Reduced precision computations in the SL-AV global atmosphere model

Mikhail Tolstykh[1,2,3], Gordey Goyman[1,3,2], Ekaterina Biryucheva[2], Vladimir Shashkin[1,2,3] , Rostislav Fadeev[1,3,2],

1- Marchuk Institute of Numerical Mathematics RAS,

2- Hydrometcentre of Russia,

3 – Moscow Institute of Physics and Technology

26/09/2023

Russian Supercomputing days`23

Global atmosphere model

- **Dynamical core:** solving 3D Reynolds-type equations (averaged Navier-Stokes equations) at the rotating sphere.

  - requires some degree of implicit time integration (can be a semi-implicit scheme or locally vertical solvers)

  - 25-45 % of total elapsed time.

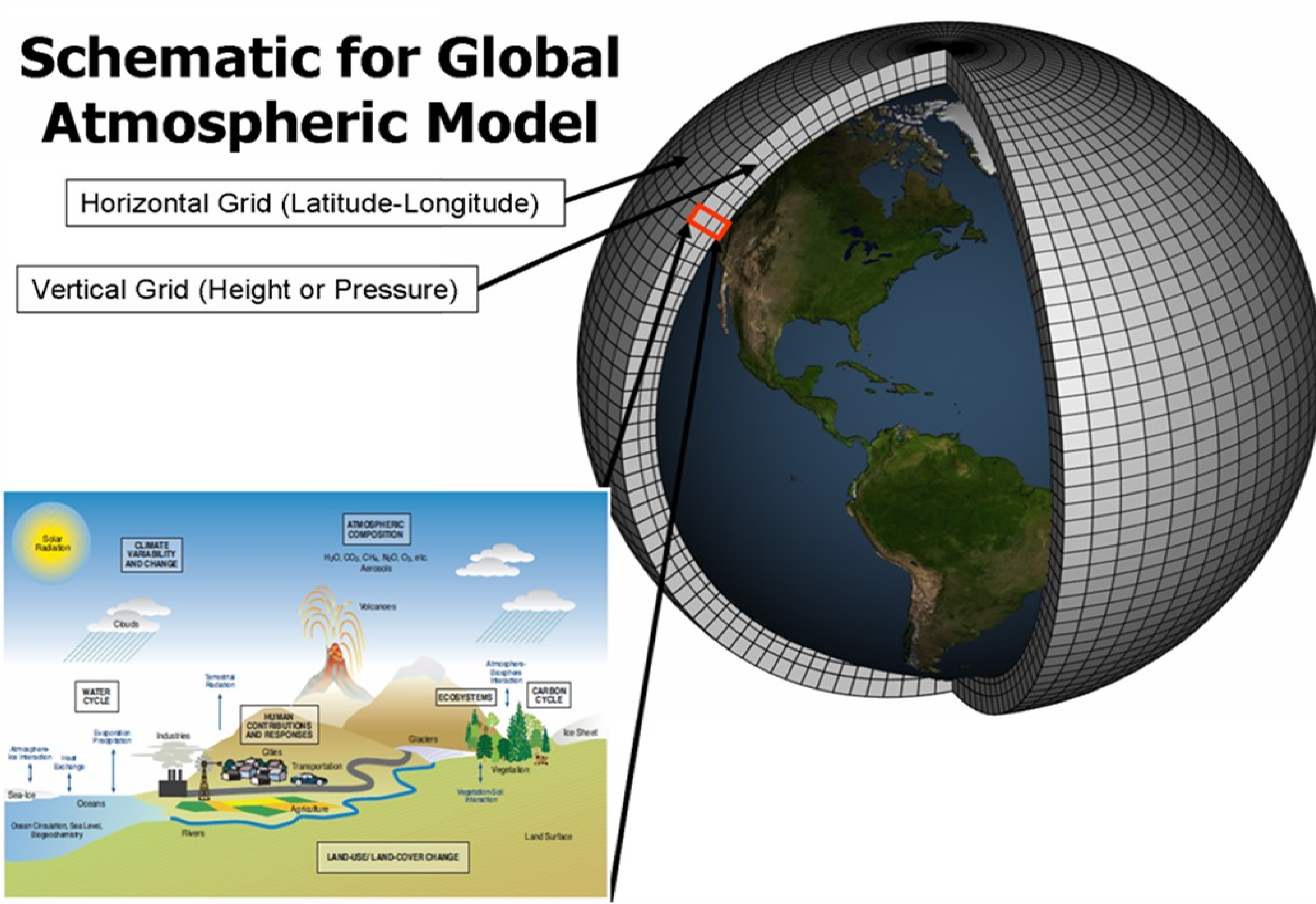
- **Right hand sides (parameterizations of subgrid scale processes)**:

  - usually locally 1D in vertical

  - the values at gridpoint (k,i,j) depend only on the values from (1:Kmax,i,j)

  - 55-75 % of total elapsed time

Atmospheric model schematics



PERSPECTIVE
https://doi.org/10.1038/s43588-021-00023-0

nature computational science

Check for updates

The digital revolution of Earth-system science

Peter Bauer[1]✉, Peter D. Dueben[1], Torsten Hoefler[2], Tiago Quintino[3], Thomas C. Schulthess[4] and Nils P. Wedi[1]

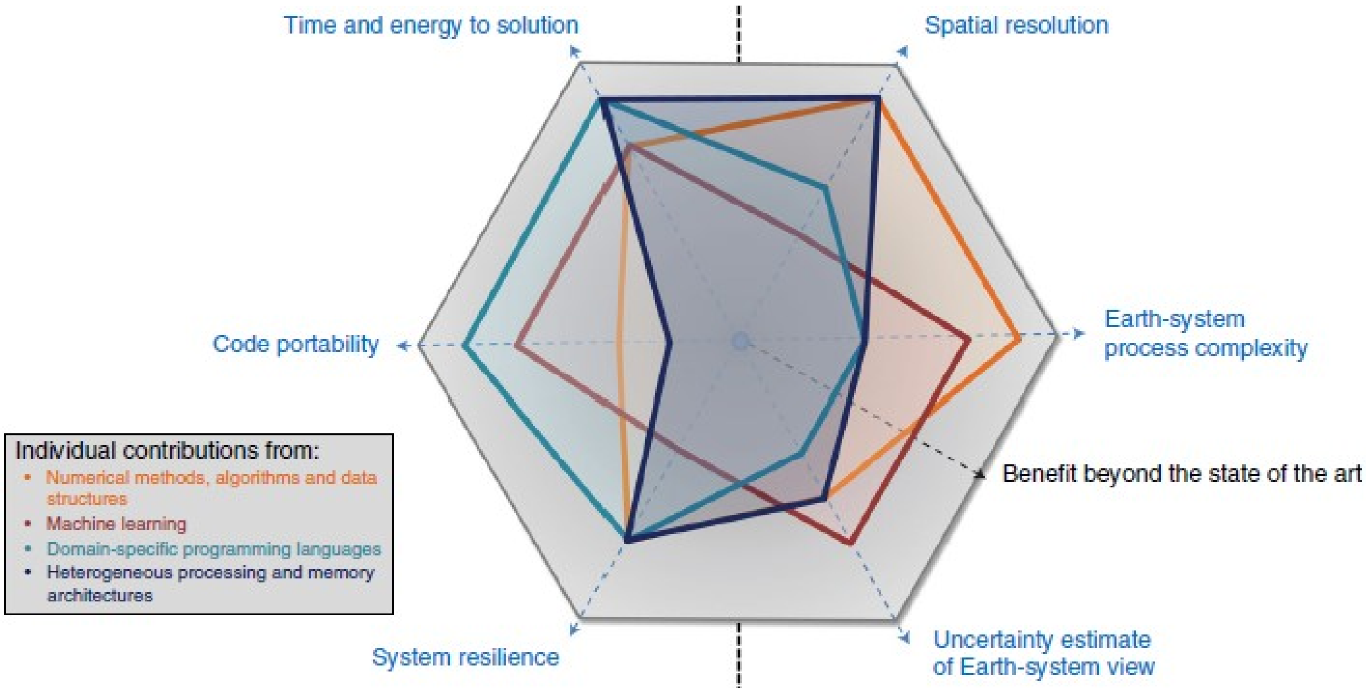


Fig. 4 | Expected contribution of main system developments necessary to achieve key science and computing technology performance goals. The distance from the center of the hexagon indicates the magnitude of the individual contributions towards enhanced efficiency for increased spatial resolution, more Earth-system complexity and better uncertainty information provided by ensembles as well as resilient, portable and efficient code and workflow execution, respectively.

Russian operational SLAV model

Federal Service for Hydrometeorology and Environmental Monitoring

HYDROMETEOROLOGICAL CENTRE OF RUSSIA

10-days operational medium range forecasts

0.1° in lon, 0.08°-0.13° in lat, 104 levels.

LETKF-based ensemble prediction system

0.9° lon, 0.72° lat, 96 levels.

New system 0.225° lon, 0.15-0.24° lat, 51 levs

Subseasonal and seasonal probabilistic forecast

(WMO S2S Prediction project)

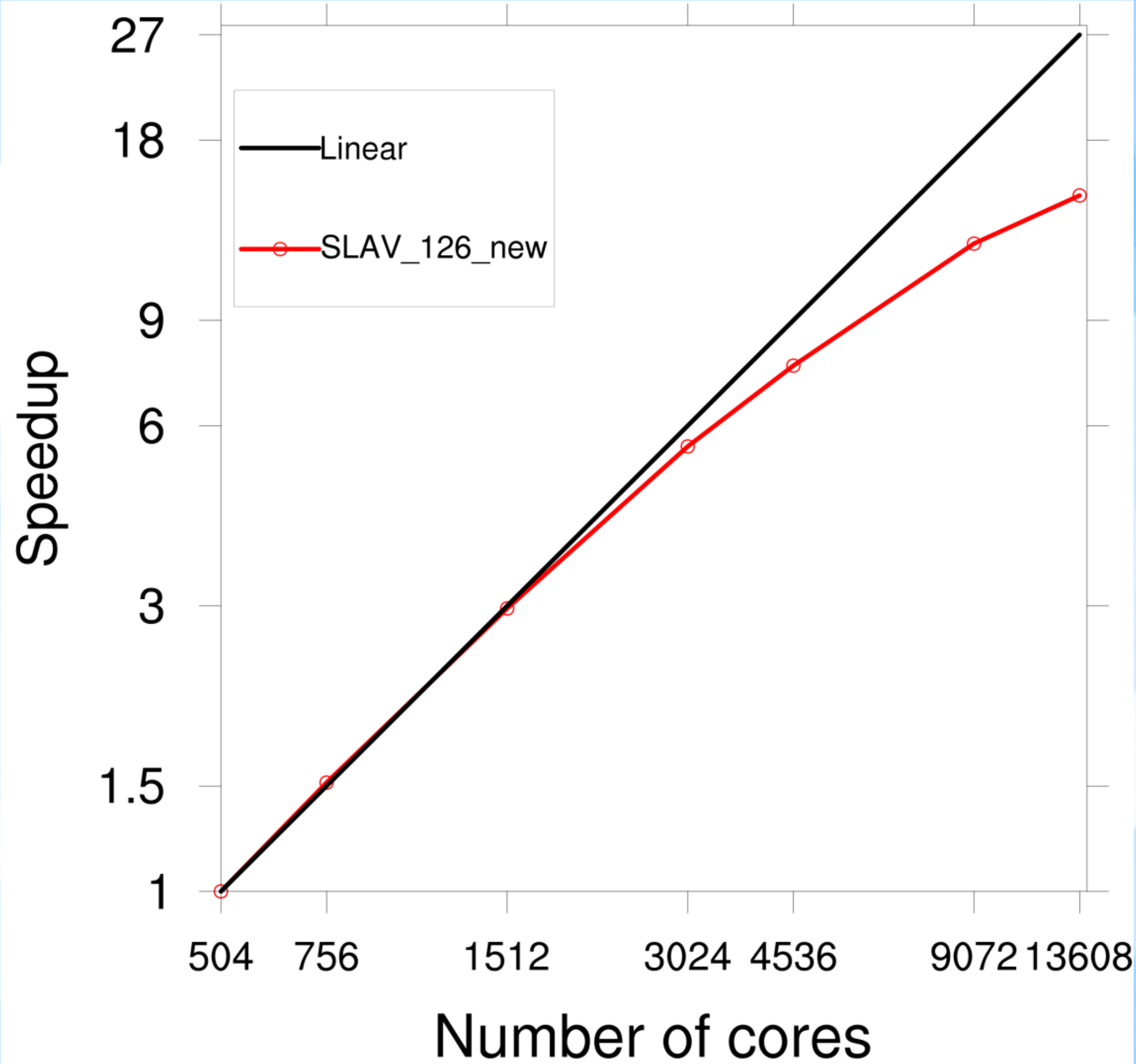
0.9°x0.72°L96

SL-AV global atmosphere model

SL-AV: Semi-Lagrangian, based on Absolute Vorticity equation

- Finite-difference semi-implicit semi-Lagrangian dynamical core (Tolstykh et al, GMD 2017). Vorticity-divergence formulation, unstaggered grid (Z grid), 4th order finite differences. Possibility to use variable resolution in latitude.
- Many parameterizations algorithms for subgrid-scale processes developed by ALADIN/ALARO consortium.
- Parameterizations for shortwave and longwave radiation: CLIRAD SW + RRTMG LW.
- INM RAS- SRCC MSU multilayer soil model (Volodin, Lykossov, Izv. RAN 1998).

SL-AV code parallel speedup at Cray XC40 w.r.t to 504 cores (2018 !)



Horizontal grid of 3024x1513 points (~13 km). 126 vertical levels

Отключение России от источников информации:

- Начальные и граничные условия для ICON; уменьшение кол-ва данных для COSMO
- Данные по ТПО и морскому льду OSTIA
- Европейские спутники с шагом в 15 мин
- Всемирный центр зональных прогнозов

(гражданская авиация)

РОССИЯ
АКЦИОНЕРНОЕ ОБЩЕСТВО "АВИАКОМПАНИЯ "РОССИЯ"

ул. Пилотов, д.18, корп. 4, Санкт-Петербург, 196210;
Ленинский проспект, д.15А, Москва, 119071
Тел. +7 (812) 633 37 00, Факс. +7 (812) 633 38 10
head_office@rossiya-airlines.com

ОКПО 01128564 ОГРН 1117847025284
ИНН 7810814522 КПП 997650001

05.05.2022 № 004/62

На исх. № ________ от ________

Руководителю Федерального
агентства службы по
гидрометеорологии и мониторингу
окружающей среды

Шумакову И.А.

Нововаганьковский пер., д. 12
г. Москва, 125993

Об анализе метеорологических
данных

Уважаемый Игорь Анатольевич!

Во исполнение пункта 6 протокола совещания Росавиации 60/05-ПР от 04.04.2022 в АО «Авиакомпания «Россия» был проведен сравнительный анализ метеорологических данных, предоставляемых Росгидрометом, на картах ветра и температуры и картах особых явлений погоды с данными, наблюдавшимися в полете.

По докладам экипажей воздушных судов расхождения данных не отмечено. Предоставляемая графическая метеорологическая информация может использоваться при планировании и выполнении полетов.

Летный директор Г.В. Баринов

Previous works on code optimization

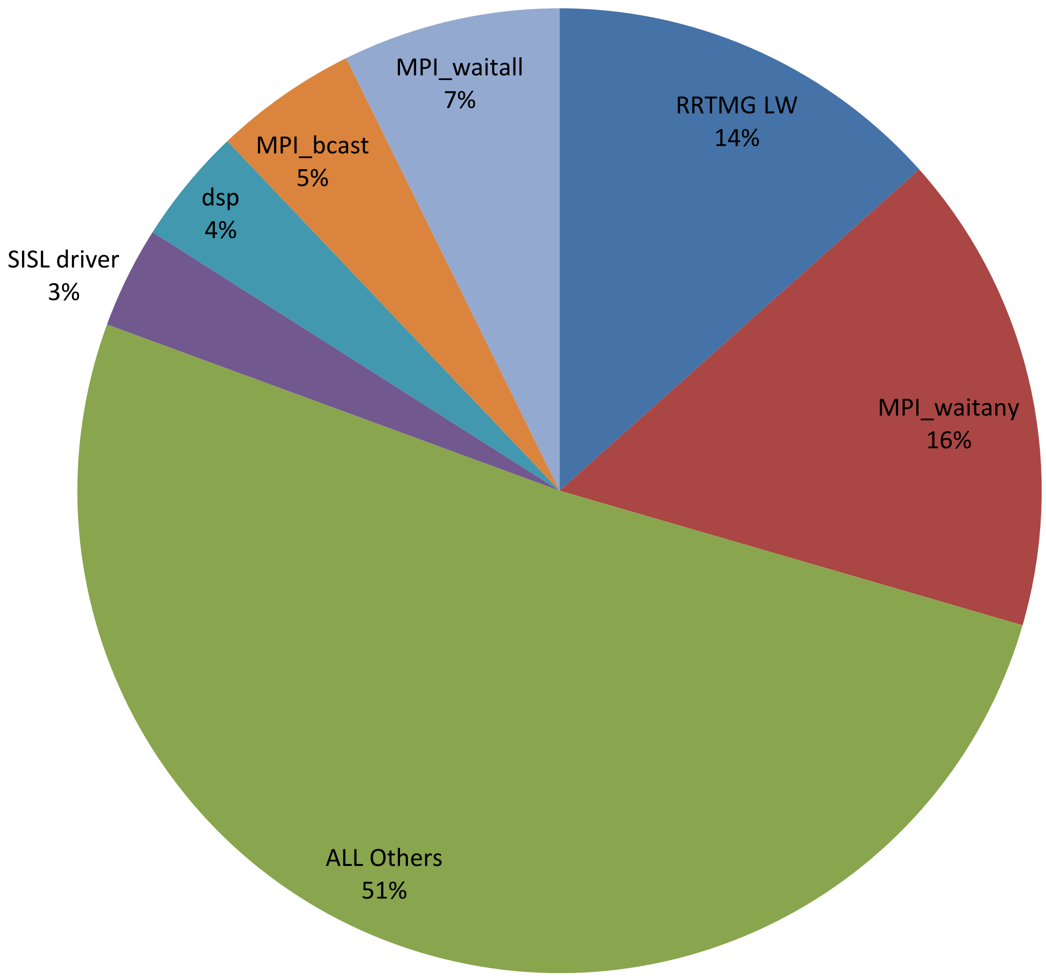
- Algorithmic improvements in dynamics – increase of the timestep 2.3 times
- NetCDF-based parallel I/O in the operational technology
- Memory access optimizations
- Single precision in data trasposition
- => 3.6 times acceleration of the forecast computations, 14 min per 24h forecast
- New version becomes more expensive after these works…

Motivation

Operational resources are limited to ~3000 processor cores taking into account other applications

1. New version SLAV10 with ~10 km horizontal resolution (3600x1946x104 grid) - needs to compute forecast for 24 hours preferably in 10 min.
2. Long-range ensemble prediction – extraction of weak signal from strong noise. Needs tuning in climate mode and large ensembles. (400x250x96 grid, ~75 km resolution)
3. Medium-range ensemble prediction system needs to increase the resolution to ~25 km

Profiling of SLAV10 at 2916 cores, Cray XC40



RRTMG LW – radiation
DSP – non-orographic gravity wave drag

Moving computations to single precision

```
real(kind=8) ::  var
```

becomes

```
real(kind=RB) ::  var
```

```
a(i) = 10.0 * b(i) + 42.0_8
```

becomes

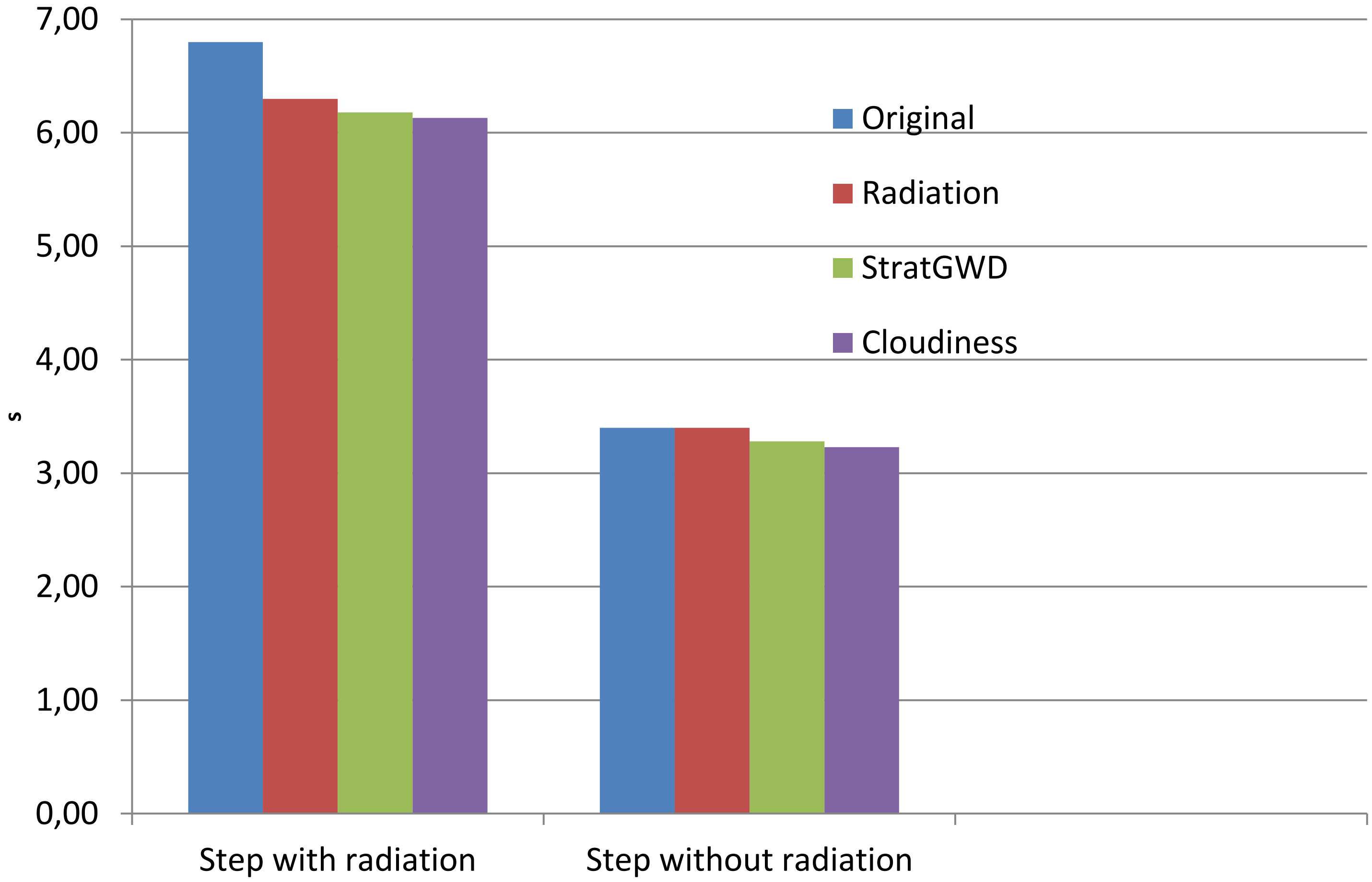
```
a(i) = 10.0_RB * b(i) + 42.0_RB
```

```
interface func
function func_sp(a)
real(kind=4), intent(in) :: a
real(kind=4) :: func_sp
end function func_sp

function func_dp(a)
real(kind=8), intent(in) :: a
real(kind=8) :: func_dp
end function func_dp

end interface
```

Problems in implementation of single precision (Vana et al MWR 2017)

- Expressions of type a/b for small a and b
- Many limiters in the code such as 1.e-13
- Spherical trigonometry and all related computations cannot be moved to single precision.
- Some expressions involving standard functions (exp, log,…) need special attention
- Mixture of double and single precision computations adds conversion cost.

Reduction of time per time step



Improving memory access with OpenMP

```
!$omp parallel do private(ist, iend, j)
do thread_idx = 1, threads_num
ist  = 1 + (NLON / threads_num) * (thread_idx-1)
iend = ist + (NLON / threads_num) - 1
 do j = jbeg, jend
    call calc_rhs(ist, iend, u(1,ist,j),
v(1,ist,j),…)
 end do
end do
```

```
!$omp parallel do private(ist, iend, j,
block_idx)
do idx = 1, blocks_num * (jend - jbeg + 1)
 j = jbeg + (idx - 1) / blocks_num
 block_idx = mod(idx - 1, blocks_num) + 1
 ist  = 1 + NLON / blocks_num * (block_idx - 1)
 iend = ist + NLON / blocks_num - 1
 call calc_rhs(ist, iend, u(1,ist,j),
v(1,ist,j),...)
end do
```

- NLON/blocks_num=16

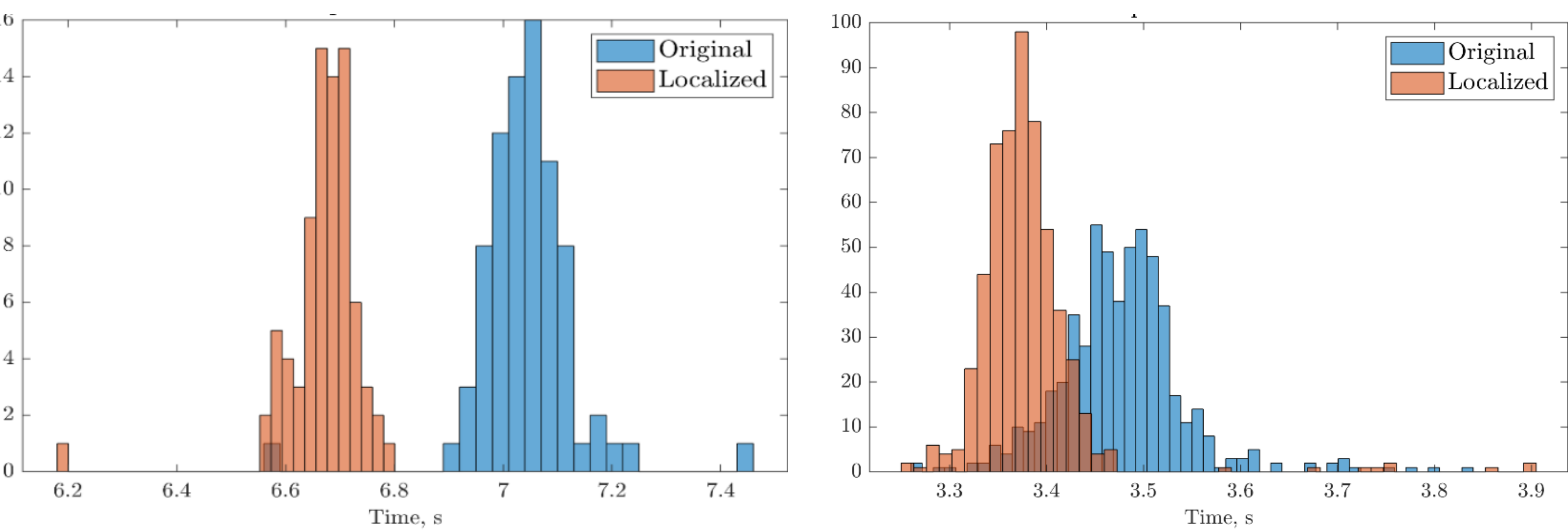
Array localization

```
real (kind=RB) :: A(NLON,NLEV)
```

is replaced with

```
real (kind=RB) :: A(NLON/blocks_num,NLEV)
```

time step with radiation (left) and without (right)



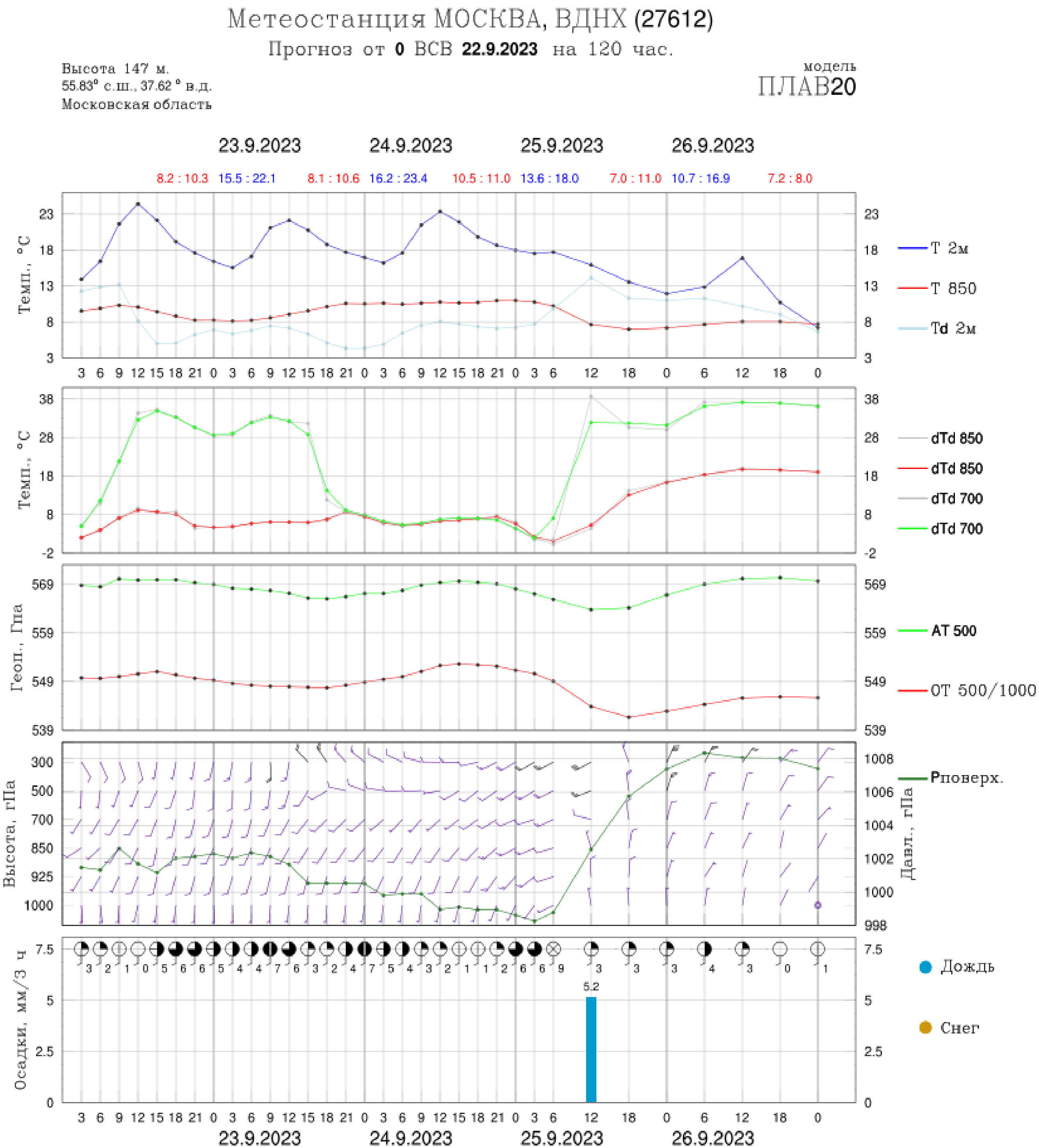
Current results at 2916 cores

- Earlier, increase of time from 14 to 16 min (output every 12hrs) as model complexity is increased
- Reduction from 16 to 14 min due to single precision computations and memory access improvements.
- Further work is ongoing (incl. FFTW instead of text FFTs, ..)

Conclusions

- Acceleration of SL-AV10 code has allowed to offset the increase in model complexity

- Further code acceleration is needed, especially to launch the new ensemble prediction system, the work is underway



Thank you for attention!