Requirement of model simulation for big data assimilation

Hirofumi TOMITA

and

SCALE developers

Adachi,S.A., Nishizawa,S., Sato, Y., Yamaura,T., Yashiro,H., Yoshida,R.

RIKEN/AICS

The 4th International Symposium on Data Assimilation (23-26 Feb. 2015)

What I talk about?

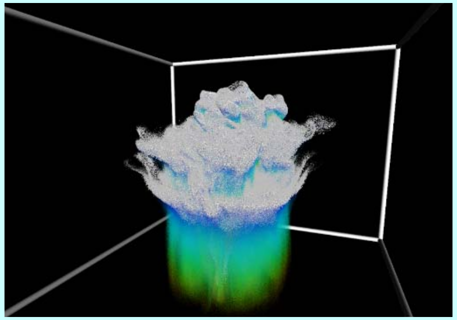
- What does our modeling activity contribute to BDA?
  - In a real time BDA, speed of simulation code is a tiny issue?
  - Challenging in the model!
- Our modeling activity:
  - Development of a high resolution model with a library suitable to HPC
  - An example of recent result
    - regional climate simulation for assessment of Hyogo/Kobe area
- Summary

Project Overview

Himawar-8
PAWR

LETKF

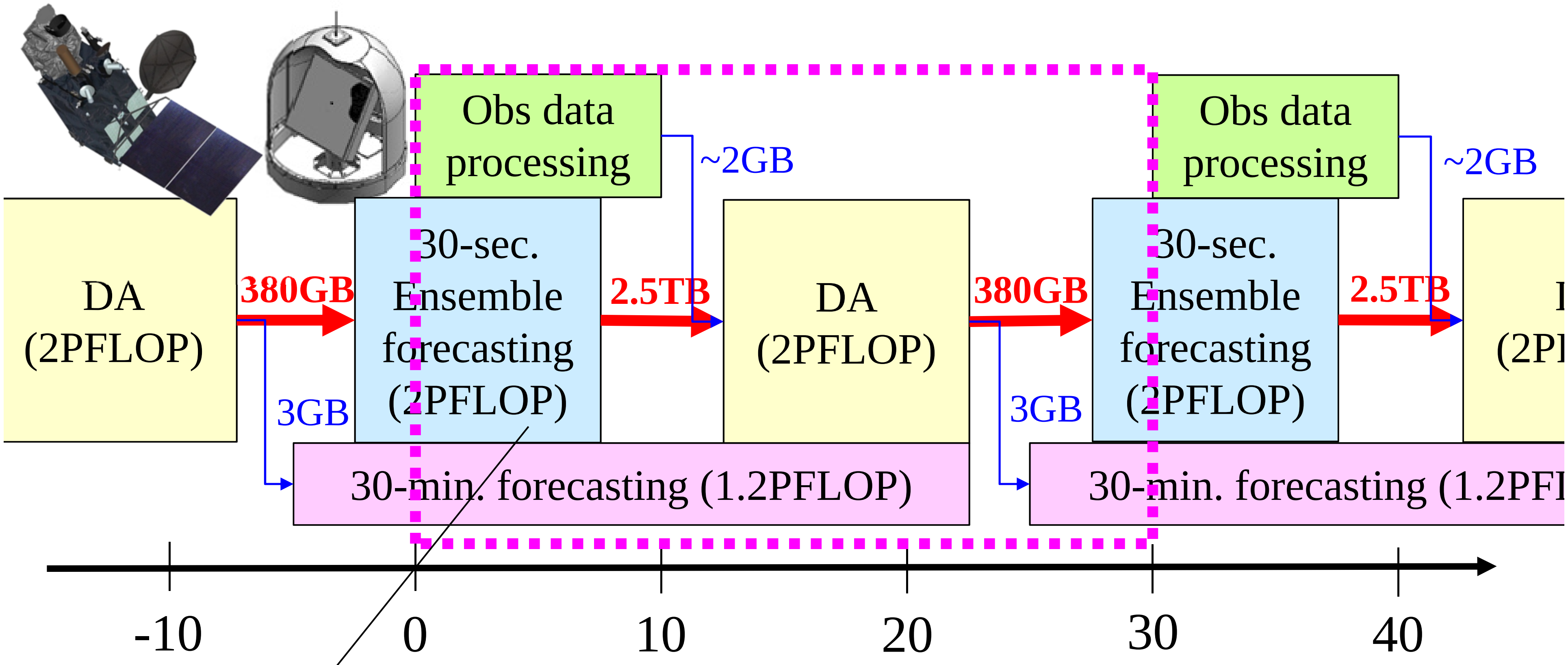
NHM SCALE



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!

Revolutionary super-rapid 30-sec. cycle



Time (sec.)

Computing requirement: 250TFLOPS (effective)

Equiv. to 1/4 of the K computer

Requirement to model from the viewpoint computational aspect

<u>120 times more rapid</u> than the hourly "Rapid Refresh"

Requirement of model simulation on real time forecast

Physical Challenge!

Is each cumulus preditable by BDA?

Cumulus is a chaotic phenomenon.

- A research of environment for convection occurrence is a useful approach
  - Criteria of convection occurrence?
- Fortunately, we will have a very frequent and huge observation data near the future.
  - We can watch the detail environment in real time.
  - e.g. PAWD, Himawar-8

Computational Challenge!

Finish huge computation within a limited resource?

- To use observational Big Data efficiently, how should the model going on?
  - 1. High resolution
  - 2. High sophistication of physics
  - 3. Large number of ensembles

Very huge computation within a given time.

Their balancing is a key!

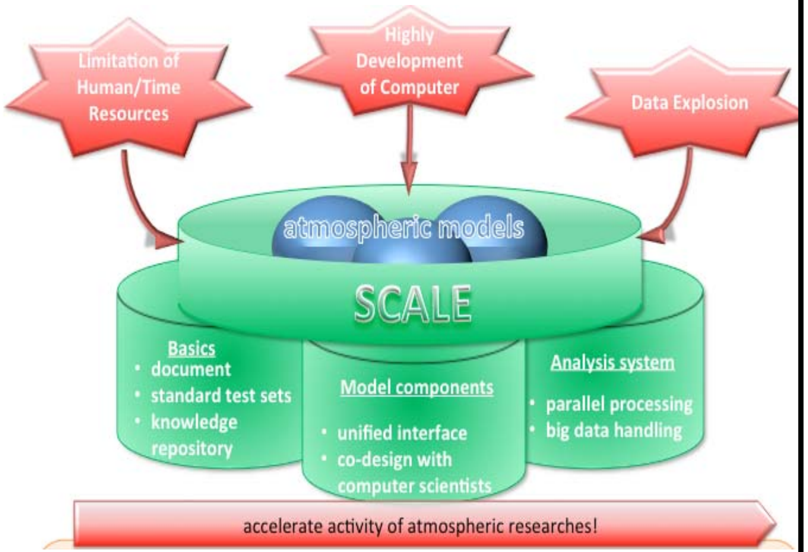
Computational resource is limited! (even when the exa-scale era)
High resolution CFD, especially, becomes tight for memory bandwidth.
Too much Sophisticated physics is not needed

OUR MODELING ACTIVITY

SCALE library

<u>SCALE (Scalable Computing for Advanced Library and Environment)</u>

- from x86 PCs to super computers (K and post K)
- Collaborate with computer scientists
- Open source with the 2 clause BSD license
  - http://scale.aics.riken.jp/

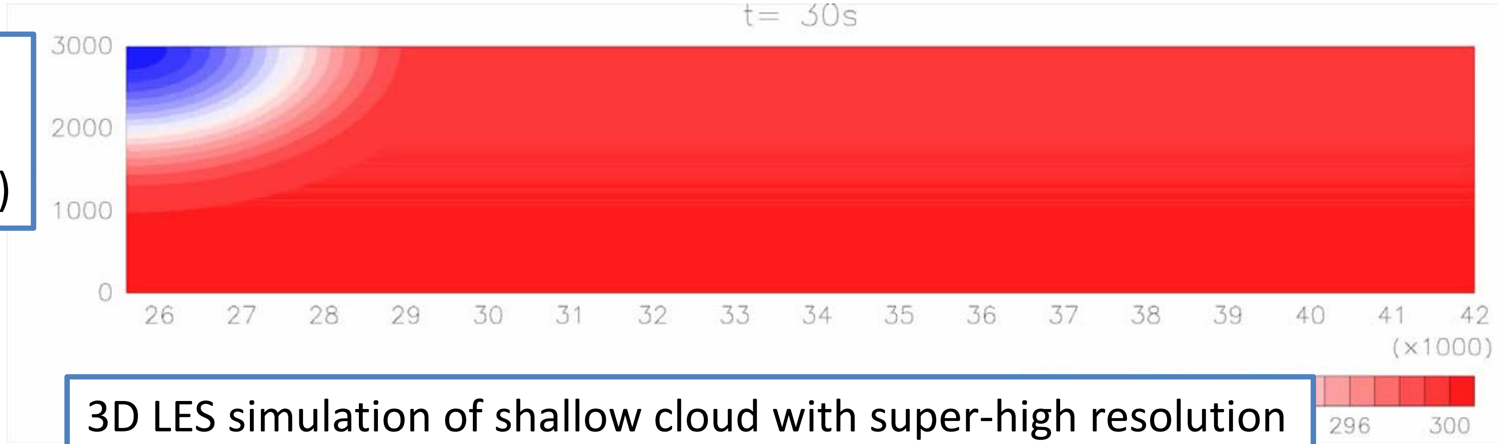


- Feature:
  - model components (unified API) cover …
    - dynamical cores framework: HE-VE, HE-VI, HI-VI
    - physics: many microphysics, turbulence, radiation, surface flux
    - misc: I/O, communication, logger, timer etc
  - Enough documentation
    - model description
    - knowledge repository (future work)
  - A lot of test cases
    - component unit test
    - standard benchmarks
  - pre- and post-processing tools
    - parallel handling (future work)

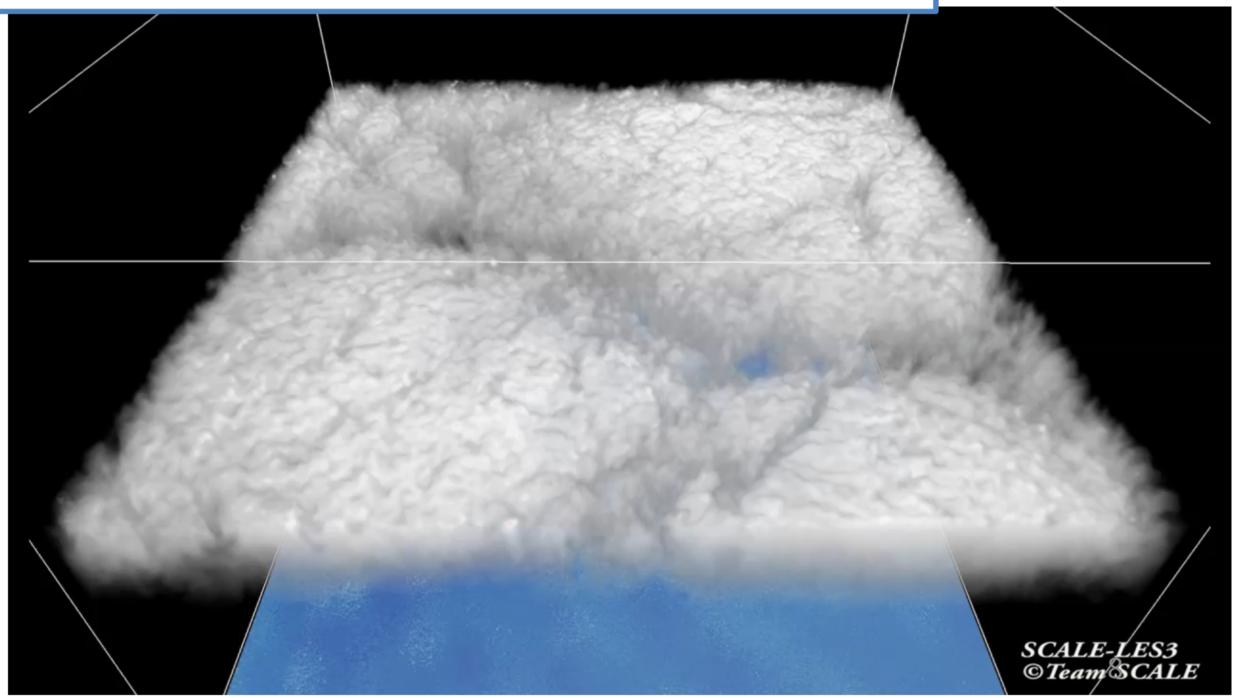
SCALE-LES (based on SCALE library)

A meteorological Large-Eddy simulation model using SCALE

2d test case of dynamics (cold bubble)



3D LES simulation of shallow cloud with super-high resolution



- SCALE-LES is an application of SCALE
- Target scales:
  - O(10-100)km^2 domain with O(10-100)m resolution
  - Going to a global LES?

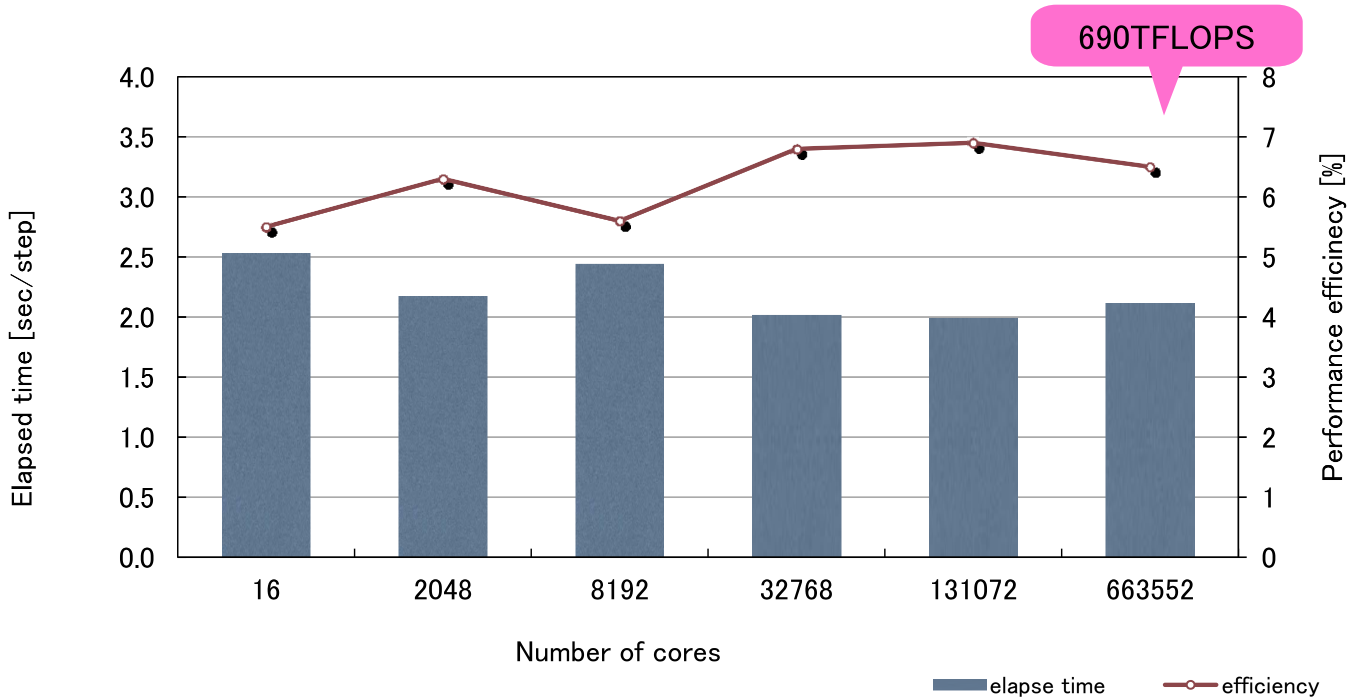
An example of experimental configurations

Regional Climate Assessment (present and future) in Hyogo/Kobe area
(funding by Hyogo prefecture and Kobe city COE project)

- Numerical schemes
  - Spatial difference
    - Terms related to acoustic wave: 2nd order central difference
    - Advection terms: 4th order central difference (tracers FCT) : requirement from LES
  - Temporal integration
    - Dynamical variables: 3 steps RK, tracers: Euler (satisfies CWC)
    - Horizontally explicit vertically implicit (without split explicit for reduction of memory load/store)
- Physical processes
  - Turbulence: Smagorinsky type (Brown et al. 1994, Scotti et al. 1993), MYNN level2.5
  - Cloud microphysics: 6 category single or double moment bulk (Tomita 2008 or Seiki and Nakajima 2014)
  - Radiation: MSTRN-X (Sekiguchi and Nakajima 2008)
  - Surface flux: Monin-Obukhov similarity model (Beljaars 1991, Wilson 2001)
  - Land: multi-layer bucket
  - Ocean: slab
  - Urban canopy: single layer (Kusaka et al. 2000)

Computational performance on K

- performance @ K computer
  - over 10% of peak performance (dynamical core)
  - Note that 5~8% for full simulation (including physics and I/O)
  - 100% weak scaling to full system (663,552 cores)



Our mission(modeling group) in real-time Big Data Assimilation

① Finish 100 ensembles of 30sec integrations within 15 sec!

② Finish a 30min forecasting within 5min!

- High resolution with LES!
- High degree of physical performance!

① Feasibility for 30 sec run w/ 100 members

▪Elapse time to spend on 30 sec run → 15 sec

- Domain/resolution settings is same as 30 min forecast.
- Assume using 18,000 nodes (180 nodes x 100 members)

Demand Performance		Current performance
≈ 15 sec w/ **180 node**	VS	= 15 sec w/ **2592 node**

with 4 domains online nesting
(7.5km > 2.5km > 500m > 100m)
＊96km x 96km size in 100m domain

must be 14 times faster!!

How solve the problem・・・

② Feasibility for 30 min Forecast

▪ Experiment Configurations

Domain settings :

- Resolution: 100 m
- Domain: 120 x 120 x 25 km
- Grid: 1200 x 1200 (horiz.), 80 (atmosphere) 5 (land) layers

▪ Elapse time to spend on 30 min forecast

Demand Performance	VS	Current performance
≈ 5 min		= 15 min

with 2592 node

4 domains online nesting

(7.5km > 2.5km > 500m > 100m)

meet the demand by 7,500 nodes approx.

＊assume perfect strong scaling

A Plan for improvement of the Performance Problem

1. Model improvement

- Take a longer dT (challenge to numerical stability)
- Computational Optimization

Ensemble cycle is much tightest work

2. Reconsideration of the arrangement of configurations

- Reduce spatial resolution
  → **x8 speed-up** w/ 200m, or **x3.4 speed-up** w/ 150m
- Change microphysical scheme: lead to reduce number of tracers
  → **x1.5 speed-up** w/ single moment
- Reduce domain size
  → **x4 speed-up** w/ $60km^2$
- Use more nodes
  → **x2 speed-up** w/ 36,000 nodes

- These combination leads to the feasibility of this ensemble cycle on K computer.

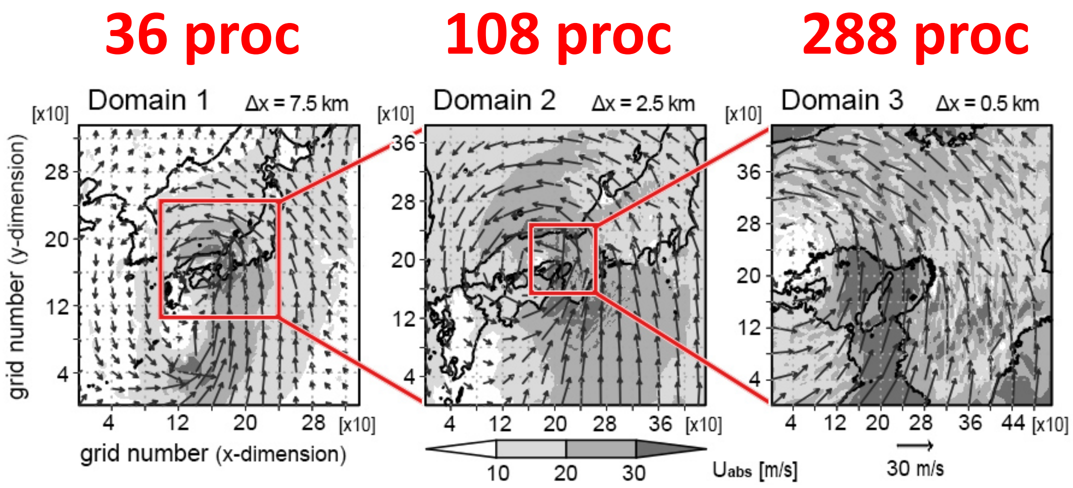
The design will be finished within this year

NOTE that : in post K, higher resolution, larger domain-size, and more sophisticated physics are available.

Online-Nesting design for high applicability of HPC

One advantage of SCALE-LES over other models

Give freedom
on configuration!



- **allow to specify MPI processes to each domain** **(not in WRF etc.)**
  - → MPI processes divided by MPI_COMM_SPLIT
  - → User can specify MPI proc. considering computational efficiency.

- **inter-domain communication is overlapped behind time-integration** **(not in WRF etc.)**
  - → asynchronous communication with double-buffer (only for 1-way nesting)

etcetera: associated with experimental setup

- allow to specify vertical grid coordinate to each domain (not in WRF etc.)
- no limitation for model configuration among domain, except for dt

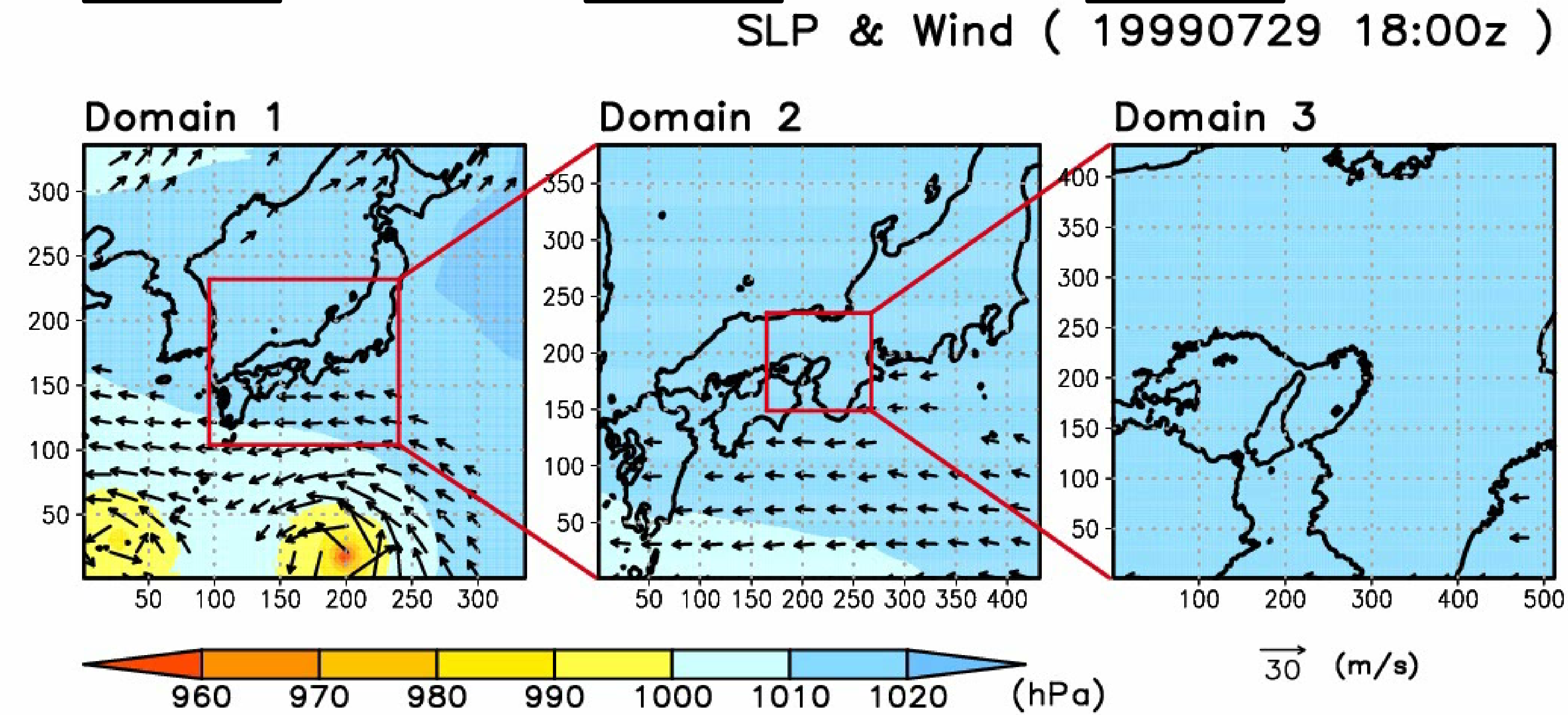
Nesting system of two typhoons ： SLP & wind field

Nesting system work well even when severe Typhoon case
Seamless simulation is ready!!

7.5km

2.5km

0.5km



Model performance of SCALE-LES : Summertime precipitation
30yr climatology .vs. 1yr model result (30yr ongoing)

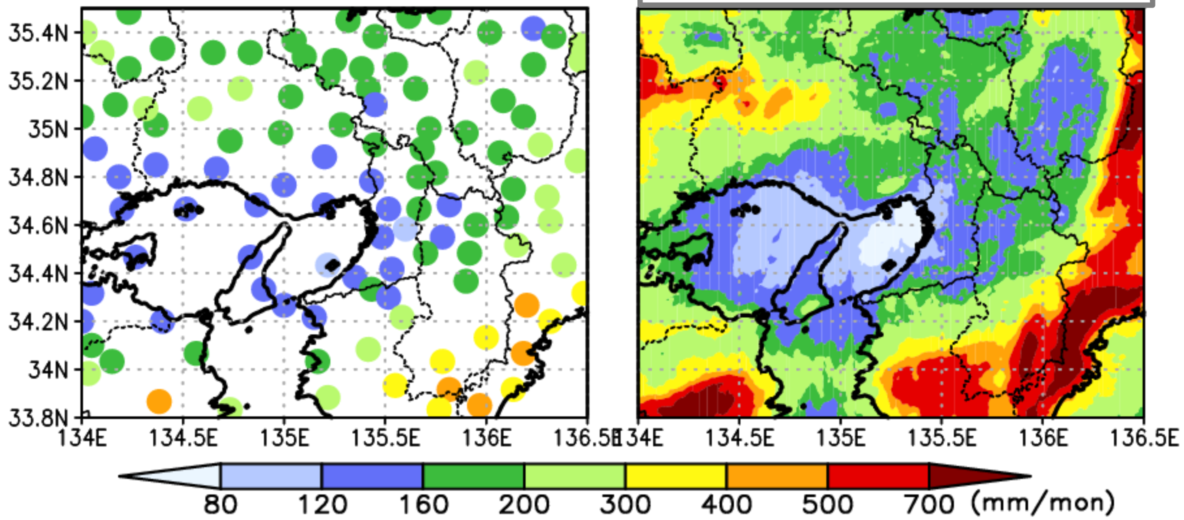
This is the first production result.
=> Improvement in the next year

Observation
30-yr mean（1981-2010）

SCALE-LES (Δx=500m)
1-yr mean

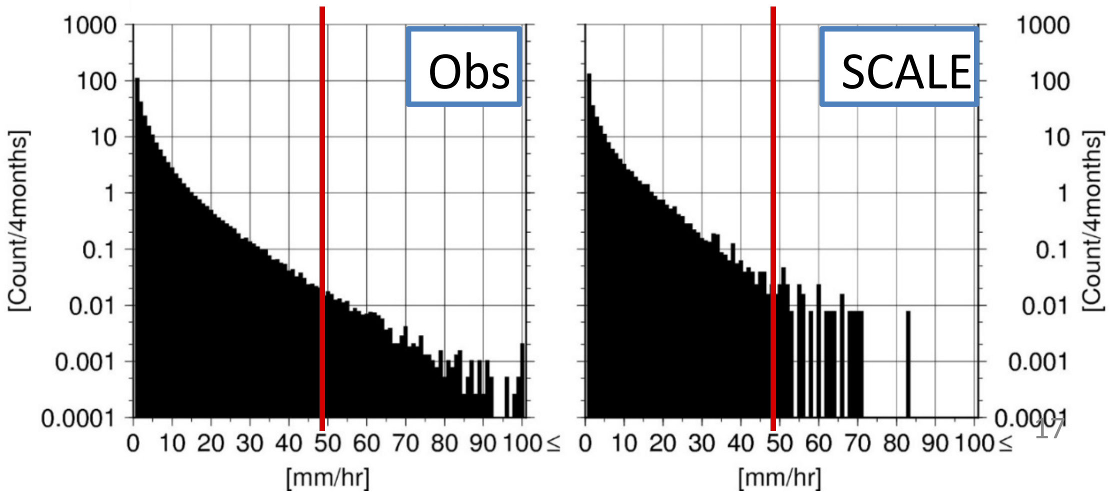
monthly mean Precipitation

Precipitation pattern is reasonable, though the improvement is necessary.

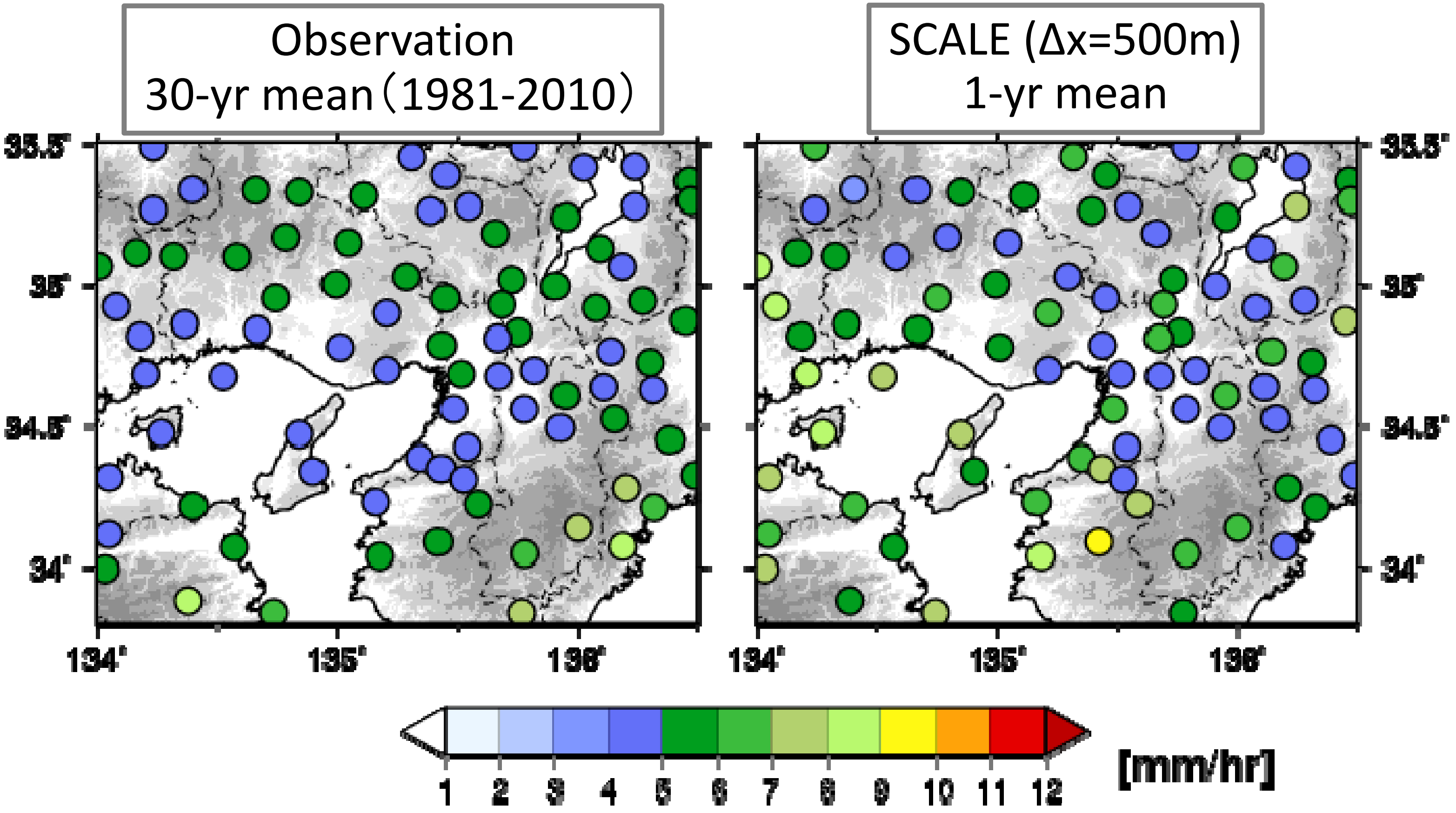


Occurrence frequency of hourly precipitation

In <50 mm/hr, quantitative agreement with obs.
In >50mm/hr, # of samples are less



Distribution of 85th percentile on hourly precipitation



- Absolute value of 85th percentile in SCALE is comparable with observation.
- Distributions differ between observation and model in the mountain and coastal areas, because of short sample years in SCALE.

Summary

- In the real time Big Data Assimilation, model simulation also requires big computer power.
  - as well as the work flow, data assimilation part, network ability from external site.
  - Total war involving meteorological science & computer science to resolve this subject!
- Our modeling group corresponds to provide….
  - Framework for fast nesting system
    - Inter-domain communication is hidden behind the calculation.
    - Debatable point : 2way nesting is needed from the viewpoint of efficient use of data assimilation?
  - High potential for predictability
    - Continuing to take efforts for improvement of physical performance by the climate simulation.
- In K computer, 100 member ensembles is still tight work.
  - For genuine application, post-K class computer power is needed.
  - But, we can demonstrate the usefulness of this BDA system in K computer era within 3 years!

Thank your for you attention!