Measurements of real computational performance of ESMs: the CPMIP Project

4th ENES HPC Workshop
Toulouse, FRANCE

V. Balaji and many others

NOAA/GFDL and Princeton University

7 April 2016

What computational performance metrics matter for science?

- For a given experimental design, what can I afford to run?
- If I add complexity (such as adding a biogeochemistry component to an AOGCM), what will I have to sacrifice in resolution?
- How much computing capacity do I need to participate in a campaign like CMIP6? How much data capacity?
- Do the queuing policies on the machine hinder the sustained run of a long-running model?
- During the spinup phase, how long (in wallclock time) before I have an equilibrium state?

Typical performance metrics, such as Flops and GHz, do not answer such questions.

Real model performance: some considerations

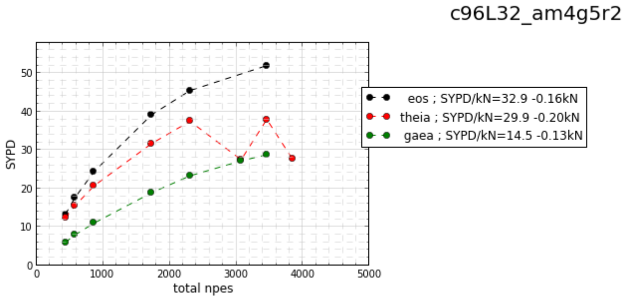
- Productions runs may be configured for capability (minimizing time to solution or SYPD) or capacity (minimizing allocation or CHSY).
- Computing resources can be applied to resolution or complexity: what is a good measure of model complexity?
- ESM architecture governs component concurrency: need to measure load balance and coupler cost.
- Codes are memory-bound: locate bloat (memory copies by user or compiler).
- Models configured for scientific analysis bear a significant I/O load (can interfere with optimization of computational kernels). Data intensity (GB/CH) is a useful measure for designing system architecture.
- Actual SYPD tells you if you need to devote resources to system and workflow issues rather than optimizing code.

Analysis of several ESMs

- Measure overall computation cost for capability (Speed) or capacity (Throughput) configurations.
- Measure complexity as number of prognostic variables in the model. (There are probably better measures based on cluster coefficients, etc.)
- Measure coupler cost and load imbalance separately.
- Measure memory bloat as actual memory (resident set size) compared to ideal memory (number of variables $\times$ data domain size).
- Measure I/O load by rerunning model with diagnostics off. (input files and restart files are considered an unavoidable cost and aren't counted here.)
- Measure actual SYPD for a complete run (from when you typed run to when the last history file was archived).

The CPMIP Metrics: Performance

The Computational Performance MIP (CPMIP) is a proposal for systematic collection of CP-related metrics from CMIP6 (via ESDOC).



- SYPD simulated years per day.
- CHSY compute hours per simulated year (NP=CHSY*SYPD/24)
- Speed (S) and throughput (T) modes: models can be configured for maximum SYPD or minimum CHSY meeting scientific requirements.

The CPMIP Metrics: Model Characteristics

The number of degrees of freedom of a model is the number of spatial degrees of freedom (resolution) $\times$ the number of prognostic variables (complexity).

- Resolution:

$$G_c \equiv \mathrm{NX} \times \mathrm{NY} \times \mathrm{NZ} \quad (1)$$

$$G \equiv \sum_c G_c \quad (2)$$

- Complexity: $S_c \equiv$ size of restart file in bytes. For a model in double precision:

$$V_c \equiv S_c/G_c/8 \quad (3)$$

$$V \equiv \sum_c V_c \quad (4)$$

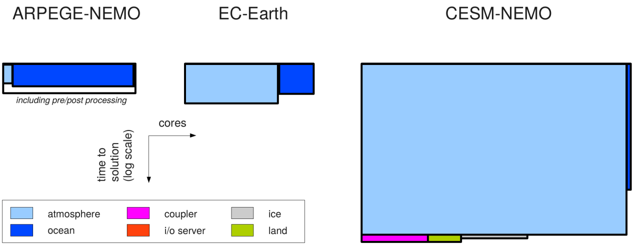
  Weighted towards counting 3D variables only.
- Also provide representative grid resolution Δx_c and Δz_c for common-language comparison.

The CPMIP Metrics: Coupling cost

While load imbalance and the actual time spent in the coupler (e.g regridding) can be separately measured, we initially require only the sum of the two: coupling cost.

$$C \equiv \frac{T_M P_M - \sum_c T_c P_c}{T_M P_M} \tag{5}$$

where T_c excludes waiting times. This is the “white area” below:



The CPMIP Metric: Memory, I/O and workflow

- Memory bloat: compare $M \equiv$ Resident Set Size (RSS) high water mark with ideal memory M_i

$$M_i \equiv \sum_c S_c \tag{6}$$

$$B \equiv \frac{M - M_i}{M_i} \tag{7}$$

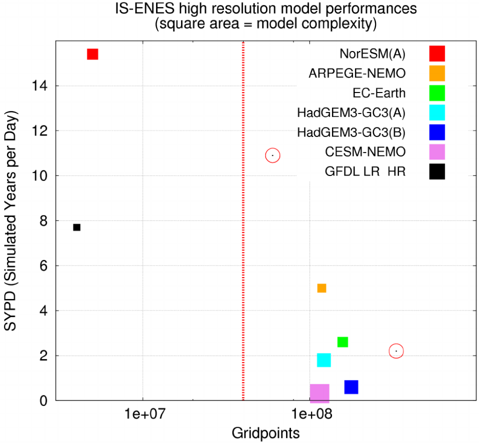
- I/O cost: measured differently for synchronous and asynchronous I/O (e.g XIOS).

$$D \equiv \frac{CHSY - CHSY_{\mathrm{noI/O}}}{CHSY} \qquad D \equiv \frac{P_M - P_{\mathrm{I/O}}}{P_M}$$

  Measured on science runs with full I/O load.
- Data intensity: GBSY/CHSY (GB of output per CH).
- ASYPD for science run of N years: timestamp of (lastfile-firstfile)/N

CPMIP preliminary results

SYPD vs resolution and complexity:



(Inconsistent complexity measurement: to be corrected)

CPMIP preliminary results: Haswell vs Opteron

Question: NOAA replaces 122,000 Opteron nodes on Cray Gemini network by 48,000 Haswell on Cray Aries: what is the relative performance?

Model	Machine	Resol	SYPD	CHSY
CM4 S	**`c2`**	1.2E8	4.5	16000
CM4 S	**`c3`**	1.2E8	10	7000
CM4 T	**`c2`**	1.2E8	3.5	15000
CM4 T	**`c3`**	1.2E8	7.5	7000

Actual comparison about 2.2X in CHSY, cannot be inferred from PFLOPs, GHz, etc. But total cores down by 2.5X.

Summary

- The CPMIP Project proposes a set of common measures of computational performance for Earth System Modeling.
  - universally available from current ESMs, with any underlying numerics, on any underlying hardware.
  - representative of actual performance of the ESMs running in a science setting, no idealizations, no kernels..
  - performance across the entire lifecycle of modeling: computation, data, and workflow,
  - easy to collect, no specialized instrumentation or software, gather during routine production computing.
- Defines a computational profile for ESMs and its evolution.
- Reflect scientific concerns on performance: planning of experiments, design of machines suitable to the profile.
- Proposal to include these metrics in ESDOC, collect systematically for CMIP6.

Contributors

- V. Balaji
- Eric Maisonnave
- Uwe Fladrich
- Joachim Biercamp
- Marie-Alice Foujols
- Arnaud Caubel
- Niki Zadeh
- Rusty Benson
- Jeff Durachta
- Garrett Wright
- Seth Underwood
- Bryan Lawrence
- Giovanni Aloisio, and others...

Anyone else who wishes to contribute to the preliminary study is welcome! GMD paper in the works...