 4 km
  - TLADJ configurations as a precursor to full data assimilation (to be released in August)
  - GPU enabled components: spectral transforms, physics, wave model (more available in September)
- **AIFS:**
  - More details: `https://arxiv.org/pdf/2406.01465`
  - **itt380** branch forked off main in July 2025
  - 4 km training benchmark
  - 4 km inference benchmark
  - 14 km (ERA6) training benchmark in development (August release)