https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

Enhancing Single-Precision with Quasi Double-Precision: Achieving Double-Precision Accuracy in the Model for Prediction Across Scales-Atmosphere (MPAS-A) version 8.2.1

Jiayi Lai[1], Lanning Wang[1,2], Qizhong Wu[1,2], Yizhou Yang[3], and Fang Wang[4]

[1]College of Global Change and Earth System Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[2]Joint Center for Earth System Modeling and High Performance Computing, Beijing Normal University, Beijing 100875, China
[3]Zhongguancun Laboratory, Beijing 100081, China
[4]CMA Earth System Modeling and Prediction Centre (CEMC), Beijing 100081, China

Correspondence: Lanning Wang (wangln@bnu.edu.cn)

Abstract. The development of numerical models are constrained by the limitations of high performance computing (HPC). Low precision computations can significantly reduce computational costs, but inevitably introduce rounding errors, which affect computational accuracy. Quasi double-precision algorithm can compensate for rounding errors by keeping corrections, thereby achieving the low numerical precision while maintaining result accuracy. This paper applies the algorithm to the Model for Prediction Across Scales-Atmosphere (MPAS-A) and evaluate its performance across four test cases. The results demonstrate that, after reducing numerical precision to single precision (from 64 bits to 32 bits), the application of quasi double-precision algorithm can achieve results comparable to double-precision computations. The round-off error of surface pressure is reduced by 68%, 75%, 97%, 96% in cases, the memory has been reduced by almost half, while the computation increases only 2%, significantly reducing computational cost. The work substantiates both effectiveness and inexpensive computation in numerical models by using quasi double-precision algorithm.

1 Introduction

Since the advent of modern computers in the 1950s, numerical simulation-based weather and climate modeling has emerged as one of the most effective methods for exploring weather and climate systems, providing a new platform for numerical model research (Bauer et al. 2015). However, in order to achieve more accurate and precise simulation results, numerical weather and climate models are evolving towards higher resolutions and more complex physical parameterization schemes (Bauer et al. 2015). With the integration of increasingly complex modules to meet diverse requirements, numerical weather and climate models have developed rapidly, and the next generation of these models will feature unprecedented resolution and complexity (Hatfield et al. 2019). In this context, the demand for more powerful high-performance computing (HPC) systems and more efficient computational methods has become particularly urgent. As noted by Bauer et al. (2015), the computational tasks of future numerical model prediction (NMP) systems are expected to be 100 to 1000 times greater than those of current sys-

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

tems. The development of future high-performance computing is crucial for the continued advancement of numerical weather forecasting. Therefore, the design of code and the selection of algorithms must prioritize the optimization of floating-point operations and memory usage to meet this technological challenge (Hatfield et al. 2019).

Mixed precision is a critical research direction in optimizing computational resources within numerical models. By reducing the bit-width required for number representation and thereby lowering the precision of floating-point numbers, mixed precision methods enable storage and computations to be performed with fewer bits. This approach not only significantly decreases memory requirements but also substantially reduces the computational and communication costs in numerical software projects such as climate modeling. Employing lower precision numerical representations is a feasible option for reducing the computational costs of complex numerical models (Dawson et al. 2017). Consequently, the study of mixed precision techniques has emerged.

In recent years, notable advancements have been made in the application of mixed-precision computing in numerical weather and climate models. Váňa et al. (2016) investigated the implementation of mixed-precision computing in the Integrated Forecast System (IFS) prediction model. They employed double precision in certain regions while utilizing lower precision in others. This approach significantly enhanced computational efficiency by an average of 40% while maintaining acceptable error margins, thereby providing a crucial reference for subsequent researchers. Dawson et al. (2018) expanded the scope of mixed-precision methods, demonstrating their applicability to simple thermal diffusion models, provided that key state variables are stored and updated with higher precision. For more complex real-world land surface schemes, they showed that using lower precision for the majority of computations while ensuring high-precision processing of state variables could still meet the requisite accuracy standards. Concurrently, Nakano et al. (2018) conducted an in-depth study on the dynamical core of the global compressible non-hydrostatic model, particularly in the baroclinic wave tests by Jablonowski and Williamson. Nakano et al (2018) opted to use double precision for grid geometry calculations and single precision for other components. The results indicated that this strategy not only successfully simulated the growth of baroclinic waves with minimal error also reduced runtime by 46%. This study further corroborated the efficacy of mixed-precision computing in dynamical core calculations. Hatfield et al. (2019) applied mixed-precision computing to the Legendre transform in the IFS, successfully implementing half-precision computations. Remarkably, this modification reduced the computational cost to 25% of that in the double-precision reference test, significantly lowering computational overhead. This achievement underscored the substantial potential of mixed-precision computing in large-scale numerical prediction models. In the same year, Oriol Tintó et al. (2019) applied mixed-precision methods to the European ocean simulation core (NEMO). They discovered that 95.8% of the 962 variables could be computed using half precision. Additionally, in the Regional Ocean Modeling System (ROMS), all 1146 variables could be computed using single precision, with 80.7% of them even using half precision. This finding suggests that mixed-precision methods have extensive applicability in ocean modeling. Cotronei at al. (2020) converted the radiation component of the atmospheric model ECHAM to a single-precision algorithm, resulting in an approximately 40% acceleration in radiation calculations. This result indicates that applying single-precision computing in atmospheric models can significantly enhance computational efficiency while preserving computational accuracy to a reasonable extent. Paxton at al. (2022) further investigated the feasibility of reduced-precision computing. He conducted tests in the Lorenz system, shallow water approx-

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

imation over a ridge, and the simplified parameterized coarse-resolution spectral global atmospheric model (SPEEDY). The findings revealed that single precision (23 bits) sufficed for most computational needs, and in numerous cases, half precision (10 bits) could also achieve the desired results. This provides an important reference for adopting lower-precision computing in various models in the future. This year, Hugo et al. (2024) further substantiated the effectiveness of mixed-precision methods in the regional weather and climate model COSMO. He found that the differences between double-precision and single-precision simulations were minimal, typically detectable only in the initial few hours or days of the simulation. However, single-precision simulations reduced computational costs by approximately 30%.In the same year, Chen et al. (2024) applied the principle of limited iterative development to identify equations that were insensitive to precision in weather and climate modeling tests, modifying them from double precision to single precision. This optimization resulted in a reduction of the runtime of the model's hydrostatic solver, non-hydrostatic solver, and tracer transport solver by 24%, 27%, and 44%, respectively, thereby substantially enhancing computational efficiency.In summary, mixed-precision computing exhibits broad application prospects and potential advantages in numerical weather and climate modeling. By flexibly applying varying precision computing methods while ensuring predictive accuracy, it is feasible to significantly enhance computational efficiency and reduce computational costs.

When utilizing mixed-precision computation, low-precision calculations inevitably introduce rounding errors, particularly when adding numbers with significantly different magnitudes. In such scenarios, the limited precision can cause the larger number to effectively "swallow" the smaller number, thereby compromising the accuracy of the result. For instance, consider the variables $A = 0.7315 * 10^3$ (a large number) and $B = 0.4506 * 10^-5$ (a small number). If the precision of the result is reduced to 4 significant digits, the outcome will be $0.7315 * 10^3$, with the large number effectively overshadowing the small one. This phenomenon is especially pertinent in numerical modeling, where the introduction of biases into fundamental fields often necessitates the addition of large and small numbers, inherently causing rounding errors. These errors can accumulate over successive computations, leading to a degradation in model accuracy or even complete failure. Therefore, addressing the rounding errors induced by low-precision computations is a critical area for further research.

In 1951, Gill (1951) proposed a fourth-order, four-step explicit Runge-Kutta method aimed at correcting rounding errors during computation. This method constructs auxiliary variables at each step to compensate for the rounding errors generated, thereby further refining the results to achieve higher precision. However, this method is not applicable to other forms of numerical solutions. In addition to this, compensated summation methods can enhance the accuracy of summation by utilizing the floating-point precision supported by lower-level hardware (Higham 1996). These methods rely on recursive summation and incorporate correction terms to reduce rounding errors. In 1965, Møller (1965) and Kahan (1965) respectively proposed the quasi double-precision method and the Kahan method. The primary idea behind both methods is to make slight adjustments to the total sum to avoid the precision loss caused by adding a small, precise value to a much larger one in floating-point addition. The quasi double-precision method has been validated in solving ordinary differential equations using the fourth-order Runge-Kutta method (Møller 1965), where the error after precision reduction is essentially minimized to zero.

Currently, methods for compensating rounding errors are primarily employed in the step-by-step integration of ordinary differential equations (Thompson et al. 1970; Tomonori et al. 1995; Dmitruk et al. 2023). However, their validation in numerical

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

```
1 :    u ≔ initial u;
2 :    c ≔ 0;
3 : L: v ≔ (< evalution of v >) + c;
4 :    s ≔ u + v;
5 :    c ≔ (v − (s − u)) + (u − (s − (s − u)))
6 :    u ≔ s
7 :    go to L:
```

Figure 1. The quasi double-precision algorithm in case of a step-by-step integration.

models remains uncertain. Considering the broader applicability of the quasi double-precision method, which can be utilized for recursive summation in any format, and its superior performance in high-performance computing environments compared to the Kahan method (Kahan 1965), this study aims to implement the quasi double-precision method in MPAS-A model. By addressing the sum of large and small numbers during the time integration process, the single-precision version improved with the quasi double-precision algorithm achieves basically consistent results comparable to those of double precision. The structure of this paper is as follows: Section 2 introduces the quasi double-precision algorithm, the MPAS model, application of quasi double-precision algorithm in MPAS-A, and the experimental design. Section 3 provides case study in MPAS. Section 4 presents conclusions and discussion of the experiments.

2 Methodology, model and experiments

2.1 Quasi double-precision algorithm

The quasi double-precision algorithm, proposed by Møller et al. (1965), aims to address the precision loss that occurs when adding small values to large values in floating-point arithmetic. This precision loss typically arises from coarse truncation operations. The quasi double-precision algorithm reduces round-off errors by keeping corrections. Primarily applied in the step-by-step integration of ordinary differential equations, the algorithm significantly corrects rounding errors in sum, particularly in computers where truncation operations are not followed by proper rounding.

A brief introduction to the algorithm is as follows, with a detailed derivation available follows Møller et al. (1965). Define the floating-point numbers u, v, s, and c, where in each step of the time integration, s=u+v. By introducing a correction variable c before computing sum (s) of u and v in each step, the final s is adjusted to reduce rounding errors. This algorithm is illustrated in Figure 1.

The process can be viewed as v being continuously accumulated onto u; however, in numerical models' computations, it is impossible to ensure that u is always greater than v. To enhance the precision of the correction process, a precondition of magnitude comparison is added to the algorithm, as shown in Figure 2.

It is important to note that the applicability of the quasi double-precision algorithm has been thoroughly analyzed (Møller et al. 1965), cases of inapplicability are exceedingly rare. Considering the numerous sum algorithm and integration involved

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

```
1 :   c ≔ 0;
2 :L: if u > v then;
3 :       u ≔ initial u;
4 :       v ≔ (< evalution of v >) + c;
5 :       s ≔ u + v;
6 :       c ≔ (v − (s − u)) + (u − (s − (s − u)))
7 :     else
8 :       v ≔ initial v;
9 :       u ≔ (< evalution of u >) + c;
10 :      s ≔ u + v;
11 :      c ≔ (u − (s − v)) + (v − (s − (s − v)))
12 : go to L:
```

Figure 2. The quasi double-precision algorithm adding a precondition of magnitude.

in numerical models, even if a few inapplicable instances occur, their impact on the overall result is negligible. Therefore, in practical applications, these infrequent cases are typically not considered.

2.2 MPAS-A

MPAS-A is a compressible, non-hydrostatic atmospheric numerical model developed by NCAR. It employs an unstructured centroidal Voronoi grid (mesh or tessellation) and a staggered C-grid for state variables as the basis for horizontal discretization in the fluid flow solver. MPAS-A consists of two main components: the model, which includes atmospheric dynamics and physics, and the initialization component, which generates initial conditions for the atmosphere and land surface, updates for sea surface temperature and sea ice, and lateral boundary conditions. Both components (model and initialization) are integral constructs within the MPAS software framework and utilize the same drivers and software infrastructure.

The dynamical core of MPAS-A solves the fully compressible, nonhydrostatic equations of motion (Skamarock et al. 2012). These fully compressible nonhydrostatic equations are transformed based on geometric height vertical coordinates. The solver employs a split-explicit time integration scheme as described by Klemp et al. (2007) . The time integration scheme utilizes Runge-Kutta methods with a large time step, while for the acoustic modes, a smaller time step forward-backward method is employed (Wicker and Skamarock 2002).

MPAS-A currently offers two time integration schemes: a second-order Runge-Kutta method and a third-order Runge-Kutta method, which can be configured through namelist parameters. The default setting is the second-order Runge-Kutta method, and this study uses the default setting for experiments. The numerical schemes used in MPAS-A are very similar to those used in the Advanced Research WRF (ARW) model. The main differences are that the ARW model uses a rectangular grid and hydrostatic pressure (mass) vertical coordinates. Additionally, MPAS employs a vector-invariant form of the horizontal momentum equations and a more general version of the WRF transport scheme as given by Skamarock and Gassmann (2011).

The MPAS-A solver utilizes the physics suite from the Advanced Research WRF (ARW) model, with a particular focus on the physical configurations used in the ARW Nested Regional Climate Model (WRF-NRCM) applications and tropical cyclone prediction experiments.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

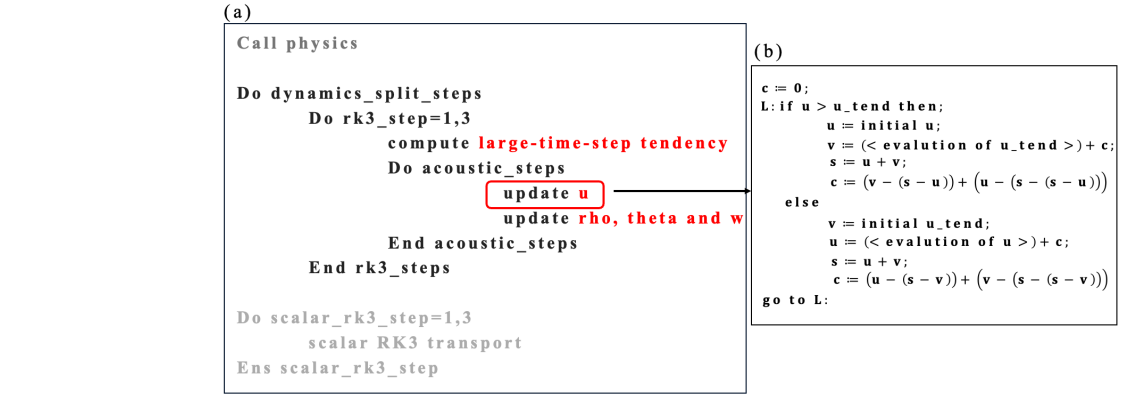

Figure 3. The application of quasi double-precision algorithm in MPAS-A. (a): The application framework of MPAS-A. (b): An example of adding quasi double-precision algorithm in MPAS-A.

In this study, the version 8.2.1 of MPAS-A was used for the following reasons: (1) This research primarily focuses on the accumulation of variables in time integration, specifically the accumulation of time integration variables within the dynamical core. Version 8.2.1 supports the option to close physical processes during model construction, preventing the influence of physical processes on the results of the dynamical core. Therefore, this version was chosen. It should be noted that all cases in this study have closed physical processes. (2) This version supports single-precision operations, reducing the repetitive work of code modification.

2.3 Application of quasi double-precision algorithm in MPAS-A

Quasi double-precision has been validated in the time integration process of differential equations (Møller et al. 1965). The primary objective of this section is to demonstrate how to apply the quasi double-precision algorithm to the time integration process in MPAS-A.

By analyzing the application framework of MPAS-A (Figs. 3a), it can be observed that the modules containing the time integration schemes are: (1) the gravity wave and acoustic wave calculation module, and (2) the scalar transport process. This study focuses on dynamic core, involving the gravity wave and acoustic wave, so we close the scalar transport in all cases. In the gravity wave and acoustic wave, the core variables calculated through time integration are horizontal momentum at cell edge (u) , dry air density (rho), potential temperature (theta) and vertical velocity at vertical cell faces (w).

These variables involve two time integration processes: In acoustic_steps, the small quantity is calculated, and is then added to big quantity in large-time-step-tendency, and the two parts both use quasi double-precision algorithms respectively. Figure 3 shows an application method adding u_tend to u and as an example (Figs. 3b) , other variables are the same.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

2.4 Experimental design and configuration

This study aims to investigate whether the quasi double-precision algorithm can effectively compensate for the rounding errors that caused by reduced numerical precision. Setting the double-precision version (DBL) as the benchmark experiment. Two control experiments are also established: the first control experiment uses the single-precision (SGL), and the second control experiment applies the quasi double-precision algorithm to the single-precision (QDP). By comparing the root mean square error (RMSE) between these two control experiments and the benchmark experiment, this study evaluates the effectiveness of the quasi double-precision algorithm in reducing rounding errors.

To assess the application effect of the quasi double-precision algorithm, this study employs four test cases, including two idealized cases (Jablonowski and Williamson baroclinic wave and super-cell) and two real cases (with initial conditions generated using GFS data at 2014-09-10_00) using two different resolutions. To prevent the influence of other factors, the basic parameters of all cases are kept consistent, including the Number of acoustic steps per full RK step, config dynamics split steps, and config number of sub steps (integer), among others.

3 Results and analysis

In this section, we show results across four cases, include two ideal scenarios: Jablonowski and Williamson baroclinic wave and super-cell, as well as two real case (with initial conditions generated using GFS data) using two different resolutions. By using the RMSE for quantitative comparison, the differences between the benchmark and control experiments are used to evaluate the effectiveness of the quasi double-precision algorithm in reducing round-off error.

3.1 Jablonowski and Williamson baroclinic wave

This case is a deterministic initial-value test case for dry dynamical cores of atmospheric general-circulation models(Jablonowski and Williamson 2006), assesses the evolution of an idealized baroclinic wave in the northern hemisphere. The initial zonal state is quasi-realistic and entirely defined by analytical expressions, which are steady-state solutions of the adiabatic, inviscid primitive equations in a pressure-based vertical coordinate system (Jablonowski and Williamson 2006). The configuration follows the specifications published on the MPAS website, with a time step of 450 seconds, 26 vertical levels, resolution of 120 km × 120 km, and an integration period of 15 days.

The round-off error begins to appear at the tenth day. Starting from the tenth day, the round-off error of kinetic energy and surface pressure caused by SGL can be reduced by using Quasi double-precision (Figs. 4a, 4b). Unlike SGL, where the error increases rapidly after more than 10 days, QDP has a very small error compared to double precision. Therefore, it can be considered that QDP can be used to replace double precision in medium range weather forecast.

It can be found that SGL can increase the round-off error in all regions(Figs. 5a), especially in high-latitude regions, such as Southern Ocean westerly belt, its high wind speed increase round-off error caused by SGL, but instability caused by high wind speeds is more important. Surprisingly, the round-off error can be reduced significantly in QDP(Figs. 5b), it means that

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

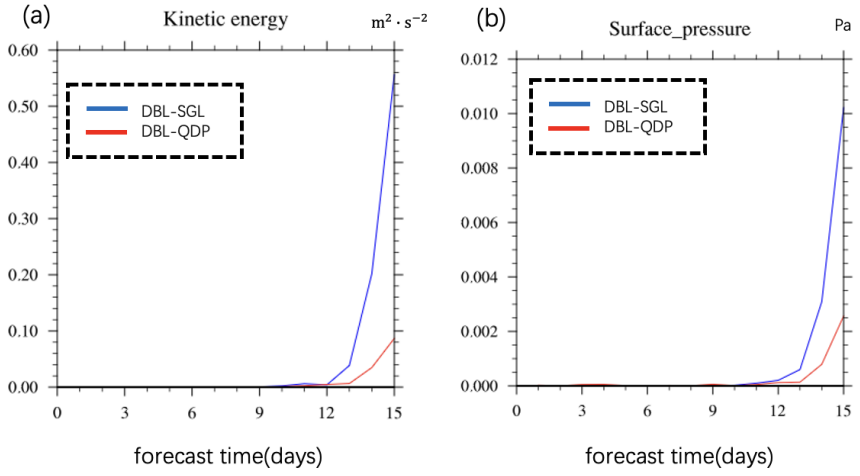

Figure 4. The time evolution of difference between DBL and SGL, as well as difference between DBL and QDP of (a) kinetic energy, (b) surface pressure in case of Jablonowski and Williamson baroclinic wave.

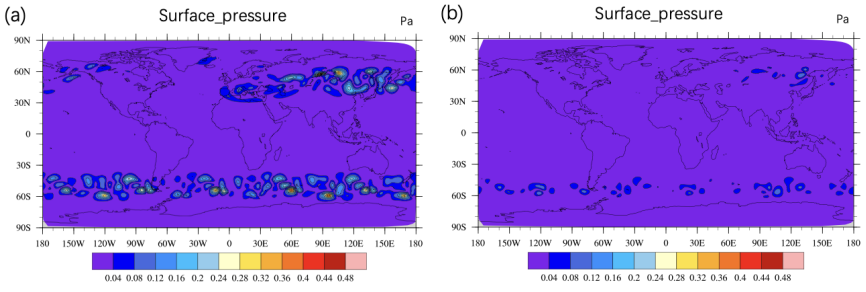

Figure 5. Spatial distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulations, (b) QDP simulations in case of Jablonowski and Williamson baroclinic wave.

QDP can improve stability compared to SGL. It should be emphasized that, this does not mean that the higher the wind speed, the better the improvement effect, but rather that the improvement effect is more pronounced in areas with larger errors. The spatial RMSE of surface pressure between DBL and SGL is $3.42 * 10^-2$ Pa, as well as $1.09 * 10^-2$ Pa between DBL and QDP, the error reduced by 68%.

3.2 Super-cell case

The test case (Klemp et al. 2015) is on a reduced-radius sphere, can evaluate the behavior of nonhydrostatic processes in nonhydrostatic global atmospheric dynamical cores provided the simulated cases exhibit good agreement with corresponding

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

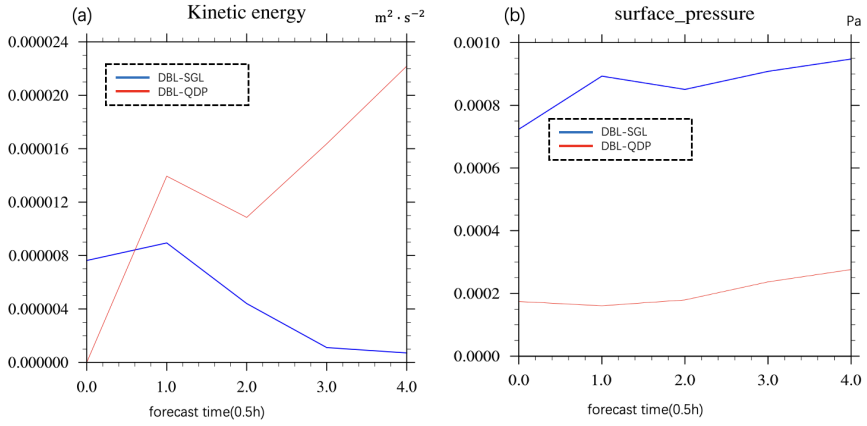

Figure 6. The time evolution of difference between DBL and SGL, as well as difference between DBL and QDP of (a) kinetic energy, (b) surface pressure in case of super-cell.

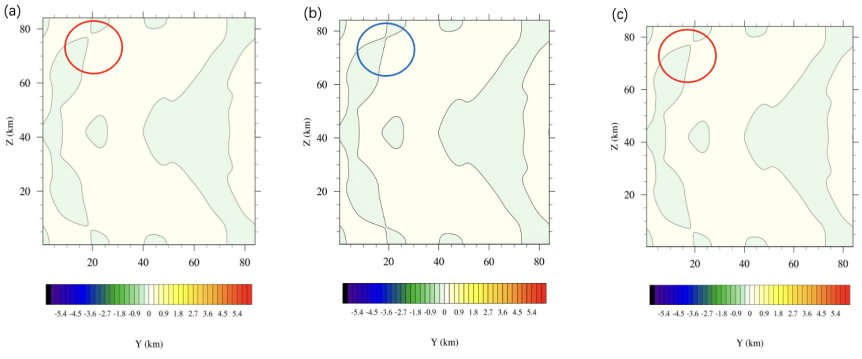

Figure 7. Perturbation theta in super-cell development at 5400s in the (a) DBL simulation, (b) SGL simulation and (c) QDP simulation, the circle represent the most clear error.

flows in a Cartesian geometry, and for which there are known solutions. The settings include a time step of 3 seconds, 40 vertical levels, resolution of 84 km × 84 km, and an integration period of 2 hours.

In this case, the reduction of kinetic energy in round-off error is not significant in QDP (Figs. 6a), except for the initial time, all others showed larger errors than SGL. But the errors of both are very small and can be ignored. For surface_pressure (Figs. 6b), the round-off error caused by SGL can obtain effective improvement in QDP. This improvement exists throughout the entire integration period.

Figure 7 shows the spatial distribution of perturbation theta, an important variable in numerical models, when reducing the numerical precision from double (Figs. 7a) to single (Figs. 7b), it displays differences, it indicates a significant increase in

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

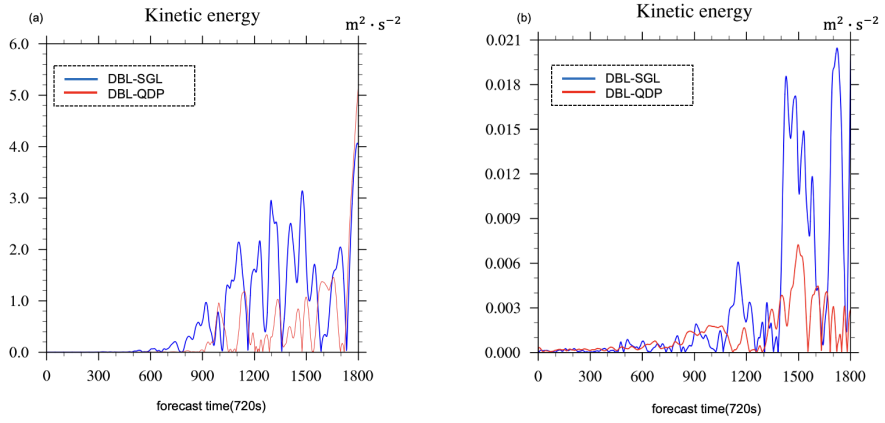

Figure 8. The time evolution of difference of kinetic energy between DBL and SGL, as well as difference between DBL and QDP, with resolution of (a) 240 km × 240 km, (b) 120 km × 120 km.

round-off error. In QDP, this difference can be compensated(Figs. 7c). The spatial RMSE of surface pressure between DBL and SGL is $8.95 * 10^{-}4$ Pa, as well as $2.19 * 10^{-}4$ Pa between DBL and QDP, the error reduced by 75%.

3.3 Real data cases

In this section, we will show the results from two cases using different resolution.The settings include a time step of 720 seconds, 55 vertical levels, resolution of 240 km × 240 km and 120 km × 120 km, and an integration period of 15 days. (Except for the resolution, all other configurations are exactly the same)

At the initial stages of the integration process(Figs. 8), both the SGL and QDP have minimal rounding errors in different resolutions. Differences in error begin to emerge after 500 steps. QDP can reduce errors generated by SGL within certain integration time, although not consistently throughout. Overall, QDP demonstrates an ability to reduce rounding errors caused by SGL. The case with resolution of 240 km × 240 km (Figs. 8a) show the larger error than 120 km × 120 km (Figs. 8b), and the error can be reduced in QDP caused by SGL.

Figure 9 and 11 show spatial distributions of surface pressure with different resolution, 240 km × 240 km (Fig. 9) and 120 km × 120 km (Fig. 11). The error has reduced throughout the all region, and the improvement effect is very obvious. From a spatial perspective, the case of SGL with resolution of 240 km × 240 km (Figs. 9a) show the larger error than 120 km × 120 km (Figs. 11a), and the errors both can be reduced by QDP (Figs. 9a and 11a). The spatial RMSE of surface pressure with 240 km × 240 km between DBL and SGL is $6.68 * 10^{-}2$ Pa, as well as $2.25 * 10^{-}3$ Pa between DBL and QDP, the error reduced by 97%. The spatial RMSE of with 120 km × 120 km between DBL and SGL is $6.33 * 10^{-}2$ Pa, as well as $2.25 * 10^{-}3$ Pa between DBL and QDP, the error reduced by 96%.

Figure 10 and 12 show spatial distributions of 500 hPa height with different resolution, 240 km × 240 km (Fig. 10) and 120 km × 120 km (Fig. 12), The error improvement effect is consistent with surface pressure. The spatial RMSE of 500 hPa height

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

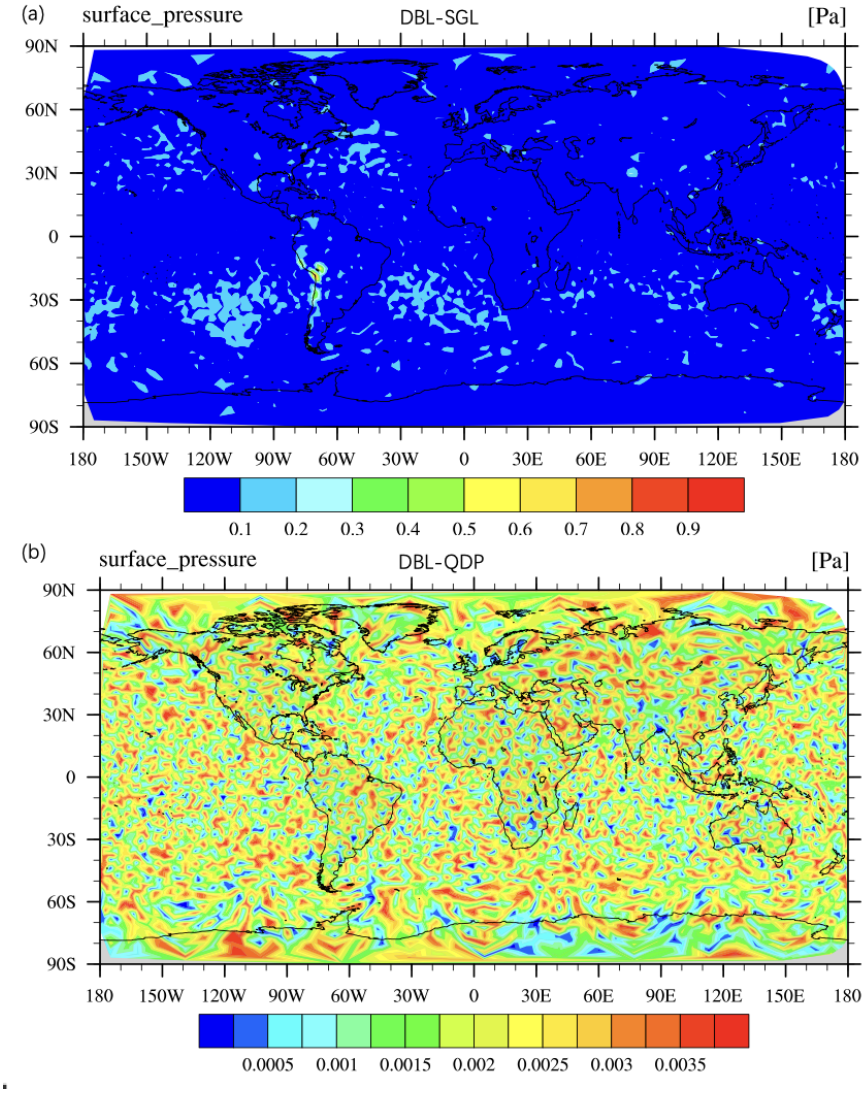

Figure 9. Spatial distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 240 km × 240 km). The RMSE of surface pressure between DBL and (a) SGL simulation is $6.68 * 10^{-}2$ Pa, (b) QDP simulation is $2.25 * 10^{-}3$ Pa.

with 240 km × 240 km between DBL and SGL is $2.80 * 10^{-}1$ m, as well as $1.40 * 10^{-}1$ m between DBL and QDP, the error reduced by 50%. The spatial RMSE of with 120 km × 120 km between DBL and SGL is $4.35 * 10^{-}3$ Pa, as well as $1.90 * 10^{-}3$ Pa between DBL and QDP, the error reduced by 56%.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

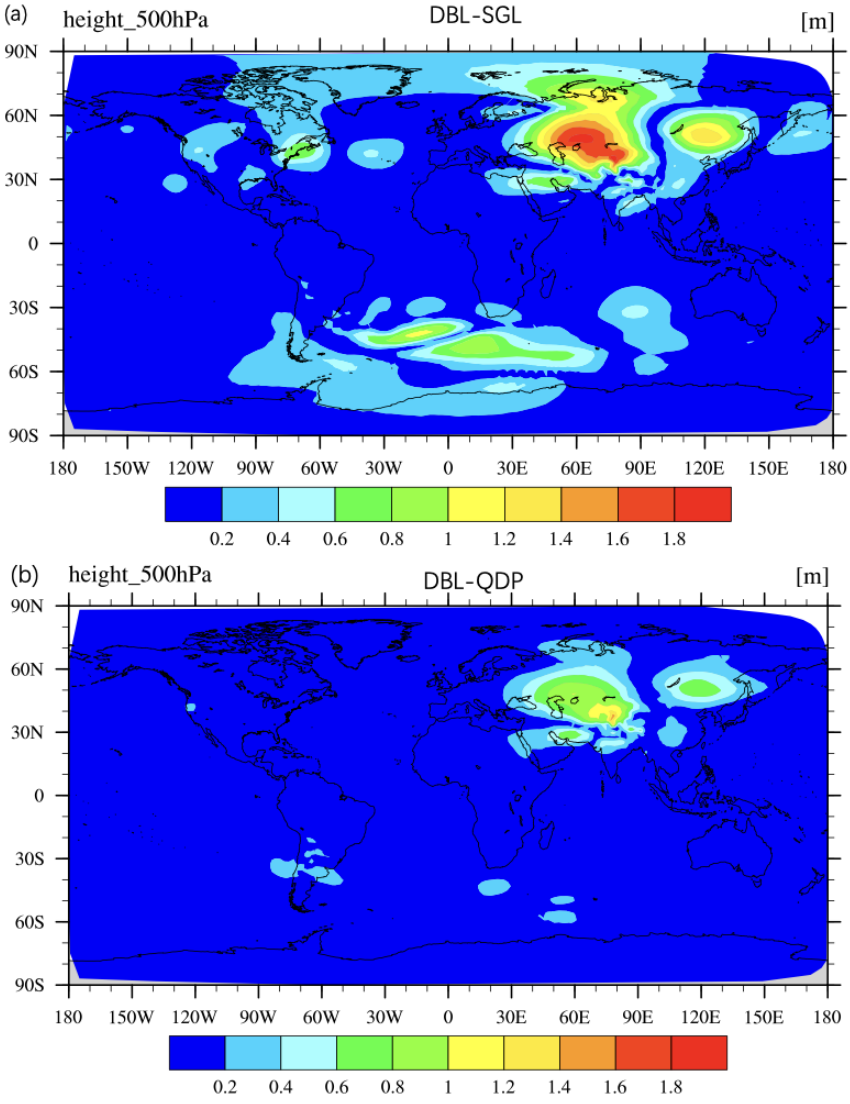

Figure 10. Spatial distributions of averaged (1-15days) difference of 500 hPa height (units: m) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 240 km × 240 km). The RMSE of 500 hPa height between DBL and (a) SGL simulation is $2.80 * 10^{-}1$ m , (b) QDP simulation is $1.40 * 10^{-}1$ m.

4 Conclusions and discussion

In this study, we applied the quasi double-precision algorithm to MPAS-A. we discovered that, the quasi double-precision algorithm can effectively reduce the errors introduced by using low precision through the iterative process of time integration. In different cases (including idealized and real data cases), the number of iterations of the correction variables varies are

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

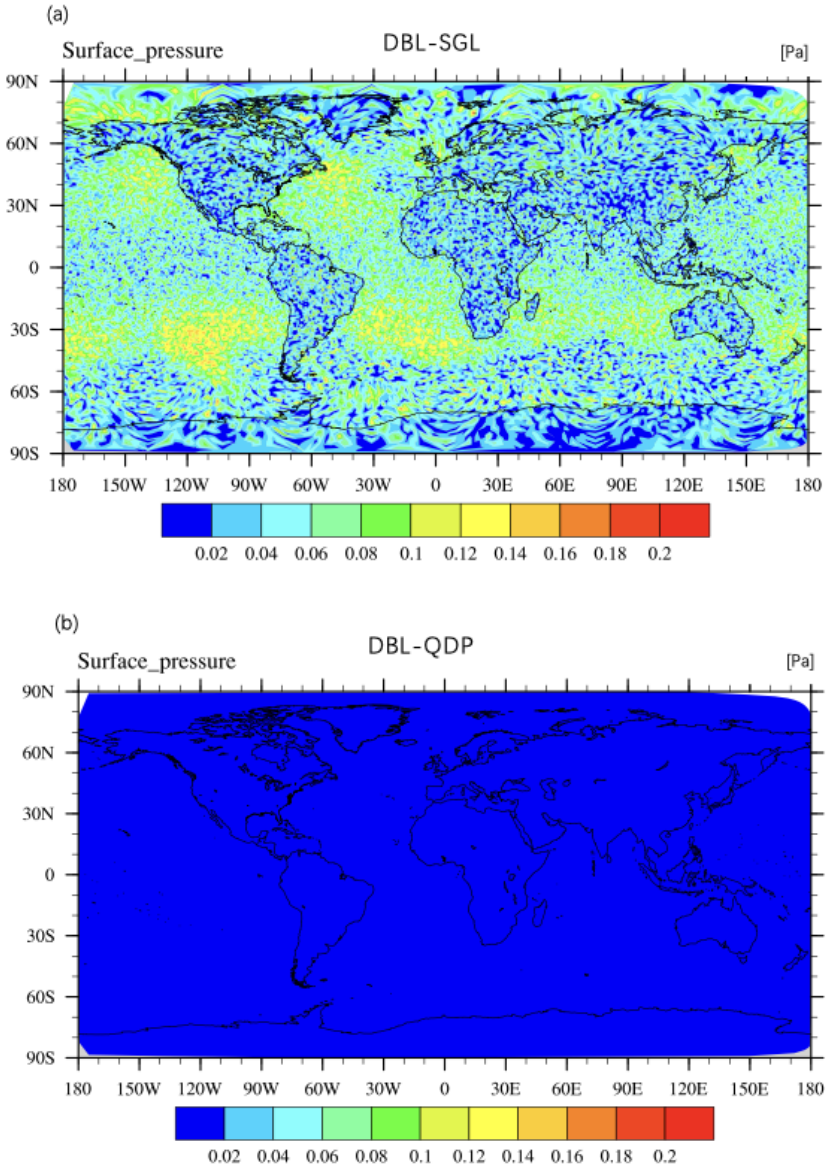

Figure 11. distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 120 km × 120 km). The RMSE of surface pressure between DBL and (a) SGL simulation is $6.33 * 10^{-}2$ Pa, (b) QDP simulation is $2.25 * 10^{-}3$ Pa.

different, leading to differences in error improvement. The error of surface pressure of 4 cases are reduced by 68%, 75%, 97%, 96% (see Section 3). Overall, QDP using quasi double-precision algorithm demonstrates higher accuracy than the SGL, suggesting the potential for applying quasi double-precision algorithm in numerical models.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

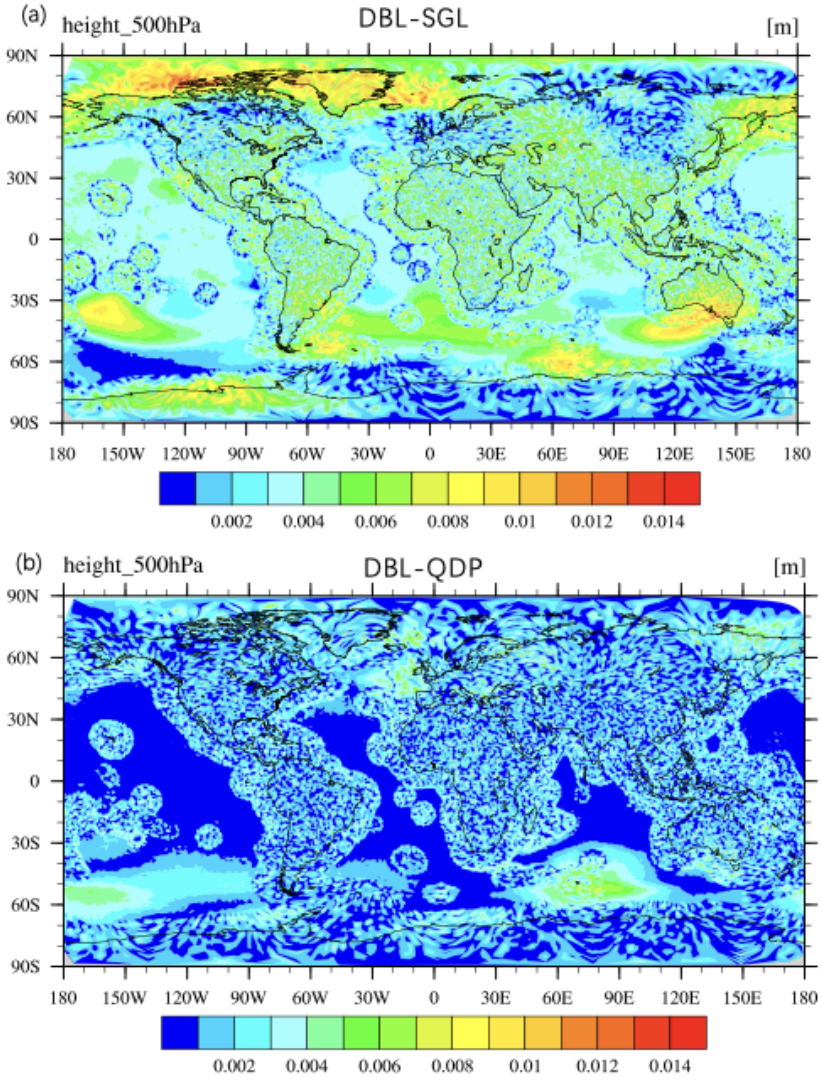

Figure 12. Spatial distributions of averaged (1-15days) difference of 500 hPa height (units: m) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 120 km × 120 km). The RMSE of 500 hPa height between DBL and (a) SGL simulation is $4.35 * 10^{-}3$ m , (b) QDP simulation is $1.90 * 10^{-}3$ m.

When applied the quasi double-precision algorithm in MPAS-A, we achieved to reduce all double precision to single precision, although increased 3 local variables and 1 array in every time-integration variables, these have little impact on the overall memory reduction. In general, memory has been reduced by almost half, while the computation increases only 2%.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

Nevertheless, there are some limitations to the application of quasi double-precision algorithm. Firstly, the algorithm relies on iterative process of time integration, its effectiveness is dependent on the number of time iterations, the more iterations, the better the error compensation. Secondly, although the quasi double-precision algorithm partially reduces the round-off errors of low-precision calculations, it still shows error compared to the double-precision version, making it less suitable for experiments requiring high precision. Additionally, applying quasi double-precision algorithm must bring other variables, increasing the complexity to a certain degree.

Currently, the quasi double-precision algorithm is only applied in the time integration scheme in dynamic core of MPAS-A model, without considering tracer transport, but the process is more sensitive for the precision. Future research will attempt to apply the quasi double-precision algorithm to this part.

Currently, we have applied quasi double-precision algorithm to ideal and real data tests with low and medium resolution, the impact on high resolution has not been studied yet. On the other hand, the tracer is also a part of the atmosphere dynamic core, which describes the transport of tracer and may be more sensitive to accuracy. In the future, we will apply quasi double-precision algorithm to the tracer to analyze and study its sensitivity.

Code and data availability. Model code and plotting data related to this manuscript is available at: https://doi.org/10.5281/zenodo.13765421.

Author contributions. JYL and LNW developed the code of applying quasi double-precision algorithm to MPAS-A and design structure of the manuscript. JYL carried out the simulations and analyzed the results with help from LNW, YZY, FW, QZW and HQC. All authors gave comments and contributed to the development of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors would like to thank the administrator of Beijing Normal University High Performance Computing for providing the high-performance computing (HPC) environment and technical support.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

References

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525(7567):47-55, doi:10.1038/nature14956, 2015.

Cotronei, A., Slawig, T.: Single-precision arithmetic in ECHAM radiation reduces runtime and energy consumption, Geoscientific Model Development, 13(6):2783-2804, dio:10.5194/gmd-13-2783-2020, 2020.

Chen, Siyuan., Zhang, Yi., Wang, Yiming., Liu, Zhang., Li, Xiaohan., and Xue, Wei.:Mixed-Precision Computing in the GRIST Dynamical Core for Weather and Climate Modelling: Geosci. Model Dev., doi: https://doi.org/10.5194/gmd-2024-68, 2024.

Dawson, A., and Peter D, Düben.: rpe v5: An emulator for reduced floating-point precision in largenumerical simulations, Geoscientific Model Development Discussions, 10(6):1-16, doi:10.5194/gmd-2016-247, 2017.

Dawson, A., Peter D, Düben., Macleod D, A., and Palmer, TN.: Reliable low precision simulations in land surface models, Climate dynamics.DOI:10.1007/S00382-017-4034-X, 2018.

Dmitruk, B., Przemysaw, Stpiczyński.:Improving accuracy of summation using parallel vectorized Kahan's and Gill-Mller algorithms, Concurrency and Computation: Practice and Experience, doi:10.1002/cpe.7763, 2023.

Dmitruk, B., Przemysaw, Stpiczyński.: Improving accuracy of summation using parallel vectorized Kahan's and Gill-Mller algorithms, Concurrency and Computation: Practice and Experience, doi:10.1002/cpe.7763, 2023.

Gill, S.: A process for the step-by-step integration of differential equations in an automatic digital computing machine, Proc. Cambridge Philos. Soc. 47, 96-108, 1951.

Hatfield, S., Peter D, Düben., Palmer, T., and Chantry, M.: Accelerating High-Resolution Weather Models with Deep-Learning Hardware, Platform for Advanced Scientific Computing, doi:10.1145/3324989.3325711, 2019.

Higham, N.: Accuracy and Stability of Numerical Algorithms. SIAM; 1996.

Hugo, Banderier., Christian, Zeman., David, Leutwyler., Stefan, Rüdisühli., Christoph, Schär.: Reduced floating-point precision in regional climate simulations: an ensemble-based statistical verification: Geosci. Model Dev., 17, 5573–5586, 2024 doi: https://doi.org/10.5194/gmd-17-5573-2024, 2024.

Kahan, W.: Pracniques: further remarks on reducing truncation errors. ACM, 1965.

Møller, O.: Quasi double-precision in floating point addition, Bit Numerical Mathematics, 5(1):37-50, doi:10.1007/BF01975722, 1965.

Nakano, M., Yashiro, H., Kodama, C., and Tomita, H.: Single Precision in the Dynamical Core of a Nonhydrostatic Global Atmospheric Model: Evaluation Using a Baroclinic Wave Test Case, Monthly Weather Review, MWR-D-17-0257.1, dio:10.1175/MWR-D-17-0257.1, 2018.

Oriol Tintó, Prims., Acosta, M. C., Moore, A. M., Castrillo, M., and Doblas-Reyes, F. J.: How to use mixed precision in ocean models: exploring a potential reduction of numerical precision in NEMO 4.0 and ROMS 3.6, Geoscientific Model Development, 12(7):3135-3148, dio:10.5194/gmd-12-3135-2019, 2019.

Paxton, E. A. D. A. M., CHANTRY, M. A. T. T. H. E. W., KLOWER, M. I. L. A. N. SAFFIN., and L. E. O. PALMER, T. I. M.: Climate Modeling in Low Precision: Effects of Both Deterministic and Stochastic Rounding, Journal of Climate, 35(4):1215-1229, 2022.

S, Gill.: A process for the step-by-step integration of differential equations in an automatic digital computing machine, Mathematical Proceedings of the Cambridge Philosophical Society, doi:10.1017/s0305004100026414, 2008.

https://doi.org/10.5194/egusphere-2024-2986
Preprint. Discussion started: 27 September 2024

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S. H., and Ringler, T. D.: A Multiscale Nonhydrostatic Atmospheric Model Using Centroidal Voronoi Tesselations and C-Grid Staggering, Monthly Weather Review, 240(9):3090-3105, doi:10.1175/MWR-D-11-00215.1, 2011.

Thompson., Robert, J.:Improving round-off in Runge-Kutta computations with Gill's method, Communications of the Acm, 13(12):739-740.DOI:10.1145/362814.362823, 1970.

Tomonori, Kouya., Hideko, Nagasaka.: On the correction method of round-off errors in the Yang's Runge-Kutta method, The Japan Society for Industrial and Applied Methematics, 1995.

Váňa, F., Düben, P., Lang, S., Palmer, T., Leutbecher, M., Salmond, D., and Carver, G.: Single Precision in Weather Forecasting Models: An Evaluation with the IFS, Monthly Weather Review, 145, 495-502, doi:10.1175/MWR-D- 16-0228.1, 2016.