High-resolution modeling and big data analysis at RIKEN AICS

S. Nishizawa,

H. Yashiro, Y. Sato, Y. Miyamoto, R. Yoshida,

T. Yamaura, S. A. Adachi, Y. Kajikawa, and H. Tomita

RIKEN Advanced Institute for Computational Science

SCALE
Scalable Computing for Advanced Library and Environment

Cartesian dynamical core, physical processes

For regional weather/climate simulations

Nishizawa et al (2015), Sato et al. (2015)

NICAM DC

Icosahedral dynamical core

For global climate simulations

Tomita et al. (2001,2002), Tomita and Satoh (2004)

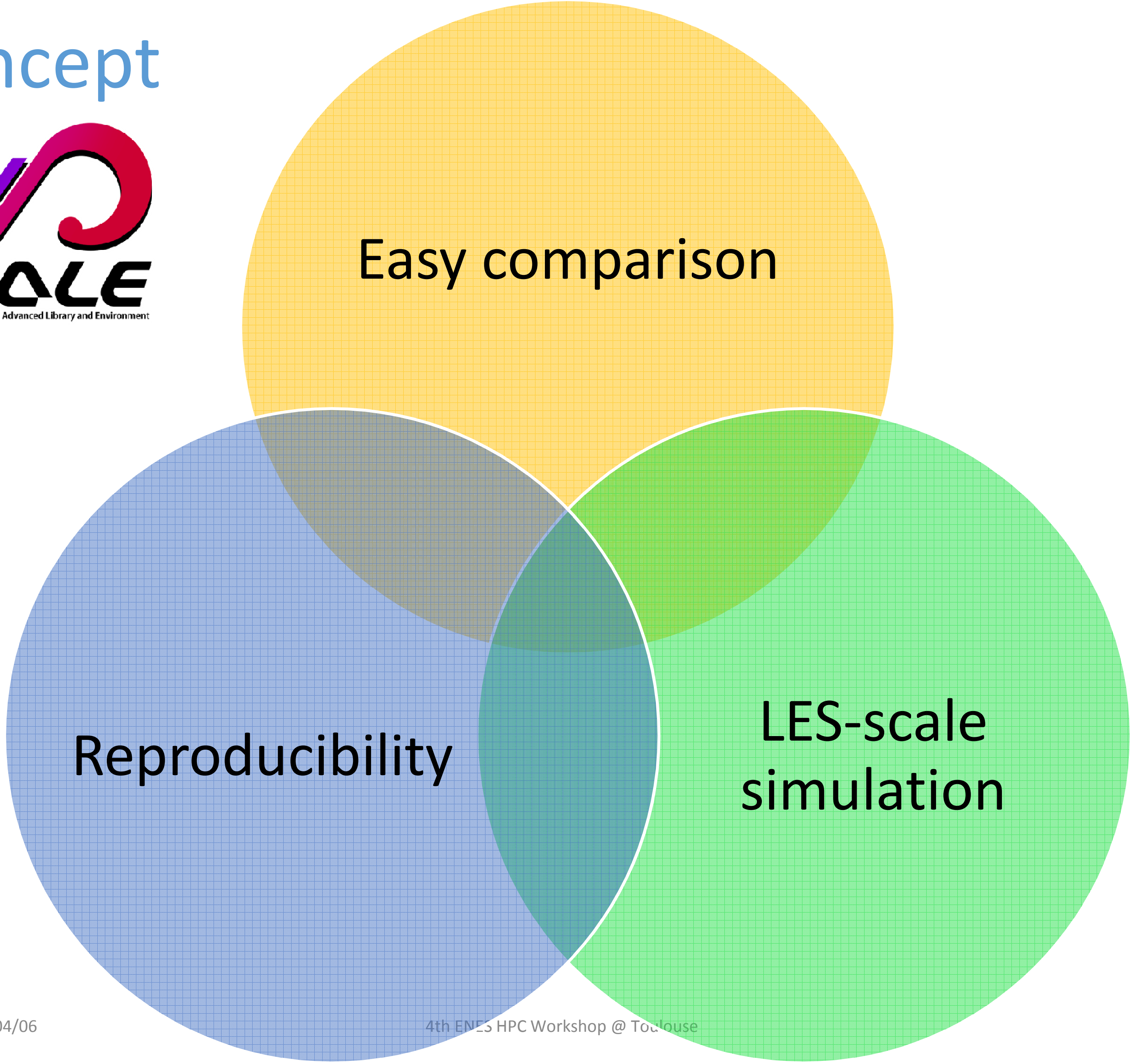
To be merged!

SCALE
Scalable Computing for Advanced Library and Environment
http://scale.aics.riken.jp/

SCALE

Scalable Computing for Advanced Library and Environment

Concept



Reproducibility

Scientific products should be able to be replicated for verification and reliability.

Openness

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

Sharing know-how

- Predecessors' undocumented knowledges have often been lost.
- We try to publish knowledge of our experiences, e.g., parameter tuning, limiter…

Easy Comparison

Comparison is a key in evaluation of the reliability of the meteorological numerical simulations.

Uncertainty of meteorological simulation

- not a first-principle simulation
- many empirical rules / hypotheses
- tones of tunable switches

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

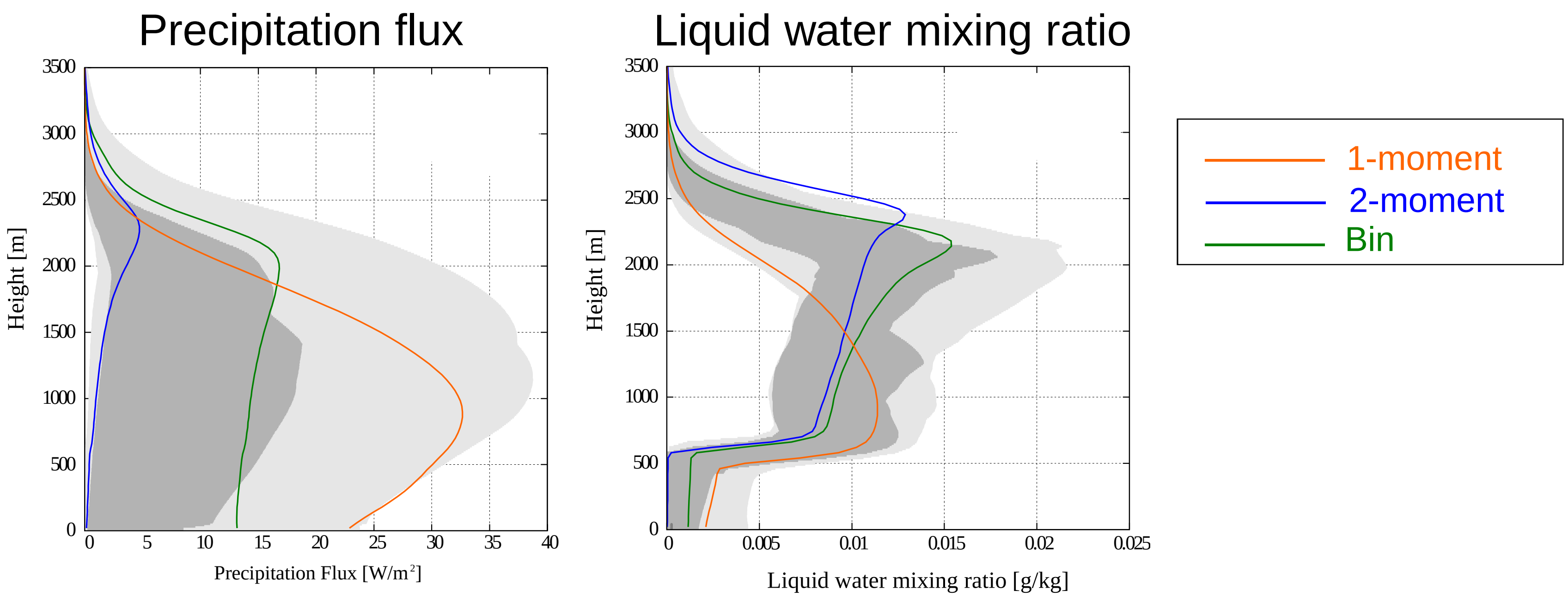
individual schemes

- physical processes, e.g.,
  - cloud microphysics: one/two moment bulk, spectral bin, super-droplet
- dynamical cores, e.g.,
  - discretization schemes
  - order of accuracy of difference scheme
  - implicit and explicit temporal integration schemes
- combination of the schemes
- tunable parameters
- precision of floating point

<u>Differences are relatively easy to be understood</u>

Comparison between cloud-microphysical schemes

RICO experiment (van Zanten et al. 2011)



We can conclude that

these differences are originated from the cloud-microphysical schemes.

<u>1-moment:</u> The faster drop is due to saturation adjustment and quick autoconversion.

2-moment: The small fall velocity is due to difficulty in growth of huge droplet

<u>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.</u>

LES-scale simulations

Several added values are expected
in high-resolution large-eddy simulations.

Smaller uncertainty, On more physical principles

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

Better representation of extremes

- finer topography / surface conditions
- less spatial averaging

Issues

Validity of parameterization

- assumption of parameterizations
- scale-dependent parameters

Computational efficiency

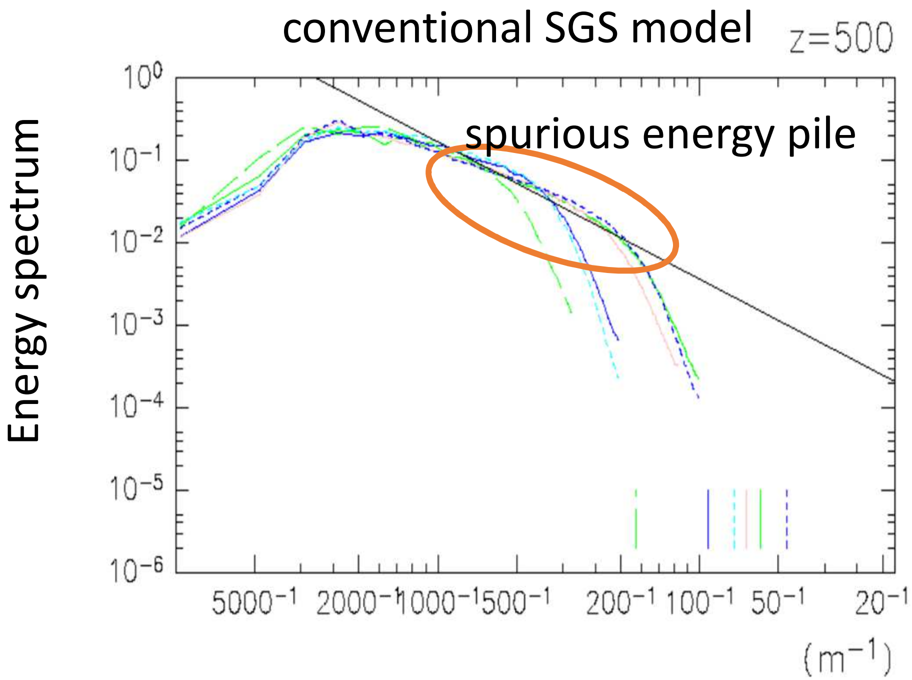
- efficient use of computational resources
- scaling at massive parallel computer

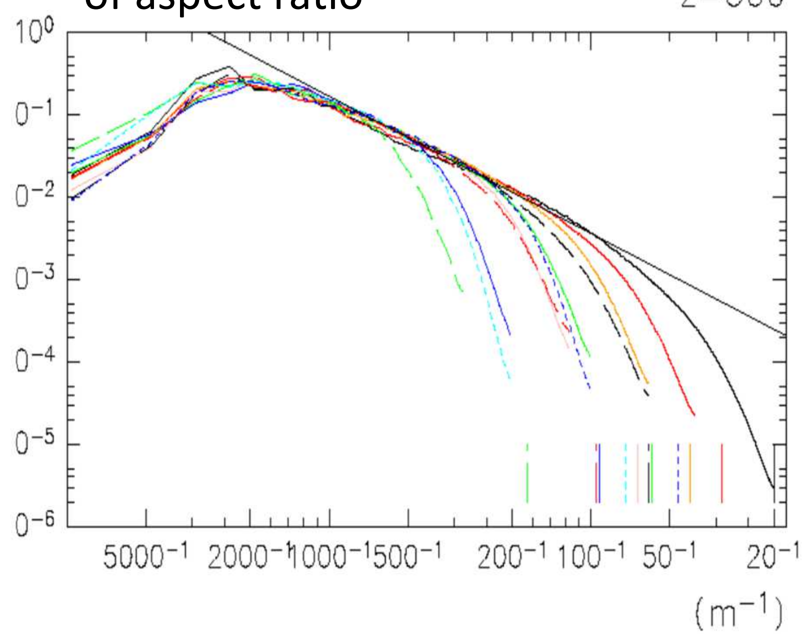
Data explosion

- better data handling in pre/post processes

Validation of large grid aspect ratio (dx/dz) in LES

Unstable PBL turbulence experiment





<u>conventional SGS model:</u> spurious energy pile due to small mixing length

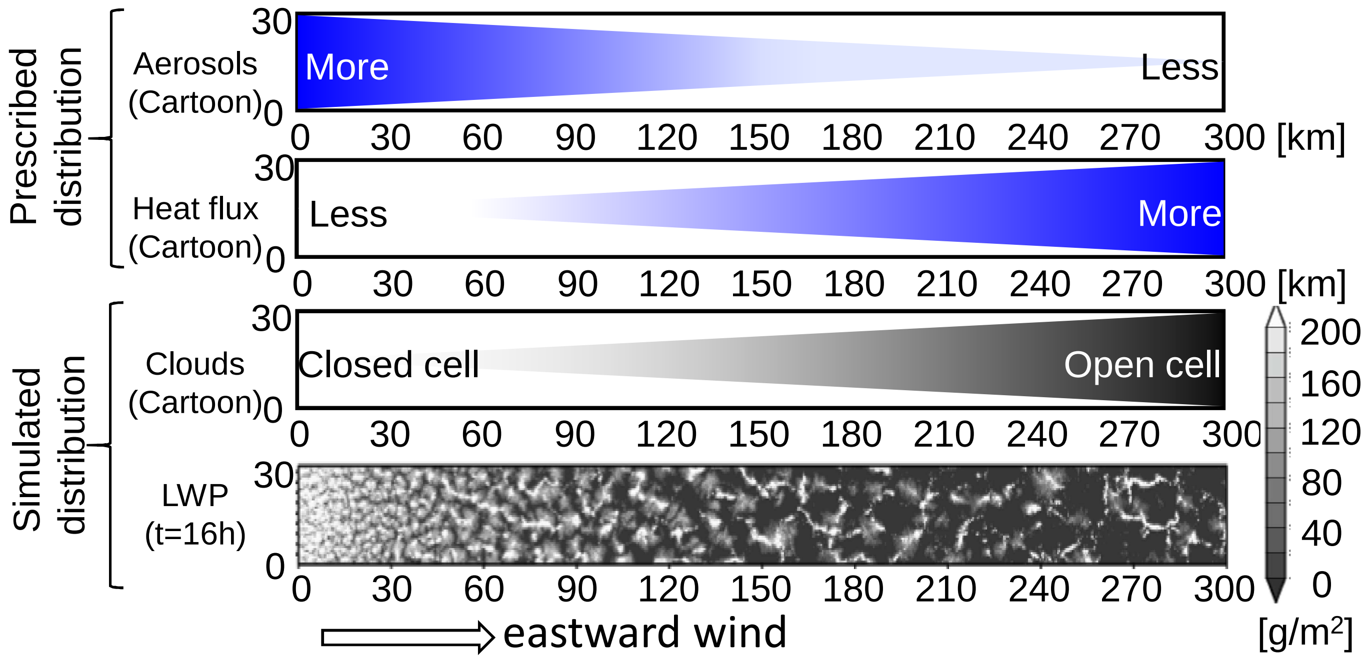
<u>Nishizawa et al. 2015: Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations, GMD, 8, 3393-3419.</u>

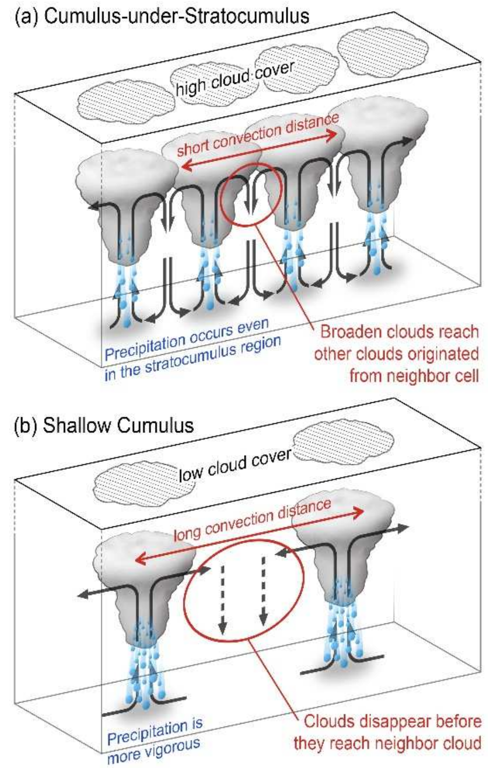
Challenge to meso-scale LES

Huge domain with high resolution LES

- 300 km x 30 km domain with Δx=50 m, 275 layers
  - 1 billion grids
- 16 h integration (dt= 0.01 sec)
  - 138 h with 221,184 cores @K computer
- total 120TB output

Transition from closed to open cell of the stratocumulus





Cloud cover determined by the balance between distance of each cumulus and cloud broadening distance at the cloud

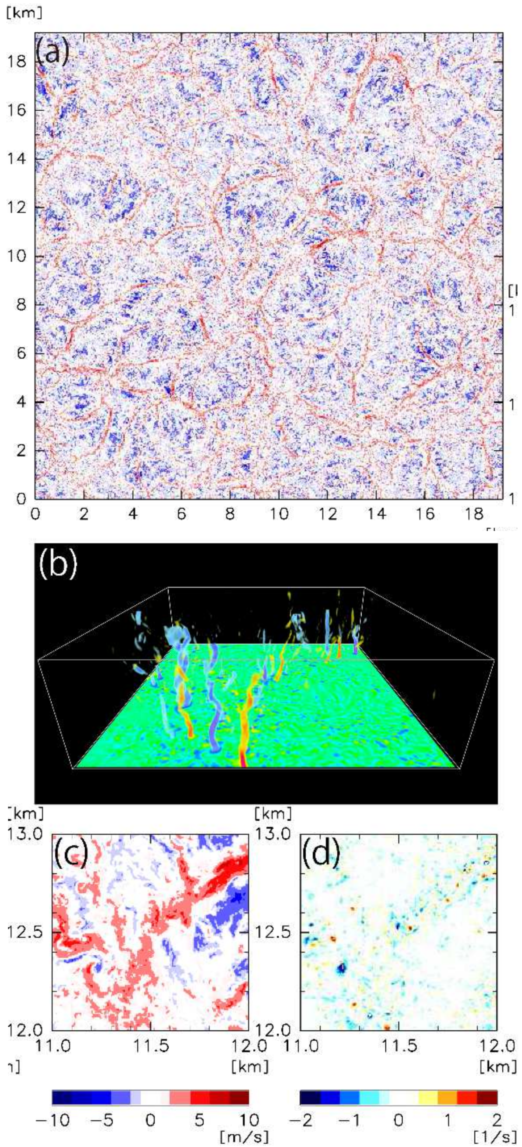
Sato et al. 2015: Horizontal distance of each cumulus and cloud broadening distance determine cloud cover, SOLA, 11, 75-79.

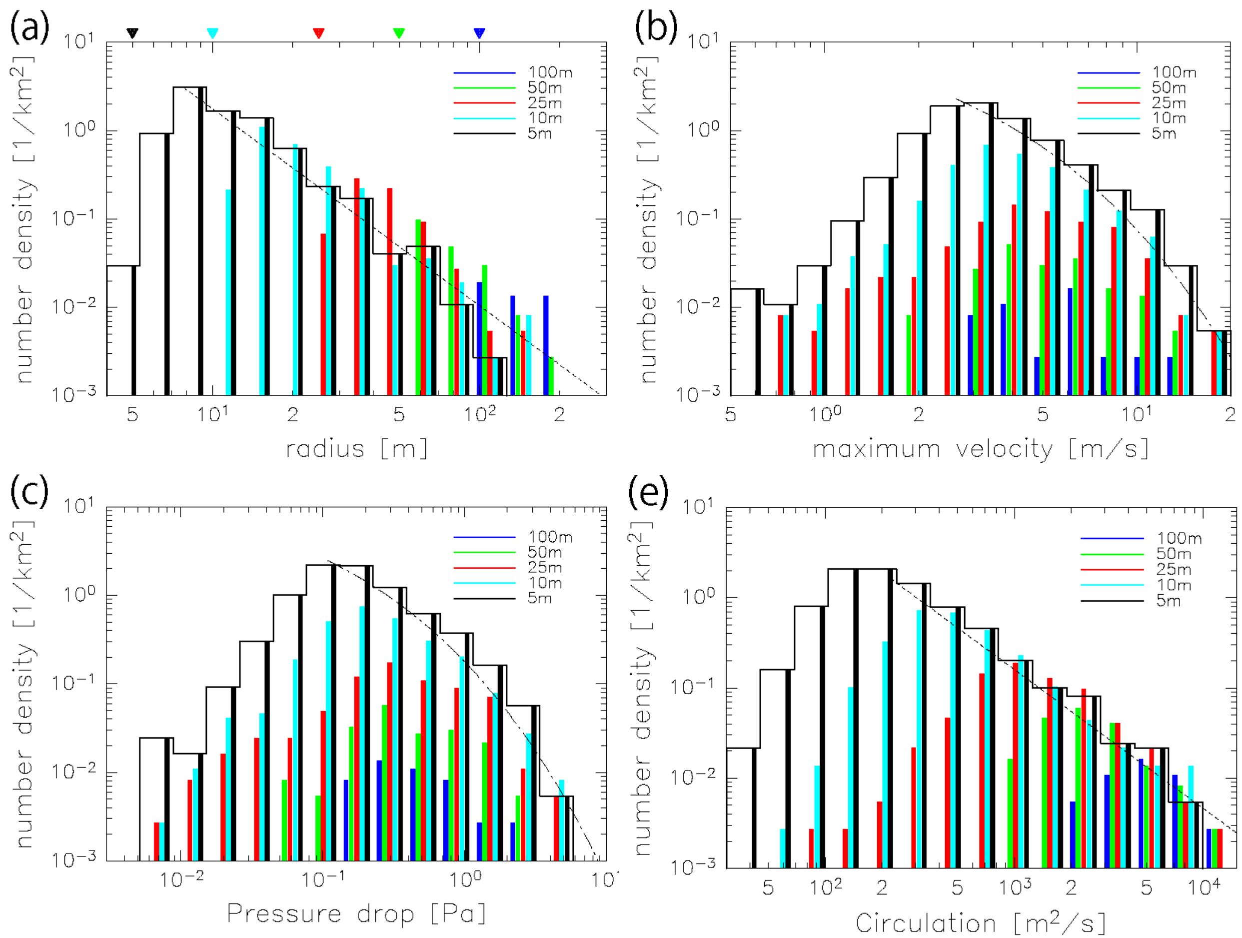
Other planets

Highest resolution on Martian PBL experiment

- 19.8 km^2 domain with Δx=5 m, 3,300 layers
  - 50 billion grids
- 1 h integration (dt= 0.006 sec)
  - 200 h with 57,600 cores @K computer
- total 60TB output

Statistics of Martian dust devils





Frequency distribution of dust devils are identified.

Nishizawa et al.: Martian dust devil statistics from high-resolution large-eddy simulations, GRL, in revision.

Revolutionary super-rapid data assimilation

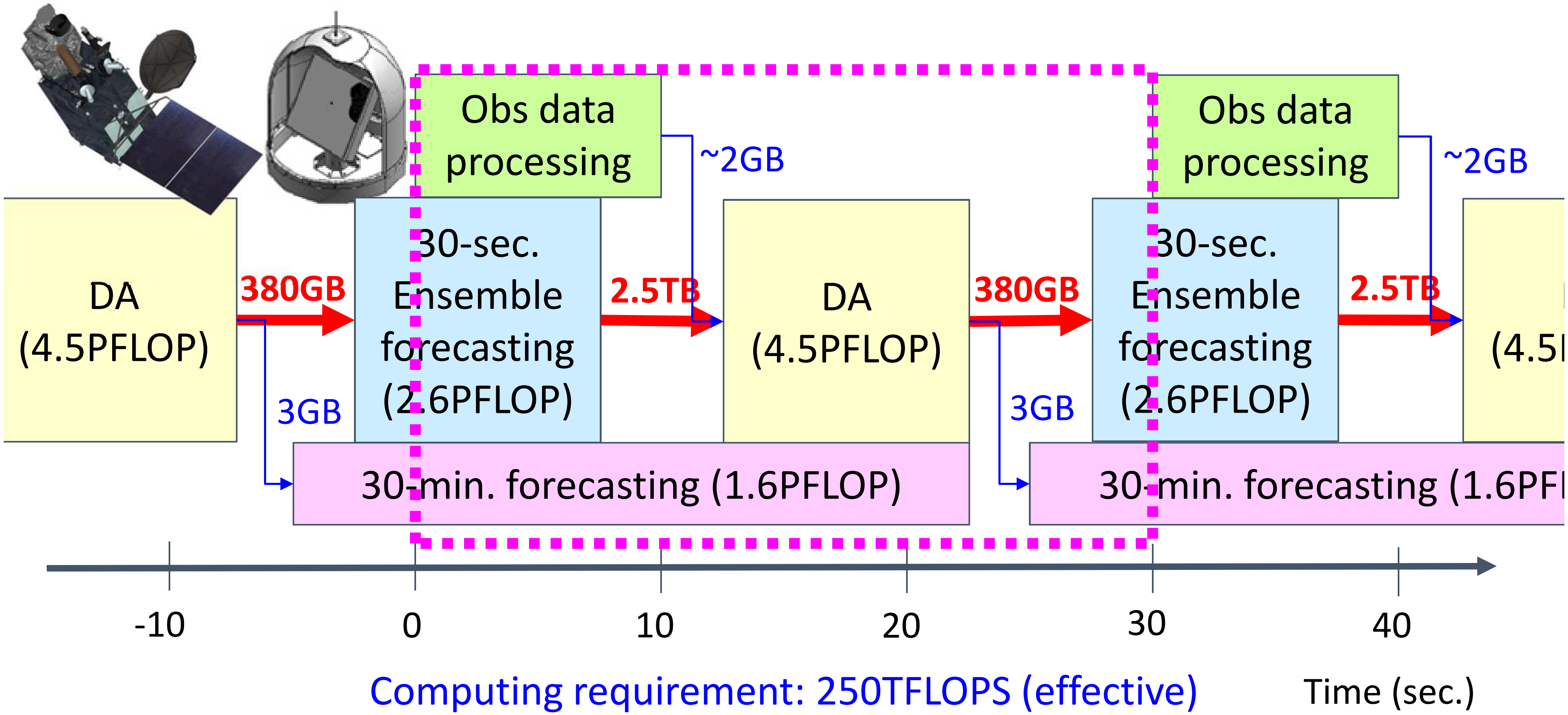
Local Ensemble Transform Kalman Filter ***(Hunt et al. 2007)***

Pinpoint (100-m resol.) forecast of severe local weather by updating 30-min forecast every 30 sec!

collaborate w/ AICS data assimilation Team, JMA, NICT, and Osaka Univ.

Miyoshi et al. : "Big data assimilation" Revolutionizing severe weather prediction, BAMS, accepted.

30-sec. assimilation cycle system



Computing requirement: 250TFLOPS (effective)

Equiv. to 1/4 of the K computer

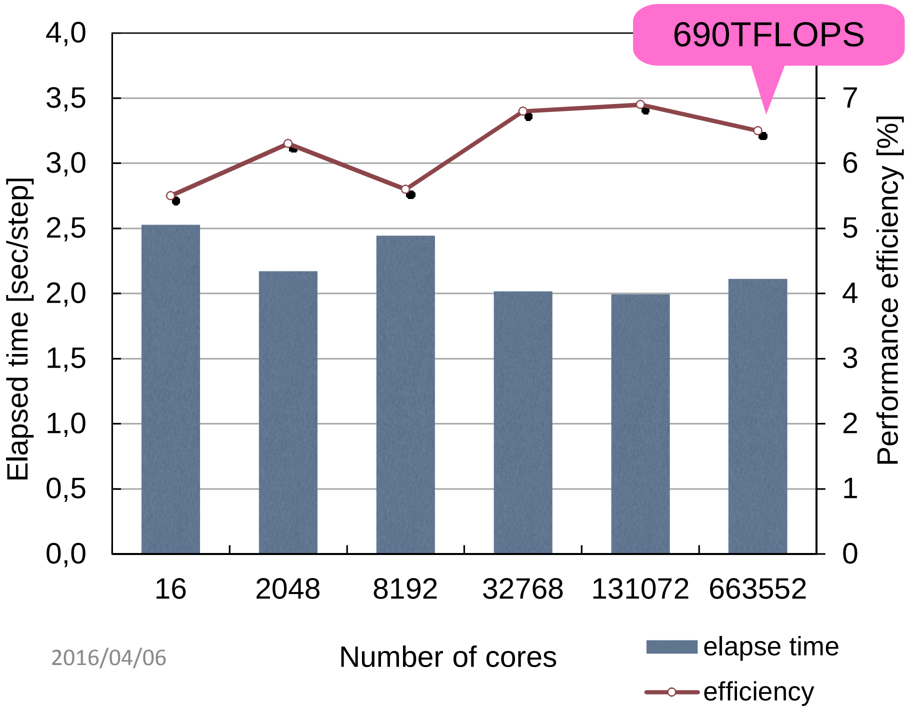
<u>120 times more rapid</u> than hourly update cycles

SCALE Computational performance

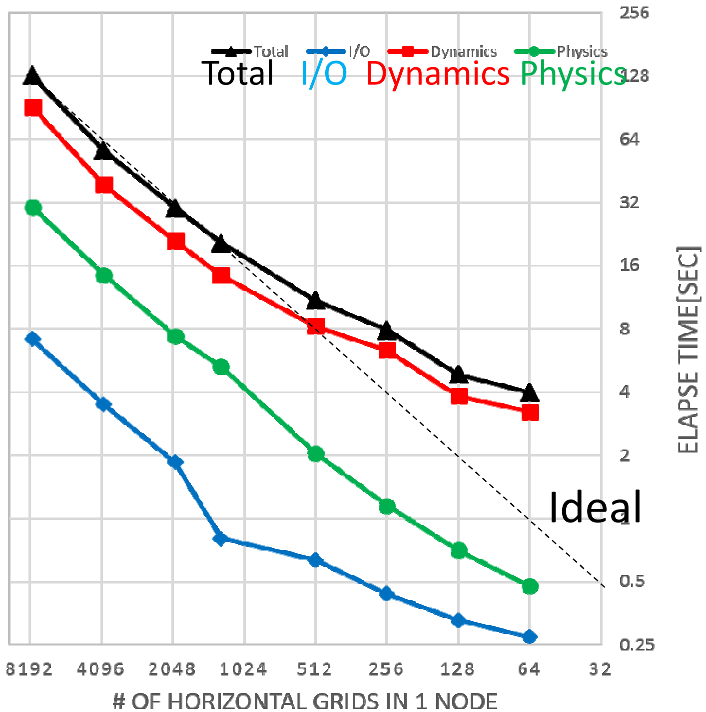
performance @ K computer

- above 10% of peak performance (dynamical core)
  - 5~8% for full simulation (including I/O)
- about 100% weak scaling up to full system (663,552 cores)
- good strong scaling

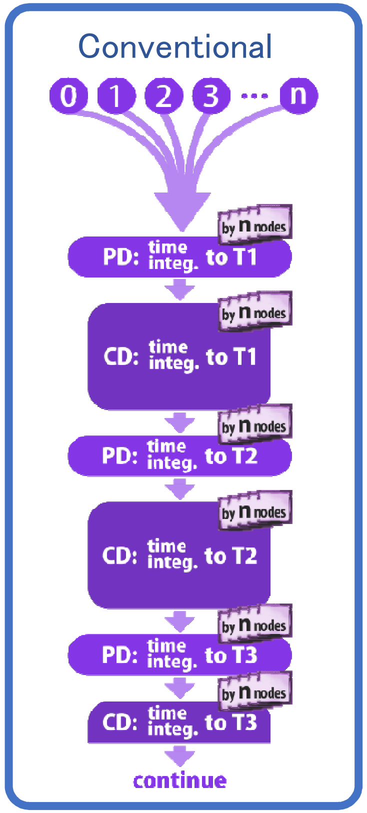
Weak scaling

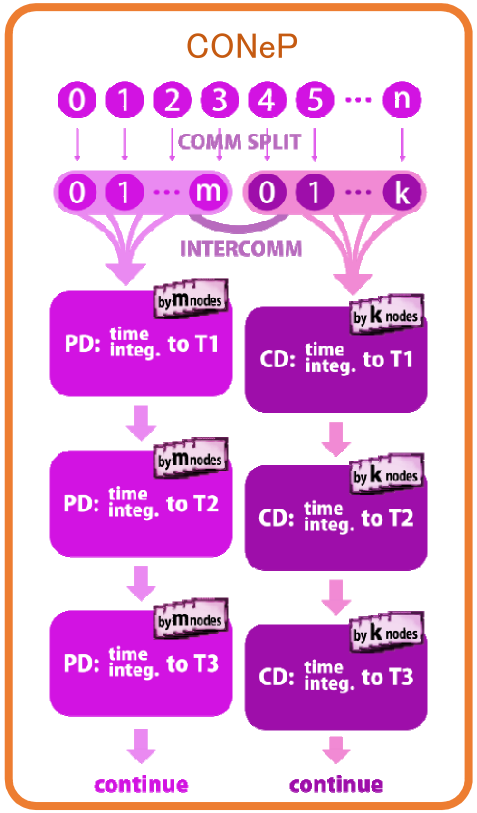


Strong scaling



A Cost-effective Online Nesting Procedure





Common	domain 1	domain 2	domain 3
grid space	27 km	9 km	3 km
#grid (nx, ny, nz)	80, 80, 48	80, 80, 64	80, 80, 80
time interval	27 s	9 s	3 s
time steps	50	150	450

CNV	domain 1	domain 2	domain 3
#node (mx, my)	10, 10	10, 10	10, 10
#grid /tile (lx, ly)	4, 4	6, 6	10, 10

CONeP	domain 1	domain 2	domain 3
#node (mx, my)	2, 2	4, 4	8, 10
#grid/tile (lx, ly)	40, 40	20, 20	10, 8

Performance Experiment on K computer for 1350s time integration

Three domains	CNV	CONeP
elapsed time	20.9 s	16.8 s

20% faster!

Four domains	CNV	CONeP
elapsed time	61.3 s	44.9 s

27% faster!

Yoshida et al.: CONeP: A cost-effective online nesting procedure for regional atmospheric models, Parallel Computing, submitted.

NICAM

Dynamical Core Package

Challenge! (explicit expression of cloud)

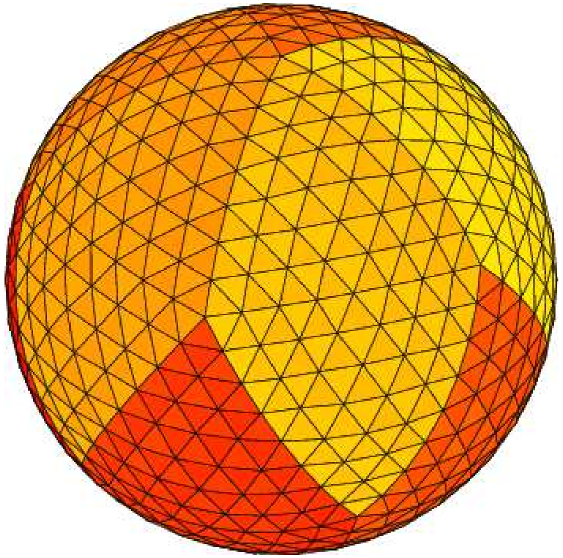
Our research community (NICAM research community)’ approach:
Resolve the cloud system & related process over the globe

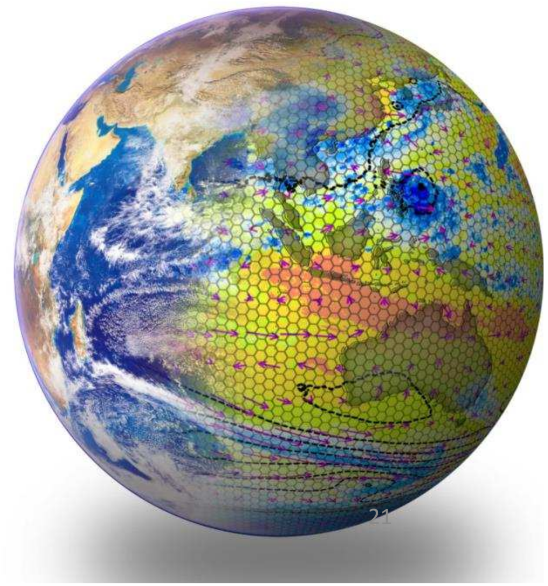
NICAM development : ~2000

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
    - Nonhydrostatic 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





Grand Challenge on the K computer

Sub-km global simulation!

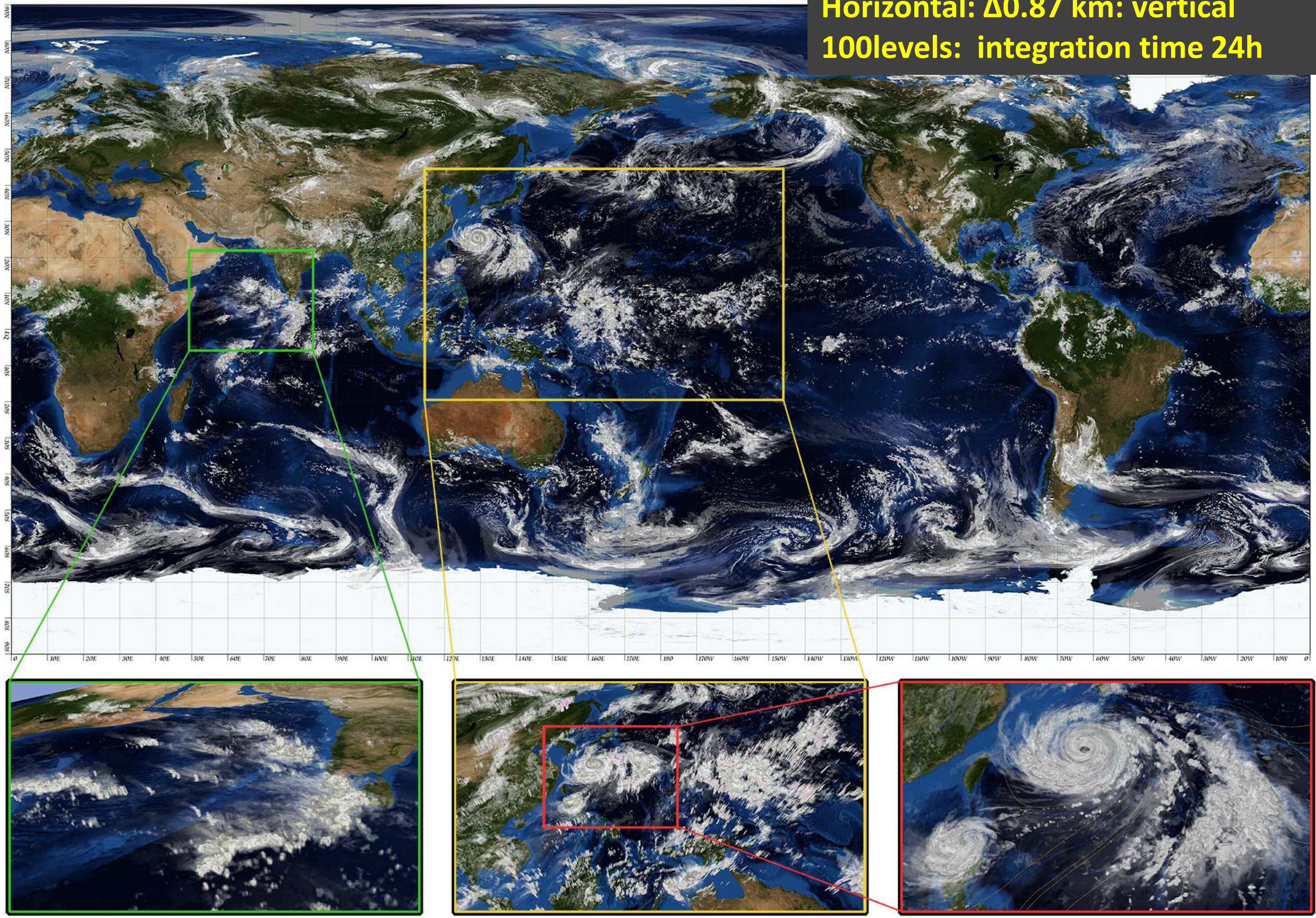
- Δx=870m, 94 layers
  - 63 billion grids
- 48 h integration (dt=2 sec)
  - 220 h with 163,840 cores @K computer
- total 320TB output

- ~~200-day post process on Xeon cluster~~

  ⇒ analysis on the K (163,840 cores)

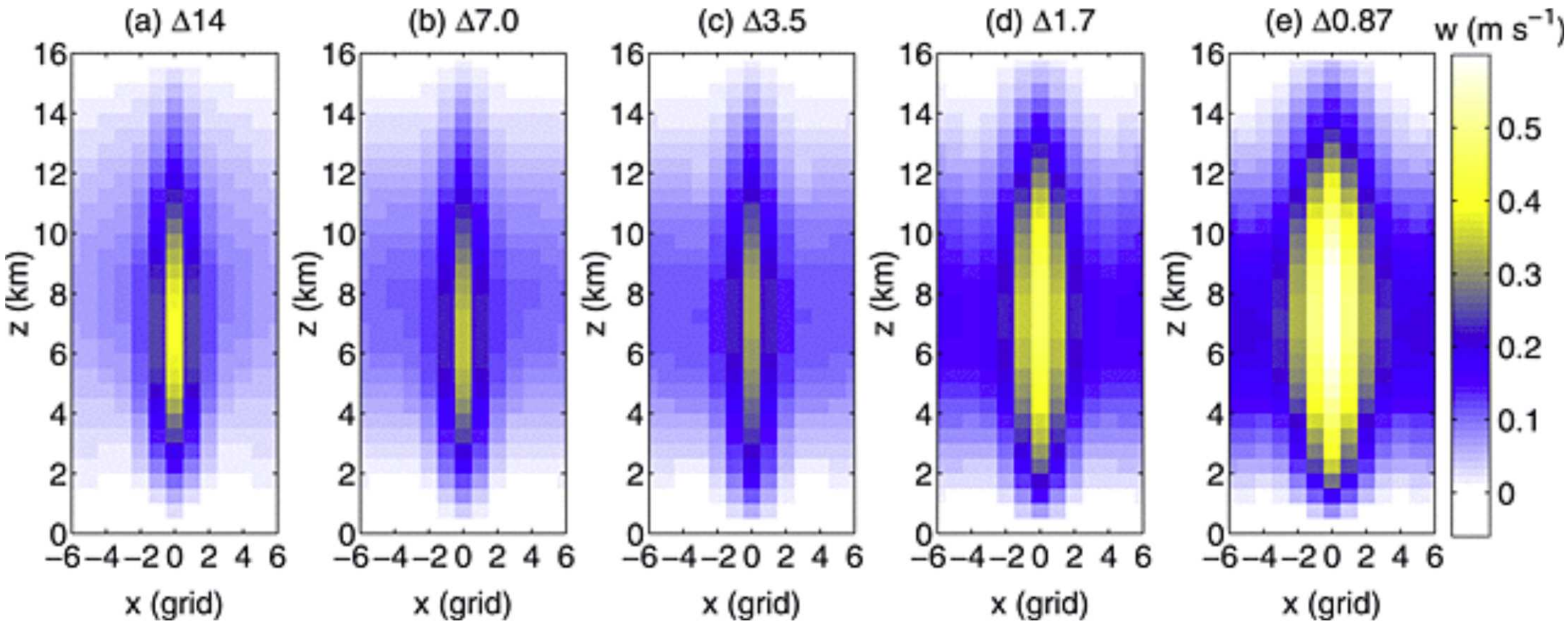
A snapshot of sub-km AGCM

Horizontal: Δ0.87 km: vertical 100levels: integration time 24h



Convergence of convections with resolution

- Global composite of deep convection (vertical velocity)
  - Δx<2km: convection is represented at multiple grids



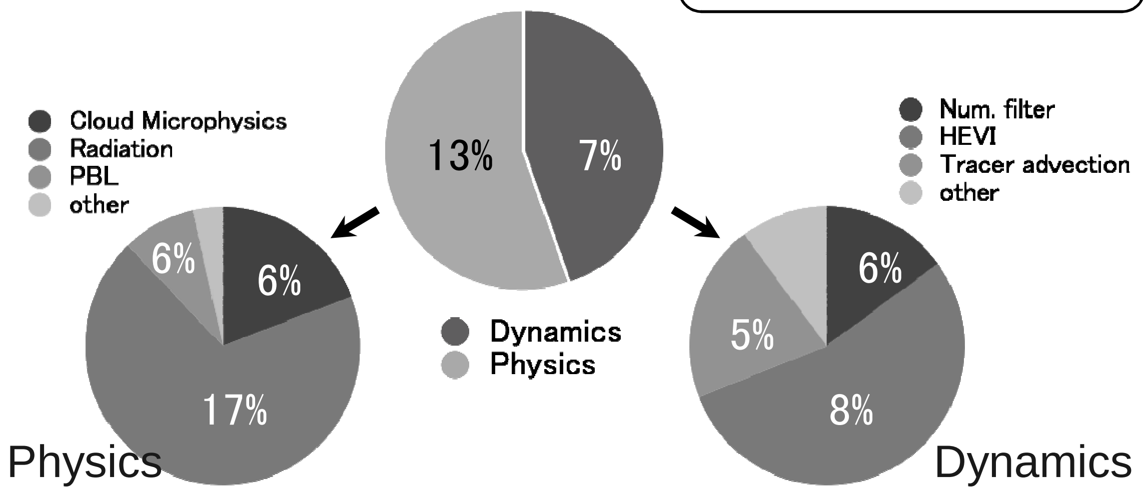
<u>Miyamoto et al. 2013: Deep moist atmospheric convection in a subkilometer global simulation, GRL, 40, 4922-4926.</u>

Efficiency of NICAM on K Computer

H. Yashiro
(RIKEN/AICS)

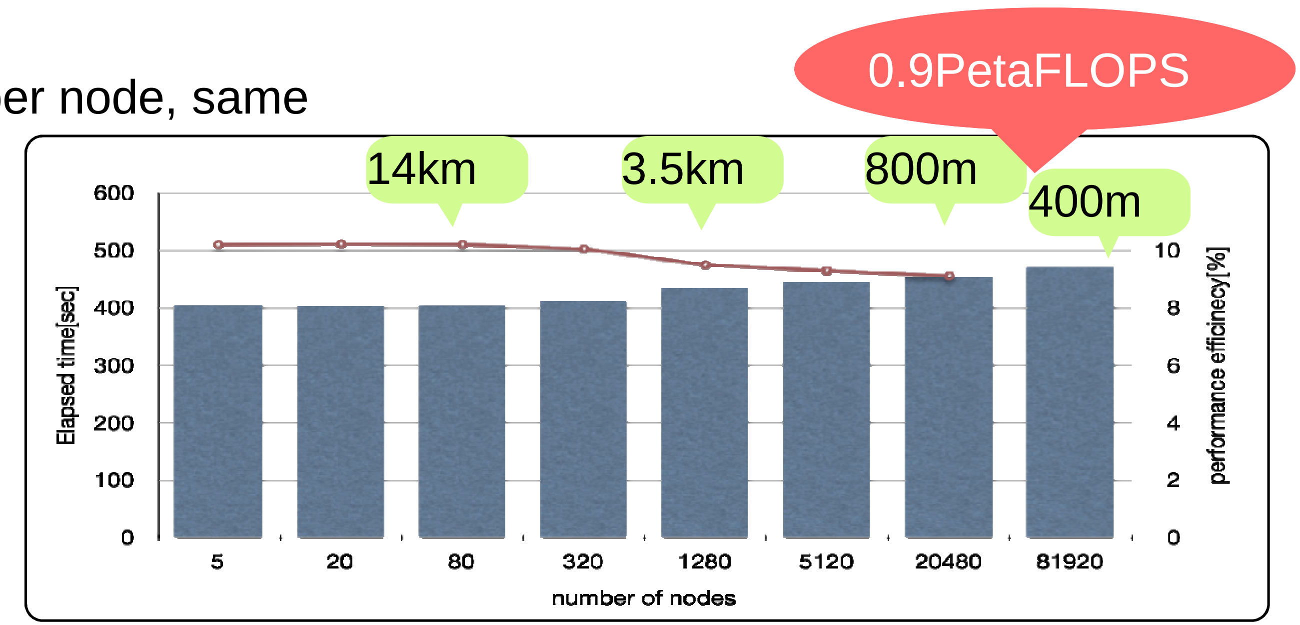
Performance efficiency

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

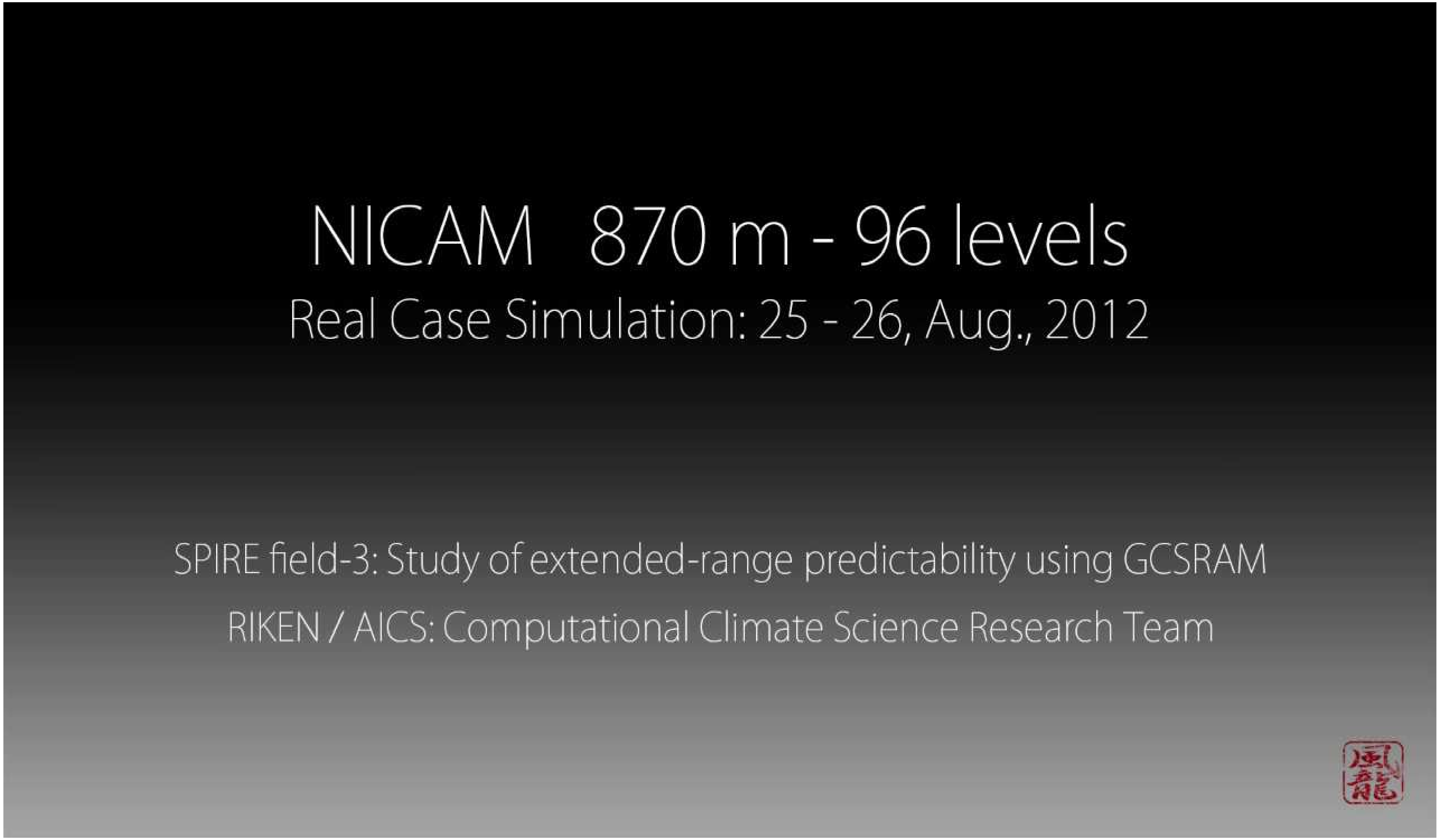


Weak scaling test

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



NICAM870m/L96 animation



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きた。***2014***年学位

Direction of our research in AICS in next 5 years

- **Infrastructure:**
  - <u>**Extension of basic library SCALE**：</u>
    - **Massive parallel analysis routines for acceleration of scientific output, social outcome**
      - <u>Not only acceleration of simulation itself but also acceleration of analysis phase</u>
    - **Easy programing and high performance computing:**
      - DSL(Domain Specific Language)? e.g. stencil DSL
        - w/ the Japanese next flagship computer project

Direction of our research in AICS in next 5 years

- **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 issue**
    - **Moist LES theory**

Brief description of SCALE

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 :
  4^{th} order central difference
- Topography :
  Terrain-following
- Positive definitive :
  FCT scheme

Other

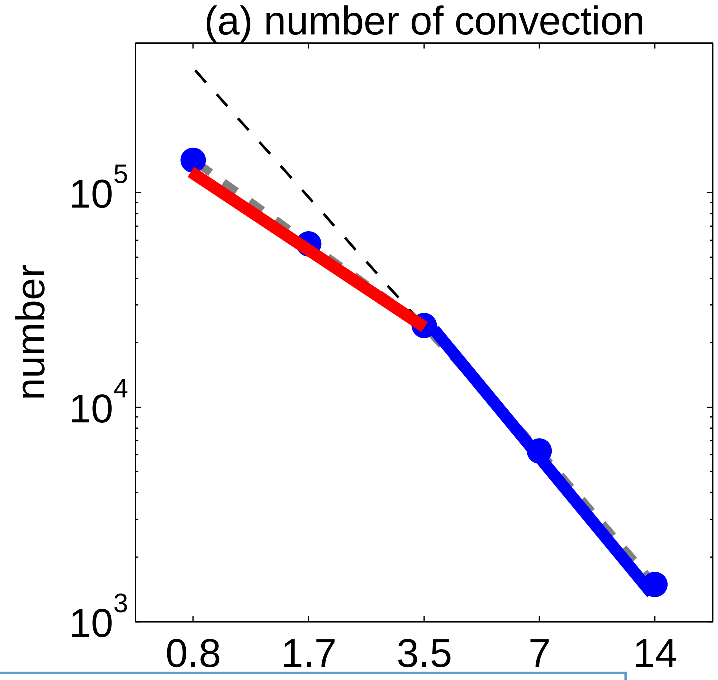
- Offline/Online nesting system
- LETKF assimilation system

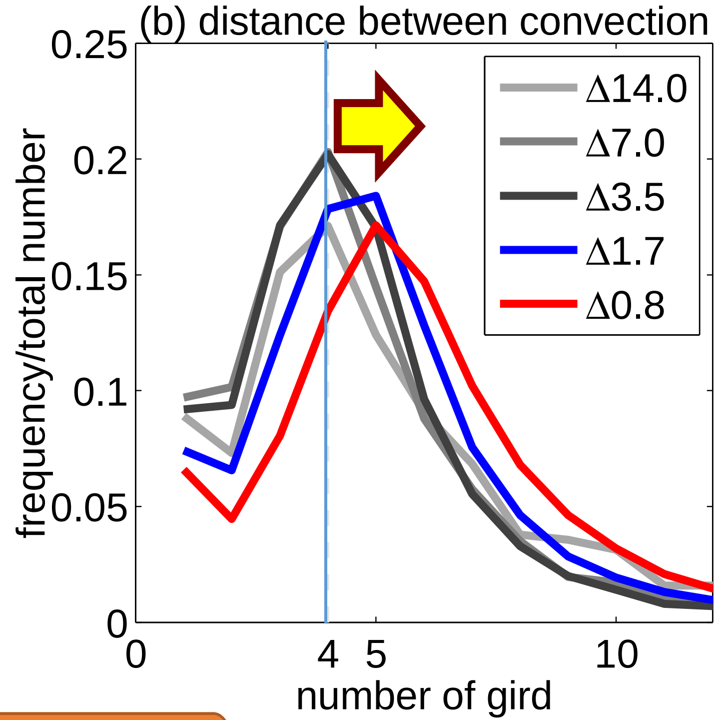
Physical schemes

- 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)

Convergence of 1. number of convection 2distance of neighboring convection

Miyamoto et al.2013 Geophys. Res. Lett.





Δx ≧ 3.5 km:

- # of conv.: increase by factor of 4
- Conv. distance between convection: 4 grids => unphysical?

Δx ≦ 1.7 km:

- # of conv.: decrease in increasing rate
- Conv. distance:

  >5 grids => close to the nature

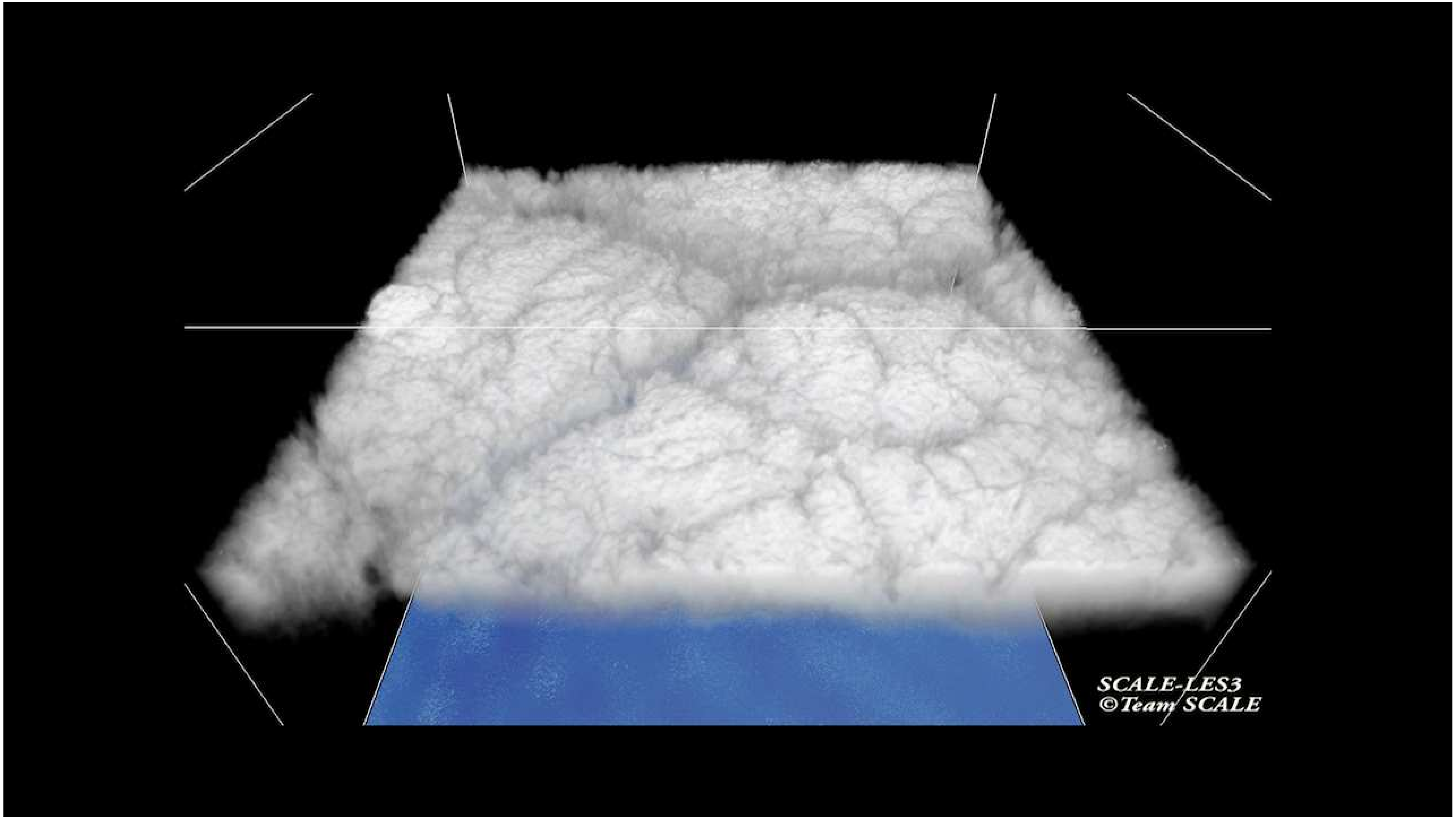
conclusion

Convection features (structure, number, distance

change between Δ3.5 km ⇔ Δ1.7 km

- Δx should be 2.0〜3.0 km to resolve convection in global models

Resolution of 2km is tipping point!



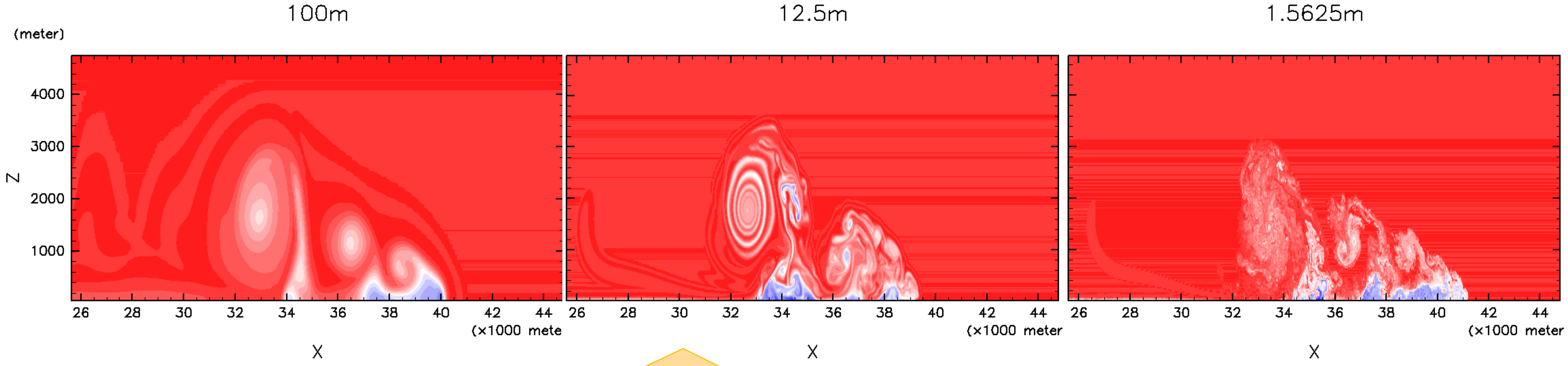
back

Validation of higher resolution simulation

Density current test case

51.2 km x 6.4 km (2-D domain)

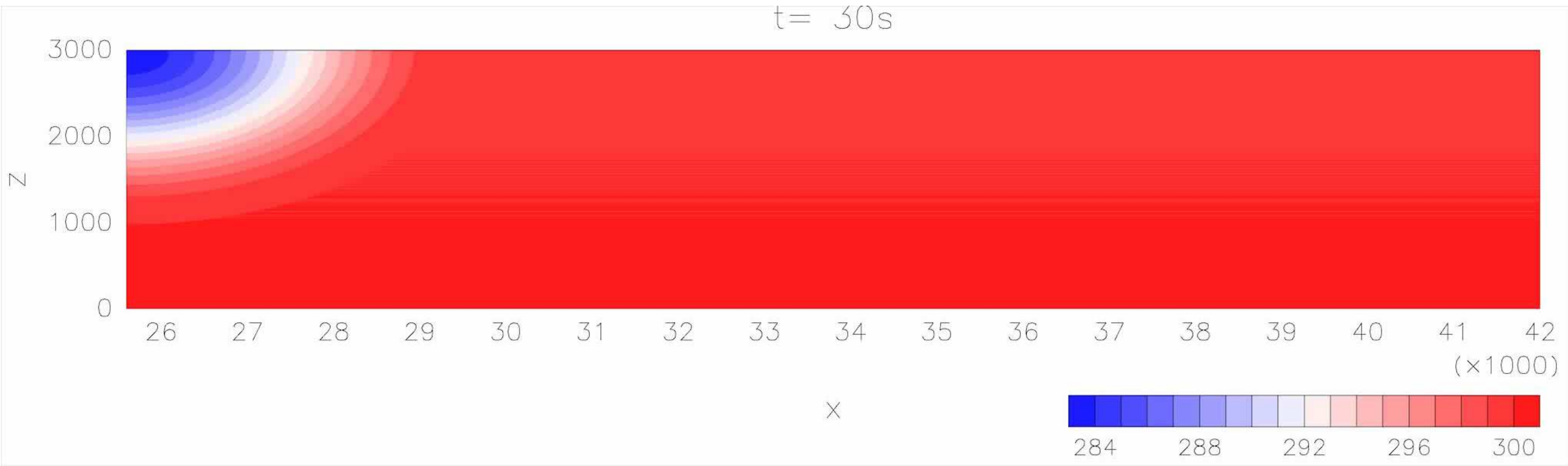
Same setting as Straka et al. (1993) but no physical diffusion.



The bigger spiral structure still remain
due to absent of the smaller scale instability.
It does not mixed well.

animation

Higher resolution: not always better than lower one without appropriate treatment /parameterization.


t= 30s
3000
2000
1000
0
Z
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
(×1000)
X
284 288 292 296 300

back