

Improving the Computational Efficiency of the Global SL-AV Numerical Weather Prediction Model

Mikhail A. Tolstykh^{1,2} , *Rostislav Yu. Fadeev*^{1,2} ,
Vladimir V. Shashkin^{1,2} , *Gordey S. Goyman*¹ 

© The Authors 2021. This paper is published with open access at SuperFri.org

The recent works on improving the efficiency of the Russian SL-AV global numerical weather prediction model both for medium- and long-range forecasts are described. The algorithmic improvements of SL-AV dynamical core, implementation of parallel I/O and several code optimizations are presented. We investigate the impact of single precision computations in some parts of the code on present climate simulations. As a result of efforts described in this article, we are now able to compute a 24-hour forecast for the model version having about 10 km horizontal resolution and 104 vertical levels in 13 min using 2916 processor cores of Cray XC40 system. This timing allows multiple experiments for tuning this new model and fits the requirements for operational weather forecast. The single long-range forecast with low-resolution SL-AV version now takes just 89 minutes instead of 111. We have also verified that the partial utilization of single precision computations produces approximately the same model climate as the previous version with fully double precision computations.

Keywords: numerical weather prediction, global atmosphere model, computational efficiency, I/O optimization.

Introduction

The common ways to improve medium-range (3–10 days) numerical weather prediction is, first, to increase the prognostic model resolution, second, to take into account the model uncertainty, and, third, to replace the atmosphere model with the coupled many-component model incorporating atmosphere, ocean sea-ice models called Earth system model. Accounting for model uncertainty is accomplished by the ensemble prediction that uses 20–100 runs of the same model incorporating some perturbations and starting from the perturbed initial conditions [12]. Using ensemble technique makes it also possible to produce a probabilistic forecast of area-averaged anomalies of weather parameters for months ahead. Both medium-range and long-range forecasting is usually done with the same global atmosphere model. As smaller scales are less predictable than larger scales, a lower-resolution model is applied for long range prediction. All the above mentioned applications require huge computer resources for timely forecast delivery. Many computer systems of the world weather forecasting centres are present in the supercomputer Top500 list [13].

A modern global atmosphere model should be able to use efficiently up to hundred of thousands of processor cores. At the same time, the new concerns about climate change require to weigh the advantages in weather prediction quality gained by the increase of resolution and/or model complexity with respect to electric power consumption [5]. All these considerations lead to increasing demands to the parallel efficiency of the atmosphere model code with different resolutions, along with its portability to different architectures.

In this paper, we describe recent works on improving the computational efficiency of the SL-AV global numerical weather prediction model both for medium-range and long-range forecasts

¹Marchuk Institute of Numerical Mathematics RAS, Moscow, Russian Federation

²Hydrometeorological Research Center, Moscow, Russian Federation

applications. This model is developed at Marchuk Institute of Numerical Mathematics Russian Academy of Sciences and Hydrometcenter of Russia. This model is applied for operational medium-range and long-range weather forecasts [17]. Algorithms and their parallel implementation using one-dimensional MPI decomposition and OpenMP loop parallelization are described in [15, 16]. The earlier code version demonstrated 53% efficiency at 9072 processor cores for $3024 \times 1513 \times 126$ grid (without I/O) [20].

There is a new version of SL-AV model with horizontal resolution of about 10 km and 104 vertical levels (SL-AV10). Some works on its optimizations are presented in [18, 19]. The elapsed time necessary to run a 24-hour forecast had reached 32 min at 4000 processor cores, without I/O. The results achieved earlier are not sufficient for operational application of this version of the SL-AV model that requires the 24-hour forecast to be computed in less than 20 minutes using less than 3000 processor cores. Furthermore, it is very time-consuming to carry out complex tuning of all model parametrizations for subgrid-scale processes that are mostly resolution-dependent. Such a tuning requires multiple numerical experiments involving a series of forecasts for different seasons.

In this paper, recent algorithmic improvements of SL-AV dynamical core (Section 1), implementation of parallel I/O (Section 2) and some code optimizations (Section 3) are presented. In Section 4, we study the impact of the partial use of single precision computations introduced earlier [19] on present climate simulation. All these works are summarized in Conclusions.

Cray XC40 system installed at Roshydromets Main Computing Center is used in all the tests described in this article. It consists of 936 nodes with two Intel Xeon E2697v4 18-core CPUs and 128 GB memory. All the nodes are connected with Cray ARIES inter-connect. The peak performance is 1.29 PFlops. The system includes Lustre parallel file system. We use Cray Fortran compiler version 10.0.3. We have also tried Intel Fortran Compiler version 19.1.254, similar results are obtained. We also use NetCDF library version 4.7.4 in this study.

1. Algorithmic Improvements

For a long time during the development of SL-AV model, we observed a noise in the numerical solution over the mountainous regions when using large values of the time step. To alleviate this problem, we had to decrease the time step that compromised model efficiency. Actually, the resulting time step value in operational medium range weather forecast was more than two times smaller than in the ECMWF IFS model [7] that is similar to SL-AV model in many aspects. In this section, we describe the modifications that allowed to get rid of the noise and consequently increase the time step value and hence improve model efficiency.

Upon inspecting the orographic noise in the model, it turned out that it consists of approximately 100–200 km-scale stationary wave modes in geopotential height field. Initially, we attributed this behaviour to the spurious orographic resonance – the known problem of semi-implicit semi-Lagrangian atmospheric models [11]. However, the spurious modes were insensitive to the common techniques of spurious orographic resonance damping [10, 11] (e.g., time off-centering of Crank-Nicolson scheme).

It was noted that the noise amplitude is sensitive to the settings of the horizontal diffusion block. Application of this kind of fields filtering is a common practice in atmospheric modelling caused by the need to avoid enstrophy clustering near the smallest resolved scales due to non-linear cascade [8]. The biharmonic hyper-diffusion operator with the implicit time-integration scheme is applied in SL-AV model [16].

The interesting fact is that this noise amplifies with the increase of diffusion coefficients, contrary to the behavior one could normally expect. We then come to the conclusion that the orographic noise in SL-AV model depends on the details of the diffusion implementation. To investigate this effect, the diffusion-driven stationary orographic noise model is developed. This model is based on the linearized shallow water model of [9] used to investigate properties of the spurious resonant response of a semi-implicit semi-Lagrangian model to the orographic forcing. The considerations [9] are modified to account for SL-AV specifics, the vorticity-divergence representation of the flow and horizontal diffusion implementation.

We start from 1D non-linear shallow water equations system:

$$u_t = -uu_x - gh_x - gb_x, \quad (1)$$

$$h_t = -uh_x - hu_x, \quad (2)$$

where u is the flow speed, h is the fluid layer thickness, b is bottom elevation (terrain height), g is the gravity acceleration, subscripts indicate partial derivatives in space (x) and time (t).

This system is linearised with respect to the reference state with uniform fluid thickness H , and wind speed U (u and h are now perturbations to the reference state):

$$u_t = -Uu_x - gh_x - gb_x, \quad (3)$$

$$h_t = -Uh_x - Hu_x. \quad (4)$$

We apply then the standard semi-implicit semi-Lagrangian time integration scheme based on the Crank-Nicolson method:

$$u^{n+1} = -\frac{\Delta t}{2} (gh_x^{n+1} + gb_x) + A \left(u^n - \frac{\Delta t}{2} (gh_x^n + gb_x) \right), \quad (5)$$

$$h^{n+1} = -\frac{\Delta t}{2} Hu_x^{n+1} + A \left(h^n - \frac{\Delta t}{2} Hu_x^n \right), \quad (6)$$

where A is the linearised semi-Lagrangian advection operator: $(Af)(x) = f(x - U\Delta t)$. Equations (5), (6) are linearised shallow water counterparts of the 3D equations used in SL-AV model. The SL-AV model reformulates this system in terms of vorticity and divergence (in 1D case only divergence is relevant variable). The prognostic equation for the divergence $D = u_x$ is obtained after differentiation of wind equation (5) in x :

$$D^{n+1} = -\frac{\Delta t}{2} (gh_x^{n+1} + gb_x)_x + \left[A \left(u^n - \frac{\Delta t}{2} (gh_x^n + gb_x) \right) \right]_x. \quad (7)$$

The height equation is also formulated using divergence, i.e. D is substituted for u_x type terms.

Consider system (6), (7) for one Fourier harmonic e^{ikx} : $(D, h)^T = (\hat{D}, \hat{h})^T e^{ikx}$ and $b = \hat{b} e^{ikx}$. The resulting equations are:

$$\hat{D}^{n+1} = -\frac{\Delta t}{2} (ik)^2 \left(\hat{h}^{n+1} + \hat{b} \right) + e^{-ikU\Delta t} \left(\hat{D}^n - (ik)^2 \frac{\Delta t}{2} \left(\hat{h}^n + \hat{b} \right) \right), \quad (8)$$

$$\hat{h}^{n+1} = -\frac{\Delta t}{2} H \hat{D}^{n+1} + e^{-ikU\Delta t} \left(\hat{h}^n - \frac{\Delta t}{2} H \hat{D}^n \right), \quad (9)$$

where $e^{-ikU\Delta t}$ is the Fourier image of semi-Lagrangian advection operator A .

The SL-AV solution procedure for system (8, 9) is as follows. First, \hat{h}^{n+1} is excluded from (8) using (9) and the Helmholtz problem is solved for the divergence:

$$\hat{D}^{n+1} + k^2 \frac{\Delta t^2}{4} gH \hat{D}^{n+1} = \hat{R}_{helm}, \quad (10)$$

where \hat{R}_{helm} is the combination of known time step n terms.

Second, artificial biharmonic diffusion (small scale filtering) is applied to the divergence:

$$\hat{D}^{n+1*} = \hat{D}^{n+1} / (1 + \Delta t C k^4), \quad (11)$$

where C is the diffusion coefficient. After divergence filtering, \hat{h}^{n+1} is calculated from equation (9) by substitution of \hat{D}^{n+1*} for D^{n+1} . After all, h^{n+1} is filtered using the same scheme as for divergence (11) (in the 3D model h is not filtered, the temperature that is closely related to h by hydrostatic equation is filtered instead).

From shallow-water considerations, it may seem that filtering divergence at the same time with h after solution of system (8, 9) will be more consistent. However, this early divergence filtering makes sense in 3D model because it implies filtering of updates for some derived fields like surface pressure and vertical velocity which will be left unfiltered otherwise (see Section 4 of [16]).

The calculation of \hat{D}^{n+1} , \hat{h}^{n+1} using the procedure described above can be summarized by the following equation:

$$\begin{pmatrix} \hat{D} \\ \hat{h} \end{pmatrix}^{n+1} = Q \begin{pmatrix} \hat{D} \\ \hat{h} \end{pmatrix}^{n+1} + R \begin{pmatrix} \hat{b} \\ 0 \end{pmatrix}, \quad (12)$$

where Q and R are the matrices with complex entries describing the solution procedure. We are interested in stationary orography-forced solutions of this system and their dependence on the wave number k and diffusion coefficients. Stationary solutions can be found with setting $\hat{D}^{n+1} = \hat{D}^n = \hat{D}_s$, $\hat{h}^{n+1} = \hat{h}^n = \hat{h}_s$ in equation (12):

$$\begin{pmatrix} \hat{D}_s \\ \hat{h}_s \end{pmatrix} = (I - Q)^{-1} R \begin{pmatrix} \hat{b} \\ 0 \end{pmatrix}. \quad (13)$$

We investigate stationary orographic response properties for three different options of diffusion application. The first is original SL-AV diffusion where the divergence is filtered right after the Helmholtz problem solution and h field is not filtered at all (the diffusion coefficient for the temperature in SL-AV is significantly smaller than for divergence). This option will be referred to as ‘reference’. With the second option, the divergence is filtered after the calculation of h^{n+1} , h itself is not diffused (this will be referred to as ‘semi-consistent’). The third option is diffusion application for both D^{n+1} , h^{n+1} after their calculation with the same coefficient (‘consistent’ option).

Figure 1 shows the amplitude of height field response to orographic forcing obtained numerically using equation (13). The following non-dimensional parameters are used: $U = 0.3$, $H = 1$, $g = 1$, $\Delta t = 0.1$, $\hat{b} = 1$, $C = 10^{-7}$. The black curve in the figure shows the exact response amplitude, independent of the wavenumber k . The most spectacular feature in Fig. 1 is the blue curve showing perfect spurious orographic resonance in the absence of any dissipation mechanisms. The response amplitude reaches infinite values at about wave number $k = 29\pi$ that corresponds to the shortest scale motions in the real model. However, this curve is not relevant for the

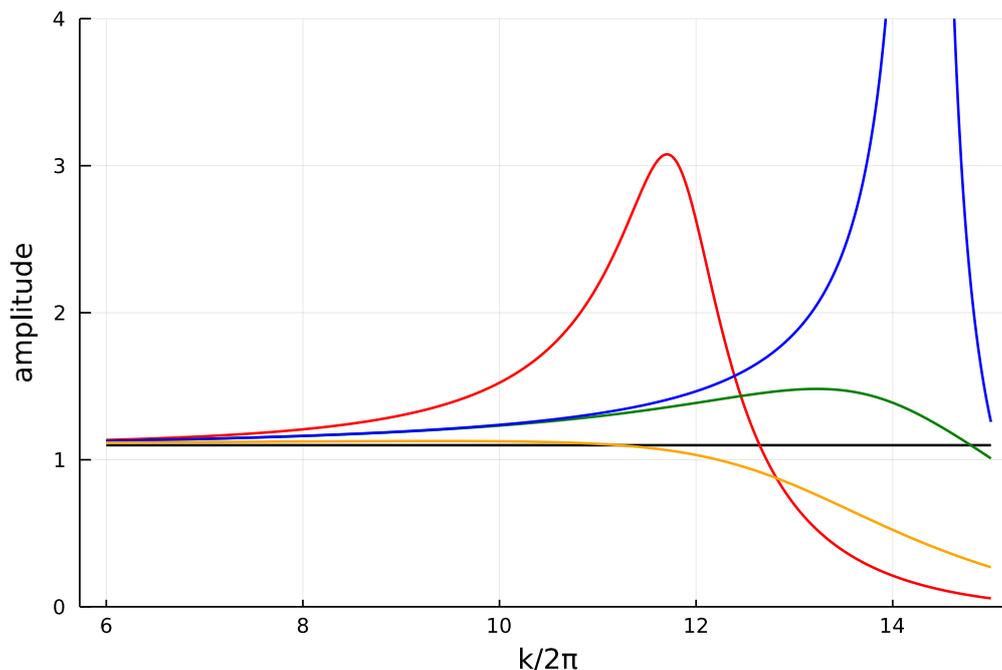


Figure 1. The amplitude of height field response to orographic forcing. Red curve – ‘reference’ scheme of diffusion application, green – ‘semi-consistent’ diffusion, blue – no diffusion, orange – ‘consistent’ diffusion, black – analytic response amplitude

real SL-AV model because both explicit (diffusion, Crank-Nicolson off-centering) and implicit (advection scheme damping) numerical dissipation will prevent the model from instability.

The red curve shows the response of the ‘reference’ scheme. This curve shows weak resonance with wavelengths that are short enough, but still far from the shortest scales resolved on the grid. We believe that exactly this scenario takes place in the model simulations. The reason for the amplitude growth is the inconsistent application of diffusion breaking the balance between D and h that makes orographic mode stationary and this is compensated by the exaggerated growth of h amplitude. At the shortest scales, the resonance is effectively eliminated.

The green and orange curves in Fig. 1 show the response amplitude for ‘semi-consistent’ and ‘consistent’ schemes. The response of ‘consistent’ scheme is very close to the exact curve for large and intermediate wavelengths, orographic waves are dumped out at the shortest scales. The ‘semi-consistent scheme’ shows very weak amplification of response for intermediate and short wave lengths.

Linearised shallow-water study, therefore, indicates that the ‘reference’ diffusion scheme can spuriously amplify the orographic response at the scales well-resolved on the grid. ‘Semi-consistent’ diffusion leads to much more accurate solutions. The best result is achieved with ‘consistent’ diffusion, suggesting that using the same diffusion coefficients for all fields might be favourable.

The difference between diffusion application schemes can be noticed in the non-linear 3D SL-AV model simulations as well as in the shallow-water model. The typical picture is given in Fig. 2 that compares two SL-AV20 (approximately 24 km horizontal resolution) six-hour forecasts of 500 hPa geopotential height field. Both forecasts are computed using the time step value $\Delta t = 540$ s that is 2.25 times greater than the operational time step. The forecast using ‘reference’ diffusion depicted in Fig. 2b suffers from evident orographic noise over North-Eurasia.

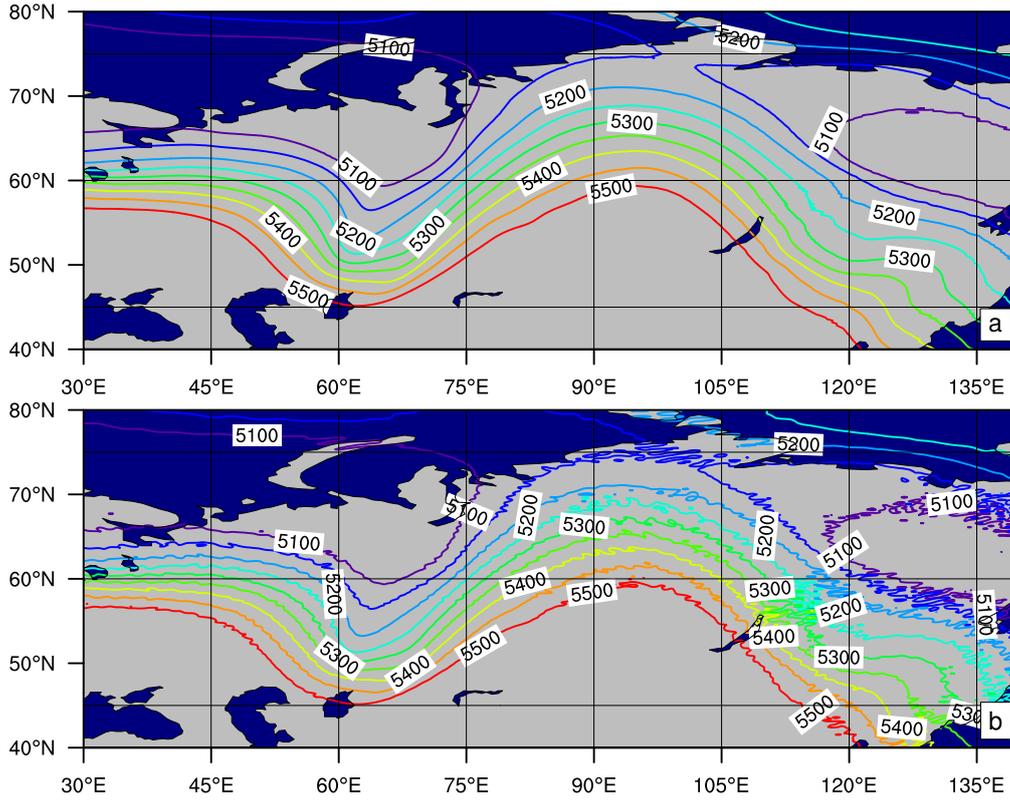


Figure 2. 6-hour forecast of 500 hPa geopotential height field using SL-AV20 model configuration with a) ‘semi-consistent diffusion’, b) ‘reference diffusion’

At the same time, the run with the ‘semi-consistent’ diffusion (Fig. 2a) is free of this deficiency. The run with the ‘consistent’ diffusion scheme (not shown) is very similar to the ‘semi-consistent’ one. That means there is no strong evidence for using the same diffusion coefficients for all fields in 3D model. The implementation of the ‘semi-consistent’ horizontal diffusion has allowed to increase the time-step size of the SL-AV model by a factor of 2.25 with respect to the one previously used. The elapsed time of the 24-hour forecast (without I/O) has reduced by the same factor.

2. Code Optimizations

Historically, SL-AV model used a single array for the state vector of the model $\vec{\varphi} = (u, v, T, q, D, \xi, \ln p_s)^T$ containing zonal and meridional components of wind field, temperature, specific humidity, horizontal divergence, relative vorticity, logarithm of the surface pressure, respectively. This array had the indices arrangement as $(Nlon, 6Nlev + 5, Nj)$ in earlier versions of the model, where $Nlon$ – is the number of grid points along longitudinal direction, $Nlev$ – is the number of vertical levels and Nj – is the number of grid points along latitudinal direction for a given MPI-process (the second dimension number is explained by the fact that the surface pressure is a two-dimensional variable stored along with its derivatives). Such a data storage organization was convenient in terms of computations parallelization (MPI and OpenMP parallelization along the latitudinal direction was used) since it allowed to perform MPI exchanges for all state vector fields at a time without using buffer arrays. A demand to use more processor

cores required to switch to use OpenMP loop parallelization along the other dimension (longitude or Fourier space wave numbers). Taking into account the large volume of the model source code, the simplest and most efficient way to organize such a transition required changing the structure of the state vector array to $(6N_{lev} + 5, N_{lon}, N_j)$. This allowed to increase the maximum theoretical number of cores used from 865 to more than 10000 for the version of the SL-AV model with the horizontal resolution of about 24 km [15]. However, new indices arrangement is not optimal in terms of memory access, since it spoils the localization in memory of a given grid-point field, especially in SL-AV10 model with about 10 km horizontal resolution. Indeed, previously $N_{lon} \cdot N_{lev}$ values laid sequentially in memory, while now it is only N_{lev} values. This led to a slowdown in the execution time of individual parts of the code, which at that time was not very significant and was an acceptable price to pay for increasing the model scalability.

In the last few years, a number of works were carried out [18–20] that allowed to significantly speed up the subgrid scale parametrizations and semi-Lagrangian advection blocks, being the most time consuming parts of the model. The recent profiling of the model showed that the slowdown associated with the use of a non-optimal state vector storage structure can no longer be considered insignificant. Thus, we decided to replace the above mentioned state vector with the individual arrays having dimensions (N_{lev}, N_{lon}, N_j) for each grid-point field in the state vector.

The implementation of these changes results in a 16–22% speed up of the time step elapsed time for the model with the grid dimensions $400 \times 251 \times 96$. In particular, the time needed to compute a single long-range forecast (in fact, single ensemble member) has decreased from 111 to 89 minutes. The effect of these optimizations is even more significant for the version of the SL-AV model with the horizontal resolution of about 10 km and 104 vertical levels ($3600 \times 1946 \times 104$ grid dimension). Experiments using 2916 processor cores (81 nodes with 6 MPI-processes and 6 OpenMP threads) show a decrease in execution time of a model time step without I/O by about 30%, which leads to a 7-minute reduction of a runtime needed to deliver a single 24-hour forecast.

3. I/O Optimizations

The typical horizontal resolution of a modern global atmosphere model (7–10 km) having the problem size of order 10^9 requires high I/O efficiency as the typical size of the initial data file is about some tens of gigabytes. Indeed, from 10 to 12 3D variables and about 20 2D variables need to be stored in this file. 3D variables in a modern atmosphere model include wind speed components, temperature, specific humidity, 4–5 hydrometeors (i.e., rain droplets and ice particles concentrations), ozone concentration, turbulent kinetic energy. Then the output forecast information with a size of 3 GB needs to be stored every 1–3 hours depending on forecast lead-time. This information usually consists of five 3D fields defined at isobaric surfaces (geopotential, temperature, wind speed components, relative humidity) and 2D fields (precipitation, near surface temperature, wind components, relative humidity, snow depth, etc).

It is known since long ago that the implementation of parallel input-output of data in an atmospheric model can significantly accelerate its execution. Gradual establishment of MPI-IO moved the focus towards interfaces convenient for atmosphere and Earth system models. So the incorporation of parallel capabilities based on MPI-IO into NetCDF freeware library commonly used in Earth system models and its model components [3] was natural. NetCDF file contains

meta-data information making it portable and searchable. This format is supported by many software packages used to manipulate, analyse and plot.

Historically, the GRIB (and GRIB2) formats [1] are generally accepted in the numerical weather prediction community, contrary to the NetCDF format widely used in climate modelling [6]. GRIB2 format allows to significantly compress data thus reducing the file size. Typically, a file in GRIB2 format is 2–3 times smaller than the file with the same single-precision information written in NetCDF4 format without compression. Unfortunately, the compression algorithm used in GRIB is essentially sequential. Recent NetCDF libraries (starting from version 4.7.4) include parallel compression [2, 3], still the size of NetCDF file obtained with parallel compression is significantly larger than GRIB2 file size.

There are advanced parallel I/O systems based on NetCDF format applied for many Earth system models coupling many component models (atmosphere, ocean, sea ice, etc.) [4, 21].

We have earlier implemented parallel I/O in the SL-AV model using parallel NetCDF standard library routines [14]. We have now implemented an improved version of the parallel I/O in the model but also in all technological accompanying code (preprocessing, postprocessing of the output fields). The parallel I/O is tested for the new SL-AV medium-range forecast version having horizontal resolution of about 10 km and 104 vertical levels.

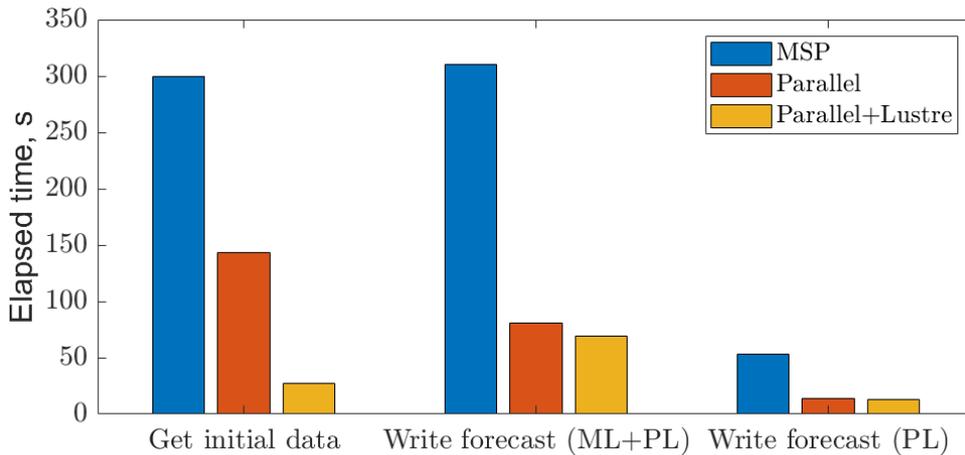


Figure 3. Elapsed time in seconds for different I/O steps of SL-AV model code while using 2916 cores at Cray XC40. ML means writing the information at model levels; the content mostly coincides with the initial data file. PL means writing prognostic information at pressure levels. MSP means sequential I/O at the master process with gather/scatter data from/to all other processes

The elapsed time for different I/O tasks in the SL-AV model using sequential and parallel I/O is shown in Fig. 3. The results of using Lustre file system capabilities for accelerating parallel I/O are also shown there.

One can see that using parallel I/O significantly accelerates this part of the code. Further acceleration is achieved while using Lustre file system options. For example, the procedure of reading the file with initial data is accelerated by a factor of 2.1 for parallel I/O alone and by factor of 10 if this file is physically located at different hard disk drives as set by `lfs setstripe` command.

Now the breakdown between different I/O components is as follows. Reading initial data at the beginning of the forecasts takes approximately 30 seconds. 70 seconds is required to write a file similar to the file with initial data which is used as a first guess file at the next forecast

cycle (6-hour forecast), and just 14 seconds is needed to write postprocessed data at pressure levels for the forecasts at all other lead times. This can be compared with the elapsed time of the usual model time step that takes 2.3 s for the time steps without radiation computations and 4.9 s for the time step including radiation calculations (this is every eighth step). Given that the output of forecast fields at pressure levels is required every three hours, the I/O elapsed time per forecast day is reduced from 440 sec to 120 sec for all forecast days other than the first one. The similar numbers for the first forecast day are 715 and 182 sec, respectively.

4. Evaluating Single Precision Computations for Climate Simulation

Earlier, the single precision calculations were introduced in some parts of the SL-AV model, namely, in the parts solving elliptic equations on the sphere, semi-Lagrangian advection and respective parallel data exchanges [19]. These parts of the SL-AV model are time consuming and include intensive parallel communications. We have found the impact of reduced accuracy on medium-range forecasts to be negligible [18], however, the impact of these changes on the model climate has not yet been investigated. We address this issue in this Section.

Details on implementation of single precision in the SL-AV program complex are presented in [19], so, we only briefly outline the main points here. The algorithm for solving elliptic equations requires a global data transposition (parallel communication of the all-to-all type), before and after the execution of this part of the code. To perform this communication, buffer arrays with compile-time defined data type (single or double precision) are used. That is, when using single precision, typecasting occurs during copying data to and from the buffer array within parallel communication phase. In the semi-Lagrangian advection block, the values of the grid-point fields to be interpolated to the departure points of the particles trajectories are stored at the single array. This array is directly used when performing parallel halo exchanges in this block, and its data type also can be switched to single precision. This allows to halve the size of data to be sent.

The following versions of the SL-AV model are considered in this Section: the model version with the reduced accuracy in the above mentioned parts, and a reference version of SL-AV with double precision computations in respective parts. Experiments are also performed for the ‘intermediate’ versions of the SL-AV model, where single precision is used either in the data transposition procedure (hereafter RATRAN experiment) or in the semi-Lagrangian scheme (RASL experiment). The horizontal resolution is 0.72 by 0.9 degrees in latitude and longitude respectively, the model has 96 vertical levels. The SL-AV model with this resolution is integrated for 5 years of model time, using I/O setup typical for long-range forecasts. It is worth to note that the new version of the model completes a year integration in 146 minutes as compared to 167 minutes for the version with double precision computations.

We define here the relative deviation of function f_2 from function f_1 as the l_2 norm $f_2 - f_1$ divided by the l_2 norm of f_1 . Figure 4 illustrates the relative deviation evolution for some model characteristics for the new version of the SL-AV model with reduced accuracy with respect to its reference version with double precision. The following fields are shown in this figure: the zonal wind at 250 hPa, precipitation, zonal wind, temperature and geopotential at 850 hPa. The averaging was performed over a period of integration time. It can be seen that the relative deviation decreases with time for all of the variables (including those not presented in the figure).

The maximum values of the relative deviation are reached on variables with large gradients: the zonal wind and precipitation. Large gradients in these fields are achieved at the boundaries of jets (for the zonal wind) and mountains (for precipitation). It should be noted that the relative deviation of the zonal wind at 850 hPa is much larger than at 250 hPa, where much stronger wind values are achieved. This means that the sensitivity of the atmospheric circulation in the SL-AV model with respect to the accuracy of the calculations decreases with height.

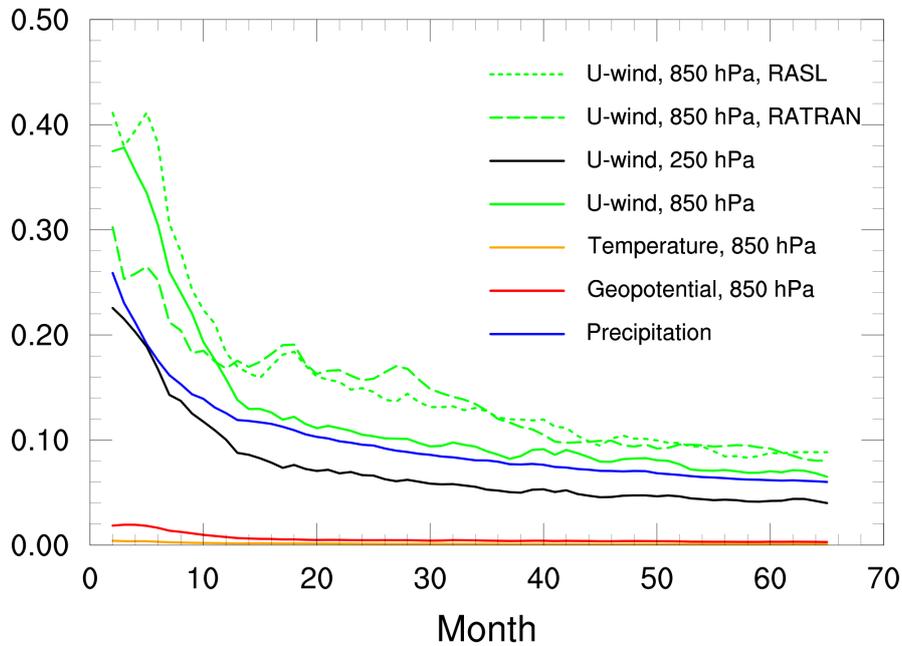


Figure 4. Relative deviation as a function of time (months) of the new version of the SL-AV model with reduced accuracy with respect to its reference version with double precision

The maximum value of the relative deviation of the 5-year averaged fields does not exceed 10%. Figure 5a illustrates the 5-year averaged zonal wind in the experiment based on the new version of the SL-AV model with reduced accuracy. The deviation of this variable from that obtained using the reference version of the SL-AV model with double precision is shown in Fig. 5b. It can be seen that the maximum value of the relative deviation of the averaged fields does not exceed 10% and is achieved due to a small shift of the jets. However, the magnitude of the shift is not large and therefore it does not significantly affect most of the forecast fields.

The dashed and dotted green curves in Fig. 4 correspond to the relative deviation of the time-averaged zonal wind at 850 hPa obtained in experiments performed using a version of the SL-AV model with a partial transition to the reduced accuracy. The dotted curve here corresponds to the application of single precision in the transposition procedure (RATRAN experiment), and the dashed curve corresponds to single precision in the semi-Lagrangian scheme (RASL experiment). It can be seen that the relative deviation of individual modifications compared to the reference version of the SL-AV model is larger. The cumulative effect of introducing single-precision calculations leads to a reduction in the relative deviation.

One can conclude that the introduction of the single precision calculations in the above mentioned parts of the model does not affect the model climate to a significant extent.

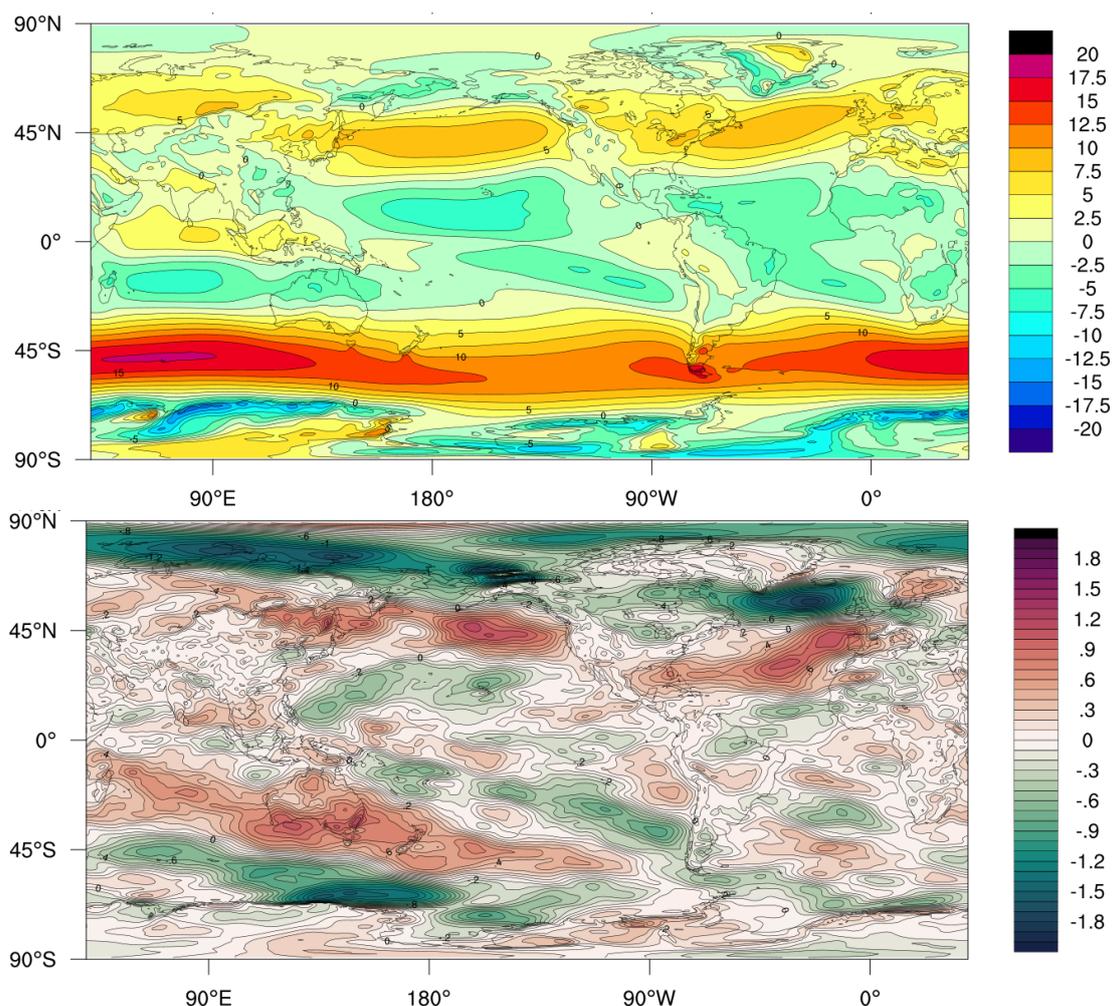


Figure 5. The 5-year time-averaged zonal wind at 850 hPa (top) and its absolute deviation in the version of the SL-AV model with the reduced accuracy compared to its reference version (bottom)

Conclusions

Initially, the elapsed time necessary to compute the 24-hour forecast with the SL-AV10 model with the horizontal resolution of about 10 km and 104 vertical levels using about 4000 processor was 42 minutes without time for I/O. The efforts undertaken in 2020 reduced this time to 32 min [19]. As mentioned above, this was still too much as the operational requirements impose a limit of no more than 20 minutes. The number of processor cores had to be reduced to less than 3000. As a result of the efforts described in this article, we are now able to compute the 24-hour SL-AV10 forecast in 13 min using 2916 processor cores. This timing allows multiple experiments for tuning this new model and fits operational requirements. Also important is the fact that the single long-range forecast (i.e., one ensemble member) with low-resolution SL-AV version now takes just 89 min instead of 111 min.

We have investigated the impact of the partial use of single precision computations on present climate simulations. It turns out that these model changes do not affect the model climate significantly.

The results described in this paper allow us to extend and accelerate the work on further model improvements.

Acknowledgements

The authors are grateful to Vassily Mizyak, Radomir Zaripov, Svetlana Travova, Vladimir Rogutov from Hydrometcenter of Russia who participate in model development and research. The numerical experiments with the SL-AV model are carried out using Cray XC40 installed at the Roshydromet Main Computer Center (MCC). Most of study (except for Section 4) was performed at Hydrometcenter of Russia and supported by the Russian Science Foundation, project 21-17-00254, <https://rscf.ru/project/21-17-00254/>.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 3.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Manual on Codes - International Codes, Volume I.2, Annex II to the WMO Technical Regulations: Part B Binary Codes, Part C Common Features to Binary and Alphanumeric Codes WMO-TD 611. Geneva, Switzerland (1994)
2. Learning HDF5. <https://portal.hdfgroup.org/display/HDF5/Learning+HDF5> (2021), accessed: 2021-10-21
3. Network common data form (NetCDF). <https://www.unidata.ucar.edu/software/netcdf/> (2021), accessed: 2021-10-21
4. Xios. XML-IO-Server. <https://portal.enes.org/models/software-tools/xios> (2021), accessed: 2021-10-21
5. Bauer, P., Stevens, B., Hazeleger, W.: A digital twin of Earth for the green transition. *Nature Climate Change* 11, 80–83 (2021). <https://doi.org/10.1038/s41558-021-00986-y>
6. Dennis, J.M., Edwards, J., Loy, R., et al.: An application-level parallel I/O library for Earth system models. *Int. J. High Perf. Comput. Appl.* 26, 43–53 (2012). <https://doi.org/10.1177/1094342011428143>
7. Hortal, M.: Aspects of the numerics of the ECMWF model. In: *Procs. of the ECWMF Seminar, September 7-11, 1998*. Reading, UK (1999)
8. Jablonowski, C., Williamson, D.L.: The Pros and Cons of Diffusion, Filters and Fixers in Atmospheric General Circulation Models, chap. 13, pp. 381–493. Springer (2011). https://doi.org/10.1007/978-3-642-11640-7_13
9. Lindberg, K., Alexeev, V.A.: A study of the spurious orographic resonance in semi-implicit semi-Lagrangian models. *Monthly Weather Review* 128(6), 1982–1989 (2000). [https://doi.org/10.1175/1520-0493\(2000\)128<1982:ASOTS0>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<1982:ASOTS0>2.0.CO;2)
10. Ritchie, H., Tanguay, M.: A comparison of spatially averaged Eulerian and semi-Lagrangian treatments of mountains. *Mon. Wea. Rev.* 124, 167–181 (1996). [https://doi.org/10.1175/1520-0493\(1996\)124<167:ACOSAE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<167:ACOSAE>2.0.CO;2)

11. Rivest, C., Staniforth, A., Robert, A.: Spurious resonant response of semi-Lagrangian discretizations to orographic forcing: Diagnosis and solution. *Monthly Weather Review* 122(2), 366–376 (1994). [https://doi.org/10.1175/1520-0493\(1994\)122<0366:SRR0SL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1994)122<0366:SRR0SL>2.0.CO;2)
12. Slingo, J., Palmer, T.: Uncertainty in weather and climate prediction. *Phil. Trans. R. Soc. A* 369, 4751–4767 (2011). <https://doi.org/10.1098/rsta.2011.0161>
13. Strohmaier, E., Dongarra, J., Simon, H., Meuer, M.: Top500 list. <https://top500.org/lists/top500/> (2021), accessed: 2021-10-21
14. Tolstykh, M., Fadeev, R., Mizyak, V.: Parallel program complex for numerical weather prediction and climate modeling. In: *CEUR Worskshop Proceedings. RuSCDays 2015 - Proceedings of the 1st Russian Conference on Supercomputing Days 2015, Moscow, Russia, September 28-29, 2015*. vol. 1482, pp. 356–367 (2015)
15. Tolstykh, M., Goyman, G., Fadeev, R., Shashkin, V.: Structure and algorithms of SL-AV atmosphere model parallel program complex. *Lobachevskii Journal of Mathematics* 39(4), 587–595 (2018). <https://doi.org/10.1134/S1995080218040145>
16. Tolstykh, M., Shashkin, V., Fadeev, R., Goyman, G.: Vorticity-divergence semi-Lagrangian global atmospheric model SL-AV20: dynamical core. *Geoscientific Model Development* 10(5), 1961–1983 (2017). <https://doi.org/10.5194/gmd-10-1961-2017>
17. Tolstykh, M., Fadeev, R., Shashkin, V., et al.: Multiscale global atmosphere model SL-AV: the results of medium-range weather forecasts. *Russ. Meteorol. Hydrol.* 43, 773–779 (2018). <https://doi.org/10.3103/S1068373918110080>
18. Tolstykh, M., Goyman, G., Fadeev, R., et al.: Development of the global multiscale atmosphere model: computational aspects. In: *Journal of Physics: Conference Series*. vol. 1740, p. 012074. IOP Publishing (2021). <https://doi.org/10.1088/1742-6596/1740/1/012074>
19. Tolstykh, M.A., Goyman, G., Fadeev, R., Shashkin, V.V.: Implementation of SL-AV global atmosphere model with 10 km horizontal resolution. In: *Supercomputing - 6th Russian Supercomputing Days, RuSCDays 2020, Moscow, Russia, September 21-22, 2020, Revised Selected Papers. Communications in Computer and Information Science*, vol. 1331, pp. 216–225. Springer (2020). https://doi.org/10.1007/978-3-030-64616-5_19
20. Tolstykh, M.A., Goyman, G., Fadeev, R., et al.: SL-AV model: Numerical weather prediction at extra-massively parallel supercomputer. In: *Supercomputing - 4th Russian Supercomputing Days, RuSCDays 2018, Moscow, Russia, September 24-25, 2018, Revised Selected Papers. Communications in Computer and Information Science*, vol. 965, pp. 379–387. Springer (2018). https://doi.org/10.1007/978-3-030-05807-4_32
21. Yang, R., Ward, M., Evans, B.: Parallel I/O in Flexible Modelling System (FMS) and Modular Ocean Model 5 (MOM5). *Geosci. Model Dev.* 13, 1885–1902 (2020). <https://doi.org/10.5194/gmd-13-1885-2020>