Development of Library for Future Meteorological/Climate Simulations: SCALE

Computational Climate System Research Team
Hirofumi TOMITA
Outline of my talk

• Introduction of SCALE, NICAM
  – What is NICAM/SCALE?
  – Example of large scale computation on K Computer:
    • SCALE: model-intercomparison, shallow cumulus
    • NICAM: Grand challenge like GCM simulation with very high resolution

• Possibility of collaboration
  – Infrastructure
  – Science
Our main tools: SCALE+NICAM-DC

Physical process library
Reginoal dynamical core
Nishizawa et al. (2015,GMD)
Sato et al. (2015,PEPS)

Global dynamical core
Tomita et al. (2001,2002,JCP)
Tomita & Satoh (2004,FDR)

To be merged!
Our direction / concept of SCALE

- Easy comparison
- Reproducibility
- LES-scale simulation
① Easy model comparison

Model inter-comparison is a key in evaluation of the reliability of the meteorological numerical simulations.

Why model intercomparison is needed?

Estimation of uncertainty of meteorological simulation

• The model is not always based on first-principle.
• The model includes many empirical rules / hypotheses
• The model has many tunable switches, (in physical parameterization)

Difficulty in validation of simulations

• limitation of observations (coverage, resolution, quantity)
• paleo/future climate, or other planets
Inter-model comparisons

- total performance

Intra-model comparison

- individual schemes

We want to have just one model including all key components. => If so, Intra-model comparison is possible?

- Cloud microphysics, cumulus parameterization, radiation process, turbulence, and so on.
- Dynamical cores, e.g.,
  - Discretization schemes
  - Order of accuracy of difference scheme
  - Implicit and explicit temporal integration schemes
- In addition,
  - Tunable parameters
  - Precision of floating point

The difference of model results are easily understood.

1. From which does the difference come?
2. What is a key for representation of target phenomena?
Example: Intra model comparison

RICO experiment (van Zanten et al. 2011)

We can conclude (very very roughly ) that these differences are strongly depending on the cloud-microphysical schemes.

1-moment: The faster precipitation drop: due to saturation adjustment and quick autoconversion.

2-moment: The small and slow precipitation: due to in growth of huge droplet.

*Sato et al. 2015: Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation? PEPS, 2:23.*
② Reproducibility/traceability

Scientific products should be able to be reproduced for the later verification => Reliability.

Openness of code, setting, and results to anyone

- SCALE is available to anyone as an open source software.

Sharing know-how

- Predecessors’ knowledges have often been unpublished. (tuning parameter reasonable limiter of filter etc.)
- In our policy, we publish all knowledge of those, e.g., How does tune the parameter tuning, and how does set limiter.
Several added values are expected in high-resolution simulations (e.g. LES).

If more fundamental physical principles can be used, uncertainty can be expected to be reduced.

- cumulus parameterization -> cloud microphysics
- RANS -> LES

Better representation of extreme detail

- finer topography / surface conditions
- less spatial averaging
Example: Validation of large grid aspect ratio \((dx/dz)\) in LES

Unstable PBL turbulence experiment

- **conventional SGS model:** spurious energy pile due to small mixing length
- **large aspect ratio:** artificial large skewness at the top of the PBL

---

Nishizawa et al. 2015: Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations, GMD, 8, 6021-6094.
Computational performance

performance @ K computer
- above 10% of peak performance (dynamical core)
- 5~8% for full simulation (including I/O)
- Almost perfect weak scaling up to full system (663,552 cores)
- good strong scaling

Weak scaling

Strong scaling
Future Issues to HPC

Still, validity of parameterization should be continued

- How does assumption of parameterizations affects results?
- Easy framework for this is needed.

Computational efficiency should be pursued

- Efficient use of computational resource (week/strong scaling)
- Currently, a key issue is still bandwidth in DC.

Data explosion should be considered

- Better data handling in pre/post processes
- Analysis also should be in parallel.
Current SCALE: if you are interested...

### Dynamics

- **Governing equations**: 3-dimensional fully compressible
- **Grid system**: Arakawa-C type
- **Temporal integration**: HEVE, HEVI, HIVI
- **Temporal difference**: 3 steps Runge-Kutta scheme
- **Spatial difference**: 4th order central difference
- **Topography**: Terrain-following
- **Positive definitive**: FCT scheme

### Physical schemes

- **Cloud microphysics**:
  - Kessler (Kessler, 1969)
  - 1-moment bulk (Tomita et al., 2008)
  - 2-moment bulk (Seiki and Nakajima, 2014)
  - 1-moment bin (Suzuki et al., 2010)
  - super droplet method (Shima et al., 2009, *experimental*)
- **Turbulence**:
  - Smagorinsky SGS (Brown et al. 1994, Scotti et al. 1993)
  - MYNN level 2.5 (Nakanishi and Niino 2004)
- **Cumulus parameterization**:
  - Kain-Fritsch (*in preparation*)
- **Radiation**:
  - MSTRN-X (Sekiguchi and Nakajima, 2008)
- **Aerosol microphysics**:
  - 3-moment bulk (Kajino et al., 2013, *experimental*)
- **Surface flux**:
  - Louis-type (Uno et al. 1995)
  - Beljaars-type (Beljaars and Holtslag 1994, Wilson 2001)
- **Land**:
  - Slab model with a bucket model
- **Ocean**:
  - Slab ocean model
- **Urban**:
  - Single-layer urban canopy model (Kusaka et al., 2001)

### Other

- Offline/Online nesting system
- LETKF assimilation system
**Challenge! (explicit expression of cloud)**

Our research community (NICAM research community)’ approach: Resolve the cloud system & related process over the globe

---

### NICAM development: ~2000

**still development is continuing!**

---

**Conceptual development philosophy**

- **Explicit resolving the cloud itself**

  - **Use of Icosahedral grid**
    - To get a quasi-homogeneous grid for computational efficiency
  
  - **nohydrostatic DC**
    - To resolve cloud scale (deep convection, shallow cloud etc.)

- **Sophistication of cloud expression:**
  - To avoid the ambiguity of cumulus parameterization and understand the cloud dynamics
Recent results on K computer (two landmark works)

---

Super-high simulation: sub-km grid spacing (Miyamoto et al. 2013, 2014 GRL, ASL): capability computing
Many ensembles by GCRM: MJO predictability (Miyakawa et al. 2014 Nature comm.): capacity computing

Grand Challenge on K computer!

- Horizontally 860m resolution, vertically 100 levels
  - Use of ¼~full system of K-computer
  - First ever simulation with sub km horizontal grid AGCM.
- Purpose
  - One reference solution to coarser grid simulation.
    - How is the convergence?!
  - Computationally, check the scalability at the use of full resource.
- Scientific scope:
  - How is Global “picture” of deep convections?

---

Why challenging? What is a challenge?

Even current High-end machine, sub-km GCM may be a demonstration simulation: However, next generation HPC enable us to integrate the long time simulation.
A snapshot of sub-km AGCM (NICAM)

Horizontal: Δ0.87 km: vertical 100 levels: integration time 24h
Convergence of 1. number of convection 2 distance of neighboring convection


(a) number of convection

\[ \Delta x \geq 3.5 \text{ km:} \]
- # of conv.: increase by factor of 4
- Conv. distance between convection: 4 grids => unphysical?

(b) distance between convection

\[ \Delta x \leq 1.7 \text{ km:} \]
- # of conv.: decrease in increasing rate
- Conv. distance: > 5 grids => close to the nature

Convection features (structure, number, distance)

change between \( \Delta 3.5 \text{ km} \) \( \cong \) \( \Delta 1.7 \text{ km} \)
- \( \Delta x \) should be 2.0〜3.0 km to resolve convection in global models

Resolution of 2km is tipping point!
Performance efficiency

- Just after porting from ES: ~4%
- Cache optimization to stencil operators: ~5%
- Cleaning the time-wasting codes: ~7%
- Modify conditional branches, refactoring: ~10%

Weak scaling test

- Same problem size per node, same steps
- Good scalability

Efficiency of NICAM on K Computer

H. Yashiro (RIKEN/AICS)
Where shall we go after this?

1. Higher resolution
2. Much more ensembles
3. More Sophisticated physics

Ultimately, one direction,..., and challenging issue is higher resolution.

GCRM => GLES?!
<table>
<thead>
<tr>
<th>Resolution Grid interval/level</th>
<th>Total FLOP for 1-day simulation</th>
<th>Machine</th>
<th>Efficiency (%)</th>
<th>Elapse time for 1-day simulation</th>
<th>Elapse time for 1 month simulation</th>
<th>What’s resolved? What is meaningful for scientific advance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5km/L40</td>
<td>230P</td>
<td>131TFLOPS (ES2)</td>
<td>15%</td>
<td>3.2hour</td>
<td>4day</td>
<td>Meso-scale convection system. Cold pool dynamics</td>
</tr>
<tr>
<td>800m/L100</td>
<td>36800P</td>
<td>10PFLOPS (K computer)</td>
<td>10%</td>
<td>10hours</td>
<td>12.5days</td>
<td>Convection resolving?</td>
</tr>
<tr>
<td>400m/L100</td>
<td>295000P</td>
<td>1EFLOPS</td>
<td>10%</td>
<td>50min</td>
<td>24days</td>
<td>Definitely convection resolving(expected)</td>
</tr>
<tr>
<td>200m/L100</td>
<td>2360E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Breakthrough does not exits. But good expression of deep cloud</td>
</tr>
<tr>
<td>100m/L100</td>
<td>18880E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Insufficient for LES</td>
</tr>
<tr>
<td>50m/L200</td>
<td>302Z</td>
<td>100EFLOPS</td>
<td>10%</td>
<td>50min</td>
<td>24hour</td>
<td>Global LES??</td>
</tr>
</tbody>
</table>

We are here

Exa scale era

Tentative goal?

Assumption: sustained performance 10% (we wish)
Possible Collaboration issue
~ along introduction of our team mission~
Direction of our research in AICS in next 5 years (Candidates of collaboration)

• **Infrastructure:**

  – **Extension of basic library SCALE:**

    • **User friendly library**
      – How does standard interface determine?
        » Exchange of subroutine level is very useful for model inter or intra comparison.
      – E.g. CBLEAM activity in Japan (initiated by AICS our team)

    • **Massive parallel analysis routines for acceleration of scientific output, social outcome**
      – Not only acceleration of simulation itself but also acceleration of analysis phase:
      – Adding them to SCALE

  • **Easy programing and high performance computing:**
    – DSL(Domain Specific Language)? e.g. stencil DSL?
Direction of our research in AICS in next 5 years
(Candidates of collaboration)

- **Science:**
  - **BIG DATA assimilation:**
    - Now, developing....
      - NICAM + LETKF (with DA research team & post K priority subject)
        » Many satellite data is available.
        » One goal: Reanalysis data by cloud resolving model
      - SCALE+LETKF (with DA research team)
        » PA data provides tremendous information in time and space.
        » We are tackling to each cumulus with 30min lead time
  - **Reginal Climate assessment! : downscale to city level**
    - Disaster prevention and mitigation, adaptation
      - Multi-model ensemble (SCALE can do it!) drastically reduce the uncertainties for the future climate assessment in the regional model
      - Model bias reduction by data assimilation
        » e.g. Determination of unknown parameters
  - **Planetary science**
    - Generalization of earth knowledge
  - **Theoretical issuuu**
    - Moist LES theory
風龍：本名吉田龍二、理研AICS複合系気候科学チーム所属、博士（理学）
2011年彗星のごとく現れ、京を用いた計算の可視化において、数々の名作を生み出していた。2014年学位取得後、その技にますますの磨きがかかり、業界（？）でも引く手あまた。今後の活躍（研究も？）が期待される若手のホープである。