HPC for climate models: Lessons from IS-ENES projects

Sylvie JOUSSAUME, CNRS, Institut Pierre Simon Laplace, Coordinator of IS-ENES2
A network of European groups in climate/Earth system modelling

Launched in 2001 (MOU)

Ca 50 groups from academic, public and industrial world

Main focus:
discuss strategy
to accelerate progress in climate/Earth system modelling and understanding

Several EU projects
ENSEMBLES, COMBINE, EUCLIPSE, EMBRACE, SPECS
PRISM, METAFORE, IS-ENES (1& 2)
Collaboration with PRACE

IS-ENES
Infrastructure for ENES

FP7 European projects
IS-ENES 2009-2013
IS-ENES2 2013-2017

Infrastructure
Models & their environment
Model data (ESGF)
Interface with HPC ecosystem

Users:
Climate modelling community (Global & regional)
Impact studies
## Earth System modelling in Europe

### CMIP5 in Europe
- 7 European modelling groups
- 17 models

<table>
<thead>
<tr>
<th>NCC</th>
<th>NorESM1-M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NorESM1-ME</td>
</tr>
<tr>
<td>MPI-M</td>
<td>MPI-ESM-LR</td>
</tr>
<tr>
<td></td>
<td>MPI-ESM-MR</td>
</tr>
<tr>
<td></td>
<td>MPI-ESM-P</td>
</tr>
<tr>
<td>MOHC</td>
<td>HadCM3</td>
</tr>
<tr>
<td></td>
<td>HadGEM2-A</td>
</tr>
<tr>
<td></td>
<td>HadGEM2-CC</td>
</tr>
<tr>
<td></td>
<td>HadGEM2-ES</td>
</tr>
<tr>
<td>EC-EARTH</td>
<td>EC-EARTH</td>
</tr>
<tr>
<td>IPSL</td>
<td>IPSL-CM5A-LR</td>
</tr>
<tr>
<td></td>
<td>IPSL-CM5A-MR</td>
</tr>
<tr>
<td></td>
<td>IPSL-CM5B-LR</td>
</tr>
<tr>
<td>CNRM-CERFACS</td>
<td>CNRM-CM5</td>
</tr>
<tr>
<td>CMCC</td>
<td>CMCC-CESM</td>
</tr>
<tr>
<td></td>
<td>CMCC-CM</td>
</tr>
<tr>
<td></td>
<td>CMCC-CMS</td>
</tr>
</tbody>
</table>

### CMIP5 Evaluate/Understand/Projections
- 3400 simul. yrs up to > 12000 yrs
- 50 expts up to > 160 expts
- 1000 – 3000 Tbytes (CMIP3: 36)

- 29 modelling groups
- 61 models

### IPCC AR5
Earth System modelling in Europe

![Diagram of Earth System modelling in Europe]

**EC-Earth**
- **France**
  - IPSLCM5
  - CNRM-Cerfacs
- **Germany**
  - MPI-ESM
- **Italy**
  - C-ESM
- **UK**
  - HadGEM2
- **Norway**
  - NorESM

**Institut Pierre Simon Laplace**

**Enes**

**Table**

<table>
<thead>
<tr>
<th>Country</th>
<th>name of model (CMIP5)</th>
<th>Atmosphere</th>
<th>Ocean</th>
<th>Sea Ice</th>
<th>Coupler</th>
<th>Land Surface *Vegetation</th>
<th>Atmospheric Chemistry</th>
<th>Ocean Bio-geochemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Consortium</strong></td>
<td>EC-EARTH</td>
<td>IFS</td>
<td>NEMO</td>
<td>LIM</td>
<td>OASIS</td>
<td>HITESSEL</td>
<td>TM5</td>
<td>PISCES</td>
</tr>
<tr>
<td><strong>France</strong></td>
<td>IPSLCM5</td>
<td>LMDz</td>
<td>NEMO</td>
<td>LIM</td>
<td>OASIS</td>
<td>ORCHIDEE</td>
<td>INCA</td>
<td></td>
</tr>
<tr>
<td><strong>France</strong></td>
<td>CNRM-Cerfacs</td>
<td>ARPEGE</td>
<td>NEMO</td>
<td>GELATO</td>
<td>OASIS</td>
<td>SURFEX</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td>MPI-ESM</td>
<td>ECHAM5</td>
<td>MPIOM</td>
<td>MPIOM</td>
<td>OASIS</td>
<td>JSBACH*</td>
<td>HAM</td>
<td></td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td>C-ESM</td>
<td>ECHAM5</td>
<td>NEMO</td>
<td>LIM</td>
<td>OASIS</td>
<td>SILVA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UK</strong></td>
<td>HadGEM2</td>
<td>UM</td>
<td>UM</td>
<td>CICE</td>
<td>OASIS</td>
<td>TRIFFID*</td>
<td>UKCA</td>
<td></td>
</tr>
<tr>
<td><strong>Norway</strong></td>
<td>NorESM</td>
<td>NCAR</td>
<td>MICOM</td>
<td>CICE</td>
<td>CPL7</td>
<td>CLM</td>
<td>Chemistry</td>
<td>HAMOCC</td>
</tr>
</tbody>
</table>

EC-Earth Con Netherlands, Sweden, Ireland, Denmark, Spain, Portugal, Italy, Belgium
IS-ENES: Infrastructure for ENES
FP7 project « Integrating Activities »

1st phase: March 2009 - Feb 2013 (7.6 M€), 18 partners
2nd phase: Apr 2013 - March 2017 (8 M€), 23 partners

http://is.enes.org/

Support to international coordinated experiments for IPCC

Better understand and predict climate variability & changes
Foster:

- The integration of the European ESM community
- The development of ESMs and their environment
- High-end simulations
- The application of ESM simulations for climate change impacts

CMIP5
7 european models

CORDEX
Euro-cordex
Med-cordex
Africa Cordex
Infrastructure Strategy for the European Earth System Modelling Community 2012-2022

Drivers: Science & Society
From understanding to development of “Climate Services”

Grand challenge: towards global 1 km climate models
(resolve convective clouds)

Recommendations:
1) Access to world-class HPC for climate - «tailored» up to «dedicated»
2) Develop the next generation of climate models
3) Set up data infrastructure (global and regional models) for large range of users from impact community
4) Improve physical network (e.g. link national archives)
5) Strengthen European expertise and networking
HPC for climate models

HPC facilities
Resolution, Complexity, Ensembles, Duration
Tier 1: Mainly national facilities, dedicated or general-purpose
Tier 0: projects on PRACE
Towards an agreement with PRACE for high-end experiments for CMIP6
ENES HPC Task Force

Models and their environment
Improve model performance
parallel I/O, coupler, tests
Future model generation:
dynamical cores, numerics, algorithms
Physics (eg radiation)
Share best practices: Workflows

Data
storage, distribution & analyses
Parallel I/O
Efficiency of post-processing: CDO
Distributed database (ESGF):
data & metadata standards, developments

Technology Tracking
Share experience on accelerators
Strengthen interactions with vendors

http://is.enes.org/
IS-ENES HPC Workshops
Lecce 12/2011
Toulouse 01/2013
Hamburg, 03/ 2014
ENES Workshops on HPC for climate models

2nd Workshop Toulouse, 30/01-01/02/2013
Model performance, dynamical cores, use of PRACE
EU exascale projects

Improve model performance:
- to be driven by science
- Performance intercomparison needed: identify & share best practices
- Need interdisciplinary teams: climate and computational

Technology tracking: not convinced of GPUs

Prepare future models
- Need for new dynamical cores to enhance scalability
- Separate science from technical software?

Data challenge: exabyte even more challenging than exaflops

BAMS, workshop report, André et al., May 2014
Scalability issue

Limitation of Scalability

e.g. at resolution 25-30 km for the atmosphere

G. Riley et al., IS-ENES (2012)
Model performance:
Need to revisit dynamical cores

Collaboration NCAR-Sandia, Dennis et al. (IJ HPC appl, 2012)

Europe: 3 projects
ICON (DE)
DYNAMICO (FR)
LFRIC (UK)

Cour. T. Dubos et Y. Meurdesoif (IPSL)

On-going international projects:
G8 exascale project ICOMEX
Dynamical Core MIP

Cubed-sphere (CAM-SE)

CESM1, 0.25°, BGP

CAM SE 0.25°

CMIP5

CMIP3

Low resolution Aquaplanet with full physics

DYNAMICO

At 1/3° 28 SYPD 184 000 cores
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.
<table>
<thead>
<tr>
<th>Model</th>
<th>Resolution</th>
<th>Cmplx.</th>
<th>SYPD</th>
<th>CHSY</th>
<th>Coupler</th>
<th>Load Imb.</th>
<th>I/O</th>
<th>MBoat</th>
<th>ASYPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM2.6 S</td>
<td>A0.5L32 O0.1L50</td>
<td>18</td>
<td>2.2</td>
<td>212,465</td>
<td>5.71%</td>
<td>20%</td>
<td></td>
<td>12%</td>
<td>1.6</td>
</tr>
<tr>
<td>CM2.6 T</td>
<td>A0.5L32 O0.1L50</td>
<td>18</td>
<td>1.1</td>
<td>177,793</td>
<td>1.29%</td>
<td>60%</td>
<td>24%</td>
<td>12%</td>
<td>0.4</td>
</tr>
<tr>
<td>CM2.5 T</td>
<td>A0.5L32 O0.25L50</td>
<td>18</td>
<td>10.9</td>
<td>14,327</td>
<td>17%</td>
<td>0%</td>
<td></td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td>FLOR T</td>
<td>A0.5L32 O1L50</td>
<td>18</td>
<td>17.9</td>
<td>5,844</td>
<td>0%</td>
<td>57%</td>
<td>5.1%</td>
<td>31%</td>
<td>12.8</td>
</tr>
<tr>
<td>CM3 T</td>
<td>A2L48 O1L50</td>
<td>124</td>
<td>7.7</td>
<td>2.974</td>
<td>0.5%</td>
<td>41%</td>
<td>14.76%</td>
<td>3%</td>
<td>4.9</td>
</tr>
<tr>
<td>ESM2G S</td>
<td>A2L24 O1L50</td>
<td>63</td>
<td>36.5</td>
<td>279</td>
<td>8.91%</td>
<td>1%</td>
<td></td>
<td>34%</td>
<td>25.2</td>
</tr>
<tr>
<td>ESM2G T</td>
<td>A2L24 O1L50</td>
<td>63</td>
<td>26.4</td>
<td>235</td>
<td>2.63%</td>
<td>22%</td>
<td></td>
<td>34%</td>
<td>11.4</td>
</tr>
</tbody>
</table>

These are very preliminary results (many thanks: Rusty Benson, Seth Underwood, Niki Zadeh!) but seem to provide a basis for analyzing results across models and platforms.
Technology tracking: accelerators

Rich Loft (NCAR): missing x 20 at 2020 for 1 km
Scalability, Memory,
Node performance (accelerators?)

Review by W. Sawyer

Talks on Xeon Phi
Still difficult to optimise

Summary, final thoughts

- Disappointment depends on your expectations...
- Numerous efforts in porting to GPUs, many are demonstrative and consists of bits and pieces
- Dynamics: GPU performance mirrors increase in memory bandwidth, e.g., 2x for K20x vs. dual-socket SNB
- Physics: increased computational intensity gives larger performance benefit on GPU
- At least one model is near production status on GPU
- Acceptance of GPU paradigms by model developers is a problem

Order of x 2
Better with regards to energy

New ways of structuring codes?
Separating Science/technology layers
Data issues

Data challenge:
exabyte even more challenging than exaflops

Different levels
I/O servers (XIOS, CDI-IO);
Compression; Storage;
Post-processing;
Distributed archive (ESGF)

CMIP5  \approx 2 \text{ to } 3 \text{ PB}
CMIP6  \times 30 \, ?

Earth System Grid Federation

G8 ExArch project
Climate analytics on distributed exascale data archives

Overpeck et al. (Science 2011)
Climate
Launch of next international coordinated experiments CMIP6
National / Europe / International

Prepare next generation of climate models
Dynamical cores …

H2020 opportunities for climate on HPC issues:
Center of Excellence & Future and Emerging Technologies
Feb 27th 2014 meeting with DG Connect on climate and weather
https://ec.europa.eu/digital-agenda/news-redirect/14942

Towards a sustained European infrastructure

Europe
Strategic nature of HPC for Europe
Technology (ETP4HPC)/ HPC ecosystem (PRACE) / Applications (CoE)

Start of H2020 European Framework 2014-2020
Centers of excellence & Emerging technologies (exascale)
Enjoy the meeting!