Center report from JMA

Numerical Prediction Division,
Japan Meteorological Agency

Summary

- Model development: our recent experience
  - Improvement of vertical transport processes: incorporated into the operational Local Forecast Model (LFM) in March 2021
  - Utilizing deliverable of the COORDE project to parameterization development for Global Spectral Model (GSM)
- HPC readiness
  - Improvement of parallelization can be a important key toward high resolution global spectral model
  - Research on GPU started
    - Toy models which imitate /extract actual models can be useful tool to estimate and understand benefits of GPU
- ML for model development: familiarization ongoing. (extra slides)

Introducing the Leonard Term to LFM

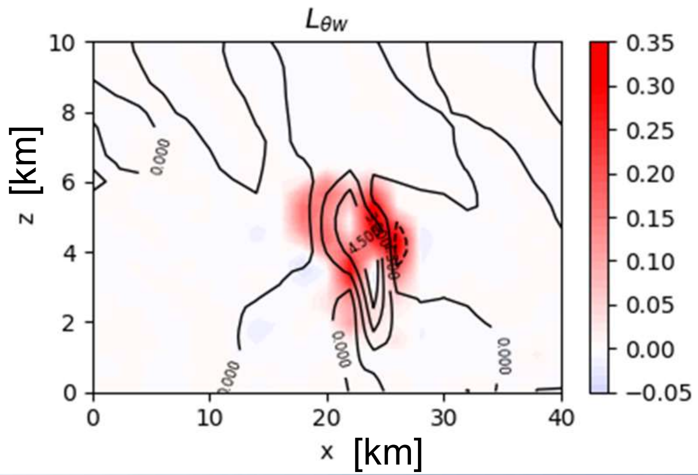
- The vertical subfilter flux is parameterized following Moeng et al. (2010) and Verrelle et al. (2017):

$$\tau_{wc} = \underbrace{\frac{K_L}{12}\left(\Delta x^2 \frac{\partial w}{\partial x}\frac{\partial c}{\partial x} + \Delta y^2 \frac{\partial w}{\partial y}\frac{\partial c}{\partial y}\right)}_{\text{the Leonard term}} + \underbrace{\tau_{wc,Kgrad}}_{\text{the Reynolds term}}$$

the Leonard term

It depends on the horizontal gradient (different from a vertical column model).

the Reynolds term
as parameterized by a traditional vertical column model

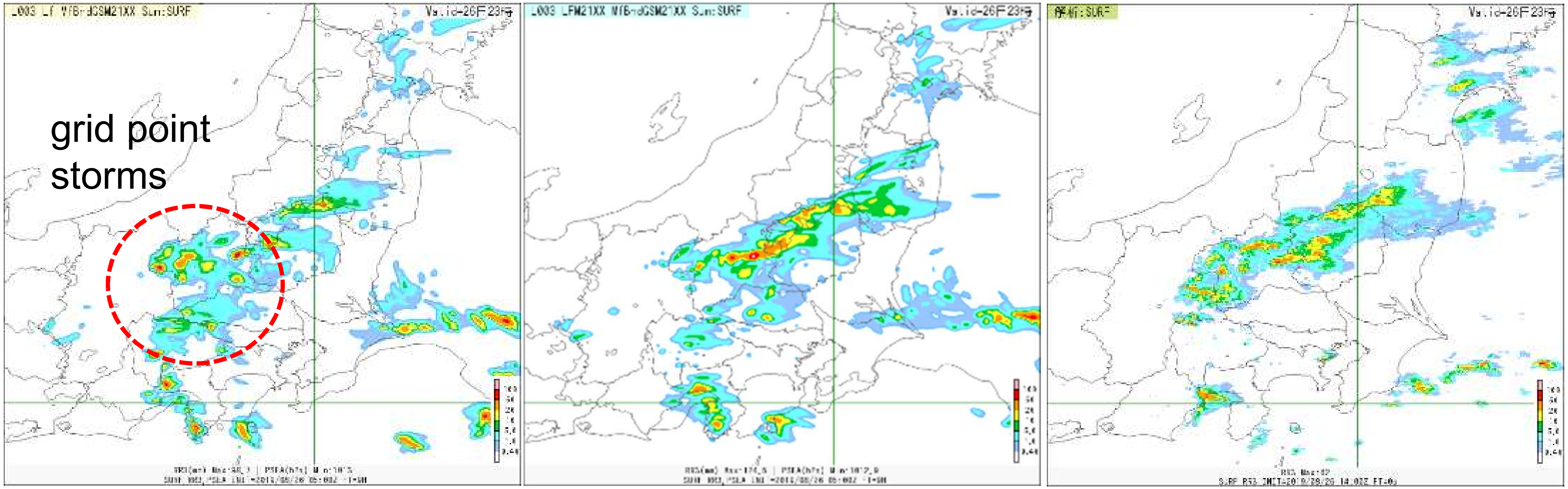


Contour: vertical velocity
Shade: vertical heat flux
by the Leonard term

3-hour accumulated precipitation

LFM:

LFM2003 | LFM2103 | Radar/Raingauge-Analyzed Precipitation



- The heat and vapor fluxes due to the Leonard term contribute to reduce artificial grid point storms.

Optimization of parameters for subgrid orographic schemes using the COORDE-type experiments

Implication from COORDE

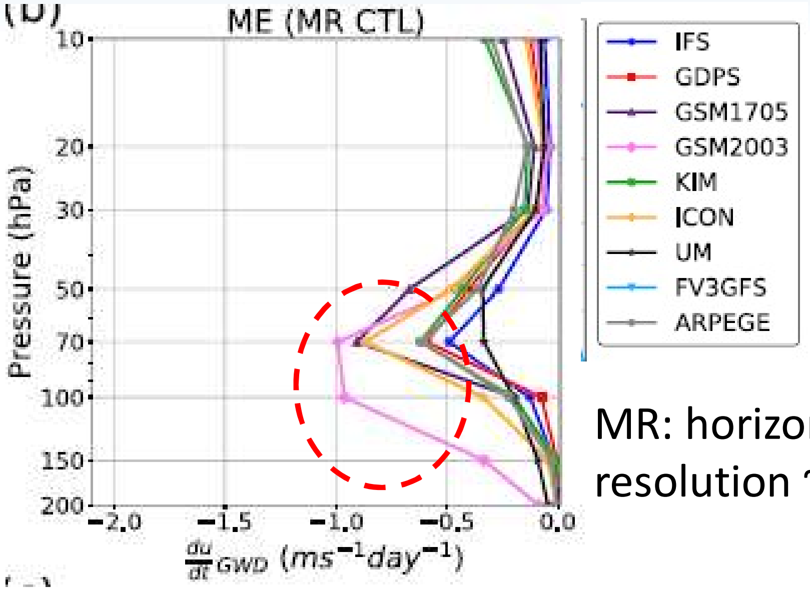


Fig.7 of van Niekerk et al. (2020)

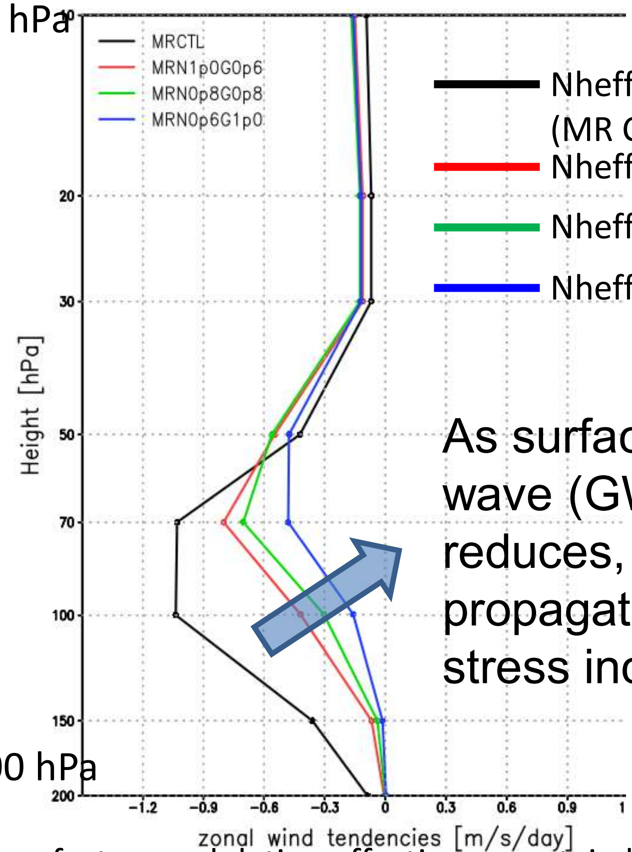
MR: horizontal resolution ~40km

The COORDE paper pointed out that:

"GSM2003 has a large deceleration in the lower stratosphere that is at a lower altitude than in the other models."

Sensitivity of parameters to subgrid GWD

GW drag over the Middle East stratosphere



Nheff=2.4, G=0.25, α=12 (MR CNTL)
Nheff=1.0, G=0.6, α=12
Nheff=0.8, G=0.8, α=12
Nheff=0.6, G=1.0, α=12

As surface gravity wave (GW) stress reduces, vertically propagating GW stress increases.

Nheff: a factor modulating effective mountain height in GWD
G: a function of the mountain sharpness in GWD
α: a parameter modulating TOFD

Optimization of parameters for subgrid orographic schemes using the COORDE-type experiments

Impact of parameterized orographic drag to zonal wind

MR CNTL –MR NOSSO

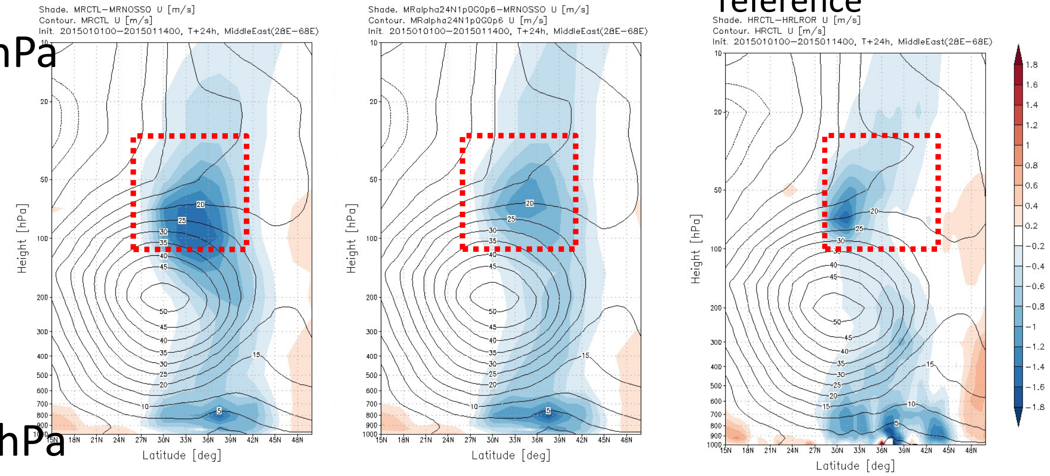
MR REVISED –MR NOSSO

HR CNTL–HR LROR : impact of resolved orography in HR as reference

10hPa

Middle East (28E-68E)

1000hPa



Zonal wind mean error

MR CNTL–Analysis

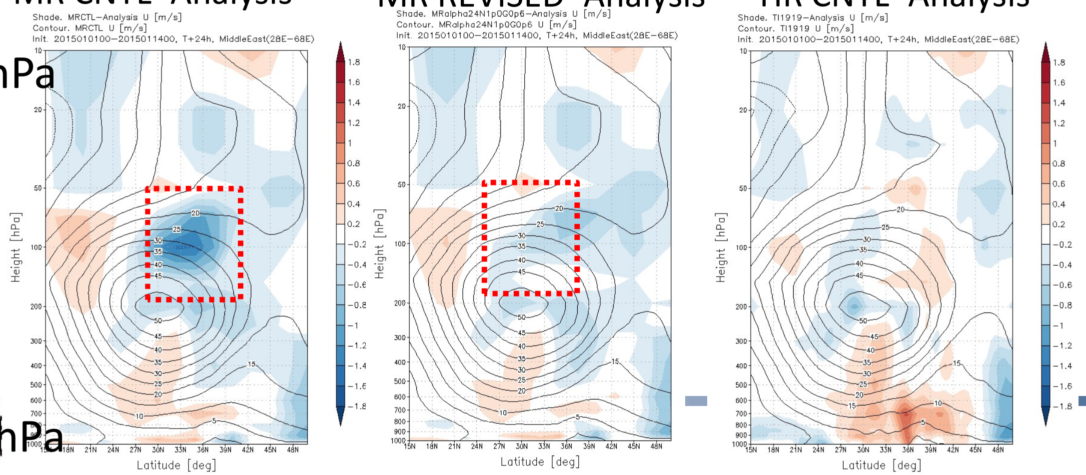
MR REVISED–Analysis

HR CNTL–Analysis

10hPa

Middle East (28E-68E)

1000hPa



*horizontal resolution:
LR: ~200km
MR: ~40km
HR: ~10km

The parametrized GW drag in MR CNTL is stronger and at lower altitude than the resolved orographic drag.

Impact of parametrized SSO drag (including the enhancement of TOFD) in MR REVISED is much closer to that of resolved drag in the lower stratosphere
And results in reduction of zonal wind meanerror.

Chihiro Matsukawa

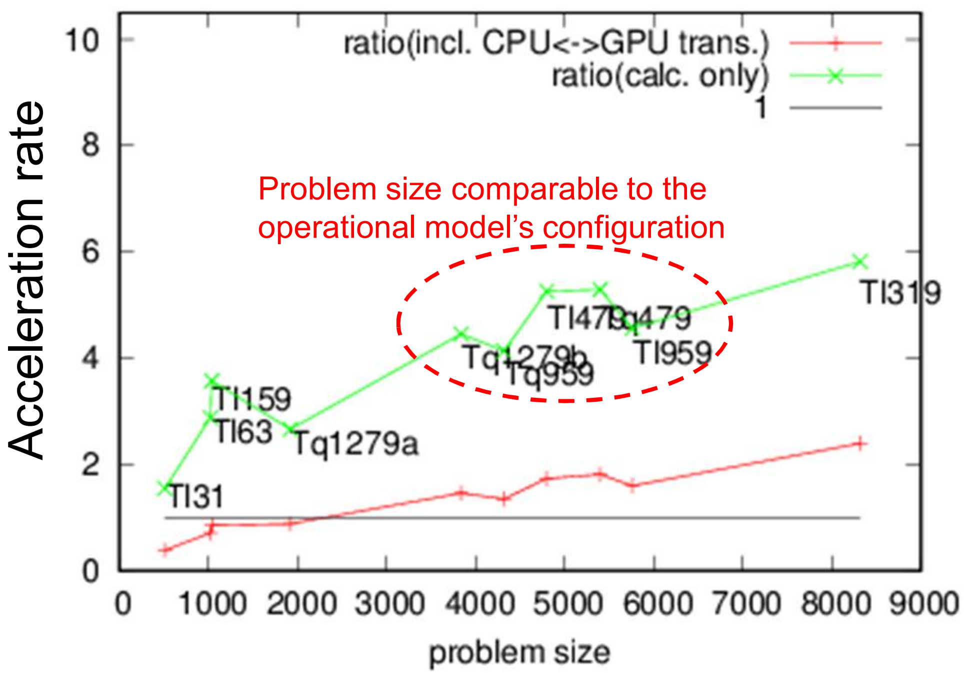
HPC readiness: parallelization

- Toward further improvement to GSM
  - Research on more efficient (and also grid model-friendly) MPI decomposition for SI-SL spectral models
    - Unifying MPI decomposition stages for physics parametrization and Semi-Lagrangian advection
      - Preliminary tests (dynamics only) showed that the new MPI decomposition reduced ~ 40% of computational cost in Tq959 jobs.
    - Also preparation for spectral/grid hybrid approach.

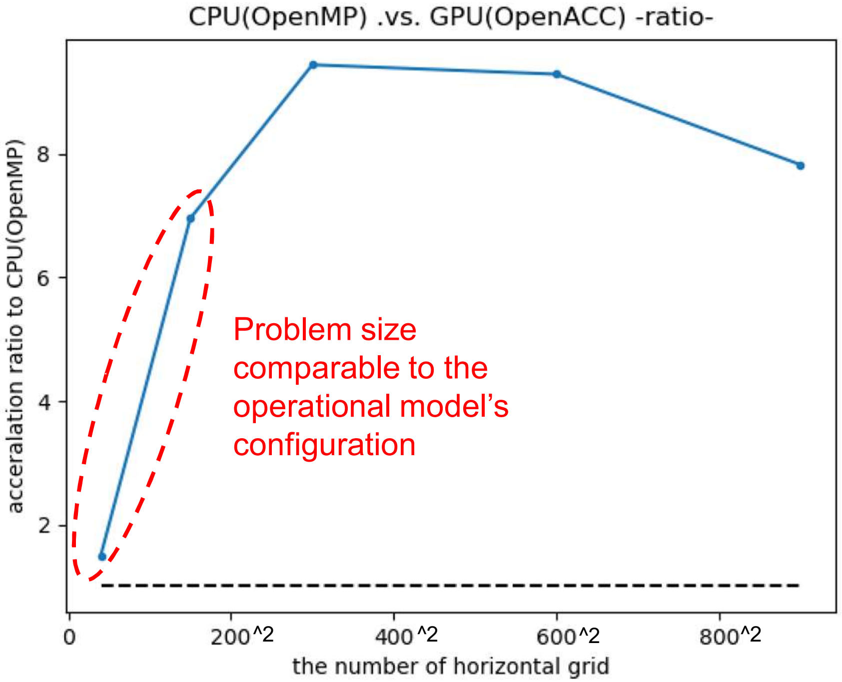
HPC readiness: GPU adaptation

Use small models which extract or imitate “bottlenecks” of the actual NWP models as a first step.

A bottleneck subroutine case in GSM’s cumulus parameterization
(against CPU with OpenMP 12 threads)



An advection model case which imitates the regional model (ASUCA) ’s bottleneck part
(against CPU with OpenMP 8 threads)



Tested on NVIDIA Tesla V100 (GPU) and Intel Xeon Gold 6226 (CPU, 2.70GHz)

Problem size = (number of horizontal grid points per MPI process)
Acceleration rate = (elapse time on CPU) / (elapse time on GPU)


EXTRA SLIDES

OPERATIONAL NWP SYSTEMS UPDATE

NWP systems upgrade since WGNE-35 (Nov. 2020)

- Global NWP (incl. Global EPS)
  - Major upgrade in Mar. 2021 (Ujiie et al. 2021, Yokota et al. 2021 and Yamaguchi et al. 2021, WGNE Bluebook)
    - Increase vertical levels of GSM, GEPS from 100 to 128
    - Improvement to the global snow depth analysis
    - Introduction of soil moisture analysis
    - Improvement to the 4DVAR-LETKF hybrid DA system
      - Increased the weight for ensemble based background error covariance
      - Increased ensemble member from 50 to 100 for LETKF
    - Increase ensemble member from 27 to 51 for GEPS
  - Improvements to observation processing in Jun. 2021
    - All-sky assimilation of microwave water-vapor sounder from GMI/GPM, ATMS/NOAA-20, Suomi-NPP, SSMIS/DMSP-F17, F18, SAPHIR/Megha-Tropiques, MWHS-2/FY-3C.
    - Revision of a bias correction method for aircraft-based observations.

NWP systems upgrade since WGNE-35 (Nov. 2020)

- Meso-scale NWP system
  - Improvements to observation processing (May. 2021)
    - Assimilation of Metop-C/AMSU-A, MHS, and radar reflectivity observed by a dual-polarized radar at Tokyo

NWP systems upgrade since WGNE-35 (Nov. 2020)

- Local NWP system
  - Major upgrade in Mar. 2021 (Kusabiraki et al. 2021, WGNE Bluebook)
    - Increase vertical levels of LFM from 58 to 76
    - Physics parameterization upgrade
      - Introduction of a cloud fraction diagnostics scheme based on Wilson and Ballard (1999)
      - Introduction of a mixed subgrid scale turbulent transport scheme by Moeng et al.(2010)
      - Update of ancillary data for the land surface process (thermal roughness over urban area and land surface albedo climatology).
  - Improvements to observation processing (May. 2021)
    - Assimilation of radar reflectivity observed by a dual-polarized radar at Tokyo

Introducing the Leonard Term

- The vertical subfilter flux is parameterized following Moeng et al. (2010) and Verrelle et al. (2017):

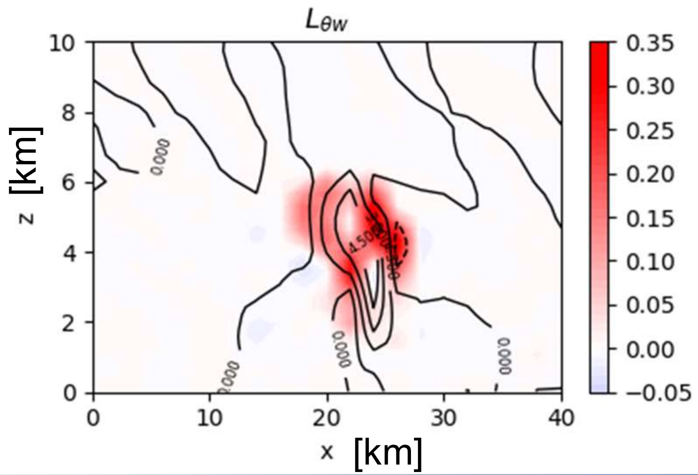
$$\tau_{wc} = \underbrace{\frac{K_L}{12}\left(\Delta x^2 \frac{\partial w}{\partial x}\frac{\partial c}{\partial x} + \Delta y^2 \frac{\partial w}{\partial y}\frac{\partial c}{\partial y}\right)}_{\text{the Leonard term}} + \underbrace{\tau_{wc,Kgrad}}_{\text{the Reynolds term}}$$

the Leonard term

It depends on the horizontal gradient (different from a vertical column model).

the Reynolds term

as parameterized by a traditional vertical column model

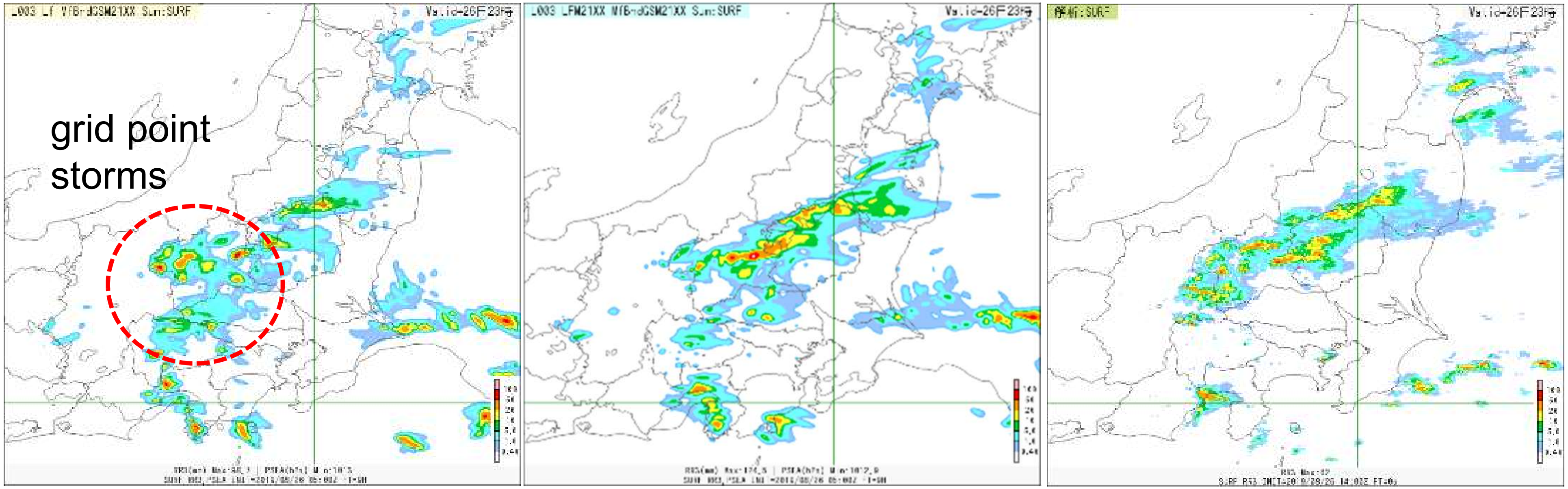


Contour: vertical velocity
Shade: vertical heat flux
by the Leonard term

3-hour accumulated precipitation

LFM:

LFM2003 | LFM2103 | Radar/Raingauge-Analyzed Precipitation



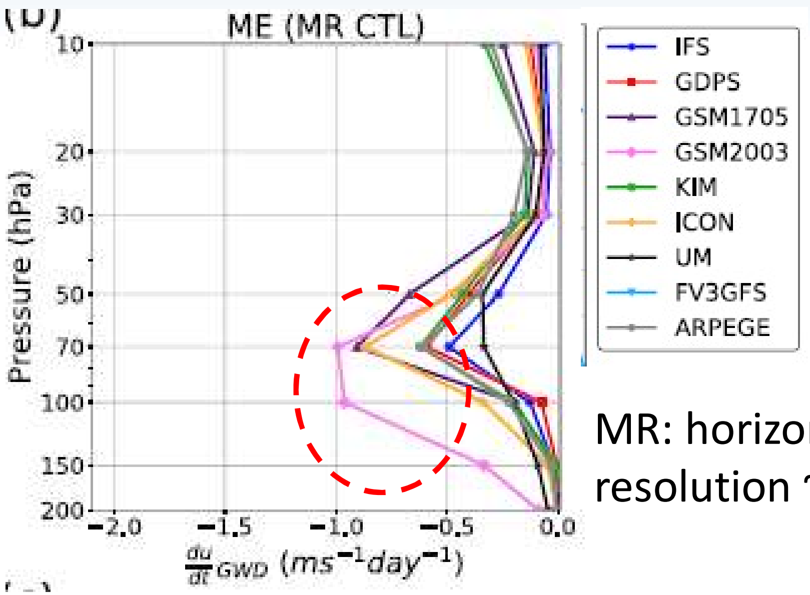
- The heat and vapor fluxes due to the Leonard term contribute to reduce artificial grid point storms.

Recent development topics

- Global NWP system
  - Increase resolution from Tl959(~20km) to Tq959(~13km) with improving effective resolution
    - Introduction of quadratic grid to mitigate spectral blocking
    - Reduce numerical diffusion
    - Further weaker filter for model orography
  - Physics parameterization upgrade
    - Optimization of parameters for subgrid orographic schemes
      - Utilize deliverables of the COORDE project
    - Diagnostics of effective size of cloud ice for radiation
    - Update of ozone climatological data
    - Improve treatment of lake temperature.
- Global EPS
  - Increase resolution from Tl479(~40km) to Tq479 (27km)
  - Update of the two-tiered SST approach
    - Start using predicted SST from CPS3 at Day+6

Optimization of parameters for subgrid orographic schemes using the COORDE-type experiments

Implication from COORDE



MR: horizontal resolution ~40km

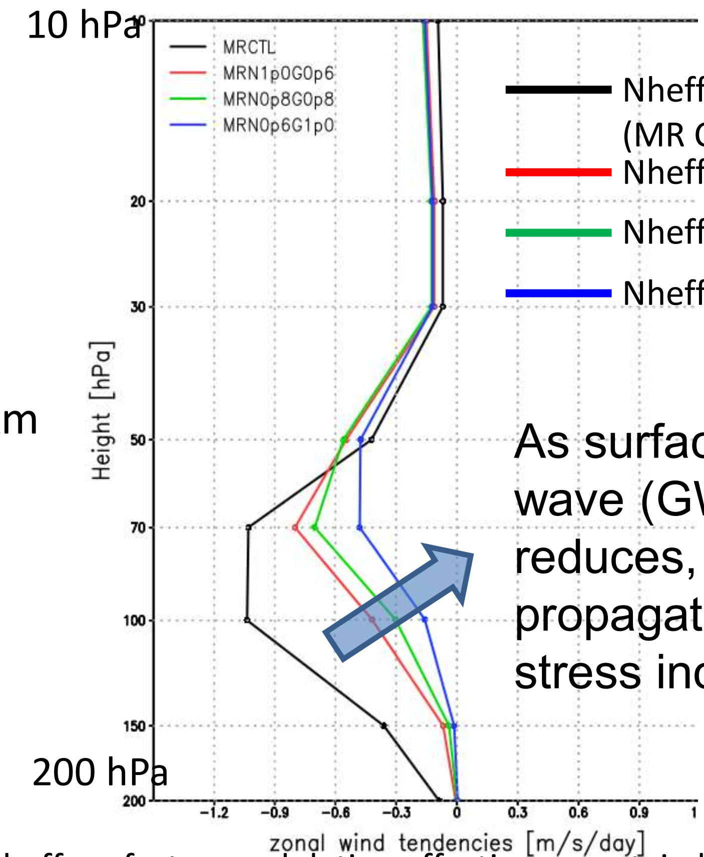
Fig.7 of van Niekerk et al. (2020)

The COORDE paper pointed out that:

"GSM2003 has a large deceleration in the lower stratosphere that is at a lower altitude than in the other models."

Sensitivity of parameters to subgrid GWD

GW drag over the Middle East stratosphere



As surface gravity wave (GW) stress reduces, vertically propagating GW stress increases.

Nheff: a factor modulating effective mountain height in GWD
G: a function of the mountain sharpness in GWD
α: a parameter modulating TOFD

Optimization of parameters for subgrid orographic schemes using the COORDE-type experiments

*horizontal resolution:
LR: ~200km
MR: ~40km
HR: ~10km

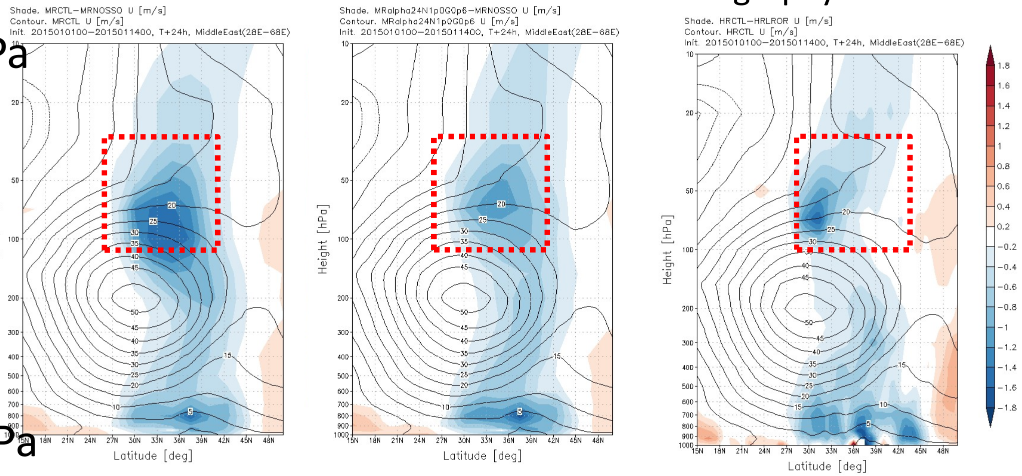
Impact of parameterized orographic drag to zonal wind

MR CNTL –MR NOSSO | MR REVISED –MR NOSSO | HR CNTL–HR with LR orography : impact of resolved orography in HR

Middle East (28E-68E)

10hPa

1000hPa



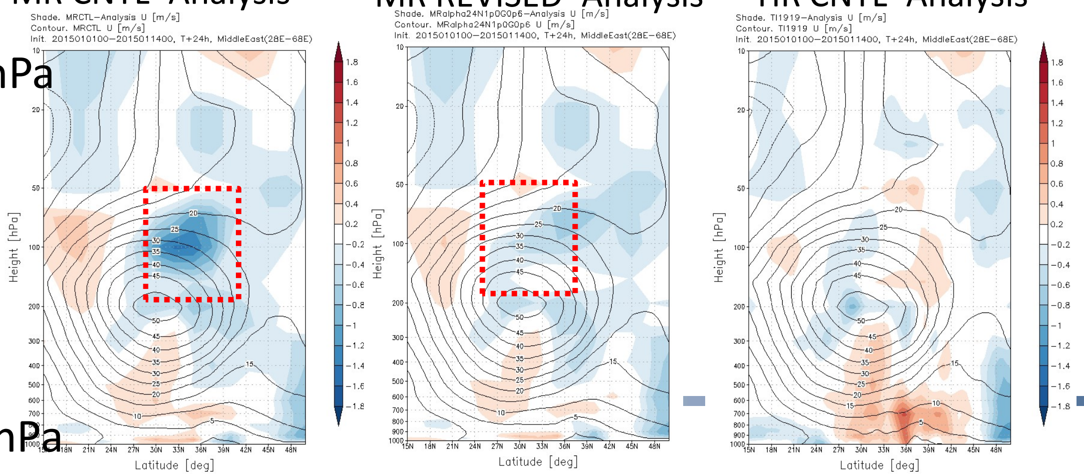
Zonal wind mean error

MR CNTL–Analysis | MR REVISED–Analysis | HR CNTL–Analysis

Middle East (28E-68E)

10hPa

1000hPa



The parametrized GW drag in MR CNTL is stronger and at lower altitude than the resolved orographic drag.

Impact of parametrized SSO drag (including the enhancement of TOFD) in MR REVISED is much closer to that of resolved drag in the lower stratosphere
And results in reduction of zonal wind meanerror.

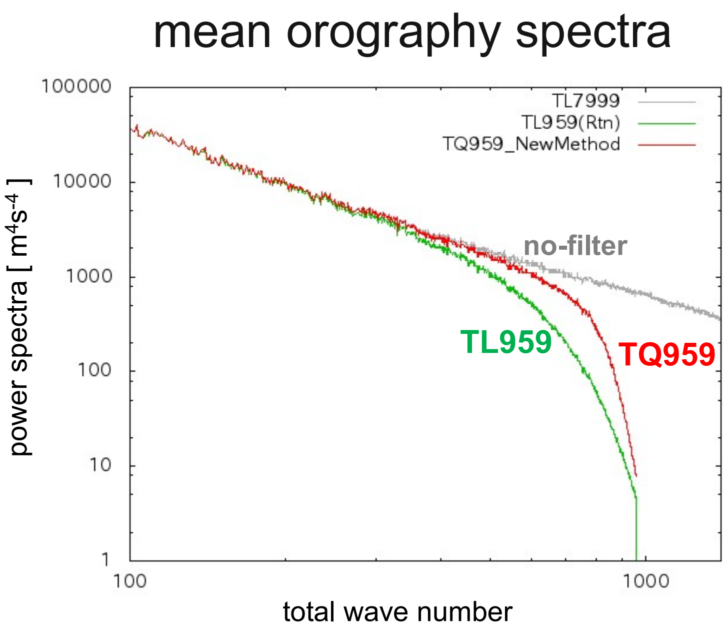
Recent development topics

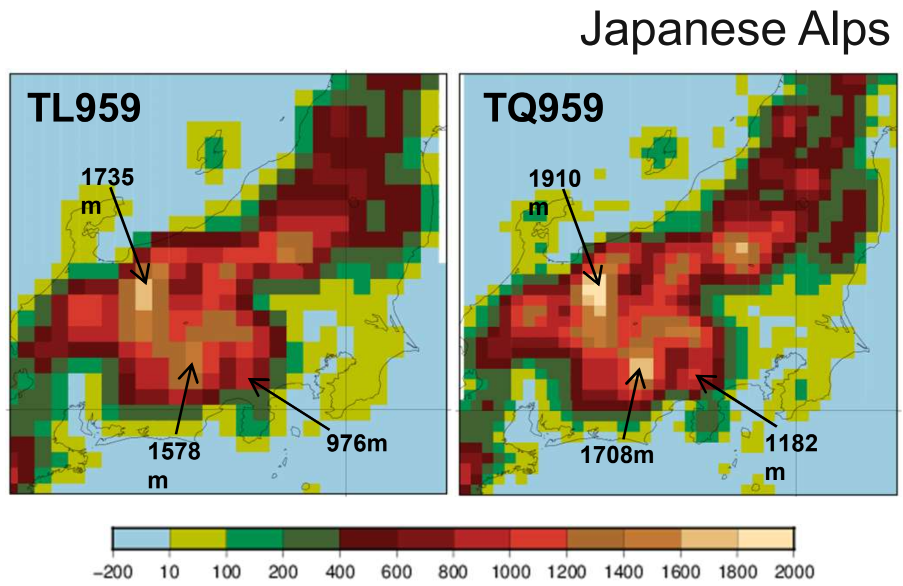
- Meso-scale NWP system
  - Increase vertical level of MSM from 76 to 96 (with a topmost level raised from 21.8km to 37.5km)
  - Physics paramterization upgrade such as boundary layer, cloud macro and micro physics, introduction of a ocean mixing model etc .
  - Extended forecast range from 51 to 78 hours for TC prediction
  - Assimilation of stratosphere- sensitive channels of microwave sounders along with the model-top extension.
- Local NWP system
  - Introduction of a hybrid-3DVAR DA system.
    - The ensemble-based B obtained from the Meso-scale EPS

Recent development topics

- Coupled Prediction System for seasonal forecast and El-Nino prediction
  - Upgrade the system from CPS2 to CPS3
    - Atmospheric and ocean models upgrade
      - Increase resolution (AGCM: 110km-> 50km, OGCM:1deg.->0.25deg)
      - Upgrade of physics parameterization schemes
    - Introduction of 4DVAR to the ocean DA system
    - Planned to be installed in FY2021.

Update of GSM's mean orography





- Smaller scales are now resolved in TQ959.

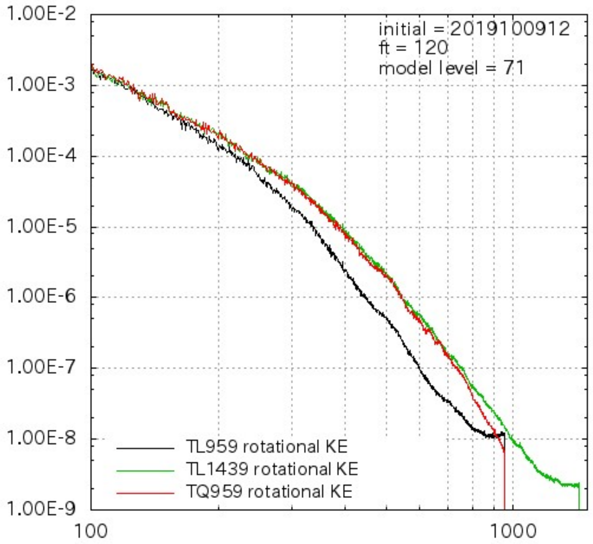
Yukihiro Kuroki

The power spectra of wind and temperature

Yukihiro Kuroki

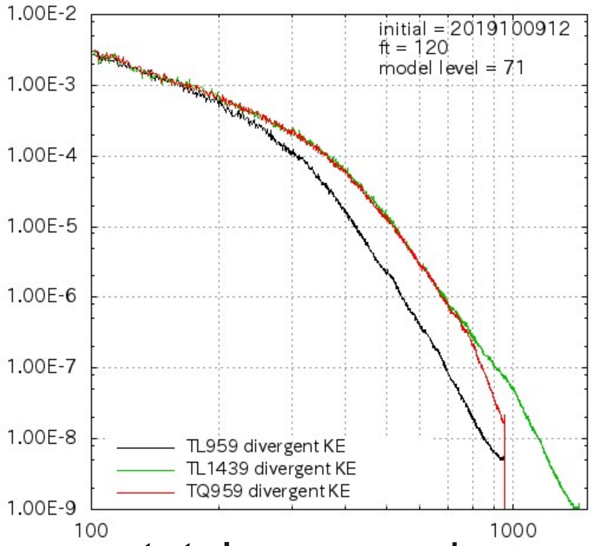
T+120 initialized at 12Z 09 Sep. 2019

Wind (rotational component)



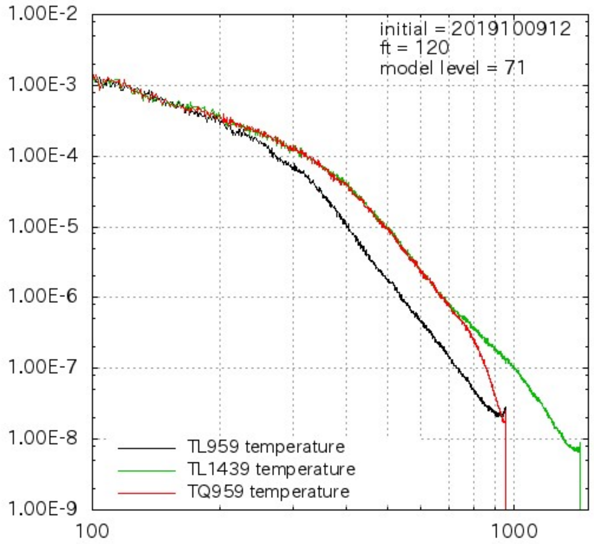
total wavenumber

Wind (divergent component)



total wavenumber

Temperature



total wavenumber

TL959 TL1439

TQ959+improvement to numerical diffusion and model orography

- Quadratic grid eliminates spectral blocking in the rotational components of the wind
- Improvement to the model orography and mitigate the damping of the spectra in the mid to high wavenumber compontent
- The power spectra of TQ959 is comparable to that of Tl1439 up to wavenumber 800.

HPC READINESS

HPC readiness: parallelization

- Improvement to parallelization in GSM (JMA operational Global Spectral Model)
  - Increase I/O ranks from 1 to 3
    - Speed up by ~200s for ~25min. Jobs (Tq959 (~13km) forecast up to 5.5day)
    - Output almost overlapped with calculation

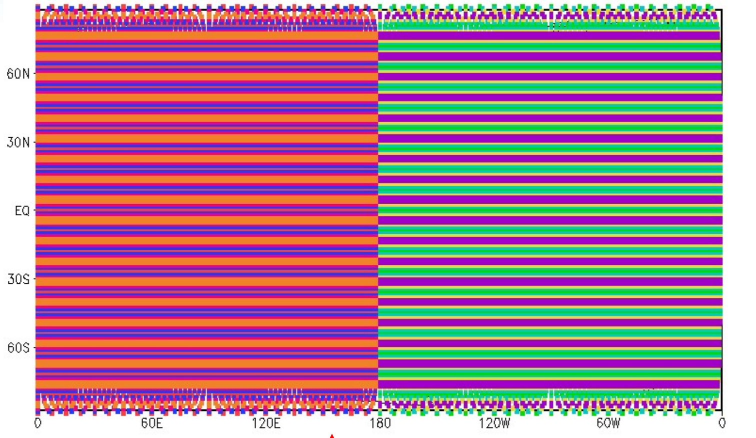
HPC readiness: parallelization

- Toward further improvement to GSM
  - Research on more efficient (and also grid model-friendly) MPI decomposition for SI-SL spectral models
    - Unifying MPI decomposition stages for physics parametrization and Semi-Lagrangian advection
      - Preliminary tests (dynamics only) showed that the new MPI decomposition reduced ~ 40% of computational cost in Tq959 jobs.
    - Also preparation for spectral/grid hybrid approach.

In a case of Tl159 2x4 MPI decomposition

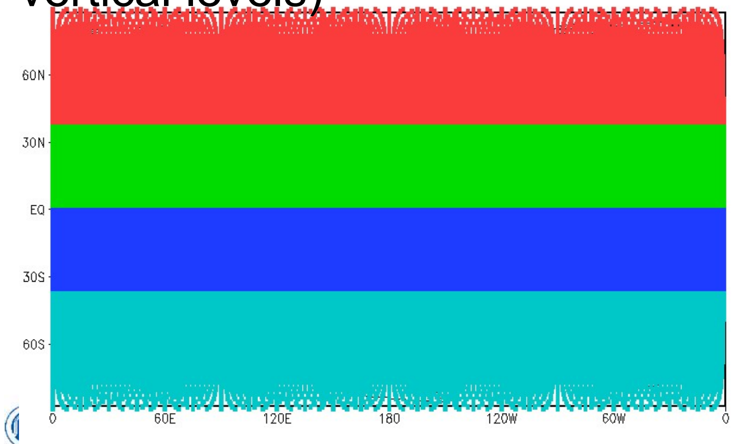
Current method

For parameterization, I/O and grid to spectral

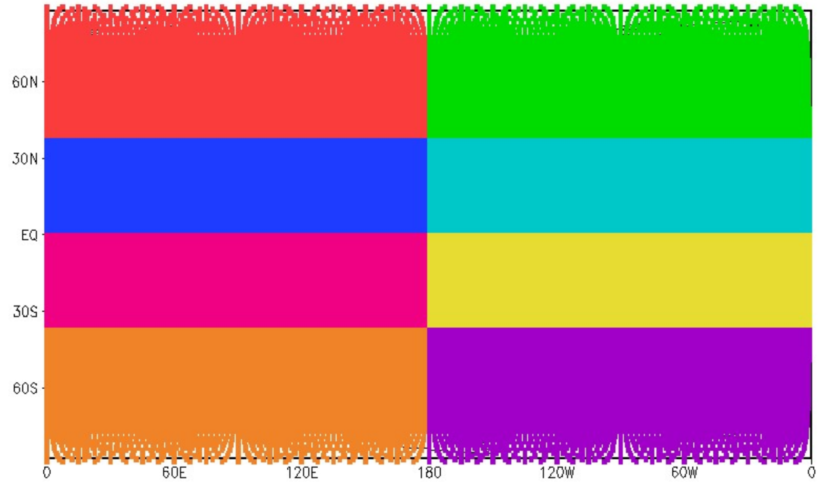


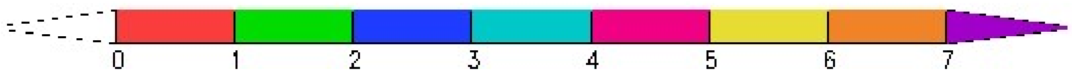
Transpose (with MPI communication) required every time step

For SL advection (first to Kmax/2 th vertical levels)



Unified grid stage





MPI rank

HPC readiness: GPU adaptation

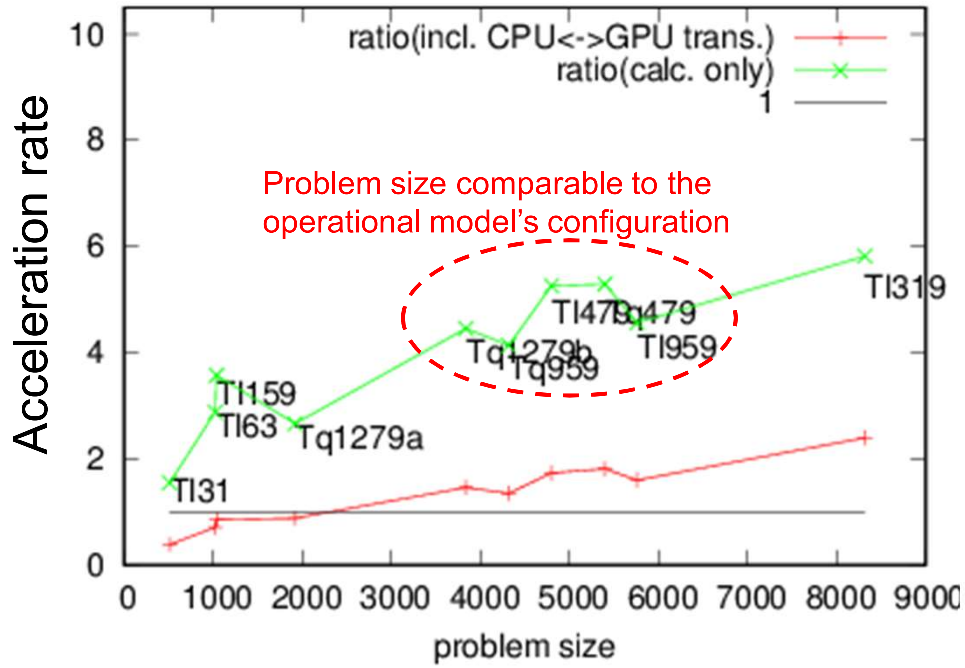
- GPU adaptation research using OpenACC directives started.
  - Use small models which extract or imitate “bottlenecks” of the actual NWP models as a first step.
    - Tested on NVIDIA Tesla V100 (GPU) and Intel Xeon Gold 6226 (CPU, 2.70GHz)
    - Regional model (ASUCA) : An advection scheme based on Koren (1993)’s flux limiter function and a tri-diagonal matrix solver for elliptic equations
      - In the same problem size as in the NWP configurations, OpenACC is ~5 faster than CPUs with OpenMP parallelization
    - Global model (GSM): A bottleneck subroutine in a cumulus parametrization scheme.
      - (Only in calculation part) In the same problem size as in the NWP configurations, OpenACC is x5 faster than CPUs with OpenMP parallelization
      - However tests also suggested that cost of GPU <-> CPU transfer can be another bottleneck.
    - Note that acceleration ratio of GPU to CPU depends on problem size
- Held an OpenACC hands-on seminar by NVIDIA (to familiarize with GPU)

Acceleration by GPU

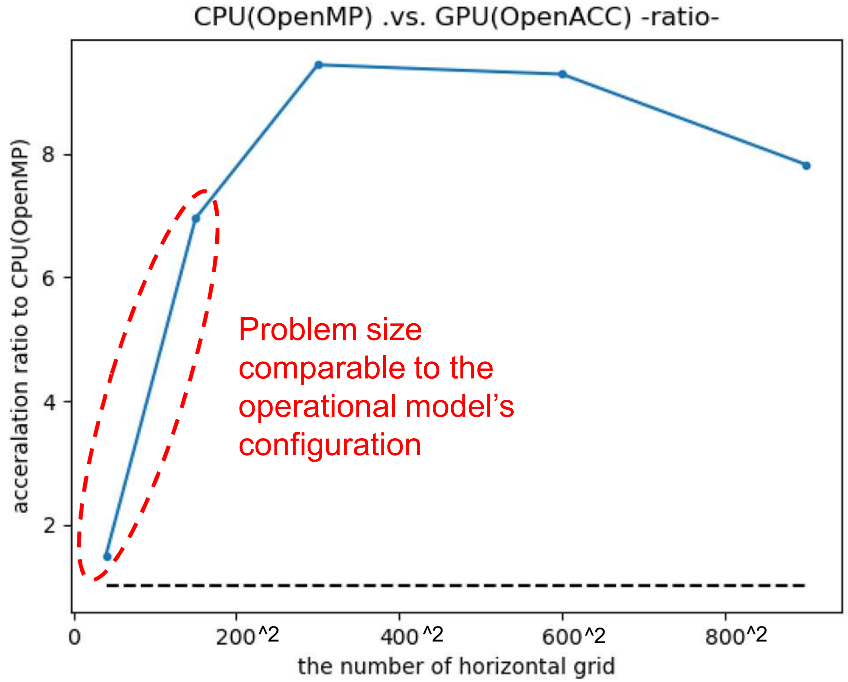
Acceleration rate = (elapse time on CPU) / (elapse time on GPU)

Problem size = (number of horizontal grid points per MPI process)

A bottleneck subroutine case in GSM's cumulus parameterization
(against CPU with OpenMP 12 threads)



An advection model case which imitates the regional model (ASUCA) 's bottleneck part
(against CPU with OpenMP 8 threads)



* When the problem size is comparable to that in actual forecast models, GPU can accelerate calculation by ~ x5 against OpenMP parallelized case on CPU.
* Large CPU->GPU transfer cost

Ryoga Ishida, Yuji Kitamura and Masashi Ujiie

MACHINE LEARNING FOR MODEL DEVELOPMENT

Machine learning for model development: JMA's activity

- Familiarization with machine learning for model developers as a first step
  - Toy model approaches using a multiscale Lorenz96 model
    - Testbed of replacing subgrid parameterization with neural net work methods (NN)
      - » Influence of a small scale model to a large scale model is the target of paramterization
    - Learn know-how of "tuning" (e.g. training methods, periods, etc) in ML
  - Planned to hold a training course on AI / ML by external lecturers

Emulating subgrid scale parameterization based on a multiscale Lorenz96 model

A toy model in which models with different scale are coupled to represent interaction between scales (Lorenz 2006, Thorns, et al. 2017).

Low-res. High-res. Truth

$$\frac{dX_k}{dt} = X_{k-1}\left(X_{k+1} - X_{k-2}\right) - X_k + F - \frac{hc}{b}\sum_j Y_{j,k}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k - \frac{he}{d}\sum_i Z_{i,j,k}$$

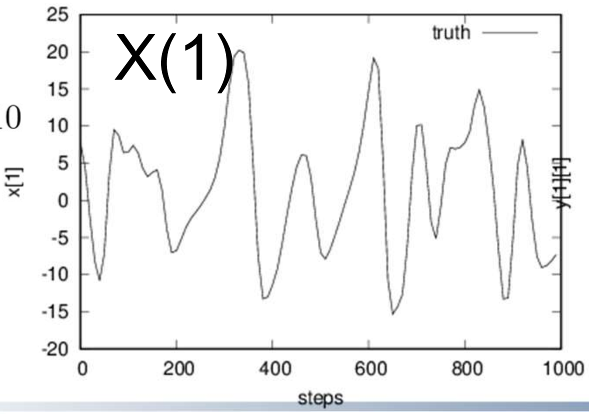
$$\frac{dZ_{i,j,k}}{dt} = Z_{i-1,j,k}\left(Z_{i+1,j,k} - Z_{i-2,j,k}\right) - geZ_{i,j,k} + \frac{he}{d}Y_{j,k}$$

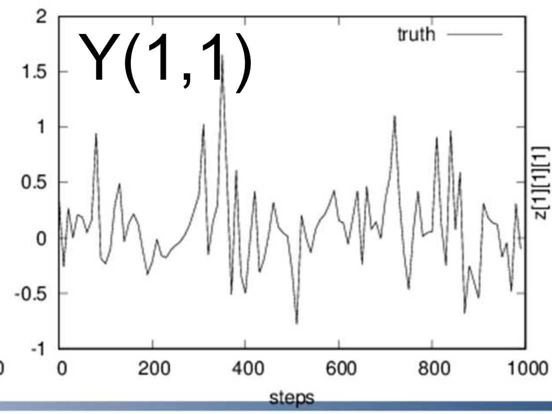
X (8 grid points)

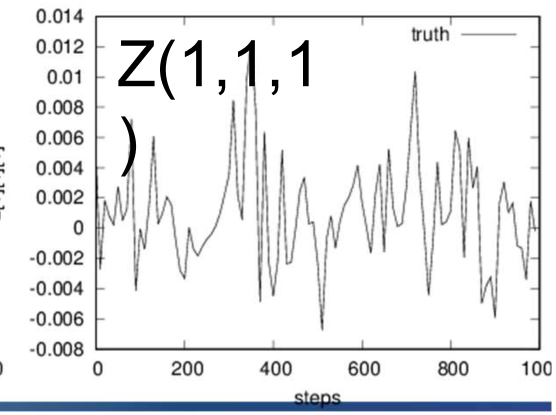
Y (8 grid points per X grid point)

Z (8 grid points per Y grid point)

$$b = c = d = e = g = 10$$
$$h = 1$$
$$F = 24$$
$$\Delta t = 0.05$$



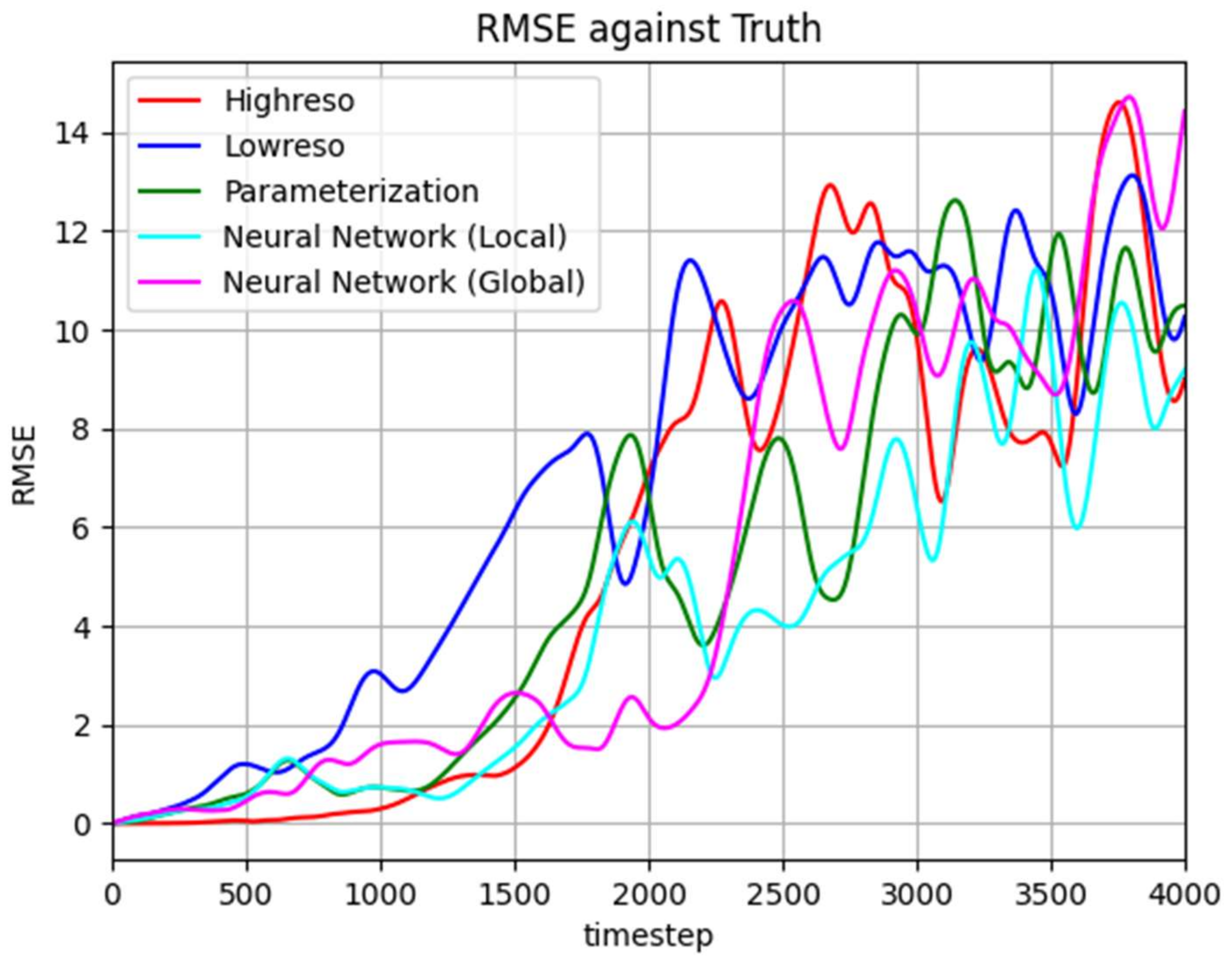




$$\frac{dX_k}{dt} = X_{k-1}\left(X_{k+1} - X_{k-2}\right) - X_k + F - \frac{hc}{b}\sum_{j} Y_{j,k}$$

Parameterize [ ] by regression (fourth-order polynomial of X) and NN

* Truth as training data



Parameterizations by NN results in better score than a low resolution model and is comparable to parameterization by a regression formula